

A NOVEL CONVEX AUTOREGRESSIVE MODEL FOR LIGHT FIELD DENOISING ON RIEMANNIAN SPACE

Mansi Sharma, Rohan Lal

Department of Electrical Engineering, Indian Institute of Technology Madras, India

ABSTRACT

Existing light field cameras are susceptible to produce low-quality results as sensor noise can dominate measurements. Thus, denoising is a critical step in the light field (LF) subsequent analysis and processing. This paper presents a novel LF denoising framework based on an adaptive parallax and auto-regressive model analysis. The novel procedure first creates a view-dependent LF stack by compensating parallax variation, employing an extended variational flow technique on a set of LF intensity and depth features. Further, it takes advantage of the spatial similarity across the registered LF stack and reduce the noisy observations. The output is, further, improved by formulating the denoising as a novel adaptive autoregressive (AR) stochastic problem. The proposed convex AR model averaged view-specific spatial energies of stacked LF images on Riemannian manifolds by a depth-directed maximum likelihood AR parameter estimation process. Lastly, scale intensity of refined AR LF predicted view by the average intensity of the superpixel in each LF stacked image. The experiments show that proposed AR LF denoiser outperforms standard algorithms in terms of visual quality and in the preservation of parallax details.

Index Terms — Light field, denoising, auto-regressive, Riemannian manifold, parallax analysis, variational flow estimation

1. INTRODUCTION

Imaging with micro-lens arrays is challenging as sensor noise introduced during the acquisition process and sparse sampling negatively interfere with the quality of raw LF outputs. Therefore, denoising is a mandatory procedure on the LF analysis pipeline.

The motivation of this work comes from Fu et al. [1] work that exploits high correlation across spectra, sparsity across the spatial-spectral domain, non-local self-similarity over space to denoise the hyperspectral images (HSIs). They formulated HSIs denoising as a spatial-spectral dictionary learning problem using an iterative numerical technique. However, the naive approach to LF denoising employ independent image denoisers on LF sub-aperture images (LF-SAIs) [2]. The standard denoiser BM3D [3] seek signal redundancy among non-locally matched groups of 2D image samples. Though it is one of the best image denoisers, but gives sub-optimal results with light field due to unexplored strong correlations embedded in the angular domain of LF-SAIs. Thus, we harness correlations across the light field images by structuring the aforementioned observation spaces as Riemannian manifolds and formulating denoising as an AR prediction process. This can significantly improve the reduction of noise. Certainly, it should be noted that aspects of light field data and hyperspectral images differ significantly. Different from hyperspectral images, noise notably suppresses anisotropic parallax characteristics for LF structural synthesis [11]. Thus, two major challenges we addressed are: 1) how to account perspective information of scenes among

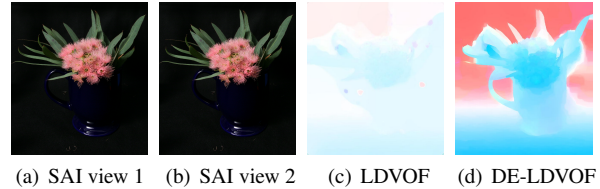


Figure 1: Comparison between 2D optic flow (LDVOF) [6] and proposed 3D scene flow method (DE-LDVOF) for light fields.

different SAIs of a LF ? 2) how to effectively recover the high frequency perspective components of a LF from noisy measurements ? A simple formulation by aggregating and averaging like monochromatic hyperspectral images over superpixels of mean band will not work with LF sub-aperture images (LF-SAIs) [4]. It will not account directional measurements of the focusing light rays and cancel out the noise pixels. The proposed scheme accounts both spatial-angular structures and redundancies of the LFs while denoising. The three major novel contributions of this work are: 1) an end-to-end processing scheme for LF denoising, 2) an anisotropic parallax analysis based on a 3D variational scene flow model, 3) an AR model adaptive to the local geometry structures and characteristics of accompanying LF images for recovering denoised outputs from lower quality imperfect inputs.

