2023 European Conference on Computing in Construction
40th International CIB W78 Conference
Heraklion, Crete, Greece
July 10-12, 2023

THE POTENTIAL FOR CREATING A GEOMETRIC DIGITAL TWIN OF ROAD SURFACES USING PHOTOGRAMMETRY AND COMPUTER VISION

Jialei Ding[1,2] and Ioannis Brilakis[1]
[1]The University of Cambridge, Cambridge, UK
[2]Loclab Consulting, Darmstadt, Germany

Abstract

Current methods to create 3D models of roads are not scalable. Advances in photogrammetry mean that they have emerged as a realistic alternative to LiDAR for producing 3D spatial data. As photogrammetry is based on image data, methods developed in the area of computer vision can be utilised to segment or find relationships between assets in the point cloud. This could in turn be leveraged to generate 3D models. This paper presents the various advances in computer vision that can be applied to generate 3D models of the road surface, which could be used as a basis for creating a digital twin of a road.

Introduction

This paper is about how advances in the field of computer vision (CV) can be used to create Road Geometric Digital Twins (RGDT)s of road surfaces using Photogrammetry. In industry and research, the consensus is that a DT is a digital replica of a physical asset that accurately reflects the as-is status of the physical asset (Ariyachandra, 2021). The road surface forms the basis for modelling other assets on the road. This is therefore the asset that will be the focus of the rest of this paper. This paper focuses on models of road surfaces that do not have any existing 3D models. Ancillaries to the road, such as street furniture are therefore not addressed. Photogrammetry is the science and technology of deriving 3D spatial data from a set of 2D image data. Computer vision is the field within computer science that concerns itself with teaching machines to “see”.

Information on asset condition and intuitive access to this data forms the basis for decision making relating to the maintenance and operations of roads (Flintsch & McGhee, 2009). However, it is common for this data to be outdated, in silos or stored in legacy databases. This makes them difficult to access, particularly in scenarios where interdisciplinary collaboration is required (Shah, et al., 2017). This directly impacts the ability to take a responsive and coherent approach to planning.

Development of a DT can allow the coordination of complex processes using a single digital resource (Jiao, et al., 2013; Khaddaj & Srour, 2016; Lindkvist, 2015). It allows for a database that can be intuitively and visually accessed. A key component of a DT is its 3D geometry. It may serve as the interface between the data and the user. However, current industry practices to generate 3D representations of infrastructure assets are either too costly to be scalable or too low in detail to be used in common use cases relating to the built environment.

Using visual data of the road to generate 3D models is currently an underutilized technique in the built environment industry. Sophisticated algorithms and Convolutional Neural Networks (CNNs) could be used to identify assets, their shapes, and their positions in a road scene.

The next section outlines current modelling methods in industry and compares them against a set of user requirements. Areas that may benefit from the use of CV and photogrammetry are highlighted. An overview of the relevant areas of CV and photogrammetry and how they have been used in previous research to create digital representations or road surfaces is given. Gaps in using CV for creating 3D models are then derived and summarised. Lastly, further research areas to advance this area are suggested.

Background

A DT can be used in a wide variety of applications within the Built Environment sector, such as visualising proposals, maintenance planning and use in simulations. Each use case will have unique requirements. The overarching requirements for a DT should allow for these requirements to be met. Broadly, these requirements can be listed as follows:

The model must…

- …reflect the as-is condition of the road.
- …display road assets at an appropriate level of detail for a given use case.
- …be object-based and able to store attribute information.
- …be data light.
- …be capable of accurately depicting larger stretches of road.
- …be affordable to develop and make.

The first type of digital representation considered is large-scale 3D maps such as Google Maps’ 3D view (Alphabet). The 3D model is a mesh created using photogrammetry. The mesh is then overlain with images used in the photogrammetry process. These maps are generated at a massive scale (Farr, et al., 2007). However, the level of detail of the model is low. The models are unsegmented meshes, meaning they are not object-based and cannot be used to store asset information. Large-scale 3D models are an excellent resource for high-level planning but fall short in applications that require more detailed asset information.

Geospatial databases are another common form of digital representation used for roads. GIS programs, such as ArcGIS (Esri) and QGIS (open source), link data to a unique object that is graphically displayed. A wide variety of data formats, both open standards and proprietary, are supported. This includes object-based file formats. The majority of GIS is in 2D. In these cases, the nuances of a 3D representation are lost (USC Spatial Sciences Institute, 2021). 3D GIS has also been developed and is in use in industry. Some applications embed 3D models of specific assets as a data layer. Other applications can show entire city scenes in a 3D view. However, these models are very low in detail and do not extend to road surfaces. While 3D GIS attempts to portray asset information in a 3D space, it does not do so for roads at the required level of granularity.

