Tree-structured algorithm for efficient shearlet-domain light field reconstruction

S. Vagharshakyan, R. Bregović and A. Gotchev, "Tree-structured algorithm for efficient shearlet-domain light field reconstruction," *Proc. GSIP 2015: Symposium on Signal Processing on Graphics Processing Units and Multicores*, Orland, Florida, Dec. 2015

TREE-STRUCTURED ALGORITHM FOR EFFICIENT SHEARLET-DOMAIN LIGHT FIELD RECONSTRUCTION

Suren Vagharshakyan, Robert Bregovic, Atanas Gotchev

Department of Signal Processing, Tampere University of Technology, Tampere, Finland

ABSTRACT

This article considers techniques for accelerating a light field reconstruction algorithm operating in shearlet domain. In the proposed approach, an independent reconstruction of epipolar images (EPIs) is replaced with a consecutive tree-structured reconstruction. It aims at decreasing the number of iterations necessary for an EPI reconstruction by using already processed EPIs as initial values in the reconstruction stage. Two algorithms for structuring such processing trees are presented. The reconstruction performance of the proposed algorithms is illustrated on a real dataset. The underlying differences between the algorithms are discussed and numerical results of computation speeds are presented.

Index Terms— Light field, sparse reconstruction, shearlet, image based rendering

1. INTRODUCTION

One of the fundamental problems related to visualization of 3D content on a 3D display is the rendering of arbitrary views out of a given, typically small, set of images (views) that are acquired by a sparse set of cameras. This is referred to as image based rendering (IBR) and is used in various applications such as refocused image generation [1], depth estimation [2], [3], novel view generation [4], holographic stereogram [5], to name a few. The existing IBR methods are based on two, fundamentally different, approaches. First approach is based on getting an explicit information about the scene geometry in the form of depth map(s) estimated from a set of given images [6], [7], [8]. The desired views are then synthesized by reprojecting the available images using the estimated scene geometry [9], [10]. Second approach is based on the light field (LF) concept [11], [12] in which each pixel of the available views is considered as a sample in a multidimensional LF function. In this case the problem of novel view synthesis can be considered as a problem of reconstructing the continuous multidimensional LF function from discrete samples. In general, the generation of novel view is prone to ghosting type of distortions and their handling requires a large number of images [13]. It has been shown in [14] that novel views of a sufficient quality can be synthesized by a simple linear interpolation if the LF is sampled such that the disparity between nearby views is always less than one pixel. Hereafter we will call such sampling, *densely sampled* LF. Capturing a densely sampled LF would require an impractical amount of cameras, with number of cameras being related to the camera resolution and scene depth [14]. A more practical way is to find a generic domain for LF representation, which allows reconstructing the densely sampled LF out of LF samples taken by several cameras only (i.e. a coarse set of cameras).

Recently, we proposed a method for reconstruction of a densely sampled LF that utilizes the sparse representation of EPIs in shearlet domain [16]. In that method, the available data (captured views) are interpreted as known rows in the EPI's. By applying an iterative inpainting technique on every EPI, we were able to reconstruct all unknown samples. Each EPI has been processed separately and independently. Since the number of EPIs is equal to the number of rows in the image, the overall method is computationally demanding.

In this paper we address the computational optimization of the above approach and show that by proper hierarchical ordering of EPIs for their further parallelized processing, one can achieve considerable computational savings and at the same time improve the reconstruction quality.

