Steered Mixture-of-Experts Approximation of Spherical Image Data

Ruben Verhack*†, Nilesh Madhu*, Glenn Van Wallendael*, Peter Lambert*, and Thomas Sikora†
*Ghent University - imec, IDLab, Department of Electronics and Information Systems (ELIS), Belgium
†Technische Universität Berlin, Communication Systems Group, Germany

***Abstract*—Steered Mixture-of-Experts (SMoE) is a novel framework for approximating multidimensional image modalities. Our goal is to provide full Six Degrees-of-Freedom capabilities for camera captured content. Previous research concerned only limited translational movement for which the 4D light field representation is sufficient. However, our goal is to arrive at a representation that allows for unlimited translational-rotational freedom, i.e. our goal is to approximate the full 5D plenoptic function. Until now, SMoE was only applied on Euclidean spaces. However, the plenoptic function contains two spherical coordinate dimensions. In this paper, we propose a methodology to extend the SMoE framework to spherical dimensions. Furthermore, we propose a method to reduce the parameter space to the same two dimensional Euclidean space as for planar 2D images by using a projection of the covariance matrices onto tangent spaces perpendicular to the unit sphere. Finally, we propose a novel training technique for spherical dimensions based on these observations. Experiments performed on omnidirectional 360° images show that the introduction of the dimensionality-reduction projection step results in very low quality loss.**

***Index Terms*—Steered Mixture-of-Experts, 360 images, image approximation, plenoptic function**

I. Introduction

Our goal is to enable full translational-rotational navigation freedom in camera captured image content. Currently, only rotational navigation (e.g. 360° video) is widely available. This is very limited compared to the navigational freedom in computer generated image content (e.g. in gaming), in which a user is allowed the so-called *Six Degrees-of-Freedom* (6DoF), i.e. three translational movements (walking around) combined with three rotational movements (head movements). Having such freedom without the knowledge of geometrical scene information is currently a very active field of research.

At the moment, MPEG is conducting standardization efforts for a 6DoF format [1]. Their envisioned process consists of two steps: (1) find the most important views on a scene, and (2) encode these views using well-known difference and transform coding approaches. At decoder side, views are synthesized potentially by using extra transmitted geometrical side-information. However, we argue that 2D regular sampling grids are not optimal representations for storing high-dimensional data. Furthermore, we believe that the view synthesis process could shift considerable computational complexity towards the decoder.

The 2D images observed by humans at each angle are processed versions of the higher-dimensional light information in the scene. In terms of signal processing, we are presented with a high-dimensional sampling problem with nonuniform and nonlinear sample spacing and high-dimensional spatio-directionally varying sampling kernels [2]. This high dimensional space is parametrized by multiple definitions. In the case of no occlusions, i.e. the "open space" assumption, the high dimensional space can be reduced to the 4D light field [3], [4]. This is currently a widely used simplification, however, this assumption allows only limited movement. In order to allow for complete navigational freedom without assumptions on the scene, the 5D plenoptic function is necessary [5].

Steered Mixture-of-Experts (SMoE) is a novel framework for the approximation of image modalities regardless of their dimensionality. It has several applications, such as coding, scale-conversion, and description. It has previously been successfully applied to images, videos, light field images and light field video [6]–[9]. Consequently, SMoE is able to approximate the 4D light field and to provide a limited 6DoF experience. However, the "open space" assumption does not always hold. We would therefore like to approximate the 5D plenoptic function. This function describes the color and intensity of a light ray arriving in a 3D point in space for each 2D angle. However, SMoE has only been considered on Euclidean spaces, and not on spherical dimensions.

Multivariate versions of directional distributions exist in the field of directional statistics, such as the *von Mises-Fisher* (vMF) and *Kent* distributions [10]. The vMF distribution is analogous to the symmetric Gaussian distribution, and thus cannot be steered. It was later generalized towards the steerable Fisher-Bingham distribution but is mathematically inelegant and lacks a natural interpretation of the parameters [11]. Kent suggested an alternative with more interpretable parameters and which is more flexible than the vMF distribution [12]. However, the normalization constant is not solvable in closed-form and the approximation of the constant is not always applicable [11]. Furthermore, the parameters are still less flexible and interpretable compared to the multivariate Gaussian. Fitting mixtures of Kent distributions has been proposed, but relies on the approximate normalization constant and remains computationally complex [13].

