# Reconstruction and Web-based Editing of 3D Objects
# from Photo and Video Footage for Ambient Learning Spaces

David Bouck-Standen, Alexander Ohlei, Sven Höffler, Viktor Daibert, Thomas Winkler, and Michael Herczeg

Institute for Multimedia and Interactive Systems
University of Luebeck
Luebeck, Germany
email: [bouck-standen, ohlei, daibert, winkler, herczeg]@imis.uni-luebeck.de, sven.hoeffler@student.uni-luebeck.de

*Abstract*—**In ambient and mobile learning contexts, 3D renderings create higher states of immersion compared to still images or video. To cope with the considerable effort to create 3D objects from images, with the NEMO Converter 3D we present a technical approach to automatically reconstruct 3D objects from semantically annotated media, such as photos and more importantly video footage, in an automated background process. Although the 3D objects are rendered in a quality acceptable for the scenario presented in this article, they still contain unwanted surroundings or artifacts and will not be positioned well for, e.g., augmented reality applications. To address this matter, with 3DEdit we present a web-based solution allowing users to enhance these 3D objects. We present a technical overview and reference pedagogical background of our research project Ambient Learning Spaces, in which both the NEMO Converter 3D and 3DEdit have been developed. We also describe a real usage scenario, starting by creating and collecting media using the Mobile Learning Exploration System, a mobile application from the application family of Ambient Learning Spaces. With InfoGrid, a mobile augmented reality application, users can experience the previously generated 3D objects placed and aligned into real world scenes. All systems and applications of Ambient Learning Spaces interconnect through the NEMO-Framework (Network Environment for Multimedia Objects). This technical platform features contextualized access and retrieval of media.**

*Keywords—Mobile media; Mobile learning; Ambient Learning Spaces; 3D Conversion; 3D Editing.*

## I. INTRODUCTION

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

Contemporary pedagogical approaches follow the assumption that humans learn individually and during all of their life. In our research context, one important goal is to offer these ubiquitous learning environments; we called them *Ambient Learning Spaces (ALS)*, as described by Winkler et al. [2]. The relatedness of body and space supports the individuality of the learning process. Together with the loss of spatial distance and the exponentially growing quantity of

information, this induces new technical requirements, as a single individual is no longer capable of consuming and structuring the globally and permanently available information in its entirety [3]-[5]. ALS enables learners to structure information themselves using ambient technology in web-based applications like in mobile contexts on their smartphones. This seems to be fostering the construction of sustainable and mindful knowledge. In this setting, 3D renderings empower imagination, creativity and learning compared to still images or video [6]. In our research project funded for more than seven years by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), ALS consist of a mixed reality where body and space are extended by digitally artifacts. These are represented by peripheral, tangible, mobile, and wearable media [2] and are illustrated in Figure 1, which shows the proximity relationship of each class of media to a learner in ALS.

In ALS, media in general and especially mobile media become the carrier of information utilized in various contexts [7]. This supports the following learning objective: the learner creates contextualized and personalized media stored as digital data, which can be enriched by digital properties. Together, media and their digital properties form will be called *enriched media*.

For the backend of ALS, we have developed the *Network Environment for Multimedia Objects (NEMO),* which is used
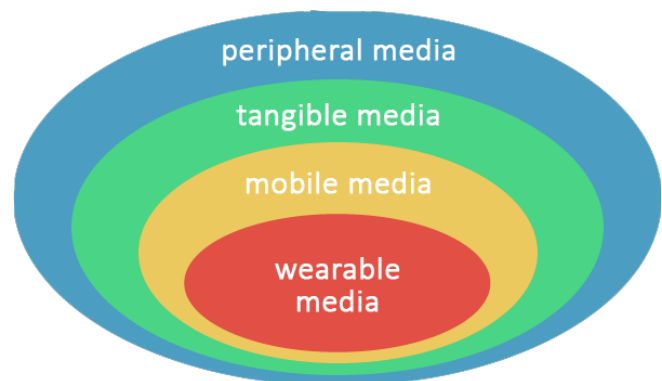


Figure 1. ALS shell model of media visualizing the proximity relationship between media classes and the learner [2]. The system concept of ALS is based on this structure.

as technical foundation of ALS in its latest implementation [8] based on the original concepts of Lob et al. [9]. NEMO is a web-based framework serving, among other capabilities, as a repository for enriched media. For this purpose, NEMO stores media such as text, still images, video, 3D objects, animations, and audio. Digital properties are stored in dynamic sets of semantic annotations and extend these media [8]. Apart from serving as a media repository, NEMO provides computational logic through a service-based interface for other applications developed for ALS. Using NEMO together with these *ALS Applications* provides a digital overlay for physical objects using enriched media within ALS.

Together with ALS applications, the NEMO framework stores media from users as enriched media, as outlined. These enriched media also contain images or video footage, which often originate from the same physical objects, only differing in visual angle, lightning or framing. With the *NEMO Converter 3D (NOC3D)* this paper presents a solution to make use of such footage collectively for example created by school students during field trips, converting these media into single 3D objects automatically without the need of user supervision. Additionally, the learner's experience of digital 3D worlds can further be enhanced by using the 3D Editor for ALS (3DEdit) making the use of a 3D editing program on a dedicated graphics computer system obsolete.

In this contribution, in Section II, we regard related work. In Section III, we present a real scenario using a smartphone application from the ALS application family. In Section IV, we introduce these ALS applications in more detail. The NEMO-Framework is outlined in more detail in Section V. In Section VI, we focus on the realization of NOC3D and describe 3DEdit in more detail in Section VII. In Section VIII, we present our findings and conclude with a summary and outlook in Section IX.

## II. RELATED WORK

Semantic media comprises the integration of data, information and knowledge. This relates to the Semantic Web [10] and aims at allowing computer systems as well as humans to make sense of data found on the web. This research field is of core interest since it yields naturally structured data about the world in a well-defined, reusable, and contextualized manner.

The field of metadata-driven digital media repositories is related to this work [11] as well. Apart from the goals of delivering improved search results with the help of meta information or even a semantic schemata, the NEMO framework distinguishes itself from a mere repository by containing and using repositories as internal components, delivering more sophisticated features through the NEMO logic described below.

NEMO facilitates collecting, consuming and structuring information by interacting device-independently with enriched media, whereas the linked data research targets sharing and connecting data, information and knowledge on the Web [12].

Various implementations exist in order to reconstruct 3D objects from photographic images. Those we have examined in our work have in common that they are not integrated into a fully automated web-based framework making use of semantically annotated data in mobile contexts providing background services for ambient learning environments.

In addition, various implementations exist in order to edit 3D objects, also in web-based applications. However, the implementations examined provide features for advanced and professional users and are directed at creating and editing 3D objects. Performing manipulations required 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.

