The MuseLearn Platform: Personalized Content for Museum Visitors Assisted by Vision-Based Recognition and 3D Pose Estimation of Exhibits

G. Styliaras[1], C. Constantinopoulos[1], P. Panteleris[2], D. Michel[2], N. Pantzou[1], K. Papavasileiou[1], K. Tzortzi[1], A. Argyros[2], and D. Kosmopoulos[1](✉)

[1] University of Patras, Patras, Greece
{gstyl,kkonstantino,cpapavas,ktzortzi,dkosmo}@upatras.gr, padeler@ics.forth.gr

[2] Foundation for Research and Technology – Hellas (FORTH), Patras, Greece
npantzou@upatras.gr, {michel,argyros}@ics.forth.gr

Abstract. MuseLearn is a platform that enhances the presentation of the exhibits of a museum with multimedia-rich content that is adapted and recommended for certain visitor profiles and playbacks on their mobile devices. The platform consists mainly of a content management system that stores and prepares multimedia material for the presentation of exhibits; a recommender system that monitors objectively the visitor's behavior so that it can further adapt the content to their needs; and a pose estimation system that identifies an exhibit and links it to the additional content that is prepared for it. We present the systems and the initial results for a selected set of exhibits in Herakleidon Museum, a museum holding temporary exhibitions mainly about ancient Greek technology. The initial evaluation that we presented is encouraging for all systems. Thus, the plan is to use the developed systems for all museum exhibits as well as to enhance their functionality.

Keywords: Museum guide system · Recommender system · Pose estimation · Content management system

1 Introduction

Museums are perhaps the most important institutions for the preservation and promotion of the world's cultural heritage and act as powerful learning environments as well, contributing to social and economic development. Traditionally, museums used linear, non-interactive ways of presenting exhibits and guide material, e.g. audio tours or QR codes that offer static information. Multimedia technologies have recently begun to be exploited, but they require considerable investment on the part of museums and often have limitations on the extent of the presented material but also on the number of people who can use them. Also, a key concern for museums has been the evaluation of their exhibitions and activities, as it is critical to understand how satisfied visitors are about the exhibits, the additional digital content and the way guidance material is presented.

Published by Springer Nature Switzerland AG 2020
I. Maglogiannis et al. (Eds.): AIAI 2020, IFIP AICT 583, pp. 439–451, 2020.
https://doi.org/10.1007/978-3-030-49161-1_37

In this paper, we present the initial results of the project MuseLearn, the main objective of which is the development of an innovative guide platform in a museum for mobile devices that will provide personalized additional multimedia content by using exhibit detection, as a basis for providing recommendations to visitors. A key aim of the project is to increase the number of museum visitors and their satisfaction.

The project is implemented in the Herakleidon Museum, Athens [1], which is a museum organising temporary exhibitions with focus on ancient Greek technology. The current exhibitions (2019) deal with the themes of ancient automata (such as the Antikythera Mechanism, and ancient war technology (such as a battering ram). The exhibits are of educational character, in the sense that they take the form of informative material, representations and reconstructions based on original artifacts.

Regarding the presentation of the initial results of MuseLearn, we first present the developed in Sect. 2. In Sect. 3 we discuss existing approaches and work related to these components. Section 4 presents the management of additional multimedia content that is offered to visitors. Section 5 describes the proposed recommender system along with visitor tracking techniques. Section 6 details the exhibit detection technique used in the project. Section 7 describes the implementation of pilot content and initial evaluation