This paper proposes a method for SMoE on spherical image data. Such image data has some specific properties that we can exploit. Firstly, we have a relatively uniform distribution of points on the unit sphere. Secondly, we typically work with mixture models with a high number of kernels that have small spatial variance. Consequentially, the unit sphere is

ISBN 978-90-827970-1-5

approximated by the tangent planes defined by the unit vectors that defines the sample location, analogous to a circle that is approximated by a polygon with a high number of edges.

Based on the above observations, we choose to interpret the spherical data as data with a 3D coordinate laying on the unit sphere. We model the data using a *Gaussian Mixture Model* (GMM) with a 3D Euclidean coordinate space. However, it is clear that this is a redundant parametrization as all the data lay on a manifold, i.e. the unit sphere. We show that the GMM can be projected onto a 2D coordinate space by locally projecting each kernel's covariance matrix onto the tangent plane defined by the kernel center. We implement this idea for omnidirectional (360°) images, thus having two spherical coordinate dimensions and three Cartesian color dimensions, i.e. RGB.

II. Steered Mixture-of-Experts

A. Introduction

Steered Mixture-of-Experts (SMoE) is a novel framework for approximating image modalities with many applications, such as image modality coding, scale conversion (e.g. frame interpolation), and image description (e.g. depth estimation) [6]–[8]. Due to the sparse structure in SMoE, it is readily extendable towards higher dimensional image modalities, such as 6DoF content. This is in stark contrast to traditional image coding schemes which rely on dense sample-grid structures. Moreover, it departs significantly from the conventional coding methods by operating in the spatial domain and thus not using any kind of transform coding. Instead of storing exactly the samples or the transform coefficients that define the image, this method relies on modeling the underlying generative function that could have given rise to the samples.

The function approximation of the underlying generative function is done by identifying coherent, stationary regions in the image modality. Each segment is modeled using a single N-dimensional entity, which we call a *kernel* or *component*. SMoE is based on the divide-and-conquer principle that is present in all *Mixture-of-Experts* (MoE) approaches [14]. Firstly, the input space is divided in soft-segments using a gating function. Secondly, local regressors (or *experts*) are sought that locally approximate the function optimally. Consequently, the gating function lets experts collaborate in segments where they are trustworthy.

SMoE is based on the Bayesian, or "alternative" definition of the MoE model [14]. The Bayesian MoE approach models the joint probability of the input space X and the output space Y using a GMM. Each Gaussian kernel then simultaneously defines the gating function (soft-segmentation of X) and the local regressors (through the conditional probability function $Y|X$).

In SMoE, where the input space is the *coordinate space* (i.e. sample locations) and the output space is the *color space* (i.e. sample amplitudes), one such Gaussian then corresponds to one kernel as mentioned above. The gating function is thus defined by the probability of a coordinate to belong to a Gaussian, and each Gaussian simultaneously defines an expert

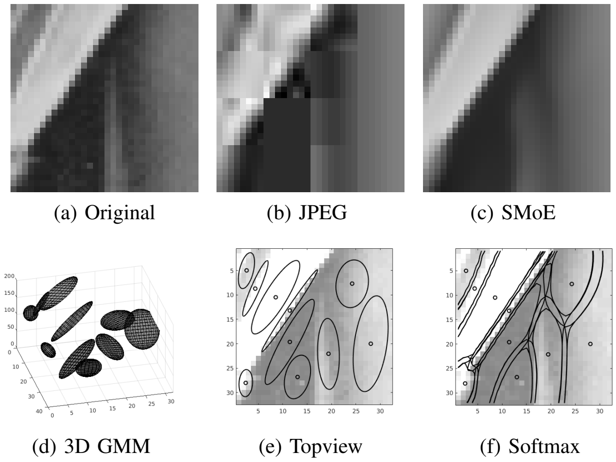


Fig. 1. An example of the modeling with 10 components and reconstruction of a 32x32 pixel crop from *Lena* (1a). For a grayscale image, the coordinate space X is 2D and the colorspace Y is 1D. Modeling the joint probability function of both X and Y using a Gaussian Mixture Model results in 3D Gaussian kernels (1d). Each kernel thus defines a 2D gradient as the *expert* function ($X \mapsto Y$). The gating function is defined by the soft-segmentation (1f). Both JPEG (1b) and SMoE (1c) are coded at 0.35 bpp [6].

