IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 May 2023, accepted 13 June 2023, date of publication 20 June 2023, date of current version 28 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3287920

RESEARCH ARTICLE

Efficient Propagation Method for Angularly Consistent 4D Light Field Disparity Maps

MARYAM HAMAD, (Graduate Student Member, IEEE), CAROLINE CONTI, (Member, IEEE), PAULO NUNES, (Member, IEEE), AND LUÍS DUCLA SOARES, (Senior Member, IEEE)

Instituto de Telecomunicações, Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon 1649-026, Portugal

Corresponding author: Maryam Hamad (maryam.hamad@lx.it.pt)

This work was supported by the Fundação para a Ciência e Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) through the National Funds under Project UIDB/50008/2020 and Project PTDC/EEI-COM/7096/2020.

ABSTRACT Light Field (LF) imaging, since it conveys both spatial and angular scene information, can facilitate computer vision tasks such as depth/disparity estimation. Although disparity maps can be estimated for all LF views, most existing methods merely estimate depth/disparity for the central view and do not adequately deal with other LF views. However, having depth/disparity maps for all LF views can be useful for enhancing immersive multimedia applications, such as 3D reconstruction and LF editing. To overcome this limitation, in this paper, an efficient and occlusion-aware disparity propagation method is proposed. The proposed method generates disparity maps for all LF views given a single disparity map for one reference view (e.g., the central view). The disparity map for the reference view is propagated first into the four corner views to ensure angular consistency. Afterwards, an off-the-shelf existing disparity estimation model is used to fill any remaining holes in the corner views. Finally, disparity maps for the remaining views are recursively generated through a fast propagation step, which is followed by a final refinement step to regularize the generated disparity maps. The proposed method not only generates disparity maps for all LF views but also handles occlusions and ensures angular consistency. Experimental results on synthetic and real LF datasets with different disparity ranges, using several accuracy and angular consistency metrics, show outperforming or competitive results compared to the benchmark methods with a significant complexity reduction.

INDEX TERMS Light field disparity estimation, angular consistency, fast disparity propagation, deep learning.

I. INTRODUCTION

Light Field (LF) imaging has attracted increasing attention from researchers due to its ability to capture not only light intensity but also ray directions [1], [2], [3]. 4D LFs can be represented as an array of views (a.k.a. sub-aperture images) $I(x, y, u, v)$, where (x, y) are the spatial coordinates and (u, v) are the angular coordinates of each view. By fixing one angular and one spatial coordinate, an Epipolar Plane Image (EPI) (i.e., the unique 2D spatio-angular LF slice typically containing a regular structure with several oriented lines [1], [4]) can be obtained, as illustrated in Fig. 1. By exploiting the rich information captured by LFs and the possible LF representations, new capabilities are enabled, such as post-capture refocusing and depth

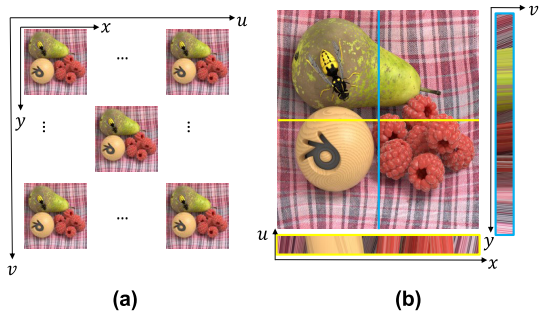

FIGURE 1. Example of LF representations. a) 4D LF represented as an array of views; b) Horizontal and vertical EPIs.

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

estimation. Additionally, disparity maps can be estimated from LFs to represent the displacement of corresponding pixels in several LF views, which is inversely proportional to the depth [4]. As humans, depth/disparity information estimated by our brains is exploited to efficiently process the surrounding world. Similarly, machines can benefit from depth/disparity information to enhance the processing of captured images. Therefore, many interesting applications for 2D content rely heavily on the use of depth maps (e.g., obtained from sensors) as an additional feature, besides the visual appearance, to apply efficient processing and editing.

In the case of 4D LF applications, besides achieving accurate editing, ensuring angular consistency is also essential. This is especially important when navigating between LF views using virtual reality headsets or LF displays. Therefore, generating angularly consistent disparity maps for all LF views has become a task of growing interest to guide several computer vision applications, such as LF segmentation, view synthesis, 3D scene reconstruction, and augmented/mixed reality [3], [5], [6], [7].

Several LF disparity estimation methods have already been proposed in the literature, e.g., [8], [9], [10], [11], [12], [13], [14], [15], [16], and [17], as briefly reviewed in Section II. Most existing LF disparity estimation methods estimate disparity maps only for the central view. However, having disparity maps for all LF views can be useful for enhancing several applications, such as 3D reconstruction and LF editing. The few methods that consider estimating disparity for all LF views, e.g., [10], [16], and [17], either are not adequately considering consistency across LF views, are computationally complex, or are only suitable for densely sampled LFs.

In this context, the main contribution of this paper is an efficient disparity propagation method that generates angularly consistent disparity maps for all LF views, which works for both densely and sparsely sampled LFs. The proposed method exploits the correlations across LF views and propagates a given disparity map from only one reference view into all LF views in an occlusion-aware manner, while also ensuring angular consistency. The proposed propagation method starts by propagating the reference view disparity map into the corner views and assigning disparity values to all their pixels. Afterwards, disparity values are recursively propagated to the remaining LF views with the guidance of the reference and corner views in both horizontal and vertical angular directions. Finally, a last refinement step is included to smooth the disparity maps. Experiments using different accuracy, consistency and complexity metrics show outperforming or competitive results when compared to the existing methods, while reducing the computational complexity.

The remainder of this paper is organized as follows. Section II briefly reviews the related work on 4D LF disparity estimation, Section III describes the proposed method in detail and Section IV evaluates its performance through a series of experimental results. Finally, Section V concludes the paper with some final remarks and directions for future work.

II. RELATED WORK

In recent decades, several 4D LF disparity estimation methods have been proposed. Existing methods can be classified as either classical or deep learning-based 4D LF methods depending on the used disparity estimation approach:

A. CLASSICAL 4D LF DISPARITY ESTIMATION METHODS

Classical 4D LF disparity estimation methods exploit different LF representations and analyze the geometry to estimate disparity information using manually designed features. This type of methods can be further classified into three categories, according to the used LF representation:

- **EPI-based methods:** the methods in this category rely heavily on the EPI regular structure. In the EPI space, a 3D point is represented by a line whose slope is inversely proportional to its disparity value [1], [4]. Wanner and Goldluecke [8] analyzed EPIs using structure tensors to locally estimate disparity values. Zhang et al. [9] proposed a spinning parallelogram operator for depth estimation on EPI space. Khan et al. [10] proposed a disparity estimation method to compute disparity maps for all LF views by detecting EPI edges and diffusing them spatially within the central view and then propagating the central view into all LF views. EPI-based methods typically achieve high estimation accuracy, but only for densely sampled LFs.
- **Sub-aperture image-based methods:** the methods in this category rely on matching corresponding pixels between LF views, i.e., stereo matching, using a robust patch-based block-matching approach. A cost volume is usually constructed to measure the similarity and angular consistency between LF views. Jeon et al. [11] proposed a disparity estimation method by applying the phase shift theorem. Huang et al. [12] proposed an empirical Bayesian framework for computing LF disparity for both dense and sparse LFs. While matching corresponding pixels to estimate disparity maps is widely used, in dense LFs with a quite narrow baseline, sub-aperture image matching can lead to poor accuracy and occlusions can cause impossible correspondences [2].
- **Focal stack-based methods:** the methods in this category produce a focal stack from LFs and rely on defocus cues to estimate the disparity. They assume that in-focus points are projected at the same spatial location in the different views [13], [14], [15]. Lee and Park [15] proposed a unified model for depth estimation by combining focus, defocus and matching corresponding pixels. The methods that rely on LF focal stack are robust to occlusions and noise. However, they may suffer from ambiguities due to the used patch and

aperture sizes, making the approach not as accurate as most methods in the previous categories.

