

An Approach to Auto-Enhance Semantic 3D Media for Ambient Learning Spaces

David Bouck-Standen, Alexander Ohlei, Thomas Winkler and Michael Herczeg

Institute for Multimedia and Interactive Systems (IMIS)

University of Luebeck

Luebeck, Germany

email: [bouck-standen, ohlei, winkler, herczeg]@imis.uni-luebeck.de

Abstract—In this contribution, we present an approach in enhancing 3D objects, which are automatically reconstructed from semantic media in a cloud-based ambient platform. Due to the automatic background process of 3D reconstruction, the objects contain artifacts from the reconstruction process and are not aligned and not positioned well for direct use in mobile augmented reality apps, such as our InfoGrid system. The goal is to automate the process of enhancing these 3D objects. In our approach, we monitor users' interactions with a Web-based 3D editor. From these interactions, we derive constraints and show, that for our scenario these parameters can be generalized and applied to other 3D objects, in order to process them automatically in the background. This continues previous work and extends the Network Environment for Multimedia Objects (NEMO), a Web-based framework used as the technical platform for our research project Ambient Learning Spaces (ALS). NEMO is the basis for ALS and among other features provides contextualized access and retrieval of semantic media. In various contexts of ALS, compared to still images or video, 3D renderings create higher states of immersion. We conclude this article with a discussion of our findings and with a summary and outlook.

Keywords—Mobile media; Web frameworks; Ambient Learning Spaces; 3D object editor.

I. INTRODUCTION

Today, people are living in a networked society within digitally enriched environments consisting of various individual interconnections between physical and digital spaces. Together with ambient and mobile technology, these play an important role and affect our being and acting in the physical world. This is also accompanied by the creation of technology-assisted environments, like through the creation, manipulation and visualization of 3D objects [1].

In our research project, with *Ambient Learning Spaces (ALS)* [2] we create ubiquitous learning environments that feature personal and collaborative learning spaces interconnecting learners with their personal digital media and the media of others. It is assumed and under current research that in context of schools and museums, the relatedness of body and spaces supports the learning process, in which ALS enables learners to interact with media. ALS enables learners to structure information themselves using ambient technology in Web-based applications, which are especially available in mobile contexts on personal smartphones. In this setting, 3D renderings empower imagination, creativity and learning compared to still images or video [3]. This seems to be fostering the construction of useful and sustainable

knowledge. In our research projects funded since 2008 by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), we developed an ALS framework to create mixed reality learning spaces where body and physical space are extended by digital artifacts in the form of media. These media in general and especially mobile and ambient media become carriers of information utilized in various learning contexts. Learners and instructors extend these media by digital properties to construct semantic media [4].

As the backend of ALS, we developed the *Network Environment for Multimedia Objects (NEMO)* as a technical foundation of ALS [5]. NEMO is a Web-based framework and storage for semantic media. It stores media, such as text, still images, video, 3D objects, animations, and audio extended by digital properties organized through ontologies. Using NEMO together with some of the applications from the ALS family provides a digital overlay for physical objects using these semantic media.

Media, such as still images and videos from users stored in NEMO, are used in automated background processes by NEMO to create 3D objects from these media without the need of user interaction [6]. In our approach, we use these 3D objects to enhance learner's experiences by placing them into context, e.g., with the application InfoGrid, with which Web-based Augmented Reality (AR) tours for smartphones can be experienced [7]. Due to the process of automatic 3D reconstruction, 3D objects contain unwanted residues or artifacts and do not align well for the use in AR. With 3DEdit, we enable users to enhance 3D objects manually [1]. 3DEdit is a Web-based editor, which runs on the ALS platform, interfaces with NEMO, and does not require the installation of 3D editing programs on a dedicated graphics computer system.

