

LIGHT FIELD STYLE TRANSFER WITH LOCAL ANGULAR CONSISTENCY

Dónal Egan, Martin Alain, Aljosa Smolic

V-SENSE Project, School of Computer Science and Statistics, Trinity College, Dublin

ABSTRACT

Style transfer involves combining the style of one image with the content of another to form a new image. Unlike traditional two-dimensional images which only capture the spatial intensity of light rays, four-dimensional light fields also capture the angular direction of the light rays. Thus, applying style transfer to a light field requires to not only render convincing style transfer for each view, but also to preserve its angular structure. In this paper, we present a novel optimization-based method for light field style transfer which iteratively propagates the style from the centre view towards the outer views while enforcing local angular consistency. For this purpose, a new initialisation method and angular loss function is proposed for the optimization process. In addition, since style transfer for light field is an emerging topic, no clear evaluation procedure is available. Thus, we investigate the use of a recently proposed metric designed to evaluate light field angular consistency, as well as a proposed variant.

Index Terms— Light field, style transfer

1. INTRODUCTION

Light fields emerged as a new imaging modality, enabling to capture all light rays passing through a given amount of the 3D space [1]. Compared to traditional 2D imaging systems which only capture the spatial intensity of light rays, the common two-plane parameterisation of light fields also contains the angular direction of the rays. A light field can be represented as a 4D function: $\Omega \times \Pi \rightarrow \mathbb{R}$, $(s, t, u, v) \rightarrow L(s, t, u, v)$ in which the plane Ω represents the spatial distribution of light rays, indexed by (u, v) , while Π corresponds to their angular distribution, indexed by (s, t) . A practical way to visualise a light field is to consider it as a matrix of $M \times N$ views, also called sub-aperture images (SAI), where each image represents a 2D slice of the light field over the spatial dimensions (u, v) . Another common representation of light fields is through Epipolar Plane Images (EPI), which are 2D slices of the 4D light field obtained by fixing one spatial and one angular dimension (us - or vt -planes, see Fig. 1). Applications of light fields notably include rendering novel images, either corresponding to new viewpoints [1] and/or with new focus distance and depth-of-field [2–4], or depth estimation [5].

Style transfer involves combining the content of one image (e.g., a photograph) with the style of another (e.g., a painting) to form a new image. Doing so requires us to be able to define, separate and extract the content and style of an image. Gatys et al.

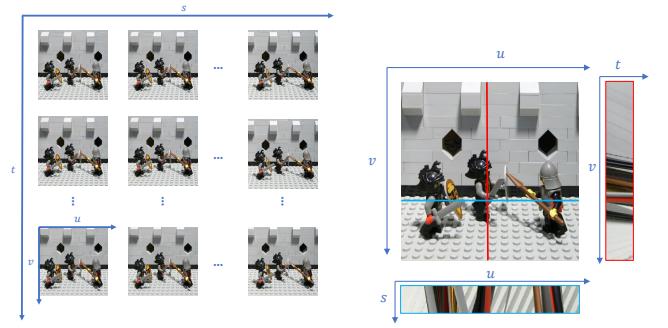


Fig. 1: Examples of light field representations: matrix of views (left); and Epipolar Plane Images (EPI) (right) shown below and on the right of the centre view.

first introduced *neural style transfer* [6], which uses the outputs from the hidden layers of a deep convolutional neural network pre-trained for object recognition (VGG [7]) to construct content and style representations of an input image, which can be used to define corresponding loss functions. Starting from white noise, the stylized image is iteratively generated through an optimization process so that a total loss function combining the style and content loss functions is minimised. While this seminal approach relies on a pre-trained network, its advantage is that any target content or style images can be combined without having to actually train any deep neural network. The cost of this flexibility is that a slow optimization process has to be run to generate each stylized image. Johnson et al. later addressed this issue by fully training a feed-forward *image transformation network* [8], using the same content and style loss functions. At inference time, the stylization process is thus orders of magnitude faster. However, this approach is not as flexible as the previous method as it requires to train a deep neural network for each target style. Following the feed-forward approach of Johnson et al., Liu et al. later introduced an additional depth loss to train the image transformation network [9], producing stylized results which better retain the original content layout and depth relationship. The depth loss is computed as the per pixel loss between the depth estimated from the input content image and the stylized image, using a pre-trained single-image depth perception network [10].

