# POINT CLOUD SEGMENTATION USING IMAGE PROCESSING TECHNIQUES FOR STRUCTURAL ANALYSIS

V.Barrile [1], G. Candela [1], A. Fotia [1].

[1] Dept. of Civil and Environmental Engineering, Mediterranea University of Reggio Calabria, 89128 Reggio Calabria, Italy - (vincenzo.barrile, gabriele.candela, antonino.fotia)@unirc.it

**KEY WORDS:** Aerial Photogrammetry, UAV survey, Point cloud Segmentation, Image semantic segmentation, CNN

**ABSTRACT:**

Modern surveying techniques, with the combined use of Unmanned Aerial Vehicles (UAV) with low-cost photographic sensors, and photogrammetric techniques, allows obtaining a precise virtual reconstruction of environment and object with centimetre accuracy. Recently, the diffusion of UAV allows the survey of extensive areas significantly reducing survey time and costs. The raw output obtainable from such survey operations consists of a three-dimensional point cloud. Numerous applications in architecture, monitoring and surveying and structural analysis require objects identification in the 3d scene to classify different element in the acquired scene and extract relevant information. Point cloud analysis, and in particular segmentation and classification techniques, are actually used to identify objects within the scenes, assign to a specific class and use them for subsequent studies. These techniques represent an open research theme and the key to add value to the entire process. Actual methodologies are based on 3d spatial analysis on the point cloud. In this paper, starting from photogrammetric reconstruction, a methodology for segmentation and classification of point cloud based on image analysis is presented. The object identification on the image's dataset is performed using a Neural Network and subsequently the identified object on dataset are transfer into the 3d environment. This classification is performed to segment structural parts of bridges and viaduct, acquire geometric information, and perform a structural analysis to preserve relevant and ancient structure. A case study for the segmentation of the point cloud acquired with an aerial survey of a Viaduct is presented. The performed segmentation allows obtaining structural elements of different type of viaduct and bridges, is propaedeutic to verify the health of the structure and schedule maintenance intervention. The methodology can be applied to different type of bridges, from reinforced concrete to ancient masonry to preserve the state of conservation.

## 1. INTRODUCTION

### 1.1 Aerial photogrammetry using Unmanned Aerial Vehicle

The aerial photogrammetry can be considered as the principal means through the photogrammetry has developed. It was the basic data source for making maps by photogrammetric means. In last years the use of UAS, from military to geomatics field became common thank to the application in close-range aerial domain introducing a low-cost alternative to manned aerial photogrammetry. This rapid development can be explained by the spreading of low-cost platform combined with digital cameras and GNSS system, and the rising of digital photogrammetry (Linder, 2016). Today the use of UAS compared with traditional airborne platform decrease the operational costs, reduce the risk of access in harsh environment and still maintain high accuracy potential (Remondino, Barazzetti, Nex, Scaioni, & Sarazzi, 2012); moreover, the use of VTOL UAV, without the need of runway to take-off, allows to quickly derive high temporal and spatial resolution images in rapid response to emergency situation when critical information is needed were quick.
Aerial photogrammetry with UAV it's possible to obtain a map, digital model and 3d data of the surveyed data (Colomina & Molina, 2014) (Nex & Remondino, 2014).

## 2. 3D RECONSTRUCTION USING PHOTOGRAMMETRY

### 2.1 SFM MVS algorithms

Photogrammetric principles and algorithms allow, as discussed, the reconstruction of 3d scene starting from different images acquired respecting stereographic criteria. In particular, the well-known computer vision algorithm *Structure from Motion* is the most reliable and utilized algorithms for the generation of a valuable 3d model from 2d imagery. *Structure from motion* (SFM) is a topographic survey technique that has emerged from advances in computer vision and traditional photogrammetry since the 1980s with the development of software GUI (Graphical User Interface). It can produce high-quality, dense, three-dimensional point clouds of a scene/area with minimal cost. In contrast to traditional photogrammetry, *SFM* uses algorithm to identify matching features in a collection of overlapping digital images and calculates the camera location and orientation from the differential position of multiple matched features. Based on these calculations overlapping imagery can be used to reconstruct a "sparse" 3d point cloud model of the acquired scene. Later the model from Sfm is refined to a much finer resolution using *Multi-Stereo-View* methods, with cheap hardware and software and fast times compared with other digital surveying techniques.
The application of SFM in geoscience field, geomatics and survey recently spread thanks to the great advantages and applicability combined with new technologies. Today the use of SFM in geoscience became relevant thanks to the emergence of affordable commercial user-friendly software and rapid developments of UAVs and SFM platform. Moreover, thanks to