In this contribution, we continue our previous work [1] [6] and present our latest work on automating the process of enhancing these 3D objects. From the user's interaction with 3DEdit we derive constraints and parameters, which can be applied to other 3D objects, in order to process them automatically. In Section 2, we discuss related work. In Section 3, we illustrate the scenario for this work and in Section 4, explain the realization. In Section 5, we present the latest approach for automatic 3D object enhancement and conclude with a summary and outlook in Section 6.

II. RELATED WORK

The research field of the Semantic Web [8] is still of core interest. It yields "naturally" structured data about the world in a well-defined, reusable, and contextualized manner. The

NEMO framework is related to metadata-driven digital media repositories [9] but is different compared to conventional repositories by delivering more sophisticated features through the NEMO logic [1].

The research field of e-learning is strongly related to our work as well. Although contemporary approaches make use of semantic structures [10], the understanding of the purpose of e-learning systems in general focuses mainly on tracking the learner’s performance, provide learning materials or manage and collect assignments [11]. In contrast, our work focuses on linking educational contents with the lifeworld (Lebenswelt) and thus engaging learners in communicative and collaborative processes by expanding the learning environment through contextualized, personalized, body- and space-related semantic media. For this purpose, NEMO provides means of connecting learning spaces inside and outside of schools, museums, urban spaces and natural habitats. ALS create a motivating environment, in which learning is able to flow and develop in whichever way and direction the learner’s like to follow. These digitally augmented learning environments transpose learning from formal to informal learning. They interlace real-physical spaces of everyday life with digital augmentations that facilitate individual and collaborative learning [12].

The field of human-computer interaction with regard to 3D objects is of interest for this research. We have examined various implementations featuring 3D object editing, also in Web-based applications. The implementations studied have in common to provide features for advanced or professional users and are directed at creating and editing 3D objects. Performing manipulations needed for our scenario would require many complex steps of interaction and would not allow editing 3D objects in a touch-only application as needed in our learning scenarios.

With regard to automatically edit, enhance and orient 3D objects, Fu et al. [13] and also Jin et al. [14] presented related work focusing on placing 3D objects upright on a planar level. This approach gives interesting insight into the algorithmic perception of 3D objects. However, it relies on an equilibrium calculation based on discriminative attributes linking the shape to function, which is assessed by machine learning algorithms. As this method works well with complete and closed “man-made” 3D objects, the properties of these 3D

objects do not match the ones of objects we receive from 3D reconstruction for complex natural environments.

III. SCENARIO

For the following scenario, apart from NEMO, we refer to our currently implemented and usable ALS applications. As a starting point for the scenario, school students already created semantic media during their excursions, using the Mobile Learning Exploration System (MoLES), a smartphone app featuring the task-based creation and gathering of media. From these media, NOC3D has already reconstructed 3D objects, as illustrated in our previous work [1]. These 3D objects are now accessible on the InteractiveWall, an installation of large multi-touch screens typically placed in school foyers or public museum spaces [4], giving access to semantic media and other applications.

After a field trip, Sarah, a 14 year old student prepares a short presentation of her findings from her field trip. For this task, she decides to use media she created with MoLES, one of the mobile ALS apps to collect data and media in the field. She logs on to the InteractiveWall located in the foyer at school. She browses through automatically transferred media from her mobile using swipe gestures on the multi-touch screen. She is surprised to discover that among her media, one sculpture she took photos of is meanwhile available as an automatically created 3D object. Using the InteractiveWall, she views the 3D object in full-screen mode, which is shown in an upright position and in a face-view. Sarah wants to incorporate the 3D object into her presentation. For this purpose, she uses 3DEdit, which is embedded into the InteractiveWall, and does some fine tuning of the orientation of the 3D object. After finishing, she is happy with the rendering and incorporates the 3D object into her presentation.

During the presentation, some of her classmates are astonished that they did not notice the sculpture themselves before. With the help of the mobile application InfoGrid, they are now able to take a closer look at it from all sides. They get curious through Sarah’s presentation on what she thinks how the sculpture “Neighbors in Conversation” is related to the current topics off their project.

