Transparent object sensing with enhanced prior from deep convolutional neural network

Jing Wang[a], Jian Bai*[a], Xiao Huang[a], Xiangdong Zhou[a], Lei Zhao[a], Kun Yan[a], Jing Hou[b] and Kailun Yang[a]

[a] State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, 38#, Zheda Road, Hangzhou, 310027, China; [b] Research Center of Laser Fusion, Mianyang, 621900, China

ABSTRACT

In recent years, with the development of new materials, transparent objects are playing an increasingly important role in many fields, from industrial manufacturing to military technology. However, transparent objects sensing still remains a challenging problem in the area of computational imaging and optical engineering. As an indispensable part of 3-D modeling, transparent object sensing is a long-standing research topic, which aims to reconstruct the surface shape of a given transparent object using various kinds of measurement methods. In this paper, we put forward a new method for the sensing of such objects. Specifically, we focus on the sensing of thin transparent objects, including thin films and various kinds of nano-materials. The proposed method consists of two main steps. Firstly, we use a deep convolutional neural network to predict the original distribution of the objects from its recorded intensity pattern. Secondly, the predicted results are used as initial estimates, and the iterative projection phase retrieval algorithm is performed with the enhanced priors to obtain finer reconstruction results. The numerical experiment results turned out that, with the two steps, our method is able to reconstruct the surface shape of a given thin transparent object with a high speed and simple experimental setup. Moreover, the proposed method shows a new path of transparent object sensing with the combination of state-of-art deep learning technique and conventional computational imaging algorithm. It indicates that, following the same framework, the performance of such method can be significantly improved with more advanced hardware and software implementation.

Keywords: Transparent object sensing, 3-D reconstruction, phase retrieval, deep learning, convolutional neural network, computational imaging

1. INTRODUCTION

With the development of new materials and manufacturing technology, transparent objects such as nano-materials, micro-optics and thin films are widely used in different research fields [1-3]. However, these objects are usually very thin such that it is hard to sense their shape and distribution using traditional methods. Besides, compared with opaque objects, their nature of reflection, refraction, absorption and scattering of incident light makes them more difficult to be accurately reconstructed. For example, the 3-D distribution of opaque objects can be more easily attained through the measurement of only reflected light [4]. On the other hand, when it comes to transparent objects, if only reflected light is considered, a lot of surface information will be missed in a practical situation [5].

Taking account of their unique characteristics, the sensing of such thin transparent objects is still very challenging. The existing methods for normal 3-D transparent object sensing can be categorized into two groups: intrusive methods and non-intrusive methods. Intrusive methods usually detect the object by covering it with a coating or immersing it into refractive liquids [6,7]. However, these methods may be harmful for the detected objects and are incapable of handling difficult situations. In recent years, image-based non-intrusive methods begin to show their potential to solve such problem as computer vision and optical engineering are undergoing a monumental boom. These techniques, on the other hand, intend to reconstruct the transparent object from the light which propagates through it, usually by inverse calculation of the captured images with the help of specially designed setups and computational imaging algorithms. Many preliminary studies have been carried out in this domain and most of them are able to solve this problem to some extent. We categorize them as follows.

One of the most popular techniques that are widely used is phase-shifting digital holography. This technique uses phase-shifted reference waves and retrieves complex shape of transparent objects from the recorded holograms [8-10]. It requires precise adjustment of the experimental setups to produce reference waves with shifted phases and is therefore not very convenient to deploy in some adverse environments. In comparison with holography, deflectometry is a technique which is more simple to implement, but it can only achieve limited quality if the computational cost is reduced to an acceptable level. It realizes transparent object sensing using a LCD panel to display the coded patterns and measures the distortion of the patterns with spatial gradient calculation or vector analysis [11-13]. Another commonly applied technique is Time-Of-Flight (TOF) distortion. This method is designed to recover a transparent object from a single view using a TOF image sensor [14-16], but it is actually limited to low resolution which is bounded by the TOF sensor. Other attempts to solve this problem include tomographic imaging [17,18], phase contrast light microscopy [19], inverse raytracing with polarization imaging [20] and scatter trace photography [21].