these characteristics the impact of SFM is arguably going to be greater than that associated with airborne Light Detection and Ranging (Lidar) because this technique and developed technology democratise data collection and development of fine-resolution 3d models at all scales; in fact to produce advanced data products, very little input data are required: as little as a photograph set from an uncalibrated, compact and often cheap, camera.

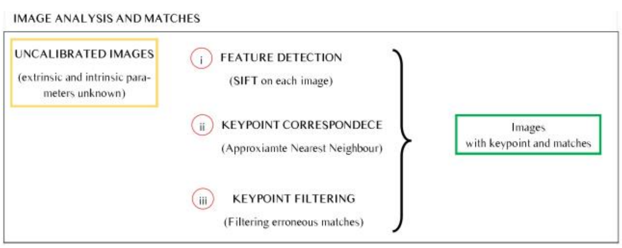
### 2.2 3d reconstruction workflow

SFM, as applied in geoscience and survey is more than a single technique: can be properly defined as a workflow employing multiple algorithms developed from traditional photogrammetry, survey techniques and three-dimensional (3d) computer vision. The full workflow is known as Structure from Motion Multi-View Stereo (SFM – MVS) to account for the Multi-View stereo algorithms used in final stages. Many commercial SfM-MVS software packages do not detail specific procedure applied to solve the problem. The basic concept for 3d reconstruction starting from uncalibrated imagery are presented, for deep understanding of mathematical formulas the interested reader can find relevant information on (David G Lowe, 2004). The basic process to reconstruct the 3d scene geometry from a set of images where the extrinsic and intrinsic calibration parameters are unknown, could be divided into three main steps: 1) Image analysis for and matches for estimation of unknown camera parameters, 2) Application of Structure from Motion (SFM) algorithm and 3) Multi-View Stereo for 3d dense cloud generation. The detailed workflow for 3d reconstruction (Snavely, Seitz, & Szeliski, 2008) is summarized as follows:

1) Image analysis and matches:
   - (i) Detect image features on key point;
   - (ii) Keypoint correspondence between different images;
   - (iii) Identify geometrically consistent matches;

2) Structure From Motion:
   - (iv) SFM of simultaneously estimating 3d scene geometry: camera pose and internal camera parameters through bundle adjustment;
   - (v) Scaling and georeferencing the resultant scene geometry;
   - (vi) Optimizing the identified parameters in the bundle adjustment using know Ground Control Points (GCP);

3) Multi View Stereo:
   - (vii) Clustering image sets for efficient processing;
   - (viii) Apply MVS algorithms