In the research field of e-learning, other work connecting semantic structures with learning can be found [13]-[15]. In contrast, our work focuses on linking educational contents with the living environment (Lebenswelt) and thus engaging learners in communicative processes through contextualized and personalized enriched media. For this purpose, NEMO provides means of connecting formal and non-formal learning inside schools or outside of schools like in museums. NEMO is not used to examine or track the learner's performance, provide standard learning materials or collect homework, such as Moodle [16].

## III. SCENARIO

Michelle, a fourteen year old student, joins a field trip through the Hanseatic City of Luebeck at school. Prior to the excursion, Michelle's teacher prepared some exercises for the students with the help of the ALS-Portal. They have to answer questions like "*What is communication?*" using the ALS application MoLES. While exploring the city, Michelle answers this question with MoLES running on her smartphone. Michelle uses MoLES and takes photos and tapes videos of what she thinks is related to the question at hand. In this case, for instance, she discovers a sculpture of four adults, two standing next to each other, and two sitting on a bench. From their body language, it seems that they are talking to each other. Michelle uses MoLES on her smartphone to take a few photos and to record videos. For every medium she creates, Michelle also takes notes using MoLES in form of some keywords and sentences to remember her thoughts later on. MoLES uploads these enriched media automatically into NEMO over a secure connection.

Back in school, Michelle prepares a short presentation of her findings from the field trip. In the meantime, Michelle found out, that the sculpture she took photos of is called "Neighbors in Conversation" and belongs to a series of artistic work. She logs on to the InteractiveWall located in the foyer at school. She browses through her media using swipe gestures on the multi-touch screen. Among her media, she discovers that the sculpture she took photos of is meanwhile available as an automatically created 3D object. Using the ALS InteractiveWall, she views the 3D object in full screen mode and notices that the sculpture is shown from its backside. She rotates the 3D object using swipe gestures until she is satisfied with the position. Looking at the sculpture from a bird' eye view, she notices that there are disturbing artifacts around the sculpture. Using 3DEdit, which is embedded into the InteractiveWall, she cleans up the model

Figure 2. School students use MoLES to take photos and record videos to answer questions during a field trip.

by selecting only the sculpture and thereby removes the artifacts. After finishing, she is happy with the orientation and presentation of the rendering.

Now, Michelle incorporates the 3D object into her presentation. During the presentation, some of her classmates are surprised 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 are astonished to hear from Michelle's presentation on what she thinks is associated with the topic at hand.

Sometime later, Michelle is engaged in another school project and logs on to the ALS Portal. She browses through her media to find something suitable to use for the new project. She again comes across the sculpture of the "Neighbors in Conversation" and notices that a woman sitting on a bench is holding her child. She uses the 3D editor module inside the ALS Portal she used before on the InteractiveWall to cut out only the child-holding woman she then uses in her new project. She is satisfied that she can use the media again in other contexts.

IV. ALS APPLICATIONS

In the following section, we are describing the frontend applications in ALS already mentioned and referred to in the scenario in more detail.

A. Mobile Learning Exploration System

The *Mobile Learning Exploration System (MoLES)* is a mobile ALS application running on smartphones and was originally introduced by Winkler et al. [17]. In a mobile context, students create enriched media to answer given questions for a specific task assigned by their teacher whilst conducting an exploration outside of school. They take photos
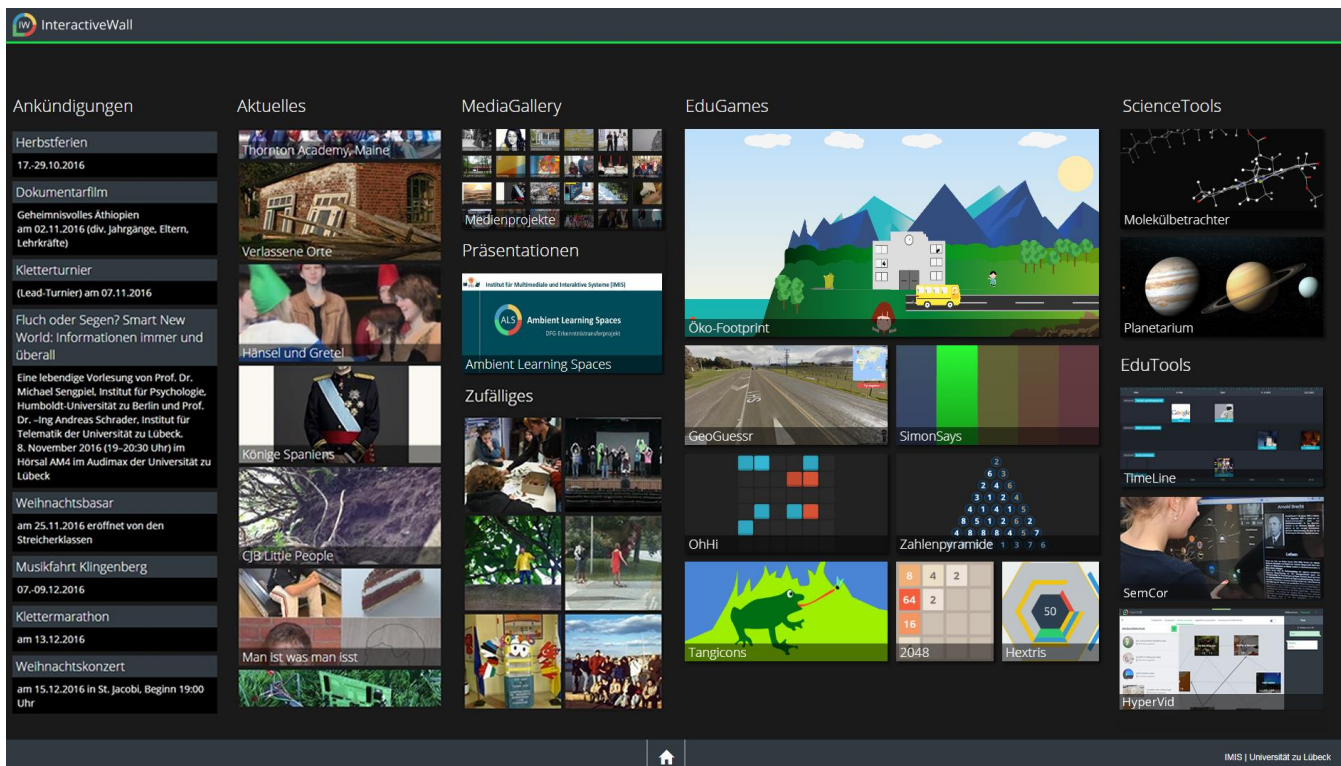


Figure 3. Start screen of the InteractiveWall, taken from one of our project partners. Apart from learning games and educational tools, the InteractiveWall visualizes media created by school students. These are shown in sections 'Latest' as well as 'Random' directly on the schools InteractiveWall. MediaGallery gives access to an overview of the entire media school students have access to.

and record audio or video footage from objects they encounter and add textual notes within MoLES by annotating the media, as illustrated in Figure 2.

After finishing a field trip, the students use the enriched media they created to reflect and present their findings to their fellow students. All media created with MoLES is stored in NEMO.