When it comes to thin transparent objects, because they can be regarded as thin phase objects, the technique of phase recovering and wave-front sensing can be also employed. Previous studies include detecting the distribution of transparent objects using traditional phase retrieval algorithm [22] and Shack-Hartmann wave-front sensor [23]. These methods merely use conventional algorithms so the speed and accuracy of reconstructed results are not that satisfactory.

To improve the present performance on thin transparent object sensing, in this paper, we put forward a method that combines the state-of-art deep learning technique with the conventional computational imaging algorithm. By using a convolutional neural network to provide an optimized initial estimate to the iterative projection algorithm, we successfully reconstruct the shape of the given thin transparent object through its recorded intensity pattern. Numerical experiment results turned out that, with the enhanced prior from neural networks, the traditional phase retrieval algorithm can achieve the sensing of transparent objects with a relatively higher speed and accuracy.

2. METHOD

2.1 Problem formulation

As mentioned in the previous section, objects that are transparent to light can be actually regarded as phase objects. When placed in an optical imaging system, they will introduce extra phase to the light propagating through them. Therefore, the surface shape of such objects can be then calculated from the measured phase distribution. In this paper, we simplify the problem by assuming that the refractive index of the material is already known, so the profile of the object can be calculated from its surface height. We describe the surface shape of the sample object as:

$$z = s(x, y) \tag{1}$$

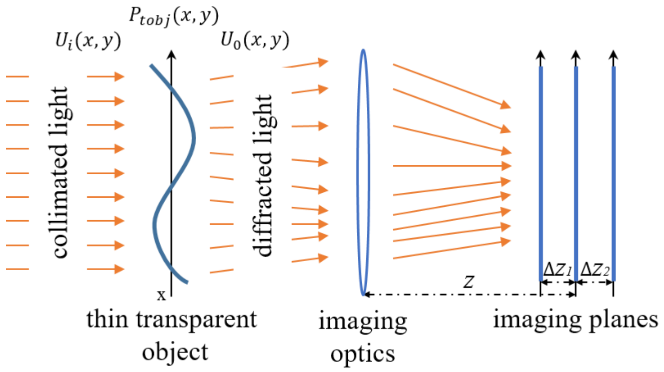

Figure 1. Ray schematic. The collimated light going through the thin transparent object is diffracted. Then the imaging optics are applied to propagate it to different imaging planes.

As shown in Figure 1, supposing that $U_i(x, y)$ and $U_o(x, y)$ are the complex amplitudes of the light before and after propagating through the sample, $U_o(x, y)$ can be formulated as

$$U_o(x, y) = U_i(x, y) P_{tobj}(x, y) \tag{2}$$

where $P_{tobj}(x, y)$ denotes the transmission coefficient and it can be calculated from the phase introduced by the sample:

$$P_{tobj}(x, y) = e^{i\varphi(x,y)} \tag{3}$$

$$\varphi(x,y) = ks(x,y)(n-1) \tag{4}$$

The definition of $s(x,y)$ is mentioned above, where k is the wave vector of incident light and n is the refractive index of the material. A lens finally images the diffracted light to imaging planes where a movable CMOS camera is placed and can be moved along the optic axis. Then the problem becomes: given the recorded patterns at imaging planes, how can we retrieve the phase $\varphi(x,y)$ of the objects? In this way, we can finally obtain their surface shapes from the phase. Traditional phase retrieval algorithm may be a solution, but its performance is very limited without sufficient knowledge of the original distribution. Therefore, in this paper, we firstly use a deep convolutional neural network to predict an estimate of $\varphi(x,y)$, and then use the iterative projection algorithm to attain a finer reconstructed result. Figure 2 shows a schematic diagram of the proposed transparent object sensing system. In the following sections, we will demonstrate how these two steps are actually achieved to solve this problem.