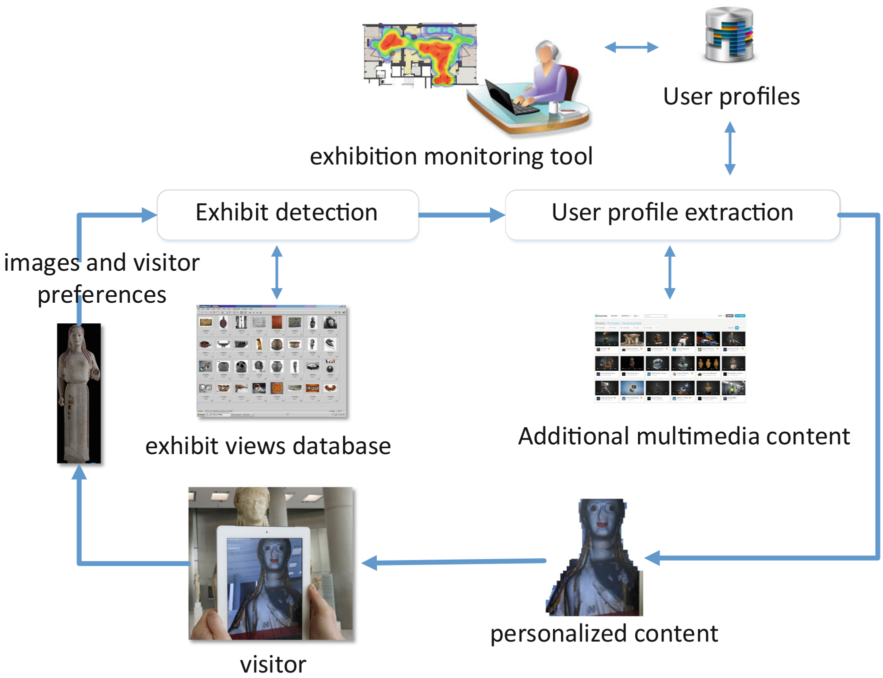

Fig. 1. Outline of the MuseLearn platform. The visitor can see the exhibit through a mobile device. The exhibit is then recognized by the system (based on the acquired image and the stored exhibits) and the appropriate digital material is suggested based on the user's profile. Concurrently the users' choices are recorded. Asynchronously, the curator may analyze the exhibition statistics.

results in all components of the project. Finally, Sect. 8 discusses the key conclusions of the paper and presents plans for future work.

2 Platform Architecture

The MuseLearn platform (see Fig. 1) consists of the following basic systems that are presented in the following sections:

- A **content management system** that stores multiply structured multimedia material that is offered to visitors through augmented reality for better understanding the exhibits of the museum.
- A **recommender system** that allows adaptation of the information to the needs and requirements of visitors.
- An **exhibit detection** and pose estimation system that can detect the exhibit the visitor is interested in and provides identification (exhibit name/id) and orientation information (3D pose) using the camera of a mobile device (phone/tablet).

According to the typical usage scenario, the visitor sees the exhibit via a portable device. The exhibit is then detected by the pose estimation system (based on the images of stored exhibits) and the recommender system suggests appropriate digital material stored in the CMS to the visitor according to their profile. At the same time, visitor choices are recorded for further refining their profile.

3 Related Work

Content Management Systems (CMS): Regarding content management systems (CMS) for museums should offer administrators the ability to structure content in various multimedia formats for artifacts, concepts and supplementary material that is necessary for their presentation [2]. A research of commercial and research CMS that are mainly or indirectly targeted for museums is found in [3]. Argus [4] is a web-based Collections Management Solution (CMS) that can be used for managing and presenting artifacts and objects. Argus provides a tabular visual interface organized in rows for accessing and editing content. The retail oriented RetailPro platform [5] supports responsive design, portable content, reports, transformation of content parameters, reorganization of collections and organization of virtual collections. Content is mainly presented in semantically colored tables. The Museum System (TMS) [6] is a CMS that supports planning and managing of exhibitions and generates reports; it contains a digital asset management module and offers administrative support. MuseumPlus [7] supports managing collections and exhibitions and is in use in over 900 museum sites worldwide. Unlimited images and other multimedia content can be linked to objects, artists, addresses and other record types. CollectiveAccess [8] is a web-based open source CMS that embraces Dublin Core, PBCore, VRA Core etc. The embedded media viewer allows to enlarge and inspect uploaded images, video and audio playback with time-based annotations and PDF viewers. Museum Anywhere [9] offers collection management and tile-based presentation through mobile devices and supports integrating content from other museum management software. The proposed CMS aims to overcome limitations of existing museum CMS by introducing a simple but flexible and expandable structure for organizing content about exhibits and concepts, which can be visualized by using modern web technologies and exploiting the spatial and interface capabilities of mobile devices.

