Hybrid Meta-Filtering System for Cultural Monument Related Recommendations

Eftychios Protopapadakis, Nikolaos Doulamis and Athanasios Voulodimos
National Technical University of Athens, 9 Iroon Polytechneiou, 15780, Athens, Greece

Keywords: Collaborative Filtering, Content-Based Filtering, Semi-supervised Learning, Recommendation System, Cultural Heritage.

Abstract: A two-phase monument recommendation concept is presented. The system ranks the alternative destinations by using the point and click technique during the process. The core of the system is a hybrid image filtering mechanism, which utilize both collaborative and content-based filtering. At first, the user profile is modelled in the form of a distance matrix, exploiting the user's annotations over a small set of descriptive images. At the same time, user's profile is compared to other profiles; the k closest profiles are utilized to refine the distance matrix. Then, the system provides relevant images to the user asking him/her to select few. The selected images are used in order to rank the alternative monuments.

1 INTRODUCTION

Cultural heritage has always been an intriguing domain for personalization applications (Ardissono et al., 2011); visitors differ and their visit experience is composed of the physical, the personal, and the socio-cultural context, and identity-related aspects (Spero, 2013). Hence they may benefit from individualized support that takes into account contextual and personal attributes (Doulamis et al., 2013). Moreover, visitors' behavior may not remain consistent during the visit and this may require ongoing adaptation.

Personalization implies modeling the user's way of thinking. Consequently, we would like to identify the user's needs by employing an easy to understand and effortless initialization procedure that requires only few minutes from the person's time, as in (Protopapadakis and Doulamis, 2014). The simplicity of the proposed approach does not expose the user to personalization related risks (Toch et al., 2012); no private information is required.

A hybrid recommendation system exploiting both collaborative and content-based filtering, for cultural heritage monuments sightseeing is presented. User's requirements are modeled according to his/her selections over a small set of representative images, using semi supervised learning approaches. The annotated set serves also as a profile descriptor, allowing the identification of other similar profiles. The system takes under consideration all the available information in order to provide relevant results.

Hybrid recommendation system(s) is a family of techniques that extents the traditional approaches of collaborative-filtering (CF) and content-based (CB) recommendation systems; e.g. CF techniques and contextual information for deriving improved recommendations in pervasive environments (Gavalas and Kenteris, 2011) or CF merged with personalized skyline operators (Bartolini et al., 2011).

Suggested approaches imply that there is a specific motivation in the user's behavior, which is known a-priori; e.g. monument 3D reconstruction (Makantasis et al., 2014). Thus, prior to any recommendation system application, the feature space is already defined.

Our work extents the approach of (Protopapadakis et al., 2014) in the following two crucial points: the feature interpretation during sampling and the user profile regularization according to other users' profiles. As a result, user preferences are expresses in the form of a distance matrix, which will be the core of the recommendation system.

In particular, the more descriptors employed the likely for the user to retrieve an image of interest; the representative images are selected for each of the

Protopapadakis E., Doulamis N. and Voulodimos A.
Hybrid Meta-Filtering System for Cultural Monument Related Recommendations.
DOI: 10.5220/0006347104360443

Figure 1: Representative image selection illustration for various monuments. Each row demonstrates few random images, from the representative image set. Each raw corresponds to different feature descriptors.

