

Consistent Disparity Synthesis for Inter-View Prediction in Lightfield Compression

Author:

Li, Y; Mathew, R; Ruefenacht, D; Naman, A; Taubman, D

Publication details:

2019 Picture Coding Symposium, PCS 2019

v. 00

pp. 1 - 5

9781728147048 (ISBN)

2330-7935 (ISSN); 2472-7822 (ISSN)

Event details:

2019 Picture Coding Symposium (PCS)

PEOPLES R CHINA, Zhejiang Univ, Ningbo Inst Technol, Ningbo

2019-11-12 - 2019-11-15

Publication Date:

2019-11-01

Publisher DOI:

<https://doi.org/10.1109/pcs48520.2019.8954506>

License:

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/unsworks_73652 in <https://unsworks.unsw.edu.au> on 2024-04-26

Consistent Disparity Synthesis for Inter-View Prediction in Lightfield Compression

Yue Li, Reji Mathew, Dominic Ruefenacht, Aous Naman, David Taubman

School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia

Abstract—For efficient compression of lightfields that involve many views, it has been found preferable to explicitly communicate disparity/depth information at only a small subset of the view locations. In this study, we focus solely on inter-view prediction, which is fundamental to multi-view imagery compression, and itself depends upon the synthesis of disparity at new view locations. Current HDCA standardization activities consider a framework known as WaSP, that hierarchically predicts views, independently synthesizing the required disparity maps at the reference views for each prediction step. A potentially better approach is to progressively construct a unified multi-layered base-model for consistent disparity synthesis across many views. This paper improves significantly upon an existing base-model approach, demonstrating superior performance to WaSP. More generally, the paper investigates the implications of texture warping and disparity synthesis methods.

Index Terms—lightfield coding, depth synthesis, dis-occlusion handling, base models

I. INTRODUCTION

This paper is concerned with the efficient prediction of views from other views within high density multi-view imagery. The term “lightfield” is also used for such content, which provides a rich angular, as well as spatial sampling of the scene, due to its large number of views. The lightfield datasets explored in this work arise from a high density camera array (HDCA). The compression of lightfields in general, and HDCA data in particular, have been attracting increasing attention, as evidenced by the current JPEG-Pleno light field standardization activity [1].

Inter-view prediction plays an important role in such compression schemes. In general, at the decoder side, some views may be predicted from other “intra-coded” views, after which prediction residuals may be coded. Prediction can also be used to recover original or virtual views at the decoder side, compensated with coded prediction residuals if available. In this paper, we do not explicitly consider the compression aspects of such a system, but focus exclusively on the performance of the inter-view prediction process. High quality inter-view prediction is facilitated by geometric models, expressed in terms of scene depth or reciprocally in terms of disparity maps. Since there are a large number of views, it is inefficient to explicitly code depth fields or disparity maps at every view. Instead, an underlying assumption in this paper is that explicit disparity information will be available to a decoder only at a sparse set of viewpoints. This is the model that has been found most effective within the JPEG-Pleno activity, embodied by its WaSP coding framework [2]. By contrast, in 3D-HEVC [3] each view is accompanied by its own disparity map.

When disparity maps accompany only a few views, the performance of view prediction depends on the way in which sparsely communicated disparity is utilized for texture warping. In cases where a view involved in prediction does not have its own communicated disparity, new disparity may need to be synthesized from existing ones, leading to an intimate connection between view prediction and disparity synthesis.

View prediction algorithms for lightfield compression proposed in [4] and [5] encode disparity maps anchored at a small set of intra-coded views, to transfer the texture data from these views to predicted target views on a pixel-by-pixel basis. In these approaches, each source pixel generally maps to a non-integer location within the target view and is splatted to the neighbouring pixel locations. WaSP [2] also adopts a pixel-based splatting strategy for texture warping, but works under a hierarchical view prediction framework. This means that the reference views employed for a given prediction step may not coincide with communicated disparity maps, so new disparity maps must generally be synthesized as the algorithm proceeds from coarse to fine levels in the hierarchy. These algorithms do not attempt to resolve inconsistencies between multiple communicated disparity maps; they simply combine the splatted results produced using each reference view and its associated disparity data.