Museum Recommendation Systems: So far, museum recommendation systems are mainly used for supporting personalized museum tours. There are examples of recommendation systems in museums that offer unique tours based on specific interests [10, 11]. The traditional recommendation approaches are (a) Content-based filtering (e.g., [12]) gives recommendations based on similar content that was of interest to the user in the past and (b) Collaborative filtering (e.g., [13]) that are based on content that was of interest to the users that are most similar to the current one based on their selections. Combination of these leads to hybrid approaches. The latent factor models (e.g. [14]) are more recent and try to factorize the matrix that associates users and items to be selected via SVD, or sparse methods. Therefore, hidden relations between users and items may be identified. The Herakleidon Museum can provide a great deal of information to the visitors, in a variety of formats, such as texts, pictures, sounds, videos, games, etc., instead of a comprehensive description of the exhibits. Thus, arises the need for filtering, hierarchy and effective delivery of this information. Our goal is to create a recommender system to manage the large amount of multimedia information accompanying the exhibits of and to recommend visitors to explore the museum on the basis of their interest.

Exhibit Detection and 3D Pose Estimation: Regarding exhibit detection and 3D pose estimation modules, there is a large number of relevant works. Some recent relevant publications address some of the requirements of the MuseLearn pose estimation module. A very significant building block of the overall approach is the capability to localize a camera relative to the scene. Kinect Fusion [15] proposed by Newcomb et al., fuses all the data from a depth camera into a voxel space, that represents a dense 3D model of the scene to be reconstructed. ICP is used with it to track the camera position. More elaborated versions [16] allows it to operate in larger scenes. Although promising, the method requires depth input to operate and can be sensitive to outliers and specular surfaces. These requirements make it a bad fit for the MuseLearn scenarios. Mur-Artal et al. [17, 18] proposed the ORBSlam method. OrbSlam2 performs simultaneous localization and mapping of the environment. It tracks the camera position and builds a map of the environment based on ORB features. The map contains a sparse 3D point cloud, each point being associated with an ORB descriptor, and some geometric information about the view corresponding to it. A set of keyframes (i.e., representative frames) is also maintained in time, and each converted to a bag of words (BoW) descriptor [19]. This approach operates with RGBD or monocular/stereo RGB input, can make use of a static map, and works in real time, making it a possible base solution for our problem. However, the method assumes a static scene and is not robust to moving objects and large numbers of outliers. More recently, deep-learning based methods [36, 37] solve the object pose estimation problem from RGB input using Convolutional Neural Networks (CNNs). The networks are trained to predict the 2D coordinates of the projections of known 3D landmarks on the objects of interest. Subsequently, they solve the PnP problem to acquire the 3D object pose. Although promising, these approaches do not scale well with multiple exhibits and suffer in dynamic environments and in the case of partially occluded exhibits. This renders them unsuitable for the MuseLearn use case. FORTH's proposed method for exhibit identification and pose estimation borrows many ideas from [18], especially for the training phase as it will be detailed in later sections. Lourakis and Zabulis [20] detect the pose of known rigid objects in monocular RGB images, using a 3D model for each object. The model consists of 3D points each associated with a SIFT descriptor. A hand-held camera is used to capture the appearance of the object from all view angles. Then, the camera motion together with the 3D point cloud describing the object is recovered using "Shape from Motion" techniques similar to [21]. The authors report a run time of 0.6 s per frame. While the work of Lourakis et al. is a good candidate for solving our problem, the dense object representation makes it too computationally expensive, and the lack of use of background features makes it harder to detect featureless objects. However, aspects of this work such as the F2P features matching strategy and the pose estimation method (posest library) are used in the proposed method.

4 Content Management System

Modern web technologies have made possible the visualization of content on a spatial area. For example, Scalable Vector Graphics (SVG) and Cascading Style Sheets (CSS) may be employed for implementing vector shapes and animations that are presented uniformly in devices of different sizes and interface options. On the other hand, in a smartphone world, content-editing operations should resemble natural operations by introducing spatial interfaces and gestures and not relying on providing a desktop experience on a smaller screen. The Content Management System (CMS) developed for MuseLearn embraces these ideas for displaying and interacting with content spatially.

