Industrial Layout Mapping by Human-Centered approach and Computer Vision

Osmar Moreira da Silva Neto[1[0000-0002-5637-0145]] and. Marcelo Rudek[1[0000-0002-6170-3370]]

[1] Pontifical Catholic University of Paraná (PUCPR), Curitiba Paraná, Brazil
marcelo.rudek@pucpr.br

Abstract. Digital twins and immersive mixed reality have been improving the human-centered digital transformation within the automotive industry context. In computer vision, photogrammetry-based techniques for reconstructing 3D environments make it possible to create virtual models of real scenery based solely on a set of digital images. This paper addresses the problem of projecting and updating shopfloor layouts through 3D mapping of the real environment, corresponding to its digital twin. Within industrial plants, we encounter objects of various sizes, shapes, and layouts, which would typically require a significant amount of time to create a digital twin model using manual measurement devices such as tape measures and clipboards to annotate length, distances, and angles. Through photogrammetry using images, we can recreate these layouts and objects with accuracy, speed, and rich details, eliminating the need for hand-drawing them in CAD software. Therefore, we propose an automated method for digitizing shop floor layouts and creating digital twins where humans can interact and make changes to the modeled equipment structure to simulate different arrangements before implementing actual changes on the shop floor in an industrial setting. This text presents the innovations to digital environments technologies, discusses the feasibility, and evaluates the expectation of the proposed method.

Keywords: Digital Twin, Computer Vision, Industrial Layout.

1 Introduction

A recurrent industry problem is to configurate the shopfloor layout due constant changes to new products and their respective new production methods, devices and technologies. In the engineering sector this problem is a big challenge because nowadays this work continues to be done as handmade drawing sketch as an initial step to create a CAD model. Thus, this work delays several hours due to its entirely manual nature. Aiming to minimize the time required to execute this activity, we propose utilizing some new technological artifacts with greater efficiency, such as computer vision techniques and photogrammetry, together with CAD software, in order to create a corresponding Digital Twin (DT). Nowadays, in the Industry 4.0 scenario, as numerous industries undergo to digital transformation, the digital twin emerges as a support element. It is widely recognized as an indispensable asset for competitiveness and achieving significant economic advantages over competitors. By ushering in a revolutionary wave, digital twins are reshaping industries [1]. In this context, virtual simulations allow us to evaluate of different scenarios, becoming shorten the design and analysis steps, and making the process of re-designing easier and faster. Thus, after the DT implemented, it can be applied in different parts of the design of a new product design, from the conceptual the idea of the product until testing for evaluation [2] [3]. According [4], the DT modeling follows the rule of these components: real space, virtual space, and linking mechanism to get the data/information interchange between them. From normalization ISO DIS 23247 to DT environments presented in [5], we found the four mains parts of a DT standardization, as: (i) implementation study; (ii) modelling and synchronizing, (iii) analytics, and (iv) user interaction and safety. After understanding the problem's need from the initial study, the following question to be solved is about the mapping to the virtual ambient from real process. Thus, we need to build the integrated way that replicates the real scenery. And, after that to stablish the protocols of communication and synchronization rate and to define which data will be collected. In the analytical phase, the data collected are stored, analyzed and used as simulation information. The result of simulated can suit with signal of control for a plant and/or information for user. Lastly, the user and safety phase, is the definition of a way an user communicates with the digital twin, for example, between an IHM (Men-Machine Interface) or Web application [5]. Since our ongoing research is now in the phase of modelling, we present a method to time reduction in creating a Digital Twin shopfloor layout by computer vision approach instead the traditional handmade drawing layout.

2 Background

2.1 The Problem Addressed

In this research we are addressing a recurrent industry problem, that is the alteration of the shopfloor layout. The constant changes are needed due new products and its respective new production devices and technologies. In the engineering sector this problem is a big challenge because nowadays this work was doing with paper, pencil, measure tape and clipboard to handmade drawing a sketch as an initial step to create a CAD model. The flow of manual operations is presented in figure 1.