Tran et al. [6] also adopt a pixel splatting approach, but arrange for the missing texture within occluded regions to be filled in a consistent way across all the target views via a proposed coordinate alignment algorithm. Their method relies heavily upon the availability of reliable disparity maps at reference views and only deals with one-dimensional inter-view prediction. Pearson et al. [7] propose a layer-based framework for communicating disparity, which effectively allows the disparity to be consistently inferred at the predicted target view and then used to interpolate pixels found within the prediction source views, along with relevance information that identifies the degree to which each source pixel is expected to be visible in the target view. Construction of the layer-based model is a non-trivial exercise in this approach, as each layer is assigned only one global disparity.

Evidently, there are multiple ways to use a small set of communicated disparity maps to predict the texture data at a given view from other views. The purpose of this paper is to explore the implications of some of these different methods in a disciplined way, while proposing our own novel approach.

To do this, we first provide a framework for the texture warping process that allows different disparity synthesis ap-

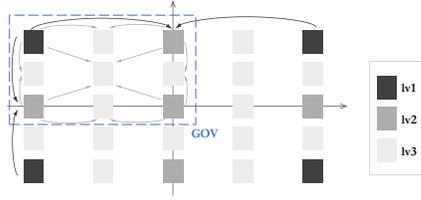


Fig. 1. Hierarchical View Prediction in HDCA Dataset

proaches to be compared in a meaningful way. We investigate forward and backward texture warping strategies, as well as the significance of consistent disparity synthesis. We then introduce our base-mesh model to address the issue of inconsistency by keeping a single description for geometric relationships across a whole group of views (GOV). First introduced in [8], the base model construction involves building an original model starting with the disparity map that governs the GOV and completing the model by progressively filling holes exposed at each view prediction step. The existing *backfill* approach in [8] suffers from excessive reliance on extrapolation for the hole filling problem. To address this, we propose a novel *disparity fusion* strategy that is superior to all the other methods explored in this paper.

II. VIEW PREDICTION USING SYNTHESIZED DISPARITY

We are concerned with the prediction of *target* views from other *reference* views, using disparity information that may be communicated at some views but must generally be synthesized elsewhere. For lightfield compression, hierarchical prediction strategies are often employed. At each level in the hierarchy, new views are predicted from views that are available at coarser levels of the hierarchy, and prediction residuals can be coded before moving to the next finer level. Benefits of hierarchical prediction include efficient access to individual views and the ability to use quality scalable compression techniques, both of which are possible due to the total number of *precursor* views on which a given decoded view ultimately depends is limited by the depth of the hierarchy.

Fig. 1 provides an example of such a hierarchy, starting from 4 intra-coded *corner* views at the root (level 0) of the hierarchy. As shown, 5 new views are introduced at level 1, each predicted from the 2 or 4 nearest views at level 0. Next, 16 new views are introduced at level 2, again each predicted from the 2 or 4 nearest views at level 1. In a compression system, prediction residuals can be coded for each predicted view, so that prediction errors can be corrected and high-quality views become available as prediction references for the next level. This paper is not concerned with the details of residual coding, focusing exclusively on the properties of the view prediction process.

Fig. 2 illustrates three different ways to predict a target view V_T from neighbouring reference views V_{R_k} , $k = 1, 2, \dots, K$. For simplicity, only $K = 2$ reference views are shown in the figure. The distinction between these three view prediction strategies lies in the way disparity information is used, which further leads to different ways of warping texture across the views based on the disparity.

