DENSE 3D OBJECT RECONSTRUCTION USING STRUCTURED-LIGHT SCANNER AND DEEP LEARNING

V. V. Kniaz^{1,2}, V. A. Mizginov¹*, L.V. Grodzitkiy¹, N.A. Fomin¹, V. A. Knyaz^{1,2}

¹ State Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia – (vl.kniaz, vl.mizginov, lev.grodzitkiy, nfomin73, knyaz)@gosniias.ru
² Moscow Institute of Physics and Technology (MIPT), Dolgoprudny, Russia

Technical Commission II

KEY WORDS: 3D object reconstruction, structured light, optical flow, 3D scanner, deep neural networks

ABSTRACT:

Structured light scanners are intensively exploited in various applications such as non-destructive quality control at an assembly line, optical metrology, and cultural heritage documentation. While more than 20 companies develop commercially available structured light scanners, structured light technology accuracy has limitations for fast systems. Model surface discrepancies often present if the texture of the object has severe changes in brightness or reflective properties of its texture. The primary source of such discrepancies is errors in the stereo matching caused by complex surface texture. These errors result in ridge-like structures on the surface of the reconstructed 3D model. This paper is focused on the development of a deep neural network LineMatchGAN for error reduction in 3D models produced by a structured light scanner. We use the pix2pix model as a starting point for our research. The aim of our LineMatchGAN is a refinement of the rough optical flow A and generation of an error-free optical flow \hat{B} . We collected a dataset (which we term ZebraScan) consisting of 500 samples to train our LineMatchGAN model. Each sample includes image sequences (S_l, S_r) , ground-truth optical flow B and a ground-truth 3D model. We evaluate our LineMatchGAN on a test split of our ZebraScan dataset that includes 50 samples. The evaluation proves that our LineMatchGAN improves the stereo matching accuracy (optical flow end point error, EPE) from 0.05 pixels to 0.01 pixels.

1. INTRODUCTION

Close-range photogrammetric techniques proved to be accurate and reliable 3D non-contact measurement in many applications beginning with industrial ones and spanning to anthropology and cultural heritage (Bosemann, 2011, Remondino, 2011, Knyaz and Maksimov, 2014). Active photogrammetric systems based on structured light demonstrate high accuracy and high performance for obtaining multiple 3D coordinates of an object's surface. Structured light scanners are intensively exploited in various applications such as non-destructive quality control at an assembly line, optical metrology, and cultural heritage documentation. While more than 20 companies develop commercially available structured light scanners, structured light technology accuracy has limitations for fast systems. Model surface discrepancies often present if the texture of the object has severe changes in brightness or reflective properties of its texture. The primary source of such discrepancies is errors in the stereo matching caused by complex surface texture. These errors result in ridge-like structures on the surface of the reconstructed 3D model. Many methods were proposed to compensate error in stereo matching for structured light systems (Curless and Levoy, 1995, Wang et al., 2016, Taylor, 2012, O'Toole et al., 2014, Wang and Feng, 2014, Chen and Shen, 2018, Bian and Liu, 2016, Knyaz, 2010). While these methods reduce surface distance error between reconstructed and the ground truth models, they could not eliminate the discrepancies caused by uneven texture brightness. The problems of stereo matching can be considered as a special case of optical flow estimation. For the structured light systems, the matches between the correspondent points result in dense optical flow from the left to the right camera.

*Corresponding author

Deep neural networks (Krizhevsky et al., 2012) have proved to be the most effective algorithms for robust optical flow estimation. Moreover, recently a new generation of neural networks has been proposed that is commonly named Generative Adversarial Networks (GANs) (Goodfellow et al., 2014). These networks could be trained for complex image-to-image translation tasks such as object transfiguration (Zhu et al., 2017), image superresolution (Zhu et al., 2017) and noise reduction (Chen et al., 2020). A GAN model consists of two networks: a generator Gand a discriminator D. Two networks are trained simultaneously for concurrent tasks. The aim of the discriminator D is to distinguish 'real' samples B from the training dataset from 'fake' samples \hat{B} produced by the generator G. The objective of the generator G is the synthesis of 'fake' samples \hat{B} that are indistinguishable from the random samples B from the training datasets.

This paper is focused on the development of a deep neural network LineMatchGAN for error reduction in 3D models produced by a structured light scanner. We use the pix2pix model as a starting point for our research. The aim of our LineMatchGAN is a refinement of the rough optical flow A and generation of an error-free optical flow \hat{B} . We hypothesize that our model can simultaneously process the object images S from the left and the right cameras and the rough optical flow A to fix errors in the preliminary stereo matching. Our proposed pipeline is given in Figure 1.

