3D and AI Technologies for the Development of Automated Monitoring of Urban Cultural Heritage

Tadas Žižiūnas[1] and Darius Amilevičius[2]

[1] Vilnius University, Sauletekio ave. 9, I house, 10222 Vilnius, Lithuania
[2] Vytautas Magnus University, K. Donelaicio str. 58, LT-44248 Kaunas, Lithuania

Abstract. New technological solutions for the effective, objective and cost-sensitive monitoring of cultural heritage are needed. Accordingly, a new methodological approach based on laser scanning, 3D photogrammetry, artificial intelligence and GIS interaction is presented in this paper. The main goal is to develop a software that could detect and compare various architectural and urban elements by comparing 2D and 3D data of objects and places of the same cultural heritage from different time periods. This represents a breakthrough technological tool for governments to track the broad-scale status of heritage and act in a timely and proactive manner.

The methodological approach was to inspect changes comprised of geometrical alterations in 3D data and pixel-based information changes in 2D data. The proposed solution was developed as part of a project financed by the Research Council of Lithuania entitled *Automated monitoring of urban heritage implementing 3D technologies*. The first results of the project are presented in this article. All pictures and tables in this paper were prepared by the authors.

Keywords: Digital Monitoring, 3D, AI, CNN, Cultural Heritage.

1 Introduction

The preservation of urban heritage is one of the main challenges in contemporary society. It is closely connected with several dimensions, including global-local rhetoric, cultural tourism, armed conflicts, immigration, cultural changes, investment flows and infrastructure development. [1] Nowadays, cultural organisations are often responsible for heritage management and must contend with a consistent lack of resources that are crucial for proper heritage preservation, maintenance and protection. This is particularly problematic for countries with a low Gross Domestic Product (GDP) or unstable political situation.

One possible solution to these problems is an automated heritage monitoring software system based on 3D data and artificial intelligence (AI) technologies to increase monitoring efficiency (scale, financial, time and data objectiveness factors). A system prototype was developed and tested by Vilnius University and Terra Modus Ltd. in the context of the project *Creation of automated urban heritage monitoring software prototype* (2014). This prototype helped to initiate a development of a full-capability system which remains under development by Vilnius University within the framework of the project *Automated urban heritage monitoring implementing 3D and AI technologies*. This project is financed by the Research Council of Lithuania (project time 2018–2022) [**2**]. It should be noted that the creation of a new method of Computer Vision is not within the scope of this paper. Instead, this article researches and combines the best monitoring practices into one single and cost-effective solution that is suitable for large scale real-life conditions. This paper presents the general pipeline and the first stage of the project.

1.1 Brief Overview of 3D and AI Technological Solutions for Heritage Monitoring

Artificial intelligence has been used in the context of cultural heritage within the last two decades [3]. Three-dimensional data analysis for machine learning procedures has been applied relatively recently, where deep convolution network-based calculations have been performed in this sector [4].

During last two to three years, these technologies have accelerated, hence convolutional neural networks (CNN) suggested more and more capabilities. On the other hand, more well-known and established 3D laser scanning and 3D photogrammetry technologies for collecting real world data have mainly been used in case studies [5], where only some technological pipelines are presented, but no analytical technological integration in more difficult tasks, such as cultural heritage maintenance, is applied.

A deeper analysis of the sparse literature suggests that there have been almost no studies on the use of 3D and AI-based technologies to monitor cultural heritage [6]. Therefore, the contribution of this paper is potentially important for this sector.

1.2 Elements of Traceable Alterations

Urban cultural heritage values are defined from a theoretical point of view. In this project, our primary target was Lithuania. This is why all the framework and terminology used to describe heritage in this paper are based on Lithuanian national law, legislation and ratified international conventions such as the Convention for the Protection of the Architectural Heritage of Europe (1985). For this reason, urban cultural heritage could be dismantled into a list of elements – valuables – that could be tracked with the digital monitoring system that is under development:

- valuables of an area

a) structure of the plan (radial, annular, regular, linear, hybrid); planned structure network (roads, streets, squares, pedestrian paths, possessions); quarters; holdings (possessions); roads, streets, squares, driveways, passages, paths; elements of nature);
b) volumetric spatial structure (structure of volumetric spatial structure); open spaces (streets, squares, squares, parks); enclosed spaces (yards, passages); panoramas; silhouettes; dominants, characteristics of arrangement of place ant its decoration, colours and surface details.