Using Mobile Laser Scanners (MLS) is a common method in industry to create 3D representations. An MLS can collect a Point Cloud (PC), which can be combined with other data sources, such as images, to create a highly detailed and accurate 3D representation of a road scene. Using an MLS has several advantages. The MLS can capture volumetric data that may not be visible on an RGB image. It is also robust against weather effects and occlusions (Farhadmanesh, et al., 2021). Companies that work within the infrastructure digitisation domain, such as Viatech (NO) and Trimble (US), provide commercial services for this process. However, these models are unsegmented and not object-based. It follows that no further information or attributes are stored in the model. Another issue of this type of system is the large file size and expensive data gathering process. Pre-processing and "cleaning" the PC can significantly reduce the data density and file size. However, there is a limit to how much the file size can be reduced without losing essential detail (Chen et.al, 2019; Tan & Li, 2019; Matsumoto et.al., 2019). While these types of models offer high levels of detail and accuracy, a more data-light and cost-effective approach is required.

Creating models manually is still common within practice. This is especially true in situations where the required model needs to be highly detailed and object-based. Models are created using a combination of Google Maps, GIS, old technical drawings, or design models as references. Additional data, such as PCs and, high-resolution images can be collected from the site to further enrich the model. The final models are tailored to fulfil the EURs of a given project (DEGES, 2022). These models are expensive and labour intensive to produce. Due to this, the frequency of their use is limited.

Powerful 3D rendering engines and modelling tools developed for the games industry are being utilised for applications in the built environment sector. These models are also object-based, and a wide variety of file formats are supported. An increasing number of software vendors in the built environment industry are supporting the export of BIM files into game engines for visualisation purposes (Mathworks, 2022; Witteveen+Bos, 2022; Loclab, 2022). However, game engine-based 3D models are labour intensive to produce if using traditional approaches.

Efforts have been made to streamline the modelling process. One example of this is the use of an object library to represent common assets, such as a standard stop sign. This removes the need to manually model these assets. Each instance of the 3D model is an object and can be enriched with attributes or other data required by the end user. These models rely on the object library, which must be curated and created manually. Many assets, however, are not suited to this type of modelling and still require manual modelling. This is slow and expensive to do. Examples include large, complex assets that may be custom made for a project. The road surface itself is an example of this. The width of the road changes, lanes branch off and on and there are often irregularities such as bus stops or chicanes present.

Levers to reduce costs

Many attributes of a company affect the viability of using RGDTs. This includes budget prioritisations, information silos within a company, management decisions and the cost of generating the RGDT itself. The rest of this paper focuses on the last aspect of this list. The specialised equipment needed, the large size of data and the high number of skilled labour hours required means that it is often uneconomical to create an RGDT. Current methods for modelling a road surface are not practical for large-scale use in industry. A scalable solution is therefore required.

Three main levers are considered to reduce this cost: The quality of the model, the cost of data collection and equipment and the cost of labour. Quality should be fixed as they are set by the user requirements. The cost of data collection alone often prohibits the creation of an RGDT. Reducing this high initial cost would lower the barrier to entry. Photos, videos, and their derived data are already being effectively used in industry as a cost-effective alternative for laser-scanned PCs. The largest cost associated with creating an RGDT is the labour cost. This can be achieved by increasing the level of automation in creating an RGDT.

One can leverage the image or video data to help increase automation. As PPCs are derived from images, it is possible to use the information found about the images to enhance the segmentation of the PPC. This has the potential to significantly improve segmentation results and computational efficiency. The next section will therefore explore the state of research in automating the modelling of on-road assets using photo, video, and their derived data, Photogrammetric Point Clouds, (PPCs).

Photogrammetric point clouds

In addition to photogrammetry, there exists a wide variety of related technologies derived from it. The most common of these are Structure from Motion, Simultaneous Localisation and Mapping, and Stereovision. As photogrammetry is the underlying mechanism for these methods (Fraser, 2018), the overarching technique of photogrammetry will be referred to.

The basic principle of photogrammetry is triangulation. All features in a set of images are matched to their corresponding positions in the other images (Jain et.al., 2017). Feature locations in space are reconstructed using two or more converging rays projected from the images. Figure 1 shows a schematic of how 2D coordinates in images relate to a 3D coordinate in space. A dense PPC can be reconstructed using this method (Forsyth & Ponce, 2012). These reconstructed PPCs contain both spatial and colour information. However, they do not contain semantic information that distinguishes between objects or asset categories (Chen et.al., 2019).

Previously, generating a PPC used to require large amounts of manual intervention (Brilakis et.al., 2011) to be a viable alternative to LiDAR. Improved photogrammetry techniques and the lower equipment costs compared to LiDAR have made photogrammetry a more attractive choice than before (Farhadmanesh et.al., 2021).