We collected a dataset (which we term ZebraScan) to train our LineMatchGAN model. The dataset consists of 500 samples. Each sample includes image sequences (Q_l, Q_r) , object images (S_l, S_r) , ground-truth optical flow B and a ground-truth 3D model. We evaluate our LineMatchGAN algorithms on a test split of our ZebraScan dataset that includes 50 samples. The evaluation proves



Figure 1. The pipeline of the developed method.

that our LineMatchGAN improves the stereo matching accuracy (optical flow end point error, EPE) from 0.05 pixels to 0.01 pixels. This improvement allows reducing the surface error from 0.05 mm to 0.015 mm for a working volume of $300 \times 300 \times 300$ mm³. Our pipeline can be used to improve the accuracy of existing hand-held laser scanners. The evaluation of our LineMatchGAN model using our DeepScan scanner demonstrates that the proposed approach is robust against complex object textures.

The inference time of our LineMatchGAN is 1 second for input with a resolution of 1024×768 pixels using NVIDIA Jetson TX2 GPU. Such computational efficiency allows using the proposed pipeline to improve the accuracy of existing hand-held scanners for on-line data processing. We made our dataset and model publicly available¹.

2. RELATED WORK

2.1 Structured Light Scanners

Today the use of structured light scanners is widespread in various tasks of photogrammetry and optical metrology. The works of (Akca et al., 2007) and (Chalmers et al., 2001) demonstrate the application of 3D scanners for documentation and visualization of cultural heritage. The first work presents the results of 3D modelling of two cultural heritage objects, where a closerange coded structured light system was used for generation of 3D models. In the second work a low-cost 3D scanner based on structured light projection with versatile coloured stripe pattern approach was designed. It adopts a set of patterns produced by recursive subdivision, which mixes thin stripes (from which the shape is reconstructed) and coloured bands that are used to reindex the stripes.

Structured light scanners are also used for a shape measurement and a 3D object surface reconstruction. Lin et al. have proposed

¹http://www.zefirus.org/en/linematchgan

an automatic 3D color shape measurement system (Lin, 2020) based on images recorded by a stereo camera was developed. Also 3D shape measurement techniques have been widely used in industrial inspection, intelligent manufacturing, reverse engineering, and many other aspects (Salvi et al., 2010, Pribanic et al., 2010, Gupta et al., 2011). Pribanic et al. have developed a multiple phase-shifting method (Pribanic et al., 2010). This method is not influenced by wrapped phase computation inaccuracies as the original approach and it is faster than common LUT-based (search) methods. Xiao et al. have proposed a structured light measurement technique (Xiao et al., 2020) based on the reverse photography. The technique includes an auxiliary reverse camera installed behind the structured light system. The camera allows to unify the local 3D shape data acquired from multi-view structured light measurements to a global frame. Such approach allows to achieve a holistic 3D shape data fusion.

Finally, using a structured light scanner in conjunction with deep learning methods allows to generate depth maps of real-world scenes. Li et al. have proposed a novel method (Li et al., 2019) that combines structured light and deep learning stereo matching to calculate the depth map. To prevent the holes in the textureless areas during the stereo matching, a depth map is predicted by a convolutional neural network. Then, a fine and accurate depth map is obtained by phase matching. The proposed method can generate a high precision depth and relieve the occlusion in the structured light system.

2.2 Mobile Scanners

Increasingly, the need arises for accurate and reliable reconstruction of three-dimensional objects in tasks where it is impossible to use expensive and bulky equipment. Recent progress and availability of the small and accurate industrial cameras have allowed the creation of low-cost and mobile systems for reconstruction of 3D objects. Piccirilli et al. developed a mobile sensor based on the fringe projection techniques (Piccirilli et al., 2016). The goal of the developed sensor is acquisition of the 3D model of a face and its color texture using a smartphone device. The data acquisition, pattern generation, and reconstruction of the final 3D point cloud are all driven only by the smartphone. This system is used for 3D face acquisition.

Donlic et al. have proposed a 3D reconstruction framework (Donlic et al., 2017) implemented on a tablet mobile device. The method leverages the mobile device's accelerometer and magnetometer for 3D registration. Kniaz et al. recently have developed a new method (Kniaz, 2019) for robust 3D model reconstruction. The methods combines a mobile structured light 3D scanner with a deep learning technique for online 3D reconstruction. The deep learning-based approach allows to perform stereo matching with a 3D scanner and achieve a sub-millimeter accuracy in the object space.