Following the success of style transfer for 2D images, applications to other imaging modalities have emerged, such as video [11, 12], omnidirectional imaging [13], or stereo imaging [14, 15]. The video style transfer approach of Ruder et al. [11] builds on the optimization approach of Gatys et al., enforcing the temporal consistency by initialising a stylized frame from a warped stylized previous frame and introducing a temporal con-

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776.

sistency loss penalizing deviations between consecutive stylized frames. The approach was later extended in [13], notably including an application to omnidirectional image stylization. Building on feed-forward network approaches, stereoscopic style transfer approaches were introduced in order to improve the stylized consistency of stereo pairs. While Gong et al. proposed a dual path network based on a feature aggregation strategy [14], Chen et al. introduced a stereoscopic disparity loss [15]. The first style transfer for light fields was proposed very recently by Hart et al. [16], extending the pair-wise stereoscopic style transfer method of Chen et al. [15] in order to preserve the angular consistency of the light field stylized views. For this purpose, a feed-forward network is trained to enforce pairwise consistency between the centre view and others. The training and testing datasets consist of light fields captured with a lenslet Lytro Illum camera [17, 18]. Due to the design of the Lytro Illum camera, such light fields are densely sampled, *i.e.* the disparity range between neighbouring views usually does not exceed 1. This in turn means that the amount of disocclusions from any view with respect to the centre view is limited, allowing the use of this pairwise strategy.

In this paper, we present a novel method to apply neural style transfer to light fields, extending the original approach of Gatys et al. [6] and its following extension to video [11]. While feed-forward approaches are faster, we chose to use an optimization approach for its flexibility. Note that more recent work improving the speed of optimization-based approaches could also be used [19]. In our proposed method, the stylized light field angular consistency is preserved by propagating the stylization outwards from the centre view. As opposed to the method of Hart et al., our approach enforces the local angular consistency, rather than enforcing the consistency with respect to the centre view. This notably makes our method more adapted to sparsely sampled light fields. Given the subjective nature of style transfer, the evaluation of the results is known to be challenging. The challenge is even greater for light fields, as few metrics have been proposed which could be applied to this new research topic. Recent work investigating no-reference quality evaluation for light fields focuses on dense light fields [20, 21]. We thus propose to study the angular consistency metric recently introduced by David et al. [22] to evaluate sparse light field video interpolation, and assess whether it is suitable for stylized light fields. We also propose a variant of this metric which better takes into account the local angular variations. Finally, while Hart et al. were only comparing to a 2D baseline stylizing each view independently, we also study the performance of a second baseline applying video style transfer to pseudo-videos created from the light field views.

This paper is organised as follows: In section 2, we introduce in more detail the baselines and proposed approach. In section 3, we first describe the metrics we propose to assess and evaluate our approach. We conclude the paper and discuss future work in section 4.

2. STYLE TRANSFER FOR LIGHT FIELDS

Given that style transfer for light fields is still in its infancy, we first describe in this section two baseline methods, directly applying existing optimization-based 2D image style transfer techniques [6] and video style transfer techniques [11]. We then de-

scribe our proposed approach.

2.1. Baselines

As a first baseline, the original approach of Gatys et al. [6] is applied to each light field view $L_{s,t}$. More formally, after white noise initialisation, each stylized view $L_{s,t}^S$ is independently generated by optimizing the following loss function:

$$\mathcal{L}_{Im}(L_{s,t}, I_S, L_{s,t}^S) = \alpha_{Im}\mathcal{L}_c(L_{s,t}, L_{s,t}^S) + \beta_{Im}\mathcal{L}_s(I_S, L_{s,t}^S)$$

where \mathcal{L}_c and \mathcal{L}_s refer to the content loss and the style loss respectively, computed from the feature maps of VGG19 [7]. As defined in [6], we use the output from layer conv4_2 to construct the content representation of an input image and the output from layers conv1_1, conv2_1, conv3_1, conv4_1 and conv5_1 to construct the style representation of an input image. In our experiments we also used the weights recommended in [6], $\alpha_{Im} = 1.0$ and $\beta_{Im} = 1 \times 10^3$ or $\beta_{Im} = 1 \times 10^4$ depending on the input content and style images.