Two on-site surveys have been conducted in order to record the information provided by the museum and all kinds of exhibits and concepts. Information about exhibits and their related multimedia informative material has provided the basis for populating the system's database. The database of the CMS stores all information about the museum's exhibits and concepts that they may be assigned to. Database design is flexible enough so exhibits may be assigned to multiple concepts along with other multimedia material linked to them. More specifically, the main entities in the DB are:

- **Concepts:** The concepts that museum exhibits may be assigned to. For example, the exhibit "battering ram" is assigned to the concept "combative weaponry". Concepts also describe all kinds of informative material within the museum.
- **Exhibits:** The main table holding information about every exhibit, such a title and description about the exhibit, the exhibit code within the museum, the museum room where the exhibit lays and the exhibit types.
- **Links:** This table is responsible for defining all possible relations among exhibits and concepts. There may be relations between exhibits; between concepts for defining the concepts hierarchy.
- **Multimedia:** Multimedia content that is linked to an exhibit or a concept.

The database has been implemented in MySQL and content has been inserted about the presentational needs of exhibits and their supplementary content for both museum buildings along with the hierarchical structure of concepts for the certain museum and all kinds of relations among them. More specifically, the database currently stores 98 concepts of all hierarchy levels; 82 exhibits; 152 links of all types and 180 multimedia items for concepts and exhibits.

PHP code has been employed for the hierarchical visualization of museum content starting from the museum level, continuing with the two main collections ("Automata" and "Ancient War Technology") and finishing at single concept and exhibit level. For every concept, its description and multimedia content are displayed along with links to subordinate concepts that analyze it and exhibits that are related directly to the concept. Visualization is both textual, as in a subject catalog, and graph-based. The latter is implemented in SVG, so that it is accessible by all device types. Treant.js [22] free to use, graph creation library has been used for drawing content hierarchies.

5 Recommender System

For the recommender system a trial mode is scheduled to initiate a few users. Gradually and based on data from more users, the system will be retrained. The preferences will arise indirectly from the viewing time of the accompanying material. Also, when entering the museum, the user will have a profile that matches his/her demographics and/or short quote when signing up. Progressively the profile will be completed during the visit and as long as its preferences are known. We currently use alternatively latent factor approaches. For a new item or a new visitor, the new object has no score and the new user has no history, which is known as the cold-start problem. So any recommendation result could be doubtful. We handle it by using demographic data, which has been an effective method for cold-start in the past [23]. Different versions of the same content are associated with some initial and grossly-defined profiles based on a few deterministic rules, which are tested against a questionnaire as described in the following subsection. After some views the system is able to fine-tune the user profile.

The question as to how to identify and distinguish visitor profiles has been a key issue in both the museum studies (e.g., [24, 25]), and information systems literatures (see [2]). For the pilot system, it was decided to create and organize the content, addressing three broad requirements related to visitors' prior knowledge, experience and expectations about the exhibition: 'general interest', 'specialized knowledge', and 'enthusiasm about new technology'. To identify these broad requirements, a group of basic questions forms the starting section of the pilot system. These focus on visitors' demographic profile (sex, age range, highest education level), and their familiarity both with the subject of the exhibition and the use of technology. A sixth key question asks for the driving motivation behind the visit. This question draws on fundamental distinction of five identity-related categories of visitors ('explorers', 'facilitators', 'professional/hobbyists', 'experience seekers', and 'rechargers') [25]. At the pilot stage of the project, visitors' responses to these questions will offer, in combinations, 'personalized' content at the most basic level. For example, a visitor aged 17 and familiar with new technologies will be offered 'general interest' content, enhanced by interactive educational games. Visitors' responses will provide the foundations on which to fine-tune later the recommender system. These six questions are also used for the primary evaluation procedure, as described in Sect. 7.2.

6 Pose Estimation System

The developed exhibit localization and pose estimation method is tailored to the specialized requirements of MuseLearn. Specifically, the inference is performed using RGB input; the solution is scalable to large exhibitions; it supports multiple levels of pose estimation detail; is associated with small (re)configuration overhead; and achieves real time performance. The method is using a "features and reference poses" database that is generated offline, in order to identify the pose of exhibits in query images captured by mobile devices.