2. PROPOSED ADAPTIVE PARALLAX AND AUTO-REGRESSIVE LF DENOISE MODEL

The workflow showing proposed algorithm blocks is depicted in Fig. 2. The algorithm is divided into four major components: Anisotropic parallax determination and compensating large parallax using 3D variational motion flow, as presented in **Block I** and **Block II** of Fig. 2, respectively. Further, autoregressive modelling of sparse light field for denoising and reconstruction, as shown in Fig. 2 **Block III** and **Block IV**, respectively. Each component is described in the following sections.

2.1. New Depth-enhanced Variational 3D Flow Model

Suppose we have a noisy LF observation $L = \hat{L} + N$, where L , \hat{L} , and $N \in R^{w \times h \times n_h \times n_v}$ are the noisy LF, noise-free LF, and additive Gaussian noise, respectively. The w and h specify spatial resolutions of each SAI. The n_h and n_v denote the horizontal and vertical angular dimensions. The objective is to recover \hat{L} from L . The system first estimate the depth from sparse, noisy light fields input using an algorithm proposed by Williem et al. [5]. Using geometry information, we extend the 2D large displacement variational optical flow (LDVOF) algorithm [6] to 3D depth-enhanced LDVOF algorithm (dubbed as DE-LDVOF) and predicts the parallax details for the entire LF by integrating

rich intensity and depth descriptors (**Block I** in Fig. 2). The content parallaxes are strictly linear among different SAIs of a LF. Our DE-LDVOF method gives accurate anisotropic parallax details between two SAIs, say I_i^{sa} and I_j^{sa} , by solving the following proposed variational model:

$$E = E_{intensity}(\mathbf{w}) + \alpha E_{grad}(\mathbf{w}) + \beta E_{smooth}^{cd}(\mathbf{w}) + \zeta E_{sparsematch}^{cd}(\mathbf{w}, \mathbf{w}_1) + E_{desc}^{cd} \mathbf{w}_1 \quad (1)$$

where,

$$\begin{aligned} E_{intensity}(\mathbf{w}) &= \int_c \int \Psi(|I_j^{sa}(\mathbf{x} + \mathbf{w}(\mathbf{x})) - I_i^{sa}(\mathbf{x})|^2) d\mathbf{x}, \\ E_{grad}(\mathbf{w}) &= \int_c \int \Psi(|\Delta I_j^{sa}(\mathbf{x} + \mathbf{w}(\mathbf{x})) - \Delta I_i^{sa}(\mathbf{x})|^2) d\mathbf{x}, \\ E_{smooth}^{cd}(\mathbf{w}) &= \int_c \int \Psi(|\Delta u(\mathbf{x})|^2 + |\Delta v(\mathbf{x})|^2) d\mathbf{x} + \\ &\quad \int_d \int \Psi(|\Delta u(\mathbf{d}_x)|^2 + |\Delta v(\mathbf{d}_x)|^2) d\mathbf{d}_x, \\ E_{sparsematch}^{cd}(\mathbf{w}, \mathbf{w}_1) &= \int_c \int \delta(\mathbf{x}) \rho(\mathbf{x}) \Psi(|\mathbf{w}(\mathbf{x}) - \mathbf{w}_1(\mathbf{x})|^2) d\mathbf{x} \\ &\quad + \int_d \int \delta(\mathbf{d}_x) \rho(\mathbf{d}_x) \Psi(|\mathbf{w}(\mathbf{d}_x) - \mathbf{w}_1(\mathbf{d}_x)|^2) d\mathbf{d}_x, \\ E_{desc}^{cd} \mathbf{w}_1 &= \int_c \int \delta(\mathbf{x}) |S_{f_j}(\mathbf{x} + \mathbf{w}_1(\mathbf{x})) - S_{f_i}(\mathbf{x})|^2 d\mathbf{x} + \\ &\quad \int_d \int \delta(\mathbf{d}_x) |S_{f_j}(\mathbf{d}_x + \mathbf{w}_1(\mathbf{d}_x)) - S_{f_i}(\mathbf{d}_x)|^2 d\mathbf{d}_x, \end{aligned}$$

