CLASSIFICATION OF RAILWAY ASSETS IN MOBILE MAPPING POINT CLOUDS

M. Corongiu[1], A. Masiero[1,2, *], G. Tucci[1]

[1] Dept. of Civil and Environmental Engineering, University of Florence,
via di Santa Marta 3, Florence 50139, Italy - (andrea.masiero, manuela.corongiu, grazia.tucci)@unifi.it
[2] Interdepartmental Research Center of Geomatics (CIRGEO), University of Padova,
Viale dell'Università 16, Legnaro (PD) 35020, Italy - masiero@dei.unipd.it

Commission I, WG I/7

KEY WORDS: Railway, Mobile Laser Scanning, Mobile Mapping, Deep Learning Classification, Point Cloud Segmentation

ABSTRACT:

Nowadays, mobile mapping systems are widely used to quickly collect reliable geospatial information of relatively large areas: thanks to such characteristics, the number of applications and fields exploiting their usage is continuously increasing. Among such possible applications, mobile mapping systems have been recently considered also by railway system managers to quickly produce and update a database of the geospatial features of such system, also called assets. Despite several vehicles, devices and acquisition methods can be considered for the data collection of the railway system, the predominant one is probably that based on the use of a mobile mapping system mounted on a train, which moves all along the railway tracks, enabling the 3D reproduction of the entire railway track area.
Given the large amount of data collected by such mobile mapping, automatic procedures have to be used to speed up the process of extracting the spatial information of interest, i.e. assets positions and characteristics.
This paper considers the problem of extracting such information for what concerns cantilever and portal masts, by exploiting a mixed approach. First, a set of candidate areas are extracted and pre-processed by considering certain of their geometric characteristics, mainly extracted by using eigenvalues of the covariance matrix of a point neighborhood. Then, a 3D modified Fisher vector-deep learning neural net is used to classify the candidates. Tests on such approach are conducted in two areas of the Italian railway system.

1. INTRODUCTION

Thanks to its safeness, railway transportation is still one of the most used public transportation methods for short to medium length travels, whereas long travels, such as intercontinental, are usually made by means of international flights.

Maintenance and monitoring of railway systems at national level is clearly a quite challenging task: rail lines are usually thousands of kilometers long, and several objects of interest are distributed all along such lines.

Despite several train companies operate in the Italian market, the physical Italian railway system is managed by the Italian Railway Network Enterprise (RFI), which is also in charge of the maintenance and monitoring operations. Given the complexity of the railway network and of the numerous objects to be monitored, RFI recently introduced the MUIF (Unique Model of the Physical Infrastructure), aiming at generating a georeferenced spatial digital representation of all the Italian railway system, to be used for easing the above-mentioned monitoring operations. MUIF is based on a geospatial database of several objects related to the railway system, such as switches, masts, cantilevers.

The acquisition of the geospatial data to be used to populate such geospatial database is clearly a not so easy task: given the huge rail line length and the complexity of the areas of interest, in particular on train stations, data acquisition is carried out by using different approaches and sensors:

---

* Corresponding author.

- airborne photogrammetry (GSD = 0.08 m);

- airborne LiDAR survey (point density $\geq$ 4 points/m^2);

- mobile mapping system mounted on a train;

- backpack mobile mapping system for high resolution geospatial data acquisition on train stations.

Furthermore, the data provided by other mobile systems can also be integrated in the system (e.g. with sensors on trolleys or other mobile devices (Tucci et al., 2018, Poiesi et al., 2017, Masiero et al., 2016)).

Such surveying techniques provide geospatial data, which are integrated in a Geospatial Information System (GIS), with different spatial resolutions and acquired in different time instants.

It is worth to notice that the information about the objects of interest, such as switches and masts, should be extracted by such huge amount of raw data, and properly inserted in the database as well, along with their 3D characteristics of interest, e.g. mast height, pylon thickness.

Hence, such characteristics of interest should be reflected on a proper semantic structure of the geospatial database attributes, and some tools should be developed for the automatic extraction of such semantic information from the raw data. The latter is the main goal of this work, whereas the geospatial database formulation and description was already considered previously (Corongiu et al., 2018).