- valuables of a building

a) height and height levels
b) volumetric composition, volume, shape of the roofs, elements of the roof (skylights, hatches, fireproof walls, chimneys);
c) structure of facades, compositions, volumetric elements (acroteriums, fountains, awnings, oriels, balconies and railings, exterior staircases and ramps, pillars and colonnades), décor of the facades (sculptures), and other elements (rain drains, fire escape ladders, lifting elements);
d) doors, windows and gates.

During the first stage of the project, most common alterations (done legally and illegally) in urban heritage areas (e.g., Vilnius old town) are identified: doors, windows, gates, height of the building, volume of the building, roof elements (e.g. new skylights and volumetric skylights).

2 Methodology for Digital Monitoring of Cultural Heritage

The methodological framework (Table 1) mainly consisted of two aspects. Firstly, 3D technologies (3D laser scanning and digital 3D photogrammetry) were used to collect the geometrical data and other (e.g. albedo, colour). Secondly, AI technologies were used to identify and compare the valuables. 3D laser scanning is the most accurate way of collecting geometrical exterior information about buildings and their surroundings, where handheld and drone-based photogrammetry helps to collect information about roofs and closed from open access spots in the area.

This method offered additional advantages in relation to the time and costs associated with the data collection, which is usually taken with a more than 5 cm accuracy in a non-destructive way. Moreover, every time the data were collected (for comparison in time), the entire geometrical data package was gathered without distinguishing between the 'most important' and 'least important' objects and places. Performing the monitoring digitally offered additional opportunities, for example the possibility of reconstructing destroyed buildings or open access to particular cultural object for educational, tourism, augmented reality, game industry or other purposes (e.g., projects like Cyark.org/).

2.1 Theoretical Preconditions for Digital Monitoring

Digital monitoring is based on seven conditions. The first is that all objects in the monitoring process are tangible, while the second stipulates that physical valuables can be expressed as simple geometrical forms or mathematical expression. The third condition is that monitored objects can be fully scanned and photogrammetrically processed. Fourthly, the data derived from Lidar devices and from photogrammetry are same quality (e.g., density, coverage). The fifth condition is that detection of cultural heritage can be analysed by AI algorithms. The sixth is that the digitally processed results should be able to be checked in reality. Lastly, the seventh condition is that digital monitoring is based on non-destructive and non-invasive 3D and analytical technologies.

Table 1. Geometry alteration methodology for urban architectural cultural heritage monitoring.

Method applied	Criteria and technology	Results	First level of interpretation
MEASURE	• Lidar technologies • Digital camera (photogrammetry) • Satellite recipient	• Point cloud • Digital photos • GPS/Glonass data	• Point cloud data corresponds to real physical characteristics of objects • Digital photos are suitable for processing 3D models • Coordinates are precise (within a desirable accuracy)
DATA PROCESSING	• Data pre-processing • Data optimization • 2D → 3D conversion • 3D data merging with coordination	• Selection and filtration for needed 3D data • Coordinated 3D data scene or real objects	• Data suitable for AI processing
DETECTION OF ELEMENTS	• Identifying cultural objects valuables elements and localization using AI	• Database with semantic labels and location metadata	• Identified valuable elements in the point cloud or photography
COMPUTATIONAL ANALYSIS OF MEASUREMENTS	• AI algorithms for the inspection of geometrical alteration (in 3D) • AI algorithms for the inspection of pixel value and form alteration (in 2D)	• Geometrical changes • Pixel value changes	• Demolition • Sustain • Addition • More space/ volume • Less space/ volume

The information collected from the different time period measurements served as data for the AI analysis, which can automatically identify the needed valuables and their changes during a particular time period. The data comparison was conducted by interpretation.