2.3 Generative Adversarial Networks

The development of a new type of neural networks known as generative adversarial networks (GANs) (Goodfellow et al., 2014) made it possible to take a significant step forward in the field of image processing. GANs consist of two deep convolutional neural networks: a Generator network tries to synthesize an image that visually indistinguishable from a given sample of images from the target domain. A Discriminator network tries to distinguish the 'fake' images generated by the Generator network from the real images in the target domain. Generator and Discriminator networks are trained simultaneously. This approach can be considered as an adversarial game of two players.

One of the first goals solved using GANs was image synthesis. Image-to-image translation problem was solved using conditional GAN termed pix2pix (Isola et al., 2017). Such network learns a mapping $G : (x, z) \rightarrow y$ from observed image x and random noise vector z, to output y. This method also uses a sum of two loss functions: a conditional adversarial objective function and an L1 distance. However, for many tasks it is not possible to generate paired training datasets for image-to-image translation tasks.

To overcome this difficulty a CycleGAN (Zhu et al., 2017) was proposed. The CycleGAN leverage a cycle consistency loss for learning a translation from a source domain X to a target domain Y in the absence of paired examples. Therefore, the CycleGAN model detects special features in one image domain and learns to translate them to the target domain. A new StyleGAN model was proposed in (Karras et al., 2018) that provides a superior performance in the perceptual realism and quality of the reconstructed image. Unlike the common generator architecture that feeds the latent code through the input layer, the StyleGAN appends a mapping of the input to an intermediate latent space, which controls the generator. Moreover, an adaptive instance normalization (AdaIN) is used at each convolution layer. Gaussian noise is injected after each convolution facilitating generation of stochastic features such as hair-dress or freckles. The problems of the first StyleGAN model were partially eliminated in the second StyleGANv2 model (Karras et al., 2019). In this model parameters are optimized and the neural network training pipeline was adjusted. The changes made have improved the quality of the results.

3. METHOD

The aim of our LineMatchGAN model is improving stereo-matching accuracy using a deep neural network. Unlike an existing ap-

proaches (Li et al., 2019), where error correction is performed in the domain of the depth map, we perform processing in the domain of the optical flow (Wedel and Cremers, 2011). Specifically, we consider stereomatches as a sparse optical flow that can be densified and improved by a GAN model.

The rest of this section presents an overview of our 3D reconstruction pipeline. After that technical properties of our mobile scanner are discussed. Finally, we present details of our Line MatchGAN model.

3.1 Framework Overview

The first step in our pipeline is acquiring of the structured light sequences S_l, S_r using a mobile scanner. After that, we use a hand-crafted stereo matching algorithm (Knyaz, 2010) $S: (S_l, S_r) \rightarrow A$ to generate rough optical flow A. We use the rough optical flow A and object images S_l, S_r as an input for our LineMatchGAN $G: (A, S_l, S_r) \rightarrow \hat{B}$. Finally, we feed the predicted refined optical flow \hat{B} to the hand-crafted 3D reconstruction algorithm to obtain a 3D mesh. We developed a mobile hand-held 3D scanner DeepScan to train and evaluate our LineMatchGAN model.

3.2 Mobile Scanner

We use a mobile scanner developed in our previous research (Kniaz, 2019, Kniaz et al., 2020). The developed scanner is based on the assumptions made by Knyaz et al. (Knyaz, 2010, Knyaz, 2015). The whole system consists of two high-speed industrial cameras located on an aluminum beam. The cameras a separated by a basis of 300 mm. The structured light illumination is provided by a mobile multimedia projector. The projector has an autonomous power supply capable for 40 minutes of operation.

We use an external synchronization clock to synchronize the cameras and the projector. The total weight of our system is 1.3 kg. It allows to use it in the filed for online scanning of archeological objects and cultural heritage. The complete system is presented in Figure 3. Technical specifications of the system are presented in Table 1.

Table 1. Technical specifications of our mobile structured light scanner.