More specifically, this work aims at developing and testing a procedure to automatically detect certain objects of interest (i.e. cantilever masts and portals) from point clouds provided by the train and backpack mobile mapping systems, and to extract the corresponding position and semantic information to be inserted in the above-mentioned database. Fig. 1 shows an example of point cloud acquired with a mobile mapping system mounted on a train.

Figure 1. Example of train mobile mapping point cloud acquired in proximity of a station.

The rest of the paper is organized as follows: Section 2 provides an overview of the proposed procedure and a short literature review of previous work related to this subject. Then Section 3 and 4 describe some pre-processing steps of the point cloud, and Section 5 presents the neural network-based classification procedure. Finally, discussion and some conclusions are drawn in Section 7 and 8.

2. PROCEDURE OVERVIEW

The problem of extracting certain objects and information from railway point clouds has already been considered in the literature by several authors: (Neubert et al., 2008) considered the detection of rail tracks, catenary and contact cables from orthophotos, whereas (Elberink, Khoshelham, 2015) extracted such information from points clouds by using template matching techniques (Arastounia, Oude Elberink, 2016). Similarly to the case considered in this work, most of the authors exploited mobile laser scanning systems (Pastucha, 2016) or on airborne LiDAR (Arastounia, 2017) for the data acquisition.

The strategy implemented in this work takes advantage of a pre-processing step, that aims at selecting a proper set of candidate objects of interest. Such candidates are then fed as inputs of a deep learning classifier, which determines the class of each object.

In particular, each candidate is selected based on the local property of the point cloud. Then, eigenvalue-based segmentation (Maalek et al., 2018, Weinmann et al., 2015) is used to discriminate between the different parts of the object, and, more specifically, to discard ground and vegetation point.

Convolutional neural networks (CNN, (Krizhevsky et al., 2012)) are used for the classification of objects from point clouds: this work exploits modified Fisher vectors for the object classification (e.g. masts) from point clouds (Ben-Shabat et al., 2018).

Fig. 2 shows certain of the objects to be detected from the mobile mapping point cloud. The extracted semantic information in the considered case studies is compared with that already stored and validated in the MUIF database (Tucci et al., 2019).
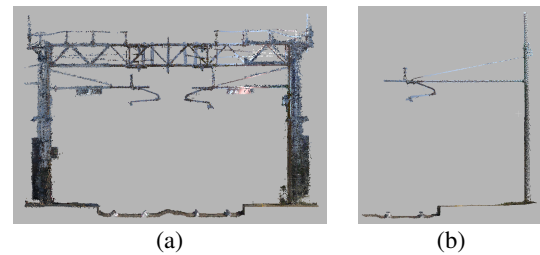

Figure 2. Examples of objects to be classified (rails are included in order to ease the readability of the figures on the top).

3. CANDIDATE EXTRACTION

This work considers in particular the problem detecting cantilever masts and mast portals. To such aim, a simple procedure is first implemented to extract a set of candidates to be tested. Such candidate set computation is mainly motivated by the need of reducing the computational burden related to the execution of the classification step via neural networks.

Since in this work the objects of interest are typically quite high, the candidate extraction procedure is mostly based on the identification of areas with high (planar) point density per square meter (Fig. 3(a)), and with quite high differences of altitudes between points in such areas.

The procedure described above is applied to a voxelized version of the point cloud. Then, connected components are computed and the centroid of each connected region is considered as a potential object of interest candidate (Fig. 3(b)).
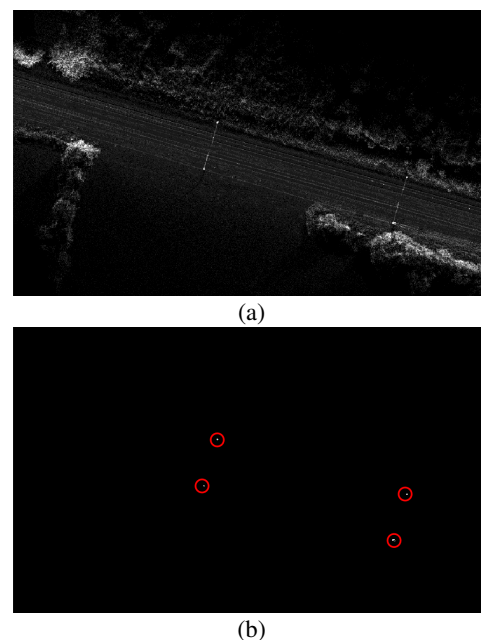