and $\mathbf{x} := (x, y)^T$ denotes a pixel in intensity domain Ω_c , $\mathbf{d}_x := (d_x, d_y)^T$ denotes a pixel in depth image domain Ω_d , $w := (u, v)^T$ is the RGB-D scene flow field, \mathbf{w}_1 represents the correspondence vector obtained by descriptor matching at points \mathbf{x} and \mathbf{d} in intensity and depth domain respectively, δ_i is defined as 1, if there is a descriptor available at point \mathbf{x} in a SAI and at point \mathbf{d} in its associated depth map; otherwise δ_i is 0. The α, β, ζ are tuning parameters. The correspondence between SAIs and their associated depth maps is weighted by matching scores $\rho(\mathbf{x})$ and $\rho(\mathbf{d})$ in their respective domains. The matching scores are defined as suggested by [6]. The S_f denotes (sparse) fields of feature vectors in I_i^{sa} and I_j^{sa} , and their corresponding depth maps D_i^{sa} and D_j^{sa} , respectively.

The term $E_{intensity}$ matches color values, supplemented with gradient constraint E_{grad} in intensity domain Ω_c to adapt brightness changes. The color $E_{intensity}$ and gradient E_{grad} terms are defined similar to [6]. However, for effectively capture dense 3D motion, three new constraints $E_{sparsematch}^{cd}$, E_{desc}^{cd} , E_{smooth}^{cd} respectively are introduced to deal with occlusions and large displacements of fine structures in arbitrary deformations, motion discontinuities, and depth variations. The new regularity term E_{smooth}^{cd} penalize total variation of the RGB-D flow field. This is critical to account global regularity of joint RGB-D motion. The $E_{sparsematch}^{cd}$ and E_{desc}^{cd} enforce a smooth RGB-D scene flow with sub-pixel accuracy, incorporating rich descriptive features of color and depth from LFs. The $E_{sparsematch}^{cd}$ and E_{desc}^{cd} handles coarse-to-fine optimization and arbitrary variations in typical non-translational motions between SAIs in spatial extent of

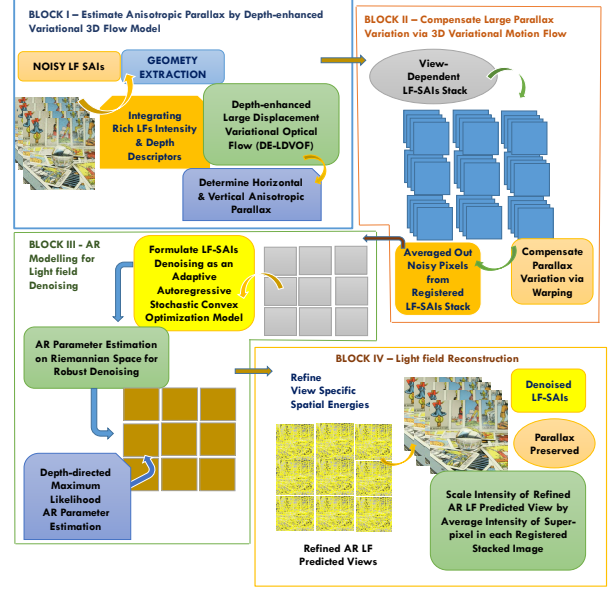


Figure 2: The workflow of proposed auto-regressive model for light fields denoising.

joint rich RGB-D descriptors. Solving proposed dense 3D variational flow model (1) on rich LFs information, noticeably produce better results than previous work, especially in computing fine-grained motion on small minute structures. A result shown in Fig. 1 makes it clear that integrating depth correspondences along with color/intensity into a variational model from descriptor matching notably improved the scene flow estimation for light fields.

2.2. Compensate Large Parallax Variation

Further, we created a view-dependent LF stack by compensating parallax variation via warping using computed 3D variational motion flow information proposed in section 2.1. We take the advantage of spatial similarity across the registered LF-SAIs stack and improve the observation by averaging out noisy pixels as

$$L_{avg}^v = \frac{1}{n} \times X_{reg}^n \quad (2)$$

where, X_{reg}^n represents noisy SAIs and superscript n indicates the SAI index. Further, we refine view-specific spatial energies of L_{avg}^v and formulate denoising in a stochastic framework on Riemannian manifolds by a depth-directed maximum likelihood AR parameter estimation process (**Block II** in Fig. 2).