PPCs are, like LiDAR PCs, a spatial dataset. They can therefore benefit from the rich body of work done in the field of PC segmentation (Brilakis et.al., 2011). Some studies have applied 3D segmentation techniques developed for LiDAR on PPCs (Chen et.al, 2019). For instance, Tan & Li (2019) were able to find the geometry of the road surface and kerbs within 1cm of accuracy using region growing in a PPC. However, the limitation of PCs is still present. These include large file sizes and high computational requirements. Additionally, directly applying 3D PC segmentation techniques does not take advantage of the other information available in the images used to generate the PPC.

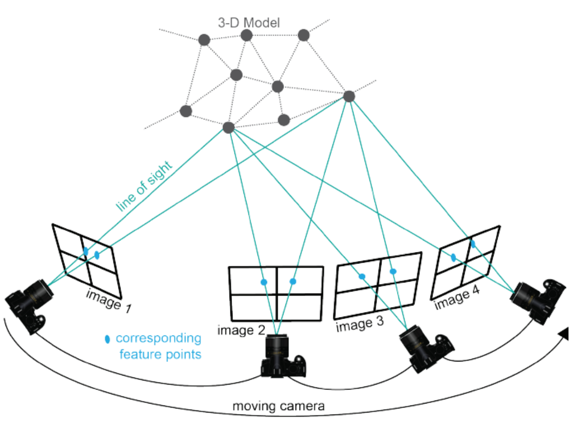

Figure 1: Schematic of 2D correspondences and 3D rays. The intersection point between rays is the point of the feature in 3D. (Shervais, 2016)

Object detection and image segmentation

In the context of a road scene, a method that can detect several objects in an image is required. Two fields of computer vision focus on identifying the placement and asset type of several objects in an image. These are Object Detection and Image Segmentation.

Object detection concerns itself with identifying bounding boxes around multiple objects and assigning each individual object or set of objects a label. Image segmentation concerns itself with partitioning an image into multiple segments or regions. The main difference between image segmentation and object detection is that labels are assigned to every pixel instead of a bounding box. This means that the shape of an object is also recognised.

Over the years, CNN-based image segmentation has phased out traditional hand-crafted methods due to much better performance (Suleymanov and Kunze, 2018; Liang et al., 2020). CNNs are a class of Deep Learning (DL) algorithms that takes an image as input. Each input image is passed through a series of convolution layers with trained weights. The final layer outputs the assigned labels of the image.

CNNs are an active area of research. Further features have been added to CNN architectures to improve performance for segmenting roads. Some issues that still need to be addressed include the lack of generalisation, large computational loads in training (Tang et.al., 2021; Hou et.al., 2019), acquiring sufficient training data (Genova et.al, 2021) and occlusion in images (Wang et.al., 2019). It is important to keep these limitations in mind when utilising CNNs and image segmentation.

Some studies have projected LiDAR PCs and PPCs into a 2D image to take advantage of segmentation technologies. These are called raster-based models. In raster-based methods, a point's intensity and range data are converted into a raster format, creating a 2D image.

CV algorithms or CNNs are then used to segment and label the points, as done in Jung et.al. (2019) and Li et.al.

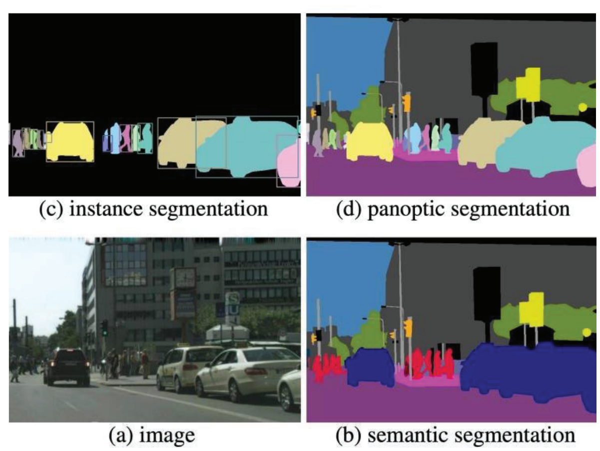

Figure 2: Differences between the three main types of image segmentation. (Kirillov et.al., 2019)

(2021). This has several advantages such as lowering the computational demand of the segmentation task (Suleymanov et.al, 2019), compared to directly segmenting PCs. However, raster-based models must be applied to a PC. A PPC must therefore first be generated using image data. After the PPC is created, it would then have to be projected back into a 2D image for segmentation. This is inefficient.

A more computationally efficient way to segment a PPC is to utilise pixel labels in pre-segmented images to label each 3D point (Zhang et.al. 2019; Li et.al., 2019; Golpavar-Fard et.al., 2015; Vineet et.al., 2015). Every point in the PPC has correspondences over one or more images. These labels are migrated from the 2D images to their corresponding points in the PPC. In some cases, a point may correspond to several labels. The final label can be selected through various methods. These include a simple voting system (Golpavar-Fard et.al., 2015), ray casting labels from images to the 3D model (Vineet et.al., 2015), and probabilistic methods, such as Bayesian progressive label migration (Zhang et.al., 2019). In some cases, the PPCs have been converted into grid voxel maps (Li et.al., 2018; Yang et.al., 2017; Kundu et.al., 2014) for further computational savings.