6.1 Training Phase

Creating the features and reference poses database is a critical step for the correct operation of the pose estimation pipeline. In order to maximize the accuracy, we employed the use of RGB-D cameras in the training phase. This type of sensors (i.e. MS Kinect2, Asus Xtion, ORBEC) provide apart from the usual red, green and blue channels an additional channel for depth. This channel provides distance information for each imaged point of the scene.

Training Sequence Acquisition: For each exhibit we capture an RGB-D video sequence. The goal is to cover a wide range of possible visitor viewing directions.

Feature and Pose Extraction: We apply a SLAM method to extract features and key frames and camera poses. For the purposes of MuseLearn we chose ORBSlam2 [18] as the SLAM method and the ORB [26] features and descriptor for feature extraction.

Exhibit Annotation: An annotator selects a reference frame, pose and bounding volume for each exhibit in the museum coordinate frame. The common coordinate frame is required in order to report pose information to the other MuseLearn modules.

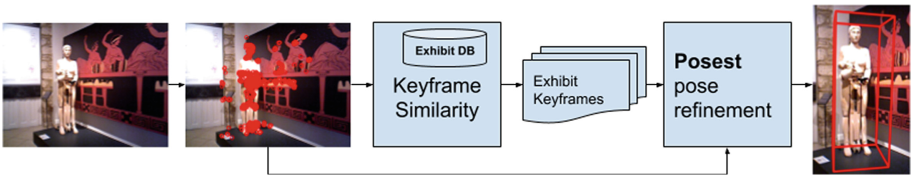

Fig. 2. The pipeline of the pose estimation method during inference. From left to right: (i) the query image captured using the camera on the mobile device (ii) ORB features are detected on the query image and their descriptors are computed (iii) The BoW representation of the ORB features is used to find the most similar (high similarity score) keyframes in the exhibit db (iv, v) Pose refinement using exhibit frames and matching ORB features from the query image (vi) the exhibit pose with the highest score is returned.

Exhibit DB: In the final step we automatically select frames from the training set where each exhibit is fully visible. The ORB features from these "keyframes" are converted into a bag of words (BoW) representation and are inserted in the exhibit database.

6.2 Exhibit Identification and Pose Estimation

The first step in the pose inference pipeline is to compute the 2D features $O(I)$ in the input image I. From these, a BoW representation q is created. We compute the similarity score S between the BoW of the query image and all the keyframes K in the exhibit database. The set C of keyframes that have a similarity score above a threshold S_i (q, K_i) > T_s are the initial "coarse" estimations. If the query from the mobile device requires only exhibit identification (no pose estimation), the exhibit ID of the keyframe in C with the highest similarity score is returned. The pipeline is depicted in Fig. 2.

The pose estimation is a refinement step over C using the features $O(I)$. Initially we compute C' that is the subset of keyframes in C that belong to the identified exhibit. Using the ORB descriptors and matching described in [26] we find the best correspondences $M(O(I), K_j)$ between $O(I)$ and the ORB features of each K_j keyframe in C'. Posest [20] is applied to all matches were $|M| < T_m$. T_m is the minimum number of features on the keyframe that belong to the exhibit (i.e., not in the background) and have a good match in $O(I)$. Posest uses a RANSAC scheme to iteratively select a subset of M and compute the rigid transformation that best explains the camera motion with respect to a given keyframe pose. The quality of the transformation is measured using the reprojection error of the matched ORB features from the query frame to the keyframe. The pose with the highest score is selected as the camera pose with respect to the exhibit.

7 Pilot Implementation and Primary Evaluation Procedure and Results

During the pilot implementation, we had the chance to test the various systems of the MuseLearn platform in the Herakleidon museum. Pilot multimedia-rich content has been developed for certain exhibits. When these exhibits are identified by the pose estimation system through visitors' mobile devices, the additional material will be displayed on visitors' devices based on the suggestions of the recommender system.

