Deep learning-based image reconstruction for multi-aperture diffractive lens

A V Nikonorov\textsuperscript{1,2}, R V Skidanov\textsuperscript{1,2}, V V Evdokimova\textsuperscript{1,2}, M V Petrov\textsuperscript{1,2}, A P Alekseyev\textsuperscript{1,2}, S A Bibikov\textsuperscript{1,2} and N L Kazanskiy\textsuperscript{1,2}

\textsuperscript{1}Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
\textsuperscript{2}Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: vkutikova94@gmail.com

Abstract. Multi-aperture imaging systems are one of the modern trends for imaging devices. The paper presents the results of developing a multi-aperture imaging system based on single diffractive lenses and postprocessing pipeline based on deep learning. The pipeline includes two processing steps: color correction using a generative adversarial network and chromatic deblurring by a convolutional neural network.

1. Introduction

Recently, there has been growing interest in multi-aperture imaging systems. In this paper, we address the problem of restoration of images captured by a multi-aperture imaging system based on long-focus single diffraction lenses. Diffractive optical elements are lighter and less expensive to manufacture than the conventional refractive optics. However, the quality of these images suffers from strong distortions of various origins. Reconstruction of images taken with the single-lens diffractive system has already been presented in previous works \cite{1, 2}, but \cite{3} is the first to use diffractive optics in the multi-aperture imaging system.

In this paper, we suggest fully convolutional image reconstruction pipeline which consists of a generative adversarial network (GAN) and a convolutional neural network (CNN). The pipeline restores images according to the distortion model described in \cite{4}. Pix2Pix \cite{5} – the state-of-art image-to-image GAN – corrects the color shift. To remove chromatic blur we use the 6-channel VDSR modification which was successfully applied in \cite{3}. The reconstruction pipeline differs from the prior works [1-3] in the color correction technique that was inspired by [6]. In [6] GANs had successfully applied to generate the white balanced images. The best results were achieved using Pix2Pix we had taken.

In summary, we propose the combination of Pix2Pix and 6-channel VDSR to restore images captured by the multi-aperture diffractive imaging system. Our reconstruction technique allows achieving a mean PSNR value about 25.93 dB that outperforms the results obtained in \cite{3}. 
2. Optical schemes of multi-aperture vision systems based on diffractive optics

As shown in previous works, diffractive optics can significantly reduce the weight of long-focus lenses [7]. Such systems are extremely relevant for solutions in such areas as space remote sensing, light UAVs, and video surveillance systems [1]. For the multi-aperture systems, the weight of the optical part is critical. Carles at al. decreased the optical path length by four times using nine apertures in the IR range [8]. However, the complexity of manufacturing and the total weight of the system led to the fact that the authors present only theoretical calculations of such a system. In this paper, we propose two schemes of the multi-aperture systems based on diffractive optics aimed at widening the viewing angle of the system and at increasing the resolution of the system.

The schematic diagrams of the lenses are shown in figures 1 – 4. As a basic optical scheme, in figure 1 a simple diffractive lens is shown. Such a long-focus diffractive lens is a component of two other optical schemes.

The general drawing of a multi-shutter multi-aperture lens is shown in figure 2. In this scheme, images from three parts of the observed object are transmitted to the sensor through three diffractive lenses. From the central lens, the image is obtained directly by the sensor, while from the other two lenses, images deflected from the axis are obtained. The shutters placed behind the lenses are triggered sequentially, so images of three areas of the object are sequentially formed on the sensor, forming a composite frame of individual frames captured at different consecutive time intervals.

In the proposed scheme, we use a feature of the diffraction lens, consisting in the fact that the optical path in such a lens is 99.9% free of optical elements. Thus, when assembling a complex imaging system, optical paths may intersect, which is impossible in a conventional lens consisting of a large number of refractive optical elements [9]. Also we use the fact that when shooting 25 frames per second, the formation of each frame occurs within 40 ms. At the same time, typical shutter speeds sufficient for video recording in an open space are 2–5 ms. In such timing, a multi-lens configuration can provide for the formation of a composite frame from 8–20 individual frames, while composite frames will be formed at a frequency of 25 frames/sec. The image deflection in such systems is performed by known method [1, 9], an example of the relief of a diffractive element with a deflection feature is shown in figure 4.

Figure 3 shows a schematic diagram of a multi-aperture lens aimed at improving image resolution through the use of the multi-frame superresolution approach. In such a scheme, each individual lens of a multi-aperture scheme ensures the registration of an image on a part of the sensor [10]. In this case, the overall image with a higher resolution is obtained as a result of computational reconstruction. This paper discusses the reconstruction of images using a convolutional neural network.
3. Image reconstruction in multi-aperture diffractive optical systems

In this work, we use the same optical system as in [3]. The system consists of two independent diffractive lenses (the right and the left lenses) which produce images of a real scene from different positions. The images captured by the right lens are transformed into the coordinate system of the left images. Further image reconstruction scheme is shown in figure 5.