function, namely the conditional color amplitudes, given a coordinate. In general, SMoE allows to query the model at any sub-pixel coordinate to yield the most optimal amplitude in a Bayesian sense.

SMoE thus arrives at a sparse representation. The whole image modality is represented as a set of Gaussian kernels. These kernels are defined by their centers and steering parameters. The coordinate space is 2D, 3D, 4D, or 5D in the case of respectively images, video, static light fields, and light field videos [6]–[9]. The color space for color images is conventionally represented as a 3D space, e.g. RGB or YCbCr.

In this paper, we interpret the spherical data as data laying on the unit sphere in a 3D coordinate space. We use the GMM to model the joint probability of the 3D coordinate and 3D color space, we thus arrive at 6D Gaussian kernels. The parameters of these kernels are typically estimated using computational efficient variations of the *Expectation-Maximization* (EM) algorithm [15]. Due to this likelihood optimization, kernels will steer along the dimensions of the highest correlation, e.g., along spatial or temporal consistencies.

Fig. 1 shows an example of the compression capability of the SMoE approach for coding a 32x32 pixel crop of Lena at 0.35 bits/sample in comparison to JPEG at same rate. Clearly, the edges are reconstructed with convincing quality and sharpness, using merely 10 components [6].

B. Theory

The goal of regression is to optimally predict a dependent random vector $Y \in \mathbb{R}^q$ from a known random vector $X \in \mathbb{R}^p$. In SMoE, X corresponds to pixel coordinates (i.e. the 3D coordinate space) and Y to the pixel amplitudes (i.e. the 3D color space). The joint probability function of the coordinate space X and color space Y is modeled as a multi-modal, multi-variate GMM. Each Gaussian kernel then defines a soft-segment in X and a local regressor ($X \mapsto Y$). The local

ISBN 978-90-827970-1-5

regressor is defined by a measure of central tendency (e.g. the mean, median, mode) of the conditional pdf $Y|X$. In this paper, we will limit the case to the mean-estimator, i.e. $\mathbf{E}[Y|X = \boldsymbol{x}]$.

Let us assume $D = \{\boldsymbol{x}^i, \boldsymbol{y}^i\}_{i=1}^N$ to be N pixels to be modeled with coordinates $\boldsymbol{x}$ and amplitudes $\boldsymbol{y}$:

$$p_{XY}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\mu}_j, R_j) \tag{1}$$

and $$\sum_{j=1}^{K} \pi_j = 1, \boldsymbol{\mu}_j = \begin{bmatrix} \boldsymbol{\mu}_{X_j} \\ \boldsymbol{\mu}_{Y_j} \end{bmatrix}, R_j = \begin{bmatrix} R_{X_j X_j} & R_{X_j Y_j} \\ R_{Y_j X_j} & R_{Y_j Y_j} \end{bmatrix}$$

The parameters of this mixture model with K Gaussian distributions are $\Theta = [\theta_1, \cdots, \theta_K]$, with $\theta_j = (\pi_j, \boldsymbol{\mu}_j, R_j)$, being the population densities, centers, and covariances respectively.