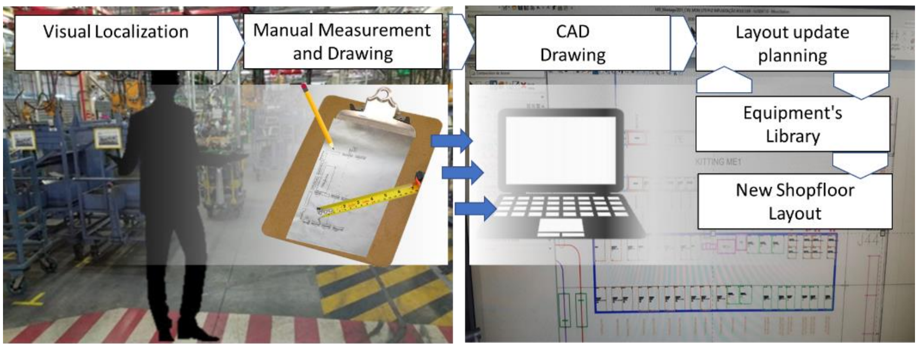
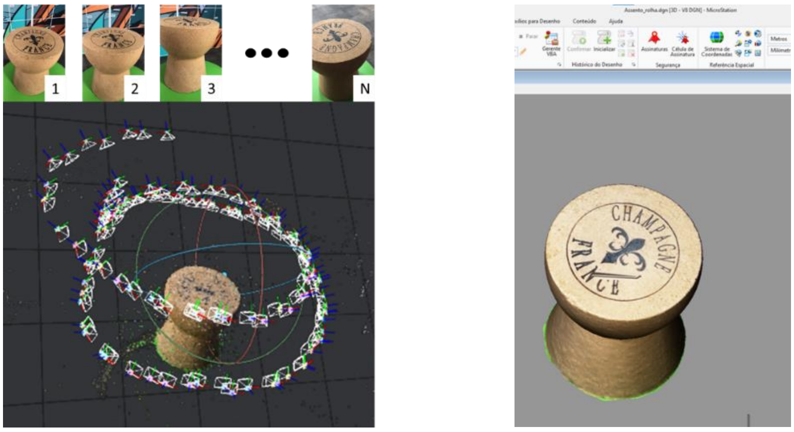


Figure 1. The manual process steps from visual inspection to layout creation.

As presented in figure 1, we are representing a real process whose operation involves the technical drawing which initially is used to represent the objects and the machines clearly and precisely identified from shopfloor [6]. This process takes several days because an employed should measure all layout area with the measure-tape and drawing it. After, the handmade sketch should be sent to the specific engineering department to be draw in a CAD system also in a handmade way. The remain of process is planning the layout update by accessing a specific equipment library and then proponing a new shopfloor layout based on the free or occupied area initially measured. This manual process is not a difficult task by itself, just a clipboard for hand drawing the layout in a sketch and after redrawing it to a CAD 2D system. However, the time consumption added to perform the drawing twice and the respective final time accumulated is a bottleneck in the process of layout reprograming. In this paper will be shown how computer vision and 3D reconstruction become feasible a digital twin to transform a real layout in your digital representation by an automatic way from video sequence images.

2.2 The Computer Vision and the 3D Reconstruction

Computer vision is the process that intends to represent a real scenery to corresponding 2D or 3D digital formatted image trough sensors, as RGB cameras, ToF cameras, Lasers and Lidars, and so one. In order to mapping the environments, both the coordinates and distance from each image pixel to the sensor in the tridimensional space must be computed faster and accurate [7][8]. The data captured by range sensors generates a cloud of points that represents an object or a complex scenery, and the identification of all elements compounding the scene can be computed by deep learning approaches. In order to overcome the high costs of the laser-based range sensors (LIDAR, ToF) the cloud of points from 3D space can be recovered by photogrammetry based on multiple views [9] from a single RGB camera, as in figure 2.



(a) (b)

Figure 2. (a) Camera's positions to *1,2,3, …, N* images from the object snapshots. (b) The equivalent 3D model from reconstructed images sequence.

With the development of VR (virtual reality) technology and the continuous emergence of new devices, three-dimensional reconstruction has become an usual research topic in the field of computer vision [8][10]. The main task is to build a three-dimensional model [9] of the real physical world based on the data collected by various sensors, using mathematical tools such as multiview geometry, probability statistics, and optimization theory, and building a bridge between the real world and the virtual world [11][12]. These theories are encapsuled in the reconstruction systems, as the example in figure 2, the equivalent 3D model from reconstructed images sequence in the software Micro Station [13].