### B. InteractiveWall

The *InteractiveWall* features an installation consisting of large wall-mounted multi-touch displays [7]. Students use the InteractiveWall to browse their digital media, the media created by others, and present their finding to their fellow students. The InteractiveWall features various applications in context of informal learning in schools, as Figure 3 of the starting monitor depicts and the users interact with the InteractiveWall through touch gestures only.

In our scenario presented in this contribution, Michelle uses the *MediaGallery* to browse the digital media she recorded during the field trip. On the InteractiveWall, among her media she discovers the sculpture she took pictures and recorded video of automatically reconstructed as 3D object by NOC3D and she uses 3DEdit to enhance the 3D object.

The InteractiveWall connects to NEMO. Thus, NEMO serves all media displayed on the InteractiveWall.

### C. InfoGrid

*InfoGrid* is an augmented reality application for smartphones used in mobile context in ALS [18]. InfoGrid recognizes visual markers and detects Bluetooth beacons both triggering the display of images and video, the playback of



Figure 4. Augmented Reality display with InfoGrid. On the left, the sculture "Birds" by Günter Grass and on the right, the skeleton of a blue whale as 3D object is shown.

audio or the augmented reality presentation and alignment of 3D objects. Figure 4 illustrates the display of InfoGrid.

The use of InfoGrid ranges from basic scenarios where markers are augmented by static media to more complex scenarios, where NEMO selects enriched media from its repositories with InfoGrid guiding the users on a dynamic narrative path.
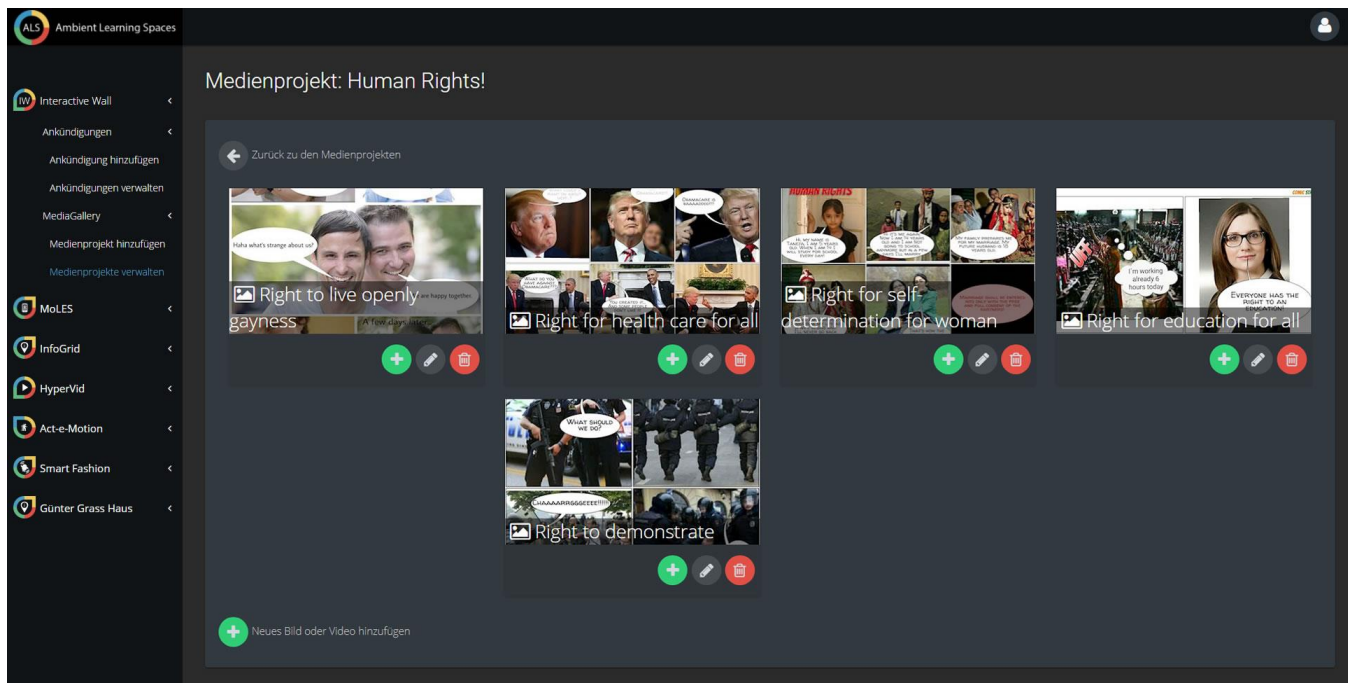


Figure 5. Screenshot taken from the ALS Portal of a school. The left menu offers a module selection. The current view shows an overview of a media gallery for the InteractiveWall. The media can be created, arranged, edited and deleted from this view by teachers or students.

In our scenario, school students use a single marker to view 3D objects automatically reconstructed by NOC3D in their classroom. For this, NEMO maps multiple media to this marker and delivers them to InfoGrid.

### D. ALS Portal

The *ALS Portal* features the management of enriched media [6]. The ALS Portal is a modularized web-based platform. For each ALS application, a dedicated module within the ALS Portal allows to manage these enriched media, depending on the user's access rights and permissions.

As illustrated in Figure 5, the ALS Portal allows editing, e.g., a media gallery within the module MediaGallery. Media entered here is stored in NEMO together with semantic annotations an also made available by NEMO in other contexts of ALS.

In our scenario, Michelle is able to browse the media recorded with MoLES inside the ALS Portal. She also uses the editor 3DEdit to enhance the 3D object of the sculpture of the "Neighbors in Conversation".

All ALS applications interconnect through NEMO. This means, that enriched media created with MoLES, is also available on the InteractiveWall, can be edited using the ALS Portal.

The ALS Portal, InteractiveWall, and MoLES are web-based and implemented as ASP.NET applications. They connect to NEMO via Web Services.

### V. NEMO

NEMO is a web-based framework for ALS. As depicted in Figure 6, the framework primarily consists of three main layers: (1) the NEMO Application Programming Interface (API) layer giving ALS applications access to NEMO, (2) the NEMO Logic layer and (3) the NEMO Core layer. NEMO as well as NOC3D have been implemented in C# running on Windows Server and Microsoft .NET architecture, also making use of the Windows Communication Foundation (WCF) framework.

