19.3
FAST IMAGE LABELLING USING 3D RECONSTRUCTION

A. Carreaud[a], K. Lis[b], A. Gressin[a]

[a] *University of Applied Sciences Western Switzerland (HES-SO / HEIG-VD), Insit Institute - (antoine.carreaud, adrien.gressin)@heig-vd.ch*
[b] *Computer Vision Laboratory (CVLAB), Swiss Federal Institute of Technology (EPFL), krzysztof.lis@epfl.ch*

1. INTRODUCTION

Deep learning can solve many problems in computer vision such as image classification, objects detection and segmentation, image enhancement (colorization, super-resolution), but it requires large training dataset with high quality labels, which is a timeconsuming and expensive task. When training a neural network for objects detection, it is common to acquire the same object from numerous points of view. In this specific case, structure from motion algorithms can be used to obtain 3D representation of such objects (as seen in Figure 1) that are sometimes used when depth maps are needed, for example to train new neural networks which try to create a depth map from a single image. In this paper, we propose an innovative semi-automatic method to take advantage of such 3D reconstruction from multiple points of view in order to speed up manual labelling for object or instance segmentation, by labelling 3D data and project it on each image. Finally, our method is applied on a real project, for pipeline segmentation.

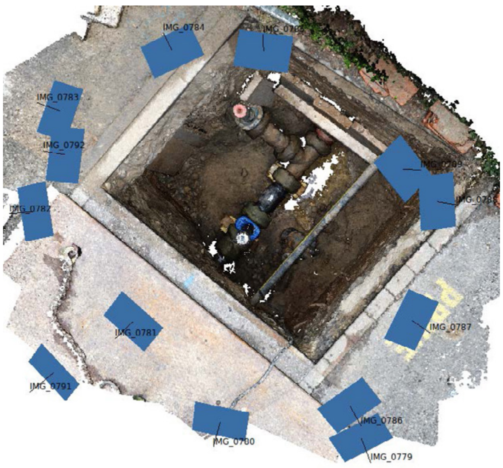

Figure 1: 3D reconstruction of pipelines objects from multiple points of view images.

2. STATE OF THE ART OF LABELLING

2.1 Image labelling

Image labelling is a laborious work consisting in manually drawing the shape of all the objects of interest visible on an image, this stage requires a lot of time in order to obtain accurate ground truth, which is a mandatory step before training a deep learning algorithm. Several papers propose methods to optimise/ accelerate this step, for example by working on improving the human-machine interface or by using image segmentation algorithms to automatically extract object envelopes (Wada, 2016, Chen et al., 2014, Gao et al., 2001). Otherwise, (Braun and Borrmann, 2019) propose to fuse photogrammetry 3D data and a BIM model to automatically label images. However, the use of the convex hull to group 3D point in image geometry does not allow to correctly deal with complex objects (like ”U” shape). The other limitation of their method is the fact that they do not separate objects instances.

2.2 Point cloud labelling

Research in various fields, for example autonomous cars, requires labelled 3D data, thus several labelled point cloud datasets have been created (SEMANTIC3D.NET, Stanford Large-Scale 3D Indoor Spaces Data Set, Paris-Lille-3D, etc.). This requires the optimisation of the time-consuming labelling step and various methods have appeared in recent years. Several softwares allow to manually label a point cloud with more or less sophisticated tools. One of the most efficient is Terrascan module from Terrasolid, but as licences are very expensive, open source solutions, such as CloudCompare, are a good alternative. Other automatic or semi-automatic methods exist, for example using labelled images or using active learning solutions. We decided here to use a simple manual solution by labelling the point clouds with Cloud- Compare.

3. METHOD

Our method is composed of three main steps. Firstly, a 3D reconstruction by photogrammetric processing provides localized and oriented images with a dense point cloud (3D data). This process have been fully automated using the Agisoft

Metashape Python API. Then, the point cloud is manually labelled by an operator in CloudCompare software. In order to save time, no instance separation is made manually on the 3D data. Instances will be computed automatically in the last step. Finally, 3D labelled data is cleverly projected on each image to obtain labelled images (in COCO format). This last stage is the core of our method (see Figure 2): (1) point cloud is sorted by classes; (2) DBSCAN algorithm is employed on each classes to create object instances; (3) depthmap (from the 3D reconstruction) are used to project only visible points on the image; (4) alpha-shape algorithm allows to create the hull around the objects; (5) finally labelled images and annotation files in COCO format are computed and exported. All processing have been automated in a Python script.