In order to change the manual drawing layout is needed integrate these concepts with available technologies tools, as software language like for instance the Python or C++ and it is OpenCV library to handling images and format them to be used as input to the simulation environments. Also, to build the virtual environment we addressed to apply Unity, Blender, etc. engines to recreate the similar scenery.

3 The Proposed Method

In order to overtake the handmade process as mentioned previously in figure 1, we proposed the automated step inside the Digital Twin that; analytical phase, the data collected are stored, analyzed and with based on desired be simulated. The result of simulated can suit with signal of control for a plant and/or information for user [5]. Described in the steps of figure 3 as the strategy to reduce the time and the to bring a real corresponding layout for engineering analyzes and for this analyzes is necessary verify the size of object and capture the images or create a video when the ambient has not movement people or machines because this confounding the software in the process of create a Digital Twin.

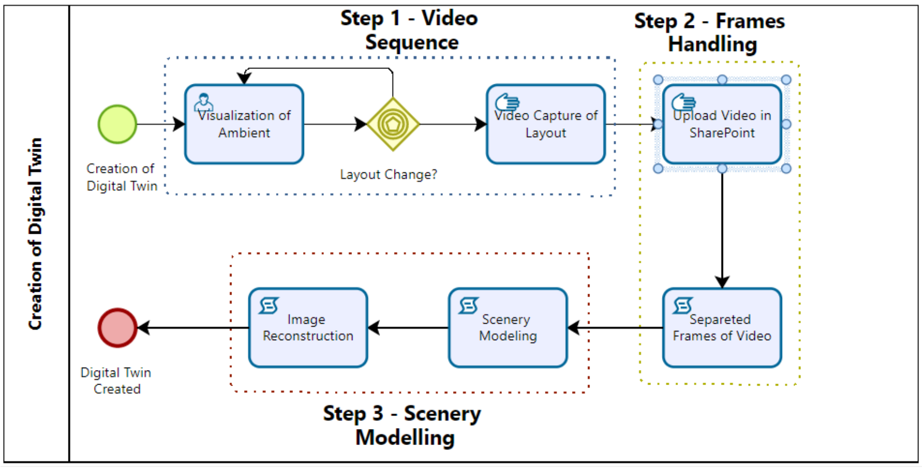


Figure 3. The proposed method with respective steps to automated layout generation.

The proposed method is human centered because we have interaction between the system and the human operator as supervisor in all steps in order to decision maker about the best strategy to produce the video sequence. We replace the sketch drawing by a video sequence captured from a single RGB camera, like embedded in a smartphone.

3.1 Step 1 – The Video Sequence

The first step is to generate a video sequence of the factory area. The video needs to have minimum tree minutes in order to obtain enough images to data generation, and a maximum of seven minutes due the size of the number of pictures to process. Also, it is ideal that this video be doing like a “caracole” sequence around the area/objects capturing all the layout up, middle down and a little of floor for reference (see figure 2). For more reliability, it was experienced that on the frames sequence should appear the one meter of floor, one meter of the extreme side left, one meter of extreme right side and one meter of up the captured area. The video sequence, as presented in figure 4, can be taken by a human operating a smartphone or piloting a drone around the objects area. By this way we can prepare the multiple views from the object/scenery in the following step.

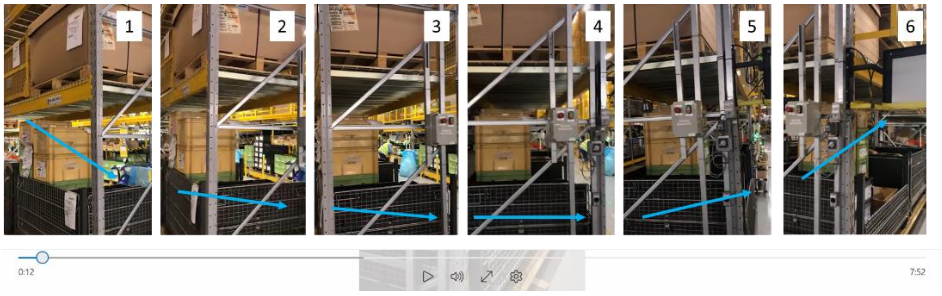


Figure 4. Multiple views taken by a RGB camera. The arrows represent the observer viewpoint.

