WIRE STRUCTURE IMAGE-BASED 3D RECONSTRUCTION AIDED BY DEEP LEARNING

Vladimir V. Kniaz[1,2,*], Sergey Yu. Zheltov[1], Fabio Remondino[3], Vladimir A. Knyaz[1,2], Artyom Bordodymov[1], Armin Gruen[4]

[1] State Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia – (vl.kniaz, zhl, knyaz, bordodymov)@gosniias.ru
[2] Moscow Institute of Physics and Technology (MIPT), Dolgoprudny, Russia
[3] Bruno Kessler Foundation (FBK), Trento, Italy – remondino@fbk.eu
[4] ETH Zurich, Switzerland – armin.gruen@geod.baug.ethz.ch

Commission II, WG II/8

KEY WORDS: structure from motion, wire structures 3D reconstruction, segmentation, deep learning, Shukhov Radio tower

ABSTRACT:

Objects and structures realized by connecting and bending wires are common in modern architecture, furniture design, metal sculpting, etc. The 3D reconstruction of such objects with traditional range- or image-based methods is very difficult and poses challenges due to their unique characteristics such as repeated structures, slim elements, holes, lack of features, self-occlusions, etc. Complete 3D models of such complex structures are normally reconstructed with lots of manual intervention as automated processes fail in providing detailed and accurate 3D reconstruction results.
This paper presents the image-based 3D reconstruction of the Shukhov hyperboloid tower in Moscow, a wire structure built in 1922, composed of a series of hyperboloid sections stacked one to another to approximate an overall conical shape. A deep learning approach for image segmentation was developed in order to robustly detect wire structures in images and provide the basis for accurate corresponding problem solutions. The developed `WireNet` convolution neural network (CNN) model has been used to aid the multi-view stereo (MVS) process and to improve robustness and accuracy of the image-based 3D reconstruction approach, otherwise not feasible without masking the images automatically.

1. INTRODUCTION

Wire structures, such as radio poles, spider webs, wire jewelry, etc., pose challenges for active and passive 3D reconstruction techniques. Complicated interweaved wire structures usually have a large number of holes, repeated patters, textureless surfaces, specular reflections, ambiguities and thin elements that could be too small to be detected by laser scanners or accurately matched by multi-view stereo (MVS) algorithms. Therefore, such complex structures are normally reconstructed with lots of manual intervention as automated processes fail in providing detailed and accurate 3D reconstruction results. Even if dense point cloud can be derived, the modeling steps necessitate many manual interventions due to the complexity of the structures.

Inspired by the progress of deep learning techniques in solving challenging tasks in photogrammetry and computer vision, this work tries to exploit a "human-like" machine learning approach to perform object masking in images and improve the MVS process. Image masking is a very time-consuming part of the image processing 3D pipeline and often the only way to achieve detailed 3D results.

1.1 Aims of the work

The paper presents a methodology where the traditional image-based 3D reconstruction pipeline is aided by a deep convolutional neural network (CNN) to improve the dense image matching process. We propose a modified neural network architecture, named `WireNet`, for the automatic semantic labelling of foreground wire structures in UAV images. The oriented and segmented images are then processed within a MVS method to derive dense point clouds. The methodology is applied to some 500 UAV images acquired to perform the 3D reconstruction of the Shukhov tower in Moscow (Russia), included in the World Monument Watch[1] since 2016.

2. STATE OF THE ART

2.1 3D reconstruction of wire structures

Remote sensing techniques based on laser scanning or multi-view stereo (MVS) are widely used for non-contact documentation, monitoring and inspection of industrial constructions. All methods offer successful performances for objects having relatively large and smooth surfaces such as buildings, pipelines, bridges, etc. On the other hand, wire-like structures are more complicated for automated 3D reconstruction due to limited sensor resolutions and other problems mentioned before.

Some previous works obtained a 3D reconstruction of wire structures in the form of individual curve segments (Teney, Piater, 2012, Usumezbas et al., 2016). In case a dense point cloud can be produced, either with laser scanning or MVS methods, Huang et al. (Huang et al., 2013) presented an automated solution to extract curve skeletons based on the L1-medial axis.

