Compression Artifact Reduction for 360-degree Images using Reference-based Deformable Convolutional Neural Network

Hee-Jae Kim, Je-Won Kang, Byung-Uk Lee
Department of Electronic and Electrical Engineering, Ewha Womans University
Graduate Program in Smart Factory
heejaekim@ewhain.net, jewonk@ewha.ac.kr, bulee@ewha.ac.kr

Abstract

In this paper, we propose an efficient reference-based compression artifact reduction network for 360-degree images in an equi-rectangular projection (ERP) domain. In our insight, conventional image restoration methods cannot be applied straightforwardly to 360-degree images due to the spherical distortion. To address this problem, we propose an adaptive disparity estimator using a deformable convolution to exploit correlation among 360-degree images. With the help of the proposed convolution, the disparity estimator establishes the spatial correspondence successfully between the ERPs and extract matched textures to be used for image restoration. The experimental results demonstrate that the proposed algorithm provides reliable high-quality textures from the reference and improves the quality of the restored image as compared to the state-of-the-art single image restoration methods.

1. Introduction

Deep learning (DL) based image restoration has been studied actively in computer vision. The performance has been improved substantially with several DL modules consisting of reconstruction blocks [1, 2]. The quality of an image could be improved further via reference-based image restoration (RefIR) [3, 4, 5] methods, by transferring high-quality textures from the reference to the current image. Although the RefIR seemed to be a plausible solution for 360-degree images to display omnidirectional views, successful models needed to establish accurate correspondence between the target and the reference images. Otherwise, RefIR could not utilize the corresponding high-quality textures, and, accordingly, the performance was severely degraded as compared with the single image restoration methods.

In this paper, we aim to develop a reference-based compression artifacts reduction (RefCAR) network to restore the quality of a 360-degree image from the compression artifacts, by referring to reliable textures in the reference.

360-degree images contain omnidirectional spatial information. When a spherical image is projected into a two-dimensional plane, equi-rectangular projection (ERP) is commonly used as a de facto standard. Although the projection incurs geometric distortion, the previous RefIR methods have neglected such the problems. To address this issue, we adopt a deformable convolution that is suitable to an ERP 360-degree image in the disparity estimator. Specifically, the proposed convolution adaptively extends the receptive field according to the latitude to successfully estimate the large and distorted disparity between two ERP images. Moreover, we subsequently remove misaligned textures in the warped reference feature with an estimated occlusion mask.

In the following Section 2, we describe the proposed network, the training details in Section 3, the experimental results in Section 4, and the conclusion in Section5.

2. Proposed Method

Figure 1 shows an overall architecture of the proposed network. It takes E^{LQ}, as a low-quality image to be enhanced, and E^{Ref}, as its reference for the input 360-degree image and produces E^{HQ} as the restored high-quality image for the output. E^{LQ} and E^{Ref} are both compressed with various quantization parameters.

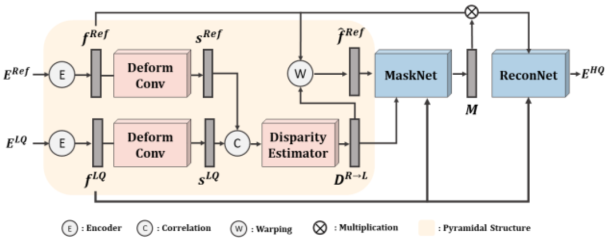

Figure 1 Overview of the proposed RefCAR network, including an adaptive disparity estimator and a deformable convolution operation.

To estimate the large and distorted disparity $D^{R \to L}$ in a coarse-to-fine manner, we adopt a pyramidal structure similar to [6]. In each level, we first generate robust features, s^{LQ} and s^{Ref}, by conducting a deformable convolution to the target feature f^{LQ} and the reference feature f^{Ref}, which are the features extracted from the conventional convolution layers. Then the correlation between the s^{LQ} and s^{Ref} is computed to be utilized as a feature of the disparity estimator. We align the reference feature to the target view by warping f^{Ref} to $\hat{f}^{Ref}$ using $D^{R \to L}$.

Note that the misaligned textures from the reference degrade the performance of RefCAR severely. In this regard, we subsequently estimate the occlusion mask M to adaptively utilize reliable high-quality information from the reference and suppress others.

3. Training Details

To conduct RefCAR in the ERP domain, we use Unity software to simulate 4,356 groups of target and references. We also take 251 groups of real 360-degree images using Samsung Gear360 in various indoor and outdoor environments. Then we adopt HEVC HM16.16 with various quantization parameters to generate compressed input. We only use Y channel in all experiments. The size of the image is 256×512 and 512×1024 for training and testing, respectively.

We first train the proposed network using all synthetic dataset and fine-tune the network with 186 real 360-degree images. We use the other 65 real image groups and the ERA [8] dataset for testing. For RefCAR, we select appropriate full ERP format images from the ERA dataset.