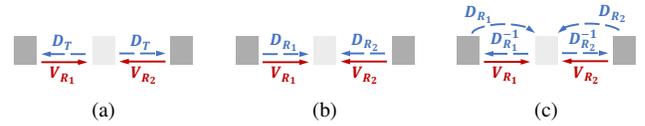


Fig. 2. View prediction using synthesized disparity. (a)*tgt-bwd*: backward warping with disparity anchored at the target; (b)*ref-fwd*: forward warping with disparity anchored at references; (c)*ref-bwd*: backward warping with disparity anchored at references

Disparity information can be available at target view V_T or reference views V_{R_k} . Fig. 2a shows a *target-anchored strategy* where view prediction is based on disparity D_T described at the target view. In this case, the target view is predicted by projecting D_T onto the reference views and bringing back the texture - this is known as *backward warping*. We refer to this as the *tgt-bwd* method. This is typical of coding frameworks such as 3D-HEVC where disparity information is explicitly communicated at the target view. Furthermore, the method proposed by Pearson et al. [7] can also be viewed as a *tgt-bwd* method.

By contrast, the prediction strategies in Fig. 2b and 2c are based on disparities D_{R_k} described at the reference view locations. We identify these as *reference-anchored strategies*. In the first case, texture pixels from each reference view are warped according to their respective disparity maps, splatting each translated texture sample to the nearest pixel on the target view - known as *forward warping*. The scheme in Fig. 2c is also reference-anchored, but the texture warping process is more elaborate: each reference disparity map is first mapped to the target view, so that backward warping can be used to import texture samples from the reference view, after which the warped views produced by each reference are combined. We refer to these two strategies as the *ref-fwd* and *ref-bwd* methods, respectively. Lightfield coding algorithms [2] [4] [5] cited in the introduction are based on *ref-fwd* methods; as we shall see, however, the *ref-bwd* method provides superior performance.

The three view prediction methods can be expressed as:

$$[tgt-bwd] \quad \hat{V}_T = \sum_k \lambda_k \cdot \mathcal{W}_{D_T}^{-1}(V_{R_k}) \quad (1a)$$

$$[ref-fwd] \quad \hat{V}_T = \sum_k \lambda_k \cdot \mathcal{W}_{D_{R_k}}(V_{R_k}) \quad (1b)$$

$$[ref-bwd] \quad \hat{V}_T = \sum_k \lambda_k \cdot \mathcal{W}_{D_{R_k}^{-1}}^{-1}(V_{R_k}), \quad (1c)$$

In these equations \mathcal{W} denotes a forward warping operation, while \mathcal{W}^{-1} denotes backward warping. The subscript indicates the disparity information the warping process is based upon. In backward warping, pixels in the target view are mapped to arbitrary locations in the reference views and a disciplined interpolation strategy is used to recover their values from neighboring reference samples. By contrast, forward warping involves a splatting process; splatting to a nearest neighbor is less desirable from a signal processing perspective, but has the benefit that regions of dis-occlusion (holes) are readily identified as locations that are never hit.

The λ_k fields in these equations denote visibility factors; with only two reference views, the visibility at each location,

from each view k , is 0, 1 or $\frac{1}{2}$, and some locations may be invisible from all views – these locations are filled in via a final in-painting step that is not shown. For the *tgt-bwd* method, texture warping by interpolation is straightforward, but visibility determination is much more involved. If the disparity is explicitly coded at the target view, visibility information can be encoded at the same time, as in 3D-HEVC. However, if one wants to avoid coding disparity information at every target view, a key challenge in using the *tgt-bwd* method is to infer visibility information. The *ref-fwd* method makes visibility determination simple, even if somewhat unreliable, identifying the visible pixels as those to which a nearest neighbor splat occurs. In the backward warping of *ref-bwd*, visibility is determined during the construction of the backward pointing disparity map $D_{R_k}^{-1}$ from D_{R_k} .

While the view prediction methods summarized here are not exhaustive, they form an important basis for investigating different disparity synthesis strategies.

Tgt-bwd and *ref-bwd* should have equivalent performances in view prediction if the reference-anchored and target-anchored disparities employed are all consistent. With reference to Fig. 2, this consistency property means that $D_T(\mathbf{x}) = D_{R_1}^{-1}(\mathbf{x}) = D_{R_2}^{-1}(\mathbf{x})$ for all pixel locations \mathbf{x} in the target view V_T that are visible from both V_{R_1} and V_{R_2} . In the next section we describe base-mesh models which ensure this consistency. The base-mesh approach proposed in this paper can be considered a *tgt-bwd* method, where visibility determination is efficiently carried out at all views within a GOV through a consolidated base model. Additionally, *Ref-fwd* and *ref-bwd* can be compared with the same disparity maps, even if inconsistent, to explore the benefits of backward warping in comparison with forward warping of texture.