Fusing methods using image segmentation with pure PC segmentation methods has been attempted. Zhong et.al. (2017) labelled a PPC after acquiring semantic labels for the images using a CNN. This achieved a mean IoU of 67.56% and point-wise accuracy of 80.13%. In this work, the LiDAR PC was also collected and segmented. The segmented PC and PPC were then fused together to improve the segmentation results. However, it was found that the improvement from fusing these two PPCs was negligible (<1%), for all evaluation parameters, compared to directly segmenting through a voting process.

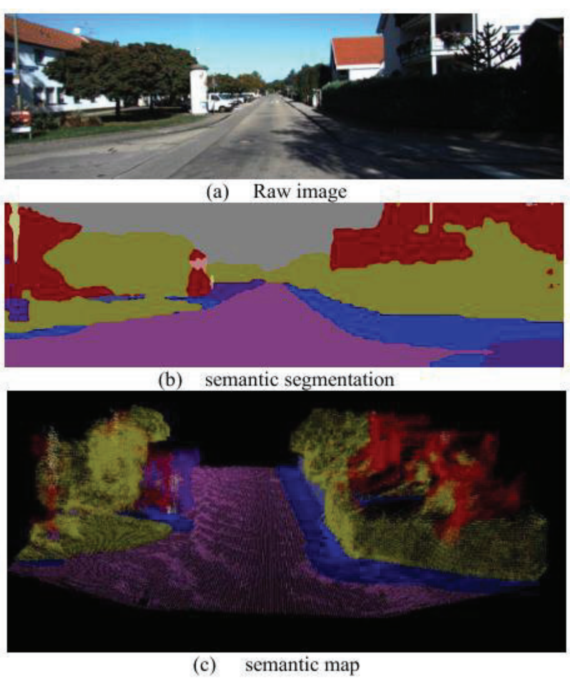

Figure 3: Labelling of PPC. Labelling a PC using a segmented image means that one doe does not need to perform 3D segmentation. (Cheng et.al., 2019)

Incorporating road asset relationships

One feature of roads is that they are inherently structured and hierarchical in nature. Assets are placed in predictable locations relative to each other. It is possible to use this prior knowledge to improve the detection and segmentation of these assets.

One attempted method is to split a PC into specific regions of interest (ROIs) and only search for certain assets within this ROI (Li et. al., 2021). However, an asset may be present in an ROI it has not been assigned to, meaning that it will not be recognised. This may explain why the median IoU score across all asset classes was 71.34% in Li et.al. (2021), which is comparable with methods that do not first find ROIs.

Prior knowledge can be used to fill in occluded parts of an image. For instance, Suleymanov et.al. (2019) used the assumption of a kerb being continuous to fill in missing kerb sections in a partially occluded scene. The described method had robust results, with an F1 score of 0.97 for a test dataset that combined both occluded and non-occluded scenes. However, this assumption is often not true. For instance, there may be discontinuities in a kerb due to a traffic crossing etc. In these cases, it would be wrong to assign those spaces as being a kerb.

The two hand-crafted methods described above hold assumptions that may not generalise to many configurations. While they perform well under certain scenarios, there is a need for more flexibility in describing road asset relationships.

Conditional Random Fields (CRFs) are a type of probabilistic graph model that takes neighbour prediction context into account in the current prediction. CRFs can be jointly implemented with the segmentation process. (Vineet et.al., 2015; Kundu et.al., 2014). However, these methods have been phased out in favour of CNN-based approaches. In more recent works, CRFs are used at the end of CNN segmentation pipelines to enforce semantic (Wang et.al., 2019; Yang et.al., 2017) and spatial (Li et.al., 2019; Zhang et.al., 2019) consistency on a pre-segmented image or model. Methods that incorporate a CRF as a final process usually see a 2-3% improvement in the average IoU of the segmented model.

Other methods have utilised the hierarchical nature of roads more explicitly. These models are based on the idea that an object can be a collection of parts and rely on object detection, rather than image segmentation. Bounding boxes for each individual part can be found and identified. These can be combined into larger bounding boxes representing the whole object (Davies, 2022).

