Sparse Needlets for Lighting Estimation with Spherical Transport Loss: Supplementary Materials

Fangneng Zhan

1. Appendix Outline

Lighting estimation is a classic challenge in computer vision and computer graphics, and it is critical for realistic relighting in objects insertion and image synthesis [12, 1, 10, 14, 29, 2, 28, 30, 21, 22, 31]. On the other hand, the recent works aim to estimate lighting from images by regressing representation parameters [3, 6, 13] or generating illumination maps [8, 16] based image translation [19, 4, 15, 17, 34, 26, 33, 24, 25, 23, 32, 27, 20].

This appendix provides more details on the derivation of sparse needlets, cubature points, spatially-varying illumination, and spherical transport distance.

2. Derivation of Sparse Needlets

We derive the sparse Needlet function from a Bayesian framework and form the problem as a maximum posterior estimator. We assume that the needlet coefficients of light sources s follows Laplace distribution prior as Laplace prior is well adapted to model sparse signals [11]. For HDR images, the ambient is mainly determined by the light sources in the scenes. We assume that the needlet coefficients of ambient contributed by each light source follow Gaussian distribution with 0 mean [5]. The needlet coefficients of ambient can thus be treated as several independent Gaussian distributions. The needlet coefficients β of the illumination map can be modelled as follows:

$$\beta = \underbrace{s}_{light\ sources} + \underbrace{\phi}_{ambient} + \underbrace{\eta}_{noise} \tag{1}$$

where s denotes the needlet coefficients of sparse light sources which follow Laplace distribution, ϕ denotes the needlet coefficients of ambient which follow Gaussian distribution, and η denotes noises that follow a Gaussian distribution. According to [18], the Bayesian formulation of the problem aims to maximize $P(s|\beta)$:

$$P(s|\beta) \propto P(s)\mathcal{L}(\beta|s) = P(s)\int \mathcal{L}(\beta|s,\phi)P(\phi)d\phi \tag{2}$$

As the noise η follows Gaussian distribution and the background follows n independent Gaussian distributions, we can get the formulations:

$$\begin{aligned} &\mathcal{L}(\beta|s,\phi) = N(\beta;\phi+s,M_\eta) \\ &P(\phi) = N(\phi,0,M_\phi) = \exp\left[-\frac{1}{2}\phi^T M_\phi^{-1}\phi\right] \end{aligned} \tag{3}$$

where M_η and M_ϕ denote the covariance matrices of the noise and Guassian distribution, respectively. Thus we can obtain:

$$\begin{aligned} &\mathcal{L}(\beta|s,\phi)P(\phi) \propto \exp\left[s^T M_\eta^{-1}\beta - \frac{1}{2}s^T M_\eta^{-1}s - \frac{1}{2}\beta^T M_\eta^{-1}\beta\right] \\ &\cdot \exp\left[-\frac{1}{2}\phi^T(M_\phi^{-1}+M_\eta^{-1})\phi + \phi^T(M_\eta^{-1}\beta - M_\eta^{-1}s)\right] \end{aligned} \tag{4}$$

According to the Gaussian integration,

$$\int \exp\left[N^T x - \frac{1}{2}x^T M x\right]dx = \sqrt{\frac{2\pi}{\det M}}\exp\left[\frac{1}{2}N^T M^{-1} N\right] \tag{5}$$

we can derive:

$$\begin{aligned} &\mathcal{L}(\beta|s) = \int \mathcal{L}(\beta|s,\phi)P(\phi)d\phi \propto \\ &\exp\left[-\frac{1}{2}\beta^T M_\eta^{-1}\beta + s^T M_\eta^{-1}\beta - \frac{1}{2}x^T M_\eta^{-1}s\right]\cdot \\ &\exp\left[\frac{1}{2}(M_\eta^{-1}\beta - M_\eta^{-1}s)^T(M_\eta^{-1}+M_\phi^{-1})(M_\eta^{-1}\beta - M_\eta^{-1}s)\right] \end{aligned} \tag{6}$$