As a second baseline, we adapt video style transfer [11] by creating a pseudo-video from the light field views in which each frame is denoted by L_i where i is the frame index. In our experiments pseudo-videos were created following a horizontal snake-like scanning order. The first frame of the pseudo-video L_0 is then stylized using the single image approach described above. The following frames are then initialised by warping the previously stylized frame using optical flow computed between the original frames. Each stylized frame L_i^S is then generated by optimizing the following loss function:

$$\begin{aligned} \mathcal{L}_{Vid}(L_i, I_S, L_i^S) = & \alpha_{Vid}\mathcal{L}_c(L_i, L_i^S) + \beta_{Vid}\mathcal{L}_s(I_S, L_i^S) + \\ & + \gamma_{Vid}\mathcal{L}_t(L_i^S) \end{aligned}$$

where \mathcal{L}_c and \mathcal{L}_s refer to the content loss and the style loss introduced above and \mathcal{L}_t is a temporal loss function first described in [11]:

$$\mathcal{L}_t(L_i^S) = \sum c_{i-1}^i \cdot (\omega_{i-1}^i(L_{i-1}^S) - L_i^S)^2$$

where c_{i-1}^i are per-pixel weights defined to be 0 for disoccluded regions and motion boundaries between the original frames and 1 elsewhere and ω_{i-1}^i is the function that warps a given image according to the optical flow between the original frames L_{i-1} and L_i . In our experiments, we used $\alpha_{Vid} = 5.0$, $\beta_{Vid} = 1 \times 10^4$ and $\gamma_{Vid} = 2 \times 10^5$.

As shown in section 3, while the video-based baseline greatly improves results compared to the image-based baseline, both baselines fail at preserving the full angular consistency of the light field. In the next section, we describe our proposed approach for light field style transfer which explicitly takes into account the angular consistency of the stylized light field.

2.2. Proposed optimization approach

The main idea of the proposed approach is to adapt the video style transfer method described previously to light fields by propagating the style outwards from the centre view while always ensuring that angular consistency is preserved, as illustrated in Fig. 2. We achieve this through the introduction of a new initialisation method for the image-optimization process and a new angular loss function.

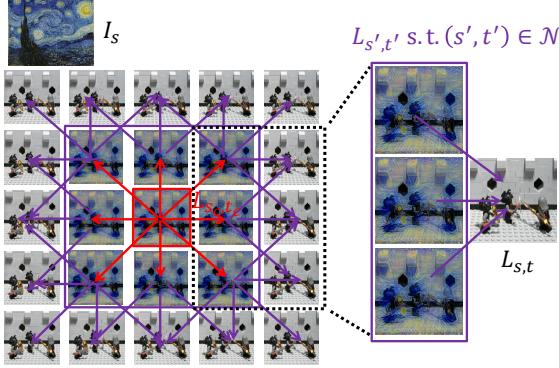


Fig. 2: Proposed approach: starting with the centre view L_{s_c, t_c} , the style transfer is iteratively propagated outwards while always ensuring that angular consistency with previously stylized view points is preserved. Red and purple arrows illustrate the first and second propagation iterations respectively.

