Calculation of a binary diffractive optical element to increase the imaging system depth of field in the task of classifying images by a neural network

A P Dzyuba¹, S N Khonina¹,², A V Nikonov¹,², P G Serafimovich²

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
²IPSI RAS – Branch of the FSRC “Crystallography and Photonics” RAS, Molodogvardeyskaya st. 151, Samara, Russia, 443001
e-mail: pavel.serafimovich@gmail.com

Abstract. Using an example of a real-world data set, it is shown that the accuracy of the image classifier based on a convolutional neural network does not deteriorate when using only one color channel. The binary diffractive optical element was calculated, which allows to increase the imaging system depth of field by several times. This is achieved by using the different color channels for various defocus values. A comparison of the MTF curves of the original and apodized imaging systems for a given minimum acceptable value of image contrast is presented.

1. Introduction

Image classification refers to one of the tasks in the implementation of computer vision. For example, an image classification algorithm may be developed to determine if the image of the number on the container in the warehouse corresponds to the number of the container that the robot should move. Although such a task is trivial for humans, reliable classification of images is still a problem in computer vision applications. The task is even more complicated if the analyzed image is distorted, in particular defocused.

Optical systems are known to be sensitive to defocusing and chromatic aberration. An increase in the depth of field of the optical system makes it possible to weaken this sensitivity and its negative consequences in blurring defocused images. However, a simple increase in the depth of field (DOF) by reducing the pupil or numerical aperture of the system leads to deterioration in resolution. One way to increase DOF without degrading resolution is to “encode” the wavefront, which is actually a phase apodization of the pupil of the lens [1-3]. As a rule, apodization is accompanied not only by positive effects (an increase in DOF and a decrease in the size of the focal spot), but also by a significant change in the structure of the point scattering function (PSF) and the growth of side lobes that worsen the image properties. Apodization is widely used in focusing [4, 5] and scanning [6-9] optical systems, as well as in microscopes to increase contrast [10-12], as well as in various applications for resolving two nearby radiation sources [13-14]. In imaging systems, the use of apodization requires additional, as a rule, digital decoding operation [15-21]. Despite the development of various decoding methods,
including neural networks [22-25], when optimizing the apodizing function, it is desirable to maintain a compromise between increasing DOF and PSF distortion.

In the task of classifying images, specific requirements are imposed on the imaging system. In particular, in this paper it is shown that for a relatively reliable classification of the image, it is sufficient to have a focused image in only one color channel. This observation is based on the study of only one data set. However, it logically follows from the very specifics of the task, in which the structure of the object is important, not its color.

2. Radially symmetric binary phase apodization
The amplitude $S$ in the focal plane of the imaging system can be written as follows:

$$ S(r, z) = 2 \int_0^1 P(\rho) J_0(rp) \exp(izp^2/2) \rho dp, \quad (1) $$

where $P(\rho)$ is the pupil function, $r$ and $z$ are the radial and axial coordinates:

$$ r = \frac{2\pi nsina}{\lambda}, \quad z = \frac{\pi 2sin^2(\alpha/2)}{\lambda}. $$

Here $\sin \alpha$ is a numerical aperture, and $R$ and $Z$ denote the normalized radial and axial coordinates. Along the optical axis, the amplitude of the light field is written as follows:

$$ S(0, z) = 2 \int_0^1 P(\rho) \exp(izp^2/2) \rho dp = \int_0^1 Q(t) \exp(izt/2) dtpdp. \quad (2) $$

We assume that $P(\rho)$ is a real function. This assumption is justified, because we consider binary phase elements, where the phase takes the values 0 and $\pi$, therefore the amplitude is 1 or -1. In this case, you can write the following ratio:

$$ S^*(0, z) = \int_0^1 [Q(t)]^* \exp(-izt/2) dt = \int_0^1 Q(t) \exp(-izt/2) dt = S(0, -z). \quad (3) $$

3. Results of simulation
To classify the images we use a VGG-like [26] convolutional neural network (CNN). The CNN architecture contains seven convolutional layers and two fully connected layers (Figure 1). Figure 1a shows initial layers of CNN and Figure 1b demonstrates end layers. To train the network, we use the Adam solver with a fixed learning rate of 0.0001. The loss function algorithm is categorical Cross-entropy. CNN parameters number is about 3.5 M. Epochs number is 15 and batch size is 64. We use framework Keras with Tensorflow2.0 backend.

Figure 1. Architecture of convolutional neural network
As training data for CNN, the Street View House Numbers (SVHN) [27] dataset is selected. SVHN is a real-world image dataset for developing machine learning and object recognition algorithms with minimal requirement on data preprocessing and formatting. This data set contains about 70000 digits for training, 25000 digits for testing. CNN was consistently trained in three conditions: (1) on color images (3 color channels), (2) on images that contained two color channels, and (3) on images that consisted of one color channel. Unused color channels were set equal to zero. For all cases, the classification accuracy on the test data was about 93% (table 1). Thus, it was concluded that one color channel is sufficient to classify the images of the SVHN dataset. Given the variety of lighting conditions, color palettes, and image position angles in SVHN, this conclusion can probably be generalized to other similar data sets, where the structure of the object is important for classification, not its color.

Table 1. Classification accuracy on the test data.

| Color channels | RGB | GB  | RB  | RG  | R   | B   | G   |
|----------------|-----|-----|-----|-----|-----|-----|-----|
| Accuracy, %    | 93.5| 92.8| 93.8| 93.4| 92.7| 92.4| 94.0|

For the last CNN layer, the softmax activation function was used, which calculates the probability of correct image classification. This, firstly, made it possible to demonstrate a decrease in CNN “confidence” in the correct classification for defocused images. Secondly, the calculated probabilities for each color channel allow you to choose the most reliable channel for classification. Figure 2 shows 9 test images and bar graphs of classification probabilities. In the upper right corner is a defocused image of “3”. The probability bar graph to the right of the image reflects the uncertainty of CNN.