7.1 Content for Pilot Implementation

We used the CMS described in Sect. 4. First, a pilot extended scenario of 75 screens for a specific section of the exhibition "EUREKA Science, Art and Technology of the Ancient Greeks" was drafted. Focus was centered on exhibits related to armors and weapons from the Mycenean to the Hellenistic Period. We decided to fully deploy content for five exhibits belonging to the "Warrior and Armor/Equipment" section of the exhibition and for three assigned categories of users/testers (see Sect. 5). The five exhibits were selected on the basis of their associative and historical value, as well as their potency to enhance the storytelling, increase visitor engagement and unfold stories about individual artifacts or specific object groups.

Content creation decisions were taken by accounting for users of different needs, interests and knowledge (see [27]). Hence, two levels of content exist for each exhibit: specialized and basic content, enhanced occasionally by interactive applications. This content includes texts, images, videos, hyperlinks and games. More precisely, as in most guide systems, the user can access descriptions of the exhibits, information about their function and the historical and the social context of their use and production from a variety of sources, as well various interpretations and representations, both scientific and popular (see [2]) by using the content gathered as described in Sect. 4. The system also allows users/testers to zoom in and observe object details, consult the glossary for unknown scientific terms, place names and archaeological periods, and search for related content from the museum collection or from other physical or virtual spaces with the help of hyperlinks and object recommendation. Finally, interactive applications through matching, ordering and sequence games and object identification provide users with the opportunity to test their knowledge and acquire new, get acquainted with typologies, explore objects' correlation/interconnection and interact in a playful way.

7.2 Primary Evaluation Procedure and Results for Visitor Satisfaction

Nowadays, evaluation construes integral component of museum practice and research (i.e. [28–30]), but also of mobile application designing and multimedia development (i.e. [28, 31, 32]). To be beneficial, evaluation should be approached as "an ongoing process" [28]. The primary evaluation procedure was carried out by a group of six undergraduate students from the University of Patras. At this preliminary stage, both qualitative and quantitative methodologies were employed. First, a questionnaire was drafted both in English and Greek consisting of 15 questions and organized in two sections. The aim of this survey is to assess visitors' motivation, familiarity with the exhibition topic as well as information and communications technology, their level of satisfaction with information provided through various interpretive means, as well as their potential interest in using a mobile application during their visit.

The questionnaire is self-administered. Therefore, an introductory note informed the participants about the goals and context of this survey. As expected, the last section collected information about the age, gender, higher level of education and place of origin of the respondents. Initial feedback was encouraging. Visitors expressed their strong interest in using a mobile application to navigate around the exhibition space.

The primary evaluation procedure was complemented by observation of visitor behavior, which has become a key element in feedback studies on museum performance (see for example [24, 33–35]). The aim of this research was to observe how visitors move, explore and use space and display in the exhibition 'Technology of War in Ancient Greece', without the use of the mobile guide system. These data will then be used for a comparative study after the implementation of the platform and potentially lead to a deeper understanding of how the platform impacts on visitor experience.

Visitor behavior, in particular their patterns of moving and viewing, is analyzed using established techniques. First, the arrangement of the display is recorded on the building layout as the basis for designing the observation record sheet for mapping visitors' movement and interactions with the displays. Traces of the paths of visitors, who are randomly selected, spread across time periods and have consented to take part in the research, are recorded for their whole visit to the exhibition. When the visitor stops to look at a work, read a text, or watch a video, a stopping point is recorded on the plan of the exhibition by the observer. Other symbols are used to clarify where a visitor stops for longer periods of time. The tracking data are complemented by the recording of the total time visitors spent in the exhibition. The resulting ratings table which associates users with exhibits and their preferences will be used to initialize the system profiles and thus to address the cold-start problem in our system.

7.3 Primary Evaluation Results for the Pose Estimation System

Regarding the pose estimation system, we evaluated our baseline method using data acquired on two sites. **(i) The AMI facility:** The ambient intelligence (AMI) facility of FORTH hosts exhibition and demonstration areas. We trained our system on six (6) exhibits selected for their size and placement to be compatible with typical exhibits in a museum. Data were acquired with low- and high-quality cameras to investigate performance with different sensors. A training set and two test sets were acquired. The first test set uses the low quality RGB sensor of an Asus Xtion. The second uses a high-quality camera similar to the cameras found in modern phones and tablets. The exhibit locations and reference poses were annotated in the training dataset as well as in a test dataset. **(ii) Herakleidon museum:** We recorded a preliminary dataset of six exhibits. In both cases the exhibit locations/reference poses were annotated in the training dataset. The preliminary version of the pose estimation method focused on the exhibit identification and coarse pose estimation. Specifically, the system must correctly identify the quadrant from which the visitor is approaching the exhibit: Left, Right Front, Back.