Latent Hierarchical Parts Based (LHPB) models are an extension of these models specifically developed to interpret road scenes (Venkateshkumar et.al., 2015). They provide a framework for interpreting a scene using a tree structure, as shown in fig.4. The strong constraints

imposed by the tree structure enable the model to search for the correct configurations of objects in a scene. This allows errors at the bottom level of detection to be corrected when placed in the larger context of a road.
This was done in Victor et.al. (2021). They used an LHPB model to generate approximate 3D models. A single image was used as an input and processed using the methods described in Venkateshkumar et.al. (2015), yielding an image with objects organized in a hierarchical structure. Depth information of the image was found and corrected for perspective projection. After the position and asset type of objects in the image were found, an approximate match of the object is placed in the model from an object library. However, due to the limited number of assets covered, the resulting models are not detailed enough to be used in engineering applications. Additionally, this study only considers short stretches of road and the perspective projection can only be applied to straight roads. As this method has only been applied to single 2D images, it is unclear if this method can be applied to a structured image set (video) or a PPC.
Mapping road markings can be used to create a 3D model. They provide valuable information about the position and types of markings present on the road. Usually, CNNs are used on an image or set of images to generate a depth map that is then projected into a 2D view. Image segmentation is then used to find road markings on the road surface. (Liu et.al., 2020).

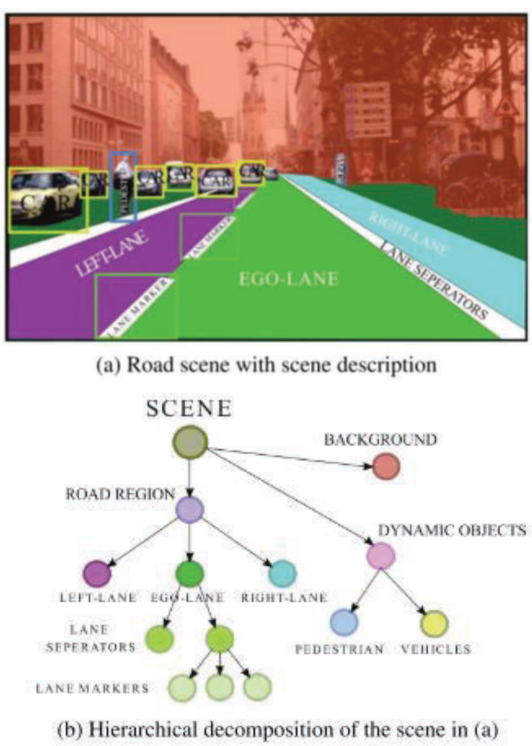

Figure 4: Road scene and hierarchical decomposition Road scenes can be split into a hierarchy. (Venkateshkumar et.al.,2015)

Another method used to infer road layout was presented in Jang et.al. (2022). The relationship between the vehicle poses and road markings was represented using a graph. This was used to identify road markings that appeared in consecutive images as well as inference of markings occluded by other vehicles. These methods were able to generate parametric layouts of complicated road scenes, such as junctions. The road markings and their relationship to one another were found rather well, exceeding 80% in accuracy and F1-score. However, their positional accuracy was relatively low. The mean IoU was typically 40-60%. This is likely due to the use of a CNN-generated depth map to project the images into 3D instead of photogrammetry-based methods. There would be a "mismatch" between the two volumetric models.
While their performance on longer stretches of road is unknown, they present a promising start to interpreting more complex road layouts.

Conversion to object-oriented 3D model

While many technologies in CV may help identify the assets in a road scene, a separate process is needed to convert these into an object-oriented model. The efforts in doing this are currently rather dispersed as the goal of the studies attempting this do not concern themselves with generating an RGDT. These studies can help inform the areas most important to focus on going forward.

The centreline of the road serves as a backbone for describing the positions of other assets on or on the sides of the road. Roads are not simple linear structures but are composed of a series of interwoven linear features: Lanes merge and unmerge, and junctions split into separate roads. Flexibility is therefore required to accurately describe a road's layout and features (Suleymanov, 2019).
There is a rich body of literature concerning itself with extracting road centrelines and road trajectories. These studies consistently report precision and recall values >90% in complex road scenes (Shao et.al., 2021; Zhang et.al., 2018; Sironi et.al., 2014).
Some studies have attempted to include converting found lines into open-data formats. Using raster-based methods Prochazka et.al. (2019) identified markings and extracted them as polyline representation, representing the lane boundaries of the road. However, the identification process fails in more complicated road scenarios as the spanning trees grow in complexity. Due to this, this method is not viable for more complex road layouts.
One could also directly extract parameters, such as road width, from a segmented PPC. This was done in Li et.al. (2021). A width estimation algorithm was used to directly extract road widths. The algorithm requires the user to input the number of width measurements to take and the length of the road section to search on. Lines, each representing one measurement, are evenly spaced, crossing the road, along the length of the road section. The length of each line is minimized to be perpendicular to the tangent of the road curvature. While robust for short stretches of road where the lane configuration is fixed, this

method is not scalable for longer stretches of road. the number of redundant measurements would be high. Essential measurements would also be missed in other places where the road width may change dramatically in a short period (junctions, merging lanes etc.).