The conditional pdf of the mixture model $Y|X$ is used to derive the regression function [16]:

$$p_Y(\boldsymbol{y}|X = \boldsymbol{x}) = \sum_{j=1}^{K} w_j(\boldsymbol{x}) \mathcal{N}(\boldsymbol{y}; m_j(\boldsymbol{x}), \hat{R}_{Y_j,Y_j}) \tag{2}$$

with mixing weights $w_j(\boldsymbol{x})$, regressors $m_j(\boldsymbol{x})$, and conditional covariance $\hat{R}_{Y_j,Y_j}$:

$$w_j(\boldsymbol{x}) = \frac{\pi_j \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_{X_j}, R_{X_j X_j})}{\sum_{i=1}^{K} \pi_i \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_{X_i}, R_{X_i X_i})} \tag{3}$$

$$m_j(\boldsymbol{x}) = \boldsymbol{\mu}_{Y_j} + R_{Y_j X_j} R_{X_j X_j}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_{X_j}), \tag{4}$$

$$\hat{R}_{Y_j,Y_j} = R_{Y_j Y_j} - R_{Y_j X_j} R_{X_j X_j}^{-1} R_{X_j Y_j} \tag{5}$$

The regression of the model is defined as the expected value y given a sample location x through the conditional. From Eq. 2 and 3 follows the regression function $m(\boldsymbol{x})$:

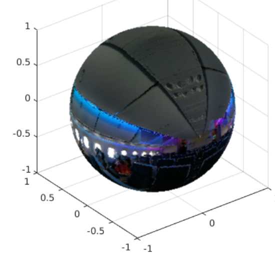
$$\hat{\boldsymbol{y}} = m(\boldsymbol{x}) = \mathbf{E}[Y|X = \boldsymbol{x}] = \sum_{j=1}^{K} w_j(\boldsymbol{x}) m_j(\boldsymbol{x}) \tag{6}$$

A signal at location x can be predicted by the weighted sum over all K mixture components (Eq. 6). Every component in the mixture model is considered as an expert and the experts collaborate towards the definition of the regression function.

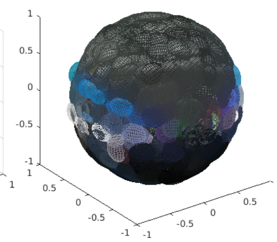
III. SMoE for Spherical Image Data

In order to perform SMoE on samples with two spherical dimensions, each coordinate is first translated into a 3D unit vector. Consequently, we then operate in a Cartesian space in which we can straightforwardly apply SMoE. However, given that all the data lays onto the unit sphere, this parametrization is redundant. In this section, we propose a projection method in order to prove that such a 3D coordinate space can be projected onto a parameter space of equal size as for planar 2D images. Fig. 2 illustrates a SMoE model trained on image data laying on the unit sphere.

(a) Equirectangular projection (P10)

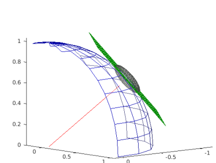


(b) Pixels on the sphere

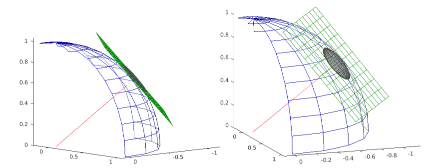


(c) Approximated by GMM

Fig. 2. Example of a SMoE model on the unit sphere without projection. Only the coordinate space is visible on the axes. The color space is visualized by the color of the ellipsoids.



(a) Original



(b) Projected onto the tangent plane

Fig. 3. Illustration of the projection of a single covariance matrix (a) onto the tangent plane $P^{\perp}$ (green plane). A small eigenvalue ϵ is added corresponding to the eigenvector that is defined by the coordinate center μ_X (red vector).

A. Projection of the covariance matrix

As mentioned in Sec. II, a kernel is defined by its prior π, center $\boldsymbol{\mu} = [\boldsymbol{\mu}_X; \boldsymbol{\mu}_Y]$ and covariance matrix R. The coordinate center $\boldsymbol{\mu}_X \in \mathbb{R}^3$ can be seen as a vector radiating out from the center of the unit sphere. The sub-space orthogonal to this vector, at the surface of this sphere, is necessarily tangential to the sphere and is given by $P_{\perp} = I - P$, where $P = \boldsymbol{\mu}_X \boldsymbol{\mu}_X^T / (\boldsymbol{\mu}_X^T \boldsymbol{\mu}_X)$ [17]. We approximate R_{XX} by projecting the covariance matrix onto $P_{\perp}$ as follows

$$R_{XX} \approx P_{\perp} R_{XX} P_{\perp} + P R_{XX} P, \tag{7}$$

where the first term is a projection of the coordinate covariance matrix onto the 2D tangent plane and the second term is the contribution along $\boldsymbol{\mu}_X$. We note that although Eq. 7 is an approximation, it will be demonstrated in Sec. V that it is sufficiently accurate for our purpose.