Parameter	Specification	Unit
Camera Resolution	1328×1048	pixels
Camera Maximal FPS	120	pixels
Camera Lens Focal Length	8	mm
Camera Exposure	16.6	ms
Camera Baseline	300	mm
Camera Object Distance	~ 700	mm
Stereo Field of View (FOV)	$400\times 300\times 200$	mm
Camera CMOS	Sony IMX035	
Camera Lens Aperture	5.6	
Projector Resolution	1280×720	pixels

3.3 LineMatchGAN

Our LineMatchGAN is based on the pix2pix model (Isola et al., 2017). Specifically, it is a conditional generitve adversarial framework with a generator and a discriminator. Our model works by translating an input tensor $X \in \mathbb{R}^{W \times H \times 8}$ into a corrected optical flow $B \in \mathbb{R}^{W \times H \times 2}$, where X is a concatenation

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B2-2020, 2020 XXIV ISPRS Congress (2020 edition)



Figure 2. Examples from the ZebraScan Synthetic dataset.



Figure 3. Mobile structured light 3D scanner.

of a rough optical flow $A \in \mathbb{R}^{W \times H \times 2}$ and images for left and right cameras $S_l \in \mathbb{R}^{W \times H \times 3}$. Therefore, our generator G learns a mapping $G : (S_l, S_r, A) \rightarrow \hat{B}$ from the input data to the corrected optical flow.

We made two main contributions to the pix2pix model. Firstly, we added an inverted residual block (Sandler et al., 2018) after each convolutional layer to increase the number of training parameters and improve the optical flow reconstruction accuracy. Secondly, we add a preprocessing module to convert the input rough optical flow to the [-1, 1] range suitable for the training pipeline.

	Algorithm 1: Stereo Matching Algorithm					
	Input: The camera positions \mathbf{X}_{0}^{Left} , \mathbf{X}_{0}^{Right} , image sequences					
	from left camera S_L and from right camera S_R					
	Output: Optical flow A, which show difference between S_R and					
	\mathbf{S}_L					
1	for each Pair of images I_i^{Left} from S_L and I_i^{Right} from S_R do					
2	Find on \mathbf{I}_{i}^{Left} Backlight Lines \mathbf{L}^{Left} ;					
3	Find on \mathbf{I}_{i}^{Right} Backlight Lines \mathbf{L}^{Right} ;					
4	Calculate for \mathbf{I}_{i}^{Right} Epipolar Lines from points \mathbf{P}^{Left}					

- 5 located on \mathbf{L}^{Left} ; 5 Find in \mathbf{I}_{i}^{Right} points \mathbf{P}^{Right} that are the intersection between \mathbf{E}^{Right} and \mathbf{L}^{Right} ;
- 6 Calculate the difference between \mathbf{P}^{Right} and \mathbf{P}^{Left} ;
- 7 end

3.4 Dataset Generation

To train our LineMatchGAN framework, we collected a new dataset ZebraScan. It includes synthetic and real images of four solid objects: Gnome, Vase, Nefertiti, and David. Some of models were generated using a structured light scanner (Gnome, Vase). Some of models were collected from the open source 3D models available in the internet. Synthetic images were generated using the Blender 3D creation suite. The dataset includes three spilits: real, 'synthetic full' and 'synthetic reduced'.

The real split of the dataset (Figure 4) was collected using a structured light scanner. To generate the ground truth optical flow, we imported the ground truth 3D models into the scene and simulated the optical flow from the left to the right camera using the estimated external orientation of the cameras.

The 'synthetic full' split of the dataset consists of images that simulate a structured light scanner (Figure 2). We created the images from the left and the right cameras and simulate a moving light line. The first split includes 300 pairs of images. The program code was developed that processes the synthetic images and generates rough optical flow. The algorithm for the code is presented in the Algorithm 1.

The 'synthetic reduced' split of the dataset was created to expand the training dataset (Figure 2). We used special technique that creates not only images from cameras, but also simulates the processing of sequences by a scanner. We simultaneously generated ground-truth optical flow and rough optical flow. The second split contains 90 quadruples of images. The resolution of rendering images is 1280×960 pixels. We used a PNG format for the color images and OpenEXR format for the optical flow.

4. EXPERIMENTS

4.1 Network Training

The LineMatchGAN framework was trained on the training split of the ZebraScan dataset using the PyTorch library (Paszke et al., 2017). The training split includes 450 local image patches. The training was performed using the NVIDIA 2080 RTX GPU and took 12 hours. For network optimization, we use minibatch SGD with an Adam solver. We set learning rate to 0.0002 with momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$ similar to (Isola et al., 2017).

This contribution has been peer-reviewed.

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B2-2020, 2020 XXIV ISPRS Congress (2020 edition)



Figure 4. Examples from the ZebraScan Real dataset.