(a)

(b)

Figure 3. Example of candidate extraction: (a) planar point density (b) extracted candidates.

4. PRE-PROCESSING

Each object point cloud is pre-processed before being fed into the neural network classifier. The pre-processing workflow for each candidate is as follows:

- extract a subset of the point cloud corresponding to the neighbors of such candidate (e.g. a cylinder with 2 m radius centered in such location).

- change the local reference system of the subset in such a way to align the y coordinates with the railway track direction (Fig. 4).

- detect and discard in such subset ground, vegetation and other points not related to the objects of interest (Fig. 5, 6, 7).

The latter step takes advantage of the geometric feature detection methods based on the eigenvalues of the covariance matrix of the neighborhood of a 3D point (Maalek et al., 2018, Weinmann et al., 2015). Let λ_1, λ_2, λ_3 be the three eigenvalues of such covariance matrix, and let e_i, $i = 1, 2, 3$, be the normalized version of the eigenvalues, i.e. $e_i = \lambda_i / \sum_{j=1}^{3} \lambda_j$.

Then, similarly to (Weinmann et al., 2015), the following 3D features are used to summarize the point geometrical characteristics: linearity L_λ, planarity P_λ, scattering S_λ, omnivariance O_λ, anisotropy A_λ, change of curvature C_λ:

$$L_\lambda = \frac{e_1 - e_2}{e_1} \quad (1)$$

$$P_\lambda = \frac{e_2 - e_3}{e_1} \quad (2)$$

$$S_\lambda = \frac{e_3}{e_1} \quad (3)$$

$$O_\lambda = (e_1 e_2 e_3)^{1/3} \quad (4)$$

$$A_\lambda = \frac{e_1 - e_3}{e_1} \quad (5)$$

$$C_\lambda = \frac{e_3}{e_1 + e_2 + e_3} \quad (6)$$

Such 3D features are used to separate points useful for recognizing the objects of interest from vegetation, walls, and other objects which should be discarded. Wires and other mostly metallic objects (parts of masts and portals) are usually easily identified as linear features, whereas walls are well described by planar features. The full set of computed 3D features is used as input of a support vector machine (SVM) classifier, which aims at separating vegetation from the other objects. SVM classifier is trained on 9458 randomly sampled (pre-classified) points. Accuracy, i.e. (true positives+true negatives)/number of samples, on the training set was 93.5%, whereas, on a validation set of about 100 k samples accuracy was 93.2%.

5. CLASSIFICATION

The classification step is based on a deep learning approach, and, more specifically, on the adaptation to this case study of the 3D modified Fisher vectors (3DmFV) approach proposed in (Ben-Shabat et al., 2018). The notation used to describe the mathematical foundations of the method is chosen in analogy with that of (Sánchez et al., 2013) and (Ben-Shabat et al., 2018).