Also note that the unit sphere is infinitely thin. Therefore, when K is sufficiently large, the contribution along $\boldsymbol{\mu}_X$ will become infinitely small. In this case, the covariance matrix R_{XX} is completely defined by the two eigenvectors (and corresponding eigenvalues), that lay in the 2D plane $P_{\perp}$ and the third eigenvector is along the $\boldsymbol{\mu}_X$ direction with a small eigenvalue. This small eigenvalue can be fixed to a small scalar ϵ. The second term in Eq. 7 is thus approximated by ϵP.

ISBN 978-90-827970-1-5

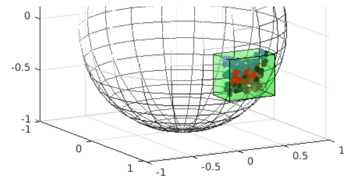


Fig. 4. Illustration of the local updating on the sphere. In this step, the likelihood of the samples (red) are only being calculated by the kernels (colored) that lay in the vicinity of these samples. The relevance window is a cube (green) surrounding the samples.

Let us define the projected covariance matrix $\tilde{R}$ as being constructed by four submatrices analogously to Eq. 1:

$$\tilde{R}_{XX} = P_\perp R_{XX} P_\perp \quad (8)$$

$$\tilde{R}_{XY} = \tilde{R}_{YX}^T = R_{XY} P_\perp \quad (9)$$

$$\tilde{R}_{YY} = R_{YY} \quad (10)$$

with $\tilde{R}_{XX}$ and $\tilde{R}_{XY}$ now being of rank-2.

B. Dimensionality reduction

In this section, we illustrate that it is possible to parametrize the two spherical dimensions with the same number of parameters as two Cartesian dimensions (planar images), i.e. having a $\dot{\boldsymbol{\mu}}_X \in \mathbb{R}^2$, $\dot{R}_{XX} \in \mathbb{R}^{2\times 2}$ and thus $\dot{\boldsymbol{\mu}} \in \mathbb{R}^5$, $\dot{R} \in \mathbb{R}^{5\times 5}$.

The coordinate center $\boldsymbol{\mu}_X$ approximates a unit vector when K goes to infinity as it is the mean of an ever decreasing amount of data laying on a small segment of the sphere. It can therefore be parametrized by two coefficients as the norm is one, i.e. $\dot{\boldsymbol{\mu}}_X \in \mathbb{R}^2$. Using the eigenvalue decomposition, we can show the following

$$\tilde{R}_{XX} = UDU^T = [\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3] \begin{bmatrix} d_1 & & \\ & d_2 & \\ & & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \boldsymbol{u}_3^T \end{bmatrix} \quad (11)$$

$$= d_1 \boldsymbol{u}_1 \boldsymbol{u}_1^T + d_2 \boldsymbol{u}_2 \boldsymbol{u}_2^T \quad (12)$$

Let $\dot{R}_{XX}$ be the desired 2x2 covariance that defines the covariance in the 2D $P_\perp$ plane. We choose to construct $\dot{R}_{XX}$ by taking the top-left four elements of $\tilde{R}_{XX}$:

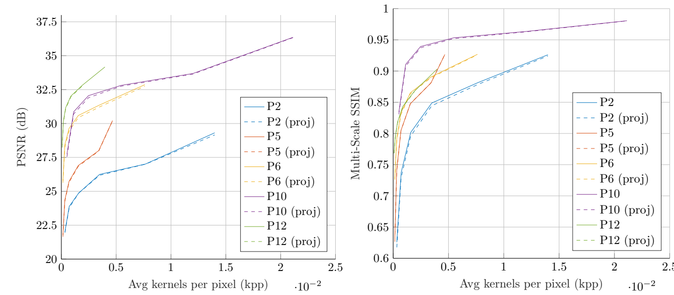
$$\tilde{R}_{XX} = \begin{bmatrix} \dot{R}_{XX} & \begin{bmatrix} a \\ b \end{bmatrix} \\ \begin{bmatrix} a & b \end{bmatrix} & c \end{bmatrix}. \quad (13)$$

At decoder side, we can find a, b, and c by solving $P\tilde{R}_{XX} = 0$ using linear operations. Finally, in order to have a small positive eigenvalue ϵ along $\boldsymbol{\mu}_X$, we add ϵP to $\tilde{R}_{XX}$.

Note that the dimension reduction is analogous for $\tilde{R}_{XY}$. However, we do not change the eigenvalues for $\tilde{R}_{XY}$, this means that there is no color gradient along $\boldsymbol{\mu}_X$. This is the information that is lost using this projection, however this is not critical since it is a color gradient along the line of sight.

IV. Proposed Modeling

We propose an iterative training method based on the covariance projection technique described above. A straightforward method would be to use batch *Expectation-Maximization* (EM) which globally fits the entire GMM onto the sample data [15].



(a) PSNR

(b) Multi-Scale SSIM

Fig. 5. Experiment results comparing the modeling with and without the projection in terms of PSNR and Multi-Scale SSIM.

However, this is computationally infeasible. The proposed training method is based on the progressive modeling approach used for training SMoE models for light field videos [9]. In this approach, global modeling is simulated by a local fitting strategy in which a local group of samples is processed by only the set of relevant kernels in its vicinity. This simulates global modeling while minimizing the computational complexity. For 360° images, the relevance window that selects the kernels becomes 3D as shown in Fig. 4. Only a minibatch of these samples is selected in each iteration in order to increase the robustness of the kernel updates and to heavily decrease the computational complexity. This also allows us to uniformly sample the sphere.

Models are initialized with K_{start} kernels spread out uniformly over the sphere and are further trained until convergence. After convergence, a portion of the top uncertain kernels are split into four smaller kernels which serves as a new initialization for the next meta-iteration. The split is performed on the tangent plane and the uncertainty is defined by the weighted conditional variance of the color space, i.e. $\pi_j \operatorname{Tr}(\hat{R}_{Y_jY_j})$. The new kernels are then projected back onto the unit sphere. This model is then further trained until convergence. The process stops when a predetermined $K_{\max}$ is exceeded. Finally, all the covariance matrices are projected onto the tangent planes and the centers are projected onto the unit sphere. Note that the projection step does not introduce considerable computational overhead.

V. Approximation Experiments

A. Setup

For the experiments, five images were selected from the *Salient360!* dataset: P2, P5, P6, P10, and P12 [18], [19]. These images were stored in equirectangular format. After remapping the pixels onto the unit sphere, these images are progressively modeled using minibatches of size 10,000 and local updates per 18°-by-18° segments. Models are initialized uniformly on the sphere with $K_{\max} = 2^{12}$ kernels. After each meta-iteration, 40% of the kernels are split based on their weighted conditional variance. The modeling stops when $K_{\max} = 2^{18}$ is reached.

ISBN 978-90-827970-1-5 © EURASIP 2018

B. Results

Fig. 5 shows the objective quality results for the indicated images in terms of PSNR and *Multi-Scale Structural Similarity* (MS-SSIM) [20]. The x-axis is expressed in *kernels-per-pixel* (kpp), as the original resolutions of the images span from 2000x4000 (P10) to 5000x10000 (P12). Note that we have a fixed $K_{\max}$, each image thus spans a different kpp range. The plots are shown up to 0.02kpp, which indicates that on average one kernel spans 50 pixels in the original equirectangle image. For P10 (Fig. 2) a 0.9 on the MS-SSIM scale is achieved at 0.001kpp, which is an average of 1 kernel per 1000 pixels.