3.2 Step 2 - Separation of Frames

With the video created, it is necessary to load it to a computer with Python language [14] [15] [16] and OpenCV library. Here we developed the code for separation of the video frames setting the FPS accordingly, as for instance 4 fps or 30 fps, looking for the influence in quantity of frames to improve the image processing. Once performed this step, we have as resulting a set of image files to be used as input to the next step.

3.3 Step 3 - Scenery Modelling

After the image uploaded in computer it is needs execute the sequential operations steps, as presented in figure 5 and described as following:

1. Images Upload – It is the stage to upload the sequenced images files to be utilized in the reconstruction process.

2. Feature Extraction – After upload the images it is necessary find and extract the pixels that remains equals between the images irrespective of rotation, translation, and scale. [18].
3. Image Matching – This section is for to find the distance of pixels equals inside the images that utilized the same scene [19].
4. Feature Matching – After find the pixel in common and the distance between pixel in images its possible separated the bad images for the good images utilizing the discretion that find in Feature Extraction and Image Matching [18].
5. Structure From Motion - The objective of this step is starting the construction the 3D object utilizing the geometric relationship utilizing the internal calibration of all the images capture and created a model if position and orientation with point clouds [20].
6. Prepare Dense Scene – In this part the software created a folder in PC where images will begin the Texturizing in scene that is transformed in another set of images [21].
7. Depth Map and Depth Map Filter – In this point each image is considered how camera as in Figure 2. Item A and its selection the best cameras will create a volume W, H, Z with many candidates that share pixels. Considering that distance between cameras is not uniform so the software utilized a filter for help in creation of volume of object [22].
8. Meshing and Mesh Filter - The objective of this step is to create a representation of scene utilizing mesh triangular that show how this object is, but in this step the object does not color yet just your format with triangles [23].
9. Texturing – In this section is painting of mesh triangular utilizing colors of images for this is applicate a filter UV in mesh [24]. In each mesh is applicate a color of accord your pixel [25], for this process be more fast its applicate a filter of low frequency and this approach is in the concept than [26] and [27].

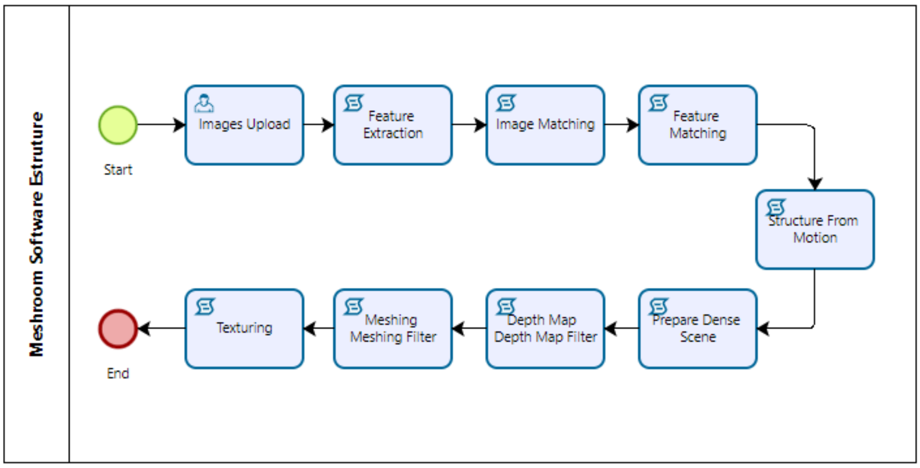


Figure 5 – The sequential operations steps to virtual image creation. This sequence can be applied, for instance, by Meshroom software structure [28].

The software Context Capture can be utilized for produce 3D models utilizing videos or photos by using these step sequences [17].

3.4 Image Reconstruction Processing

Here we need an environment to process the 3D reconstruction of scenery, for instance, using the software Context Capture. The steps performed here are the generation of the clouds of points, triangulation and reconstruction. The time of this process depends on the quantity of image frames. Recent approaches to 3D object detection exploit different data sources, and camera based approaches utilize either monocular [29] or stereo images [30]. With the introduction of digital photogrammetric techniques, certain errors associated with the subjectivity of human operators during triangulation may be avoided [31].