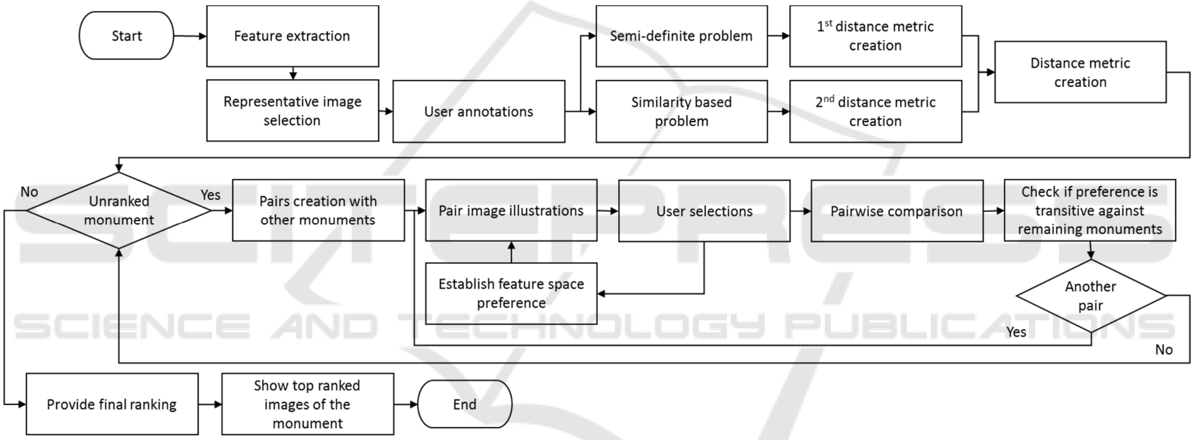

Figure 2: An illustration of the proposed, two-phase recommendation system.

calculated descriptors separately. Consequently, we have a bigger set of alternatives to show during the initialization, allowing the user for a broader search without any limitations regarding the search fields.

Secondly, after the user preferences are set, in a content-based way, according to his/her selections over presented images (see fig. 1), the preferences are slightly adjusted according to a comparison with other users' profiles, available in the data base.

At the end, a set of images is presented to the user. A simple voting mechanism is applied over the selections in order to rank the investigated monuments from the most relevant to least important to the user.

Within this paper: Section 2 describes the employed techniques regarding image selection, profiling and monument recommendation. Section 3 refers to the experimental setup and provides various metrics regarding models performance.

2 PROPOSED METHODOLOGY

The proposed approach lies between content-based and collaborative filtering. In particular for any user, we exploit a brief profile initialization process together with profile-similarity metrics in order to build a personalized suggestion system.

Assume a set of monuments $\mathcal{M} = \{m_1, \ldots, m_o\}$. A two-step process is employed in order to provide a ranking, given the user's profile. The first stage can be seen as a "quick" look in each available monument. The second stage is the selection among alternative monuments.

At first, we need to capture the user preferences, in a limited time. Given a monument, m_i, we form the appropriate distance metrics (see sec. 2.2 and 2.3). The updated distance metrics will be used during the image retrieval at the monument comparison stage.

Then, all possible combinations in pairs of two are formed. For each monument pair (e.g. m_i, m_j), few images are presented to the user. These images are randomly selected among the most descriptive ones, (see sec. 2.1), and the top ranked ones, according to eq. 7, (see sec. 2.4), spanning all possible feature spaces.

User selections allow the system to:

1. Identify which monument is more appealing to the user; i.e. has more images selected.
2. Establish the appropriate feature space, which will be used for further recommendations.

All the user sees are just images to select. However, these images are selected according to the user defined distance metrics and his/her feature space of interest. The process terminates when all the employed monuments are ranked. An illustration of the process is shown in fig. 2.

2.1 Initialization

The first step is actually a sampling approach. Given a set of feature vectors, since each image is described via many descriptors, we select a small subset of descriptive images. In particular, assume that we have n_D available descriptors. Thus, we perform n_D times the sampling process. Each time we obtain a different set of descriptive images $\boldsymbol{I}_s^{(i)} = [I_1, \dots, I_m]$, $i = 1, \dots, n_D$. Note that the number of representative images in each set, m, varies, depending on the feature vectors we use.

In order to extract the most important (descriptive) ones, the work of (Elhamifar et al., 2012) around sparse modeling for finding representative objects is employed. Their work is summarized through the following formulation:

$$\min \lambda \|\boldsymbol{C}\|_{1,q} + \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{C}\|_F^2 \quad s.t. \quad \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T \tag{1}$$

where $\boldsymbol{Y}$ and $\boldsymbol{C}$ refer to data points and coefficient matrix respectively. This optimization problem can also be viewed in a compression scheme, where we want to choose a few representatives that can reconstruct the data.