The NEMO API (cf. Figure 6) provides access for applications such as MoLES, interacting through Web services in an authenticated context over a secure connection. Each application accesses a specific Web service, which achieves a higher layer of transparency and maintainability with regard to the system's architecture. With the NEMO API Client Model, we created a model for a well-defined data interchange between NEMO and any application in ALS through the Web, following the idea of knowledge representation in a formal and explicit way. From experience, we expect any information entered by a learner to be incomplete, as he or she is still engaged in a process of gathering, structuring, and memorizing, thus NEMO is able to handle incomplete and uncertain information [19] in the NEMO Logic and Core layers. Therefore, the model for any client application is independent of any internal model used by the NEMO framework. In addition, this minimizes the learning curve for ALS application development, as no detailed knowledge of semantic modeling is required when developing an application accessing NEMO. NEMO also provides cross-device capabilities [20].
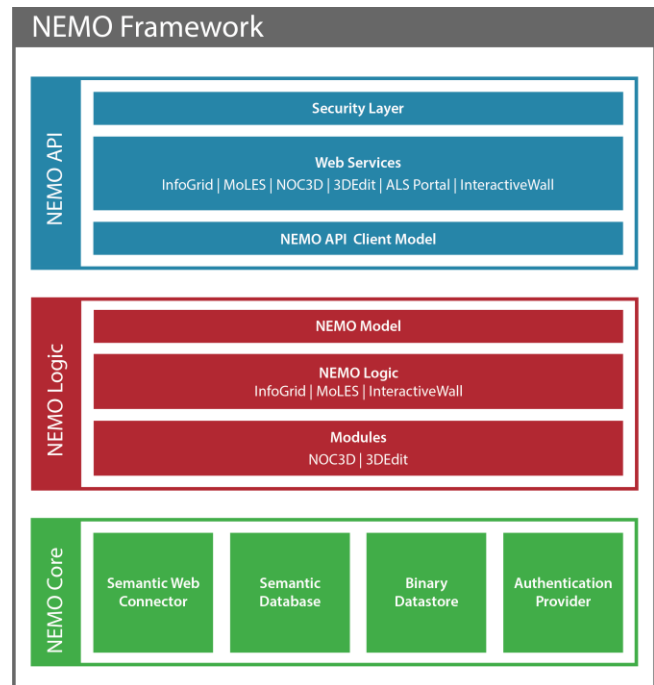


Figure 6. The NEMO-Framework [1]. The NEMO Logic computes, e.g., coherences, semantic models, and data mapping and a modularised interface for feature extendibility.

In the NEMO Logic, we implemented the NEMO Model, which abstracts ALS as a semantic model. Here, the computational logic resides. It initiates and controls semantic search and context analysis in the NEMO Core. In the NEMO Logic, mappings are conducted between the NEMO Model and the NEMO API Client Model through a modified Semantic Object Relational Mapping (SORM). For any Web service, the NEMO Logic holds the specific application logic and thus interconnecting the applications accessing the NEMO framework semantically through an extendable modular structure with loose coupling. We have already developed extensions for NEMO, e.g., the NEMO Converter (cf. Figure 6), which delivers media in device-specific formats and resolution as requested. For research purposes, another extension tracks all requests, actions, as well the corresponding application state of the NEMO framework. NOC3D also extends NEMO Logic. All data collected is stored anonymously due to the sensitivity of the data and legal regulations for public organizations like schools and museums. In a defined context of an evaluation, personal information may be collected synonymously. As we develop NEMO with technical scalability and diversity in mind, NEMO also runs in multiple interconnected instances.

The NEMO framework is based on the NEMO Core where enriched media is stored (cf. Figure 6). A semantic database provides internal storage for any digital entity in the form of semantic annotations. Through the Semantic Web Connector, (cf. Figure 6) any semantic database can be used as internal data store, thus developing applications accessing NEMO requires no knowledge of the respective database query language. Any query result of the internal or any external
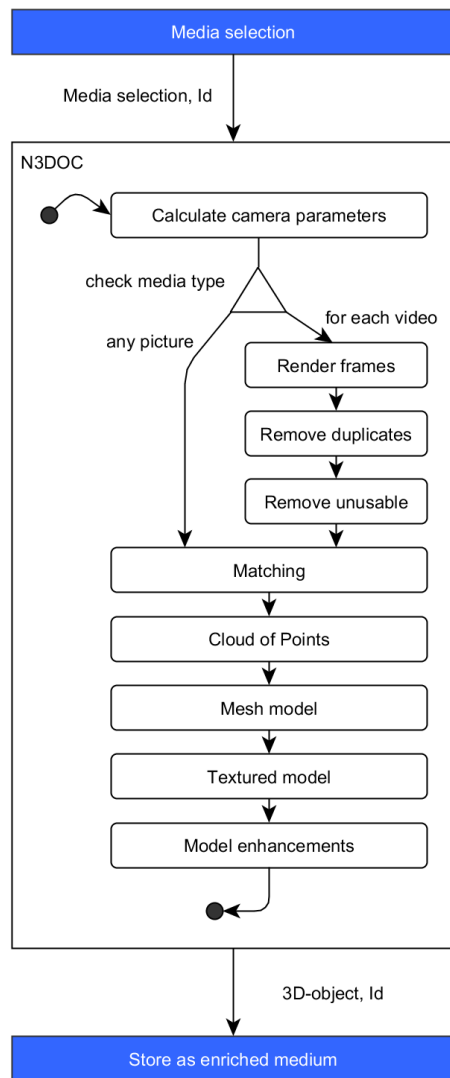
Figure 7. Pipeline of the NOC3D algorithm, as more detailed described in section V.A. Media selection as well as storing the 3D object is performed externally from NOC3D by NEMO.



Figure 8. Screenshot of a 3D object reconstructed with NOC3D from 225 images automatically extracted from semantically annotated videos. The blue background is rendered by the 3D object viewer.

- NOC3D runs in an autonomous mode as a background service without any user interaction required.
- Images and videos are taken with different camera models, mostly with smartphones, from various angles and may contain only sections of the object. Therefore, an input for NOC3D can most certainly not be described as "ideal" or "complete". The cameras have not been calibrated.
- The process of media creation does not require additional markers, only steady surroundings around the object. Every photo or video has to contain surroundings around the object, which is subject to reconstruction.
- An object for reconstruction has dimensions between 5cm and 5m in both width and height.
- Images and videos may not contain multiple objects and only one object will be reconstructed per run.

### A. 3D Reconstruction

In general, the algorithms used in each step and data they require or provide as input and output determine the sequence of steps of the 3D reconstruction process. For our scenario in an ambient context, we have enhanced their combination and derived parameters from the tests we conducted.

At first, from manually entered and automatically generated semantic annotations, such as GPS coordinates, date and time and with regard to different calendrical seasons, NEMO compiles a selection of images and videos, which possibly show the same object which might be reconstructed.

Using GPS coordinates, date and time as well as data from Exif information [21], such as camera make and model, enhances the picture selection in NEMO.

All media selected by NEMO is transferred to the NOC3D module, as shown in Figure 7. As NOC3D provides a web-based API, NOC3D may be set up on a dedicated server, still being part of NEMO [8]. An identifier passed additionally allows NEMO to link the original media with the 3D object after the asynchronous task of NOC3D completes.