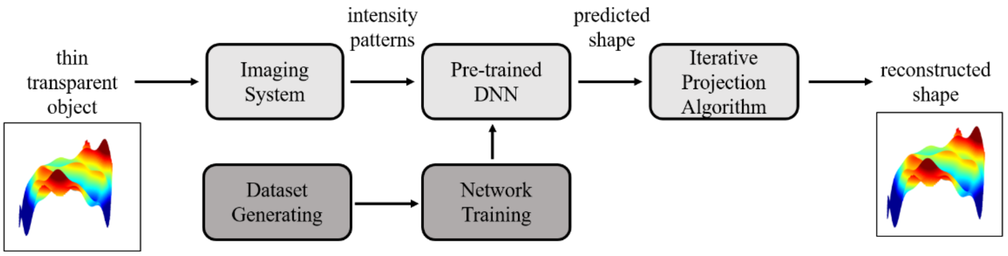

Figure 2. A schematic diagram of the proposed transparent object sensing system.

2.2 Neural network training

As is shown in Figure 2, in our transparent object sensing system, the CMOS camera will capture the intensity patterns at the image planes. Then one of these intensity images, which is captured near the focus of the imaging system, will be pre-processed and fed into the deep convolutional neural network. The neural network will use this intensity pattern to approximately predict an estimate of $\varphi(x,y)$. In this part, we chose Inception-Resnet-v2[24] as a basic network to fulfill feature extraction, considering its outstanding performance in the ImageNet classification competition [25]. We adapt it to perform regression tasks by removing the top layers, and adding a Global Average Pooling layer and a Dropout layer. The architecture of the network is shown in Figure 3.

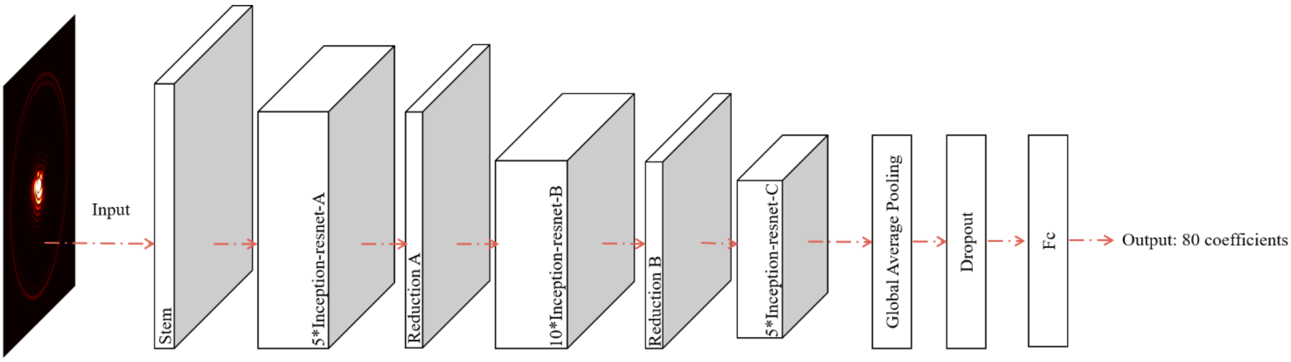

Figure 3. The adapted architecture of our neural network based on Inception-Resnet-V2 to perform regression tasks.

The network takes the recorded intensity image captured by the CMOS camera as input. The input intensity image is augmented to enhance some dim features that cannot be seen clearly from the original picture. The output of the network are 80 Zernike coefficients which are used to approximately describe the surface distribution $s(x,y)$. In this way, $s(x,y)$ is transformed into polar coordinate $s(r,\theta)$ and represented with a Zernike series z_i with coefficients a_i [26]:

$$s(r,\theta) = \sum_{i=1}^{80} a_i z_i(r,\theta) \tag{5}$$

where r is the radial distance and θ is the azimuthal angle on the pupil plane.

This network is trained on a synthetic dataset that we generated a large number of transparent object surfaces and corresponding intensity patterns through a two-step angular spectrum propagating method. We use Adam optimization method [27] during the network training with an initial learning rate of 0.0001 and a batch size of 8. The network is trained over 60 epochs. The dropout ratio is set to 0.2 in the last dropout layer in order to avoid over-fitting. Moreover, the 80 coefficients are randomly distributed within the range of [-0.5,0.5] to generate arbitrary surface shapes. There are 150000 and 6000 surface-intensity pairs in the training and validation dataset respectively and another 100 pairs are generated for the purpose of testing. Some examples of the generated surface shape and corresponding intensity patterns are shown in Figure 4 row (a) and row (b) respectively. It takes some time to train the network from scratch, but after the training process is accomplished, in a real transparent object sensing system, the network takes only a few milliseconds to perform inference on an ordinary desktop computer with a modern GPU.