Maximizing $P(s|\beta) = \mathcal{L}(\beta|s) * P(s)$ is equivalent to minimizing $\partial_s(-log(P(s|\beta)))$. As we assume that s follows a Laplace distribution with 0 mean, namely, $P(s) \propto \exp\left[-\lambda||s||\right]$, we can obtain the partial derivative with respect to x as follows:

$$\begin{aligned} &\partial_s(-log(P(s|\beta))) = -(M_\eta + M_\eta M_\phi^{-1}M_\eta)^{-1}s + M_\eta^{-1}s \\ &\quad + (M_\eta + M_\eta M_\phi^{-1}M_\eta)^{-1}\beta - M_\eta^{-1}\beta + \lambda\partial_s||s|| \end{aligned} \tag{7}$$

So sparse function to maximize $P(s|\beta)$ is:

$$s = \beta - \left[M_\eta^{-1} - (M_\eta + M_\eta M_\phi^{-1}M_\eta)^{-1}\right]^{-1}\lambda\partial_s|s| \tag{8}$$

where $\partial_s(|s|) = \begin{cases} 1 & if\ s>0 \\ -1 & if\ s<0 \end{cases}$. Thus we can obtain the thresholding as follows:

$$sgn(\beta)(|\beta| - \left[M_\eta^{-1} - (M_\eta + M_\eta M_\phi^{-1}M_\eta)^{-1}\right]^{-1}\lambda)_+ \tag{9}$$

which is essentially a soft thresholding operator with threshold $(M_\eta^{-1} - (M_\eta + M_\eta M_\phi^{-1}M_\eta)^{-1})\lambda$.

As described in [7], we can separate light source and ambient region by thresholding illumination maps. The needlet coefficients of ambient ϕ and covariance matrice M_ϕ can thus be computed, and s can be further computed by Eq. 8 with known M_ϕ. Computing s directly on thresholded light-source region is the same as hard thresholding (HT), which performs worse than the proposed sparse function (soft thresholding) as shown in Table 3 of the manuscript. Specially, we only apply sparse function to high-frequency coefficients (j=3 in this work). Therefore, the coefficients s of high frequency bands are sparse but those of low-frequency bands (i.e. s+ϕ which preserves the ambient ϕ) are not sparse.

3. Cubature Points

Cubature points and cubature weights are provided by the HEALPix discretization of the sphere [9]. The HEALPix grid discretizes the sphere into N_{pix} pixels with equal area, where $N_{pix} = 12N_{side}^2$ and Nside is required to be a power of two which measures the discretization resolution. We specify the cubature weights λ_{jk} as $\lambda_{jk} = \frac{4\pi}{N_{pix}}$. There are 1, 12, 48, 192 needlet coefficients for frequency order $j = 0, 1, 2, 3$ respectively. The spatial localization of needlet coefficients are indicated by cubature points on a unit sphere. Fig. 1 illustrate cubature points for $j = 1, 2, 3$ by panoramas of spheres.

4. Spatially-varying Illumination

For Laval Indoor dataset [8], spatially-varying illumination cannot be predicted directly as there is no corresponding ground truth in this dataset. We thus employ cubature points to approximate the spatially-varying illumination, more details to be described as follows.

We recall the definition of needlet basis ψ_{jk} and needlet coefficients β_{jk}:

$$\begin{aligned}
\psi_{jk}(x) &= \sqrt{\lambda_{jk}} \sum_{l=\lceil B^{j-1} \rceil}^{\lfloor B^{j+1} \rfloor} b(\frac{l}{B^j}) \sum_{m=-l}^{l} Y_{lm}(\xi_{jk}) \overline{Y}_{lm}(x) \\
\beta_{jk} &= \sqrt{\lambda_{jk}} \sum_{l=0}^{\infty} b(\frac{l}{B^j}) \sum_{m=-l}^{l} a_{lm} Y_{lm}(\xi_{jk})
\end{aligned} \tag{10}$$

where $x \in \mathbb{S}^2$, ξ_{jk} and λ_{jk} are pre-defined cubature points as shown in Fig. 1 and the associated cubature weights, respectively. The illumination map $I(x)$ can be reconstructed via $I(x) = \sum_{j,k} \beta_{jk}\psi_{jk}(x)$. Obviously, the illumination map is reconstructed based on cubature points. Thus we approximate spatially-varying illumination by moving the coordination cubature points.