(Morioka et al., 2013) firstly extract from a point cloud the topology of a wire structure as a graph and then the extracted structure is presented as a combination of cylindrical surfaces centered

*Corresponding author

[1] `https://www.wmf.org/project/shukhov-tower`

(a) (b) (c) (d)

Figure 1. The world's first diagrid hyperboloid water tower (37 m height) built by V. Shukhov for the All-Russian Exposition in 1896 in Nizhny Novgorod, Russia (a), now located in Polybino (b). The Shukhov radio tower, also known as the Shabolovka tower, build between 1919 and 1922 in Moscow, Russia (c) and some details of the wire hyperboloid structure (d).

along the edges of the graph. Using Delaunay tetrahedralization, the initial edges are then created and simplified by applying iterative edge contractions to extract the graph representing the wire topology. Finally, an optimization technique is applied to the positions of the surface in order to improve the geometrical accuracy of the final reconstructed object.

(Su et al., 2018) retrieved the topology of a spider web developing an innovative experimental method to directly capture the complete digital 3D spider web architecture with micron scale image resolution. The authors built an automatic segmentation and scanning platform to obtain high-resolution 2D images of individual cross-sections of the web that were illuminated by a laser sheet. Processing these images, the digital 3D fibrous network of the spider web was reconstructed.

(Martin et al., 2014) presented a method to reconstruct thin tubular structures from a dense set of images using physics-based simulation of rods to improve accuracy. They used a 3D occupancy grid to disambiguate 2D crossings of cables.

(Liu et al., 2017) presented a novel image-based reconstruction method of wire objects based on 3 images as input which exploits unique characteristics of wire objects (simplicity - the object is composed of a few wires, and smoothness - each wire is bent smoothly) to recover the global 3D wire decomposition. A project aimed at preserving information about the Shukhov Shabolovka radio tower has been carried out by means of laser scanning (Leonov et al., 2015).

2.2 Deep convolutional neural networks

In the last years, deep convolutional neural networks (CNNs) started to be employed within the 3D image-based pipeline in order to boost the processing and facilitate some steps. According to their role in the 3D reconstruction pipeline, semantic segmentation networks could be divided intro three broad groups: CNNs for single-photo 3D reconstruction, CNNs for feature matching and CNNs for semantic segmentation and boosting of SfM/MVS procedures.

Single photo 3D reconstruction has been recently intensively studied. Multiple neural network models were proposed for reconstruction of objects and buildings from a single image using conditional generative adversarial networks – GAN (Girdhar et al., n.d., Shin et al., 2018, Choy et al., 2016, Xie et al., 2019, Shin et al., 2019, Knyaz et al., 2018, Kniaz et al., 2019). While deep models such as `Pix2Vox` (Xie et al., 2019) and `Z-GAN` (Kniaz et al., 2019) proved to reconstruct complex structures from a single photo, but a large training dataset is required to achieve the desired quality. However, no public datasets of wire structures is available to date to train such models.

Feature matching networks (Yi et al., 2016, Ono et al., 2018, Christiansen et al., 2019, Shen et al., 2019, Kniaz et al., 2020) seems to outperform handcrafted feature detectors/descriptor methods. Still, their performance is closely related to the similarity of local image patches in the training dataset with respect to the images used during inference. However, repeating metal beams of wire structures are not present in modern datasets.

Another application of deep learning and CNNs is the semantic segmentation of images (Ronneberger et al., 2015, Sandler et al., 2018, Minaee et al., 2020, Kniaz, 2018, Kniaz, 2019). Thanks to this, CNNs have also demonstrated their potential for multi-view stereo (Huang et al., 2018, Kuhn et al., 2019, Stathopoulou, Remondino, 2019).

3. PROJECT BACKGROUND

3.1 Shukhov and his hyperboloid structures

Vladimir Shukhov (1852-1939) was a genius Russian engineer, scientist and architect renowned for his pioneering works in the area of world hyperboloid and diagrid shell structures.

Shukhov invented the world's first hyperboloid structure in 1890 and he built the first diagrid tower for the All-Russian Exhibition in Nizhny Novgorod (Russia) in 1896 (Figure 1a). Later Shukhov designed the so called Shabolovka tower, which was built in Moscow under his direction in 1920-1922. The Shukhov radio tower in Moscow is a landmark in the history of structural engineering and an emblem of the creative genius of an entire generation of modernist architects in the years that followed the Russian Revolution. Shukhov wanted to build a light but solid construction higher than the Eiffel Tour (Paris, France) but much lighter. His approach allowed to achieve a high degree of rigidity of the tower and to reduce significantly the weight of the construction. The planned height of the nine-sectioned hyperbolic

tower was 350 meters but, due to Civil War and lack of resources, the project was revised and the height reduced to ca 150 m, with a weight of 240 tons and some 7,500 of individual connecting wire elements. The Shabolovka tower has played a very important role in the history of radio and TV broadcasting in the USSR and Russia and was recognized as monument of international heritage (Shukhov et al., 1990).