A preliminary set of suggested images $\boldsymbol{I}_D = \left\{\boldsymbol{I}_s^{(i)}\right\}$, $i = 1, \dots, n_D$, which describes the entire set, has been created. The $\boldsymbol{I}_D$ set is the basis for the profile definition in every new user case. However, despite the sparse selection the number of images could be quite troublesome for achieving a fast initialization profile setup. In that case, a subset is randomly selected over $\boldsymbol{I}_D$.

2.2 Content-Based Filtering

The representative objects retrieved, are shown to

Figure 3: Illustration of the user profiling steps (from left to right) over the Knossos monument. The user's interest appear to be structural elements (support pillars). Note the existence of non-relevant images when using CB retrieval techniques, as well as, their partial elimination using the hybrid approach.

Figure 4: An illustration of the monuments comparison via image selection. On the left are the images from the Knossos monument. On the right are the images from Parthenon. In both cases, the presented images have the higher rank scores according to eq. 7. User is asked to point and click over images he/she likes the most. Each selection counts as one vote for the corresponding monument.

user, who can define the relevance to his current search. User can select any number of the representative images for annotation, as long as there is *at least one relevant and one non-relevant image in the end*. Even if user decide to annotate all the suggested images that will not be troublesome due to the small number that sparse modeling indicates.

The exploitation of many image descriptors allow the selection of a wider range of images; it is very likely that some images will be rather appealing to the user. Once the user denotes for some of the displayed images the relevance to the current search, we update the distance metrics.

For any two given data points x_i and x_j, let $d(x_i, x_j)$ denote the distance between them. To compute the distance, let $\boldsymbol{A} \in R^{m \times m}$ be a symmetric matrix, we can then express the formula of distance measure in a generic form:

$$d_{\boldsymbol{A}} = \sqrt{(x_i - x_j)^T \boldsymbol{A} (x_i - x_j)} \qquad (2)$$

Similar to the approach of (Hoi et al., 2008), the distance metric learning (DML) problem is to learn an optimal $\boldsymbol{A}$ from a collection of data points C on a vector space R^m together with a set of similar pairwise constraints $\mathcal{S}$ and a set of dissimilar pairwise constraints D. The two sets of user defined pairwise constraints among data points have the form:

$$\begin{aligned} \mathcal{S} &= \{(x_i, x_j) | x_i \ relevant\ to\ x_j\} \\ \mathcal{D} &= \{(x_i, x_j) | x_i \ irrelevant\ to\ x_j\} \end{aligned} \qquad (3)$$

The problem formulation is stated as (Hoi et al., 2008):

$$\begin{aligned} \min_{A}\ & t + \gamma_s tr(\boldsymbol{A} \cdot \boldsymbol{S}) - \gamma_d tr(\boldsymbol{A} \cdot \boldsymbol{S}) \\ s.t.\ & tr(\boldsymbol{X L X}^T \boldsymbol{A}) \leq t \\ & \boldsymbol{A} \in S_+ \end{aligned} \qquad (4)$$

Thus, the DML problem has been approached as a semi-definite problem, which can be solved efficiently with global optimum using existing convex optimization packages.

2.3 Collaborative Filtering

User's initial selections, $\boldsymbol{U}^{(i)} = \{I_j\}$, $j = 1, \dots, m$, where $m \ll n_D$ so that $\boldsymbol{U}^{(i)} \subset \boldsymbol{I}_S^{(i)}$, are actually a signature vector, whose similarity to the other data entries is exploited. In particular, for a known monument $\mathcal{M}$, we identify the k closest entries to the ones of the current user. The similarity among to entries, e_i,e_j is defined as:

$$S(e_i, e_j) = \boldsymbol{U}_{e_i}^{(i)} \cap \boldsymbol{U}_{e_j}^{(i)} \qquad (5)$$

Thus, the more common elements the higher the similarity is. Then, we form another distance matrix $\boldsymbol{A}$, denoted as $\boldsymbol{A}_{cf}$ according to the following equation:

$$\boldsymbol{A}_{cf} = \frac{1}{C} \sum_{i=1}^{k} c_i \cdot \boldsymbol{A}_i \qquad (6)$$

where $C = \sum_i^k c_i$ and $c_i \in [0,10]$ is a user assigned value, which denotes the satisfaction of the i-th user from the retrieved results according to his personalized content-based filtering matrix $\boldsymbol{A}_i$.