The main steps are:

1. Frames/Point Clouds – In this step insert the images taken with the same camera configuration (image dimension, focal length and sensor size). [32].
2. Camera Properties – Keep the knowledge of the configuration of camera that was utilized in project.
3. Aero triangulation – In this step it is the selection of all images with points in common between your characteristic is views from above. [33] [34].
4. Reconstruction Settings – Configure the parameters about reconstruction of 3D model (depends on which software is used).
5. Production – In this step is generated a model, with error feedback, progress monitoring and a possible option for retouching. In this point its possible export the file for different formats how: 3D meshes, point clouds, orthophotos/DSM [32].

After all operations we can visualize the resulting virtual image corresponding with real environment, as presented in figure 6.

Figure 6. Resulting 3D layout image after reconstruction.

4 Results

The process of creation for one industrial layout how showed in example of figure 6 if to utilize the today methods that be manuals methods wait around the tree business days of work utilizing the tools: paper, pencil, measure tape and clipboard. The propose of solution in this paper it is reduce the time for create the layout utilizing a cell phone it will do a movie of layout about the 7 minutes after in the computer wait more 1 minute for code separated frames in the movie and with this photography get in the software context capture that late 35 minutes for processing these images and generated the Digital Twin with a 3D model of layout. The time difference between these techniques was the 16h:20m – 0h:43min = 15h:37m. Thus, the proposed method saves time to around 2 full workdays as presented in the table 1.

Table 1. Comparing of Manual Process X Automated Process for Creation of One Digital Twin

Task / Activities for One Layout	**Manual Process for One Activity (current)**	**Automated Process for One Activity (Proposed method)**
Separated Tools	20 minutes	1 minute
Drawing Layout in Clipboard	8 hours	0 minute
Capture Layout	No applicable	<= 7 minutes
Drawing Layout in Software CAD 2D	8 hours	0 minute
Drawing Layout in Software CAD 3D	No applicable	35 minutes
Total time consumption	16:20 hours	43 minutes

The contribution of this research is that the process is centralized of human in the loop concept just of in this process anything object can captured and transformed in a Digital Twin all depends on the time of video.

5 Conclusion

In this paper we investigate some digitalization techniques can be applied to virtual modelling of an environment. This environment can be mapped for a 3D digital scenery representation creating a simulation scenario, from that we can identify which equipment are present in the image. Once equipment is identified, a human operation can reallocate it in a virtual sketch to evaluate the processes, (for example, like some sequential operation is working or to define a new layout configuration to improve a factory process). Once analyzed it in virtual space, the real space can be rearranged with a new devices' positioning. This paper shows the digitalization phase and the interaction between real and virtual is the ongoing research. For now, we can evaluate the process to build the virtual model by analyzing the tools as the meshroom and context capture software and evaluate the gain of time in the process. The next part of the research is to finish the second step of a DT creation, by including the synchronization activity between real and virtual worlds. For the future we will test how to utilize virtual reality for walk inside the layout utilizing an oculus VR. Also, we can embed Intelligence Artificial routines to define better layout configurations.