B. DEEP LEARNING-BASED 4D LF DISPARITY ESTIMATION METHODS

Deep learning-based 4D LF disparity estimation methods have been recently proposed to improve the performance of existing classical methods while greatly reducing the disparity estimation time. These methods rely on deep learning techniques and most of them are supervised by ground truth disparity maps to estimate disparity information. Heber and Pock [18] proposed the first convolutional neural network model to learn an end-to-end mapping between a 4D LF and its corresponding depths. Afterwards, Heber et al. proposed a U-Net architecture with 3D convolutions to estimate LF disparity maps for the central EPIs [19]. The EPI-based fully-convolutional neural Network (EPINet) and Multi-scale Aggregated Network (MANet) proposed in [20] and [21] significantly improved the disparity estimation accuracy for the central view and heavily rely on the EPI structure in densely sampled LFs. Shi et al. [16] overcame this limitation by proposing a framework that can be used for both dense and sparse LFs. While this method can estimate a disparity map for any LF angular location, angular consistency across views is not ensured. Jiang et al. [17] proposed a disparity estimation method starting from the four corner views. After that, the disparity is propagated into all other LF views and a 3D reconstruction method is used to fill the holes (i.e., remaining regions after propagation without disparity values). Although it can estimate disparity maps for both dense and sparse LFs, relying on the initial estimation of corner views can significantly affect the disparity estimation performance for wide baseline LFs. Wang et al. [22] proposed a generic mechanism for LF processing including disparity estimation using domain-specific convolutions. Recently, Chao et al. [23] proposed a disparity estimation method called SubFocal that learns the disparity distribution of dense LFs and estimates a smooth disparity map for the central view by using cost volumes at the sub-pixel level. Supervised deep learning-based methods achieve state-of-the-art results. However, they require a large number of training LFs with ground truth disparity maps, which are challenging to obtain in the real world. Moreover, training deep learning models using only synthetic LF datasets may not adequately handle the domain shift between the real world and synthetic datasets. Therefore, several unsupervised methods are proposed to handle this challenge, although the performance is slightly reduced [24].

III. PROPOSED DISPARITY PROPAGATION METHOD

The proposed method comprises three main steps as illustrated in Fig. 2. To start, two inputs are required, namely a 4D LF and an estimated disparity map of one reference view with respect to its adjacent right view (estimated by any available method). In this paper, the central view is selected as a reference view since it is equidistant from all corner views. Hence, represents a good compromise in terms of the remaining holes after propagating its disparity map into all corner views. Therefore, from hereinafter in this paper, the central view will be considered the reference view. Notice, however, that the proposed method can use any disparity estimation method and any angular location for the reference view, though the results may be affected accordingly, as explained in Section IV.

To apply the propagation into all LF views, in the first step, the input reference view disparity map is propagated into the four corner views in an occlusion-aware manner. The remaining holes in the corner views after propagation are filled by estimating their disparity values. Any disparity estimation method that can estimate disparity for any angular location for dense and sparse LFs can be used to fill the holes. In the second step, the disparity maps for the remaining LF views (i.e., all LF views except the reference and corner views) are generated via a recursive propagation in both horizontal and vertical directions separately. Afterwards, disparity maps from both horizontal and vertical propagation are fused for each view using their arithmetic mean.

In the last step, the disparity values of any remaining holes are computed, and a final edge-preserving refinement is applied to further regularize the output. The following subsections describe these steps in more detail.

A. DISPARITY PROPAGATION FOR CORNER VIEWS

To ensure angular consistency across LF views, the reference view disparity map, d^{ref}, is initially propagated into the four corner views (since they typically include most of the scene information) – in this paper, the central view located in the angular location (u_c, v_c), is used as reference view; thus $ref=(u_c, v_c)$. The propagation is achieved by assigning the same disparity value of each pixel in d^{ref} to the corresponding pixel in each corner view, which are computed using d^{ref} itself, as shown in (1):

$$\begin{cases} x^{(u,v)} = x^{ref} + d_{hor}^{ref\rightarrow(u,v)}, \\ y^{(u,v)} = y^{ref} + d_{ver}^{ref\rightarrow(u,v)}, \end{cases} \tag{1}$$

where x^{ref}, y^{ref} are the spatial coordinates from which the propagation is applied; $x^{(u,v)}$, $y^{(u,v)}$ are the corresponding spatial coordinates of x^{ref}, y^{ref} in view (u, v); $d_{hor}^{ref\rightarrow(u,v)}$, $d_{ver}^{ref\rightarrow(u,v)}$ are the horizontal and vertical disparity values located in the spatial position $\left(x^{ref}, y^{ref}\right)$ from the reference view to view (u, v). To ensure integer positioning, rounding is applied to the projected coordinates. Assuming a regular arrangement of cameras with a parallel optical axis and uniform camera baseline and focal length, as assumed in [25], and [26], the horizontal and vertical disparities from the reference view into any other LF view (u, v) is computed using (2). Equations (1) and (2) hold under the above assumption. Otherwise, camera parameters must be

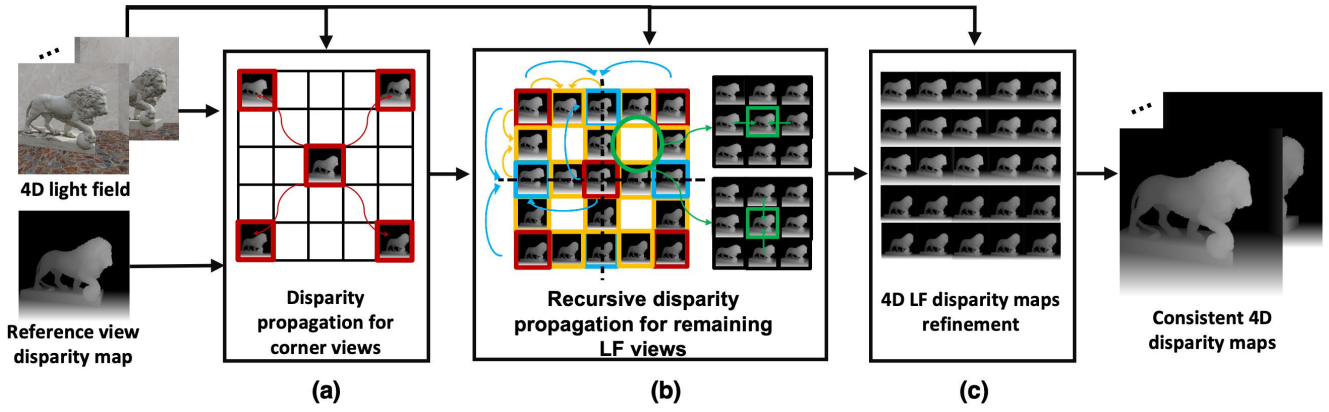

FIGURE 2. Illustration of the proposed disparity propagation method: a) Disparity propagation for corner views; b) Recursive disparity propagation for all remaining LF views to ensure angular consistency; c) Disparity map refinement for all LF views.

considered.

$$\begin{cases} d_{hor}^{ref\to(u,v)} = d^{ref} \times (u - u_c), \\ d_{ver}^{ref\to(u,v)} = d^{ref} \times (v - v_c)\,. \end{cases} \tag{2}$$

To detect occlusions, a binary occlusion mask is initialized for each LF view for all pixels, where each pixel is initially labeled as occluded. After estimating the disparity map of the reference view and propagating it into other LF views, all spatial locations that have disparity values are labeled as non-occluded and the remaining ones keep the initial occluded label.