In recent years, more work has been done to make BIM for transport infrastructure reach the same maturity as the building industry (Barbosa, et al., 2017). IFC-Road is a schema that is currently in development under this project. Due to the recentness of this schema, there have been very few studies that attempt to extract road model parameters that are suitable to be converted into that format. In fact, to the author's knowledge, no studies that use image-derived data to generate IFC-compliant models for road surfaces exist.

The spatial nature of PPCs means that methods developed using LiDAR PCs can be applied to them. In the context of generating an Ifc model, road trajectories and centrelines can be found for either the road as a whole or for each individual lane. Soilán et.al. (2020) developed a semi-automated method to extract the alignment and centreline of road lanes. In this work, the input was a LiDAR PC, and the goal was to automatically generate an IFC-compliant file. The road surface was found through region growing. The road markings are then found using Euclidian distance clustering based on the intensity values. This information was used to generate an Ifc model. In this work 20% of the road had to be manually edited, taking 69% of the total modelling time. In particular, the authors noted that the method struggles to identify the alignment of curved road sections and road markings denoting merging lanes.

Justo et. al. (2021) extended this work and used a modified approach. The approach to finding the alignments and centrelines was therefore simplified. The number of lanes was treated as fixed and known. This removed road markings that do not strictly follow the trajectory of the road, for example, due to lane merging etc. A very slight decrease in manual modelling time was recorded, taking 65% of the time to model 20% of the road. While no numerical values are provided, visual inspection reveals that the alignments that were automatically found were also more accurate. However, the shortfalls of the previous work were not addressed. While a promising start, the authors also recognise it as a scaffolding for future work.

It is interesting to compare the work done on finding road alignments to that of rail. Following up on their previous work, Soilán et.al. (2021) also developed a method to find the centreline for rail tracks and export these into IFC-Rail. The rail tracks were segmented according to height and not intensity. Rail tracks have a constant width, and the sleepers are also of a consistent shape. These factors had a major impact on the result. While the methods developed for roads are described as semi-automatic, the method for rail is described as fully automatic. Additionally, the found alignment for rail had an error of only < 3%. It can be concluded that factors that contribute to uniformity in the road profile, such as turns and lane merging, are a major roadblock to fully automating the generation of Ifc-compliant RGDTs. In the next section, other challenges are also discussed and future areas of research are suggested.

Discussion

Many methods can be borrowed from CV that could be applied to creating 3D models of road surfaces. However, as these technologies have not been developed for use in the built environment sector nor tailored to any of the use cases therein, at the current state, they cannot be directly used to create object-oriented 3D models for DTs.

For instance, road surface segmentation methods are usually not sufficiently generalised. Current methods rely on assumptions that are designed for very specific cases. For instance, there are no methods that can generate models of long stretches of road that are at least a few km long in a computationally realistic way. All the methods reviewed had difficulties in situations where the road layout was more complex, such as at a crossing or in a place where lanes merge. This is a major issue as most roads contain these types of features. Additionally, a limited amount of asset types can be reliably found. Existing studies only extract a limited number of asset types. There is certainly a tension between scalability and model complexity.

It is also unknown if and to what degree state-of-the-art methods for segmenting and mapping the road surface can improve current segmentation methods or to what scale of road this can be applied. Many of the technologies presented have also not been applied to use on road surfaces yet. For instance, road marking maps have not been generated using PPCs. CRFs that purely define the relationship between assets on the road surface have not been attempted and it is unknown if LHPB models can be effectively applied to PPCs, on non-straight roads or on long stretches of road.

Current methods also struggle to reliably extract road surface parameters on road sections with an ununiform layout, and then store them in an object-oriented format. Even with an ideal segmented PC, a significant amount of development is needed to extract geometric parameters from these PCs.

Lastly, using image segmentation to label PCs is completely dependent on how well the image is segmented. Image segmentation and object detection are currently much more advanced than 3D segmentation. Further exploring the applications of this technology in the context of creating 3D models, therefore, seems like a worthwhile pursuit.

Conclusion

Traditional methods for creating 3D models of roads either do not meet the user requirements of a DT or are too expensive to be a scalable solution. Using images and videos as a data source for creating a 3D model can greatly reduce data collection costs. Utilising this type of data also allows for the application of CV techniques to help identify and label the PPC generated from the image data.

Overall, there are challenges both to identifying the modellable assets in the input data and to converting these into a 3D model file format for final use. While current efforts have made a promising start, they have not had generating an RGDT as their final goal. The technology needs to be further tailored for use in the built environment industry. This could in turn contribute to the development of a DT that may assist various stakeholders to make more informed decisions. In turn, contributing to future-proofing road networks.

Acknowledgements

This work is part of the Cloud-based Building Information Modelling (CBIM) project, a European Training Network. The CBIM project receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant with agreement No 860555.