2. EPI RECONSTRUCTION IN SHEARLET DOMAIN

Consider a camera moving with a dense sampling step in horizontal direction, thus creating a set of perspective images. Stacking all images and taking a slice along view dimension creates an EPI - a regular structure of stripes with different thickness and slope [15]. Fig. 1(a) shows the stack of perspective images. A slice of it (e.g. the yellow line) results in an EPI Fig. 1(b)). where each line corresponds to a scene point with the line slope being inversely proportional to the distance of the point to the camera plane (assuming only Lambertian reflectance in the 3D scene). In the spectral domain, a line at constant depth is mapped to a line passing through the origin. Consequently, the spectrum of an EPI has a bow-tie type shape and in the case of a densely sampled LF is always contained in a region similar to the one highlighted in Fig. 1(c) [16]. In Fig. 1(b) the dense EPI is superimposed with green lines showing the available measured lines corresponding to input camera views. The goal is to reconstruct the dense LF in EPI domain out of the

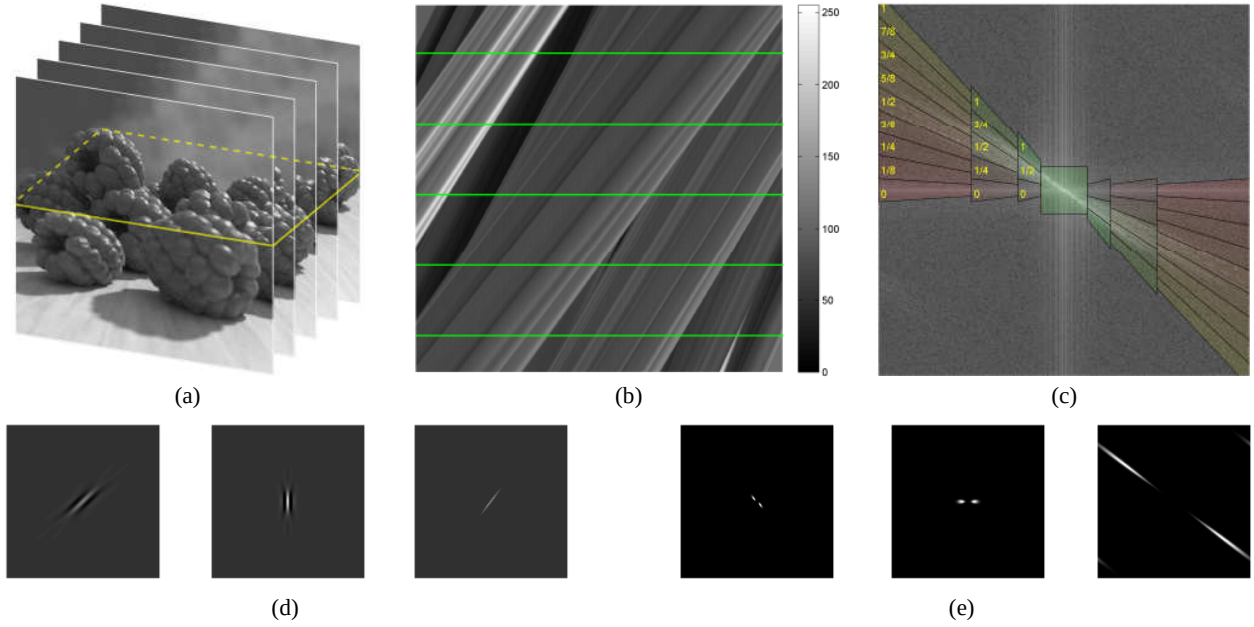

Figure 1. (a) Example of scene image.(b) Example of densely sampled light field EPI corresponding to row highlighted in yellow in (a). (c) Frequency domain characteristics of EPI with desirable frequency domain truncation, presented in 3 scales and central low pass filter with disparity values of corresponding shears. (d) Example of several constructed shearlet atoms in spatial and (e) frequency domains.

given measured lines. Based on the spatial and spectral properties of EPIs the reconstruction is done in shearlet domain. The shearlet analysis and synthesis frames are formed by elements being translation-invariant functions indexed by scale and shear (direction) indexes. Their frequency support allows tiling the EPI spectrum for some finite number of scales and directions as shown in Fig. 1(c). We favor the use of compactly supported shearlets [17], which have compact support in spatial domain as illustrated in Fig. 1(d, e) and reduce the ringing artifacts during reconstruction. While constructing a frame of compactly supported shearlets, different directional filters for different scales have been used for improving the directional properties at lower scales [16], [17], [18].