To ensure occlusion-aware propagation, the input 4D LF and corresponding texture variation maps are used. Therefore, the input 4D LF is converted to the CIELAB color space. Then, a per-pixel texture variation map is generated from the L channel by computing the local standard deviation of a (3×3) neighborhood for all the pixels in each view. The texture variation maps are used to guide the propagation when different objects share either the same color or the same disparity values. Disparity propagation is applied only if the color and texture difference, D, between pixels in *ref* view and corresponding pixels in another view, as in (3), is less than or equal to a pre-defined threshold, τ, (i.e., $D \leq \tau$). In this paper, τ is set to 0.01 after extensive experiments to allow for a reasonable difference due to rounding and lighting differences in each view.

$$D = \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2 + (t_i - t_j)^2}, \tag{3}$$

where l, a, b are normalized color channel values (using min-max normalization [27]) in CIELAB color space; i, j represent the original pixel in *ref* view and corresponding pixels in (u, v) view, respectively; and t is the normalized texture value (using min-max normalization) for each pixel. The above thresholding operation is beneficial in preventing inaccurate projections into other views if the reference disparity maps have inaccurate values. Different values of τ are tested to study their effect in Section IV.

During propagation, the occlusion mask is checked for each spatial location, and when it has already a non-occluded label (i.e., another candidate has been propagated into the same location due to rounding, or inaccurate estimation), the maximum disparity value between the previous and current candidates is kept only if $D \leq \tau$, otherwise the disparity value will not be changed. The rationale for keeping the maximum disparity value comes from the observation that foreground objects, which are typically not occluded, have larger disparity values.

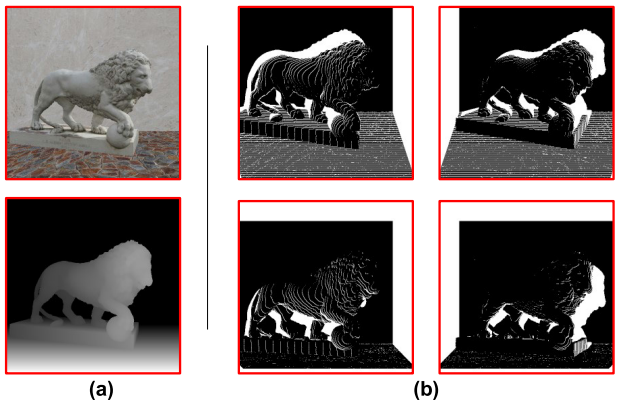

FIGURE 3. Example of occlusion masks after projecting the disparity map of the central view into other LF views: a) The central view and its disparity map; b) Binary occlusion masks of the four corner views after propagating the central disparity map into each one. White pixels indicate pixels without disparity values, i.e., occluded/invisible pixels (relatively to the central view).

The remaining holes in the corner views (white regions in Fig. 3b), i.e., regions without disparity values, need to be filled next. Instead of applying a blind filling/inpainting to those holes, the actual disparity values are truly estimated from the input 4D LF. Therefore, any existing disparity estimation method that can compute disparity maps for any angular location, and not only for the central view, for both dense and sparse LFs, can be used to fill the remaining holes. Different disparity estimation methods are used and evaluated in Section IV to study their effect on the estimated disparity maps. After assigning disparity values for all pixels in corner views, the corner views are used to guide the propagation for the remaining LF views, as explained in the next step.

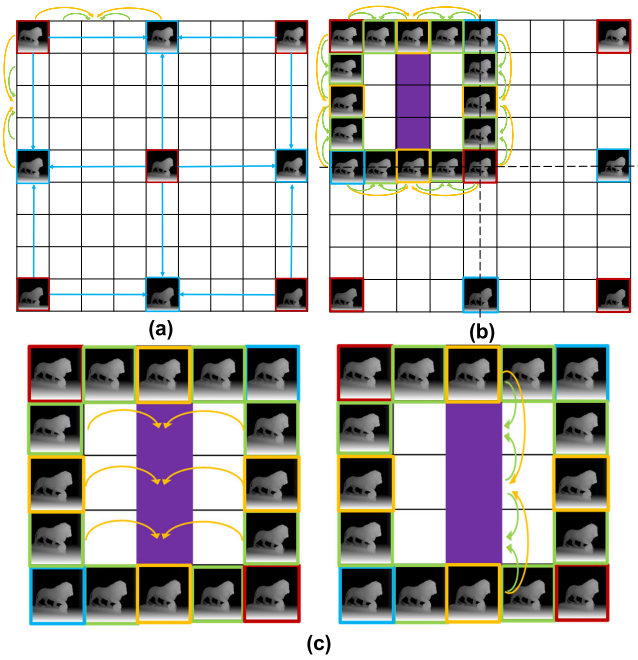

FIGURE 4. Disparity map estimation via propagation: a) Middle views between any two corner views (blue squares); b) Border and central crosshair views (yellow and green squares) using a recursive propagation; c) Internal views, such as purple shaded squares, have two different propagations from horizontal and vertical directions independently, then both computed disparity maps are fused to create one disparity map for each internal view.

B. RECURSIVE DISPARITY PROPAGATION FOR REMAINING LF VIEWS

In this step, disparity maps of all other LF views are obtained in three stages as presented in Fig. 4.

First, a disparity map is assigned to each middle view located halfway between any two corner views (blue squares in Fig. 4a). To achieve that, the reference view and the nearest two corner views are propagated into each middle view as shown in Fig. 4a (considering occlusions as in the previous step). Those three propagated disparity maps are then fused by considering their arithmetic mean.

Second, a disparity map is assigned to each remaining border and crosshair view, i.e., views in the central horizontal and vertical angular coordinates (yellow and green squares in Fig. 4b). To achieve that, a recursive propagation from two reference views located in the same horizontal or vertical angular dimension into the middle view located halfway between them is applied, as shown in Fig. 4b, until no more middle views without disparity values remain. At this point, the LF is divided into four quadrants and no disparity maps are still assigned for the internal views of each quadrant.

Third, a disparity map is assigned to the internal views in each quadrant by applying a recursive horizontal and vertical propagation independently, as shown in Fig. 4c (the same way as it has been done in the second stage for each row or column of internal views). Both disparity maps generated from horizontal and vertical recursive propagation are then fused for each internal LF view by considering their arithmetic mean.

TABLE 1. Test datasets used in the experiments.

4D LF dataset for testing	Disparity range
EPFL real world [29]: Fruits, Swans	$[-1, 1]$
Stanford real world [30]: Lego, Bunny	$[-3, 3]$
HCI synthetic [31]: Buddha, Papillon, Horses, StillLife	$[-4, 4]$
HCI* synthetic [28]: Table, Dishes (* for distinguishing the two different HCI datasets)	$[-4, 4]$
Inria synthetic sparse LFs [16]: Lion, Electro devices	$[-20, 20]$

C. 4D LF DISPARITY MAPS REFINEMENT

In this step, all LF views already have a disparity map. However, remaining hole locations, caused by occluded regions that do not exist in either the reference view or the corner views, or just caused by rounding the coordinates to integer indexing, need to be filled. Therefore, the four nearest left, right, top and bottom spatial neighbors that have disparity values for each pixel are considered in each view. The disparity value corresponding to the minimum difference $D_k < \tau, k = 1, \ldots, 4$ as in (3), is assigned as the disparity value of that location. If more than one neighbor has an equal D value, the minimum disparity value is considered. The reason for considering the minimum disparity is that the remaining holes, typically belonging to occluded regions, cannot be seen in the reference or corner views, are deeper than frontal objects (i.e., occluding objects), and hence have lower disparity values. After filling the remaining isolated pixels and holes, a simple and fast 2D edge-preserving median filtering using a (5×5) kernel size is applied spatially for all LF views to refine the estimated disparity maps.