References

Ariyachandra, M. M. R. F., 2021. Automated Generation of Geometric Digital Twins of Existing Rail Infrastructure, Cambridge: s.n. Brilakis, I., Fathi, H., Rashidi, A., 2011. Progressive 3D reconstruction of infrastructure with videogrammetry. Autom. Constr. 20, 884–895.

Barbosa, F., Woetzel, J., Mischke, J., Ribeirinho, M.J., Sridhar, M., Parsons, M., Bertram, N., Brown, S., 2017. Reinventing Construction: A Route To Higher Productivity. Mckinsey Glob. Institute 20.

Chen, M., Feng, A., McAliden, R., Soibelman, L., 2019. Photogrammetric Point Cloud Segmentation and Object Information Extraction for Creating Virtual Environments and Simulations. Journal of Management in Engineering 36.

Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., Pan, C., 2017. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. IEEE Trans. Geosci. Remote Sens. 55, 3322–3337.

Davies, E. R., 2022. The dramatically changing face of computer vision. In: E. R. Davies & M. A. Turk, eds. Advanced Methods and Deep Learning in Computer Vision. s.l.:Elsevier, pp. 1-91.

Farhadmanesh, M., Cross, C., Mashhadi, A.H., Rashidi, A., Wempen, J., 2021. Highway asset and pavement condition management using mobile photogrammetry. Transp. Res. Rec. 2675, 296–307.

Flintsch, G. W. & McGhee, K. K., 2009. Quality Management of Pavement Condition Data Collection, s.l.: Transportation Research Board.

Farhadmanesh, M., Cross, C., Mashhadi, A.H., Rashidi, A., Wempen, J., 2021. Highway asset and pavement condition management using mobile photogrammetry. Transp. Res. Rec. 2675, 296–307.

Farr, T. G. et al., 2007. The Shuttle Radar Topography Mission. Review of Geophysics, 45(2).

Forsyth, D. A. & Ponce, J., 2012. Computer Vision: A modern approach. 2 ed. s.l.:Pearson.

Fraser, C., 2018. SLAM, SfM and Photogrammetry: What's in a name?. Riva del Garda: The International Society for Photogrammetry and Remote Sensing.

Genova, K., Yin, X., Kundu, A., Pantofaru, C., Cole, F., Sud, A., Brewington, B., Shucker, B., Funkhouser, T., 2021. Learning 3D Semantic Segmentation with only 2D Image Supervision. Proc. - 2021 Int. Conf. 3D Vision, 3DV 2021 361–372.

Hou, Y., Ma, Z., Liu, C., Loy, C.C., 2019. Learning lightweight lane detection CNNS by self attention distillation. Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob, 1013–1021.

Jain, S., Kumar, B. L. S., Shettigar, R., 2017 "Comparative study on SIFT and SURF face feature descriptors," International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017, pp. 200-205

Jang, W., Hyun, J., An, J., Cho, M., Kim, E., 2022. A Lane-Level Road Marking Map Using a Monocular Camera. IEEE/CAA J. Autom. Sin. 9, 187–204.

Jiao, Y. et al., 2013. A cloud approach to unified lifecycle data management in architecture, engineering, construction and facilities management: integrating BIMs and SNS. Advanced engineering informatics, Volume 27, pp. 173-188.

Jung, J., Che, E., Olsen, M.J., Parrish, C., 2019. Efficient and robust lane marking extraction from mobile lidar point clouds. ISPRS J. Photogramm. Remote Sens. 147, 1–18.

Justo, A., Soilán, M., Sánchez-Rodríguez, A., Riveiro, B., 2021. Scan-to-BIM for the infrastructure domain: Generation of IFC-complaint models of road infrastructure assets and semantics using 3D point cloud data. Autom. Constr. 127.

Khaddaj, M. & Srour, I., 2016. Using BIM to retrofit existing buildings. Precedia engineering, Volume 145, pp. 1526-1533.

Kirillov, A., He, K., Girshick, R., Rother, C., Dollar, P., 2019. Panoptic segmentation. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June, 9396–9405.

Kundu, A., Li, Y., Dellaert, F., Li, F., Rehg, J.M., 2014. Joint semantic segmentation and 3D reconstruction from monocular video. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8694 LNCS, 703–718.

Li, B., Ouyang, W., Sheng, L., Zeng, X., Wang, X., 2019. GS3D: An efficient 3D object detection framework for autonomous driving. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June, 1019–1028.

Li, H.T., Todd, Z., Bielski, N., Carroll, F., 2021. 3D lidar point-cloud projection operator and transfer machine learning for effective road surface features detection and segmentation. Vis. Comput.

Li, L., Liu, Z., Özgüner, Ü., Lian, J., Zhou, Y., Zhao, Y., 2018. Dense 3D Semantic Slam 3–8.