Later, Sarah is engaged in another school project and logs into the ALS Portal. She browses again through her media to find something suitable to use for the new project. As she



Figure 1. Two sculptures reconstructed from 186 pictures (left) and 176 pictures (right). Both are not oriented well for use in augmented reality applications and contain artifacts and unwanted surroundings.



Figure 2. Orienting (left) and cutting by placing a volume (right) in the web-based editor 3DEdit we used to monitor the subject’s interactions.

comes across the sculpture of the “Neighbors in Conversation” she notices that a woman sitting on a bench holding her child. She uses the 3D editor module inside the ALS Portal she used before on the InteractiveWall to cut out the child-holding woman to use this 3D element for her new project. She is satisfied that she can use the media again in a quite different context.

IV. REALIZATION

In ALS, we have implemented different components, which have been used as a basis for this work.

A. NEMO and 3D Object Reconstruction

NEMO is a Web-based framework for ALS. The NEMO API provides access for ALS learning applications [1][4]. The NEMO Converter 3D (NOC3D) extends NEMO Logic as a module. It runs in an autonomous mode as a background service without any user interaction required. All data processed are images and videos taken with different camera models. These are stored as semantic media in NEMO and selected by NEMO for the process of 3D reconstruction [6]. NEMO stores the reconstructed 3D object together with semantic annotations of the media used for reconstruction. These 3D object is now available for use through NEMO. Other ALS applications can now retrieve the 3D object, e.g., for further editing, or to view the object in various contexts, such as AR [7], as the scenario outlines.

B. Web-based 3D Editor

As NOC3D cannot determine which specific parts of the footage are relevant for the user, or the actual physical object, the reconstructed 3D objects may include superfluous parts of the physical object’s environment, such as artifacts of the surrounding ground surface. Similarly, the desired focus of the object cannot be extracted reliably from the information available in semantic media. Thus, when viewed, the reconstructed 3D objects may appear with an offset to the side or turned away from the viewer (cf. Figure 1). The intuitive solution to isolate the physical object prior to reconstruction is not an option, as the process of 3D reconstruction requires surroundings to be part of the original footage [6].

For this reason, with 3DEdit we have developed a Web-based application, which offers functionalities to solve these issues for our scenario: (1) a function to cut extraneous parts of the 3D object and (2) a function to reorient the 3D objects, which sets the object’s center and default orientation according to the user’s requirements. 3DEdit (cf. Figure 2) offers a browser-based interface, which can also be used on the InteractiveWall or on mobile devices, like smartphones or tablets. Through this interface connected to NEMO, the necessary functionalities are simplified and automated to the point where they only require a single input by the user. The actual manipulation of the 3D object is conducted inside NEMO, so there is no special hardware required for this purpose.

The manual process of editing a 3D object is illustrated in Figure 2. In order to orient the misaligned 3D object (cf. Figure 2, left), the view is rotated until a suitable angle is found. This angle depends on the use case, in which the 3D object is used. Through experiments, we found that rotating the object toward a head-on view is sufficient for our scenario. Unwanted surroundings can be removed by placing a clipping volume around the area of the object to keep (cf. Figure 2, right). Any polygon outside the volume will be omitted or cut along the selected edge. After the 3D object is aligned and all unwanted surroundings and artifacts have been removed in the editor of 3DEdit, the backend of 3DEdit takes the necessary actions to calculate the final 3D object [1].

V. AUTOMATING 3D OBJECT ENHANCEMENT

We have introduced a technical solution to make the results of the automatic 3D object reconstruction in NEMO usable in other applications, such as the AR app InfoGrid. To achieve this, with cutting and orienting 3D objects we have identified two tasks we simplified into dedicated functions into 3DEdit, which do not require the complexity and extent of user interactions like typical 3D editors.

Now our goal is to fully automate the steps in order to enhance 3D objects. For this, we need to automatically (1) detect unwanted surroundings and (2) fix the orientation of the object and (3) perform the process of cutting and rotation.