5.1 Fisher vectors

Let $X = \{x_1, \ldots, x_T\}$ be a set of T observations of a certain process, whose statistical behavior is assumed to be statistically described by a probability density u_λ, which depends on a set of parameters λ. To be more specific, u_λ describes the generative process of the observations.
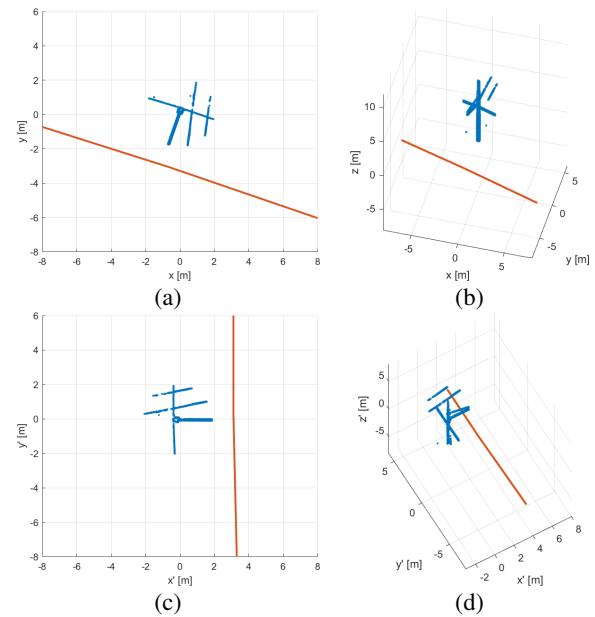

Figure 4. Example of point cloud pre-processing: mast (blue) and railway track (red line). (a) and (b) original data, (c) and (d) local reference system with y' coordinate aligned with the railway track.
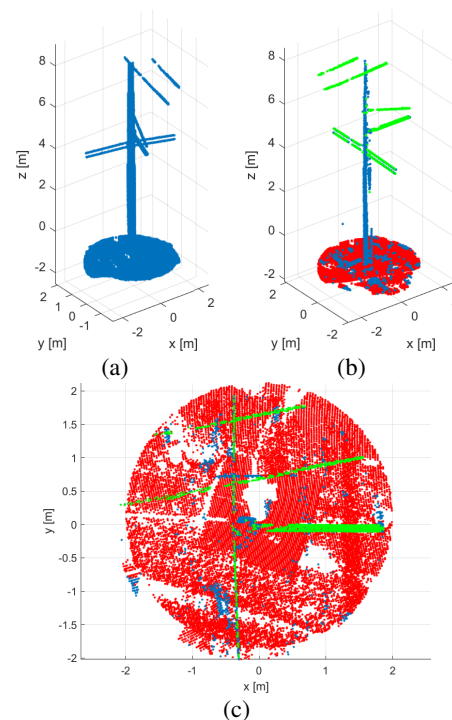

Figure 5. Example of point cloud pre-processing: (a) original data, (b) segmentation side view, mast (blue), cantilever and wires (green), ground (red), (c) segmentation re-oriented top view.

The contribution of each parameter to the generative process can be assessed by taking into account of the value of the partial derivative of $\log u_\lambda$ with respect to such parameter. Generalizing to the all parameter set, we can compute the gradient values $G_\lambda(X)$:
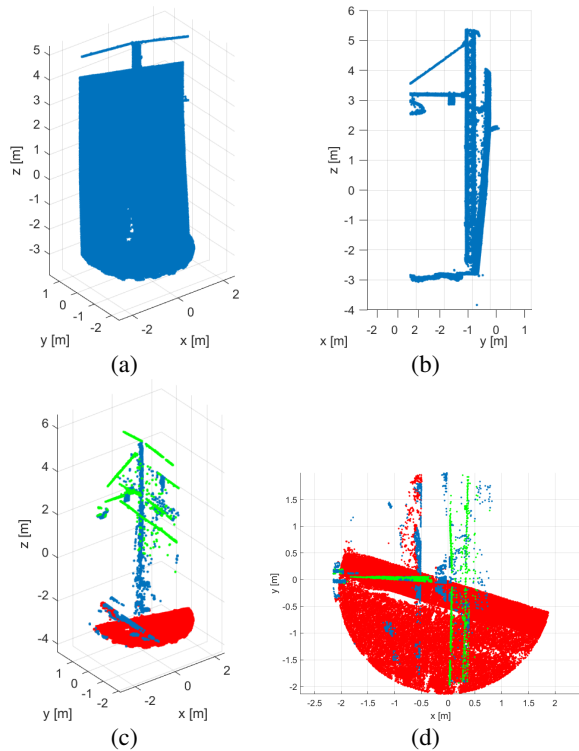
(a)   (b)