III. UNIFIED BASE-MESH MODEL

In this paper, we propose to decompose the whole HDCA dataset into groups of views (GOV) and associate a communicated disparity map with each such GOV. A unified disparity model is built for the description of geometric flows across the whole GOV. Synthesized disparities within the GOV are derived as needed from this single model, being consistent with each other. The model itself is constructed by starting with the original communicated disparity map and progressively introducing new elements only as required to provide complete disparity maps for target view prediction operations. Our model is represented in the form of a multi-layered triangular mesh where new mesh elements that are progressively introduced constitute underlying layers in the regions of dis-occlusion.

A. Multilayered Base-mesh Model Construction

Our proposed mesh model, written as \mathcal{M} , is a collection of triangular mesh elements, whose nodes carry spatial locations and associated disparity values. Spatial locations are defined on the so-called *base-view* V_B associated with the corresponding disparity map D_B .

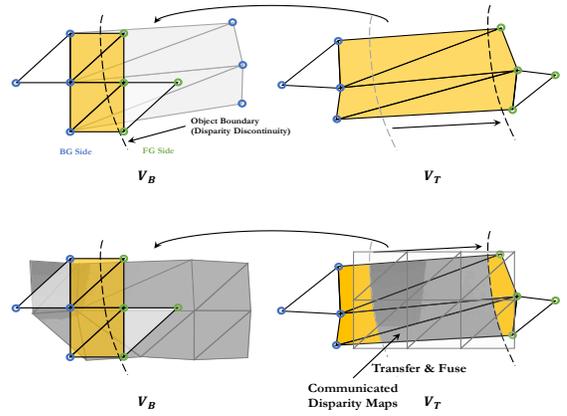


Fig. 3. Backfill approach (top), novel disparity fusion approach (bottom)

When the mesh model is projected from the base-view to other views, regions of dis-occlusion (holes) arise naturally in the vicinity of object boundaries. These holes are identified by the mesh elements in the base-view whose edges straddle the object boundaries (i.e., disparity discontinuities). Ideally, these edges have vanishing length in the base-view, undergoing infinite expansion when projected between views, as depicted in Fig. 3a. Accordingly, we denote these as ∞ elements.

While the ∞ elements do provide projected disparity values within holes in the target view, this disparity information is physically meaningless. To address this, we introduce new mesh elements to cover the exposed ∞ regions in the target view V_T , mapping these back to the base-view V_B where they contribute new underlying layers to the base-mesh model \mathcal{M} .

More formally, the base model is initialized to \mathcal{M}^0 using the base-map D_B . Then target views to be predicted within the GOV are visited in a prescribed order, starting from the coarsest level of the prediction hierarchy and working down, and within each level visiting views in order of decreasing angular distance from the base-view. Writing V_{T_l} , $l=1, 2, \dots$ for the l^{th} such prediction target, the exposed ∞ regions within V_{T_l} , denoted as Ω_l , are covered with a new set of mesh elements, denoted by $\mathcal{B}(\Omega_l)$, where \mathcal{B} is an operator that synthesizes new mesh elements. These are mapped back to the base-view using the disparity $D_{T_l}(\Omega_l)$ within regions covered by newly synthesized base-mesh elements, so that the model update can be expressed as

$$\mathcal{M}^l = \mathcal{M}^{l-1} \cup \mathcal{W}_{D_{T_l}(\Omega_l)}(\mathcal{B}(\Omega_l)) \quad (2)$$

with $\mathcal{M} = \lim_{l \rightarrow \infty} \mathcal{M}^l$.