We train the network with reconstruction loss, L_{rec}, and warping loss, L_{warp}. We adopt Charbonnier penalty function for both loss terms which are computed as follows:

$$L_{rec} = \sqrt{\|E^{GT} - E^{HQ}\|^2 + \rho^2} \quad (1)$$

$$L_{warp} = \sqrt{\left\|E^{GT} - \hat{E}^{Ref}\right\|^2 + \rho^2} \quad (2)$$

where E^{GT} and $\hat{E}^{Ref}$ are the ground truth and the warped reference ERP image, respectively. We set ρ to 0.001.

4. Experimental Results

We compared the proposed RefCAR method with RDN [7] as a state-of-the-art CAR method. We train the RDN with the same dataset. In Table 1, the proposed RefCAR is denoted by “Ours” and “Ours+” depending on the quality of the reference. Specifically, “Ours” utilized a low-quality reference image while the quality of the reference in “Our+” is relatively high. The proposed method outperforms RDN about 0.36 dB on the real dataset when using a high-quality reference. The experimental results show that the proposed RefCAR not only achieved better quantitative results but also provides visually pleasing results as shown in Figure 2.

Dataset	Real dataset		
Method	PSNR	SSIM	WS-PSNR
HEVC	30.80	0.848	27.41
RDN [7]	31.08	0.854	28.74
Ours	31.29	0.859	29.70
Ours+	**31.44**	**0.862**	**30.46**
Dataset	ERA [8] dataset		
Method	PSNR	SSIM	WS-PSNR
HEVC	30.93	0.852	28.70
RDN [7]	31.12	0.856	29.54
Ours	31.24	0.859	30.16
Ours+	**31.47**	**0.866**	**31.28**

Table 1 Quantitative comparison with RDN [7] on real and ERA dataset.

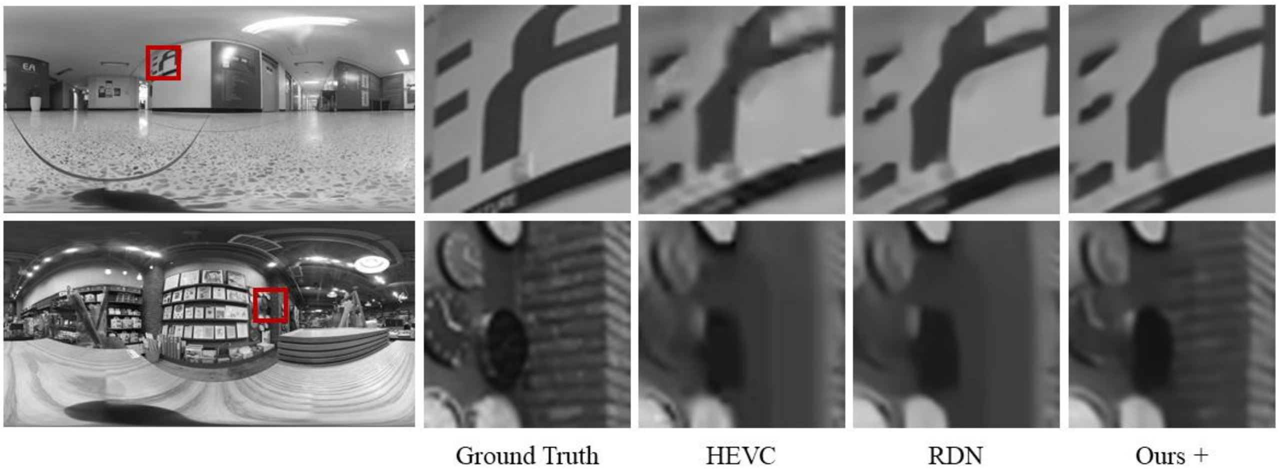

Figure 2 Qualitative comparison with RDN [7] on real dataset.

5. Conclusions

In this paper, we presented a reference-based compression artifact reduction algorithm for 360-degree images. The deformable convolutions generate robust features in the disparity estimator to search large and distorted disparity between the ERPs. Then we estimate the occlusion mask using convolutional layers to remove unreliable textures in the warped reference feature and improve the performance. The experimental results demonstrated that the proposed algorithm provides superior performance to the state-of-the-art restoration method.

Acknowledgement

This work has been supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(No.NRF-2019R1C1C1010249). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2020-0-00920, Development of Ultra High Resolution Unstructured Plenoptic Video Storage/Compression/Streaming Technology for Medium to Large Space)

Reference

[1] Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144).

[2] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV) (pp. 286-301).

[3] Zheng, H., Ji, M., Wang, H., Liu, Y., & Fang, L. (2018). Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In Proceedings of the European conference on computer vision (ECCV) (pp. 88-104).

[4] Zhang, Z., Wang, Z., Lin, Z., & Qi, H. (2019). Image super-resolution by neural texture transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7982-7991).

[5] Yang, F., Yang, H., Fu, J., Lu, H., & Guo, B. (2020). Learning texture transformer network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5791-5800).

[6] Sun, D., Yang, X., Liu, M. Y., & Kautz, J. (2018). Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8934-8943).

[7] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2020). Residual dense network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(7), 2480-2495.
[8] Yang, W., Qian, Y., Kämäräinen, J. K., Cricri, F., & Fan, L. (2018, August). Object detection in equirectangular panorama. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 2190-2195). IEEE.