SimFIR: A Simple Framework for Fisheye Image Rectification with Self-supervised Representation Learning

Hao Feng[1,2] Wendi Wang[1] Jiajun Deng[3] Wengang Zhou[1,4,*] Li Li[1] Houqiang Li[1,4,*]
[1] CAS Key Laboratory of Technology in GIPAS, EEIS Department, University of Science and Technology of China
[2] Zhangjiang Laboratory, Shanghai, China [3] The University of Sydney
[4] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
{haof,wendiwang}@mail.ustc.edu.cn, jiajun.deng@sydney.edu.au, {zhwg,lil1,lihq}@ustc.edu.cn

36	35	34	33	32	31	30	29	29	30	31	32	33	34	35	36
35	28	27	26	25	24	23	22	22	23	24	25	26	27	28	35
34	27	21	20	19	18	17	16	16	17	18	19	20	21	27	34
3	26	20	15	14	13	12	11	11	12	13	14	15	20	26	33
32	25	19	14	10	9	8	7	7	8	9	10	14	19	25	32
31	24	18	13	9	6	5	4	4	5	6	9	13	18	24	31
30	23	17	12	8	5	3	2	2	3	5	8	12	17	23	30
29	22	16	11	7	4	2	1	1	2	4	7	11	16	22	29
29	22	16	11	7	4	2	1	1	2	4	7	11	16	22	29
30	23	17	12	8	5	3	2	2	3	5	8	12	17	23	30
31	24	18	13	9	6	5	4	4	5	6	9	13	18	24	31
32	25	19	14	10	9	8	7	7	8	9	10	14	19	25	32
33	26	20	15	14	13	12	11	11	12	13	14	15	20	26	33
34	27	21	20	19	18	17	16	16	17	18	19	20	21	27	34
35	28	27	26	25	24	23	22	22	23	24	25	26	27	28	35
36	35	34	33	32	31	30	29	29	30	31	32	33	34	35	36

Figure 1. An illustration of our position map when dividing a 256×256 image with 16×16 patches.

Figure 2. Qualitative results of SimFIR on non-square fisheye images in high-resolution video surveillance scenarios.

A. Position Map

As shown in Fig. 1, we present the resulting position map generated utilizing our default setting, in which a 256×256 image is subdivided into non-overlapping regions, each with dimensions of 16×16. Within the position map, individual values signify particular distortion pattern classifications associated with the corresponding patches. Identical values within the map imply that the respective patches exhibit equivalent degrees of distortion.

B. Generalization to Non-square Images

By default, our method is trained and evaluated on square-shaped fisheye images, following existing methods in the field. Furthermore, we demonstrate the generalization capabilities of our method when applied to non-square images. As evidenced in Fig. 2, our SimFIR exhibits a robust ability to effectively rectify these types of distorted images. It is worth noting that these distorted images are commonly encountered in high-resolution video surveillance scenarios, further emphasizing the practical applicability of our method in real-world situations.

C. More Qualitative Comparisons

In this section, we supplement more qualitative comparisons with existing methods, including SC [45], Blind [33], DeepCalib [6], DR-GAN [35], DDM [36], MLC [34], and PCN [55]. As shown in Fig. 3 and Fig. 4, the images rectified by our method display reduced distortion while simultaneously preserving a greater degree of textural detail.