2.3. New AR Model for Light field Denoising

The AR model has been applied in several practical applications like super-resolution, interpolation, time-series analysis, video coding, etc [13]. This signifies that use of AR predictors is versatile as long as the AR models be designed properly. In this paper, we attempted first at AR modeling of sparse light fields (**Block III** in Fig. 2). The following novel LF AR convex optimization model is proposed for denoising on Riemannian manifolds:

$$A_{RM}(\tilde{L}_{AR}) \triangleq \min_{\hat{L}} \lambda \times \prod_x \|(L_{avg_x}^v - \hat{L}_{avg_x}^v)\|_* + \prod_{y \in N(x)} c_{x,y} \|\hat{L}_{avg_y}^v\|_F^2 \quad (3)$$

Table 1: Light field denoising performance comparison

	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
BM3D	36.153/0.929	32.283/0.869	29.986/0.769
VBM4D	36.569/0.949	33.346/0.899	31.156/0.809
HyperFan	32.951/0.889	30.922/0.789	27.901/0.699
OUR	38.539/0.976	34.994/0.956	32.586/0.931

where, $L_{avg_x}^v$ is L_{avg}^v at location x , $N(x)$ is neighborhood of pixel x , and $c_{x,y}$ LF AR coefficient for pixel y in the neighborhood $N(x)$. The $c_{x,y}$ is computed as $c_{x,y} = \frac{1}{F_x} c_{x,y}^I c_{x,y}^D$ where, $c_{x,y}^D$ and $c_{x,y}^I$ are the depth term and the intensity term of L_{avg}^v respectively, with normalization factor F_x . We model the behaviour of proposed LF AR model (3) coefficients $c_{x,y}$ on Riemannian space for robust denoising. Modeling on non-Euclidean space is amenable to analysis and parameter estimation. Inspired from [7], an optimal $c_{x,y}$ can be found on Riemannian manifold M^{Rn} by analysing AR coefficients as a stochastic source. For simplicity, the Riemannian manifold M^{Rn} is assumed to be connected and geodesically complete. The sufficient condition for this to happen is the compactness of M^{Rn} . Mathematically, we say, for every point p in M^{Rn} , the exponential map Exp_p is defined on the entire tangent space $T_p M^{Rn}$. Thus, the AR coefficients in proposed formulation are computed as:

$$c_{x,y}^k = Exp_{c_{x,y}^{k-1}}(\Delta_k), Exp_{c_{x,y}^{k-1}}(\epsilon\Delta_k) = c_{x,y}^{k-1} + \epsilon\Delta_k, \forall \epsilon \geq 0 \quad (4)$$

where, $c_{x,y}^k$ is found by moving along the geodesic which deviates from $c_{x,y}^{k-1}$ in the direction of the tangent vector $\Delta_k \in T_{c_{x,y}^{k-1}}$.

Assuming $c_{x,y}^k$ be an AR process of order p evolving on R^n , the sequence of increments $\Delta_k = c_{x,y}^k - c_{x,y}^{k-1}$ is considered as a stochastic rule $\Delta_k = S_1\Delta_{k-1} + \dots + S_p\Delta_{k-p} + noise$, generating current increment Δ_k from the previous one. Here each S_i denotes an $n \times n$ matrix. Each increment lies in a distinct tangent space, which is computed by constructing a LF AR process $\{c_{x,y}^k\}$ on M^{Rn} as

$$\Delta_k = \prod_{i=1}^k S_i Q_{c_{x,y}^{k-i-1}, c_{x,y}^{k-1}}(Log_{c_{x,y}^{k-i-1}}(c_{x,y}^{k-i})) + V_k, \quad c_{x,y}^k = Exp_{c_{x,y}^{k-1}}(\Delta_k) \quad (5)$$

where, the Q denotes a canonical isomorphism. The V_k is identified as a random tangent vector in $T_{c_{x,y}^{k-1}} M^{Rn}$. The maximum likelihood estimate of the LF AR parameter is then obtained, given time-series $\{c_{x,y}^{k-1}, c_{x,y}^{k-2}, \dots, c_{x,y}^k\}$ as

$$\hat{c}_{x,y}^{k,ML} = arg \max_c p(c_{x,y}^{k-1}, c_{x,y}^{k-2}, \dots, c_{x,y}^k, c) \quad (6)$$

The estimated coefficients in proposed linear AR model (3) are adaptive to the local geometry and effectively signalize high correlation, sparsity across spatial-angular domain of LFs.