At first, in the following section we conduct an analysis of the 3D objects at hand and compare our setting with related work in terms of compatibility. In Section V.B, we focus on our setting, in which we monitor interactions of 31 subjects with 3DEdit. In Section V.C, we detail our observations from the monitoring setting and in Section V.D, derive constraints to develop an automated algorithm for our goal.

A. Analysis

At first, a general classification of the 3D objects we receive from NEMO is necessary before proceeding further. In our ongoing research project, all NEMO-generated 3D objects have in common to be of a minimum of 15 cm and a maximum of 2.5 meters in height and width. The figures inside the 3D object can be classified as statues or of similar shape, meaning that the figures are generally standing on a planar surface and this surface is part of the 3D object. The figures are standing freely, which means they are not part of a larger coherent construction (cf. Figures 1-2). Consequently, these 3D objects have a base or similar, stand in an upright orientation. Additionally, sculptures typically have a front side, which can be found by rotating around the vertical axis in an upright perspective.

In the ideal case, the artifacts from reconstruction are only small parts of the objects, but from the experimental reconstruction of a total of 51 real objects we find found out that most reconstructed 3D objects differ from the ideal case, as illustrated in Figures 1-2. We have not defined a measure for the amount of artifacts, but they contain a considerable amount of artifacts, which do not belong to the figure to be preserved inside its digital representation.

The approaches of the related work analyze the given 3D objects by geometrical and other mathematical means [15]. In these approaches, shapes and orientation are analyzed and abstracted into orientation candidates, e.g., according to static equilibrium [13]. For each candidate, a set of discriminative attributes link the shape to function. In this approach, an assessment function of these attributes is learned from a training set using a combination of Random Forest and Support Vector Machine classifiers [13]. As the reconstructed

3D objects however may contain artifacts that are actually larger than the preferable figure itself, this algorithm cannot be used.

In this ongoing work, our approach is to monitor user interactions for typical patterns and derive constraints for 3D objects that yield heuristics, which we can transform into parameters for the automation of 3D object enhancement. Thus, in the following, we describe the setting, in which we monitored the users of 3DEdit in an evaluation with a total of 31 subjects.

B. Monitoring Setting

We conducted the monitoring of user interactions in an evaluation setting, which corresponds to our scenario. We simultaneously validated the technical system of the integration of 3DEdit into NEMO. Concerning the use of 3DEdit in a school environment running on the InteractiveWall installed in, e.g., a crowded and noisy school foyer, it is crucial that the task of enhancing the 3D object is accompanied by comfortable and predictable interaction. For this reason, our evaluation strategy focuses on the intuitive stage and hence, apart from the minimal on-screen help and a short introduction, no instructions, explanations and demonstrations of the usage and functions or techniques were presented to the subjects. The subjects used computer systems provided by us and were free to choose between mouse, touchpad or on-screen multi-touch interaction. 3DEdit runs in a Web-browser and logs every step of user interaction through NEMO. The software was deployed on our servers in a live multi-user setting. For this setting, we have chosen the 3D object illustrated in Figure 1 (right). The subjects had two tasks, which were to (1) remove artifacts from the 3D object and (2) rotate it into head view, as illustrated above. In the beginning, the subjects saw a short introduction explaining the system and their task description, which was displayed on-screen. From this view, they were directed to 3DEdit, which initially shows the 3D object in its default orientation and the on-screen help.