IV. RESULTS AND EVALUATION

In this section, the proposed method is compared to several benchmark methods, namely: i) Shi et al. method [16], which is applied for each LF view independently since it only estimates one disparity map for any angular location; ii) Jiang et al. method [17]; and iii) Khan et al. method [10]. Both [10] and [17] create disparity maps for all LF views.

Moreover, several LF datasets with different disparity ranges are used (see Table 1). Notice that HCI and HCI∗ datasets are both synthetic dense LF datasets, however, they are different in the disparity ranges and in the spatial resolution. The entire (9×9) views for all datasets are considered. Only synthetic LFs with Ground Truth (GT) disparities for all LFs are used for the quantitative evaluation. The EPFL and Stanford LF datasets do not have GT disparity maps and, hence, the quantitative evaluation is not applied to them.

To quantitatively evaluate the proposed method, three different metrics are used, namely: i) Mean Square Error

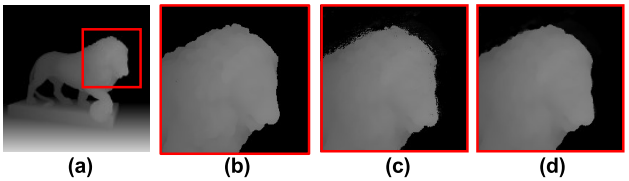

FIGURE 5. A visual example of the refinement step using edge-preserving filtering. a) GT disparity of view (3, 6); b) GT disparity enlargement; c) Proposed without refinement; d) Proposed with final refinement.

(MSE) [28]; ii) Percentage of Bad Pixels (BP) (i.e., the percentage of pixels with a disparity error above a certain threshold; the typically used Bad Pixels error thresholds of 0.01, 0.03 and 0.07 are used) [28]; and iii) View Consistency Error (i.e., where disparity maps of all LF views are projected into each angular location; then the variance of all projected disparity maps is computed for each angular location (81 values) as defined, formulated and implemented in [10]).

As explained in Section III, in the proposed method. a reference view disparity map is required as input. Additionally, the holes in corner views after propagating the reference view disparities need to be filled by estimating their disparities. Therefore, any existing method that can estimate disparity maps for any angular location in dense and sparse LFs can be used (more accurate is favored). In this paper, to achieve both accuracy and angular consistency for the generated disparity maps, Chao et al. SubFocal method [23] is adopted to estimate the input disparity map. The reason for choosing the SubFocal method is that it ranks first place, as reported by the authors, among other 99 submitted methods on the HCI 4D LF benchmark [28] considering different metrics. However, the SubFocal method can estimate a disparity map only for the central view and has been trained for dense LFs with a disparity range of $[-4, 4]$. Therefore, the remaining holes after propagation to the corner views are filled by using the pre-trained model of Shi et al. [16] that fine-tuned the optical flow estimation network (a.k.a. FlowNet 2.0) [32] for LF disparity estimation in any angular location. Moreover, to consider sparse LFs, the model in [23] is retrained in our experiments by using LFs with a wider disparity range (i.e., $[-20, 20]$). To retrain the SubFocal method for sparse LFs, the Inria synthetic sparse LF dataset in Table 1 was used for training (the same number of LFs was used for training, i.e., 16 LFs, as in [23]). The hyperparameters were kept the same as in [23], except for the disparity sampling step size which was set to 2.5 instead of 0.5 to reduce the cost volume complexity. In this paper, the retrained model is tested for sparse LFs using Lion, and Electro devices test LFs. In the experiments, different methods are also used to study the effect of the selected disparity estimation method on the proposed propagation method (including the use of the method [16] for both the reference and corner views).

Notice that some results are not available (indicated n/a in Fig. 7, Fig. 8 and Fig. 10) since the EPI-based method in [10] does not support LFs with large disparity ranges. Additionally, the SubFocal method [23] is used to estimate the input reference view disparity map for most datasets. However, disparity maps of the HCI dataset [31] are estimated using the Shi et al. method [16] since the SubFocal method [23] generates inaccurate disparity estimations for this dataset due to the different resolutions and the domain shift of training and testing, as reported by the authors and shown later in Fig. 9.

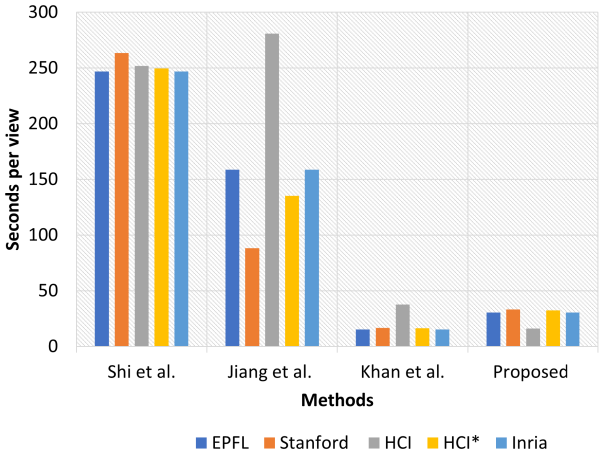

FIGURE 6. Average CPU time in seconds per view.

The proposed method is implemented using MATLAB and all results ran on a desktop computer with a 64-bit Ubuntu operating system, AMD® Epyc 7282 16-core CPU, NVIDIA GeForce RTX 3090 and 256 GB RAM.

A. QUANTITATIVE AND QUALITATIVE RESULTS

In this section, the proposed method results are presented and compared to the benchmark methods using several datasets with various disparity ranges. Notice that only the datasets that have ground truth disparity maps are used in the quantitative results namely, HCI [31], HCI* [28] and Inria [16] LF datasets.

Initially, the effect of the parameter τ is studied by using different values and finding the evaluation metrics accordingly. Therefore, 4 different experiments are conducted where different values of τ are used, i.e., 0.001, 0.01, 0.1 and ∞, where ∞ refers to the case where the visual consistency is discarded during the propagation. As can be seen from Table 2 using different τ values can slightly affect the accuracy and the CPU time. As illustrated in Section III, the occlusion-aware propagation step considers the disparity values of foreground and background regions. However, to avoid wrong projection for objects that have similar disparity values but are different in color, the τ value is set to 0.01. This value prevents inaccurate projection from occurring due to discrete sampling, rounding errors or inaccurate estimated values. The chosen value allows for reasonable visual differentiation across LF views, accounting for varying lighting conditions. Simultaneously, it strikes a reasonable balance between accuracy, efficiency, and prevention of inaccurate projections.