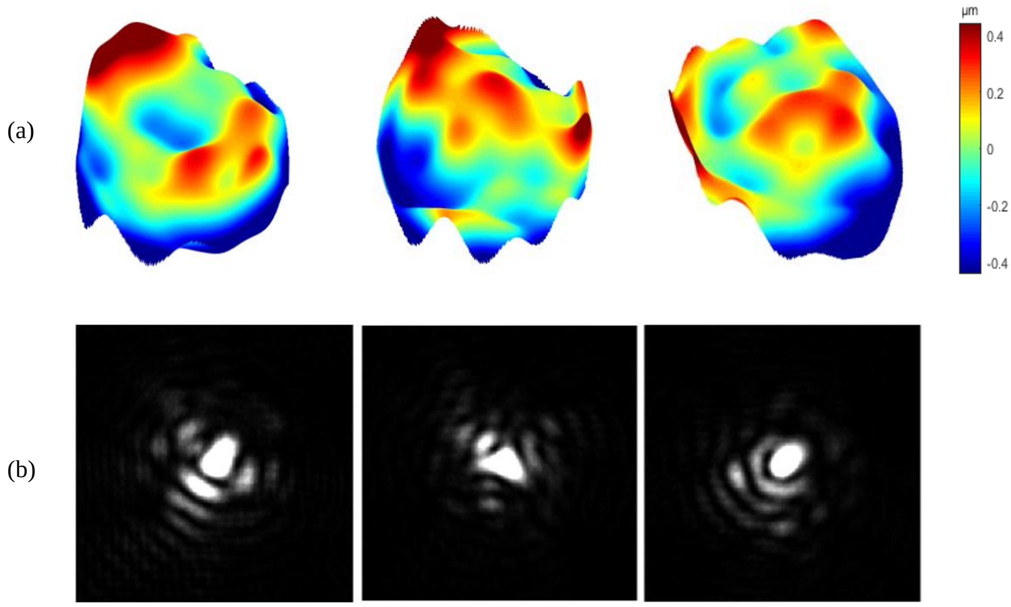

Figure 4. Examples of (a) generated arbitrary surfaces of transparent objects and (b) corresponding intensity patterns in the network training dataset.

2.3 Iterative projection phase retrieval algorithm

Once the recorded intensity pattern is taken as the input of the neural network, it will give out the results of 80 Zernike coefficients that approximately predict the distribution of original surface shape. After that, we use these coefficients to generate a corresponding phase map and it can be used as an optimized initial estimate of the traditional phase retrieval algorithm. A Non-linear Optimization Phase Retrieval (NOPR) algorithm is employed in the proposed method.

This algorithm uses three intensity measurements near the focus to simultaneously reconstruct the phase and amplitude of the input complex light field. As illustrated in Figure 1, this algorithm firstly starts from an initial guess of light filed in the object plane and propagates it to the imaging plane using two-step angular spectrum propagating method at the distance z. This plane is then referred to as a master plane and the other two are chosen as slave planes. We use the measured diffraction pattern of the master plane to reconstruct the phase on this plane by iteratively propagating it to slave planes and back. Each plane is employed as the master plane successively while each time only a small number of iterations are performed. Then the light is propagated back to the object plane to start another iteration. Root Mean-Squared (RMS) error is used to evaluate the difference between the original and reconstructed shape in the simulation. The algorithm will go through several iterations to acquire a reconstructed phase distribution on the object plane until the residual RMS error meets the threshold. In the field of phase retrieval, capture range, which implies how the initial estimate is close to the ground truth, is a commonly known problem. Normal applications of this technique requires a good starting estimate of original surface shape. Therefore, the convolutional neural network described in section 2.2 is used to provide enhanced priors. In this way, we can achieve phase-retrieval-based transparent object sensing.