Operating on the media selection passed on by NEMO, NOC3D at first calculates camera parameters, which will be used for the process of reconstruction later on. 3D object reconstruction starts by calculating match points of all images and grouping them using VisualSFM [22]. This is necessary

semantic database is mapped into the NEMO Model. In the NEMO Logic, this data will be processed as described above. Binary media is stored in the Binary Storage (cf. Figure 6), which is linked to the internal semantic database in order to retrieve the stored object as enriched media again and also serves as cache in order to reduce on-the-fly conversion time of the NEMO Converter.

An authentication module provides an interface to connect to different authoritative systems in order to check application or media-specific permission settings and user access rights.

### VI. THE NEMO CONVERTER 3D

NOC3D was developed as a component for the NEMO Logic under the following assumptions, which are partly derived from the scenario described above:

in order to find the object for 3D reconstruction in necessary in order to find the object for 3D reconstruction in the images automatically. Every two images with at least 40 match points are grouped. To receive a high quality result from later steps, all images with a resolution below 1200x1200px are discarded at this point. A group with less than 10 images is discarded as well, because these will not be of any use for further processing. We found these parameter values through experimental testing during development. Running VisualSFM on the group of images, until no other image of the selection can be grouped repeating all steps outputs a group of images containing the object for 3D reconstruction.

In the next step, depicted in Figure 7, the Center for Machine Perception Multi-view Reconstruction Software (CMPMVS) [23] calculates the cloud of points using the camera parameters from the first step [24][25]. CMPMVS transforms the point cloud into a mesh model and separately calculates a preliminary texture.

The textured model is handed over to MeshLab [26]. Small artifacts are removed. In addition, MeshLab is used to close polygon gaps in the reconstructed model, remove devious edges and smooth the entire model. The result of an exemplary 3D object is illustrated in Figure 8.

For web-based, mobile device, and browser compatibility, in this stage an additional 3D object is created, in which the number of polygons is reduced to 30.000.

After conversion into a NEMO-compatible file format, NOC3D hands over the completed 3D object to NEMO. NEMO stores the 3D object together with the semantic annotations of the images used to reconstruct the model. The user may have to adjust these annotations, depending on the variety of annotations of the original media. Afterwards, the 3D object is available in NEMO. From here, it can now be retrieved by ALS applications, like the ALS Portal for further editing, or by the presentation applications such as InfoGrid and the InteractiveWall, to view the object in various contexts, as the scenario outlines.

### B. 3D Reconstruction from Video Footage

In general, in the process of 3D reconstruction more images from different angles lead to qualitatively better results. During the development of NOC3D through qualitative evaluation with university students, we found out that taking hundreds of images (cf. Figure 9) of the same object does not integrate well with our usage scenario. In case of an entire class of 20 or more students, who take at least five images of the same object, NOC3D produces acceptable results. However, the challenge of acquiring sufficient footage with low effort remains.

The process of acquiring footage used for 3D reconstruction is simplified by supporting videos as input format. Assuming a video generally consists of at least 24 frames per second, just moving around the object taping a video will produce a sufficient amount of material. Before starting the process of reconstruction, the videos have to be pre-processed, as illustrated in Figure 7. The video frames are extracted frame-by-frame into images using FFmpeg [27] and stored temporarily. This leads to duplicate or similar images,



Figure 9. Example of the media taken as still images or extracted from video footage Michelle took during her field trip. Input images for NOC3D are not expected to be ideal or oriented.

Figure 10. 3D object obtained from automated reconstruction from 176 images through NOC3D. Figure 8 shows exemplary images used in the reconstruction process for this 3D object. This is the default unmodified view, which appears offset. The black background is rendered by the 3D object viewer, in this case 3DEdit.

e.g., when the camera movement around the object is slow. These images do not contribute usable data for the object reconstruction process.

The solution is to remove all duplicates during pre-processing using the imaging library ImageMagick [28]. In addition, unusable images like from overexposed or black frames will be removed.

After pre-processing, all images extracted from the video footage are joined with other images for reconstruction. Our tests indicate, that at least one image (e.g., photo) not taken from a video is required in order to produce acceptable results. The reason are camera parameters stored in Exif data, which are usually not separately stored with each video frame. These are required for the process of 3D reconstruction by CMPMVS. As for our scenario, smartphones used to take photos and tape videos available today produce video footage in similar quality to pictures, which are sufficient for NOC3D, as illustrated in Figure 9.

### C. Running Time Issues

With regard to 3D reconstruction, running time of the module is critical. Preparing the media for processing is performed with linear effort, including extracting usable still images from video as shown in Figure 7. All further steps require significantly more effort, depending on the number of images, the objects complexity and the image resolution [1].

During development, we found that integrating NOC3D directly on the same server with NEMO is unpractical, as 3D reconstruction in general results in high processor (CPU) utilization. Besides, 3D reconstruction performs faster on Graphics Processing Units (GPU) than on CPUs [29].

The solution we implemented is to run NOC3D on a dedicated server. Therefore, we extended NOC3D to connect with NEMO through over Web Services. As a result, prior to reconstruction NEMO transfers all footage to NOC3D, which stores all data temporarily on that server system. The process of 3D reconstruction is started after all footage has been transferred and NOC3D signals NEMO the completion of a conversion process via callback. Because we are using a dedicated server, we are now able to choose CMPMVS as

Cloud of Points algorithm, which only runs on CUDA-enabled (Compute Unified Device Architecture) GPUs.

At this point, we did not further quantify running time or have performed tests with various GPU systems, because running time largely depends on the CMPMVS algorithms. However, the optimizations, like pre-processing images described above and earlier [1] can improve running time on all systems.

3DEdit is implemented in JavaScript, also using HTML and CSS, running inside the client's browser. 3D objects are reduced to 30.000 polygons for viewing purposes. Only the reduced 3D objects are transferred to the client for display in 3DEdit. Their size usually does not exceed 4MB and transfer time depends on bandwidth. On the server-side, only the full-size 3D objects are used in the process of editing. Thus, 3DEdit performs well on current devices. As for the algorithms, with which the full-size 3D objects are manipulated on the server-side, running time performs with linear effort.

### D. Quality of Models from Automated 3D Reconstruction

As Figure 10 illustrates, the quality of 3D objects obtained from the process of automated 3D reconstruction is sufficient for usage with other ALS applications.

However, NOC3D is not capable of automatically placing a 3D object. This is required, if the 3D object was used with, e.g., InfoGrid. As InfoGrid and its augmented reality display places the 3D object on a photographic marker, it is required orienting the object first in order to display it placed correctly on that marker.

