4D Modelling in Cultural Heritage

Anastasios Doulamis[1(✉)], Nikolaos Doulamis[1], Eftychios Protopapadakis[1], Athanasios Voulodimos[1], and Marinos Ioannides[2]

[1] National Technical University of Athens, 15773 Zografou, Athens, Greece
{adoulam,ndoulam}@cs.ntua.gr,
{eftprot,thanosv}@mail.ntua.gr
[2] Cyprus University of Technology, 3036 Lemesos, Cyprus
marinos.ioannides@cut.ac.cy

Abstract. This chapter describes the main research outcomes and achievements of 4D modelling in cultural heritage. 4D digital modelling implies the creation of precise time-varying 3D reconstructions of cultural heritage objects to capture temporal geometric variations/distortions, i.e., a spatio-temporal assessment. The key research challenge for 4D modelling, was the data collection over heterogeneous unstructured web resources. Such "in the wild" data include outliers and significant noise, since they have not been created for 3D modelling and reconstruction purposes. In addition, GPS and geo-information is limited or non-existent. However, such data allow for a massive reconstruction of the content even for monuments that have been destroyed due to natural phenomena or humans' interventions. The key outcomes include (i) a Twitter-based 3D modelling of CH objects so as to reconstruct CH monuments and sites from unstructured image content, (ii) the development of a search engine and a (iii) recommendation system for different CH actors (curators, conservators, researchers), (iv) 3D reconstruction of the historic city of Calw in Germany, (v) the creation of a 3D virtual environment in real-time and (vi) launch of a 4D viewer enabling the easy handling of the 3D geometry plus the time. The results show the main innovation of the proposed 4D dimension, i.e., the time in precise modelling of the rich geometric content of the monuments.

Keywords: 4D modelling · Tangible cultural heritage
3D reconstruction, tweets analysis and recommendation systems

1 Introduction

Digitalizing cultural sites and objects and creating 3D digital models is an important task to preserve Cultural Heritage (CH) [1]. Among all CH resources, the outdoor large-scale cultural sites are mostly sensitive to weather conditions, natural phenomena (earthquakes, flooding, etc.), excavation procedures, and restoration protocols [2, 3]. This implies an imminent need for a spatio-temporal monitoring of those sites to identify regions of potential material degradation, unstable structuring conditions, localize spatial modifications and detect environmental damages [4]. A time varying 3D model should be developed to assess the spatial and temporal diversity of CH objects but again under a cost-effective framework able to be applied to large-scale sites [5].

M. Ioannides et al. (Eds.): Digital Cultural Heritage 2017, LNCS 10754, pp. 174–196, 2018.
https://doi.org/10.1007/978-3-319-75789-6_13

On the other hand, several actors are involved in CH community. (i) CH researchers and archaeologists can better document sites and objects (CH resources), relate their operational uses with past social and political structures, economical factors and past advances in science and technology [6]. (ii) Conservators can better assess the effect of different restoration methods on maintaining the structure and the nature of the cultural items, while simultaneously simulate and predict material degradation factors [7]. (iii) Curators can exploit the digital counterparts to design exhibitions and disseminate cultural knowledge to wide public [8]. (iv) Creative industries and Small Medium Enterprises (SME's) can build new services promoting Europe's culture worldwide, forging Europe's cohesion and integration through its diverse cultural legacy and boosting European economy. During the recent severe economic crisis, creative enterprises and tourism industries are some good examples of withstanding economic recession and (though slightly) contribute towards sustainable growth of Europe's economy [9]. (v) Finally, the simple visitors and the public can share unique cultural experiences on handling digital CH objects and navigating through a digital cultural world. Each of the aforementioned players has different needs regarding digital model parameters and scales. Therefore, we need to develop digital models able to respect all diverse requirements of the CH community users [10, 11].

On the other hand, the rapid progress in technology regarding visual capturing accompanying with respective progress in respective software tools has stimulated the generation of millions of image content being nowadays stored onto distributed and heterogeneous internet repositories, like Flickr, Picasa, Photosynth, etc. [12]. This content provides a unique opportunity for cultural heritage documentation, like for 3D reconstruction, through the fact that the overwhelming majority of these images have been captured for personal use and thus they are not suitable for such documentation process [13]. Thus, many of these images contain irrelevant material like views of other objects, or of the city instead of the monument itself [14]. Therefore, content-based filtering algorithms are necessary for an effective and computationally efficient e-documentation process that exploits the "wild Internet image collections".