The first level of interpretation demonstrated some information about geometrical changes. The second level depended on the particular legal status and local legislation for managing cultural heritage (i.e., the meaning of the detected changes depends on legislation). The first level of the interpretation could be evaluated by logical operators (Table 2). For example, alteration is described as ‘status quo unchanged’, ‘reduction in volume by 65%'. The second level of interpretation could be a legal analysis of the first level results (e.g., 'reduction in volume = fact of illegal demolition works').

Table 2. Logical operators of alteration detection. XYZ – initial data of X, Y, Z; +1 – modus of alteration. 'Importance' means which operator is higher in hierarchy. By optimising the analysis, second level operators could be eliminated.

Logical operator	Earlier data	Later data	Sequence of alteration	importance
destruction	XYZ	-XYZ	is→ non	first
creation	XYZ	XYZ + 1	non→ is	first
increase of area/volume	XYZ	XYZ + 1	is→ is (increase)	second
decrease of area/volume	XYZ	XYZ -1	is→ is (decrease)	second
unchanged status quo	XYZ	XYZ	is→ is	first

In this project, only the first level of interpretation was under consideration.

3 Data Analysis with Deep Learning Technology: Case Study of Vilnius Old Town

To performing the inspection of alterations in Vilnius Old Town, a database of 3D and 2D fixation was required. To conduct comparisons, further data about the same objects from different time periods was needed. The current asset condition assessment procedures are highly time consuming, laborious, expensive and can even pose a threat to the health and safety of surveyors, particularly at height and roof levels that are difficult to access. The challenges and limitations of our model and alternative solutions in real-life applications were identified.

After initial experiments, we eliminated solutions that used only photogrammetry to process 3D point clouds. This solution provides material that enables precise examination, but it is not time-effective and would be difficult to apply at a large scale, particularly in the context of an old town, as in Vilnius (area of 350 ha). Advances in Lidar sensors and the decrease in cost over the last five years meant that performing handheld scanning with systems like Geo-Slam Horizon, which was used to collect 3D information on the ground in Vilnius, was suitable to our project's needs. On the other hand, Lidar usually does not provide sufficient information about small decorative objects [7] in facades; for this reason, high-definition images of facades were also used. These photos were also required for performing semantic segmentation procedures to identify valuables in the analysed data. Consequently, the data collection (for status fixation and for semantical segmentation) of objects under the monitoring process was carried out in real-time using mobile devices, cameras, drones and Lidar scanners.

3.1 Collecting the Data

The biggest challenge was the data set creation. The Old Town of Vilnius presents a vast variety of architecture styles. For example, it has been reported that there are almost 100 different types of window shapes in the Old Town of Vilnius. In open access data [8], there are no sufficiently large data sets for training old town facades semantic segmentation's classifier. We also found that archive of the Department of Cultural Heritage under the Ministry of Culture in Lithuania was not suited to our needs. It contained a vast number of facade images, but one half were taken before World War II, and the second half were taken in present times. The first half were comprised of highly professional images, containing all element of the facades, but were very low resolution. Among the second half, the majority of the pictures were low resolution and were taken from perspectives that did not provide all elements of the facades (Figure 1).

Fig. 1. Sample 2D photos of Vilnius Old Town from the Department of Cultural Heritage under the Ministry of Culture, https://kvr.kpd.lt/#/static-heritage-search, last accessed 2020/02/11

For the purposes of the experiment, we created a small data set of old town facade images – 420 high resolution units – where the windows and doors were annotated manually by human experts. Annotation was performed using Labelbox tools, which enables annotations to be exported in a TensorFlow format. The data set is currently updated with a larger number of additional images of Vilnius Old Town and other cities' old town building facades. After completing the project, our 2D photo data set will be available as an open data set.