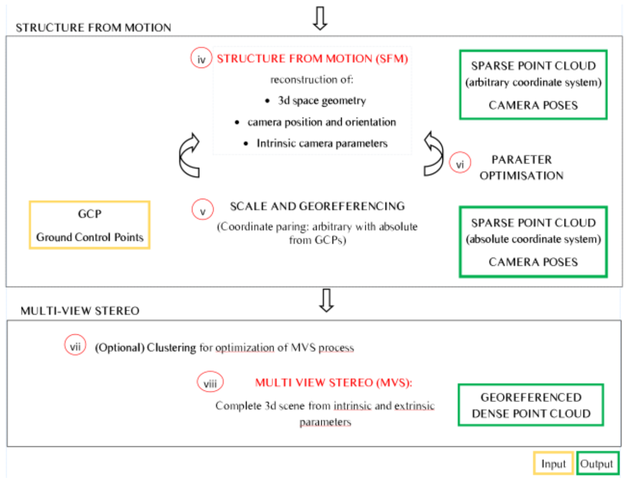




Figure 1. SFM MVS workflow for 3d reconstruction

Different photogrammetry softwares are available, both open source and commercial, that uses this process to reconstruct a 3d scene starting from uncalibrated images. It should be considered that while open source software generally offers the use of greater transparency, and the use of different algorithms can vary from different software packages, commercial software doesn't release information of algorithms used (that can be sometimes proprietary). Also, the continuous evolution of computer vision techniques, and development and refinement of algorithms in the different process steps suggest that further improvements will be implemented very quickly, reducing the memory consumption increasing elaboration speed and point cloud density and accuracy (Brook, 2017).

## 3. POINT CLOUD SEGMENTATION USING IMAGE ANALYSIS

### 3.1 Point cloud analysis

Currently, analytical tools, such as algorithm for acquisition and software for the generation and manipulation of the photogrammetric data are well developed by scientists. However, these methods are highly empirical and quantitative and qualitative data rely on analysts. Consequentially replication of the existing results can be difficult. For this reason, to ensure data quality and result's accuracy, researchers are focusing on standardization and protocol for information extractions and the integration in a complete process workflow.
The output of survey, made with Lidar or photogrammetric techniques, consists in a 3d point cloud of the captured environment with million or billion points. Despite the board availability of point cloud there is still a relevant need of an automatic method and integrated workflow for the extraction of the information needed to provide 3d data with meaningful attributes that characterize and provide significance to the objects represented in 3d. The added value of the entire process consists in fact in the capacity of automatic extracting relevant information.
Actually, *Point cloud analysis* is at its infancy, and today exists different techniques and algorithms to treat this type of data; in particular segmentation and classification of point cloud are very active research topics. *Segmentation* is the process of grouping point clouds into multiple homogeneous regions with similar properties (such as geometric, radiometric etc), while *classification* is the definition and assignments of point to specific classes, called "labels", according to different criteria. These two processes allow the extraction of relevant

information from the acquired data. Moreover, outlier elimination, spatial analysis and object simplification are other active research field.

It's important to consider that while the output of Lidar and photogrammetric acquisition is the same, the generated point cloud can be very different in terms of accuracy, resolution and information, due to the different nature of the acquisition technique. For Lidar point cloud the American Society For Photogrammetry and Remote Sensing proposed in "Las Specification" (Sensing, 2013) different standardized classes in which the objects in point cloud can divided. These classes are at the same time applicable to photogrammetric point cloud.

| Classification value | Meaning |
|---|---|
| 0 | Never classified |
| 1 | Unassigned |
| 2 | Ground |
| 3 | Low Vegetation |
| 4 | Medium Vegetation |
| 5 | High Vegetation |
| 6 | Building |
| 7 | Low Point |
| 8 | Reserved |
| | * |
| 9 | Water |
| 10 | Rail |
| 11 | Road Surface |
| 12 | Reserved |
| | * |
| 13 | Wire - Guard (Shield) |
| 14 | Wire - Conductor (Phase) |
| 15 | Transmission Tower |
| 16 | Wire-Structure Connector (Insulator) |
| 17 | Bridge Deck |
| 18 | High Noise |
| 19-63 | Reserved |
| 64-255 | User Definable |

Table 1 - ASPRS point cloud class

To extract relevant information from point cloud it's necessary to segment and classify the interested object inside the acquired scene. There are multiple research studies related to these two topics, driven by specific needs provided by the field of application (cultural heritage, building modelling, heritage documentation, robotics etc..). A non-exhaustive review of segmentation and classification methods is presented in (Ozdemir & Remondino, 2018).