In this book chapter, we recent achievements in 4D modelling (reconstruction) derived as a results of the four dimensional (4D) Cultural Heritage World Project, (4D-CH-World [15]), which aims at analyzing, designing, researching, developing and validating an innovative system integrating the latest advances in computer vision and learning, as well as, 3D modeling and virtual reality for the rapid and cost-effective 4D maps reconstruction in the wild for personal use, and support the aim of our European Commons and the digital libraries EUROPEANA and UNESCO Memory of the World (MoW) to build a sense of a shared European cultural history and identity.

Currently, 3D digital models are generated under a spatial-temporal independent framework. This means that digitalization information of common parts (surfaces) of an object is not exploited to digitalize similar surfaces of the same or other objects. Furthermore, the digitization process at a current time instance does not exploit results from reconstructions obtained at previous time instances. Last but not least, the scale dimension of each 3D model is generated differently for each users' category (e.g., CH researchers, curators, visitors) leading to the creation of independently scaled 3D models [11]. As an additional drawback, one can also refer to the lack of a semantic enrichment of the digital information to assist CH community users in their research and work. This makes 5D modelling too complex to be validated under real-life large-scale application domains.

Spatio-temporal dependency means that a 3D model at a current time instance is generated taking into account information of the same object from previous reconstructions (temporal dependency) and common surface properties (spatial dependency). Predictive refers to the ability to select regions of interest to be reconstructed differently by fusing/integrating selective 3D capturing methodologies with respect to surface features (material and geometric properties) and users' needs.

This chapter is organized as follows: The overall methodology and a summary of the key achievements is shown in Sect. 2. The techniques used to recognize objects from distributed multimedia repositories of CH objects are presented in Sect. 3. The 4D modelling search engine and recommendation system is discussed in Sect. 4. Section 5 deals with the 3D reconstruction of the historic city of Calw in Germany while the developed 4D viewer is shown in Sect. 6. Finally Sect. 7 draws the conclusions.

2 The Recent Achievements in 4D Modelling for Cultural Heritage

In the following we summarize the overall achievements of the 4D CH World project. The key achievements are organized into five main outcomes:

Groundbreaking photograph recognition and data sampling from data bases and web: A novel search engine has been developed to mine cultural data from distributed multimedia repositories such as Flickr, Picasa, and the web [16]. The search engine has been improved using tweets messages by initially proposing an algorithm which is able to identify key events from a pool of simple tweets messages [17]. The method supports two main phases; event detection from tweets and on the fly 3D reconstruction from the visual content embedded on common tweets clusters with respect to a cultural heritage object (an event) [17, 18]. For the first case, we propose a modification of existing document-based information retrieval metrics such as Term Frequency-Inverse Document Frequency (TF-IDF) criterion by including information regarding retweeting and number of followers. This new metric is then transformed in space and time adopting a Wavelet Transform. In the sequel, multi-assignment graph partitioning is proposed to localize clusters of events processing tweets messages.

Regarding the on-the fly 3D reconstruction, we proceed with outliers' removal, image clustering via a dense based scheme and key images extraction [16, 19]. The latter are selected so that they mostly represent all different orientations of CH monument at spherical coordinate system. The selected images are fed to SfM tool for 3D modelling. The fact that a small but representative number of images is selected severely reduce the time needed for reconstructing the 3D model.

Search and retrieval for further non digital content: We have investigated the combination of the aforementioned event detection approach from the Twitter on cultural heritage photos uploaded on this social medium [20]. In particular, we have analyzed tweet messages based on their textual information. Then, the embedded visual content is extracted to build up 3D CH models that represent high level semantics of the images. These high level semantic information is exploited for searching and retrieving non-digital content. The models are evolved in time and thus the four dimension is added.

In addition, a novel content-based recommendation system of CH objects was developed [21]. The system dynamically models end-users preferences, creating common profiles and then rank and filter media information exploiting both visual similarities and user's profiles. Relevance feedback mechanisms are exploited for the automatic profile estimation [22–24]. Relevance feedback is a method for dynamically updating user's preferences according to a set of relevant/irrelevant data selected from the user based on user's interaction [25–27].