6 References

1. Tao, F.: Applications of Digital Twin. In Digital Twin Driven Smart Manufacturing; Academic Press: Cambridge, MA, USA, 29–62. (2019).
2. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F.: Digital twin-driven product design, manufacturing and service with big data. Int. J. Adv. Manuf. Technol. 94, 3563–3576, (2018).
3. Grieves, M.: Origins of the Digital Twin Concept, https://www.researchgate.net/publication/307509727_Origins_of_the_Digital_Twin_Concept, last accessed 2022/9/1.
4. Grieves, M.W, F.: Product lifecycle management: The new paradigm for enterprises. Int. J. Prod. Dev. 2, 71–84, (2005).
5. Bittencourt, V., Carvalho, L., de Barros, R. C.: Evolução e estudo da aplicação do digital Twin no setor industrial. (The evolution and study of the application of the digital Twin in the industrial sector) (in Portuguese), (2020).
6. Dosovitskiy, A., Beyer L., Kolesnikov A., Weissenborn D., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations ICLR 2021, pp. 1-21,(2021).
7. Ardeshiri, T., Larsson, F., Gustafsson, F., Schön, T. B. and Felsberg, M.: Bicycle tracking using ellipse extraction, 14th International Conference on Information Fusion, pp. 1-8, (2011).
8. Silva, R.L., Canciglieri Junior, O., Rudek, M.: A road map for planning-deploying machine vision artifacts in the context of industry 4.0, Journal of Industrial and Production Engineering, 39(3), 167–180, (2022).
9. Kurka, P. R. G., Rudek, M.: Three-dimensional volume and position recovering using a virtual reference box, IEEE transactions on Image Processing 16 (2), 573-576, (2007).
10. Ding, Y., Xiaofeng, Z., Zhili, Z., Wei, C., Nengjun, Y. and Ying, Z.: Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 1-12, (2022).
11. Xin, S., Adam, M. and Bahram, J.: Three-dimensional profilometric reconstruction using flexible sensing integral imaging and occlusion removal, *Applied Optics*, vol. 56, no. 9, p. D151, (2017).
12. Lei, J., Liu, Q.: Three-dimensional temperature distribution reconstruction using the extreme learning machine, *IET Signal Processing*, vol. 11, no. 4, pp. 406–414, (2017).
13. MicroStation CONNECT Edition® Design, Model, and Manage Infrastructure Homepage, https://www.bentley.com/wp-content/uploads/PDS-MicroStation-LTR-EN-LR.pdf, last accessed 2023/1/22.
14. Reffold CN.: Teaching and Learning Computer Aided Engineering Drawing, Int. J. of Eng. Ed.14, 276-81, (1998).
15. Hai-bo., W.: Computer Aided Industrial DesignJournal of Anhui University of Tech.23-26, (2005).
16. TIOBE Software Index F.: TIOBE Programming Community Index Python. pg1, (2011).
17. Bentley Systems, Homepage, https://www.bentley.com/wp-content/uploads/PDS-ContextCapture-LTR-EN-LR.pdf , last accessed 2023/2/15.
18. David, G., Lowe, F.: Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60, 91–110 (2004).
19. Nister, D., Stewenius, H.: Scalable Recognition with a Vocabulary Tree, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), pp. 2161-2168 (2006).

20. Cheng, J., Leng, C., Wu, J., Cui, H., Lu, H.: Fast and Accurate Image Matching with Cascade Hashing for 3D Reconstruction. 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2014), pp.1-8 (2014).
21. Meshroom Homepage, https://meshroom-manual.readthedocs.io/en/latest/feature-documentation/nodes/PrepareDenseScene.html, last accessed 2023/2/15.
22. Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information, IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2005).
23. Jancosek, M., Pajdla, T.: Exploiting Visibility Information in Surface Reconstruction to Preserve Weakly Supported Surfaces, Hindawi Publishing Corporation International Scholarly Research Notices, Volume 2014, pp.1-20 (2014).
24. Lévy, B., Petitjean, S., Ray, N., Maillot, J.: Least Squares Conformal Maps for Automatic Texture Atlas Generation, ISA (Inria Lorraine and CNRS), France, pp.1-10 (2002).
25. Burt, P. J., Adelson, E. H.: A Multiresolution Spline with Application to Image Mosaics, ACM Transactions on Graphics,Volume 2, Issue 4, pp 217–236 (1983).
26. Baumberg, A.: Blending images for texturing 3D models, pp 404 – 413, BMVC (2002).
27. Allene, C., Pons, J., Keriven, R.: Seamless image-based texture atlases using multi-band blending, 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 2008, pp. 1-4 (2008).
28. Chen, X., Kundu, K., Zhang, Z., Ma, H., Fidler, S., Urtasun, R.: Monocular 3d object detection for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2147–2156 (2016).
29. Chen, X., Kundu, K., Zhang, Z., Ma, H., Fidler, S., Urtasun, R.: 3d object proposals using stereo imagery for accurate object class detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2017).
30. Helava, U.V.: Digital Comparator Correlator System (DCCS), ISPRS Journal of Photogrammetry B Remote Sensing, 44(1):37 – 47, (1989).
31. Bentley Homepage.: Context Capture User Guide Homepage, https://docs.bentley.com/LiveContent/web/ContextCapture%20Help-v17/en/GUID-B9804464-3D42-420F-ACFF-7C61C1C028BD.html, last accessed 2023/1/22).
32. Jones, A. D.: American Society of Photogrammetry (ASP), Manual of Photogrammetry, Fourth Edition, Church Falls, VA, (1980).
33. Moffitt, F.H., Mikhail, E.M.: Photogrammetry, Harper & Row, New York, (1980).
34. Sallam, K.F., Abdulhadi, A., Hawas. K.: 2023 Bentley Systems, Incorporated. Homepage, https://www.bentley.com/software/contextcapture, last accessed 2023/5/20.