Denote by $f \in \mathbb{R}^{N\times N}$ the unknown complete EPI matrix, where each row represents a corresponding image row and denote by $g \in \mathbb{R}^{N\times N}$ the decimated EPI where rows from available camera views form the input and the unknown rows are set to 0. The two two-dimensional matrices f and g are column-wise reshaped into $\mathbb{R}^{N^2}$ vectors and the same notations f and g are kept for those. The binary matrix $H \in \mathbb{R}^{N^2\times N^2}$ where $H(i,i) = 1$ if $g(i) \neq 0$ and 0 otherwise, determines the available measurements. The analysis and synthesis matrices of the shearlet frame are denoted by $S \in \mathbb{R}^{M\times N^2}$ and $S^* \in \mathbb{R}^{N^2\times M}$, respectively, where $M = \eta N^2$ and η is the number of all shears in all scales of the shearlet. The reconstruction of unknown rows of g is formulated under the prior condition for having sparse solution in the shearlet domain, i.e.

$$\min_{f\in\mathbb{R}^{N^2}} \|Sf\|_0, \text{ subject to } g = Hf \tag{1}$$

The problem (1) is solved through the following iterative thresholding algorithm [20]:

$$f_{n+1} = S^*\left(H_{\lambda_n}\left(S(f_n + \alpha(g - Hf_n))\right)\right) \tag{2}$$

where $H_\lambda(x) = \begin{cases} x, |x| \geq \lambda \\ 0, |x| < \lambda \end{cases}$, is a hard thresholding operator and α is a chosen relaxation parameter. The initial value f_0 is set to 0 everywhere and the thresholding parameter λ_n is set to decrease with the iteration number, in our case further on we assume will assume that λ_n linearly decreasing from λ_{max} to λ_{min} over L iterations, $n = 0, \dots, L$.

3. ORGANIZING EPI PLANES IN TREES

3.1. Choice of optimal parameters

The performance of the algorithm presented in Section 2 depends on the selection of some parameters, e.g. the thresholding parameters λ_n and the number of iterations L. In [16], it was assumed that each EPI is reconstructed independently from others, using optimal parameters $\left(L^{opt}, \lambda_{max}^{opt}, \lambda_{min}^{opt}\right)$. Under this assumption, the algorithm is highly parallelizable provided a large number of GPUs is available. For independent processing, good results for all EPIs are obtained without providing any specific initial EPI estimate. The thresholding parameters have been selected based on exhaustive search $\left(\lambda_{max}^{opt}, \lambda_{min}^{opt}\right)$ and in general, a

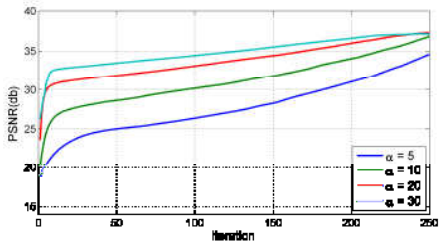

Figure 2. Example of EPI reconstruction convergence for different values of parameter α.

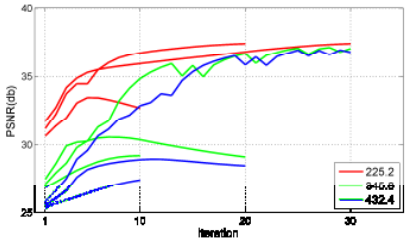

Figure 3. Example of EPI reconstruction convergence for different number of iterations and initial estimates. Different colors represent different nearby EPIs used as initial estimate.

higher number of iterations always give a better result. Still, some optimal value L^{opt} has to be chosen depending on the available computing resources, the targeted reconstruction quality. Fig. 2 shows the reconstruction quality during iterative reconstruction for different parameter α of the algorithm for an EPI from the Teddy dataset [19] validated over the even indexed views and taking odd indexed views as input for the algorithm. One can see that by using a proper selection of parameters the algorithm performance converges faster, however algorithm convergence instability should be considered for big value of α.