3.2 3D surveying

Studying the condition of such a structure with almost a century of history is a difficult task. Various specialists are required, from civil engineers to metalworkers or corrosion specialists and, of course, 3D measurements of the tower play a fundamental role to understand how its characteristics have changed over the long period of time. Unfortunately, the original documentation for the Shukhov towers has not been preserved. The last surveys of the tower were conducted in 1947 and later in 1971. In 2009, under the direction of Prof. Dr. Uta E. Hassler and Prof. Dr. Armin Gruen (ETH Zurich, Switzerland) and Prof. Dr. Rainer Graefe (University of Innsbruck, Austria), a research project named "Shukhov's strategies for thin iron constructions", was initiated. The project was dedicated to the study of Shukhov's engineering achievements and it was joined in 2011 by the Russian partners GosNIIAS (team led by Prof. S. Yu. Zheltov) and Prof. Dr. Felix L. Chernousko, Chairman of the Shukhov Committee of the Russian Academy of Sciences.

It was decided to conduct photogrammetric measurements of two structures: the 1922 Shukhov radio tower of Moscow (Russia) – also known as the Shabolovka tower, as well as the Shukhov tower built in 1896 for the All-Russian exposition in Nizhny Novgorod (Russia), now located in Polybino (Lipetsk region).

In 2012, in the year of the 90th anniversary of the Moscow tower, a team of the Institute of History of Technology of the Russian Academy of Sciences, led by Andrey Leonov, conducted a laser scanning 3D surveying of the Shabolovka tower (Leonov et al., 2015). It has resulted in a precise polygonal 3D model using both the results of the scanning and some existing drawings. The transition from an unstructured point cloud to a highly structured representation has been performed using a special semi-automated methodology to model deformed steel elements of hyperboloid sections. To reproduce the individual shapes of twisted rods and rings, pre-defined cross-sections were used, which were precisely positioned in a point cloud. The connection joints of steel elements were modeled using drawings based on measurements that were made in 1947. The combination of various 3D modeling methods for different parts of the tower allowed to visualize the geometry of the huge steel construction with high accuracy (about 10 mm). At the moment, the produced laser scanning 3D model is the most accurate in existence, but it only creates the geometry of the tower, without giving information about its state of conservation.

In the same years, two attempts were made to conduct a photogrammetric UAV survey and obtain the geometric characteristics of the tower and its textured 3D model. It took ca 1.5 years to get permission to fly around the tower, since this area is also the active TV center of Moscow with live broadcast systems. The first UAV experiments, under the guidance of Prof. Armin Gruen (ETH Zurich) found out that the large number of cellular transmitters installed on the tower create powerful electromagnetic fields that disturb the drone's navigation system. Two years later, the same team performed a second, more successful attempt, collecting some 600 images with a GSD spanning from 5 mm (lower part) to 10 mm (upper part).

Figure 2. One of the GCPs measured on the basement of the tower (left). Scheme of the 9 UAV strips flown to capture images of the tower (right).

The tower was also surveyed with a Geomax Zoom 25pro total station in order to measure 10 GCPs useful to georeference the produced 3D results. Special markers (Figure 2a) were used for labeling reference points, placed inside and outside the tower, on the foundation and at an altitude of approximately 50 meters.

4. UAV-BASED 3D RECONSTRUCTION

4.1 UAV image-based survey

The UAV images of the Shukhov tower in Moscow were acquired in 2015 using a Falcon 8 octocopter drone equipped with a Sony NEX-5T camera (16 MPixels, 16 mm focal length). The onboard camera acquired a set of ca 600 images divided in nine vertical strips with an interval of about 3 m (Figure 2b). The image GSD varies from 5 mm to 10 mm. The lack of auto-piloting and the difficult location of the mission did not provide a uniform and constant image overlap.