The results of our experiments on both sites are given in Table 1 and Table 2.

Table 1. Average precision and recall on exhibit classification accuracy.

	Herakleidon	AMI (low quality)	AMI (high quality)
Average precision	0.947	0.980	0.971
Recall	0.771	0.844	0.810

7.4 Primary Evaluation Results for the Recommender System

We started to evaluate different approaches using public data, until we collected our proprietary ones for the Herakleidon museum. More specifically we evaluated the data provided in [11]. According to this, a dataset of visitors' times spent in each thematically organised exhibit area is created, which were recorded and collected by computer-supported methodology. The recommendations concern specific thematic areas. In order

Table 2. The coarse (quadrant) pose estimation accuracy. Two metrics are shown. Left is the error in degrees relative to the ground truth. On the right we quantize the same error using 3 thresholds. The quadrant approach corresponds to "<45°".

Error in degrees on the yaw axis		% of frames within correct quadrant		
Avg error (std dev)	Median error	<22.5°	<45°	<90°
8.98 ($\pm$7.94)	6.73	93.0%	99.8%	100%

to create appropriate recommendations, the exhibits were grouped into semantic and spatially coherent areas. With this in mind, 126 thematic and naturally organised areas were created at the museum's facilities. The time spent by 158 visitors on these areas of the museum was recorded and expressed implicitly the rating.

The viewing times are transformed to logarithmic with values between 1 and 8. We applied two methods: a simple collaborative filtering with KNN and a latent factor method with SVD. The root mean square error is calculated by ten-fold cross validation. With both methods the mean RMSE is around 1 with standard deviation between 0.02 and 0.03. There are similarities and differences with our system. We similarly use the viewing times as an implicit rating mechanism; the target users are museum visitors as well. However, we provide recommendations on supplementary digital material associated to museum artifacts and not directly on museum artifacts.

8 Future Work and Conclusions

We presented the MuseLearn platform that is constituted of three innovative systems (artifact detection, content, recommendation,). Some initial evaluation results were presented after the implementation of the platform for Herakleidon Museum.

In the future, subsystems of MuseLearn will be extended both in functionality and content for covering all exhibits of Herakleidon Museum and other museums. Content-wise, further support will be implemented for inserting and linking multimedia content, such as 3D representations and virtual reality material. Towards this direction, refined visitor profiles will be developed on the basis of the empirical findings, and their preference for certain themes, in conjunction with demographic data. We also plan the creation of more educational applications based on content that will have been prepared for the presentation of exhibits.

Acknowledgement. Co-financed by the EU and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE (project: T1EDK-00502 - MuseLearn).