In addition, 3D objects might contain unwanted surroundings and artifacts (cf. Figure 11). As the currently implemented process of 3D reconstruction requires surroundings to be part of the original footage in order to work, isolating the physical object is not an option. This observation corresponds to our scenario, as images and video footage are not intentionally taken for the process of 3D reconstruction (cf. Figure 9). This requires tools of manually retouching the automatically generated 3D object, as our scenario illustrates.
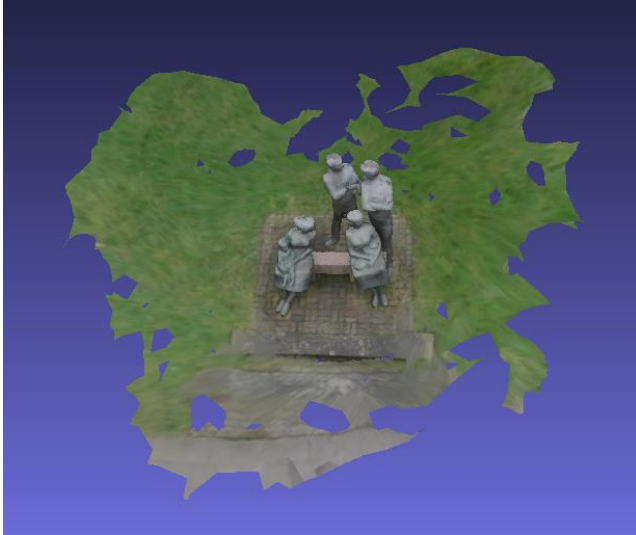
Figure 11. 3D object rotated manually from default view. The 3D object presents some artifacts as unwanted surroundings. The blue background is rendered by the 3D object viewer, in this case MeshLab.



Figure 12. Screenshot of the 3D object in reorientation mode. The users can rotate the 3D object, until they are satisfied. Afterwards 3DEdit saves the changes to a copy, keeping the original 3D object.

## VII. WEB-BASED EDITING OF 3D OBJECTS

As NOC3D is not capable of determining 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 the surrounding ground surface. Similarly, the desired focus of the object cannot be extracted reliably from the information available, e.g., from images or video footage, Exif tags, or semantic annotations. Thus, when viewed in their default orientation, the reconstructed 3D object may appear offset to the side or turned away from the viewer, as illustrated in Figure 10.

For this reason, with 3DEdit we have developed an additional web-based application, which offers two specific functionalities sufficient to solve these issues for our scenario:

(1)     A function to reorient the 3D objects, which sets the object's center and default orientation according to the user's requirements.
(2)     A function to cut extraneous parts of the 3D object.

To allow for editing in a mobile and ambient context, 3DEdit offers a browser-based interface that seamlessly integrates with the ALS Portal or the InteractiveWall and can be used on mobile devices. 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. This way, even an inexperienced user can make the necessary adjustments easily without much effort.

The web-based editor component of 3DEdit makes use of JavaScript in order to display the 3D object. The controls are minimized to the functions presented to users. For the purpose of using 3DEdit as module for both the InteractiveWall and the ALS Portal, 3DEdit uses the CSS style sheets corresponding to the application it is used in. For this reason, some figures in this article show blue controls on purpose.
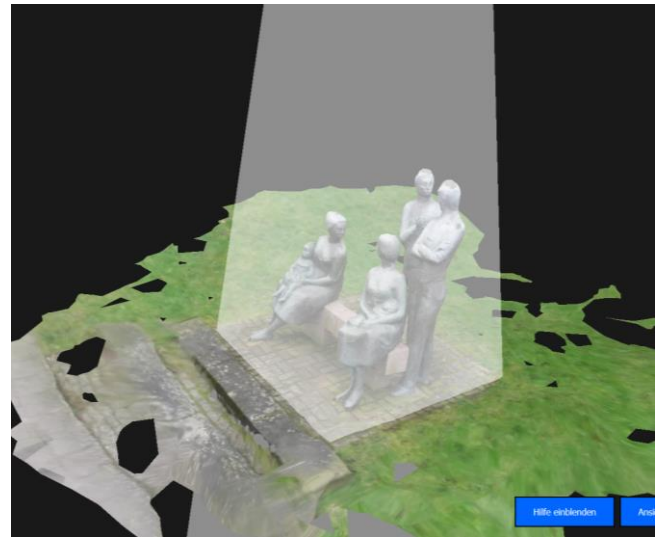


Figure 13. Screenshot of the 3D object in cutting mode. Extraneous parts of the 3D object are outside of the selected volume and not kept after applying the changes using 3DEdit.

Please also note that all screenshots of 3DEdit show 3D objects reduced to 30.000 polygons because of browser limitations. The actual 3D object is kept in NEMO in full size.

### A. 3D Object Manipulation

The actual manipulation of the 3D object is conducted inside NEMO, so there is no special hardware required in addition to the requirements of the ALS Portal and the InteractiveWall. The modules communication is handled via Web Services, which connects the frontend of 3DEdit to its dedicated backend component, which is located in the NEMO

Figure 14. Screenshot of the 3D object after reorientation and cutting is finished using 3DEdit from a web browser.



Figure 15. Screenshot of the 3D object in cutting mode. Also parts of a larger 3D object may be cut out using 3DEdit and use, e.g., in InfoGrid.

Logic layer, as outlined above. Running as part of the NEMO Logic, 3DEdit automatically makes use of the 3D platform *Blender*. This offers a consistent running time as well as a static location for depositing and retrieval of the media by ALS applications through NEMO.

The process of editing a 3D object is illustrated in Figures 12-14. In order to orient the misaligned 3D object (cf. Figure 10), the view is rotated until a suitable angle is found, as Figure 12 illustrates. 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.

All 3D objects reconstructed from footage from physical objects from our scenario such as, e.g., statues and sculptures have in common to be standing on a base level and their height is limited. Considering this, unwanted surroundings can be removed by placing a clipping volume around the area of the object to keep. Any polygon outside the volume will be omitted or cut along the selected edge. On our InteractiveWall, the user places the volume inside the viewport and modifies its size with the help of touch gestures, as Figure 13 illustrates.

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 resulting 3D object, illustrated in Figure 14. At first, the parameters from the user's selection in the web-based interface is transferred to the backend and translated into modification commands. A single user selection of a volume containing the 3D object requires multiple commands executed by 3DEdit using Blender sequentially, in order to manipulate the 3D object accordingly.

### B. Quality and Running Time

Using 3DEdit on 3D objects does not affect the objects quality, as 3DEdit only deletes or clips polygons using Blender.

The original 3D object is not displayed in the browser-based editor of 3DEdit due to limitations. In the backend, within NEMO 3DEdit manipulates the actual 3D object.

The running time for both reorientation and cutting modifications is linear. This supports the use of 3DEdit in our scenario. 3Dedit may also be used in order to extract certain



Figure 16. Screenshot of the cut out part of the 3D object. This becomes available as separate 3D object in NEMO.

Figure 17. Statue called "Dorothea" by the people of the Hanseatic City of Luebeck.
Number of images used for 3D reconstruction, from left to right: 62, 110, 233, 327.

parts of a larger model, as the scenario describes. Figures 15-16 illustrate this process. In terms of running time, there is no difference between removing unwanted surroundings or extracting parts of an existing 3D object. This can alo be observed with 3D objects, which have not been generated through NOC3D, but modeled using professional tools.