Matching captured images with ground truth is described in early works [1, 4]. After the image matching, the color shift is corrected using GAN in the color correction stage. Finally reconstructed images are obtained in the CNN processing stage which removes the chromatic blur [3].

3.1. Color correction by GAN

A GAN consists of a generator $G$ and a discriminator $D$ which are convolutional neural networks. Figure 6 shows a typical scheme of GANs. $G$ takes as input $X$, an RGB image with color shift, and outputs $Y$, the same color corrected image. The generated output is then passed to $D$ which classifies whether the input is ground truth image or is generated by $G$. The output signal provided by $D$ feed to $G$ to learn to transform distorted images to color corrected images. In this paper, we use Pix2Pix GAN [5] based on U-Net [11] architecture as the generator and fully convolutional PatchGAN [6] as the discriminator. The network is trained with the following adversarial loss [5]:

$$L_{GAN}(G, D, X, Y) = E_{y \sim p_{data}(y)} \left[ \log D(y) \right] + E_{x \sim p_{data}(x)} \left[ \log (1 - D(G(x))) \right],$$

where $X$ is an RGB image captured by the diffractive lens, and $Y$ is a ground-truth image. The discriminator $D$ receives the generated image from the generator $G$ and predicts whether it was generated image $G(x)$ or ground-truth $y$. The generator $G$ has the L1 loss to produce images closed to the ground truth ones:

$$L_{L1}(G) = \|y - G(x)\|_1.$$

3.2. CNN-based reconstruction

Figure 7 illustrates the CNN architecture, which is based on the VDSR network [12]. The PReLU is chosen as an activation function. The loss function uses the grey-edge penalty and has the same form as in [3]:

$$L(\hat{p}, p, G_0, w) = \frac{1}{N} \sum_{j=1}^N d\left(\hat{p}^{(j)} - p^{(j)}\right) + \lambda \frac{1}{N} \sum_{j=1}^N \left\|G_0^{(j)} \nabla \hat{p}^{(j)} - \nabla p^{(j)} \right\|_1 + \eta \|w\|_1,$$
where \( N \) is the number of images in one training minibatch, \( j \) is the index of an image in the sample, \( p^{(i)}_j \) is the ground truth RGB image, \( \hat{p}^{(i)}_j \) is the reconstructed RGB image, \( p^G_j \) is the green channel of the reference image. The first term of the loss function minimizes the distance \( d(\cdot) \) between the ground truth image and the network output image. The second term implements the penalty for deviation of the gradient from the green channel of the ground truth image. The last term applies constraints on the network weights, following the weight decay rule.

Figure 6. Scheme of GAN.

In the last layer of the network (the addition layer), we offer to use adding the output of the previous layer to a linear combination of input images:

\[
\text{Addition} = Y_{\text{prev}} + X_{\text{in}},
\]

\[
X_{\text{in}} = (1 - \gamma) X_{\text{left}} + \gamma X_{\text{right}},
\]

(1)

where \( Y_{\text{prev}} \) is the output of the previous layer, \( X_{\text{left}} \) is an RGB image produced by the left lens, \( X_{\text{right}} \) is an RGB image produced by the right lens, \( \gamma \) is a training parameter. This combination of input images was used in work [3] but the parameter \( \gamma \) was constant and set to zero. That is, only the left frame \( X_{\text{in}} = X_{\text{left}} \) fed to the addition layer. Also, we considered such combinations as

\[
X_{\text{in}} = X_{\text{left}} + \gamma X_{\text{right}},
\]

\[
X_{\text{in}} = \gamma X_{\text{left}} + X_{\text{right}},
\]

(2)

but (1) allows achieving better results than (2) or (1) with \( \gamma = 0 \).

The reconstruction quality is estimated using the peak signal-to-noise ratio (PSNR) which proved to be effective in [1-4, 7].

4. Results
We conducted our studies on the dataset, which consists of 118 training images and 29 testing images with corresponding ground truth ones. We present experiments trained on the training set and evaluated on the test set.

The Pix2Pix network was trained on 1024x1024 training images. We used minibatch stochastic gradient descent and applied the Adam solver [13], with a learning rate of 0.0002, and momentum parameters \( \beta_1 \) and \( \beta_2 \) set to 0.5 and 0.999 respectively.