(c)   (d)

Figure 6. Example of point cloud pre-processing (discarding wall): (a) and (b) original data, (c) segmentation side view, mast (blue), cantilever and wires (green), ground (red), (d) segmentation re-oriented top view.
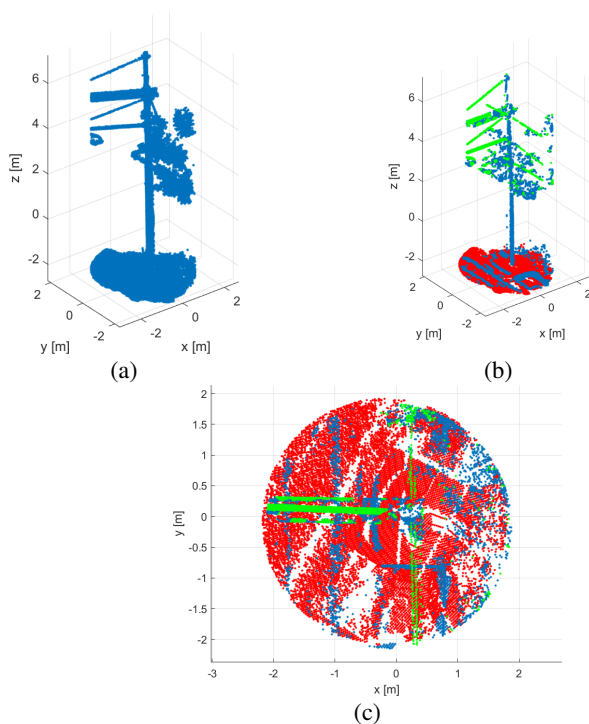

(a)   (b)

(c)

Figure 7. Example of point cloud pre-processing (vegetation elimination): (a) original data, (b) segmentation side view, mast (blue), cantilever and wires (green), ground (red), (c) segmentation re-oriented top view.

$$G_\lambda(X) = \nabla_\lambda \log u_\lambda(X) \tag{7}$$

Then, the similarity measure between two samples can be measured by considering the Fisher kernel, i.e. the inner product between gradient vectors computed as described above, weighted by the Fisher information matrix.

If the observations are independent, then $u_\lambda(X)$ can be factorized in the product of the density function for each of the observations, and consequently:

$$G_\lambda(X) = \sum_{t=1}^{T} \nabla_\lambda \log u_\lambda(x_t) \tag{8}$$

5.2 Association of Fisher vectors with Gaussian Mixture Models

Since a Gaussian Mixture Model (GMM) allows to approximate arbitrarily well any continuous distribution, a GMM seems to be well suited as generative density:

$$u_\lambda(x_t) = \sum_k w_k u_k(x_t) \tag{9}$$

where $u_k(\cdot)$ is the k-th Gaussian, whereas w_k is its weight. Mean and covariance matrix of the k-th Gaussian are μ_k and Σ_k, respectively.

Using the soft-max formalism (Sánchez et al., 2013), w_k can be substituted with α_k:

$$w_k = \frac{e_k^\alpha}{\sum_j e_j^\alpha} \tag{10}$$

Let $\gamma_t(k)$ be defined as

$$\gamma_t(k) = \frac{w_k u_k(x_t)}{\sum_j w_j u_j(x_t)} \tag{11}$$

Then, the gradients with respect to the parameters can be computed as follows

$$\nabla_{\alpha_k} \log u_\lambda(x_t) = \gamma_t(k) - w_k \tag{12}$$

$$\nabla_{\mu_k} \log u_\lambda(x_t) = \gamma_t(k) \frac{(x_t - \mu_k)}{\sigma_k^2} \tag{13}$$

$$\nabla_{\sigma_k} \log u_\lambda(x_t) = \gamma_t(k) \left[\frac{(x_t - \mu_k)^2}{\sigma_k^3} - \frac{1}{\sigma_k} \right] \tag{14}$$

where Σ_k was assumed to be diagonal, with the values on its diagonal equal to σ_k^2.