*Corresponding authors: Wengang Zhou and Houqiang Li

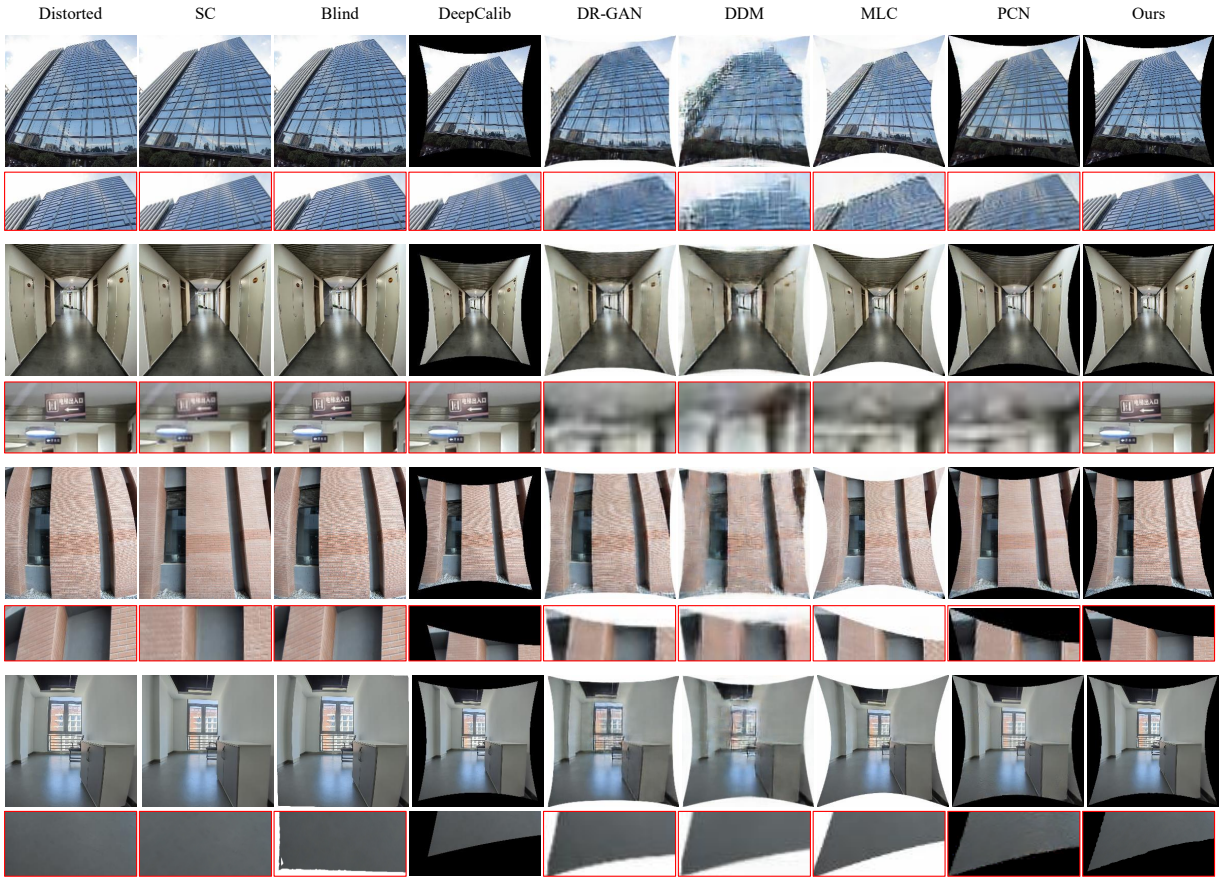

Figure 3. Qualitative comparisons on real-world fisheye images. For each comparison, we show the distorted image, the rectified results of SC [45], Blind [33], DeepCalib [6], DR-GAN [35], DDM [36], MLC [34], PCN [55], and our method. Besides, we magnify identical regions from each image, to facilitate a more detailed comparison between the methods.

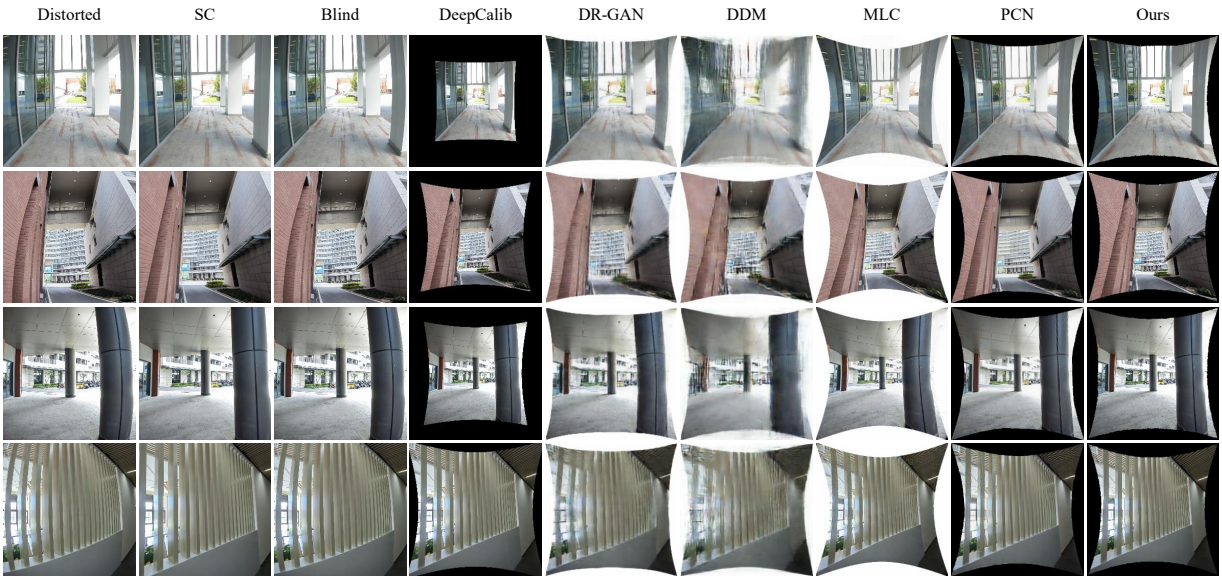

Figure 4. Qualitative comparisons on real-world fisheye images. For each comparison, we show the distorted image, the rectified results of SC [45], Blind [33], DeepCalib [6], DR-GAN [35], DDM [36], MLC [34], PCN [55], and our method.