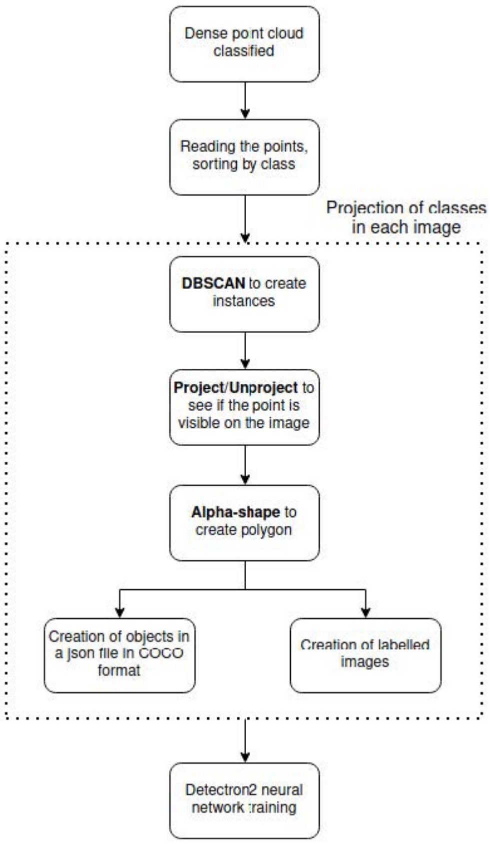

Figure 2: Processing chain of the projection step.

4. RESULTS AND DISCUSSIONS

The proposed labelling method have been compared to a fully manual method on a typical excavation site composed of 30 images with 20 objects, both in term of accuracy and time saving. The labelling accuracy is linked to the processing method but also to the accuracy of the operator who labels the point cloud. As a manual ground truth is needed to calculate an accuracy, evaluation of operator labelling is tedious, thus only the automatic process has been evaluated in this paper. The method used to estimate the labelling accuracy is Intersection over Union (IOU) which is widely used in machine learning (Hui, 2018). We found that there are unlabelled points on the images after our processing that would have been labelled by hand. When passing from the point cloud to the image, some areas that are not reconstructed in 3D (therefore non-existent in the point cloud) are not projected on the image. On the other hand objects can be very tedious to label by hand on 2D images, humans can easily take shortcuts where the machine will remain strict. As can be seen in Figure 3, the labelling of the 3D object is done by clicking 4 times in two different views (images 1 and 2), resulting on a precise labelling (image 3). Whereas in the 2D image the user should be careful to draw the envelope of the complex object. In this case, the accuracy of the proposed method will be better than that of the human.
Moreover, on this dataset, processing time has been reduced by a factor of 3 as seen in Table 1, allowing to save almost one hour with an acceptable loss of accuracy. As explained above the accuracy depends on the point cloud, if the point cloud is complete, the IOU value is close to 100 %, if there are missing areas in the point cloud, the lowest IOU was 95 %.

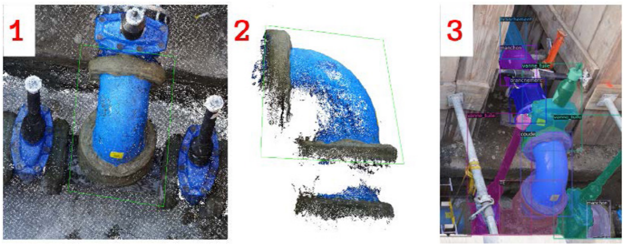

Figure 3: Manual 3D labelling process in only two steps (1+2) and result after projecting 3D labels on one image with our method (3).

Time per step	Manual	Semi-automated
Time for 3D reconstruction	X	5 min.
Time for point cloud labelling	X	25 min.
Time for reprojection process	X	5 min.
Time for labelling per image	5 min.	X
Total time (30 images)	90 min.	35 min.
Accuracy (IoU)	≈100 %	95-100 %

Table 1: Comparison of time and accuracy of the manual labelling method and our semi-automated method.

5. APPLICATION TO REAL PROJECT

The industrial services of Geneva (SIG) have approached the HEIG-VD to study the future of survey methods for updating the underground cadastre of the canton of Geneva. This concerns the possibilities of automation (in order to speed up fieldwork) as well as their impact on the profession of surveyor. This connection has given rise to a broader project: the ”Automatic Network Survey” project in partnership with the city of Lausanne and the EPFL. In particular, it seeks to find out whether it is possible to automatically recognise the objects making up the underground cadastre (water and gas networks, etc.) during excavation work, using automatic learning techniques (deep learning or artificial intelligence). After the acquisition phase, we have created an attribute table of 43 codes to identify the elements of the water network (valve, pipe, etc.). In order to train the automatic recognition algorithm we have chosen (Detectron2 (Wu et al., 2019)), we used the method presented above to label our images (20 excavation sites, 700 images), resulting on more than 5000 objects instances on 2D image. Manual labelling have been estimated to 50 hours (5 minutes per image, which is optimistic in many cases), when we successfully labelled all this dataset in less than 20 hours thanks to our method, saving about 60% of time.

REFERENCES

Braun, A. and Borrmann, A., 2019. Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning. Automation in Construction.

Chen, L. C., Fidler, S., Yuille, A. L. and Urtasun, R., 2014. Beat the MTurkers: Automatic image labeling from weak 3D supervision. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition pp. 3198–3205.

Gao, H., Siu, W. C. and Hou, C. H., 2001. Improved techniques for automatic image segmentation. IEEE Transactions on Circuits and Systems for Video Technology 11(12), pp. 1273–1280.

Hui, J., 2018. mAP (mean Average Precision) for Object Detection.

Wada, K., 2016. labelme: Image Polygonal Annotation with Python. https://github.com/wkentaro/labelme.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. and Girshick, R., 2019. Detectron2. https://github.com/facebookresearch/detectron2.