It is worth to notice that the soft assignment $\gamma_t(k)$ is usually sharply peaked, e.g. the t-th observation can be quite safely assigned to its closest Gaussian $u_k(\cdot)$.

If the above observation holds, then the Fisher information matrix becomes approximately diagonal (Sánchez et al., 2013), and

its effect can be summarized by a normalization of the Fisher vector as follows:

$$\mathcal{G}_k(X) = \frac{1}{\sqrt{w_k}} \left[\begin{array}{c} \sum_{t=1}^{T} \nabla_{\alpha_k} \log u_\lambda(x_t) \\ \sum_{t=1}^{T} \nabla_{\mu_k} \log u_\lambda(x_t) \\ \sum_{t=1}^{T} \nabla_{\sigma_k} \log u_\lambda(x_t) \end{array} \right] \quad (15)$$

where $\mathcal{G}_k(X)$ is the part of the Fisher vector related to the k-th Gaussian. The overall Fisher vector can be obtained concatenating all the $\{\mathcal{G}_k(X)\}$.

5.3 Modified Fisher Vector-based classifier

The assumption made in the previous subsections that u_λ is a generative model of the samples, allows to obtain a particularly effective classifier: actually, it can be shown that if for example the label is included as latent variable of the generative model, then asymptotically a Fisher vector-based classifiers reach the same level of accuracy as the maximum a posteriori (Jaakkola, Haussler, 1999).

However, the need of employing a model more compatible with the combination with neural nets leads to considering different alternatives. The choice of a different model with respect to a generative one, properly describing the real samples, is clearly suboptimal, however the combination with neural nets is supposed to compensate such sub-optimality.

In particular, similarly to (Ben-Shabat et al., 2018), in this work the Gaussian are assumed to be homogeneously distributed on the input domain, with each Gaussian located on a vertex of an equally spaced grid. The structure of the neural net resulting from the approach in (Ben-Shabat et al., 2018) is roughly similar to PointNet's one (Qi et al., 2017), but adapted to work on Fisher vectors.

In addition to the variables considered in (Ben-Shabat et al., 2018), this work modeled with a GMM also the point distribution on the vertical and horizontal direction (where the latter is determined as orthogonal to the railway track) close to the determined candidate locations.

6. RESULTS

Training was performed considering several thousands of cantilever masts, portals and other objects taken from the MUIF database of the province of Venice. It is worth to notice that only those objects with at least 700 points were considered.

The developed approach was tested on a railway test area, approximately 1 km long, in the province of Venice (Italy), corresponding to about 260 million points, not included in the training dataset. A total of 470 objects were included in such area (but only 4 portals).

The obtained classification results are reported in Table 1.

Fig. 8 shows two examples of classification errors, in particular two cantilever masts classified as portal and as another object.

Fig. 9 shows an example of false positive classification as cantilever mast.

Finally, Fig. 10 shows the distribution of the position error of the classified cantilever masts with respect to the positions stored in the MUIF database, which are used as reference data here.

Table 1. Classification results

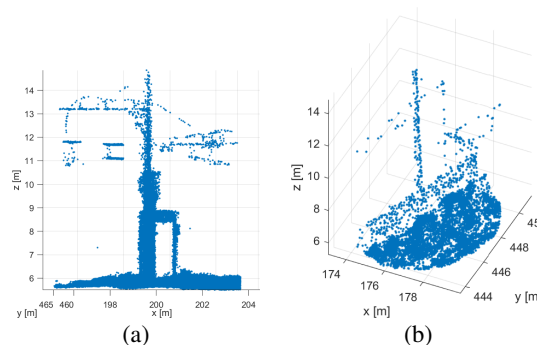| Real class	Classification results		
	Portal	Cantilever mast	Other
Portal	100%	0	0
Cantilever mast	3.0%	86.2%	10.8%
Other	0%	2.5%	97.5%