4.2 Qualitative Evaluation

We evaluate our LineMatchGAN framework qualitatively in terms of the smoothness of the reconstructed optical flow \hat{B} . For the evaluation, we use an independent test split of our ZebraScan dataset consisting of 20 images for the real split and 20 images for the synthetic split. We present results in Figure 5. The evaluation results prove the our LineMatchGAN model learns both optical flow completion and elimination of the discrepancies caused by uneven brightness of the surface texture.

4.3 Quantitative Evaluation

We evaluate our LineMatchGAN framework quantitatively in terms of L1 distance between the ground truth optical flow and the optical flow \hat{B} processed by our model. The evaluation results are presented in Table 2. The average endpoint optical flow error for the optical flow produced by the scanner (EPE base) includes discrepancies caused by the illumination and ranges from 0.5 to 0.1 pixel. Processing of the rough optical flow using our LineMatchGAN model reduces the error more then ten times.

We reconstructed 3D models of the objects using the rough optical flow and the optical flow filtered by our LineMatchGAN model to compare surface distance accuracy. The surface distance to the ground truth 3D model for the models reconstructed using a rough optical flow (SD base) was about 0.05 mm for all objects. 3D models reconstructed using the filtered optical flow have average surface distance of 0.01 mm.

Object	EPE base	EPE corr.	SD base	SD corr.
	pix	pix	mm	mm
Test field	0.162	0.002	0.025	0.005
Gnome	0.321	0.051	0.051	0.012
David	0.531	0.092	0.056	0.014

Table 2.	Optical	flow	EPE	and	surface	distance.
----------	---------	------	-----	-----	---------	-----------

5. CONCLUSION

We showed that a conditional adversarial loss function can be used to improve a 3D model reconstruction accuracy for mobile structured light scanner. Specifically, our model corrects the stereomatching errors causes by the uneven surface brightness of the object's texture.

We developed a LineMatchGAN conditional generative adversarial model for optical flow filtering. Our model receives a joint input consisting of a stereopair and the rough optical flow generated by a mobile scanner. The model attempts to reconstruct the errorfree optical flow using the learnt experience. We collected a ZebraScan dataset to train and evaluate our LineMatchGAN model. Both qualitative and quantitative evaluation demonstrates that the framework successfully removes errors in the optical flow. Moreover 3D models reconstructed using the filtered optical flow have five times lower surface distance error compared to the models reconstructed using the rough optical flow.

ACKNOWLEDGEMENTS

The reported study was funded by Russian Foundation for Basic Research (RFBR) according to the research project $N^{\rm o}$ 17-29-04509.

REFERENCES

Akca, D., Remondino, F., Novk, D., Hanusch, T., Schrotter, G. and Gruen, A., 2007. Performance evaluation of a coded structured light system for cultural heritage applications. Vol. 6491.

Bian, H. and Liu, K., 2016. Robustly decoding multiple-linestructured light in temporal Fourier domain for fast and accurate three-dimensional reconstruction. *Optical Engineering* 55(9), pp. 1 - 7.

Bosemann, W., 2011. Industrial photogrammetry: challenges and opportunities. In: F. Remondino and M. R. Shortis (eds), *Videometrics, Range Imaging, and Applications XI*, Vol. 8085, International Society for Optics and Photonics, SPIE, pp. 188 – 197.

Chalmers, E., Rhyne, T., Rocchini, C., Cignoni, P., Montani, C., Pingi, P. and Scopigno, R., 2001. A low cost 3d scanner based on structured light. *Computer Graphics Forum*.

Chen, H. and Shen, J., 2018. Denoising of point cloud data for computer-aided design, engineering, and manufacturing. *Engineering with Computers* 34, pp. 523541.

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B2-2020, 2020 XXIV ISPRS Congress (2020 edition)



Figure 5. Qualitative evaluation results.

Chen, Z., Zeng, Z., Shen, H., Zheng, X., Dai, P. and Ouyang, P., 2020. Dn-gan: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images. *Biomedical Signal Processing and Control* 55, pp. 101632.

Curless, B. and Levoy, M., 1995. Better optical triangulation through spacetime analysis. In: *Proceedings of IEEE International Conference on Computer Vision*, pp. 987–994.

Donlic, M., Petkovic, T. and Pribanic, T., 2017. On tablet 3d structured light reconstruction and registration. In: *The IEEE International Conference on Computer Vision (ICCV) Workshops*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger (eds), *Advances in Neural Information Processing Systems 27*, Curran Associates, Inc., pp. 2672–2680.