A second database (for testing) was prepared by the project team and partners (Spotland Ltd) using Lidar data of Vilnius Old Town from the summer of 2019. To ensure the completeness of measurements, a photogrammetric survey using a drone was conducted to cover architectural structures from above (e.g., rooftops, roof windows). To cover structures that could not be seen from above, the whole area was surveyed using a handheld SLAM scanner (GEO-Slam Horizon). For geo-referencing the results ground control points (GCP) were surveyed using a GPS station and theodolite. The first step in the processing was performed using Pix4D software, and the photogrammetric survey was processed using regularly distributed GCPs. The ground-based mobile scans were processed with the manufacturer's own software, GeoSLAM Hub. Geo-referencing was performed in a second step using CloudCompare. Once aerial and ground derived point clouds were processed and georeferenced, the SLAM point clouds were fine-aligned to the drone point clouds. This was carried out using the iterative closest point algorithm implemented in CloudCompare (Fine Registration ICP). As a result, ≈21 ha area was captured with detailed precision (~4 cm RMS), with ≈2.9 billion points where 1 m^2 covers ≈13400 voxels.

In addition, 2017 Lidar data from the National Land Service under the Ministry of Agriculture is the Lithuanian State was used. Unfortunately, the data density is very sparse (45 points/1m^2), hence only a few valuables such as the heights of buildings or volumetric alterations could be identified with the final system.

The third part covered the GIS data-based mapping of facade information with corresponding objects in the Lidar scanner material and also involved mapping the results to the general orthophoto of heritage objects of the old town. In this case, in the GIS layers of all Vilnius' heritage objects with attribute information (coordinates, perimeters, etc.) were filtered.

3.2 AI workflow and some results

Experiments on the first technical section were concluded. After detailed research on open source technical solutions for the facades, semantic segmentation Google DeepLab v.3+ (addition to Tensorflow) and PASCAL VOC 2012 standard to facilitate learning transferring techniques were chosen [9]. It is well documented and has a large community of developers and researchers in the internet. Image segmentation involves partitioning an image into multiple segments to facilitate the analysis of a given image. There are two different types of image segmentation: semantic segmentation and instance segmentation. Every pixel in the image belongs to one a particular class (e.g., door, roof, window). All pixels belonging to a particular class are assigned a single colour. It should be noted that classification assigns a single class to the whole image, whereas semantic segmentation classifies every pixel of the image into one of the classes. Pixel-level labelling tasks, such as semantic segmentation, play a central role in image understanding. Image segmentation is a long-standing Computer Vision problem that a number of algorithms have been designed to solve, such as the Watershed algorithm, Image thresholding, K-means clustering, and Graph partitioning methods. A number of deep learning architectures (like fully connected networks for image segmentation) have also been proposed; however, but Google's DeepLab model has provided the best results and was sufficient for our project needs, as well as being cost effective [10].

The DeepLab model is mainly composed of two steps: the encoding phase and the decoding phase. The DeepLab architecture is based on a combination of two popular neural network architectures: Spatial Pyramid Pooling and Encoder-decoder networks. Spatial pyramid pooling uses multiple instances of the same architecture, which leads to an increase in computational complexity and the memory requirements of training. To deal with this problem, DeepLab has introduced the concept of 'atrous convolutions', a generalized form of the convolution operation. Here, ASPP uses four parallel operations: a 1 x 1 convolution and 3 x 3 atrous convolution with rates (6, 12, 18). It also adds image level features with Global Average Pooling. Bilinear upsampling is used to scale the features to the correct dimensions. DeepLab V3+ uses Aligned Xception as its main feature extractor, with the following modifications: 1) all max pooling operations are replaced by depth wise separable convolution with striding, 2) extra batch normalization and activation are added after each 3 x 3 depth wise convolution, and 3) the depth of the model is increased without changing the entry flow network structure [11].

Our initial experiment of the proposed model on PASCAL VOC 2012 and our custom 2D data set demonstrated an effectiveness performance of 80%, without any post-processing.

4 Conclusion

An analytical tool was proposed for the effective monitoring of urban architectural tangible heritage. Digital monitoring is based on effective 3D laser scanners and digital 3D photogrammetry. The 2D, 3D and GIS information collected from different time periods could serve as a data for AI analysis to automatically identify needed valuable elements, its location and changes during a time period. Such monitoring could be performed remotely, non-destructively and in a cost-effective way.