A subset of the images was processed within a photogrammetric pipeline with the aid of an automated background masking in the MVS step in order to match only the wire structure of the foreground and achieve more accurate results. The segmentation of the wire structures in the images was performed with a CNN model named `WireNet` (Figure 3).

4.2 `WireNet` Model Architecture

Local patch similarity is the main problem for 3D object reconstruction using structure-from-motion pipeline. False matches result in poor quality of camera external orientation estimation and deform the resulting 3D model. The main reason for the false feature point matching are the repeating structures of the tower. Moreover, feature point matching algorithms confuse points on the foremost sections of the tower with the rear points visible through the windows in the wire structure.

Masking of the irrelevant object parts to improve the stereo matching accuracy is a well-known technique for improving the quality of 3D reconstruction. The developed approach was inspired by a recent research (Stathopoulou, Remondino, 2019) regarding semantic segmentation for improving accuracy of a structure-from-motion pipeline. Still, the total number of photos in the UAV survey exceeded 700. Therefore, manual labelling of all acquired data was impossible. A deep learning based technique was proposed to overcome this problem. Firstly, a simple `U-net` (Ronneberger et al., 2015) model was trained. However, the quality of segmentation using the `U-net` model was insufficient for correct point matching.

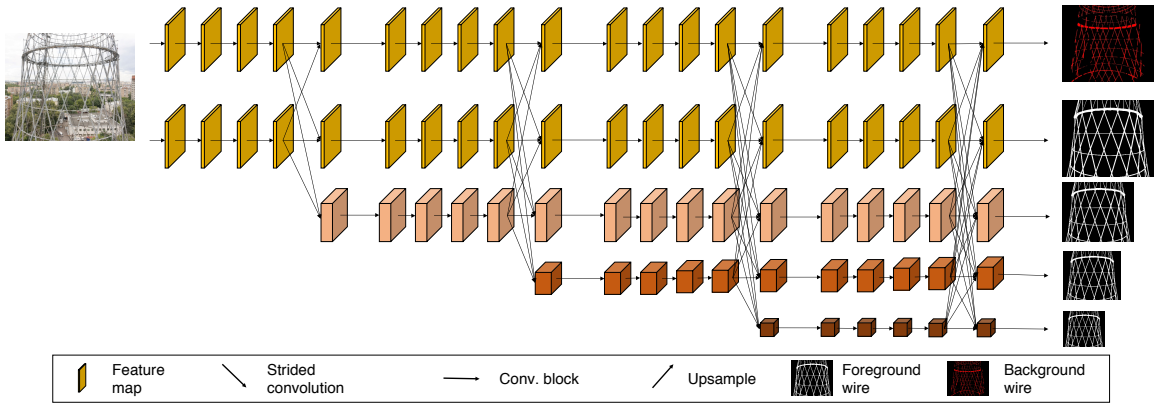


Figure 3. Overview of the `WireNet` model.

The segmentation results of a `HRNetv2` (Sun et al., 2019) was much more impressive. Still, the model was unable to distinguish between foreground and background wire structures on the images. Hence, a new model based on the `HRNetv2` was developed, that was termed `WireNet`. The idea of an `HRNetv2` model is to provide high-resolution labelling by connecting high-to-low resolution convolutions in parallel. These convolutions are repeated multi-scale and fused in multiple points of the model architecture. All in all, the architecture of `HRNetv2` includes four stages. The 2nd, 3rd and 4th stages are formed by repeating multi-resolution convolution blocks. A single multi-resolution block consists of a group convolutions and followed by a multi-resolution convolution. The multi-resolution group convolution is a simple extension of the group convolution, which divides the input channels into several subsets of channels and performs a regular convolution over each subset and over different spatial resolutions separately.

The multi-resolution convolution resembles the multi-branch fully connected regular convolution. A regular convolution can be divided as multiple small convolutions (Zhang et al., 2017) The input channels are divided into several subsets, and the output channels are also divided into several subsets. The input and output subsets are connected in a fully connected fashion and each connection is a regular convolution. Each subset of output channels is a summation of the outputs of the convolutions over each subset of input channels.

The following contributions were applied to the `HRNetv2` in the `WireNet` model to improve segmentation of the frontmost and rare parts of the tower. Firstly, an additional branch of multi-resolution convolutions was added that produces the segmentation of the rare wire structures. Secondly, a negative log likelihood loss function was used to improve the segmentation accuracy. The resulting architecture is presented in Figure 3.

