Using Big Data for Solving Big Problems of the Cultural and Historical Heritage

Orlin Kouzov[1, 2]

[1] GATE Institute – Sofia University "St. Kliment Ohridski",
125 Tsarigradsko shosse Blvd., Block 2, Floor 3, Sofia, Bulgaria
orlin.kouzov@gate-ai.eu
[2] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences,
Acad. Georgi Bonchev Str., Block 8, 1113 Sofia, Bulgaria
orlinkouzov@gmail.com

Abstract. Damage or destruction of cultural and historical objects due to time and other factors is a key problem in culturology and archaeology. The article examines the use of technologies related to processing of the so-called „big data" in search of a rational solution to this important scientific and social problem.

Keywords: Big Data, Artificial Intelligence, Digital Reconstruction, Machine Learning, Computational Model

1 Introduction

One of the main problems in the preservation of cultural and historical objects with their original characteristics and integrity is their physical vulnerability. Over time or with the deterioration of different natural, meteorological or ecological conditions, various objects - interest of cultural and historical heritage may be seriously affected or even completely destroyed and generations may be permanently deprived of the relevant historical period, creators, cultural epoch or an important stage in the development of the respective civilization or community. Adding the possibility of the same objects being damaged by natural disasters, wars, robbery or simple vandalism, we obviously need to choose an entirely new strategy for the preservation of cultural and historical heritage if we want to keep it for the future generations.

One of the obvious ways for this to happen is the digitalization (Dochev, Pavlov, Paneva-Marinova, & Pavlova, 2019). Whether it is an architectural landmark, an original mural, or a famous sculpture or painting, it is obvious that any physical matter sooner or later changes, damages, or decays, and there are no traditional methods to preserve it. On the other hand, if we have an exact digital copy of the object, then with the development of technology, we can most likely be able to reconstruct the original if something happens to it. Such technology will be used, for instance, for the restoration of the famous Notre Dame Cathedral in Paris, which was severely damaged by fire last year.

Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings. Vol. 10, Sofia, Bulgaria: Institute of Mathematics and Informatics – BAS, 2020. ISSN: 1314-4006, eISSN: 2535-0366

With the advancement of technology, digitalization will become an easier and more natural process. The 3D scanner and printer have been in the restorer's arsenal for years, and laser scanning and holographic images are now on the agenda. In any case, these are obvious technologies when it comes to existing and available objects, but how to deal with objects that have already been damaged or lost before the time their digitization was possible. It is here where supercomputers, artificial intelligence (AI) and the so-called "Big data" could help us.

2 Digital Reconstruction Using Machine Learning

We often witness authentic historical objects that are damaged or scattered in parts that researchers are trying to restore. Apart from the obvious problems that such a 3D puzzle would create, a very common situation is that some segments of the object we are trying to restore are missing. In general, the winning strategy is critically dependent on two factors - to try to find or produce somehow a maximum number of realistic parts/segments of the object, as well as to involve a sufficiently powerful technological resource in order to make them fit to each other as close as possible in order to recreate the authentic object. Of course, with a sufficiently complex items or relatively tiny segments, such an "assembly" can take years, and if we rely solely on human capabilities, much of our past would probably be lost forever, because simply most people will not live that long in order to restore it. However, if we have a sufficiently modern and efficient computer algorithm that can be run on a supercomputer or cloud environment with a sufficiently large computing capacity (Yoshinov, Pavlova, & Kotseva, 2015), then the time for processing and analysis of the data collected would probably be drastically shortened and the capabilities of the artificial intelligence would naturally complement the skills of the respective architects, culturologists and restorers. Moreover, if we have a sufficient amount of valid data in the analysis (for example, segments of similar works or other objects from the same era or artist), the computer could be "trained" to enter the "mind of the artist" and recreate the segments based on reliable simulations of the already verified objects and parts. It is known that most famous artists have a specific style that could hardly be mistaken. Creativity is usually a secluded activity, and the genius who creates the masterpiece usually has a specific style that the computer algorithm can sense and try to recreate, whether it is a sculpture, a painting or a piece of music. Respectively, even if we do not have an exact prototype, there is always some acceptable degree of reliability that gives us reasonable confidence in how valid our assumptions are. Naturally, this information can cross tested with the so-called "Big data" from the respective era on various assumptions. For example, if a building with a remarkable architecture is destroyed, but we have fragments of a drawing or mural showing the original or travelogues of eyewitnesses describing what they saw, all this information can be added to the database used by the artificial intelligence thus helping it to further refine its assumptions. Of course, much more information can be added to the existing data - for example, from the analysis of the material from which the building is made and calculations of how the time period from its origin would affect the observed changes, the process can be approximated backwards and the idea could be reached of what the original looked like. Scientific theories are based on facts - when planning the restoration we could have hundred theories about the type of the original object, then by accumulating the facts, collecting and combining different pieces of information from large data sets relevant to the topic, we can exclude most of the hypotheses as impossible, thus significantly increasing our chances to realistically restore an object that no longer exists.