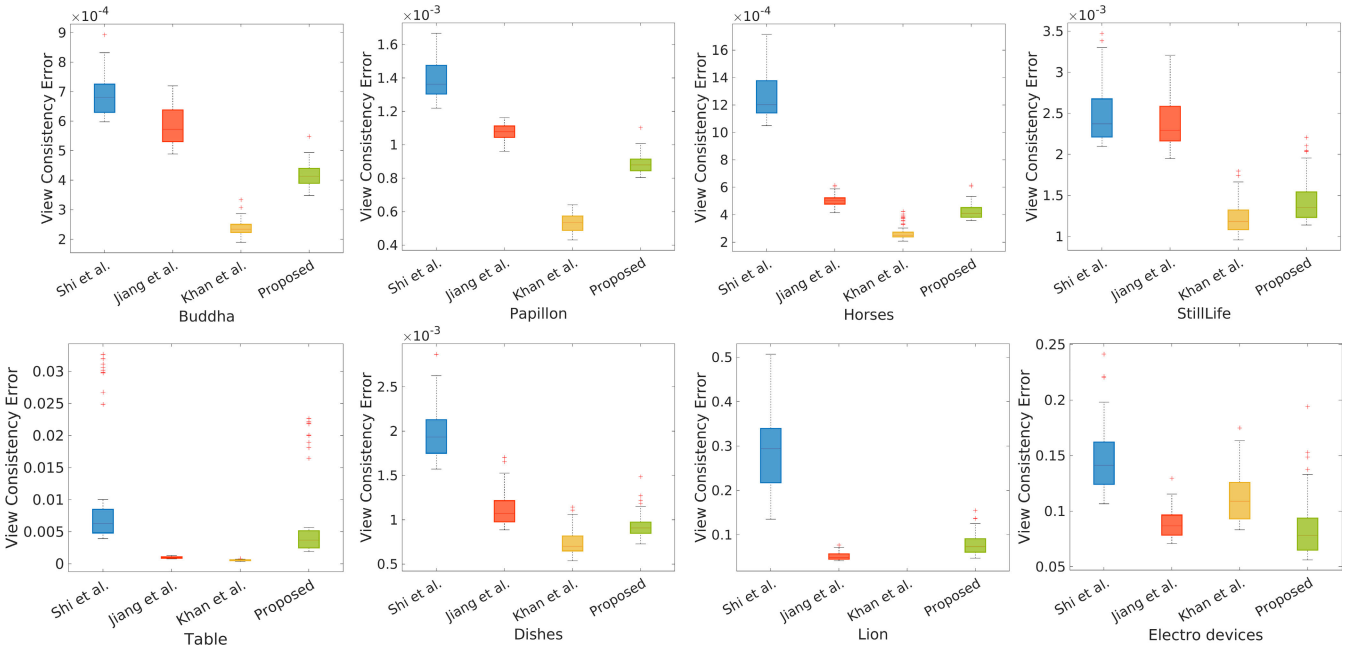

FIGURE 7. Summary statistics of view consistency error across all LF views for each test LF.

TABLE 2. Average quantitative results using the proposed method with different τ values on various LF datasets for all LF views.

Metric	LF \ τ	0.001	0.01	0.1	∞
MSE × 100 (↓)	HCI [31]	0.68	**0.66**	0.73	0.76
	HCI* [28]	1.63	1.49	0.83	**0.83**
	Inria [16]	**39.16**	41.52	44.43	45.53
BP 0.01 (↓)	HCI [31]	62.16	**62.10**	62.11	62.12
	HCI* [28]	22.32	19.74	19.60	**19.58**
	Inria [16]	**87.43**	87.87	88.01	88.03
BP 0.03 (↓)	HCI [31]	25.78	**25.75**	25.82	25.85
	HCI* [28]	10.04	9.07	8.73	**8.70**
	Inria [16]	**64.58**	65.44	65.83	65.88
BP 0.07 (↓)	HCI [31]	**16.76**	17.27	17.46	17.52
	HCI* [28]	5.35	4.94	4.69	**4.67**
	Inria [16]	**31.63**	32.09	32.64	32.73
Average CPU time in seconds per view	HCI [31]	17.11	16.01	15.97	**15.92**
	HCI* [28]	32.93	32.46	32.43	**32.41**
	Inria [16]	33.63	33.18	33.14	**33.12**

TABLE 3. Average quantitative results on various LF datasets using different 4D LF disparity estimation methods for all LF views.

Metric	LF \ Method	Shi et al. [16]	Jiang et al. [17]	Khan et al. [10]	Proposed without refine.	Proposed with refine.
MSE × 100 (↓)	HCI [31]	0.70	0.98	1.20	0.70	**0.66**
	HCI* [28]	2.14	1.66	3.95	**1.46**	1.49
	Inria [16]	67.62	2104.39	118.40	44.50	**41.52**
BP 0.01 (↓)	HCI [31]	63.11	**58.25**	75.08	62.26	62.10
	HCI* [28]	50.93	48.53	75.38	**19.40**	19.94
	Inria [16]	**82.90**	96.30	94.41	88.38	87.87
BP 0.03 (↓)	HCI [31]	27.01	**24.66**	39.68	26.05	25.75
	HCI* [28]	25.41	25.30	45.98	**8.90**	9.07
	Inria [16]	**59.09**	89.26	83.58	66.72	65.44
BP 0.07 (↓)	HCI [31]	7.50	8.28	15.87	7.04	**6.82**
	HCI* [28]	14.09	16.34	25.86	4.98	**4.94**
	Inria [16]	37.82	80.37	64.79	34.01	**32.09**

To study the impact of the final refinement step, Table 3 shows the performance of the proposed method with and without applying the edge-preserving filter to refine the generated disparity maps in the refinement step (i.e., step C in Section III). As presented in Table 3, the results are slightly improved when a simple median filter is applied to all LF views to regularize the estimated disparity maps in most datasets. The used filter is simple, fast, preserves scene edges and can reduce inaccurate disparity propagations, especially for a few pixels with wrong disparity values surrounded by pixels with accurate ones (see Fig. 5).

Moreover, the proposed method is evaluated and compared to the benchmark methods using various datasets, as presented in Table 3, Table 5, Fig. 6, Fig. 7 and Fig. 8.

To compare the computational complexity between the various methods, all methods were run using the CPU, and CPU times are reported in Fig. 6. The reported time for the proposed method includes the disparity estimation time for the reference view and for occlusions in corner views and all the steps in Section III. The CPU time spent by all the benchmark methods for all LF views is then divided by the number of views to obtain an average CPU time per view. The breakdown of the average CPU time for each step is