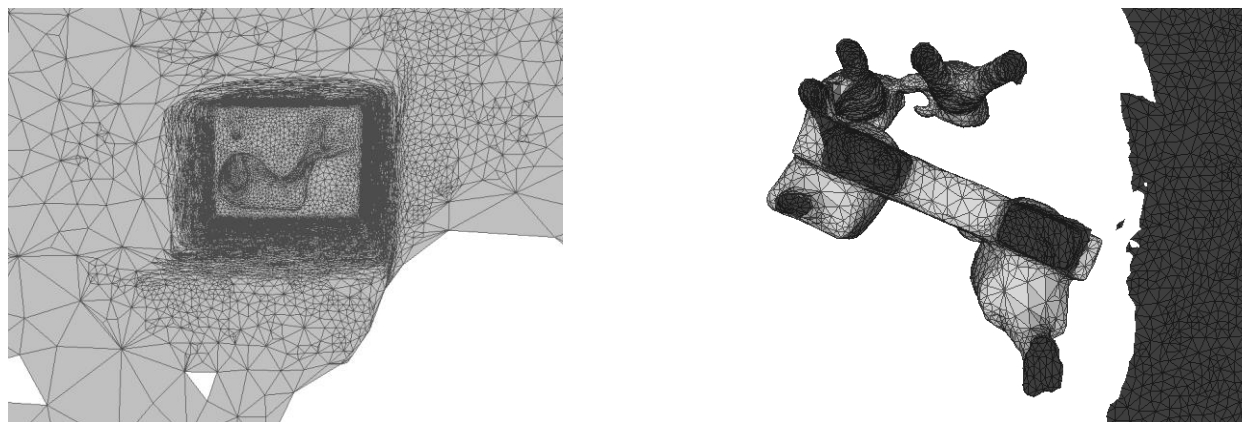


Figure 3. Polygon model of the statues introduced in Figure 1. By detecting an enclosed hollow, our approach determines the statues upright position. Implicitly, the base level is found. A threshold is necessary to identify an unlevelled base of the statue (right).

C. Observations

From the data collected during monitoring in the setting we describe above, we observed that 29 subjects rotated the 3D object into the bird’s perspective. In this view, they used the volume function and placed the square volume around the foundation of the figure in the 3D object. After placing the volume, 24 subjects rotated the view around the object multiple times. 17 subjects returned to the bird’s perspective and changed the volume selection. The other 5 subjects continued straight away. Monitoring data reveals that the subjects used a mean of 5 (min: 4; max: 12) rotating interactions, in order to rotate the object into the bird’s perspective.

Placing the object in an upright position was achieved by 26 subjects. Tracking data reveals that they rotated the object multiple times. 22 subjects oriented at first and rotated the figure until it was placed correctly on the horizontal layer. The other 4 subjects rotated the figure multiple times without any obvious pattern. From the 22 subjects, 19 continued by using the orientation tool to point at the figures’ front and saved the front-facing figure as default view for the 3D objects. The other 3 subjects were not able to complete the task successfully. Monitoring data reveals that the 19 successful subjects used a mean of 8 (min: 2; max: 14) rotating interactions, in order to rotate the object into the front facing view.

For an experienced user, with a desktop system, the interaction of rotating and cutting takes around 30s. The successful subjects completed the task in a median time of 116s (min: 29s; max: 755s). From those using a touch-screen, the median was slightly higher with 201s. This time can potentially be saved using an auto-enhance process running in the background.

D. Findings

From the observations, we derive the following algorithm, described here in functional pseudo-code.

```
rotateToBirdsView();
findBaseContours();
cutObject();
rotateUpright();
```

The bird’s view plays a vital role in the process to remove unwanted surroundings. Rotating an object to the bird’s perspective can be achieved by rotating the object until the maximum depth of an enclosed hollow is found. This enclosed hollow is formed by the figure, which’s surface forms the hull of the enclosed hollow (see Figure 3). Flipping the camera coordinates to the opposite position gives the bird’s view. From our lab experiments we derive that only 3 out of 51 3D objects cannot be rotated using this heuristic, as the unwanted surroundings in 49 reconstructed 3D objects form layers and surfaces without enclosed hollows.