The average loss over all images is 0.1 dB PSNR and 0.002 MS-SSIM. We can conclude that the projection step introduces a relatively small quality loss, which indicates that the assumptions made are valid. Note that kernels can become insignificant during the modeling. These kernels are consequently removed, which influences the shape of the plots.

VI. Conclusion

We have presented a methodology for applying SMoE to spherical dimensions by operating on the unit sphere in the Euclidean $\mathbb{R}^3$ space. A computationally cheap technique to reduce the parameter space to the same space as for planar 2D images is introduced. This is done by projecting the covariance matrices onto the 2D tangent space defined by the kernel's center. Finally, a computationally efficient modeling scheme that utilizes this projection step is presented. Experiments validate that the parameter space reduction introduces nearly no quality loss for the tested 360° dataset. As such, we have shown that SMoE can be applied efficiently to spherical dimensions as needed for approximating the 5D plenoptic function in the future.

References

[1] A. T. Hinds, D. Doyen, and P. Carballeira, "Toward the realization of six degrees-of-freedom with compressed light fields," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2017, pp. 1171–1176.

[2] I. Ihrke, J. Restrepo, and L. Mignard-Debise, "Principles of Light Field Imaging: Briefly revisiting 25 years of research," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 59–69, 2016.

[3] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. Conf. on Computer Graphics and Interactive Techniques - SIGGRAPH '96*, New York, New York, USA: ACM Press, 1996, pp. 31–42.

[4] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light Field Image Processing: An Overview," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–1, 2017.

[5] E. Adelson and J. Bergen, *The plenoptic function and the elements of early vision*. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.

[6] R. Verhack, T. Sikora, L. Lange, G. Van Wallendael, and P. Lambert, "A universal image coding approach using sparse steered Mixture-of-Experts regression," in *IEEE Proc. Int. Conf. on Image Processing (ICIP)*, IEEE, 2016, pp. 2142–2146.

[7] L. Lange, R. Verhack, and T. Sikora, "Video Representation and Coding Using a Sparse Steered Mixture-of-Experts Network," in *Picture Coding Symposium (PCS)*, 2016.

[8] R. Verhack, T. Sikora, L. Lange, R. Jongebloed, G. Van Wallendael, and P. Lambert, "Steered mixture-of-experts for light field coding, depth estimation, and processing," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, Ed., IEEE, 2017, pp. 1183–1188.

[9] R. Verhack, G. Van Wallendael, M. Courteaux, P. Lambert, and T. Sikora, "Progressive Modeling of Steered Mixture-of-Experts for Light Field Video Approximation," in *Submitted to Picture Coding Symposium '18*, 2018.

[10] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009, vol. 494.

[11] P. Kasarapu, "Modelling of directional data using Kent distributions," 2015. arXiv: 1506.08105.

[12] J. T. Kent, "The Fisher-Bingham distribution on the sphere," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 71–80, 1982.

[13] D. Peel, W. J. Whiten, and G. J. McLachlan, "Fitting Mixtures of Kent Distributions to Aid in Joint Set Identification," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 56–63, 2001.

[14] S. E. Yuksel, J. N. Wilson, and P. D. Gader, "Twenty Years of Mixture of Experts," *IEEE Trans. Neural Netw. Learn. Syst*, vol. 23, no. 8, pp. 1177–1193, 2012.

[15] T. Moon, "The Expectation-Maximization Algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[16] G. Bugmann, "Normalized Gaussian Radial Basis Function networks," *Neurocomputing*, vol. 20, no. 1-3, pp. 97–110, 1998.

[17] G Strang, *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 2016.

[18] Y. Rai, P. Le Callet, and P. Guillotel, "Which saliency weighting for omni directional image quality assessment?" In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, IEEE, 2017, pp. 1–6.

[19] Y. Rai, J. Gutiérrez, and P. Le Callet, "A Dataset of Head and Eye Movements for 360 Degree Images," in *Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys'17*, New York, New York, USA: ACM Press, 2017, pp. 205–210.

[20] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, IEEE, pp. 1398–1402.