Therefore, two loss functions govern the training process of our `WireNet` model

$$\mathcal{L} = \lambda_f \cdot \mathcal{L}_{NLL}(L_f, \hat{L}_f) + \lambda_b \cdot \mathcal{L}_{NLL}(L_b, \hat{L}_b), \qquad (1)$$

where L_f is the ground truth foreground segmentation, $\hat{L}_f$ is the predicted foreground segmentation, L_b is the ground truth back-

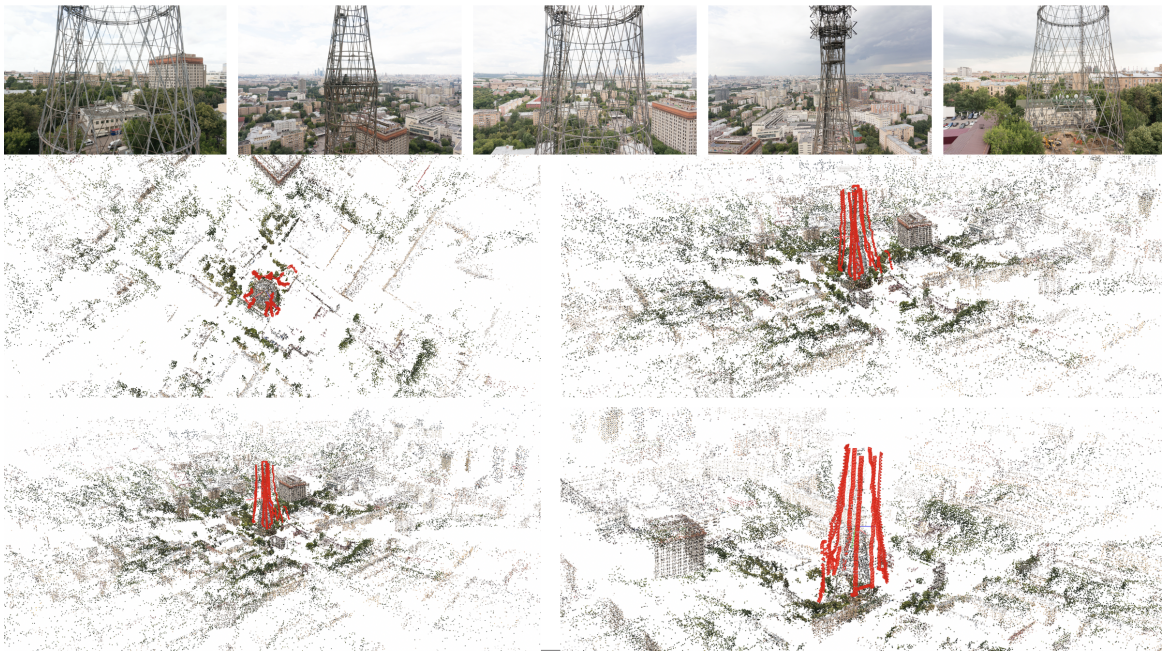


Figure 4. Some UAV images acquired to survey the tower (above). Different views of the recovered camera poses - red dots - and sparse point cloud (below).

ground segmentation, $\hat{L}_b$ is the predicted background segmentation, λ_f and λ_b are the hyperparameters, $\mathcal{L}_{NLL}(A, B)$ is a negative log likelihood loss function given by

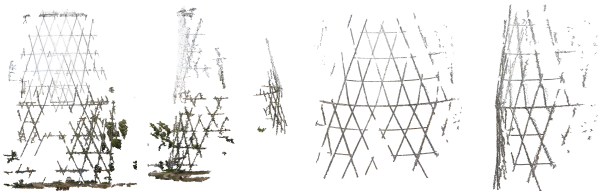
$$\mathcal{L}_{NLL}(A, B) = \frac{1}{2 \cdot w \cdot h} \sum_{x=0}^{w} \sum_{y=0}^{h} \sum_{i=0}^{1} -m_i \log(B(A(x, y), x, y),$$

where w, h are the image width and height, $A \in \{0, 1\}^{w \times h}$ is the ground truth semantic labelling, $B \in [0, 1]^{2 \times w \times h}$ is multichannel probability map defining the probability of pixel with coordinates (x, y) belonging to class i, m_i is the class weight for class i.