2.2.1. Initialisation

The centre view L_{s_c, t_c} is stylized using the single image approach described in the previous section and is thus initialised with white noise. For all other views, we warp each previously stylized neighbouring view $L_{s', t'}^S$ according to the optical flow between the original views $L_{s', t'}$ and $L_{s, t}$. We then initialise $L_{s, t}^S$ to be a weighted sum of these warped views:

$$L_{s, t}^S = \frac{1}{\sum_{(s', t') \in \mathcal{N}} w_{s', t'}^{s, t}} \sum_{(s', t') \in \mathcal{N}} w_{s', t'}^{s, t} \omega_{s', t'}^{s, t}(L_{s', t'}^S) \quad (1)$$

where \mathcal{N} is the set of neighbouring views that have already been stylized. The weights are computed so as to give more importance to angularly close viewpoints:

$$w_{s', t'}^{s, t} = ((s - s')^2 + (t - t')^2)^{-\frac{1}{2}}$$

Finally, $\omega_{s', t'}^{s, t}$ is the function that warps a given image according to the disparity between the original views $L_{s', t'}$ and $L_{s, t}$. As used by Ruder et al. for their video style transfer [11, 13], we use DeepFlow [23] to estimate the disparity between the light field views $L_{s', t'}$ and $L_{s, t}$.

2.2.2. Angular loss function

We define the *angular loss* \mathcal{L}_a for the view $L_{s, t}^S$ to be:

$$\mathcal{L}_a(L_{s, t}^S) = \sum_{(s', t') \in \mathcal{N}} \left(\sum_{u, v} c_{s', t'}^{s, t} \cdot (\omega_{s', t'}^{s, t}(L_{s', t'}^S) - \bar{L}_{s, t})^2 \right)$$

where $c_{s', t'}^{s, t}$ are per-pixel weights defined to be 0 for disoccluded regions between views $L_{s', t'}$ and $L_{s, t}$ and 1 elsewhere and $\omega_{s', t'}^{s, t}$ is defined as above. The weights $c_{s', t'}^{s, t}$ are calculated using a consistency check between the forward and backward disparity warping. Similar to Ruder et al. [11], we use the consistency check provided in [24]. The check is based on the fact that in non-disoccluded regions the backward disparity should be the opposite direction of the forward disparity.

Finally, to stylize a view $L_{s, t}$ (different from the centre view L_{s_c, t_c}), we first initialise $L_{s, t}^S$ according to (1). We then use

gradient descent to iteratively update $L_{s, t}^S$ so as to minimise the total loss function:

$$\begin{aligned} \mathcal{L}_{LF}(L_{s, t}, I_S, L_{s, t}^S) &= \alpha_{LF} \mathcal{L}_c(L_{s, t}, L_{s, t}^S) + \beta_{LF} \mathcal{L}_s(I_S, L_{s, t}^S) \\ &\quad + \gamma_{LF} \mathcal{L}_a(L_{s, t}^S) \end{aligned}$$

where \mathcal{L}_c and \mathcal{L}_s refer to the content loss and the style loss as defined in the previous section, and \mathcal{L}_a is the proposed angular loss function. In our experiments, we used $\alpha_{LF} = 5.0$ or 50 , $\beta_{LF} = 1 \times 10^3$ or 1×10^4 , and $\gamma_{LF} = 2 \times 10^5$ or 4×10^5 depending on the light field and style image used. Together, our new initialisation method and angular loss function help to preserve the angular structure of the light field during the stylization process.

3. EXPERIMENTS AND RESULTS

In this section, we first describe the metrics we propose to assess for light field style transfer quality evaluation. We then evaluate our proposed approach against the baselines and the existing light field approach using direct visual observation and compare the scores given by the metrics.

3.1. Light Field Angular Consistency metric

Given the subjective nature of evaluating stylization itself, we instead focus on evaluating the angular consistency of the stylized light fields. In [16], Hart et al. chose to compare the disparity map estimated from the stylized light field to the one estimated from the original light field. In addition, we propose to use in our evaluation the recently introduced Light Field Epipolar Consistency metric (LFEC) [22], designed to evaluate the angular consistency for light field video interpolation. The LFEC back-warps the light field views to the centre view to compute a variance map which is then used in a PSNR-like computation to output the metric score (higher is better). As back-warping to the centre view cannot take into account disocclusions for views far from the centre, we introduce in this paper a variant of this metric by computing local angular variance maps, which we denote Light Field Angular Consistency metric (LFAC). While the original LFEC implementation relies on the ground truth disparity map to perform the back-warping, we use in this paper the disparity maps estimated from the original light field. To compute the proposed LFAC, a variance map σ^2 is first obtained as:

$$\sigma^2(L) = \frac{1}{MN} \sum_{s, t}^{M, N} \frac{1}{|\mathcal{C}_{s, t}|} \sum_{(s', t') \in \mathcal{C}_{s, t}} c_{s', t'}^{s, t} \left(\omega_{s', t'}^{s, t}(L_{s', t'}) - \bar{L}_{s, t} \right)^2$$

where M, N is the number of horizontal and vertical views respectively, $\mathcal{C}_{s, t}$ contains the closest neighbouring view indices of view $L_{s, t}$, $c_{s', t'}^{s, t}$ are per-pixel weights defined to be 0 for disoccluded regions between views $L_{s', t'}$ and $L_{s, t}$ and 1 elsewhere, $\omega_{s', t'}^{s, t}$ is the function that warps a given image according to the disparity between the neighbouring view $L_{s', t'}$ and the view $L_{s, t}$, $\bar{L}_{s, t}$ is the average of all the light field neighbouring views warped to the view $L_{s, t}$:

$$\bar{L}_{s, t} = \frac{1}{|\mathcal{C}_{s, t}|} \sum_{(s', t') \in \mathcal{C}_{s, t}} c_{s', t'}^{s, t} \omega_{s', t'}^{s, t}(L_{s', t'})$$

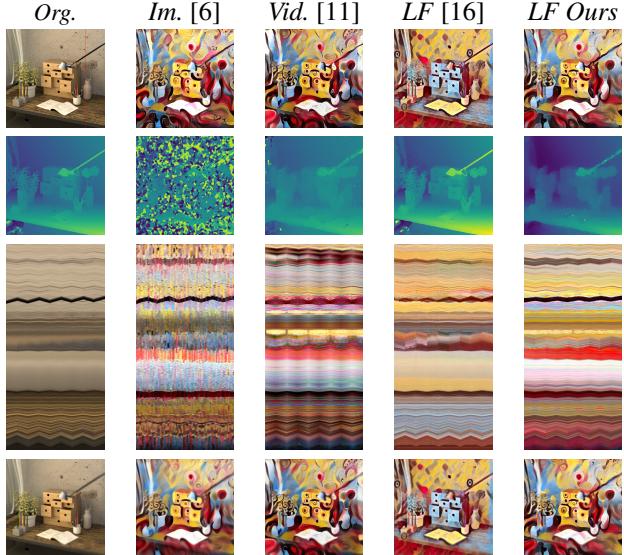


Fig. 3: Visual results on test material D. From top to bottom: centre view, estimated disparity map, stacked vertical EPI, refocused centre view.

The final metric is obtained as:

$$\text{LFAC}(L) = 10 \log_{10} \left(d^2 / \overline{\sigma^2}(L) \right)$$

where d is the range of the pixel values and $\overline{\sigma^2}(L)$ is the average of $\sigma^2(L)$.

3.2. Evaluation

In our experiments we used 6 light fields from 4 different datasets [17, 18, 25, 26], covering a range of various textures and disparity ranges, summarised in Table 1. Note that the code of Hart et al. [16] does not provide a pre-trained network for the *Starry night* style and thus we could not use it for comparison for test material C. We denote here the first image-based baseline as *Im.*, the video-based baseline as *Vid.*, the method of Hart et al. as *LF*, and our proposed approach as *LF Ours*.

Visual results for test material D are given in Fig. 3, showing the centre view, the estimated disparity map, a stacked vertical EPI, and a refocused centre view, for each of the evaluated methods and the original light field (*Org.*). It is clear from the EPIs of the image-based baseline that the angular consistency is not preserved. This can also be seen in the refocused image where the target style appears blurry even in the in-focus area. The video baseline greatly improve the consistency compared to the image baseline, but only along the horizontal snake-like path used to create the pseudo-videos. Thus, many artifacts are still present and clearly visible in the stacked vertical EPIs. Both light field methods clearly outperform the image and video baselines in terms of angular consistency. Note that these observations hold for all the test light fields used in our experiments. However, as illustrated in Fig. 3, we observe more artifacts in Hart et al.’s method than in ours for light fields with a wide disparity range such as test material D, E, and F, but no noticeable difference otherwise, as for test material A and B. Note that all the observations mentioned above are easier to visualize in videos created from the stylized viewpoints or focal stacks, which are