Figure 8. Examples of classification errors: objects classified as (a) portal, (b) "other", instead of cantilever mast.
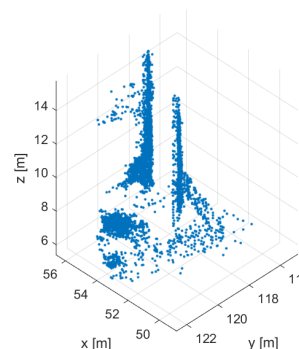

Figure 9. Example of classification error: object classified in the cantilever mast class instead of "other".
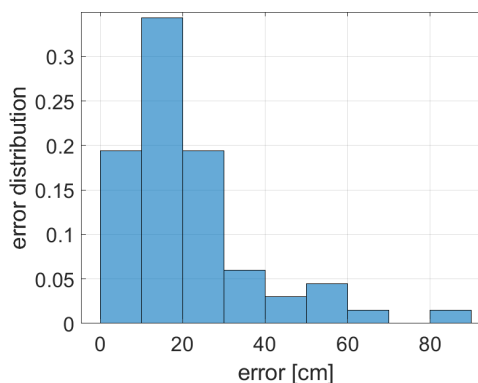

Figure 10. Distribution of the position error of the classified cantilever masts.

7. DISCUSSION

The main motivation of the proposed approach is that of introducing a kernel representation in order to avoid issues with voxelization and point ordering and reduce the sensitivity of the classification results with respect to possible variations on the input, such as different point densities, obstructions. The intro-

duction of the kernel representation and of the Fisher feature vector can allow to obtain with a linear classifier similar performance to those obtained with a nonlinear one (Sánchez et al., 2013).

The candidate extraction steps proved to be quite effective in the considered case study areas: areas related to all the assets have been properly extracted by the implemented procedure, along with a quite large set of extra candidates, not really related to the assets of interest. Despite the presence of such extra candidates, this step proved to be relatively fast, and hence fairly effective in reducing the overall computational burden, i.e. notably reducing the number of points to be examined by the neural net.

The pre-processing step also allowed to obtain quite good results in terms of reducing the number of outlier points, not related to the object of interest, i.e. vegetation, walls.

The classification performance obtained with the proposed approach was quite acceptable, in particular in the portal case (however, despite portals were quite easily identified in this dataset, a much larger number of samples should be considered for a statistically more reliable result).

Several cantilever masts were classified as "other" objects. This was probably mostly due to the presence of certain cantilevers described by really few points in the mobile mapping 3D reconstruction (Fig. 8(b)). Furthermore, cantilever masts can also be classified as portals in certain cases (Fig. 8(a)).

Furthermore, in certain cases objects with a similar appearance to a cantilever mast were classified in such class, as for instance in the case of Fig. 9.

Despite the obtained classification errors might be acceptable taking into account of only the information provided by a local subset of the overall cloud, the addition of information provided by the context should allow to reduce the rate of such errors. According to this observation, the introduction of context information in the classification procedure will be considered in our future work. Furthermore, the extension to the classification of other railway assets will be considered as well, along with a more depth analysis of the influence of the point density on the classification results. Finally, an extension to the analysis of buildings and structures close to the railway will also be considered (Park et al., 2007, Chen, 2012, Bitelli et al., 2004, Guarnieri et al., 2015, Masiero et al., 2015, Boreggio et al., 2018).

To conclude, the position of the detected objects was quite well estimated with respect to the reference one, with an error usually smaller than 30 cm, as shown in Fig. 10.

8. CONCLUSIONS

This paper presented the current state of development of an automatic approach for the extraction of information about railway assets from large point clouds collected by a mobile mapping system mounted on a train.

Such approach aims at reducing the need of human interaction needed during geospatial information extraction, potentially also speeding up the overall process. The following workflow is implemented in the considered approach: determine a set of candidate asset areas, pre-process the subset of points in

such areas and finally feed them as input to a 3DmFV neural net classifier.

The obtained results show an acceptable performance in the classification of the considered objects, however the proposed procedure shall be extended to other assets in our future work.

ACKNOWLEDGEMENTS

REFERENCES

Arastounia, M., 2017. An enhanced algorithm for concurrent recognition of rail tracks and power cables from terrestrial and airborne LiDAR point clouds. *Infrastructures*, 2(2), 8.