3. NUMERICAL EXPERIMENTS

In this section, numerical experiments were carried out to evaluate the performance of proposed transparent object sensing method. The codes were implemented in Keras and MATLAB, and executed on a computer with an Intel Core CPU running at 2.80GHz and a NVIDIA GeForce GTX1060 GPU. Firstly, before object sensing, we trained the established neural network using our synthetic dataset which includes pairs of arbitrary surfaces and corresponding intensity patterns. As mentioned above, the captured intensity patterns would firstly go through the image pre-processing step before it was input into the neural network. An example of the intensity pattern before and after pre-processing is shown in Figure 5. It is clearly that with the pre-processing step, the intensity pattern was more capable of augmenting the dim features, which was proved later by experiments that after pre-processing the loss function of the neural network converged to a relatively lower value based on the same training configuration.

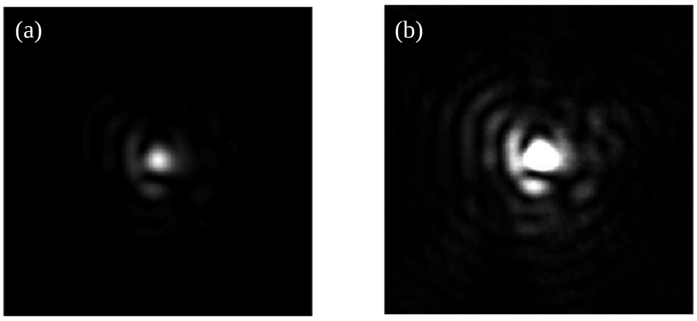

Figure 5. Intensity patterns (a) before and (b) after pre-processing.

Next, we used the proposed method to reconstruct the original shape of target thin transparent object. Experimental results from the two steps are also studied carefully. Numerical simulation results were given to evaluate the speed and accuracy. Firstly, the thin transparent object was placed in an imaging system to obtain its intensity patterns. The focal length of the imaging system was 500 mm. The distance z was therefore $z = 500\ mm$ and the displacement distances were $\Delta z_1 = -60\ mm$ and $\Delta z_2 = 40\ mm$ respectively. After this, the intensity pattern image at the focus was enhanced and normalized following the above instructions and then fed to the neural network. The neural network predicted 80 Zernike coefficients from the image and these coefficients were used to generate the predicted surface map. At the end of this, this surface map was used as an optimized initial estimate. Non-linear phase retrieval algorithm was performed starting from this initial solution, then it iteratively calculated the estimated light field in object and image plane from their recorded intensities by repeatedly propagating the light among these planes. In this step, because the calculated phase distribution was wrapped into $[0, 2\pi]$ due to the nature of NOPR algorithm, a robust two-dimensional phase unwrapping algorithm [28] was employed to address the distribution issue. With the reconstructed unwrapped phase distribution, the surface shapes can be then calculated from equation (4).

Figure 6 shows the ground truth surface shapes of original thin transparent objects, the predicted surface shapes from the deep convolutional neural network and the reconstructed surface shapes from NOPR algorithm respectively. In all of these three experiments, the sample points at object and image plane were both set to 299×299 in order to fit the input size of neural network, such that the pixel size was 33.44 μm. Besides, the original intensity images were normalized to [0, 1] before processed by the deep convolutional neural network. It can be seen that the proposed method can achieve the sensing of thin transparent objects with very sophisticated surface distributions and the reconstruction is dense. The predicted results are quite close to the original ones, while the reconstructed results are even closer. The processing speed of our method is about 1s for every single image which is fast enough for general usage. The object sensing speed can be accelerated with more advanced hardware devices and extended dataset with the use of a spatial light modulator. For objects with a larger size, they can be reconstructed by gradually moving the sample in the plane perpendicular to the incident light using the same experimental instruments.