## VIII. FINDINGS

In summary, NOC3D produces 3D objects with an acceptable quality given the mobile and ambient context of our scenario in the open standard OBJ-file-format. Due to the automatic process, it is inevitable that 3D objects may contain some surroundings, or will be misaligned with regard to their orientation. With 3DEdit we developed a solution to address these imperfections. This provides a web-based user interface that allows editing of 3D objects in alignment and removal of unwanted surroundings, as well as a backend module, which manipulates 3D objects on the server-side within the backend framework NEMO. 3DEdit can be used with a touch interface or by mouse.

In order to integrate NOC3D in a timely manner as outlined in our scenario, most importantly a multi-GPU system consisting of multiple CUDA-compatible graphic boards is recommended. In addition, free RAM capacity of at least the size of the footage used for conversion as well as hard disk storage of at least ten times the size of the footage for temporary storage is advisable. NOC3D supports images and video footage from different cameras and in different resolutions.

We have taken footage from more than 30 different statues across the Hanseatic City of Luebeck, Germany, and compiled them into different selections according to semantic annotations using NEMO. The footage taken cannot be described as 'ideal', as we cared to take mostly snapshots, e.g., only showing parts of the objects or without optimal lightning that would be used when reconstructing 3D models in, for instance, a laboratory with a special 3D scanner. Thus, our tests reflect media expected to be created by students on a field trip, matching our scenario.

With regard to the usage of MoLES in our scenario, the task illustrated was simplified for this contribution. In a real scenario, the question and tasks are accompanied with pedagogical considerations, which lead to a set of questions for each task, as outlined by Winkler et al. [30].

The process of selecting pictures by semantic annotations is not trivial, as our tests show. It depends on the quality of semantic annotations. Especially with manual annotations, the quality varies. Our observations from school students as well as teachers usage of the ALS Portal show a tendency to repeat and generalize semantic annotations. In many cases, groups of media are annotated with the same semantic annotations. We are currently working on a solution in order to encourage users to use more diverse semantic annotations when creating media in ALS. In a first approach, we additionally save and display annotations automatically created by a computer vision library, which is automatically accessed for each still image uploaded by NEMO. These automatic annotations are displayed as suggestions. However, we still have to evaluate possible impact on the users interaction and the quality of the annotations with regard to their use when selecting pictures.

The image selection process is further enhanced for best results by also using data from Exif tags, which are included in of enriched media, providing data on e.g., camera make, model, date, time and geolocation. This data is also used in the reconstruction process.

Our evaluation shows that, in our scenario, an average minimum of 110 images is required in order to be able to recognize the resulting 3D object as such, as illustrated in Figure 17. The maximum of images is limited by hardware resources, but keeping in mind the time-consuming process of

Figure 18. On the left: Statue "Panther" in a botanic garden in the Hanseatic City of Luebeck. On the Right: The output from 175 photos is hardly recognizable as a panther.

3D reconstruction should be limited to a maximum of 450 images. This value is derived from our experiments in context with our scenario and is depending on the objects complexity, desired quality, hardware capabilities, and the usage scenario. Hence, our research does not focus on optimizing the algorithms employed for 3D reconstruction within NOC3D, but we recommend setting up NOC3D on a dedicated GPU render server.

Further research and tests revealed that the quality of 3D objects is enhanced if the actual object for reconstruction fills about two-third of the entire image or video frame. This is due to the nature of the reconstruction algorithm.

Using footage from symmetric objects especially in front of symmetric or repeating backgrounds often leads to unusable 3D objects, as the example in Figure 18 shows. Using more photos does not necessarily enhance the output.

Generally and as expected, higher resolution of footage as well as using more images results in more detailed 3D objects, but also consuming more time during reconstruction. Nevertheless, using MoLES in contexts like our scenario limits students to the use of smartphones, which is why our primary focus lies on generating acceptable 3D models from smartphone-generated footage.

During our tests, we found that in some cases NOC3D aborted due to a memory overflow. This occurs due to limited hardware resources, exhausted by huge amounts of input data. Apart from upgrading the hardware, our solution is to catch the exception and remove images with the highest and lowest resolution gradually, restarting the process. With this strategy we try to keep as much information on the object and as much high quality footage as possible. This strategy may be optimized.

In total, NOC3D generates all sample models without any unexpected result or malfunctioning. Processing the sample models depends on model complexity and the amount of data to process. Using multiple GPUs is highly recommended.

We also tested the accompanying editor 3DEdit with a variety of 3D objects generated by NOC3D. In every case, it was possible to edit the generated object successfully in a matter of minutes, resulting in cleaned-up and properly oriented objects suitable for further use in other ALS applications. We have also tested 3DEdit with a number of 3D objects created with professional 3D tools containing up to 2.9M polygons.

## IX.    SUMMARY AND OUTLOOK

NOC3D is a module for NEMO that serves fully automated reconstruction of 3D objects from images and most importantly from video footage created in ambient and mobile context and is used for learning scenarios in ALS. Through NOC3D, enriched media collectively created by students using their smartphones in mobile context is converted into 3D objects. Using web technology, we integrate 3D objects seamlessly with applications from ALS which are used in mobile contexts and in context of learning with media. These 3D objects can be enhanced using two steps of the web-based 3D object editor 3DEdit, in order to reorient the 3D object, remove unwanted surroundings or cut out a certain part of the entire object.

It is our hypothesis that learning in a formal and non-formal learning space [7][30], which is digitally enriched through ambient media, fosters cognitive skills and knowledge in a communicative environment [31][32]. We are going to evaluate this in more detail in formal context within schools as well as non-formal context in museums.

3D objects play a vital role in our research. We plan to evaluate their values in a digitally enriched learning environment. In the setting of our ongoing research, InfoGrid will be deployed to our four project partners, two schools and two museums within this year. This will allow us to further develop our scenarios and to examine, what is gained from a learning perspective, by the integration of 3D objects into Ambient Learning Spaces.

For our school project partners, we are currently developing new teaching models implementing ALS technology.

For our museum project partners, we have developed scenarios integrating 3D objects exhibitions with special focus on the autonomous use inside the museums. In the museum context, NOC3D can be also used in the process of reconstruction among others for objects of cultural heritage. Although for this use case dedicated digitization systems exist, like using accurate laser scanning, in the case where only photographic or video material of an object remain for certain reasons, the use of NOC3D is imaginable. In museum contexts, with 3DEdit the curators experience during the process of selection of 3D objects to augment an exhibition can be enhanced, as 3DEdit offers a solution to enable curators to edit 3D objects from any source. From our project partners we learned that this might be the case, when only a special part of a 3D object should be exhibited. In our further work, we are going to evaluate the curators' user experiences using or ALS systems.

Especially in museum contexts, 3DEdit will also be used together with InfoGrid in order to align 3D objects on photographic markers. This gives the curator's a tool to make fine adjustments to the presentation of their digital museum contents.