Table 1: Test material for our experiments

	Content	Style	Disparity range
A	Swan [17]	Candy	[−1, 1]
B	Bikes [18]	Rain princess	[−1, 1]
C	Herbs [26]	Starry night	[−3, 1.8]
D	Table [26]	Candy	[−2, 1.6]
E	Lego knights [25]	Rain princess	[−3, 3]
F	Crystal ball [25]	Mosaic	[−8, 4]

Table 2: Disparity MSE*100 (↓)

	<i>Im. [6]</i>	<i>Vid. [11]</i>	<i>LF [16]</i>	<i>LF Ours</i>
A	209.19	2.56	1.01	2.11
B	29.28	2.09	0.91	1.94
C	96.08	11.68	n/a	8.31
D	284.98	8.45	1.39	5.20
E	256.36	22.15	13.14	19.50
F	4182.80	247.60	58.00	287.50

Table 3: LFEC / LFAC (↑)

	<i>Im. [6]</i>	<i>Vid. [11]</i>	<i>LF [16]</i>	<i>LF Ours</i>
A	18.94 / 29.58	24.38 / 40.87	30.31 / 46.48	27.05 / 42.49
B	21.64 / 33.18	22.35 / 36.63	35.63 / 44.78	26.56 / 40.58
C	22.34 / 33.11	24.39 / 38.32	n/a	27.22 / 40.58
D	18.72 / 29.55	21.24 / 36.19	28.11 / 41.09	25.24 / 39.10
E	19.23 / 30.54	20.32 / 34.92	23.21 / 36.95	22.71 / 37.45
F	15.46 / 27.36	15.26 / 28.62	18.87 / 27.24	15.80 / 28.93

available as additional results on our webpage.¹

Similar to Hart et al., we report the Mean Squared Error (MSE) between the disparity map estimated from the stylized light field and from the original one using [5] in Table 2. The LFEC and LFAC metrics are reported in Table 3. All metrics show that the video-based baseline outperform the image-based baseline significantly, and that both light field methods outperform the video-based baseline. However, the disparity MSE and the original LFEC metrics give better scores to the Hart et al. approach for all light fields, in contradiction with the observations made on test material D, E, and F. The proposed LFAC variant reports correctly on the test materials E and F which have the highest disparity range and are better handled by our method. Thus the metrics are able to discriminate when there is an obvious gap in the stylization angular consistency, but less so when the results are comparable.

4. CONCLUSION AND FUTURE WORK

We introduced in this paper a novel optimization-based approach for light field style transfer enforcing the local angular consistency. Our qualitative assessment shows that the proposed approach outperforms existing methods in terms of angular consistency, especially for sparser light fields. However, we found through our experiments that existing light field metrics like the disparity MSE and the LFEC are not suitable for quantitative evaluation of light field style transfer. While the proposed LFAC metric is better at discriminating for sparse light fields, it remains ambiguous in some cases. These findings motivate future work where novel metrics for light field style transfer need to be thoroughly investigated, and compared to subjective evaluations.