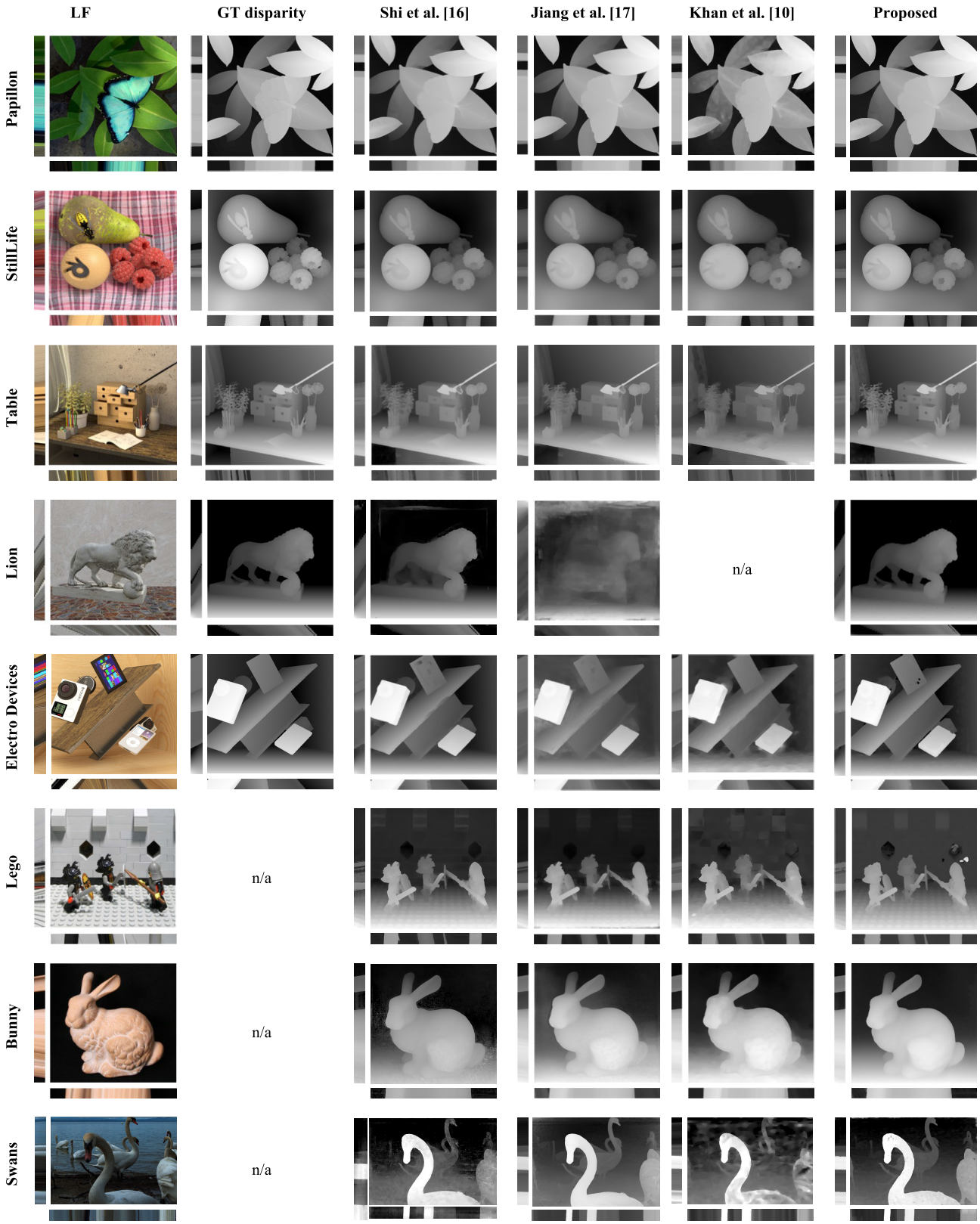

FIGURE 8. Visual comparison using the proposed disparity propagation method and the benchmark methods for dense and sparse LFs. The central view and central horizontal and vertical EPIs are shown for all LFs. Not available (n/a) results for the Khan et al. method since it does not support very sparse LFs.

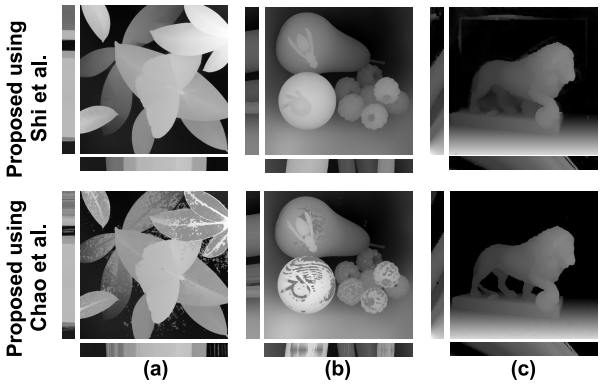

FIGURE 9. Examples to show the effect of the selected central view disparity estimation method on the proposed disparity propagation method. First row, using the Shi et al. method [16]; Second row, using Chao et al. SubFocal method [23].

TABLE 4. Breakdown of the average CPU time for the proposed method (in seconds).

4D LFs dataset for testing	(A) Disparity estimation for the reference view	(B) Disparity estimation for the 4 corner views	(C) Disparity propagation into all 81 views	Average CPU time per view $\frac{(A)+(B)+(C)}{81}$
EPFL [29]	1456.38	987.36	19.19	30.41
Stanford [30]	1613.42	1053.61	22.15	33.20
HCI [31]	251.72	1006.90	38.04	16.01
HCI* [28]	1612.37	998.05	18.70	32.46
Inria [16]	1612.22	1050.91	24.27	33.18

reported in Table 4. As in Table 3 and Fig. 6, the results of the proposed method generate competitive accuracy results while reducing the complexity when compared to the benchmark methods, especially for challenging sparse LFs.

Besides the improvements of the proposed method in terms of accuracy metrics in some datasets, a significant reduction in time is shown in Fig. 6 when compared to Shi et al. [16] and Jiang et al. [17]. Compared to Khan et al. [10], the CPU time results are still competitive. However, when the Khan et al. method is used for the reference and corner views, the proposed method requires less CPU time than [10], as shown later in this section. This reduction in time is achieved by exploiting the correlation between LF views and applying angularly consistent propagation. The CPU time for the proposed method depends on the used methods to estimate the reference view disparity map and the occluded regions as described below in this section.

Regarding the angular consistency of obtained disparity maps, Fig. 7 shows the angular consistency using boxplots where the central mark indicates the median and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Notice that, in this experiment, two different disparity estimation methods are used during the propagation steps in the proposed method (e.g., the SubFocal

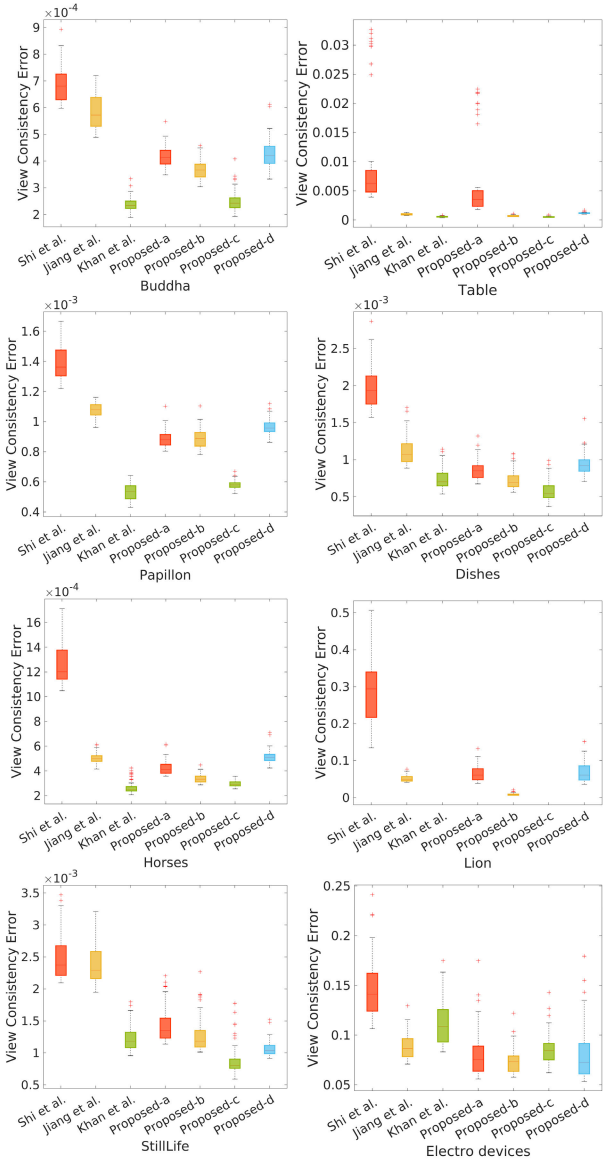

FIGURE 10. Summary statistics of view consistency error across all LF views for each test LF. In this figure, the results of the proposed method are compared using the benchmark method to estimate the disparity map for the reference view and occluded regions in the corner view namely, using Shi et al. method (Proposed-a); using Jiang et al. method (Proposed-b); using Khan et al. method (Proposed-c); and using the ground truth disparity (Proposed-d). The proposed disparity propagation method leads to better view consistency compared to the original benchmark methods.