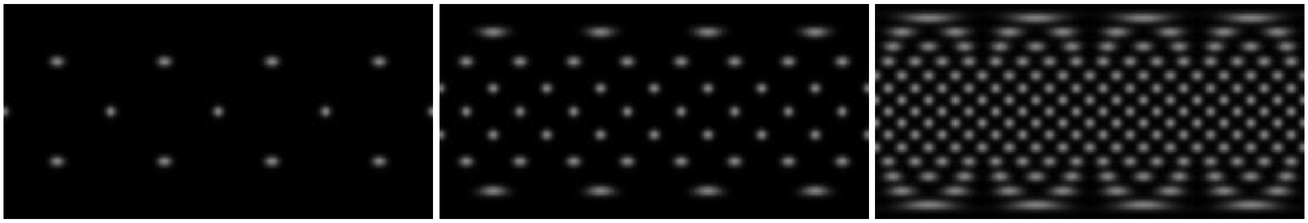
When we move the insertion position by $\nabla\xi$, the new direction of the cubature point jk can be denoted by $\xi_{jk} + \nabla\xi$. Then the needlet basis ψ'_{jk} on the new insertion position can be denoted by:

$$\psi_{jk}(x)' = \sqrt{\lambda_{jk}} \sum_{l=\lceil B^{j-1} \rceil}^{\lfloor B^{j+1} \rfloor} b(\frac{l}{B^j}) \sum_{m=-l}^{l} Y_{lm}(\boldsymbol{\xi_{jk}} + \boldsymbol{\nabla\xi}) \overline{Y}_{lm}(x) \tag{11}$$

With the predicted needlet coefficients, illumination map at a new insertion position can be reconstructed by: $I(x)' = \sum_{j,k} \beta_{jk}\psi_{jk}(x)'$. Fig. 2 shows the reconstructed spatially-varying illumination at different insertion positions.

5. Spherical Transport Distance

Spherical transport distance can compute the distance between two masses distributed on a sphere effectively. We compare the spherical transport distance and L2 distance with a simple example as illustrated in Fig 3.



(a) j=1 (b) j=2 (c) j=3

Figure 1. Visualization of cubature points ($j = 1, 2, 3$) on a unit sphere.

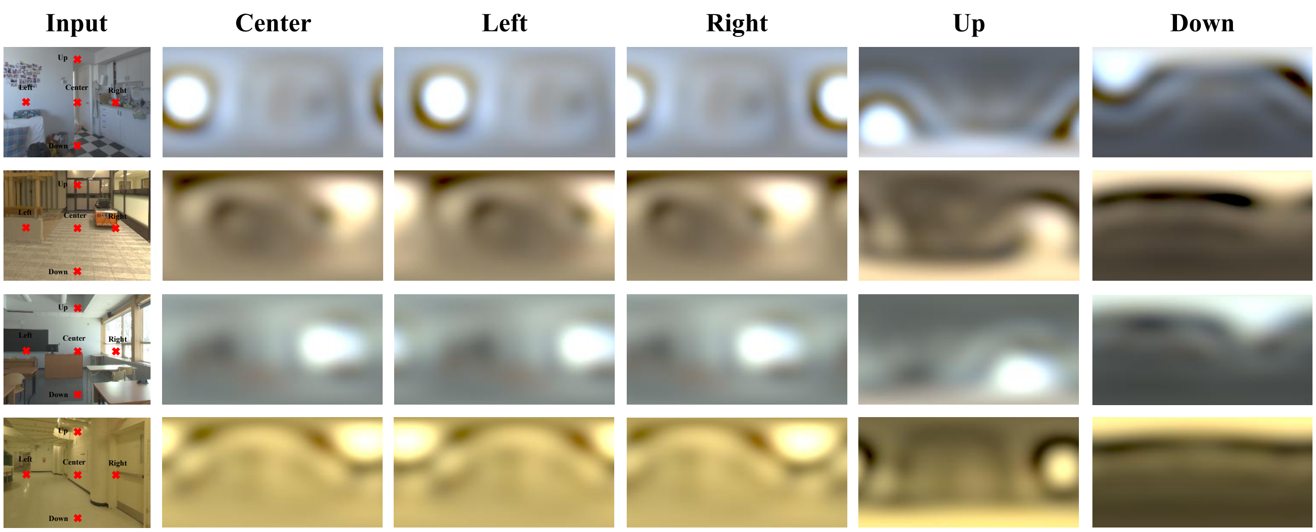


Figure 2. Estimated spatially-varying illumination maps at different insertion positions (Center, Left, Right, Up, and Down).