2.4 Providing Appropriate Image Suggestions

The final image suggestion is based on a total raking approach, for every image x_j, described by the following equation:

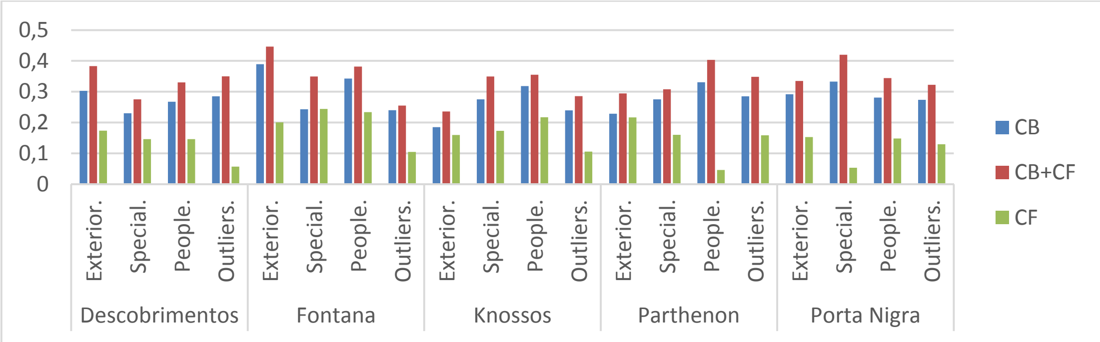

Figure 5: The hybrid approach (CB + CF) outperformed both the traditional content-based (CB) and collaborative filtering (CF) technique for every scenario.

$$r_j = \sum_{\substack{i=1 \\ i \neq j}}^{|P|} \frac{1}{d_{A_H}(x_i, x_j)} + \sum_{\substack{i=1 \\ i \neq j}}^{|N|} \frac{1}{d_{A_H}(x_i, x_j)} \tag{7}$$

where $r_j$is the overall ranking score for an image j, given its feature vector x_j, $|P|$ and $|N|$ denotes the size of user annotated images as positive and negative to current search respectively, and $d_{A_H}(x_i, x_j)$ is a distance metric defined both collaborative and content-based distance metrics.

In particular $d_{A_H}(x_i, x_j)$ is calculated according to Eq. 2, using the distance matrix $\boldsymbol{A}_H$ defined as:

$$\boldsymbol{A}_H = (1 - t)\boldsymbol{A} + t\boldsymbol{A}_{cf} \tag{8}$$

where t is a trade-off parameter, $t \in (0,1)$.

When ranking is concluded, the n higher ranked images are presented to the user. Please note that the already annotated images, are excluded from the ranking process; system recommendations are over new unseen images. An illustration of the image suggestion process is shown in fig. 3.

2.5 Ranking the Monuments

The monument ranking can be seen as a simple voting system. In each monument pair comparison set, 16 images are shown to the user. Then, user is asked to select the images that he is interested in, as shown in fig. 4.

Selected images count as votes. The monument with most votes is the winner of the pair contest. In every monument pair contest we use different images, which are relevant to the user. Also, in order to simplify the ranking approach (i.e. reduce the number of pair comparisons), we assume that the preference among alternatives is transitive. If "monument A is at least as good as monument B" and "monument B is at least as good as monument C" then "monument A is at least as good as monument C".

Finally, if there is a tie between two or more monuments (i.e. same number of votes), system provides a last set of images to select. The selection is repeated until there is a final rank score.

3 EXPERIMENTAL SETUP

Initially a large data set of images is collected from Flickr (Ioannides et al., 2013). The data retrieval was based in various parameters (including free description, tags, location, etc.). Evaluation data is specifically build around five cultural monuments. These monuments were Padrão dos Descobrimentos, Fontana dei Quatro Fummi, Knossos, Parthenon and Porta Nigra. Over 3000 images from five cultural monuments in Europe were used.