For any ALS application of the research project, among other features, NEMO provides persistent semantic storage of enriched media. Together with our project partners, two schools and two museums located in the Hanseatic City of Luebeck, the use of these applications together with NEMO

in context of mobile and ambient learning is currently being further evaluated. NEMO is running in multiple instances on-site. The ALS applications are featuring the creation, presentation, use and interaction with enriched media. The applications are developed for various platforms, in desktop, stationary and mobile contexts. Thus, NEMO and the ALS applications are connecting the learner's knowledge with an ambient context, bridging the learning environment and lived-in world (Lebenswelt) to foster sustainable learning and meaningful knowledge.

REFERENCES

[1] 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," *AMBIENT 2017 - The Seventh International Conference on Ambient Computing, Applications, Services and Technologies,* IARIA, pp. 6-11, 2017.

[2] T. Winkler, F. Scharf, C. Hahn and M. Herczeg, "Ambient Learning Spaces," *Education in a Technological World: Communicating Current and Emerging Research and Technological Efforts,* pp. 56-67, 2011.

[3] A. Lugmayr, T. Risse, B. Stockleben, K. Laurila and J. Kaario, "Semantic ambient media - an introduction," *Multimedia Tools and Applications,* vol. 44, no. 3, pp. 337-359, 2009.

[4] M. McLuhan, "Understanding Media: The Extensions of Man," McGraw-Hill, New York, 1964.

[5] A. Whitmore, A. Agarwal and L. Da Xu, "The Internet of Things - A survey of topics and trends," *Information Systems Frontiers,* vol. 17, no. 2, pp. 261-274, 2015.

[6] K. Persefoni and A. Tsinakos, "Use of Augmented Reality in terms of creativity in School learning," in *Make2Learn 2015 workshop at ICEC'15*, Trondheim, Norway, pp. 45-53, 2015.

[7] 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 Johnston, J.P. (Ed.) *EdMedia: World Conference on Educational Media and Technology.* Washington*, DC, USA: AACE, pp.* 1317–1326, 2017.

[8] D. Bouck-Standen, "Construction of an API connecting the Network Environment for Multimedia Objects with Ambient Learning Spaces," Master Thesis, Lübeck, Germany, doi: 10.13140/RG.2.2.12155.00804, 2016.

[9] S. Lob, J. Cassens, M. Herczeg and J. Stoddart, "NEMO - The Network Environment for Multimedia Objects," *ACM (Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia (IITM*

*2010), Dec. 27-30, 2010, IIIT Allahabad, India),* pp. 245-249, 2010.

[10] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web," *Scientific American,* pp. 30-37, 2001.

[11] F. Nack, "The future in digital media computing is meta," *IEEE MultiMedia,* vol. 11, no. 2, pp. 10-13, 2004.

[12] C. Bizer, T. Heath and T. Berners-Lee, "Linked Data - The Story So Far," *International Journal of Semantic Web Information Systems,* vol. 5, no. 3, pp. 1-22, 2009.

[13] S. S. Kusumawardani, L. E. Nugroho, A. Susanto, A. Kumara, H. S. Wasisto and U. Cortés, "Ontology Development of Semantic E-Learning for Final Project Course," *Advanced Science Letters,* vol. 21, no. 1, pp. 46-51, 2015.

[14] M. Masud, "Collaborative e-learning systems using semantic data interoperability," *Computers in Human Behavior,* vol. 61, pp. 127-135, 2016.

[15] P. Bouquet and A. Molinari, "A New Approach to the Use of Semantic Technologies in E-Learning Platforms," *International Journal of Advanced Corporate Learning,* vol. 9, no. 2, pp. 5-12, 2016.

[16] M. Dougiamas and P. Taylor, "Moodle: Using Learning Communities to Create an Open Source Course Management System," *World Conference on Educational Multimedia, Hypermedia and Telecommunications,* pp. 171-178, 2003.

[17] T. Winkler, S. Günther and M. Herczeg, "Moles: Mobile Learning Exploration System," in *Proceedings of Society for Information Technology & Teacher Education International Conference (SITE).* AACE. Charleston, SC, USA, pp. 3230-3234, 2009.

[18] A. Ohlei, D. Bouck-Standen, T. Winkler, J. Wittmer and M. Herczeg, "InfoGrid4Museum: A Medial Strategy for Mediation in Museums using Augmented Reality," *47. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Gesellschaft für Informatik (GI),* Available: http://hci.uni-konstanz.de/downloads/7_InfoGrid4Museums_Ohlei.pdf, 2017.

[19] P. Oliveira and P. Gomes, "Instance-based Probabilistic Reasoning in the Semantic Web," in *Proceedings of the 18th International Conference on World Wide Web.* ACM, New York, pp. 1067-1068, 2009.

[20] D. Bouck-Standen, M. Schwandt, T. Winkler and M. Herczeg, "ELBlocks - An Interactive Semantic Learning Platform for Tangibles," *Mensch und Computer 2016 – Workshopband*, doi: 10.18420/muc2016-ws10-0002, 2016.

[21] CIPA, "Exchangeable image file format for digital still cameras: Exif Version 2.3," in *Standard of the Camera & Imaging Products Association*, Camera & Imaging Products Association, 2012.

[22] C. Wu, "VisualSFM : A Visual Structure from Motion System," 2011. [Online]. Available: http://ccwu.me/vsfm/. [Accessed 20 May 2017].

[23] Center for Machine Perception, Czech Technical University, Prague, [Online]. Available: http://cmp.felk.cvut.cz. [Accessed 20 May 2017].

[24] M. Jancosek and T. Pajdla, *Multi-View Reconstruction Preserving Weakly-Supported Surfaces, CVPR 2011,* IEEE

Conference on Computer Vision and Pattern Recognition 2011, pp. 3121-3128, 2011.

[25] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multi-View Stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 32, no. 8, pp. 1362-1376, 2010.

[26] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Eurographics Italian Chapter Conference*, pp. 129-136, 2008.

[27] F. Bellard, "FFmpeg," [Online]. Available: https://www.ffmpeg.org/. [Accessed 15 January 2018].

[28] ImageMagick Studio LLC, "ImageMagick," [Online]. Available: https://www.imagemagick.org. [Accessed 12 December 2017].

[29] M. Denkowski, "GPU accelerated 3D object reconstruction," *International Conference on Computational Science, ICCS 2013,* vol. 18, pp. 290-298, 2013.

[30] T. Winkler and M. Herczeg, "The Mobile Learning Exploration System (MoLES) in Semantically Modeled Ambient Learning Spaces," *IDC '13 Proceedings of the 12th International Conference on Interaction Design and Children,* pp. 348-351, 2013.

[31] C.-C. Huang, T.-K. Yeh, T.-Y. Li and C.-Y. Chang, "The Idea Storming Cube: Evaluating the Effects of Using Game and Computer Agent to Support Divergent Thinking," *Educational Technology & Society,* vol. 13, no. 4, pp. 180-191, 2010.

[32] M. D. Dickey, *Engaging by design: How engagement strategies in popular computer and video games can inform instructional design,* vol. 53, Kluwer Academic Publishers, pp. 67-83, 2005.