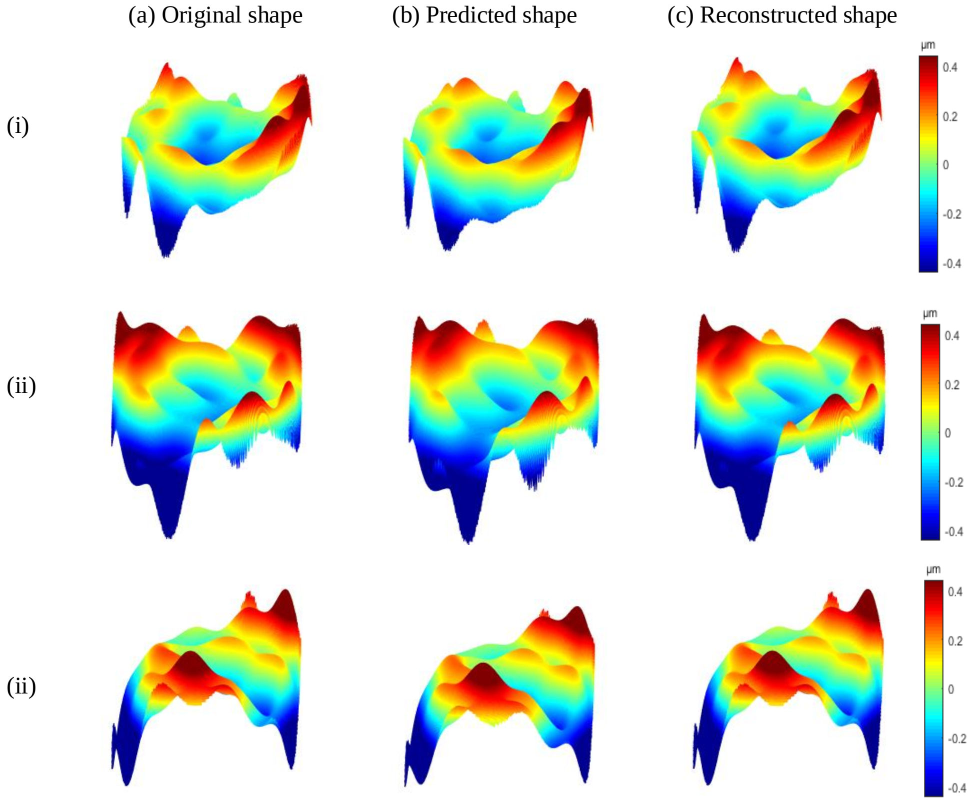

Figure 6. (a) Ground truth surface distribution; (b) Predicted surface shape from the deep convolutional neural network; (c) Reconstructed surface shape from NOPR.

Figure 7 presents the quantitative results of the 100 samples in the test dataset. It shows both the RMS errors of original surface shapes, and the residual RMS errors between original surfaces and reconstructed results. In most of the cases tested, the residual errors of our proposed method have been reduced to a relatively low value, which is calculated to be less than 1/7 of the original RMS errors with respect to the surfaces. This proves the effectiveness of our algorithm. But it is worth noticing that there are still a few failure cases which indicates the reconstructions in these samples are actually crashed down due to some reasons. We checked these data and found out that in these cases, the predicted shapes in step 1 are far from satisfactory, which means, the initial estimates provided to the iterative phase retrieval algorithm are not that good. In this regard, it could be improved by further training the neural network with a more sufficient dataset so that the accuracy of predicted results can be sufficient for the subsequent algorithm to reconstruct a better result.