Gupta, M., Agrawal, A., Veeraraghavan, A. and Narasimhan, S., 2011. Structured light 3d scanning in the presence of global illumination. pp. 713 – 720.

Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A., 2017. Image-toimage translation with conditional adversarial networks. *CVPR*.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T., 2019. Analyzing and improving the image quality of stylegan.

Karras, T., Laine, S. and Aila, T., 2018. A style-based generator architecture for generative adversarial networks. *CoRR*.

Kniaz, V. V., 2019. FringeMatchNet: effective stereo matching onboard of mobile structured light 3D scanner. In: H. Liang, R. Groves and P. Targowski (eds), *Optics for Arts, Architecture, and Archaeology VII*, Vol. 11058, International Society for Optics and Photonics, SPIE, pp. 152 – 160.

Kniaz, V. V., Mizginov, V., Grodzitsky, L. and Bordodymov, A., 2020. GANcoder: robust feature point matching using conditional adversarial auto-encoder. In: P. Schelkens and T. Kozacki (eds), *Optics, Photonics and Digital Technologies for Imaging* *Applications VI*, Vol. 11353, International Society for Optics and Photonics, SPIE, pp. 59 – 68.

Knyaz, V. A., 2010. Multi-media projector single camera photogrammetric system for fast 3d reconstruction. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII-5, pp. 343–348.

Knyaz, V. A., 2015. Scalable Photogrammetric Motion Capture System "Mosca": Development and Application. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-5/W6, pp. 43–49.

Knyaz, V. A. and Maksimov, A. A., 2014. Photogrammetric technique for timber stack volume contol. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-3, pp. 157–162.

Krizhevsky, A., Sutskever, I. and Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems* 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States, pp. 1106–1114.

Li, F., Li, Q., Zhang, T., Niu, Y. and Shi, G., 2019. Depth acquisition with the combination of structured light and deep learning stereo matching. *Signal Processing: Image Communication* 75, pp. 111 – 117.

Lin, T.-H., 2020. Automatic 3d color shape measurement system based on stereo camera. *Applied Optics*.

O'Toole, M., Mather, J. and Kutulakos, K. N., 2014. 3d shape and indirect appearance by structured light transport. In: 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3246–3253.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., De-Vito, Z., Lin, Z., Desmaison, A., Antiga, L. and Lerer, A., 2017. Automatic differentiation in pytorch.

Piccirilli, M., Doretto, G., Ross, A. and Adjeroh, D., 2016. A mobile structured light system for 3d face acquisition. *IEEE Sensors Journal* 16(7), pp. 1854–1855. Pribanic, T., Dapo, H. and Salvi, J., 2010. Efficient and low-cost 3d structured light system based on a modified number-theoretic approach. *EURASIP J. Adv. Sig. Proc.*

Remondino, F., 2011. Heritage recording and 3d modeling with photogrammetry and 3d scanning. *Remote Sensing* 3(6), pp. 1104–1138.

Salvi, J., Fernandez, S., Pribanic, T. and Llado, X., 2010. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition* 43, pp. 2666–2680.

Sandler, M., Howard, A. G., Zhu, M., Zhmoginov, A. and Chen, L., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, pp. 4510–4520.

Taylor, C. J., 2012. Implementing high resolution structured light by exploiting projector blur. In: 2012 IEEE Workshop on the Applications of Computer Vision (WACV), pp. 9–16.

Wang, J., Sankaranarayanan, A. C., Gupta, M. and Narasimhan, S. G., 2016. Dual structured light 3d using a 1d sensor. In: B. Leibe, J. Matas, N. Sebe and M. Welling (eds), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, pp. 383–398.

Wang, Y. and Feng, H.-Y., 2014. Modeling outlier formation in scanning reflective surfaces using a laser stripe scanner. *Measurement* 57, pp. 108 – 121.

Wedel, A. and Cremers, D., 2011. *Stereo Scene Flow for 3D Motion Analysis*. Springer.

Xiao, Y.-L., Wen, Y., Li, S., Zhang, Q. and Zhong, J., 2020. Large-scale structured light 3d shape measurement with reverse photography. *Optics and Lasers in Engineering* 130, pp. 106086.

Zhu, J.-Y., Park, T., Isola, P. and Efros, A. A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Computer Vision (ICCV), 2017 IEEE International Conference on.*