City area and buildings reconstruction using acquired data: Several photogrammetric and computer vision methods are investigated for city modelling and buildings reconstruction [28, 29]. The mediaeval city of Calw in Germany has been selected for demonstration of the proposed 3D modelling methods. We have reconstruct a major part of the city of Calw and embed these reconstructions on the 4D Viewer [30]. The reconstruction faces not only 3D geometry of the buildings but also time evolution.

Development of appropriate simulation model: The 4D viewer [31] supports 4D reconstructions, that is, 3D geometry plus the time. The viewer has been tested at Calw case. It has been supported by Virtual Reality and Augmented Reality functionalities. The users are able to navigate through time for selected CH objects, see semantic information and tags assigned to these objects and relate these assignments with different historic periods, manipulate the CH at different angles and views with respect to their needs and information preferences [32].

Depiction of future urban structures: Adeep learning algorithm has been developed through the exploitation of convolutional neural networks. The algorithm receives big volumes of LiDAR data of an urban region and classify the structures of the buildings and the type of the material used [33]. Urban buildings consist an integral part of cultural heritage. They shape the sense of belonging somewhere, of social traditions, of cultural identity of a history spanning centuries. Therefore, automatic detection and recognition of specific types of urban buildings is extremely important for disseminating cultural heritage to general audience.

3 Recognition and 3D Reconstruction of CH Photos from Distributed Multimedia Repositories

In this section, we deal with the recognition and 3D reconstruction of photos from distributed multimedia repositories. The key concept is to 3D reconstruct CH monuments and buildings based on the huge amount of media information found over the web. The main challenge, however, in this case, is the unstructured nature of the content which have been derived from applications different than for 3D modelling. Such a tool can stimulate a massive 3D reconstruction of CH objects of interest to face (i) looting, (ii) mankind destructions (war, fire) and (iii) environmental defects on the objects (earthquakes) [34]. We consider in the wild (unstructured) 3D reconstruction on two different multimedia repositories; the first refers to tweets messages (see Sect. 3.1) while the second on the web-based unstructured multimedia tools (see Sect. 3.2).

3.1 3D Modelling of Images from Tweets' Messages

3D reconstruction of a CH asset from images located on the Twitter is a research challenging process since Twitter images are collected for purposes different than digitization [17]. Therefore, they are (i) incomplete in the sense that several parts of the CH objects are missing, (ii) presents a lot of noise (e.g., existence of other objects), (iii) have been captured from quite different conditions in the sense of resolution, image calibration and registration [17, 20].

To reconstruct the images from tweets messages, we need first to introduce a novel event detection algorithm. This algorithm analyzes tweets signals and through the use of apt characterization metrics identifies the events. Events detection acts as a medium to refine visual information of tweets messages constrained upon textual properties. Finally, tweets of similar content are organized together to create 3D models by "structuring" image content of the same object/place. 3D models are created by initially filtering the image content through suitable features and finding correspondences between similar images.

For an efficient tweet event detection algorithm, text characterization is required. Text characterization is accomplished through the extraction of textual features that categorize the significance of a word in tweets. The information theoretic metrics used for textually characterizing a document, such as Term Frequency-Inverse Document Frequency (TF-IDF) metric [35] or distributional features [36], are not suitable for the tweets case. Tweets are very short messages (no more than 140 characters), leading to statistical inaccuracies in estimating traditional document metrics over tweet posts. In addition, Twitter is distinguished from similar websites by some key features. Users may subscribe to other users' tweets – this is known as following and subscribers are called followers. Moreover, Twitter has the re-tweet feature that gives users the ability to forward an interesting tweet to their followers. For all these reasons, new metrics are required to model tweet posts.

In our research, we have introduced three information theoretic metrics for measuring the importance of a word on a tweet post. These three metrics are described in details and mathematical formulation in [17]. The first metric, called Conditional Word Tweet Frequency (CWTF), we consider as a document the collection of tweets extracted over the k-th time interval and as a corpus of documents the assembly of tweets over the p previous time intervals. The main difference of CWTF from the classical description of TF is that here we count the number of tweets that contain a specific word within the current examined time interval k instead of counting the number of times that a word appears within a document. That is, all tweets that contain the specific word contribute the same to the calculation of CWTF.