¹v-sense.scss.tcd.ie/research/neural-style-transfer-for-light-fields/

5. REFERENCES

- [1] Marc Levoy and Pat Hanrahan, “Light field rendering,” in *Proc. SIGGRAPH*, 1996, pp. 31–42.
- [2] Aaron Isaksen, Leonard McMillan, and Steven J. Gortler, “Dynamically reparameterized light fields,” in *Proc. SIGGRAPH*, 2000, pp. 297–306.
- [3] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, “Light Field Photography with a Hand-Held Plenoptic Camera,” Tech. Rep., CSTR 2005-02, Apr. 2005.
- [4] D. G. Dansereau, . Pizarro, and S. B. Williams, “Linear volumetric focus for light field cameras,” *ACM Trans. Graph.*, vol. 34, pp. 15:1–15:20, 2015.
- [5] Changha Shin, Hae-Gon Jeon, Youngjin Yoon, In So Kweon, and Seon Joo Kim, “Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images,” in *Proc. IEEE CVPR*, 2018, pp. 4748–4757.
- [6] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge, “Image style transfer using convolutional neural networks.,” *Proc. IEEE CVPR*, pp. 2414 – 2423, 2016.
- [7] K. Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2015.
- [8] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution.,” *Proc. ECCV*, p. 694, 2016.
- [9] Xiao-Chang Liu, Ming-Ming Cheng, Yu-Kun Lai, and Paul L. Rosin, “Depth-aware neural style transfer,” in *Proc. NPAR*, 2017.
- [10] Weifeng Chen, Zhao Fu, Dawei Yang, and Jia Deng, “Single-image depth perception in the wild,” in *Proc. NIPS*, 2016, pp. 730–738.
- [11] Manuel Ruder, Alexey Dosovitskiy, and Thomas Brox, “Artistic style transfer for videos,” in *German Conference on Pattern Recognition*. Springer, 2016, pp. 26–36.
- [12] Ondřej Jamriška, Šárka Sochorová, Ondřej Texler, Michal Lukáč, Jakub Fišer, Jingwan Lu, Eli Shechtman, and Daniel Sýkora, “Stylizing video by example,” *ACM Trans. Graph.*, vol. 38, no. 4, July 2019.
- [13] Manuel Ruder, Alexey Dosovitskiy, and Thomas Brox, “Artistic style transfer for videos and spherical images.,” *International Journal of Computer Vision*, vol. 126, no. 11, pp. 1199 – 1219, 2018.
- [14] Xinyu Gong, Haozhi Huang, Lin Ma, Fumin Shen, Wei Liu, and Tong Zhang, “Neural stereoscopic image style transfer,” in *Proc. ECCV*, 2018, pp. 54–69.
- [15] Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua, “Stereoscopic neural style transfer,” in *Proc. IEEE CVPR*, 2018, pp. 6654–6663.
- [16] David Hart, Bryan Morse, and Jessica Greenland, “Style transfer for light field photography,” in *Proc. IEEE WACV*, 2020, pp. 99–108.
- [17] Martin Rerabek and Touradj Ebrahimi, “New Light Field Image Dataset,” in *Proc. QoMEX*, 2016.
- [18] Abhilash Sunder Raj, Michael Lowney, Raj Shah, and Gordon Wetzstein, “Stanford lytro light field archive,” <http://lightfields.stanford.edu/LF2016.html>, accessed: 06-10-2020.
- [19] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang, “Universal style transfer via feature transforms,” in *Advances in neural information processing systems*, 2017, pp. 386–396.
- [20] Likun Shi, Shengyang Zhao, and Zhibo Chen, “BELIF: Blind quality evaluator of light field image with tensor structure variation index,” in *Proc. IEEE ICIP*, 2019, pp. 3781–3785.
- [21] Ali Ak, Suiyi Ling, and Patrick Le Callet, “No-reference quality evaluation of light field content based on structural representation of the epipolar plane image,” in *Proc. IEEE ICMEW*, 2020, pp. 1–6.
- [22] P. David, M. L. Pendu, and C. Guillemot, “Angularly consistent light field video interpolation,” in *Proc. IEEE ICME*, 2020, pp. 1–6.
- [23] Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, and Cordelia Schmid, “DeepFlow: Large displacement optical flow with deep matching,” in *Proc. IEEE ICCV*, 2013.
- [24] N. Sundaram, T. Brox, and K. Keutzer, “Dense point trajectories by GPU-accelerated large displacement optical flow,” in *Proc. ECCV*, 2010.
- [25] “The stanford light field archive,” <http://lightfield.stanford.edu/lfs.html>, accessed: 06-10-2020.
- [26] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke, “A dataset and evaluation methodology for depth estimation on 4D light fields,” in *Proc. ACCV*, 2016.