3.2. Accelerated reconstruction

In order to accelerate the EPI reconstruction, one can attempt starting the reconstruction with a good initial EPI estimate, which would decrease the number of iterations along with the initial threshold λ_{max}. Applying a very high λ_{max} would decrease the influence of the initial estimate on the reconstruction result. We assume that the initial (maximum) thresholding parameter is selected as linearly dependent on the number of iterations

$$L < L^{opt}, \qquad \lambda_{max} = \frac{L}{L^{opt}}\left(\lambda_{max}^{opt} - \lambda_{min}^{opt}\right) + \lambda_{min}^{opt} \qquad (3)$$

Fig. 3 quantifies the quality of an EPI reconstruction from the *Teddy* dataset using reconstruction results of three neighbor EPIs as initial estimates. The distances between EPIs are calculated using l_2 norm and are given in the bottom right corner of the figure. Reconstructions for three different number of iterations $L = 10, 20, 30$ are shown, with $L^{opt} = 100$. Taking reconstruction result of closer EPI as an initial estimate proves to be beneficial. The tendency

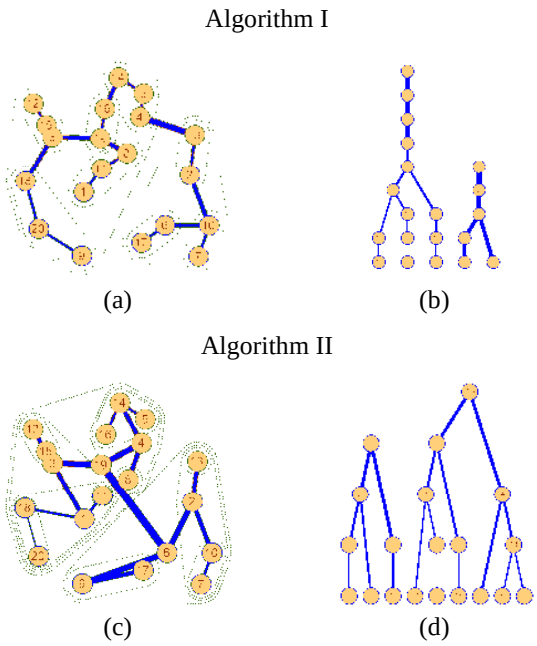

Figure 4. Illustration of clustering and processing trees. (a, b) Graph construction using Algorithm I and corresponding processing tree. (c, d) Graph construction using Algorithm II and corresponding processing tree.

demonstrated in Fig. 3 motivates us to attempt LF reconstruction in a consecutive manner such that already reconstructed EPIs are used as initial estimates for reconstructing the remaining ones. Such consecutive processing can be performed over a properly constructed tree where each node represents an EPI to be reconstructed with its initial estimate determined by the EPI at the parent node. The necessary number of iterations was chosen to be linearly depends on the distance between EPIs $d < D$, where D is an upper bound beyond which a parent EPI cannot be used as an initial estimate for its offspring, thus $N = N^{opt}\min(d, D)/D$. Following the above discussion, we propose two *tree-structured* algorithms.

3.3. Algorithm I

In this algorithm we attempt to minimize the distance between already processed and unprocessed EPIs. Assume that an EPI is a node of a set and we have found the distance between every pair of nodes. A graph is constructed to define the optimal processing order of the EPI nodes. The closest node to every node is found first and those are connected. As a result the whole set is clustered in small groups. Fig. 4 (a, b) illustrates the construction of such a graph and its corresponding processing tree. In Fig. 4(a) an example of small grouping is given by the nodes indexed as (12, 15, 3) and (1, 11, 8). Subsequently, we connect the closest pair for each small group with the nodes out of that group. By iterating the process we eventually get a connected graph. Two processing trees' roots are selected