3 When Big Data Provides Big Opportunities

There is a whole theory for the reconstruction of three-dimensional objects from 2D photographic images, and the options are either photos from different cameras or a series of photos from one camera (Zhou, Geng, & Wu, 2012). It is generally based on the depth of the pixels in the photo images, epipolar geometry is involved, and the pixels in the different images are superimposed on each other.

Fig. 1. 3D Reconstruction from set of 2D images (AI/HUB, Durham College - ORSIE, 2018)

Although this methodology is now relatively inexpensive to implement, due to the availability and increasing precision of amateur cameras, behind it lies a serous toolkit and mathematical apparatus, which covers at least 4 phases, namely: Camera Calibration, Pose Estimation, Epipolar Geometry, Depth Map from Stereo Images. We will not analyse this process in detail, because despite all its variables and conventions, it is still based on a specific photo or video, which for most of the already lost historical and/or cultural objects is obviously inapplicable. Although photography has a long history, and the first photos were taken nearly 200 years ago, they cannot turn back time and for authentic restoration of objects from, say, ancient Greece, the old Rome or even the Italian Renaissance, we must use other methods. It is here that big data and artificial intelligence appear as important development tools in the restoration process.

Let's imagine the life cycle of a historical site and how many different types of traces it actually leaves on its way. First of all, even if we are talking about objects that no longer exist, there are usually some physical traces from the objects themselves somewhere - ruins, pieces, or sometimes just dust. It is often debated even whether what was found was what we were looking for or something completely different – it is enough to recall Heinrich Schliemann and the discovery of ancient Troy, which turned out to be just one layer of more than ten settlements built on top of each other and it took years to study and obtain sufficiently reliable archaeological evidence in which layer are the remnants of the Trojan War, described by Homer.

After all, we are usually looking for specific traces of a specific historical epoch, and the more we have, the greater id the chance that we can recreate it. If an object, such as a famous statue, is of great historical or cultural significance, it is usually the work of a genius artist, and they rarely make a masterpiece without some preparation - models, layouts, sketches etc. … Even if the original object is lost, it is likely that some of the accompanying materials have been preserved, and they have serious informational value because initially they were used to make the original, so they can simply be used again, already for its restoration. In the same line of thought, the great works are usually subject of public and artistic interest and could in turn become parts of a subsequent creative process - drawings, sketches, lines and so on. It is a problem nowadays, for example, to recreate the peeing boy, which is a symbol of Brussels, as it has millions of replicas, souvenirs, images - in fact, although dating back to the early 17th century, this statue has been abducted many times over the years, but every time the city authorities replace it with a realistic copy, and most tourists don't even realize that they only know the copy, because the original has been lost long ago. The situation is similar with the famous statue of the little mermaid, located in Copenhagen - it has been severely damaged on many occasions and has been successfully restored each time.

Finally, written traces are important as a source of information - for example, descriptions of travellers, historical chronicles, family stories, etc. All this is a source of serious information that can be of scientific importance, as long as all the available data can be summarized and put into a sufficiently powerful and intelligent computing system to sift through the possibilities and "guess" what were the original type, size, material, etc. characteristics of the missing object. In short - from all the available data, we can, by assuming a relevant logical model, synthesize the actual historical item/ object, based on the matching information and the huge computational capabilities of the AI.

4 Reconstruction of a Historical Object, Based on AI

The logical structure of the model, we will use, is described in Figure 2. It illustrates the digital ecosystem, enabling us with the necessary toolkit to reconstruct objects that have been destroyed or seriously damaged throughout the centuries. In practice, the researcher's role is to ensure that the system is fed with data, relevant to his/her point of view, that can be harmonized and compared by the AI. As in the mathematical modelling of the three-dimensional objects from the two-dimensional images (Cosmas, Itegaki, & Green, 2001), this model also uses mathematical toolkit that searches for intersection points, but already in the n-dimensional space, analysing not only the spatial parameters but also any other distinctive features in the points it tries to fit.