From the bird’s perspective, the subjects recognized the contours of the figure inside the 3D object and consequently placed the volume around the contours. For an automated algorithm, we can derive that from each point of the volumes base drawn around the contours of that figure, a clipping search for polygons directed inwards along the x- and y-axis yields a value, which represents the distance of that point on

the volume’s base to the contours of the figure. As the base of a figure in 51 reconstructed objects is never totally flat, a threshold to detect the contours of the figures base is necessary to take into account. The result of this step of the algorithm is the placement of a volume from the direction of the bird’s perspective camera on to the base of the volume. As the volume’s projection across the entire object at intersects with the ground level, the figure is assumed to be standing on, a check for orthogonality verifies whether the object’s contours for the volume’s base was detected.

In the next step, the 3D object is cut according to the given volume. This method is part of 3DEdit and already implemented using Blender [16] to perform all necessary operations on the 3D object itself. Figure 4 illustrates the ideal result.

In the next step, upright rotation is achieved by transforming the 3D object according to its horizontal base and the orthogonal camera view.

In the last step, our observations yield many patterns of user interaction in order to find the face view of the 3D object. However, although 19 subjects were successful in reorienting the 3D object, we were not able to derive a general pattern an algorithm can perform without any user interaction in an ambient context. In this final step, it is again not possible to use approaches of our related work, as, e.g., the figures base is not closed. In a previous step, we exploited these characteristics of the figure in order to find its bird’s view. Apart from the figure’s base, the related work stops at rotating a figure upright and terminates in any upright position. In an initial approach, we will consider means of facial recognition, in order to find the front view of a certain class of 3D objects. Another approach would be to calculate different possibilities as preview options and offer the user a choice between the most likely options. When selected, the calculations are performed automatically.

This approach is supposed to work on figures or statues or objects of equivalent characteristics. For any other type of object characteristics, other means of automation still need to be evaluated.



Figure 4. Ideal result. This manually edited sculpture serves as reference for our approach to auto-enhance 3D object enhancement.

VI. SUMMARY AND OUTLOOK

Through NEMO, semantic photographic media created by students using their smartphones in mobile context is automatically converted into 3D objects for further use in Ambient Learning Spaces and especially the ALS mobile apps. Although we developed a solution to enhance these automatically created 3D objects in only two steps of cutting and rotating using 3DEdit, in order to remove artefacts or cut out a certain part of a 3D object, this process still requires user interaction.

In this contribution, we present our initial approach to automate the process of enhancing 3D objects in the ambient context of our scenario. Using our frontend editor 3DEdit, we illustrate how we use observations from user interactions to get valuable data in order to define a first approach to automate the process of enhancing 3D objects from 3D object reconstruction. In our ongoing work of our research project, we focus on the implementation and validation of our algorithm.

It is our hypothesis that formal, non-formal, and informal learning through ALS media in a digitally enriched learning space fosters cognitive skills and knowledge in the communicative environment [4] of ALS. We are going to evaluate this in more detail in within schools as well as in museums. Therefore, 3D objects play a vital role in our research. We plan to evaluate their values in a digitally enriched learning environment, e.g., with our school and museum project partners.

For our museum project partners, we have developed scenarios integrating 3D objects into exhibitions with special focus on the autonomous use inside the museums. In this context, the curators experience in the selection process of 3D objects to augment an exhibition can be enhanced with 3DEdit. Therefore, 3DEdit offers a solution to enable curators to edit 3D objects from any source [17]. In our further work, we are going to evaluate the curators' user experiences using our ALS systems.

For any ALS application of the research project, among other features, NEMO provides persistent storage of semantically Enriched Media. The ALS applications are featuring the creation, presentation, use and interaction with such Enriched Media. Together with our project partners, two schools and two museums located in the Hanseatic City of Luebeck, Germany, the use of these applications together with NEMO in context of mobile and ambient learning is currently being further evaluated.

ACKNOWLEDGMENT

We develop NEMO and ALS applications in the ongoing research project "Ambient Learning Spaces" supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). We also thank our school and museum project partners for their kind support.