For every monument, four recommendation schemes are considered: a) need for exterior images of the monument, b) special attributes (e.g. interior design, paintings, sculptures, etc.), c) people around the monument and d) various images without any cultural interest (e.g. animal pictures, night sky, signs, etc.).

In every scenario the relevant images are taken from one category and the non-relevant from the rest three in order to construct the pairwise constraints shown in eq. 3. In every case the ratio was 3 relevant to 3 irrelevant, leading to user feedback of 6 images in total. The trade-off factor, in eq. 8, was set as $t = 0.3$.

There was in total 350 user profiles available. Each profile had from 3(2) up to 5(4) positive (negative) annotated images for every monument. Also, the ranking order of the monument was given from each of the users.

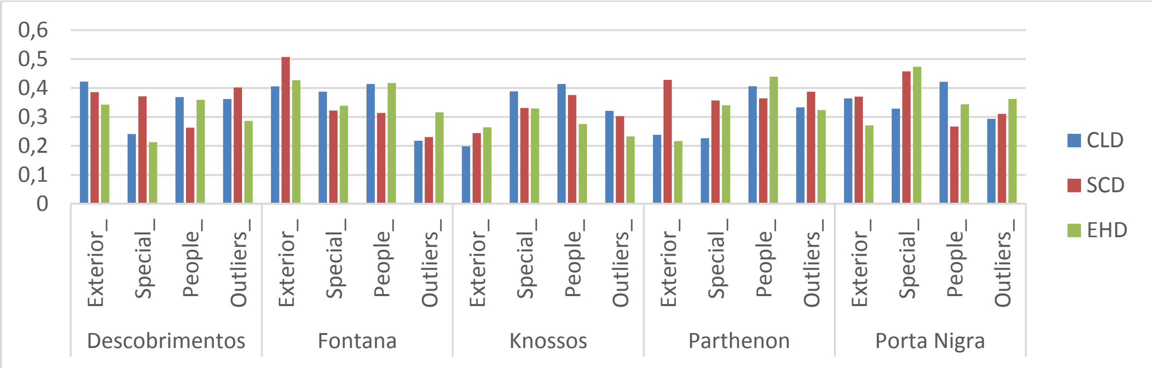

Figure 6: The impact of feature descriptor selection, Color Layout Descriptor (CLD), Scalable Color Descriptor (SCD) and Edge Histogram Descriptor (EHD), during the suggestion process.

3.1 Dataset Description

A brief description of the five selected monuments is provided in the following lines.

Padrão dos Descobrimentos is a monument on the northern bank of the Tagus River estuary, in the civil parish of Santa Maria de Belém, Lis-bon. Located along the river where ships departed to explore and trade with India and Orient, the monument celebrates the Portuguese Age of Discovery (or Age of Exploration) during the 15th and 16th centuries. The set contains 847 images and the special category refers to the square in front of the monument images.

Fontana dei Quatro Fummi (Fountain of the Four Rivers) is a fountain in the Piazza Navona in Rome, Italy. It was designed in 1651 by Gian Lo-renzo Bernini for Pope Innocent X. The set contains 133 images and the special category refers to night shots and grayscale images.

The Parthenon is a former temple on the Athenian Acropolis, Greece, dedicated to the goddess Athena. Construction began in 447 BC. It is the most important surviving building of Classical Greece, generally considered the zenith of the Doric order. The set contains 1109 images and the special category refers to support beams images.

Knossos is the largest Bronze Age archaeological site on Crete, Greece and is considered Europe's oldest city. The set consists of 1392 images and the special category refers to wall drawings. The set contains 133 images and the special category refers to night shots and grayscale images 1392.

Porta Nigra (black gate) is a large Roman city gate in Trier, Germany. It is today the largest Roman city gate north of the Alps. The set contains 690 images and the special category refers to interior images.

3.2 Descriptors Used & Sampling Performance

Once the data set for a specific monument is gathered, additional features from the images are extracted. Three MPEG-7 visual descriptors have been employed for the purposes of this research: Color Layout Descriptor (CLD), Scalable Color Descriptor (SCD) and Edge Histogram Descriptor (EHD). The specific descriptors were chosen due to their simplicity and small size, high processing speed, robustness, scalability and interoperability (Serna et al., 2011).