Liang, D., Guo, Y.C., Zhang, S.K., Mu, T.J., Huang, X., 2020. Lane Detection: A Survey with New Results. J. Comput. Sci. Technol. 35, 493–505.

Lindkvist, C., 2015. Contextualizing learning approaches which shape BIM for maintenance. Built environment and asset management, 5(3), pp. 318-330.

Liu, B., Zhuang, B., Schulter, S., Ji, P., Chandraker, M., 2020. Understanding road layout from videos as a whole. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 4413–4422.

Loclab ,2022a. Digital Twins, Digitale Zwillinge, Simulation, Prozessoptimierung, Entscheidungsunte-rstützung. Available at: https://www.loclab-consulting.com/ [Accessed: 16 February 2022].

Mathworks, 2022. Roadrunner. [Online] Available at: https://nl.mathworks.com/products/roadrunner.html [Accessed 11 08 2022].

Matsumoto, H., Mori, Y., Masuda, H., 2019. Extraction and shape reconstruction of guardrails using mobile mapping data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 42, 1061–1068.

Prochazka, D., Prochazkova, J., Landa, J., 2019. Automatic lane marking extraction from point cloud into polygon map layer. Eur. J. Remote Sens. 52, 26–39.

Shah, R., McMann, O. & Borthwick, F., 2017. Challenges and prospects of applying asset management principles to highway maintenance: A case study of the UK. Transportation Research Part A: Policy and Practice, 97(C), pp. 231-243.

Shao, Z., Zhou, Z., Huang, X., Zhang, Y., 2021. Mrenet: Simultaneous extraction of road surface and road centerline in complex urban scenes from very high-resolution images. Remote Sens. 13, 1–18.

Shervais, K., 2016. Structure from Motion (SfM) Photogrammetry Field Methods Manual for Students. Unavco 1–10.

Sironi, A., Lepetit, V., Fua, P., 2014. Multiscale centerline detection by learning a scale-space distance transform. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2697–2704.

Soilán, M., Justo, A., Sánchez-Rodríguez, A., Riveiro, B., 2020. 3D point cloud to BIM: Semi-automated framework to define IFC alignment entities from MLS-acquired LiDAR data of highway roads. Remote Sens. 12.

Suleymanov, T., Kunze, L., Newman, P., 2019. Online Inference and Detection of Curbs in Partially Occluded Scenes with Sparse LIDAR. 2019 IEEE Intell. Transp. Syst. Conf. ITSC 2019 2693–2700.

Tan, Y., Li, Y., 2019. UAV photogrammetry-based 3D road distress detection. ISPRS Int. J. Geo-Information.

Tang, J., Li, S., Liu, P., 2021. A review of lane detection methods based on deep learning. Pattern Recognition. 111, 107623.

USC Spatial Sciences Institute, 2021. All about 3D GIS. [Online] Available at: https://gis.usc.edu/blog/all-about-3d-gis/ [Accessed 16 08 2022].

Victor, A.C., Sreevalsan-Nair, J., 2021. Building 3D Virtual Worlds from Monocular Images of Urban Road Traffic Scenes. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 13018 LNCS, 461–474.

Vineet, V., Miksik, O., Lidegaard, M., Nießner, M., Golodetz, S., Prisacariu, V.A., Kähler, O., Murray, D.W., Izadi, S., Pérez, P., Torr, P.H.S., 2015. Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction. Proc. - IEEE Int. Conf. Robot. Autom. 2015-June, 75–82.

Wang, K., Yan, F., Zou, B., Tang, L., Yuan, Q., Lv, C., 2019. Occlusion-free road segmentation leveraging semantics for autonomous vehicles. Sensors (Switzerland) 19, 1–15.

Witteveen+Bos, 2022. Digital Twin. [Online] Available at: https://digitalsolutions.witteveenbos.com/en/news-/digtal-twin [Accessed 22 08 2022].

Vineet, V., Miksik, O., Lidegaard, M., Nießner, M., Golodetz, S., Prisacariu, V.A., Kähler, O., Murray, D.W., Izadi, S., Pérez, P., Torr, P.H.S., 2015. Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction. Proc. - IEEE Int. Conf. Robot. Autom. 2015-June, 75–82.

Zhang, C., Liu, Z., Liu, G., Huang, D., 2019. Large-Scale 3D Semantic Mapping Using Monocular Vision. 2019 IEEE 4th Int. Conf. Image, Vis. Comput. ICIVC 2019 71–76.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road Extraction by Deep Residual U-Net. IEEE Geosci. Remote Sens. Lett. 15, 749–753.

Zhong, Y., Wang, S., Xie, S., Cao, Z., Jiang, K., Yang, D., 2017. 3D Scene Reconstruction with Sparse LiDAR Data and Monocular Image in Single Frame. SAE Int.