We associate a layer-id l with each element in the base model, where $l=0$ corresponds to the original mesh elements derived from D_B and subsequent layer-ids l correspond to the new elements derived from Ω_l , with ∞ elements notionally assigned the layer-id $l = \infty$. Whenever the base-mesh model is warped to a view V_{T_l} , many mesh elements may map to any given location within that view. We resolve such multiple mappings by assigning absolute priority to those that arise from mesh elements with lower layer-id l , and where there are multiple mappings with the same layer-id, we adopt the

one that yields the larger disparity, since larger disparity is associated with locations closer to the camera array. This policy ensures that all mesh elements that contribute to the synthesized disparity map at target view V_{T_l} (in the well-defined sequence described above) necessarily come from \mathcal{M}^l , so the base-mesh model only needs to be partially constructed to render views at a given level in the hierarchy.

The following sub-sections describe two different operators \mathcal{B} for deriving novel mesh elements in dis-occluded regions.

B. Backfill Approach

Our first approach relies upon filling dis-occluded regions using extrapolation. Specifically, each mapped ∞ element in the target view V_{T_l} generates a new so-called *backfill* element, whose node locations within the target view correspond exactly to those of the mapped ∞ element, so that the backfill elements are guaranteed to cover the dis-occluded region Ω_l , as shown in Fig. 3a. The nodes of each backfill mesh element are of two types. Background nodes are assigned the same disparities as those of the mapped ∞ element. The remaining nodes, identified as *backfill nodes* are assigned extrapolated disparity values derived via a background splatting procedure. Details of this procedure and the methods used to distinguish between background and back-fill nodes may be found in [8].

C. Novel Disparity Fusion Approach

The backfill approach can only produce smooth extrapolated background, which may be unreliable for large holes arising at coarser levels, and especially near view boundaries. We propose a novel disparity fusion approach that takes advantages of other communicated disparity maps to infer disparities in regions of dis-occlusion.

We transfer all the communicated disparity maps onto Ω_l in the target view V_{T_l} , taking the median of these transferred disparities to produce a fused result $D_{T_l}(\Omega_l)$. As explained earlier, $\mathcal{B}(\Omega_l)$ is required provide a description of the disparity over Ω_l in terms of triangular mesh elements. For this reason, we follow the median-based fusion with a meshification step, creating a hierarchical variable sized mesh to approximate $D_{T_l}(\Omega_l)$. Fig. 3b shows how this approach is able to introduce important geometric features into regions that are occluded relative to the base-view.

Note that $\mathcal{B}(\Omega_l)$ may introduce new ∞ elements, corresponding to newly exposed object boundaries. These new ∞ elements may expose new holes in subsequent target views $V_{T_p}, p > l$, leading to further augmentation of the base model with mesh elements $\mathcal{B}(\Omega_p)$ having underlying layer-id p .

IV. EXPERIMENTS EXPLORATION

We carry out our experiments using the same HDCA datasets [9] and estimated disparity maps [10] as employed by the JPEG-Pleno standardisation activity. Specifically, we use the full HD resolution sets known as Sets 2, 6 and 9 and focus our attention on a subset of their input view arrays, these being the central 49×13 views of the original HDCA.

TABLE I
COMPARISON OF PREDICTED VIEWS IN AVERAGE PSNR_Y AND SSIM_Y

		<i>ref-fwd</i> <i>WaSP</i>	<i>ref-bwd</i> <i>ind</i>	<i>tgt-bwd</i> <i>bm-xtr</i>	<i>tgt-bwd</i> <i>bm-fus</i>
Set2	PSNR _Y	40.1423	40.5239	39.8062	40.7473
	SSIM _Y	0.9836	0.9852	0.9847	0.9861
Set6	PSNR _Y	28.0671	28.1335	27.9664	28.1400
	SSIM _Y	0.9132	0.9202	0.9178	0.9208
Set9	PSNR _Y	36.1778	36.8223	34.7901	36.8560
	SSIM _Y	0.9786	0.9831	0.9818	0.9835

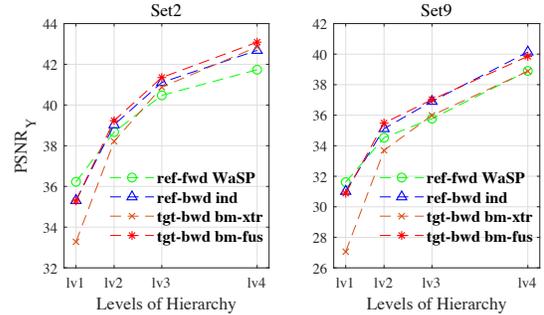


Fig. 4. Average PSNR_Y at level of hierarchy for Sets 2, 6, 9. Here, the horizontal axis reflects the cumulative number of predicted views at each successive level in the hierarchy.