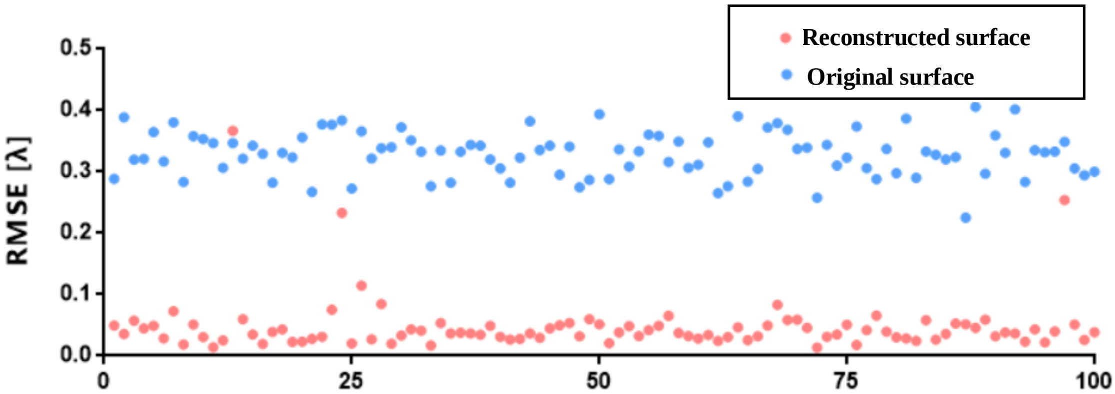

Figure 7. Original and residual RMS error of the 100 samples in the test dataset.

4. CONCLUSION

In this paper, we proposed a method for the sensing of thin transparent objects based on conventional computational imaging algorithm with enhanced priors from deep learning. In order to reach this goal, we firstly trained a deep convolutional neural network using a synthetic dataset generating from two-step angular spectrum method. Then, this network is used to predict the surface distribution of the target object from its captured intensity pattern. A Zernike polynomials description of the shape of transparent objects is presented. Finally, the predicted surface serves as an enhanced prior for the non-linear phase retrieval algorithm and the reconstructed surface is calculated from this initial estimate.

The numerical experiments turned out that the proposed method is able to achieve the sensing of thin transparent objects with complicated and arbitrary surface distributions. The computational cost mainly lies in the training of neural network and iterative projection phase retrieval algorithm but the actual processing time of the whole algorithm can be limited to within 1 second for each single image on a modern GPU processor. The accuracy of our sensing system is also studied and the residual RMS errors of reconstructed results are calculated to be less than 1/7 of original RMS errors with respect to the surfaces in most tested cases. Though it is obvious that Zernike polynomials are not enough to describe the surface shape of all various kinds of thin transparent objects, still, the results in this paper give us the critical insight that the sensing of such objects can be realized with the combination of deep learning and conventional phase retrieval methods. To reconstruct more complicated surface shapes, an image-to-image neural network like U-net could be considered. With regard to the phase retrieval, it can also process phase maps that are in other forms instead of the Zernike polynomials representation.

In the future work, more experiments will be carried out with the employment of a spatial light modulator and some noises will be added to the training dataset to enhance the robustness of this algorithm. Furthermore, other hardware and software implementations will be considered to improve the performance of our proposed method following the same framework.

ACKNOWLEDGEMENTS

This work is supported by Science Challenge Project (No. TZ2016006-0502-02).

REFERENCES

[1] Kong, Y. C., YU, D. P. and Zhang, B. et al., “Ultraviolet-emitting ZnO nanowires synthesized by a physical vapor deposition approach,” Appl. Phys. Lett. 78(4), 407–409(2001).

[2] Hata., K., “Water-assisted highly efficient synthesis of impurity-free single-walled carbon nanotubes,” Science 306(5700), 1362–1364(2004).

[3] Yu, Y., Li, C., Liu, Y. et al., “Controlled scalable synthesis of uniform, high-quality monolayer and few-layer MoS2 films,” Sci. Rep. 3, 1866 (2013).

[4] Zhang, H., Han, S., Liu, S. et al., “3D shape reconstruction of large specular surface,” Applied Optics, 51(31), 7616-7625(2012).

[5] Han, K., Wong, K. Y. K. and Liu, M., "Dense reconstruction of transparent objects by altering incident light paths through refraction." International Journal of Computer Vision, 126(5), 460-475 (2017).

[6] Trifonov, B., Bradley, D., Heidrich, W., “Tomographic reconstruction of transparent objects,” Eurographics Conference on Rendering Techniques. Eurographics Association, 55 (2006).

[7] Hullin, M. B., Fuchs, M., Ihrke, I., et al., “Fluorescent immersion range scanning,” ACM. Trans. on Graphics (SIGGRAPH’08), 27(3), 87:1 – 87:10(2008).

[8] Sasada, M., Fujii, A., Awatsuji, Y. and T. Kubota, “Parallel quasi-phase-shifting digital holography that can achieve instantaneous measurement,” International Conference Optics and Photonics in Technology Frontier, 187-188(2004).