Artificial intelligence can perform extremely precise calculations where convolution neural networks are best possible choice. Experimental testing with Vilnius Old Town demonstrated that convolution neural networks suggest vast possibilities for developing the final monitoring system.

Acknowledgements

This project has received funding from European Regional Development Fund (project No 01.2.2-LMT-K-718-01-0043) under grant agreement with the Research Council of Lithuania (LMTLT).

References

1. Managing Disaster Risks for World Heritage, http://whc.unesco.org/uploads/activities/documents/activity-630-1.pdf, last accessed 2020/01/20.
2. Informational portal of 2014–2020 European union investment in Lithuania, https://www.esinvesticijos.lt/lt/paraiskos_ir_projektai/urbanizuotu-vietoviu-paveldo-automatinis-monitoringas-panaudojant-3d-vaizdo-technologijas, last accessed 2020/02/07.
3. Patel, J., et al.: Beyond classification: the use of artificial intelligence techniques for the interpretation of archaeological data (1989), https://publikationen.uni-tuebingen.de/xmlui/bitstream/handle/10900/60942/30_Patel_Stutt_CAA_1989.pdf?sequence=2, last accessed 2018/09/27; Dries,M. H. Archaeology and the application of artificial intelligence: case-studies on use-wear analysis of prehistoric flint tools. Doctoral thesis (1998). Faculty of Archaeology, Leiden University,

https://www.narcis.nl/publication/RecordID/oai:openaccess.leidenuniv.nl:1887%2F1314, last accessed 2019/10/01; Barcelo, J. A. Computational Intelligence in Archaeology (2008).

4. Gualandi, M. L., et al.: ArchAIDE, EUROGRAPHICS Workshop on Graphics and Cultural Heritage, https://arpi.unipi.it/retrieve/handle/11568/863393/229085/ArchAIDE_short-final.pdf, last accessed 2018/10/02; Pirotti, F., et al., Detection of building roofs and facades from aerial laser scanning data using deep learning (2019), https://re.public.polimi.it/retrieve/handle/11311/1090187/376920/isprs-archives-XLII-2-W11-975-2019.pdf, last accessed 2018/09/14.
5. Taylor J., et al.: 3D imaging technology for museum and heritage applications, http://onlinelibrary.wiley.com/doi/10.1002/vis.311/full, last accessed 2017/01/31; Haddad A. Naif. From ground surveying to 3D laser scanner: a review of techniques used for spatial documentation of historic sites, http://www.sciencedirect.com/science/article/pii/s1018363911000250, last accessed 2019/11/22.
6. Žižiūnas, T.: The technological aspect in cultural heritage research: the application of a methodological model of 3D technologies and spectroscopy, Summary of Doctoral Thesis (2019), https://epublications.vu.lt/object/elaba:38980985/index.html, last accessed 2020/02/10.
7. Sun Y.: Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data (2018), https://www.sciencedirect.com/science/article/abs/pii/S0924271618301680, last accessed 2020/02/11.
8. The CMP Facade Database, http://cmp.felk.cvut.cz/~tylecr1/facade/, last accessed 2020/02/11; eTRIMS Dataset, http://www.ipb.uni-bonn.de/projects/etrims_db/, last accessed 2020/02/11; Ecole Centrale Paris Facades Database, http://vision.mas.ecp.fr/Personnel/teboul/data.php, last accessed 2020/02/11.
9. Krizhevsky A.: et al., ImageNet Classification with Deep Convolutional Neural Network. https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf; Simonyan K., et al., Very Deep Convolutional Networks for Large-Scale Image Recognition. https://arxiv.org/pdf/1409.1556.pdf, last accessed 2020/02/11.
10. Garcia-Garcia A. et al.: A Review on Deep Learning Techniques Applied to Semantic Segmentation. 2017. https://arxiv.org/pdf/1704.06857.pdf; last accessed 2020/02/11.
11. Chen Ch., et al.: Rethinking Atrous Convolution for Semantic Segmentation. 2017. https://arxiv.org/abs/1706.05587; last accessed 2020/02/11; Chen l. et alias, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. https://arxiv.org/pdf/1802.02611.pdf, last accessed 2020/02/11.