Table 1: Number of representative images for different feature descriptors.

Monument name	Number of Images	Descriptor name		
		CLD	SCD	EHD
Descobrimentos	847	25	100	6
Fontana	133	22	28	13
Knossos	1392	83	77	29
Parthenon	1109	32	35	24
Porta Nigra	690	52	70	14

Table 1 describes the representative image subset creation. It appears that, regardless the monument, one tenth of the original images suffice to adequately describe the entire image set. As such, a few images are provided to the user, allowing a fast initialization process (profiling).

Regardless the application scenario, recommended images of the hybrid system are more appealing to the user than the traditional contend based approach, as shown in fig.5. Among the exploited feature descriptors there is no dominant one (fig.6). The variance in performance suggest a low quality of feature descriptors, for the problem at hand.

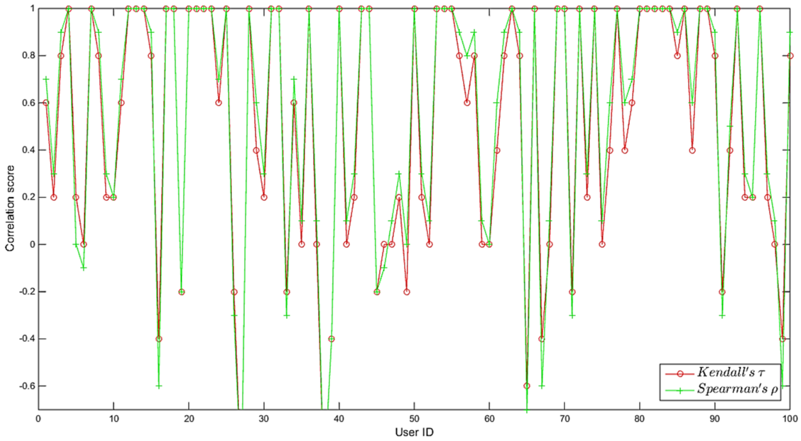

Figure 7: Spearman's ρ and Kendall's τ ranking scores for 100 user profiles.

3.3 Ranking Scores

The proposed system ranks the alternative visit destinations (monuments) according to the number of selected images. Therefore, we should measure the ordinal association between the actual rankings (known a priori by asking the user) and the systems' recommendations (i.e. rank correlation).

The performance metrics utilized are Spearman's ρ (Ornstein and Lyhagen, 2016) and Kendall's τ (Park and Stone, 2014). The individual scores for each of the test samples (users) can be found in fig.5.

The system recommendations where positively correlated with most of the actual ones. However, the negative correlations suggest that the process has to be further refined.

4 CONCLUSIONS

A hybrid recommendation system, for the cultural heritage field, has been proposed. User preferences are modeled using a fast image selection process. The process allows for a semi supervised profiling (content-based), which is further refined using other available profiles (collaborative-based). The combination of both approaches allow the encoding of the user's preferences into a distance matrix, which is the core of the recommendation system.

Monument ranking is achieved, again, via image selection. A mixed set from various monument images are shown to the user. The presented images are retrieved from a large dataset using the user defined distance metrics. The selected images count as votes; the monument with most votes is ranked first.

Future research will focus on the exploitation of better feature descriptors and the implementation in a wider system for monument suggestions. Also, we should examine the system behaviour in a larger user population with more

ACKNOWLEDGEMENTS

This work was supported by the EU H2020 TERPSICHORE project "Transforming Intangible Folkloric Performing Arts into Tangible Choreographic Digital Objects" under the grant agreement 691218. The work has been, also, partially supported by 4D-CH-World: Four Dimensional Cultural Heritage World, Marie Curie IAPP project Grant agreement number 324523.