[9] Tahara, T., Shimozato, Y., Kakue, T. et al., “Comparative evaluation of the image-reconstruction algorithms of single-shot phase-shifting digital holography,” J. Electron. Imaging 21(1), 013021 (2012).

[10] Awatsuji, Y., Ito, K., Shimozato, Y. et al., "20000-frames-per-second phase-shifting digital holography," OSA Digital Holography and Three-Dimensional Imaging DMA6 (2010).

[11] Kim, J., Reshetouski, I., and Ghosh, A., "Acquiring axially-symmetric transparent objects using single-view transmission imaging," International Conference on Frontiers of Signal Processing. IEEE, 1484-1492 (2016).

[12] Fischer, M., Petz, M., Tutsch, R., "Model-based deflectometric measurement of transparent objects," International Workshop on Advanced Optical Imaging and Metrology, 573-576 (2014).

[13] Shin, S. and Yu, Y.. "Measuring the both surfaces profiles of optical element using transmission deflectometry with liquids," Lasers & Electro-optics Pacific Rim. IEEE, 27P_14 (2015)

[14] Tanaka, K., Mukaigawa, Y., Kubo, H., et al., "Recovering transparent shape from time-of-flight distortion," The IEEE Conference on Computer Vision and Pattern Recognition, 4387-4395(2016).

[15] Kadambi, A., Whyte, R., Bhandari, A., Streeter, L., et al., "Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles," ACM Transactions on Graphics, 32(6), 1–10(2013).

[16] Shim, H. and Lee, S., "Recovering translucent object using a single time-of-flight depth camera," IEEE Transactions on Circuits and Systems for Video Technology, 2(99), 1–14 (2015).

[17] Vertu, S., Yamada, I., Delaunay, J. et al., "Tomographic observation of transparent objects under coherent illumination and reconstruction by filtered back-projection and Fourier diffraction theorem," Proc. SPIE 6861, 715-732(2008).

[18] Trifonov, B., Bradley, D. and Heidrich, W., "Tomographic reconstruction of transparent objects," Proc. Eurographics Symposium on Rendering, 51–60(2006).

[19] Knight, P., and Parsons, N., "Measurement of the width of thin, cylindrical, transparent objects by phase contrast light microscopy," Journal of Microscopy, 137(1), 10(1985).

[20] Miyazaki, D. and Ikeuchi. K., "Shape estimation of transparent objects by using inverse polarization raytracing," IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(11), 2018–2030 (2007).

[21] Morris, N. J. and Kutulakos, K. N., "Reconstructing the surface of inhomogeneous transparent scenes by scatter trace photography," IEEE International Conference on Computer Vision, 14-20(2007).

[22] Forest, C. and Claude, R., "Metrology of thin transparent optics using Shack-Hartmann wave-front sensing," Optical Engineering, 43(3), 742-760 (2004).

[23] Wu, L., Tao, S. and Xiao, S., "Phase retrieval–based distribution detecting method for transparent objects," Optical Engineering, 54(11), 113103(2015).

[24] Szegedy, C., Ioffe, S. and Vanhoucke, V., "Inception-v4, inception-resnet and the impact of residual connections on learning," AAAI Conference on Artificial Intelligence, 1602.07261 (2016).

[25] Russakovsky, O., Deng, J. and Su, H. et al., "ImageNet large scale visual recognition challenge," International Journal of Computer Vision, 115(3), 211-252 (2015).

[26] Schmidt, J. D., [Numerical Simulation of Optical Wave Propagation with Examples in MATLAB], Bellingham: Society of Photo-Optical Instrumentation Engineers, Washington, 168-182(2010).

[27] Kingma, D., Ba, J., "Adam: a method for stochastic optimization," IEEE International Conference on Consumer Electronics, 434-435 (2014).

[28] Gdeisat, M. and Lilley, F. "Two-dimensional phase unwrapping problem" Slate, June 2019, <https://www.researchgate.net/publication/333866674_Two-Dimensional_Phase_Unwrapping_Problem> (June 2019)