We manually labeled 5% of the whole UAV image dataset to train our `WireNet` model. The evaluation of the model on the independent test set demonstrated 85% accuracy for the Intersection-over-Union (IoU) metric.

5. RESULTS

After discarding the images with insufficient image quality, the photogrammetric image processing was performed on the remaining set of ca 500 images in COLMAP[2] (Schönberger et al., 2016, Schönberger, Frahm, 2016). As the images contain the far-away scene in the background of the tower, given the short baselines between the UAV images, a threshold on the ray intersection angle was imposed in order to avoid 3D points reconstructed under a very small intersection angle. Once the camera poses and sparse point cloud were derived (Figure 4), a dense image matching was applied to derive a dense point cloud of the wire tower.



(a) (b)

Figure 5. Dense point cloud (3.3 mil points) produced on a set of images without applying any masking (a). Dense point cloud (2.1 mil points) derived applying a masking (b).

The main problem preventing an accurate dense 3D reconstruction of the wire structure was the incorrect and noisy dense matching result when no masks were used in the MVS step (Figure 5(a)).

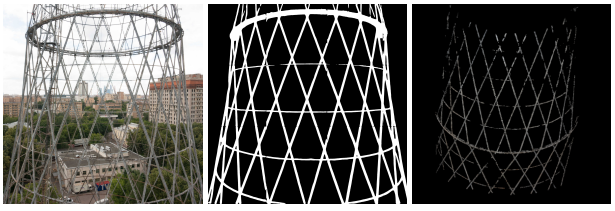
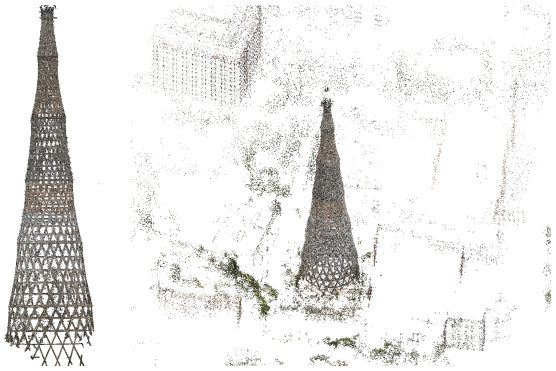


Figure 6. Results for the proposed `WireNet` model. An input image (left), the predicted semantic segmentation (middle) and dense point cloud generated using the created masks (right).

A detailed image masking was therefore necessary to constrain the patch-based MVS method. The developed `WireNet` (Section 4.2) was trained on a manually labeled 5% subset of the UAV images. The trained `WireNet` model was then used on the rest of the images for the automatic segmentation of the wire structures, providing a quick and robust masking of the background scene and also eliminating many outliers in the dense point cloud (Figure 5(b) and Figure 6).



(a) (b)

Figure 7. Photogrammetric dense point cloud of Shukhov tower (a), inserted in the sparse cloud of the surrounding area (b).

The final image-based dense point cloud on the entire tower, composed of ca 5.7 mil. points, is shown in Figure 7. As portions of the lower part of the tower were occluded by vegetation, some holes in the point cloud are present.

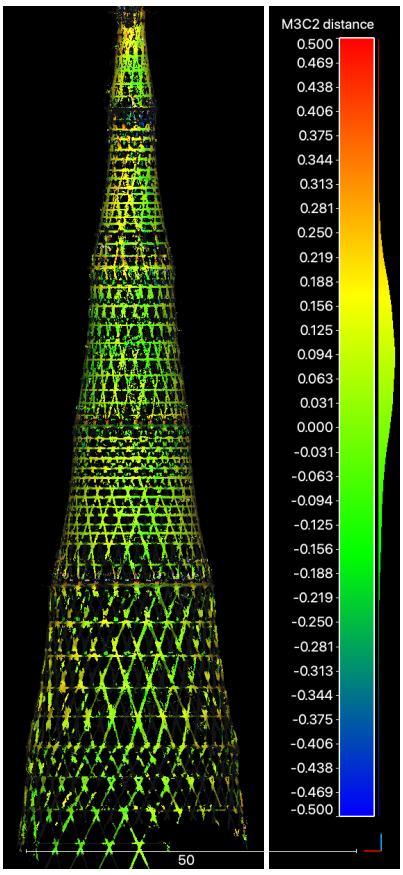


Figure 8. 3D comparison between the UAV-based point cloud and the available laser scanning cloud.