REFERENCES

Ardissono, L., Kuflik, T., Petrelli, D., 2011. Personalization in cultural heritage: the road travelled and the one ahead. User Model. User-Adapt. Interact. 22, 73–99. doi:10.1007/s11257-011-9104-x

Bartolini, I., Zhang, Z., Papadias, D., 2011. Collaborative Filtering with Personalized Skylines. IEEE Trans. Knowl. Data Eng. 23, 190–203. doi:10.1109/TKDE.2010.86

Doulamis, N., Yiakoumettis, C., Miaoulis, G., Protopapadakis, E., 2013. A Constraint Inductive Learning-Spectral Clustering Methodology for Personalized 3D Navigation, in: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Li, B., Porikli, F., Zordan, V., Klosowski, J., Coquillart, S., Luo, X., Chen, M., Gotz, D. (Eds.), *Advances in Visual Computing, Lecture Notes in Computer Science. Springer* Berlin Heidelberg, pp. 108–117.

Elhamifar, E., Sapiro, G., Vidal, R., 2012. See all by looking at a few: Sparse modeling for finding representative objects, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). P*resented at the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR*), pp. 1600–1607. doi:10.1109/CVPR.2012.6247852

Gavalas, D., Kenteris, M., 2011. A web-based pervasive recommendation system for mobile tourist guides. Pers. Ubiquitous Comput. 15, 759–770. doi:10.1007/s00779-011-0389-x

Hoi, S.C.H., Liu, W., Chang, S.-F., 2008. Semi-supervised distance metric learning for Collaborative Image Retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008. Presented at the *IEEE Conference on Computer Vision and Pattern Recognition, 2008.* CVPR 2008, pp. 1–7. doi:10.1109/CVPR.2008.4587351

Ioannides, M., Hadjiprocopi, A., Doulamis, N., Doulamis, A., Protopapadakis, E., Makantasis, K., Santos, P., Fellner, D., Stork, A., Balet, O., Julien, M., Weinlinger, G., Johnson, P.S., Klein, M., Fritsch, D., 2013. Online 4d Reconstruction Using Multi-Images Available Under *Open Access. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. II-5/W1, 169–174. doi:10.5194/*isprsannals-II-5-W1-169-2013

Makantasis, K., Doulamis, A., Doulamis, N., Ioannides, M., 2014. In the wild image retrieval and clustering for 3D Cultural heritage landmarks reconstruction. Multimed. Tools Appl. 1–37.doi:10.1007/s11042-014-2191-z

Ornstein, P., Lyhagen, J., 2016. Asymptotic Properties of Spearman's Rank Correlation for Variables with Finite Support. PLOS ONE 11, e0145595. doi:10.1371/journal.pone.0145595

Park, L.A.F., Stone, G., 2014. Inducing Controlled Error over Variable Length Ranked Lists, in: Tseng, V.S., Ho, T.B., Zhou, Z.-H., Chen, A.L.P., Kao, H.-Y. (Eds.), Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science. Presented at the *Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer International Publishing,* pp. 259–270.

Protopapadakis, E., Doulamis, A., 2014. Semi-Supervised Image Meta-Filtering Using Relevance Feedback in Cultural Heritage Applications. Int. J. Herit. Digit. Era 3, 613–627. doi:10.1260/2047-4970.3.4.613

Protopapadakis, E., Doulamis, A., Matsatsinis, N., 2014. Semi-supervised Image Meta-filtering in Cultural Heritage Applications, in: Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.-Y., Quak, E. (Eds.), Digital Heritage. Progress in Cultural Heritaage: Documentation, *Preservation, and Protection, Lecture Notes in Computer Science. Springer International Publishing*, pp. 102–110.

Serna, S.P., Scopigno, R., Doerr, M., Theodoridou, M., Georgis, C., Ponchio, F., Stork, A., 2011. 3D-centered Media Linking and Semantic Enrichment Through Integrated Searching, Browsing, Viewing and Annotating, in: Proceedings of the 12th International Conference on Virtual Reality, Archaeology and Cultural Heritage, VAST'11. *Eurographics Association, Aire-la-Ville, Switzerland, Switzerland*, pp. 89–96. doi:10.2312/VAST/VAST11/089-096

Spero, S.B., 2013. The museum experience revisited. Mus. Manag. Curatorship 28, 430–432. doi:10.1080/09647775.2013.831528

Toch, E., Wang, Y., Cranor, L.F., 2012. Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. User Model. User-Adapt. Interact. 22, 203–220. doi:10.1007/s11257-011-9110-z