Before VDSR training all images were divided into fragments of 51x51 size with striding of 51. In addition, data augmentation (rotation and adding Gaussian noise) was used. We used the same training parameters initialization as in [3] except the network depth. The VDSR model with the depth set to 18 allowed achieving the highest PSNR value.
Table 1. PSNR evaluation results.

| Epoch | Reconstructed image (dB) | Input left frame (dB) | Input right frame (dB) | Color corrected left frame (dB) | Color corrected right frame (dB) |
|-------|--------------------------|-----------------------|------------------------|---------------------------------|---------------------------------|
| 0     | 22.12                    |                       |                        |                                 |                                 |
| 1     | 23.70                    |                       |                        |                                 |                                 |
| 2     | 24.13                    |                       |                        |                                 |                                 |
| 3     | 25.17                    |                       |                        |                                 |                                 |
| 4     | 25.80                    |                       |                        |                                 |                                 |
| 5     | 25.83                    | 17.21                 | 17.19                  | 25.40                           | 25.09                           |
| 6     | 25.86                    |                       |                        |                                 |                                 |
| 7     | 25.88                    |                       |                        |                                 |                                 |
| 8     | 25.91                    |                       |                        |                                 |                                 |
| 9     | 25.93                    |                       |                        |                                 |                                 |
| 10    | 25.93                    |                       |                        |                                 |                                 |

Figure 8. Image examples – a) ground truth; b) captured left; c) captured right; d) reconstructed.

At inference time, the Pix2Pix network takes $1024 \times 1024$ images with the color shift as input and outputs $1024 \times 1024$ color corrected images. Then the VDSR model removes the chromatic blur from color corrected images. It can be run on images with different size as input.

Table 1 shows the performance of our reconstruction pipeline. The mean PSNR was calculated for input color corrected left and right images and reconstructed images from the testing set.
This reconstructed method for multi-aperture diffractive optical systems outperforms our last results in [3] where we have achieved the mean PSNR equaled to 25.51 dB on 5 testing images. Figure 8 demonstrates visual reconstruction quality. It contains examples of input, ground truth, and reconstructed images.

5. Conclusion
In this work, we have presented a fully convolutional image reconstruction pipeline for multi-aperture diffractive optical system. We have demonstrated that the combination of generative adversarial network (GAN) and convolutional neural network (CNN) outperforms the reconstruction method described in [3]. We suggest using the Pix2pix architecture GAN to correct color shift and the VDSR modification to remove chromatic blur. As a result of the experiments, we found that the CNN consisting of 18 convolutional cascades provides the best reconstruction quality and stable convergence of training. The mean PSNR value we have achieved is about 25.93 dB and is calculated using 29 testing images.

Our work shows that using the GAN for color correction problem allows improving image reconstruction quality. The improving of postprocessing pipeline for multi-aperture systems based on simple diffraction lens system is promising for further research.

6. References
[1] Nikonorov A, Petrov M, Bibikov S, Yakimov P, Kutikova V, Yuzifovich Y, Morozov A, Skidanov R and Kazanskiy N 2018 IEEE JSTARS 11 3338-3348 DOI: 10.1109/JSTARS.2018.2856538
[2] Nikonorov A, Petrov M, Bibikov S, Kutikova V, Yakimov P, Morozov A, Skidanov R and Kazanskiy N 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (Beijing) 1-9 DOI: 10.1109/PRRS.2018.8486408
[3] Nikonorov A, Skidanov R, Kutikova V, Petrov M, Alekseyev A, Bibikov S and Kazanskiy N 2019 Proc. SPIE 11146 111460Y DOI: 10.1117/12.2526923
[4] Nikonorov A, Skidanov R, Fursov V, Petrov M, Bibikov S and Yuzifovich Y 2015 Proc. of Conf. on Computer Vision and Pattern Recognition Workshops (Boston) 33-41
[5] Isola P, Zhu J-Y, Zhou T and Efros A 2018 Image-to-image translation with conditional adversarial networks Preprint arXiv:1611.07004v3
[6] Das P, Baslamisi A, Liu Y, Karaoglu S and Gevers T 2018 Color constancy by GANs: an experimental survey Preprint arXiv:1812.03085v1
[7] Carles G, Muyo G, Bustin N, Wood A and Harvey A 2015 J. Opt. Soc. of A. 32 411-419
[8] He K, Zhang X, Ren S and Sun J 2015 Int. Conf. On Computer Vision (Santiago) 1026-1034
[9] Ronneberger O, Fischer P and Brox T 2015 Int. Conf. On Medical Image Computing & Computer Assisted Intervention (Munich)
[10] Kim J, Lee J and Lee K 2016 Conf. On Computer Vision and Pattern Recognition 1646-1654
[11] Kim J, Lee J and Lee K 2016 Int. Conf. On Learning Representations (San Diego)

Acknowledgments
Methods and algorithms were designed and developed with the support of RFBR grants (projects 19-29-01235-мк, 16-29-11744-офн-м, № 16-29-09528-офн-м, № 17-29-03112-офн-м, № 18-07-01390-А, № 18-37-00457-мол а), optics and experimental studies - in the framework of the state assignment of the IPSI RAS - a branch of the Federal Scientific-Research Center "Crystallography and Photonics" of the RAS (agreement № 007-Г3/Ч3363/26).