**Segmentation** methods can be subdivided into five main classes (Nguyen & Le, 2013), according to the segmentation criteria (fig.3):

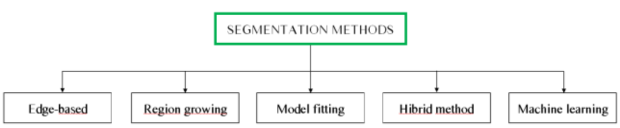



Figure 2. Point cloud segmentation methodologies

*Edge based* segmentation (Rabbani, van den Heuvel, & Vosselman, 2006) use algorithms for edge detection to outlines the borders of different region and to group points inside the boundaries to deliver final segments. *Region growing* methods start from one or more points featuring specific characteristics and then grow around neighbouring point with similar characteristics such as curvature, surface orientation etc.. (Rabbani et al., 2006). *Segmentation by model fitting* is based on the observation that many man-made objects can be decomposed into geometric primitives like planes, sphere and cylinders. Primitive shapes are then fitted onto point cloud data and the points that conform to the mathematical representation of the primitive shaper are labelled as one segment. In *Hybrid segmentation technique* more methods are combined to exploit the strength of a method and bypass the weakness of other methods. The accuracy depends on the methods used and on the target scene and objects. Finally, *Machine learning methods* are based on classification performed by machine learning algorithms (including deep learning, neural network (Zeiler & Fergus, 2014) etc). Machine learning is a specific discipline of computer vision that using Artificial Intelligence algorithms allows the computer to take decisions based on empirical and trained data. Learning that allows computer to take decision can be supervised (and reinforcement learning) or unsupervised: unsupervised learning is a class of problems in which one seeks to determine how the data are organized. In supervised learning, data are provided "labelled" to make the machine learn how to correctly perform a task. Dataset and features play a fairly important role in the use of machine learning, because these methods are based on the quantity and quality of the input data. Review of machine learning techniques applied to semantic segmentation of images and on 3d data is presented in (Garcia-Garcia, Orts-Escolano, Oprea, Villena-Martinez, & Garcia-Rodriguez, 2017). Actually, on of the more accurate and large dataset for 3d segmentation using Neural Network is 3d PointNet (Qi, Su, Mo, & Guibas, 2017), developed by Stanford University.

**Classification** of each segmented points can be achieved using three different approaches:

- *supervised approach*, where semantic categories are learned from a dataset of annotated data and the trained model is used to provide a semantic classification of the entire dataset.
- *Unsupervised approach* where data is automatically partitioned into segments based on a user-provided parametrization of the algorithms
- *Interactive approach* where user is actively involved in the segmentation/classification loop by guiding the extraction of segments through feedback.

Most of these segmentation algorithms are tailored to work with 2,5D surface model assumption, provided by Lidar-based survey, or analysis on 3d space.

### 3.2 Point cloud segmentation using image segmentation and machine learning

In this work, to perform 3d point cloud segmentation of specific object and scene, for aerial survey of infrastructure (such as bridges and viaducts), an approach based on 2d-image analysis using machine learning is presented. The elaborated procedure is based on the segmentation of the single images of the dataset, and on the transfer of the 2d segmentation on the 3d point cloud through masks. Image classification and segmentation techniques are in fact at a good point of development, algorithms are mature, and with a good training dataset it's possible to reach high precision (Lokanath, Kumar, & Keerthi, 2017). The technique used to highlight object and create an alpha mask on the image is *instance segmentation*. The object recognition can be splitted in to 4 different methodologies:

- **Image classification** is used to predict a set of labels to characterize the contents of an input image
- **Object detection** builds on image classification but allows to localize each object in an image. The image is now characterized by: 1. Bounding box (x,y coordiantes) for each box 2. An associated class label for each bounding box
- **Semantic segmentation** algorithms require to associate every pixel in an input image with a class label (including a class label for the background). While semantic segmentation algorithms are capable of labelling every object in an image, they cannot differentiate between two objects of the same class. This behaviour is especially problematic if two objects of the same class are partially occluding each other, we have no idea where the boundaries of one object ends and the next one begins.
- **Instance segmentation** algorithms compute a pixel-wise mask for every object in the image, even if the objects are of the same class label. The algorithm not only localized each individual object but predicted their boundaries as well.

Instance segmentation is well performed using Mask-RCNN (Convolutional Neural Network) (He, Gkioxari, Dollar, & Girshick, 2017) architecture that enable to segment complex objects and shaper from images and was built on previous object detection work of *R-CNN* (Girshick, Donahue, Darrell, & Malik, 2014) *Fast R-CNN* (Girshick, 2015) and *Faster R-CNN* (Ren, He, Girshick, & Sun, 2017) .