method [23] is used for the reference view and the pre-trained model of Shi et al. [16] is used for occlusions in the corner views). Despite that, the proposed method outperforms Shi et al. [16] by ensuring angular consistency during the propagation as can be noticed in Fig. 7. Additionally, the proposed method outperforms Jiang et al. method [17] in 6 LF scenes, while their method outperforms the proposed method in the remaining 2 LFs (i.e., Table, and Lion). Khan et al. method [10] outperforms the proposed method for dense LFs and ensures angular consistency across views. However, the method in [10] does not adequately support large occlusions

TABLE 5. Average quantitative results on various LF datasets using different 4D LF disparity estimation methods for all LF views. The propagation results are tested using different disparity estimation methods for the reference view and hole filling in the coroner views, including the ground truth disparity. The best results among the proposed method results using different estimation techniques are highlighted in bold.

Metric	LF \ Method	Shi et al.[16]	Proposed using Shi et al. [16]	Jiang et al. [17]	Proposed using Jiang et al. [17]	Khan et al. [10]	Proposed using Khan et al. [10]	Proposed using ground truth
MSE× 100 (↓)	HCI [31]	0.70	**0.66**	0.98	0.96	1.20	1.11	0.28
	HCI* [28]	2.14	2.14	1.66	**1.70**	3.95	3.89	0.51
	Inria [16]	67.62	**68.44**	2104.39	2129.15	118.40	111.78	8.96
BP 0.01 (↓)	HCI [31]	63.11	62.10	58.25	**58.26**	75.08	74.94	10.37
	HCI* [28]	50.93	**46.38**	48.53	48.87	75.38	75.44	5.89
	Inria [16]	82.90	**80.48**	96.30	96.28	94.41	94.12	19.20
BP 0.03 (↓)	HCI [31]	27.01	25.75	24.66	**24.69**	39.68	39.40	1.72
	HCI* [28]	25.41	**23.17**	25.30	25.54	45.98	45.76	2.62
	Inria [16]	59.09	**55.15**	89.26	89.14	83.58	82.65	7.26
BP 0.07 (↓)	HCI [31]	7.50	**6.82**	8.28	8.20	15.87	15.54	0.65
	HCI* [28]	14.09	**13.15**	16.34	16.31	25.86	25.43	1.78
	Inria [16]	37.82	**33.33**	80.37	80.07	64.79	63.50	2.11

across LF views, which is typical for sparse LFs. While the view consistency of disparity maps is essential, accuracy is also important. As can be seen from Fig. 7 and Fig. 8, for the Lion LF, for instance, achieving the best performance in terms of view consistency does not necessarily lead to better visual accuracy and in terms of other metrics. Besides the visual results shown in Fig. 8, readers are encouraged to see also the dynamic results in our GitHub repository.[1]

To study the effect of the selected disparity estimation method on the proposed propagation method, different disparity estimation methods are used for estimating the reference view and filling the occlusions in the corner views. To achieve that, the benchmark methods [10], [16], [17] and the ground truth disparity are compared to the proposed method results generated by using each benchmark method for estimating the reference view and filling the holes in the corner views. After that, the proposed propagation method is applied to compute disparity maps for all LF views.

As can be seen in Table 5, by only using one reference disparity map and by exploiting the correlations across LF views, the proposed propagation method outperforms the original benchmark methods in most test LFs and can generate competitive results in others. Moreover, the proposed propagation method ensures better view consistency than the original benchmark methods in most LFs as presented in Fig. 10. Notice that, for some LFs, using the ground truth disparity in the proposed method has lower performance, in terms of the view consistency metric, compared to the estimated ones, as shown in Fig. 10. The reason for this is that the ground truth disparity is more distinct and sharper around objects boundaries when compared to the smooth estimated ones. Hence, it generates larger and sharper holes

[1]Dynamic results for all LF views can be found at: https://github.com/MaryamHamad/LFDisparityPropagation

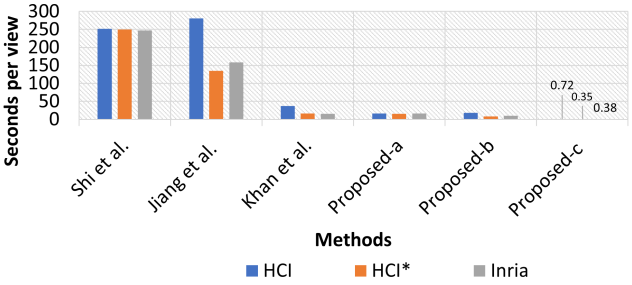

FIGURE 11. Average CPU time in seconds per view. In this figure, the results of the proposed method are generated by using the benchmark method to estimate the disparity map for the reference view and occluded regions in the corner views namely, using Shi et al. (Proposed-a); using Jiang et al. (Proposed-b); and using Khan et al. (Proposed-c). The proposed disparity propagation method significantly reduces the required time compared to the original benchmark methods.

in occluded regions that need to be filled after propagation. Small differences in filling those regions across LF views can heavily affect the consistency metric results. Finally, the proposed method can drastically reduce the average CPU time per view when compared to the benchmark methods, as shown in Fig. 11.

To sum up, the proposed disparity propagation method enables computing an accurate disparity map for each LF view only from one reference view disparity map and hole filling in the corner views. The proposed method leads to improved accuracy and view consistency for most of the LF datasets and reduces the computational complexity compared to the benchmark methods. Some limitations remain such as if the input reference view has inaccurate estimation, there is no correction step to check if the values are accurate or not, and the inaccurate values will be propagated into all other LF views as shown in Fig. 9. This limitation can be avoided by using an accurate disparity estimation method to estimate the disparity map for the reference view.

V. FINAL REMARKS

In this paper, an efficient disparity propagation method is proposed to generate angularly consistent disparity maps for all LF views. Given only one estimated disparity map of a reference view, the proposed method exploits the correlation across LF views and propagates the reference disparity map to the corner views at first. The remaining holes in the corner views are not interpolated but truly estimated by adopting an off-the-shelf disparity estimation method. Afterwards, disparity maps of the reference and corner views are propagated recursively in horizontal and vertical angular directions in an occlusion-aware manner into all remaining LF views. Finally, a refinement step is included to regularize the final disparity maps and fill any remaining holes. Since most of the existing methods estimate disparity information for the central view only, the proposed method can be used as plug and play with them to enable the generation of angularly consistent disparity maps for all LF views. Experimental results for several LF datasets with different disparity ranges show competitive results in terms of angular consistency and estimation accuracy compared to the existing methods with a significant complexity reduction.

For future work, the question of how to adaptively select the location of the reference view and the possibility of adding more reference views will be investigated to effectively consider occlusions based on the LF disparity range. Moreover, the current implementation is not optimized yet and the computational complexity of the proposed disparity propagation method can be further reduced.

REFERENCES

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

[2] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.

[3] S. Zhou, T. Zhu, K. Shi, Y. Li, W. Zheng, and J. Yong, "Review of light field technologies," *Vis. Comput. Ind., Biomed., Art*, vol. 4, no. 1, pp. 1–13, Dec. 2021.