Thanks to the availability of a laser scanning (32 mil.) point cloud of the tower ((Leonov et al., 2015), a geometric comparison between the 3D data was performed.

[2] `https://colmap.github.io`

The best-fit alignment of the 3D data resulted in a RMS of 0.26 m whereas the Cloud-to-Cloud distances are shown in Figure 8.

Taking into account the different conditions of the two surveys (including temperature, wind, etc.), the complexity of the wire structure and the unfavorable image network design, the achieved results are satisfactory and could be used for further analyses and modeling.

6. CONCLUSIONS

The paper presented the UAV-based 3D survey of the Shukhov radio tower in Moscow, Russia. The main aim of the work was to show how geometric processing can be aided (not replaced) by deep learning. A deep Convolutional Neural Network (CNN) model, named `WireNet`, was developed in order to support a MVS procedure for the 3D reconstruction of the wire large structure. The semantic segmentation allowed to automatically produce image masks and avoid a time-consuming manual labelling on some 500 images. The masking was of fundamental importance to achieve 3D point clouds not obtainable without masks.

The project highlighted difficulties in the 3D reconstruction of such complex wire structures, including the definition of a proper image network due to electromagnetic interferences of mobile phone antennas on the tower with the UAV's navigation system.

ACKNOWLEDGEMENTS

The reported study was funded by Russian Foundation for Basic Research (RFBR) according to the research project 17-29-04410. We also acknowledge the support of the SNF (Swiss National Science Foundation) and ETH Zurich for the UAV-based data acquisition.

REFERENCES

Choy, C. B., Xu, D., Gwak, J., Chen, K., Savarese, S., 2016. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII*, 628–644.

Christiansen, P. H., Kragh, M. F., Brodskiy, Y., Karstoft, H., 2019. UnsuperPoint: End-to-end Unsupervised Interest Point Detector and Descriptor. *CoRR*, abs/1907.04011. http://arxiv.org/abs/1907.04011.

Girdhar, R., Fouhey, D. F., on, M. R. E. C., 2016, n.d. Learning a predictable and generative vector representation for objects. *Springer*, 702–722.

Huang, H., Wu, S., Cohen-Or, D., Gong, M., Zhang, H., Li, G., Chen, B., 2013. L1-Medial Skeleton of Point Cloud. *ACM Trans. Graph.*, 32(4). https://doi.org/10.1145/2461912.2461913.

Huang, P., Matzen, K., Kopf, J., Ahuja, N., Huang, J., 2018. Deepmvs: Learning multi-view stereopsis. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2821–2830.

Kniaz, V. V., 2018. Conditional GANs for semantic segmentation of multispectral satellite images. L. Bruzzone, F. Bovolo (eds), *Image and Signal Processing for Remote Sensing XXIV*, 10789, International Society for Optics and Photonics, SPIE, 259 – 267.

Kniaz, V. V., 2019. Deep learning for dense labeling of hydrographic regions in very high resolution imagery. L. Bruzzone, F. Bovolo (eds), *Image and Signal Processing for Remote Sensing XXV*, 11155, International Society for Optics and Photonics, SPIE, 283 – 292.

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

Kniaz, V. V., Remondino, F., Knyaz, V. A., 2019. GENERATIVE ADVERSARIAL NETWORKS FOR SINGLE PHOTO 3D RECONSTRUCTION. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9, 403–408. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W9/403/2019/.

Knyaz, V. A., Kniaz, V. V., Remondino, F., 2018. Image-to-voxel model translation with conditional adversarial networks. *Computer Vision - ECCV 2018 Workshops - Munich, Germany, September 8-14, 2018, Proceedings, Part I*, 601–618.

Kuhn, A., Sormann, C., Rossi, M., Erdler, O., Fraundorfer, F., 2019. DeepC-MVS: Deep Confidence Prediction for Multi-View Stereo Reconstruction. *ArXiv, abs/1912.00439*, abs/1912.00439.

Leonov, A. V., Anikushkin, M. N., Ivanov, A. V., Ovcharov, S. V., Bobkov, A. E., Baturin, Y. M., 2015. Laser scanning and 3D modeling of the Shukhov hyperboloid tower in Moscow. *Journal of Cultural Heritage*, 16(4), 551 - 559. http://www.sciencedirect.com/science/article/pii/S129620741400137X.