We consider the hierarchical view prediction framework described in section II. Level 0 consists of the 4 corner views, with explicitly communicated disparity maps; level 1 includes 5 new predicted views; etc. Continuing the hierarchical prediction for 5 levels horizontally and 3 vertically yields an array of 17×5 original and predicted views. Specifically, we have 5, 16, 20 and 40 views predicted in each successive stage of the hierarchy. The full array of 49×13 views can be reached by adding further levels of hierarchical prediction, but this is less interesting for our current study, because inter-view displacements and regions of dis-occlusion become very small.

For our proposed base-mesh models, the view array is decomposed into 4 GOVs arranged to cover the quadrants, each associated with the disparity map at a corner of the array. All 4 GOVs overlap at the center of the array, one of which is illustrated in Fig. 1. It is important to highlight that our proposed method does not depend upon the arrangement of GOVs and placement of disparity maps within the view array.

We explore four different view prediction approaches. First is the view prediction employed by WaSP [2] that forward warps the texture based on independently synthesized disparity at each reference view, referred to as the *ref-fwd WaSP*¹. The second approach backward warps the texture based on independently synthesized disparity at reference views, whose visibility information is resolved by our mesh-based reasoning, named as the *ref-bwd ind* approach. The third and fourth approaches adopt our proposed unified base-mesh model for

¹The WaSP framework is more commonly used with 4 reference views for each predicted target view, whereas the arrangement in Fig. 1 involves only 2 references for some target views; however, in our experiments, a WaSP did not perform better with 4 reference views.



Fig. 5. Visual comparison between *ref-bwd ind* (left two columns) and *tgt-bwd bm-fus* (right two columns)

disparity synthesis, where the model is completed with either backfilling or the novel disparity fusion scheme respectively, written in short as *tgt-bwd bm-xtr* and *tgt-bwd bm-fus*.

Since we only focus on investigating the quality of predicted views other than introducing the added complications of quantization and coding, all the reference views used at each level of hierarchy are original images without any degradation.

PSNR_Y(dB) and SSIM_Y of predicted views are reported in Table I. Both the *tgt-bwd bm-fus* and the *ref-bwd ind* approaches that employ the backward texture warping give better results compared with the *ref-fwd WaSP* approach; in fact, the proposed *tgt-bwd bm-fus* method consistently outperforms the others. The clear superiority of the *ref-bwd ind* approach over *ref-fwd WaSP*, highlights the benefits of backward warping in comparison to forward warping of texture.

Overall averaged PSNR_Y is broken down level by level, as plotted in Fig. 4. We can see that both the *ref-fwd WaSP* (green) and the *ref-bwd ind* (blue) approach suffer degradations from level 2 onwards. This phenomenon reveals the adverse impact of independently warping texture from various reference locations using incompatible disparities maps. At the finest levels of the hierarchy, this inconsistency becomes less apparent only because inter-view displacements becomes small in general.

While the *tgt-bwd bm-fus* (red) approach does not yield dramatic improvements in objective quality compared with the *ref-bwd ind* (blue) method, we note that the *ref-bwd ind* method is not actually used in practice, because it introduces a high computational cost, requiring disparity fields to be synthesized by median consensus at the pixel level in each employed reference view and then transported to the target view using mesh-based reasoning. The proposed *tgt-bwd bm-fus* method, by contrast, progressively builds a single consistent mesh for each GOV. Furthermore, Fig. 5 demonstrates the visual improvements that can arise from consistently synthesized disparity. As can be seen, the *ref-bwd ind* method can result in object boundary misalignment or toggling between different decisions in inferring disparities for regions of dis-occlusion, producing ghosting or blurring artifacts in predicted views.