![Figure 2](image-url)

**Figure 2.** Examples of test images and bar graphs of classification probabilities

Binary diffractive optical elements introduce significant chromatic aberration. As a rule, the compensation of these aberrations requires additional efforts in optimizing the optical element. In this work, chromatic aberration was used to optimize PSF separately for each color channel in order to increase DOF. For this, the phase apodization of the aperture using a radially symmetric (ring) binary phase element is used. When optimizing this element, the value of the maximum resolvable frequency is set (in fractions of the cutoff frequency, 0.5 was set) and the value of the minimum allowable contrast at this frequency (20% was set). The optimized parameters are the height of the phase ring and its inner radius (the normalized outer radius is assumed to be 1).
Figure 3. MTF plots for the source and apodized imaging systems at distances of (a) 9.0 m, (b) 1.2 m, (c) 0.65 m, (d) 0.44 m

Figure 3 shows MTF plots for the source and apodized imaging systems. These graphs are calculated for various defocus values. We assume that the imaging system (aperture 1.8 mm) is focused at a distance of 1.2 m from the lens. Figures 3a, b, c, d show MTF at distances of 9.0, 1.2, 0.65, 0.44 m, respectively. Comparison of MTF plots for the initial and apodized imaging systems allows us to conclude that in the apodized system the focus depth is increased by about an order of magnitude.

4. Conclusions
The binary optical element for phase apodization of the pupil function of the optical system is calculated. This optical element provides an increase in the depth of focus of the optical system through at least one color channel for a certain value on the optical axis. The calculated optical element can be used in machine vision problems for image classification. For example, when using robots in storage facilities. Using the appropriate training data set, the obtained results can be generalized, for example, to the task of monitoring the technical condition of railway rolling stock or the task of classification examination of ships and offshore structures.

Acknowledgments
The work was funded by RFBR, projects number 19-29-01235 mk, 18-07-01390 a.

References
[1] Bagheri S and Javidi B 2008 Opt. Lett.33 757
[2] Reddy A N K and Khonina S N 2018 Appl. Phys. B124 229
[3] Elmalem S, Giryes R and Marom E 2018 Opt. Express26 15316
[4] Sun C C and Liu C K 2003 Opt. Lett.28, 99
[5] Liu Z, Flores A, Wang M R and Yang J J2007 Opt. Eng.46018002 (1-9)
[6] Reddick R C, Warmark R J and Ferrel T L 1989 Phys. Rev. B. 39 767
[7] Kowalczyk M, Zapata-Rodriguez C J and Martinez-Corral M 1998 Appl. Opt. 37(35) 8206
[8] Hecht B, Sick B, Wild U P, Deckert V, Zenobi R, Martin O J F and Pohl D W 2000 J. Chem. Phys. 112 7761
[9] Boruah B R and Neil M A A 2009 Rev. Sci. Instrum. 80 013705
[10] Fürhapter S, Jesacher A, Bernet S and Ritsch-Marte M 2005 Opt. Express 13 689
[11] Situ G, Warber M, Pedrini G and Osten W 2010 Opt. Commun. 283 1273
[12] Maurer C, Jesacher A, Bernet S and Ritsch-Marte M 2011 Laser Photon. Rev. 5 81
[13] Grimes D N and Thompson B J 1967 Journal of the Optical Society of America 57(11) 1330
[14] Xu B, Wang Z and He J 2018 Scientific Reports 8 15216
[15] Raveh I, Mendlovic D, Zalevsky Z and Lohmann A W 1999 Opt. Eng. 38(10) 1620
[16] Cathey W T and Dowski E R 2002 Appl. Opt. 41(29) 6080
[17] Xu L, Ren J S, Liu C and Jia J 2014 Advances in Neural Information Processing Systems 1790
[18] Asif M S, Ayremlou A, Sankaranarayanan A, Veeraraghavan A and Baraniuk R G 2017 IEEE Trans. Computational Imaging 3 1790
[19] Eilertsen G, Kronander J, Denes G, Mantuik R and Unger J 2017 ACM Trans. Graph. (SIGGRAPH Asia) 36 6
[20] Sitzmann V, Diamond S, Peng Y, Dun X, Boyd S, Heidrich W, Heide F and Wetzstein G 2018 ACM Trans. Graph. 37(4) 114
[21] Dzyuba A, Serafimovich P, Khonina S and Popov S 2019 Proc. SPIE 11516, Optical Technologies for Telecommunications 115161A
[22] Gharbi M, Chaurasia G, Paris S and Durand F 2016 ACM Trans. Graph. 35 191
[23] Rivenson Y, Göröcs Z, Günaydın H, Zhang Y, Wang H and Ozcan A 2017 Optica 4(11) 1437
[24] Zhao H, Gallo O, Frosio I and Kautz J 2017 IEEE Transactions on Comput. Imaging 3 47
[25] Hershko E, Weiss L E, Michaeli T and Shechtman Y 2019 Opt. Express 27(5) 6158
[26] Vedaldi A, Zisserman A 2016 University of Oxford 66
[27] Netzer Y, Wang T, Coates A, Bissacco A, Wu B, Ng A 2011 NIPS Workshop on Deep Learning and Unsupervised Feature Learning