Liu, L., Ceylan, D., Lin, C., Wang, W., Mitra, N. J., 2017. Image-Based Reconstruction of Wire Art. *ACM Trans. Graph.*, 36(4). https://doi.org/10.1145/3072959.3073682.

Martin, T., Montes, J., Bazin, J.-C., Popa, T., 2014. Topology-aware reconstruction of thin tubular structures. *SIGGRAPH Asia 2014 Technical Briefs*, SA '14, Association for Computing Machinery, New York, NY, USA.

Minaee, S., Boykov, Y., Porikli, F. M., Plaza, A. J., Kehtarnavaz, N., Terzopoulos, D., 2020. Image Segmentation Using Deep Learning: A Survey. *arXiv:2001.05566*, abs/2001.05566.

Morioka, K., Ohtake, Y., Suzuki, H., 2013. Reconstruction of wire structures from scanned point clouds. G. Bebis, R. Boyle, B. Parvin, D. Koracin, B. Li, F. Porikli, V. Zordan, J. Klosowski, S. Coquillart, X. Luo, M. Chen, D. Gotz (eds), *Advances in Visual Computing*, Springer Berlin Heidelberg, Berlin, Heidelberg, 427–436.

Ono, Y., Trulls, E., Fua, P., Yi, K. M., 2018. Lf-net: Learning local features from images. *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, 6237–6247.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

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

Schönberger, J. L., Frahm, J.-M., 2016. Structure-from-motion revisited. *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Schönberger, J. L., Zheng, E., Pollefeys, M., Frahm, J.-M., 2016. Pixelwise view selection for unstructured multi-view stereo. *European Conference on Computer Vision (ECCV)*.

Shen, X., Wang, C., Li, X., Yu, Z., Li, J., Wen, C., Cheng, M., He, Z., 2019. RF-Net: An End-to-End Image Matching Network based on Receptive Field. *CoRR*, abs/1906.00604. http://arxiv.org/abs/1906.00604.

Shin, D., Fowlkes, C., Hoiem, D., 2018. Pixels, voxels, and views: A study of shape representations for single view 3d object shape prediction. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Shin, D., Ren, Z., Sudderth, E. B., Fowlkes, C. C., 2019. 3d scene reconstruction with multi-layer depth and epipolar transformers. *The IEEE International Conference on Computer Vision (ICCV)*.

Shukhov, V. G., Graefe, R., Gappoev, M., Pertschi, O., Bach., K., 1990. *Vladimir G. Ŝuchov, 1853-1939: die Kunst der sparsamen Konstruktion.* Deutsche Verlags-Anstalt, Stuttgart, 195.

Stathopoulou, E.-K., Remondino, F., 2019. MULTI-VIEW STEREO WITH SEMANTIC PRIORS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W15, 1135–1140. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W15/1135/2019/.

Su, I., Qin, Z., Saraceno, T., Krell, A., Mühlethaler, R., Bisshop, A., Buehler, M. J., 2018. Imaging and analysis of a three-dimensional spider web architecture. *Journal of The Royal Society Interface*, 15(146), 20180193. https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2018.0193.

Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., Mu, Y., Wang, X., Liu, W., Wang, J., 2019. High-Resolution Representations for Labeling Pixels and Regions. *arXiv: 1904.04514*.

Teney, D., Piater, J., 2012. Sampling-based multiview reconstruction without correspondences for 3d edges. *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, 160–167.

Usumezbas, A., Fabbri, R., Kimia, B. B., 2016. From multiview image curves to 3d drawings. B. Leibe, J. Matas, N. Sebe, M. Welling (eds), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 70–87.

Xie, H., Yao, H., Sun, X., Zhou, S., Zhang, S., 2019. Pix2vox: Context-aware 3d reconstruction from single and multi-view images. *The IEEE International Conference on Computer Vision (ICCV)*.

Yi, K. M., Trulls, E., Lepetit, V., Fua, P., 2016. LIFT: learned invariant feature transform. *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI*, 467–483.

Zhang, T., Qi, G.-J., Xiao, B., Wang, J., 2017. Interleaved group convolutions. *The IEEE International Conference on Computer Vision (ICCV)*.