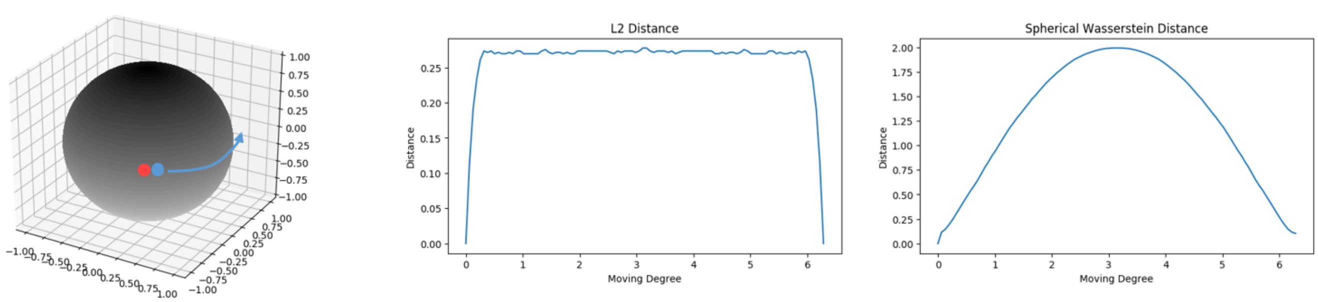


Figure 3. Comparison between spherical transport distance and L2 distance: Take the red and blue masses in the first graph as two identical spherical distributions. When the blue mass is moving along a great circle as indicated by the blue arrow, the two graphs on the right show the distances between the two distributions that are measured by L2 distance and spherical transport distance, respectively.

References

[1] Jonathan T Barron and Jitendra Malik. Intrinsic scene properties from a single rgb-d image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 17–24, 2013. 1

[2] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik Lensch, and Jan Kautz. Two-shot spatially-varying brdf and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3982–3991, 2020. 1

[3] Dachuan Cheng, Jian Shi, Yanyun Chen, Xiaoming Deng, and Xiaopeng Zhang. Learning scene illumination by pairwise photos from rear and front mobile cameras. In *Computer Graphics Forum*, pages 213–221. Wiley Online Library, 2018. 1

[4] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8188–8197, 2020. 1

[5] D.Maino, A.Farusi, C.Baccigalupi, F.Perrotta, A.J.Banday, L.Bedini, C.Burigana, G.De Zotti, K.M.Gorski, and E.Salerno. All-sky astrophysical component separation with fast independent component analysis (fastica). *Monthly Notices of the Royal Astronomical Society*, 2002. 1

[6] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-Francois Lalonde. Deep parametric indoor lighting estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7175–7183, 2019. 1

[7] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-François Lalonde. Deep parametric indoor lighting estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7175–7183, 2019. 2

[8] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. Learning to predict indoor illumination from a single image. *arXiv preprint arXiv:1704.00090*, 2017. 1, 2

[9] Krzysztof M Gorski, Eric Hivon, Anthony J Banday, Benjamin D Wandelt, Frode K Hansen, Mstvos Reinecke, and Matthia Bartelmann. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2):759, 2005. 2

[10] Yannick Hold-Geoffroy, Kalyan Sunkavalli, Sunil Hadap, Emiliano Gambaretto, and Jean-François Lalonde. Deep outdoor illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7312–7321, 2017. 1

[11] Bobin J., Moudden Y., Starck J.-L., Fadili J., and N.Aghanim. Sz and cmb reconstruction using generalized morphological component analysis. *Statistical Methodology*, 2008. 1

[12] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision*, 98(2):123–145, 2012. 1