Mask-RCNN can be used to automatically segment and construct pixel-wise masks for every object in an image: in infrastructure survey the model is trained to perform structure recognition and split the different structural part.

The alpha masks, created with the instance segmentation for each image of the acquired dataset, are then used to filter and select the point of 3d point cloud. Classification is subsequentially made with an interactive approach to assign the selected point to different categories according to structural parts. The workflow for the image segmentation and the application to the 3d point cloud is presented in *fig.4*:

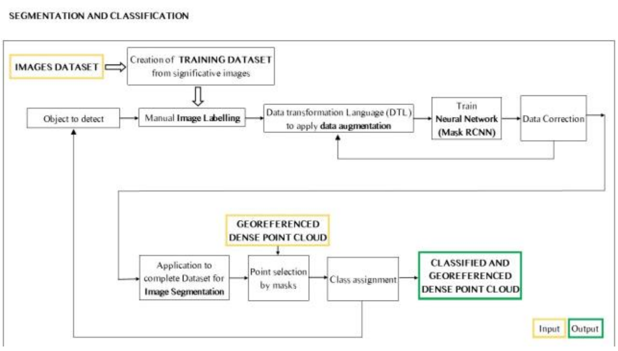



Figure 4. Workflow for image segmentation using Mask-RCNN

The procedure was applied using online services for machine learning Supervise.ly (Deep system, USA); from the starting dataset significant images where selected to obtain a train dataset in order to manually label single images. Subsequently the dataset is augmented to train the selected neural network (Mask RCNN). The neural network is then applied to the complete dataset to obtain masks for each selected element, such as bridge and piers as show in *fig.5*:

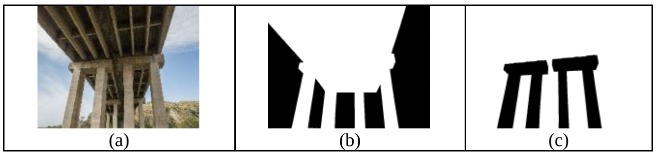



Figure 5. Mask R-CNN for bridge (b) and piers (c) segmentation

## 4. ANNUNZIATA VIADUCT CASE STUDY

The methodology was then applied to the aerial dataset, acquired with a UAV survey, of a simply supported viaduct (*fig.6*) located on A2 highway. Annunziata viaduct is located on "Autostrada del Mediterraneo" in the city of Reggio Calabria, Italy. The viaduct, built on 1970 upon the "Annunziata" river, is in a high seismic risk zone.

Figure 6. Annunziata Viaduct in A2 highway

The viaduct is made of pre-stressed reinforced concrete with 9 short-spans of 27 m, and a total length of 250 m (in curve). Curvature radius is 352 m and the medium height of the bridge is 25 m asl. The two decks (one per each direction) are sustained by a couple of piers with a common foundation.

The acquired dataset, composed by 1039 photos, was used to extract the geometrical feature of the structure to be used as input for the structural analysis. The photos of the structure (focusing on piers and deck) were captured according to photogrammetric principles with a low-cost UAV from different perspectives. Using Agisoft Metashape (Agisoft LLC, Russia) the photos were elaborated according to the described workflow, to obtain a 3d point cloud of the target scene. The application of SFM and MVS lead to the creation of the the sparse point cloud, composed by 70.000 point (*fig.7*):

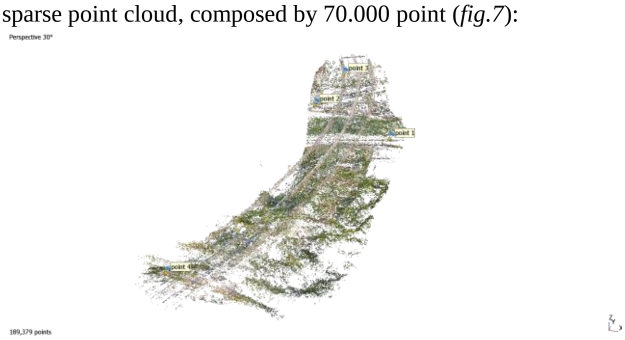



Figure 7. Annunziata Viaduct sparse point cloud

The dense point cloud after the application of the MVS stereo algorithm was reconstructed with 45 million points (*fig.8*):