REFERENCES

- [1] D. Bouck-Standen et al., "Reconstruction and Web-based Editing of 3D Objects from Photo and Video Footage for Ambient Learning Spaces," *Int. J. Adv. Intell. Syst.*, vol. 11, IARIA, pp. 94-108, 2018.
- [2] T. Winkler, F. Scharf, C. Hahn, and M. Herczeg, "Ambient Learning Spaces," *Educ. a Technol. World Commun. Curr. Emerg. Res. Technol. Efforts*, pp. 56-67, 2011.
- [3] K. Persefoni and A. Tsinakos, "Use of augmented reality in terms of creativity in school learning," *CEUR Workshop Proc.*, vol. 1450, September, pp. 45-53, 2015.
- [4] T. Winkler, D. Bouck-Standen, M. Ide, A. Ohlei, and M. Herczeg, "InteractiveWall 3.1 - Formal and Non-Formal Learning at School with Web-3.0-based Technology in Front of Large Multi-touch Screens," in *EdMedia: World Conf. on Educational Media and Techn.*, Washington, DC: AACE, Outstanding Paper Award, pp. 1317-1326, 2017.
- [5] D. Bouck-Standen, "Construction of an API connecting the Network Environment for Multimedia Objects with Ambient Learning Spaces," Master Thesis, Lübeck, Germany, 2016, doi: 10.13140/RG.2.2.12155.00804.
- [6] D. Bouck-Standen, A. Ohlei, V. Daibert, T. Winkler, and M. Herczeg, "NEMO Converter 3D: Reconstruction of 3D Objects from Photo and Video Footage for Ambient Learning Spaces," in *AMBIENT 2017 - Best Paper Award; The Seventh Int. Conf. on Ambient Computing, Applications, Services and Technologies*, IARIA, pp. 6-12, 2017.
- [7] A. Ohlei, D. Bouck-Standen, T. Winkler, and M. Herczeg, "InfoGrid: Acceptance and Usability of Augmented Reality for Mobiles in Real Museum Context (In Press)," *Mensch und Comput. 2018 - Work.*, pp. 1-8, 2018.
- [8] J. H. and O. L. T. Berners-Lee, "The Semantic Web," in *Scientific American*, pp. 30-37, 2001.
- [9] F. Nack, "The future in digital media computing is meta," *IEEE Multimed.*, vol. 11, no. 2, pp. 10-13, 2004.
- [10] P. Bouquet and A. Molinari, "A New Approach to the Use of Semantic Technologies in E-Learning Platforms," *Int. J. Adv. Corp. Learn.*, vol. 9, p. 5, 2016.
- [11] M. Dougiamas and P. Taylor, "Moodle: Using Learning Communities to Create an Open Source Course Management System," *Research.Moodle.Net*, January, pp. 1-16, 2003.
- [12] R. Arnold, "How to Teach without Instructing: 29 Smart Rules for Educators," *Rowman & Littlefield*, Lanham, 2015.
- [13] H. Fu, D. Cohen-Or, G. Dror, and A. Sheffer, "Upright orientation of man-made objects," *ACM SIGGRAPH 2008 Pap. - SIGGRAPH '08*, pp. 1-7, 2008.
- [14] Y. Jin, Q. Wu, and L. Liu, "Unsupervised upright orientation of man-made models," *Graph. Models*, vol. 74, no. 4, pp. 99-108, 2012.
- [15] B. Jiang, M. Zeng, and X. Liu, "An upright orientation detection algorithm for 3D man-made objects based on shape properties," *Journal Comput. Des. Comput. Graph.*, vol. 25, pp. 1099-1106, 2013.
- [16] Blender. [Online]. Accessed 27 July 2018. Available: <https://www.blender.org>
- [17] A. Ohlei, D. Bouck-Standen, T. Winkler, and M. Herczeg, "InfoGrid: An Approach for Curators to Digitally Enrich their Exhibitions," *Mensch u. Comput. 2018 - Work.*, pp. 345-352, 2018.