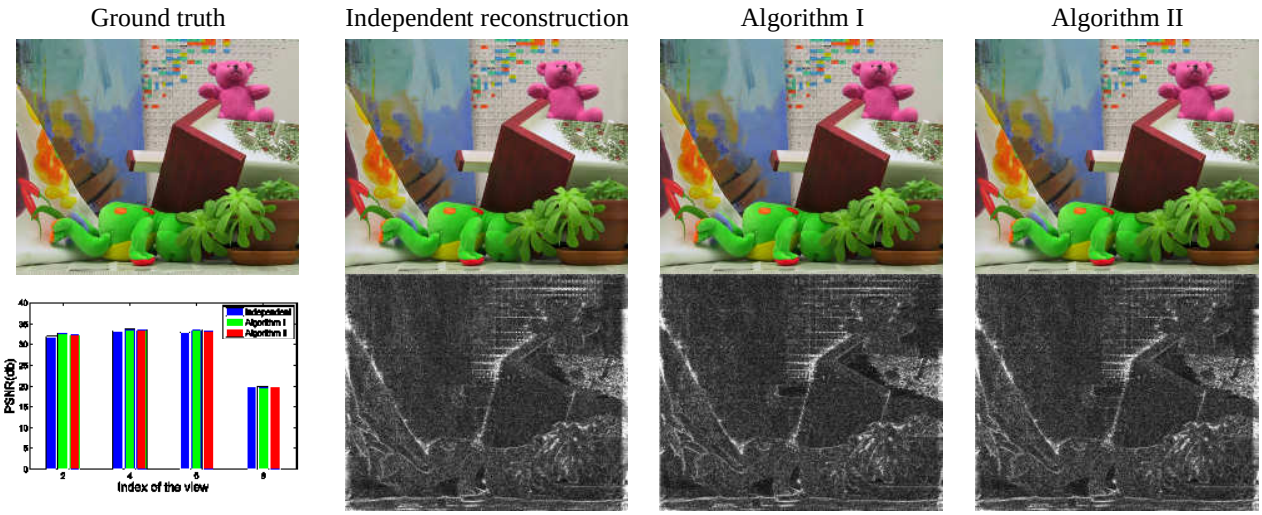

Figure 5. Top row is an example of reconstruction for three different algorithms. Bottom row presents the PSNR between GT and reconstructed images along with corresponding grayscale difference image (brighter means higher error).

by the graph elements, which were connected last (in Fig. 4(a) those are nodes 4 and 13). The trees are constructed by simply following over the graph starting from the root nodes. As a result, one gets a processing order similar to the one in Fig. 4(b).

3.4. Algorithm II

The algorithm presented above is clearly content-dependent and a parallel processing with an arbitrary number of M computing units might not be directly implementable. Therefore, we propose another algorithm, which naturally distributes the processing effort among the desired numbers of parallel branches. Let us consider the case of two processing units. At the first step we split the whole set of EPIs into two groups using a K-means clustering algorithm [21]. In each group, we find the node, which is the closest to the group center and mark it as root for all nodes in the group. Referring to the example in Fig. 4 (c, d), nodes 6 and 19 are selected as roots. The same dichotomy splitting is applied on each so-constructed group, removing the already selected parent nodes. Following this simple subdivision, after some number of separations, one constructs trees similar to the ones in Fig. 4(d). The method can be easily generalized for M divisions at each step. Fig. 4(e) presents the tree structure for a set of EPIs from the Teddy dataset, which determines the order of reconstruction.

4. EVALUATION

The performance of the proposed algorithms is quantified by calculating the number of iterations required to reconstruct the densely sampled LF taking into account that the necessary number of iterations is decreased depending on the distance between subsequently processed EPIs. For the illustrative example of the view reconstruction of Teddy dataset we choose an upper bound of distance between EPIs $D = 1500$. Table I summarizes the speed up for different color channels assuming that the processing of all EPIs independently takes a unit time.