Figure 8. Annunziata Viaduct dense point cloud

The automatic classification of the cloud using 2d masks, and the segmentation of the structural parts was used to extract the geometry of the Viaduct structural parts (deck, piers). On each photo the Mask RCNN algorithm is applied, to recognize target figure (*fig.8*). While the entire structure and the deck were recognized using segmentation algorithms, the piers were reconstructed by subtracting the two created masks.

Figure 9. Annunziata Viaduct from UAV Photo (a), and 2d masks of the bridge (b) and deck (c)

Using Supervise.ly online application the different masks to segment viaduct and deck were applied. The obtained alpha masks, with the use of Mask RCNN process, were transferred from 2d to 3d using "select by point" techniques. After the segmentation the class on the selected point were manually assigned, according to the different parts.

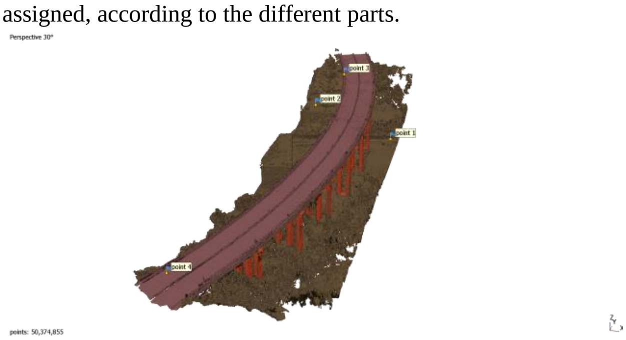


Figure 10. Annunziata Viaduct classified point cloud with deck (pink) and pillar (red)

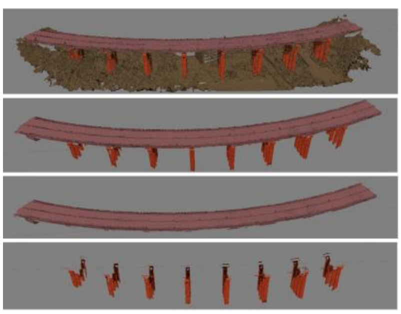


Figure 11. Annunziata Viaduct classified point cloud with deck and pillar

Finally, the obtained masks were used to segment the point cloud dataset into different structural parts, such as deck, pillar and soil, as presented in *fig.9*. The obtained point cloud, classified according to the structural parts, was used as input data, to perform risk assessment analysis. The structure geometry, separated according to the different structural parts, was used in the structural analysis, coupled with the information about age of the structure and year of construction. The main advantages in this case derives from the possibility to extract only the needed structural parts and in an automatic way, the geometrical feature, useful to perform a quick structural analysis to obtain the fragility curve of the structure.
For the evaluation of the fragility curve (*fig. 11*), it's possible to perform a structural project simulation according to the blueprint or the legislation corresponding to the age of the bridge, assuming standard mechanical characteristics.

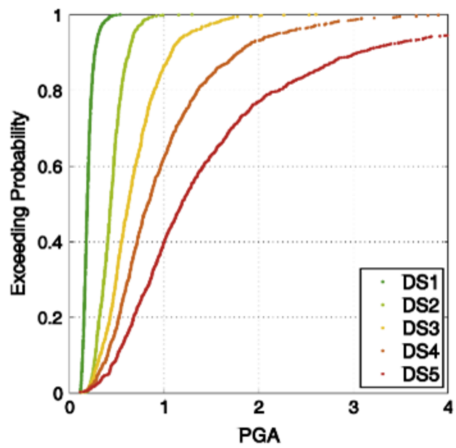



Figure 11. Fragility curve with PGA and exceeding probability

The structural analysis can be performed separately on deck and piers to evaluate:

- Piers section resistance
- Deck loads
- Interaction pier-deck and overall structure


https://doi.org/10.5194/isprs-archives-XLII-2-W11-187-2019 | 

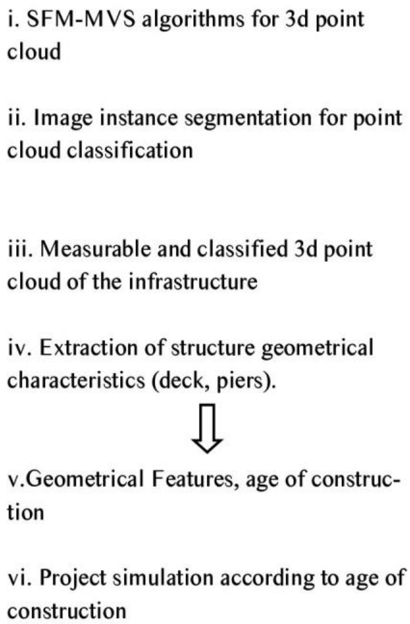



Figure 12. Workflow for structural analysis

From the fragility curve it's possible to link the failure probability related to the seismic loads in a certain location. The evaluation of the structural response can be used to estimate the level of security of the bridge, to schedule maintenance intervention, and to increase the security of the entire infrastructure.

## 5. CONCLUSIONS

In this paper a workflow and case study for point cloud segmentation using 2d image analysis is presented. Aerial survey through UAV to acquire dataset for photogrammetric 3d reconstruction are used today to perform fast and economic data gathering. The presented workflow to segment the 3d point cloud based on image segmentation can be applied to find and separate the structural parts of the point cloud to perform structural analysis for risk assessment and preservation of bridges and viaduct.

## REFERENCES

Brook, M. (2017). Structure from motion in the geosciences. New Zealand Geographer (Vol. 73). Wiley Blackwell.

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 92, 79–97.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A Review on Deep Learning Techniques Applied to Semantic Segmentation, 1–23.

Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter, 1440–1448.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 580–587.

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2017–Octob, 2980–2988.

Linder, W. (2016). Digital Photogrammetry (Fourth Edi). Springer.

Lokanath, M., Kumar, K. S., & Keerthi, E. S. (2017). Accurate object classification and detection by faster-RCNN. IOP Conference Series: Materials Science and Engineering, 263(5).

Lowe, D. G. (1999). Object recognition from local scale-invariant features. Proceedings of the Seventh IEEE International Conference on Computer Vision, 1150–1157 vol.2.

Lowe, D. G. (2004). Distinctive Image Features from. International Journal of Computer Vision, 60(2), 91–110.

Nex, F., & Remondino, F. (2014). UAV for 3D mapping applications: A review. Applied Geomatics.

Nguyen, A., & Le, B. (2013). 3D point cloud segmentation: A survey. IEEE Conference on Robotics, Automation and Mechatronics, RAM - Proceedings, (November), 225–230.

Ozdemir, E., & Remondino, F. (2018). Segmentation of 3D photogrammetric point cloud for 3D building modeling. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(4/W10), 135–142.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learning on point sets for 3D classification and segmentation. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017–Janua, 77–85.

Rabbani, T., van den Heuvel, F. a, & Vosselman, G. (2006). Segmentation of point clouds using smoothness constraint. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences - Commission V Symposium "Image Engineering and Vision Metrology," 36(5), 248–253.

Remondino, F., Barazzetti, L., Nex, F., Scaioni, M., & Sarazzi, D. (2012). Uav Photogrammetry for Mapping and 3D Modeling – Current Status and Future Perspectives. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVIII-1/, 25–31.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149.

Sensing, R. (2013). LIDAR format- American Society for Photogrammetry & Remote Sensing, (November 2011), 1–28.

Snavely, N., Seitz, S. M., & Szeliski, R. (2008). Modeling the world from Internet photo collections. International Journal of Computer Vision, 80(2), 189–210.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8689 LNCS(PART 1), 818–833.


https://doi.org/10.5194/isprs-archives-XLII-2-W11-187-2019  