Fisheye Rectified Image Super-resolution Using Deep Symmetrical Neural Network

Yuanfa Zhang¹, Weili Li², Xiaoqing Yin³* and Jinsheng Deng³

¹Naval Research Academy, Beijing, China.
²College of System Engineering, National University of Defense Technology, Changsha, China.
³College of Advanced Interdisciplinary Research, National University of Defense Technology, Changsha, China.
* Corresponding Author. Email: yinxiaoqing89@163.com

Abstract. Influenced by the fast developments of imaging technologies, fisheye cameras have been widely applied. But the resolution of images captured by fisheye cameras is limited due to lens distortions. In our work, we use a deep symmetrical network model to deal with single fisheye rectified image super-resolution task. It is shown by the experiments that the proposed method can achieve satisfactory results for fisheye rectified image super-resolution.

1. Introduction
Image super-resolution is one of the hot research areas with applications on safety surveillance [1], remote sensing [2] and image in-painting [3], etc. Fisheye cameras are often used to obtain panoramic images with large field-of-view. However, the resolution of fisheye images is relatively low due to lens distortions. Few research has been conducted on improving the resolution of fisheye images or similar types of images [4-6]. In a fisheye image, the resolution varies in different regions, which increase the difficulty of super resolution. Recently, deep learning based approaches has proved to be effective for image processing tasks, which provides new ideas for fisheye image super-resolution. Convolutional neural networks can be trained on large numbers of samples. Then the trained model can be applied to learn the mapping from input fisheye rectified image to the corresponding high-resolution result.

In this paper, we propose a deep symmetrical network model to improve the resolution of fish rectified image. Our approach can generate satisfactory results according to the experiments. The proposed method is described as follows.

2. Method
Our model is built as a symmetrical deep network with an encoding process and a decoding process. The numbers of feature maps as well as the kernel sizes for the network layers are shown in Table 1. In the proposed model, the original fisheye rectified image is the input and the corresponding super-resolution result is the output.

The encoding process includes convolution and pooling layers which processes the input fisheye rectified image and generates feature maps with various resolutions. Larger receptive field is obtained and the higher layers contains more global information, and detailed local information is contained in lower layers. On the other hand, the decoding process deals with the feature maps generated from the encoding process. The resolutions of feature maps are raised and the super-resolution result is generated.
As the input of the network is similar to the output, we add a residual link from the input to the output to accelerate the convergence of the network. The loss function in the training process is the L2 loss between the output image and the corresponding ground truth image.

**Table 1.** Number of feature maps and the kernel sizes for the network layers

| Encoding Process | Decoding Process |
|------------------|------------------|
| Network Layer    | Number of Feature maps | Kernel Size | Network Layer    | Number of Feature maps | Kernel Size |
| Conv1            | 32                | 3*3, stride 1 | Deconv1         | 256                | 4*4, stride 2 |
| Conv2            | 32                | 3*3, stride 1 | Conv_d1_1       | 256                | 3*3, stride 1 |
| Conv3            | 64                | 3*3, stride 2 | Conv_d1_2       | 256                | 3*3, stride 1 |
| Conv3_1          | 64                | 3*3, stride 1 | Deconv2         | 128                | 4*4, stride 2 |
| Conv4            | 128               | 3*3, stride 2 | Conv_d2_1       | 128                | 3*3, stride 1 |
| Conv4_1          | 128               | 3*3, stride 1 | Conv_d2_2       | 128                | 3*3, stride 1 |
| Conv5            | 256               | 3*3, stride 2 | Deconv3         | 64                 | 4*4, stride 2 |
| Conv5_1          | 256               | 3*3, stride 1 | Conv_d3_1       | 64                 | 3*3, stride 1 |
| Conv6            | 512               | 3*3, stride 2 | Conv_d3_2       | 64                 | 3*3, stride 1 |
| Conv6_1          | 512               | 3*3, stride 1 | Deconv2         | 32                 | 4*4, stride 2 |
|                  |                   |              | Conv_d4_1       | 32                 | 3*3, stride 1 |
|                  |                   |              | Conv_d4_2       | 32                 | 3*3, stride 1 |
|                  |                   |              | Conv_final       | 3                  | 3*3, stride 1 |
Figure 1. Super-resolution results.
From top to down: low-resolution fisheye rectified image, corresponding high-resolution ground truth, super-resolution results of DRCN [9] and super-resolution results of the proposed method.
3. Experiments
Firstly, fisheye rectified images are obtained based on the general fisheye model [7] with various parameter settings. Then simulated training and testing sets can be built. The training and testing dataset contain 20000 and 1000 samples, respectively. The resolution of the input fisheye rectified image and the output are 128*128 and 256*256 respectively. A training sample contains a fisheye rectified image and the corresponding super-resolution ground truth.

Our experiments are complemented based on Caffe [8]. The initial learning rate is 1e-4 and the number of iterations in the training process is 300K. After the training process is finished, we test the proposed model on the testing set. Figure 1 shows the satisfactory performance of the proposed method (From left to right: low-resolution fisheye rectified image, corresponding high-resolution ground truth and super-resolution results of the proposed method). To further test the performance of our method, we compare the proposed model with the widely used super-resolution algorithm DRCN [9]. The DRCN method is not able to deal with low-resolution problems caused by lens distortions. It is shown that details can be better recovered by the proposed model and the quality of fisheye rectified image can be effectively improved.

4. Conclusions
In this paper, we design a deep symmetrical network to improve the resolution of fisheye rectified image. It is shown by the experiments that our model can achieve satisfactory super-resolution results. In the following, we will improve our work in terms of network architecture and training strategy.

5. References
[1] L. Zhang, H. Zhang, H. Shen, P. Li, A super-resolution reconstruction algorithm for surveillance images, Signal Process. 90 (3) (2010) 848–859.
[2] D. Yang, Z. Li, Y. Xia, Z. Chen, Remote sensing image super-resolution: challenges and approaches, in: Proceedings of the IEEE International Conference on Digital Signal Processing (DSP), Singapore, 2015, pp. 196–200.
[3] O. Le Meur, M. Ebdelli, C. Guillemot, Hierarchical super-resolution-based inpainting, IEEE Trans. Image Process. 22 (10) (2013) 3779–3790.
[4] T. Takano, S. Ono, Y. Matsushita, H. Kawasaki, K. Lkeuchi, Super resolution of fisheye images captured by on-vehicle camera for visibility support, in: Proceedings of the IEEE International Conference on Vehicular Electronics and Safety, Yokohama, Japan, 2015, pp. 120–125.
[5] H. Nagahara, Y. Yagi, M. Yachida, Super-resolution modeling using an omnidirectional image sensor, IEEE Trans. Syst Man Cybern Part B (Cybernetics) 33 (4) (2003) 607–615.
[6] X. Shao, J. Xu, J. Wang, X. Chen, R. Gong, X. Bi, Design of a wide-field imaging optical system with super-resolution reconstruction, in: Proceedings of the SPIE 9501, Satellite Data Compression, Communications, and Processing XI, 2015.
[7] Juho Kannala and Sami Brandt. A generic camera calibration method for fisheye lenses. In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, volume 1, pages 10-13. IEEE, 2004.
[8] Yangqing Jia, Evan Shelhamer, Jef Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Cafe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, pages 675-678. ACM, 2014.
[9] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee, Deeply-Recursive Convolutional Network for Image Super-Resolution, Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.