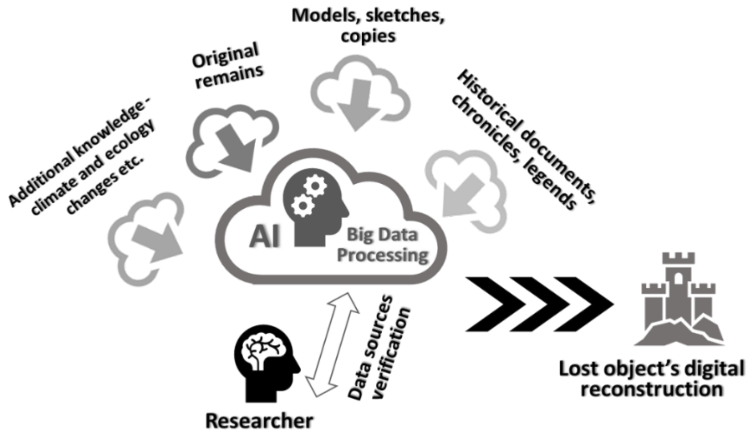

Fig. 2. Historical object reconstruction, based on big data sources and AI suggestions

For example, types of matching check-ups could be of various nature such as whether the material of a fragment corresponds to the epoch, whether the style suits the respective artist, whether the way of processing is similar to other items with guaranteed authenticity, etc. It is this multi-spatial cross-referencing that allows us to establish an increasing degree of compliance, and, for understandable reasons, the richer is the factual basis, with which we make a comparison, the higher is the probability percentage that we have reproduced the exact model. On the other hand, even a key discrepancy would already be a clear indication that our model does not correspond to the reality and we need to change the simulation parameters and restart the reconstruction.

The main idea of the model is that the result, that AI offers to the researcher, should have harmonized all the inconsistencies without contradicting the verified facts like any other scientific field or theory develops. Of course, typical for such a model is its openness, i.e. if we subsequently receive new artifacts or traces, they should be subjected to the same treatment and further refine the result. Openness can be complemented by subsequent connections with other computing clouds (Pavlova, 2019) or we can now even talk about communication at the AI level in search of the optimal research results, thus creating a self-developing intelligent algorithm.

5 Conclusions

The development of archaeology and the study of cultural heritage require more and more precise and powerful tools (Bernard, 2003), as the damage that humanity inflicts on the planet through its activities undoubtedly affects historical monuments. Similar damage could cause all the other effects of the climate, as well as the various natural phenomena. Even in the most sterile conditions, we should consider the fact that every historical object has some lifetime, whether it is a painting, sculpture, or architectural monument. Today we experience a data-driven society (Petrova-Antonova, Krasteva, Ilieva, & Pavlova, 2019) thus the ability to now effectively use big data and artificial intelligence to preserve or restore such objects is undoubtedly a turning point in our understanding of the world, and it is quite possible that thanks to these innovations we will soon be able to recreate a realistic picture of history far beyond our previous ideas and opportunities.

Acknowledgements

The result presented in this paper is part of the GATE project. This project has received funding from the European Union's Horizon 2020 WIDESPREAD-2018-2020 TEAMING Phase 2 programme under Grant Agreement No. 857155.

References

AI/HUB, Durham College - ORSIE. (2018, 10 23). *3D Reconstruction with Stereo Images -Part 1: Camera Calibration*. Retrieved from https://medium.com/: https://medium.com/@dc.aihub/3d-reconstruction-with-stereo-images-part-1-camera-calibration-d86f750a1ade

Bernard, F. (2003). Mission and Recent Projects of the UCLA Cultural Virtual Reality Laboratory. *Conference Virtual Retrospect 2003*, (pp. 65-74). Biarritz, France.

Cosmas, J., Itegaki, T., & Green, D. e. (2001). 3D Murale: A Multimedia System for Archaeology. *Conference on Virtual reality, archaeology and Cultural Heritage*, (pp. 315-322). Glyfada, Greece.

Dochev, D., Pavlov, R., Paneva-Marinova, D., & Pavlova, L. (2019). Towards Modeling of Digital Ecosystems for Cultural Heritage. *Digital Presentation and Preservation of Cultural and Scientific Heritage, 9*, 77-88.

Pavlova, D. (2019). Exploration of a distance learning toolkit through integration the capabilities of public and private clouds in a heterogeneous environment. *INTED2019 - 13th International Technology, Education and Development Conference.* Valencia.

Petrova-Antonova, D., Krasteva, I., Ilieva, S., & Pavlova, I. (2019). Conceptual Architecture of GATE Big Data Platform. *CompSysTech '19: Proceedings of the 20th International Conference on Computer Systems and Technologies* (pp. 261-268). Ruse: ACM.

Yoshinov, R., Pavlova, D., & Kotseva, M. (2015). Specifications for centralized datacenter serving the educational cloud for Bulgaria. *XII International Conference on Electronics, Telecommunications, Automatics & Informatics (ETAI)* (pp. 1-6). Ohrid, Macedonia: Ss. Cyril and Methodius University "Faculty of electrical engineering and information technologies.

Zhou, M., Geng, G., & Wu, Z. (2012). *Digital Preservation Technology for Cultural Heritage.* Beijing: Springer Science & Business Media.

Received: May 14, 2020
Reviewed: June 10, 2020
Finally Accepted: July 10, 2020