Fig. 6 reveals the substantial improvements made by the proposed novel disparity fusion approach *tgt-bwd bm-fus*, in comparison with the previous backfill scheme *tgt-bwd bm-xtr*

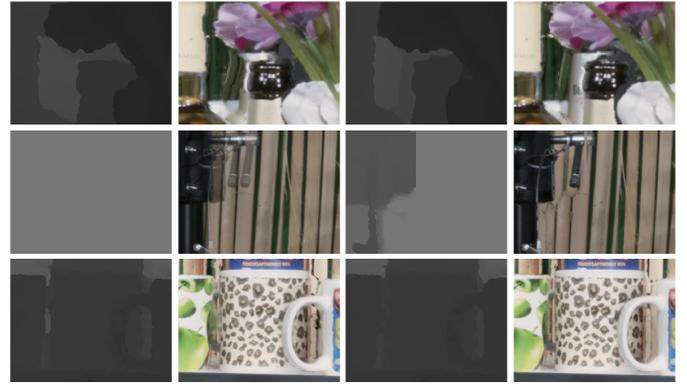


Fig. 6. Visual comparison between *tgt-bwd bm-xtr* (left two columns) and *tgt-bwd bm-fus* (right two columns)

[8]. Middle row of Fig. 6 is cropped from the edge of the view, indicating that disparities outside of the base-view boundary can be correctly inferred using our proposed approach.

V. CONCLUSIONS

The first conclusion from our work is that backward texture warping is preferable to forward warping. However, if the views involved in a prediction step do not have explicitly communicated disparity, backward warping requires mesh-based reasoning to correctly resolve visibility; this is perhaps one reason why most existing schemes employ forward warping. Our unified base-mesh model addresses these difficulties by consistently synthesizing required disparities over a large group of views. The disparity fusion approach proposed in this paper for base model dis-occlusion handling substantially improves view prediction performance, surpassing that of WaSP, which is the framework currently used in the evolving JPEG-Pleno lightfield coding standard.

REFERENCES

- [1] “Jpeg pleno call for proposals on light field coding.” 74th Meeting, Geneva, Switzerland, Jan. 2017, ISO/IEC. JTC1/SC29/WG1N74014.
- [2] P. Astola and I. Tabus, “Wasp: Hierarchical warping, merging, and sparse prediction for light field image compression,” in *2018 7th European Workshop on Vis. Inf. Proc. (EUVIP)*. IEEE, 2018, pp. 1–6.
- [3] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee *et al.*, “3d high-efficiency video coding for multi-view video and depth data,” *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3366–3378, 2013.
- [4] X. Huang, P. An, F. Cao, D. Liu, and Q. Wu, “Light-field compression using a pair of steps and depth estimation,” *Optics express*, vol. 27, no. 3, pp. 3557–3573, 2019.
- [5] X. Huang, P. An, L. Shen, and R. Ma, “Efficient light field images compression method based on depth estimation and optimization,” *IEEE Access*, vol. 6, pp. 48 984–48 993, 2018.
- [6] L. Tran, R. Khoshabeh, A. Jain, C. Pal, and T. Nguyen, “Spatially consistent view synthesis with coordinate alignment,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 905–908.
- [7] J. Pearson, M. Brookes, and P. L. Dragotti, “Plenoptic layer-based modeling for image based rendering,” *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3405–3419, 2013.
- [8] D. Rufenacht, A. T. Naman, R. Mathew, and D. Taubman, “Base-anchored model for highly scalable and accessible compression of multiview imagery,” *IEEE Transactions on Image Processing*, 2019.
- [9] “Jpeg pleno call for proposals submission process details.” 75th Meeting, Sydney, Australia, Jan. 2017, ISO/IEC. JTC1/SC29/WG1N75024.
- [10] “Unsw depth reciprocal fields for the hdca dataset.” JPEG Pleno AhG Meeting, Rio, Jan. 2018, ISO/IEC. JTC1/SC29/WG1M78000.