Toward Neural Light-Field Compression

Yuya Ishikawa
Nagoya University

Chihiro Tsutake
Nagoya University

Keita Takahashi
Nagoya University

Toshiaki Fujii
Nagoya University

***Abstract*—We propose a data compression method for a light field via a compact and computationally-efficient neural representation. We experimentally show that our method achieves a promising rate-distortion performance.**

***Index Terms*—Light field, Compression, Neural representation**

I. Introduction

A light field (a set of multi-view images of a target 3-D scene) has been utilized for various applications such as 3-D display and post-capture refocusing. Due to the large data amount, data compression for a light field is an important research issue. For this purpose, we investigate to use a neural representation [1], [2], a neural network trained to reproduce the target light field; the network parameters are quantized, entropy-coded, and transmitted to the remote site.

Our method is significantly different in concept from the traditional approaches where a light field is encoded as a set of images or a video (as a pseudo-temporal sequence) using off-the-shelf image/video codecs. Moreover, we can generate arbitrary views directly from the neural representation, which is impossible with the traditional approaches. Shi et al. [3] used a neural representation of a 3-D scene called NeRF [1] for data compression. However, a NeRF is computationally expensive, and thus, impractical for low-power devices. In contrast, our method is implemented on a more computationally-efficient neural representation called SIGNET [2], but still achieves a promising rate-distortion performance.

II. Proposed Method

As shown in Fig. 1, the target light field is written as $L(u, v, x, y)$ where (u, v) and (x, y) denote the viewpoint and pixel. We use a neural network with trainable parameters Θ:

$$F_\Theta \colon (u, v, x, y) \longrightarrow \hat{L}(u, v, x, y) \quad (1)$$

which takes the coordinate (u, v, x, y) as input and directly regresses the corresponding pixel value $\hat{L}(u, v, x, y)$. The network is trained to minimize the regression loss as

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \sum_{u,v,x,y} \|L(u, v, x, y) - F_\Theta(u, v, x, y)\|^2 \quad (2)$$

Once the training is finished, all we need to reconstruct $L(u, v, x, y)$ is the set of the network parameters, Θ^*. Therefore, Θ^* is quantized, entropy-coded, and transmitted to the remote site. For this purpose, we develop an adaptive quantization scheme that can control the trade-off between the bit-rate (data size) and reconstruction quality. Since F_Θ can be queried at arbitrary (u, v), arbitrary views can be rendered directly from the nueral representation.

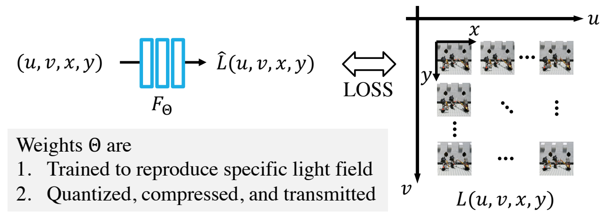

Fig. 1. Overview of proposed method.

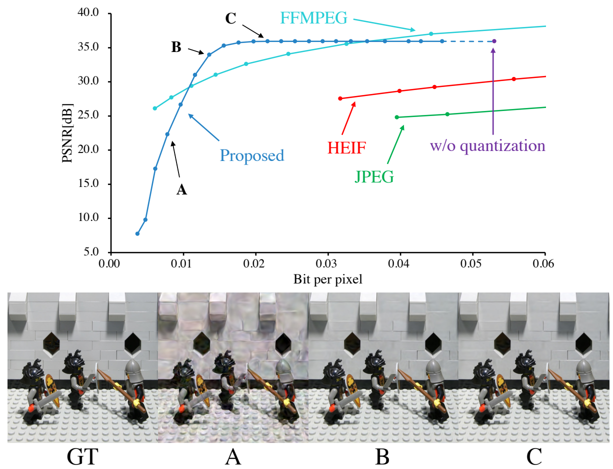

Fig. 2. Rate-distortion performance and some visual results.

III. Results and Discussion

We trained a SIGNET-based network on a light field with 9×9 views, applied our compression method, and synthesized 17×17 views (with viewpoint interpolation) from the compressed data. In Fig. 2, we compared the rate-distortion performance of our method against the traditional approaches including image-based coding (JPEG and HEIF) and video-based-coding (FFMPEG). As can be seen, our method outperformed them. We also present some visual results obtained with our method. For the case at "B" on the graph, the compressed Θ^* takes only 513k bytes, which is approximately 0.056 % of the original size of 17×17 views (909M bytes).

References

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. Ng: "NeRF: Representing scenes as neural radiance fields for view synthesis," ECCV, 2020

[2] B. Y. Feng, A. Varshney: "SIGNET: Efficient neural representation for light fields," ICCV, 2021

[3] J. Shi, C. Guillemot, "Light Field Compression Via Compact Neural Scene Representation," ICASSP, 2023