Table I

	R channel	G channel	B channel
Algorithm I	0.622	0.652	0.652
Algorithm II	0.758	0.782	0.774

Among the considered three approaches, algorithm I is fastest yet difficult to parallelize, because the constructed processing order dependents on the given dataset. Algorithm II allows a high flexibility in parallelization for a lower speed compared to Algorithm I. Results of view reconstruction are presented in Fig. 5. As one can see in the figures, the difference between reconstruction qualities for different reconstruction approaches is negligible.

5. CONCLUSION

In this article, we presented two algorithms for accelerating the LF reconstruction method introduced in [16]. In the presented algorithms, the EPIs are ordered into tree structures. Each EPI, instead of being processed independently, uses information from already processed EPIs as initial value for the reconstruction thereby reducing the computational complexity required to reconstruct the EPI. Beside increasing the speed of LF reconstruction, it turned out that the reconstruction quality also increased. Although that increase is not remarkably high, it hints that it is possible to further improve the reconstruction quality by utilizing the interaction between individual EPIs. This is a topic to be considered in future work.

REFERENCES

[1] R. Ng, "Fourier slice photography," in *ACM Transactions on Graphics (TOG)*, 2005, pp. 735-744.

[2] I. Tosic and K. Berkner, "Light field scale-depth space transform for dense depth estimation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2014, pp. 441-448.

[3] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 606-619, 2014.

[4] M. Tanimoto, "Overview of free viewpoint television," *Signal Process Image Commun*, vol. 21, pp. 454-461, 2006.

[5] J. Jurik, T. Burnett, M. Klug and P. Debevec, "Geometry-corrected light field rendering for creating a holographic stereogram," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2012, pp. 9-13.

[6] A. Gelman, P. L. Dragotti and V. Velisavljevic, "Multiview image compression using a layer-based representation," in *17th IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 493-496.

[7] J. Berent, P. L. Dragotti and M. Brookes, "Adaptive layer extraction for image based rendering," in *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*, 2009, pp. 1-6.

[8] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung and M. H. Gross, "Scene reconstruction from high spatio-angular resolution light fields." *ACM Trans. Graph.*, vol. 32, pp. 73, 2013.

[9] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Electronic Imaging 2004*, 2004, pp. 93-104.

[10] M. Li, H. Chen, R. Li and X. Chang, "An improved virtual view rendering method based on depth image," in *2011 IEEE 13th International Conference on Communication Technology (ICCT)*, , 2011, pp. 381-384.

[11] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, 1996, pp. 31-42.

[12] S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, 1996, pp. 43-54.

[13] Z. Lin and H. Shum, "A geometric analysis of light field rendering," *International Journal of Computer Vision*, vol. 58, pp. 121-138, 2004.

[14] J. Chai, X. Tong, S. Chan and H. Shum, "Plenoptic sampling," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, 2000, pp. 307-318.

[15] R. C. Bolles, H. H. Baker and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, pp. 7-55, 1987.

[16] S. Vagharshakyan, R. Bregovic, A. Gotchev, “Image based rendering technique via sparse representation in shearlet domain,” in *ICIP 2015* (accepted for publication).

[17] P. Kittipoom, G. Kutyniok and W. Lim, "Construction of compactly supported shearlet frames," *Constructive Approximation*, vol. 35, pp. 21-72, 2012.

[18] G. Kutyniok, W. Lim and R. Reisenhofer, "Shearlab 3D: Faithful digital shearlet transforms based on compactly supported shearlets," *ArXiv Preprint arXiv:1402.5670*, 2014.

[19] D. Scharstein and R. Szeliski, “High-accuracy stereo depth maps using structured light,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, vol. 1, pp. 195-202, Madison, WI, June, 2003.

[20] M. Elad, J. Starck, P. Querre and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and Computational Harmonic Analysis*, vol. 19, pp. 340-358, 2005.

[21] Lloyd, P. Stuart, "Least Squares Quantization in PCM." *IEEE Transactions on Information Theory*. vol. 28, pp. 129–137, 1982.