[4] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.*, vol. 1, no. 1, pp. 7–55, 1987.

[5] M. Hamad, C. Conti, P. Nunes, and L. D. Soares, "ALFO: Adaptive light field over-segmentation," *IEEE Access*, vol. 9, pp. 131147–131165, 2021.

[6] D. Egan, M. Alain, and A. Smolic, "Light field style transfer with local angular consistency," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Toronto, ON, Canada, Jun. 2021, pp. 2300–2304.

[7] M. Hamad, C. Conti, A. M. de Almeida, P. Nunes, and L. D. Soares, "SLFS: Semi-supervised light-field foreground-background segmentation," in *Proc. Telecoms Conf. (ConfTELE)*, Leiria, Portugal, Feb. 2021, pp. 1–6.

[8] 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.

[9] 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.

[10] N. Khan, M. H. Kim, and J. Tompkin, "View-consistent 4D light field depth estimation," in *Proc. 31st Brit. Mach. Vis. Conf. (BMVC)*, 2020.

[11] H. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1547–1555.

[12] C. 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. 2019.

[13] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 673–680.

[14] H. Lin, C. Chen, S. B. Kang, and J. Yu, "Depth recovery from light field using focal stack symmetry," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 3451–3459.

[15] J. Y. Lee and R. Park, "Complex-valued disparity: Unified depth model of depth from stereo, depth from focus, and depth from defocus based on the light field gradient," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 830–841, Mar. 2021.

[16] J. Shi, X. Jiang, and C. Guillemot, "A framework for learning depth from a flexible subset of dense and sparse light field views," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5867–5880, Dec. 2019.

[17] X. Jiang, J. Shi, and C. Guillemot, "A learning based depth estimation framework for 4D densely and sparsely sampled light fields," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 2257–2261.

[18] S. Heber and T. Pock, "Convolutional networks for shape from light field," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 3746–3754.

[19] S. Heber, W. Yu, and T. Pock, "Neural EPI-volume networks for shape from light field," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2271–2279.

[20] C. Shin, H. Jeon, Y. Yoon, I. S. Kweon, and S. J. Kim, "EPINET: A fully-convolutional neural network using epipolar geometry for depth from light field images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 4748–4757.

[21] 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)*, Barcelona, Spain, May 2020, pp. 1998–2002.

[22] Y. Wang, L. Wang, G. Wu, J. Yang, W. An, J. Yu, and Y. Guo, "Disentangling light fields for super-resolution and disparity estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 425–443, Jan. 2023.

[23] W. Chao, X. Wang, Y. Wang, L. Chang, and F. Duan, "Learning sub-pixel disparity distribution for light field depth estimation," 2022, *arXiv:2208.09688*.

[24] J. Jin and J. Hou, "Occlusion-aware unsupervised learning of depth from 4-D light fields," *IEEE Trans. Image Process.*, vol. 31, pp. 2216–2228, 2022.

[25] M. Hog, N. Sabater, and C. Guillemot, "Superrays for efficient light field processing," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1187–1199, Oct. 2017.

[26] H. Zhu, Q. Zhang, Q. Wang, and H. Li, "4D light field superpixel and segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 85–99, 2020.

[27] J. Han, M. Kamber, and J. Pei, "Data pre-processing," in *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2011, ch. 3, pp. 114–115.

[28] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Proc. 13th Asian Conf. Comput. Vis. (ACCV)*, Taipei, Taiwan, 2016, pp. 19–34.

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

[30] V. Vaish and A. Adams. (2008). *The (New) Stanford Light Field Archive*. Accessed: Jun. 24, 2020. [Online]. Available: http://lightfield.stanford.edu/acq.html

[31] S. Wanner, S. Meister, and B. Goldlüecke, "Datasets and benchmarks for densely sampled 4D light fields," *Vis., Model. Vis.*, vol. 13, pp. 225–226, Sep. 2013.

[32] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "FlowNet 2.0: Evolution of optical flow estimation with deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1647–1655.

MARYAM HAMAD (Graduate Student Member, IEEE) received the B.E. degree in computer systems engineering (CSE) from Palestine Technical University-Kadoorie (PTUK), Palestine, in 2018, covered by an excellence scholarship. She is currently pursuing the Ph.D. degree with Instituto Universitário de Lisboa (ISCTE-IUL), Portugal. During her degree, she spent one semester as an Exchange Student with the ERASMUS+ Program, Middle East Technical University (METU), Turkey. She completed her professional internship in information science and technology with the IAESTE Program, Multimedia Signal Processing Group, Instituto de Telecomunicações, Portugal, as a Researcher. Currently, she is a Researcher with the Multimedia Signal Processing Group and a member of the IEEE Women in Engineering Society, IEEE Signal Processing Society, and IEEE Young Professionals Group. Her current research interests include immersive visual technologies, such as light field imaging, digital image processing, and computer vision. She acts as a reviewer of the IEEE Access.

CAROLINE CONTI (Member, IEEE) received the B.Sc. degree in electrical engineering from Universidade de São Paulo (USP), Brazil, in 2010, and the Ph.D. degree in information science and technology from Instituto Universitário de Lisboa (ISCTE-IUL), Portugal, in 2017. Currently, she is a Senior Researcher with the Multimedia Signal Processing Group, Instituto de Telecomunicações, and an Assistant Professor with the Information Science and Technology Department, ISCTE-IUL. Her research interests include immersive visual technologies and image and video processing, including light field processing and coding. She has contributed over 25 papers to international journals and conferences in these areas. She also serves as an Associate Editor for IEEE Transactions on Image Processing, has been a Guest Editor for *Signal Processing: Image Communication* (Elsevier) journal, and actively participates as a reviewer for various IEEE and EURASIP journals and conferences.

PAULO NUNES (Member, IEEE) received the Graduate degree in electrical and computers engineering from Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1992, and the M.Sc. and Ph.D. degrees in electrical and computers engineering from IST, in 1996 and 2007, respectively. Currently, he is a Senior Researcher with the Multimedia Signal Processing Group, Instituto de Telecomunicações, Portugal. In addition, he is also an Associate Professor with the Information Science and Technology Department, Instituto Universitário de Lisboa (ISCTE-IUL), Portugal. His current research interests include 2D/3D image and video processing and coding, namely light field image and video processing and coding. He has coordinated and participated in various national and international (EU) funded projects and has acted as a project evaluator for the European Commission. He acts often as a reviewer of various ACM, EURASIP/Elsevier, IEEE, IET, MDPI, SPIE, and Springer conferences and journals and a member of the program and organizing committees of various international conferences. He has contributed more than 70 papers to international journals and conferences in these areas.

LUÍS DUCLA SOARES (Senior Member, IEEE) received the Licenciatura and Ph.D. degrees in electrical and computer engineering from Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1996 and 2004, respectively. Currently, he is a Senior Researcher with the Multimedia Signal Processing Group, Instituto de Telecomunicações, Portugal. In addition, he is also an Associate Professor with the Information Science and Technology Department, Instituto Universitário de Lisboa (ISCTE-IUL), Portugal. His research interests include image and video coding/processing, including light field coding and processing and biometric recognition. He has contributed more than 70 papers to international journals and conferences in these areas. In addition, he has participated in the development of the MPEG-4 visual standard and in several national and international projects. He is a member of the Editorial Board of the *EURASIP Journal on Advances in Signal Processing* (Elsevier). In parallel, he acts as a reviewer of several IEEE, IET, and EURASIP journals and conferences.