2.4. Light Field Reconstruction

Lastly, the proposed method scale intensity of refined AR LF predicted view \tilde{L}_{AR} by the average intensity of the superpixel in each LF registered stacked image. The complete process outputs high visual quality and preserve parallax details of reconstructed LF views (**Block IV** in Fig. 2).

3. RESULTS

We evaluated the performance of proposed AR LF denoiser and compare with the state-of-the-art LF denoisers: BM3D [3], which denoises each SAI independently; V-BM4D [8], which monitors the LF SAIs as a video sequence, and the HyperFan4D [9], which performs at the LF 4D frequency space. The testing is performed on LFs from the Lego Gantry [10], (New) Stanford Light Field Archive. The input LF data is prepared by adding additive white Gaussian noise (AWGN) of standard variance $\sigma = 10$, $\sigma = 20$, and $\sigma = 30$ on the LF SAIs. The average PSNR and Structural Similarity scores are computed between the ground truth LFs and the denoised results for different methods. The average scores estimated for 289 images of Tarot Cards and Crystal Ball LFs on a 17×17 grid are shown in TABLE 1. As it is observed, proposed AR LF shows better denoising performance at all noise levels. The performance is clearly noted under higher noise levels, *i.e.* above $\sigma = 20$. At noise level $\sigma = 30$, our model gain 2.6db and 4.6850db over BM3D and HyperFan4D respectively.

The precision recall (PR) curves are determined to give assessment on how well the parallaxes among different SAIs are preserved and restored by the proposed AR LF denoiser. The evaluation metric is designed as suggested by Chen et al. [11]. To compute PR curves, the central SAI is subtracted from all LF SAIs of Tarot Cards and Crystal Ball. The binary parallax edge map for each SAI is obtained by using different threshold values. The PR curves are plotted in Fig. 4 for all methods, comparing binary parallax edge maps between the ground truth LF and the denoised LF. As observed from the graphs, the proposed AR LF denoise model best restore the LF parallax at all noise levels.

The visual comparison results are shown in Fig. 3. The PSNR scores and SSIM (Structural Similarity Index) are also computed for the individual LFs illustrated in Fig. 3. It is clearly observed that proposed AR LF model best detach noise and maintain scene details. The details are lost and blurred out (see inside the marked regions) in the output of BM3D and VBM4D. The HyperFan4D failed to remove noise completely. The minute details of the scene also seriously degraded in the output of HyperFan4D.

4. CONCLUSIONS

We have proposed an end-to-end processing and a first-of-its-kind auto-regressive model for recovering high-quality LFs from low quality sensor measurements. The denoising problem is posed equivalent to an AR linear modeling system, and its conditioning is analyzed by a depth-directed stochastic parameter estimation process on Riemannian space. The proposed convex AR modeling framework achieves promising results on light field data.

The proposed scheme is very different from CNN based approaches like [11]. The significance over learning based denoising models is that we don't perform calculation of noise-free features by direct angular averaging of the SAIs, which is vulnerable to noise of large magnitudes. Instead, we deal with anisotropic variations among SAIs with a proposed 3D variational scene flow technique. This accounts for the high frequency perspective components and large motion. This computer vision based approach proved beneficial for registering warped SAIs and restore view dependent local energies and recreate the structural parallax details in the first step before subjecting to AR denoiser. Experiments demonstrate that our end-to-end LF processing scheme achieves high quality recovery under heavy noise degradation.

In the present framework, we made Lambertian scene assumption in analysing content parallaxes. In the future, we will further