[13] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2475–2484, 2020. 1

[14] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A dataset of multi-illumination images in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4080–4089, 2019. 1

[15] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2337–2346, 2019. 1

[16] Shuran Song and Thomas Funkhouser. Neural illumination: Lighting prediction for indoor environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6918–6926, 2019. 1

[17] Hao Tang, Dan Xu, Nicu Sebe, Yanzhi Wang, Jason J Corso, and Yan Yan. Multi-channel attention selection gan with cascaded semantic guidance for cross-view image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2417–2426, 2019. 1

[18] Flavien Vansyngel, Benjamin D. Wandelt, Jean-François Cardoso, and Karim Benabed. Semi-blind bayesian inference of cmb map and power spectrum. *arXiv:1409.0858*, 2014. 1

[19] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018. 1

[20] Fangneng Zhan, Shijian Lu, and Chuhui Xue. Verisimilar image synthesis for accurate detection and recognition of texts in scenes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 249–266, 2018. 1

[21] Fangneng Zhan, Shijian Lu, Changgong Zhang, Feiying Ma, and Xuansong Xie. Adversarial image composition with auxiliary illumination. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 1

[22] Fangneng Zhan, Shijian Lu, Changgong Zhang, Feiying Ma, and Xuansong Xie. Towards realistic 3d embedding via view alignment. *arXiv preprint arXiv:2007.07066*, 2020. 1

[23] Fangneng Zhan, Chuhui Xue, and Shijian Lu. Ga-dan: Geometry-aware domain adaptation network for scene text detection and recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9105–9115, 2019. 1

[24] Fangneng Zhan, Yingchen Yu, Kaiwen Cui, Gongjie Zhang, Shijian Lu, Jianxiong Pan, Changgong Zhang, Feiying Ma, Xuansong Xie, and Chunyan Miao. Unbalanced feature transport for exemplar-based image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1

[25] Fangneng Zhan, Yingchen Yu, Rongliang Wu, Kaiwen Cui, Aoran Xiao, Shijian Lu, and Ling Shao. Bi-level feature alignment for versatile image translation and manipulation. *arXiv preprint arXiv:2107.03021*, 2021. 1

[26] Fangneng Zhan, Yingchen Yu, Rongliang Wu, Jiahui Zhang, and Shijian Lu. Multimodal image synthesis and editing: A survey. *arXiv preprint arXiv:2112.13592*, 2021. 1

[27] Fangneng Zhan, Yingchen Yu, Rongliang Wu, Jiahui Zhang, Shijian Lu, and Changgong Zhang. Marginal contrastive correspondence for guided image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10663–10672, 2022. 1

[28] Fangneng Zhan, Yingchen Yu, Changgong Zhang, Rongliang Wu, Wenbo Hu, Shijian Lu, Feiying Ma, Xuansong Xie, and Ling Shao. Gmlight: Lighting estimation via geometric distribution approximation. *IEEE Transactions on Image Processing*, 31:2268–2278, 2022. 1

[29] Fangneng Zhan and Changgong Zhang. Spatial-aware gan for unsupervised person re-identification. *Proceedings of the International Conference on Pattern Recognition*, 2020. 1

[30] Fangneng Zhan, Changgong Zhang, Wenbo Hu, Shijian Lu, Feiying Ma, Xuansong Xie, and Ling Shao. Sparse needlets for lighting estimation with spherical transport loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12830–12839, 2021. 1

[31] Fangneng Zhan, Changgong Zhang, Yingchen Yu, Yuan Chang, Shijian Lu, Feiying Ma, and Xuansong Xie. Emlight: Lighting estimation via spherical distribution approximation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3287–3295, 2021. 1

[32] Fangneng Zhan, Jiahui Zhang, Yingchen Yu, Rongliang Wu, and Shijian Lu. Modulated contrast for versatile image synthesis. *arXiv preprint arXiv:2203.09333*, 2022. 1

[33] Fangneng Zhan, Hongyuan Zhu, and Shijian Lu. Spatial fusion gan for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3653–3662, 2019. 1

[34] Zhen Zhu, Zhiliang Xu, Ansheng You, and Xiang Bai. Semantically multi-modal image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5467–5476, 2020. 1