References

1. Herakleidon Museum. http://herakleidon-art.gr. Accessed 29 July 2019

2. Kosmopoulos, D.I., Styliaras, G.D.: A Survey on developing personalized content services in museums. Pervasive Mob. Comput. **47**, 54–77 (2018)
3. Lanir, J., Kuflik, T., Sheidin, J., Yavin, N., Leiderman, K., Segal, M.: Visualizing museum visitors' behavior: where do they go and what do they do there? Pers. Ubiquit. Comput. **21**(2), 313–326 (2017)
4. Lucidea, Argus, make your collection more visible and accessible than ever before. https://lucidea.com/argus/. Accessed 29 July 2019
5. RetailPro. http://www.retailpro.com/. Accessed 29 July 2019
6. GallerySystems, Collection management software, museum and art collections. https://www.gallerysystems.com/products-and-services/tms-suite/tms/. Accessed 29 July 2019
7. Zetcom, Museum plus. http://www.zetcom.com/en/productsen/. Accessed 29 July 2019
8. Collective Access. http://www.collectiveaccess.org/. Accessed 29 July 2019
9. Museum Anywhere. http://museumanywhere.com/. Accessed 29 July 2019
10. Keller, I., Viennet, E.: Recommender Systems for Museums: Evaluation on a Real Dataset, pp. 65–71. IARIA, Brussels (2015)
11. Bohnert, F., Zukerman, I.: Non-intrusive personalisation of the museum experience. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 197–209. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02247-0_20
12. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_10
13. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: ACM SIGIR Conference 1999, pp. 230–237 (1999)
14. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. computer. IEEE Comput. Soc. **42**(8), 30–37 (2009)
15. Newcombe. R.A., et al.: KinectFusion: real-time dense surface mapping and tracking, In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (2011)
16. Whelan, T., Kaess, M., Johannsson, H., Fallon, M., Leonard, J.J., McDonald, J.: Real-time large-scale dense RGB-D SLAM with volumetric fusion. Int. J. Robot. Res. **34**(4–5), 598–626 (2015)
17. Mur-Artal, R., Montiel, J.M., Tard'os, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. on Robot. **31**(5), 1147–1163 (2015)
18. Mur-Artal, R., Montiel, J.M.M., Tard'os, J.D.: ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. IEEE Trans. Robot. **33**(5), 1255–1262 (2017)
19. Galvez-Lopez, D., Tard´os, J.D.: Bags of binary words for fast place recognition in image sequences. IEEE Trans. Robot. **28**(5), 1188–1197 (2012)
20. Lourakis, M., Zabulis, X.: Model-based pose estimation for rigid objects. Comput. Vis. Syst. **7963**, 83–92 (2013)
21. Snavely, N., Seitz, S., Szeliski, R.: Photo tourism: exploring photo collections in 3D. ACM Trans. Graph. **25**(3), 2006 (2006)
22. https://fperucic.github.io/treant-js. Accessed 29 July 2019
23. Solanki, S., Batra, S.: Recommender system using collaborative filtering and demographic characteristics of users. Int. J. Recent Innov. Trends Comput. Commun. **3**, 4735–4741 (2015)
24. Serrell, B.: Paying attention: Visitors and museum exhibitions. American Association of Museums 1998, Washington (1998)
25. Falk, J.: Identity and the Museum Visitor Experience. Routledge, Abingdon (2009)
26. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: IEEE - ICCV, Barcelona, Spain, November 2011

27. Raptis, D., Tselios, N., Avouris, N.: Context-based design of mobile applications for museums: a survey of existing practices. In: International Conference on Human Computer Interaction with Mobile Devices & Services, pp. 153–160 (2005)
28. Economou, M.: Evaluation strategies in the cultural sector: the case of the Kelvingrove museum and art gallery in Glasgow. museum soc. **2**(1), 30–46 (2004)
29. Pontin, K., Lang, C., Reeve, J., Woolard, V.: Understanding museum evaluation. The Responsive Museum, Working with Audiences in the Twenty-First Century. Ashgate Publishing, Ltd., Farnham (2006)
30. Screven, C.G.: Exhibit evaluation: a goal-referenced approach. Curator **19**(4), 271–290 (1976)
31. Rasimah, C.M.Y., Ahmad, A., Zaman, H.B.: Evaluation of user acceptance of mixed reality technology. AJET **27**(8) (2011). https://ajet.org.au/index.php/AJET/article/view/899/176
32. Tost, L.P., Economou, M.: Exploring the suitability of virtual reality interactivity for exhibitions through an integrated evaluation: the case of the ename museum. J. Museol. **4**, 81–97 (2007)
33. Tzortzi, K.: Museum Space: Where Architecture Meets Museology. Routledge, London (2015)
34. Veron, E., Levasseur, M.: Ethnographie de l'Exposition: l'espace, le corps, le sens. Centre Georges Pompidou, Paris (1983)
35. Yalowitz, S.S., Bronnenkant, K.: Timing and tracking: unlocking visitor behavior. Visit. Stud. **12**(1), 47–64 (2009)
36. Tekin, B., Sinha, S.N., Fua, P.: Real-time seamless single shot 6D object pose prediction. In: Proceedings of the IEEE CVPR (2018)
37. Peng, S., et al.: Pvnet: Pixel-wise voting network for 6D of pose estimation. In: Proceedings of the IEEE CVPR (2019)