Arastounia, M., Oude Elberink, S., 2016. Application of template matching for improving classification of urban railroad point clouds. *Sensors*, 16(12), 2112.

Ben-Shabat, Y., Lindenbaum, M., Fischer, A., 2018. 3DmFV: Three-dimensional point cloud classification in real-time using convolutional neural networks. *IEEE Robotics and Automation Letters*, 3(4), 3145–3152.

Bitelli, G., Dubbini, M., Zanutta, A., 2004. Terrestrial laser scanning and digital photogrammetry techniques to monitor landslide bodies. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(B5), 246–251.

Boreggio, M., Bernard, M., Gregoretti, C., 2018. Evaluating the differences of gridding techniques for Digital Elevation Models generation and their influence on the modeling of stony debris flows routing: A case study from Rovina di Cancia basin (North-eastern Italian Alps). *Frontiers in Earth Science*, 6, 89.

Chen, S.-E., 2012. Laser scanning technology for bridge monitoring. *Laser Scanner Technology*, 71.

Corongiu, M., Tucci, G., Santoro, E., Kourounioti, O., 2018. Data integration of different domains in geo-information management: a railway infrastructure case study. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(4).

Elberink, S. O., Khoshelham, K., 2015. Automatic extraction of railroad centerlines from mobile laser scanning data. *Remote sensing*, 7(5), 5565.

Guarnieri, A., Masiero, A., Vettore, A., Pirotti, F., 2015. Evaluation of the dynamic processes of a landslide with laser scanners and Bayesian methods. *Geomatics, Natural Hazards and Risk*, 6(5-7), 614–634.

Jaakkola, T., Haussler, D., 1999. Exploiting generative models in discriminative classifiers. *Advances in neural information processing systems*, 487–493.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097–1105.

Maalek, R., Lichti, D. D., Ruwanpura, J. Y., 2018. Robust segmentation of planar and linear features of terrestrial laser scanner point clouds acquired from construction sites. *Sensors*, 18(3), 819.

Masiero, A., Fissore, F., Pirotti, F., Guarnieri, A., Vettore, A., 2016. Toward the use of smartphones for mobile mapping. *Geo-Spatial Information Science*, 19(3), 210–221.

Masiero, A., Guarnieri, A., Pirotti, F., Vettore, A., 2015. Semi-Automated Detection of Surface Degradation on Bridges Based on a Level Set Method. *ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3), 15–21.

Neubert, M., Hecht, R., Gedrange, C., Trommler, M., Herold, H., Krüger, T., Brimmer, F., 2008. Extraction of railroad objects from very high resolution helicopter-borne LiDAR and ortho-image data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38, 25–30.

Park, H. S., Lee, H., Adeli, H., Lee, I., 2007. A new approach for health monitoring of structures: terrestrial laser scanning. *Computer-Aided Civil and Infrastructure Engineering*, 22(1), 19–30.

Pastucha, E., 2016. Catenary system detection, localization and classification using mobile scanning data. *Remote Sensing*, 8(10), 801.

Poiesi, F., Locher, A., Chippendale, P., Nocerino, E., Remondino, F., Van Gool, L., 2017. Cloud-based collaborative 3d reconstruction using smartphones. *Proceedings of the 14th European Conference on Visual Media Production (CVMP 2017)*, ACM, 1.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.

Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J., 2013. Image classification with the Fisher vector: Theory and practice. *International journal of computer vision*, 105(3), 222–245.

Tucci, G., Corongiu, M., Flamigni, F., Comparini, A., Panighini, F., Parisi, E. I., Arcidiaco, L., 2019. The validation process of a 3D multisource/multiresolution model for railway infrastructures. *Applied Geomatics*, 1–16.

Tucci, G., Visintini, D., Bonora, V., Parisi, E. I., 2018. Examination of Indoor Mobile Mapping Systems in a Diversified Internal/External Test Field. *Applied Sciences*, 8(3), 401.

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 286–304.