The second metric, called Word Tweet Frequency (WTF), considers the frequency of appearance of the specific word w in the tweets. Finally, the third metric, called Weighted Conditional Word Tweet Frequency (WCWTF) compensates the aforementioned metric with respect to the significance of a tweet as expressed either by the number of followers or by the number of re-tweets it produces. The number of followers indicates the credibility of the author of the tweet, which also implies credibility for the textual information of the respective tweet post. The number of re-tweets is a metric for ranking the importance of the semantic information (textual content) posted by the tweet.

A wavelet transformation is the exploited to localize the signals both in time and in frequency domain. This is important since tweets are not synchronized messages. At the end, a graph-cut approach is exploited to extract the events, exploiting as similarity distance the cross correlation criterion since this can directly express the similarity of two feature vectors being invariant in scale and translation. Then, clustering is modeled as a graph partitioning problem, by adopting an optimal methodology that cuts the graph into clusters so that intra-cluster elements present the maximum coherence, while the inter-cluster elements the minimum one. In this graph partitioning problem, we contribute by allowing for a multi-assignment clustering since, in our context, one word can belong to several clusters (events). This is achieved by introducing a modification of the spectral clustering algorithm that handles multi-assignment clustering problems. Finally, tweets of similar content are organized together to create 3D models by "structuring" image content of the same object/place. 3D models are created by initially filtering the image content through suitable features and finding correspondences between similar images.

Figure 1 shows the performance of the three introduced information theoretic metrics as regards the precision and recall values. The Fig. 1 evaluates the performance when using wavelet transform or not. As we can see, the third information theoretic metric outperforms the other two ones.

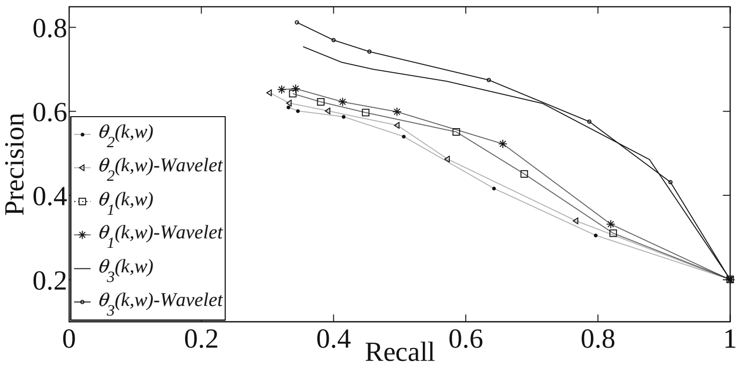

Fig. 1. Precision-recall curve of the three proposed metrics using both wavelet and non-wavelet representation (Figure created by the authors and cited in [17]).

3.2 3D Modelling of Images from Distributed Web-Based Multimedia Repositories

Nowadays, there are available extremely large collections of images and videos, most of them located on distributed and heterogeneous platforms over the web. The proliferation of billions of shared photos has outpaced the current technology for browsing such collections and entails the necessity for developing new efficient image retrieval techniques.

The presented approach takes into account both images metadata description, including geo-location and user generated tags, and visual information. It exploits image metadata to retrieve an initial set of images and then it uses visual information to perform a two-step unsupervised image clustering. Based on the assumption that the initial retrieved set will contain sufficient number of images depicting the same object, this algorithm needs neither a priori knowledge of the retrieved dataset nor a reference image to compute visual similarities and perform clustering [14, 19].

Images' visual information is encoding through the usage of local descriptors such as the ORB [37]. Then, a DBSCAN [38] dense-based clustering scheme is employed to find out compact image points and remove the outliers. Initially a set of images is retrieved by using text query. This initial set contains a lot of outliers that have to be removed. After the computation of local descriptors and the construction of the similarity matrix, DBSCAN algorithm takes care of this step and removes the most prominent outliers. We also introduce a modification of the DBSCAN, called CSP, to better reply on the problem of image clustering in such an unstructured environment. Finally, through spectral clustering, two sets of images that contain the rear and the front view of the monument, are being constructed.

Figure 2 shows the location of the ORB extracted image features and the dense of the outliers. This way, we are able to remove the outliers and improve the 3D reconstruction performance.

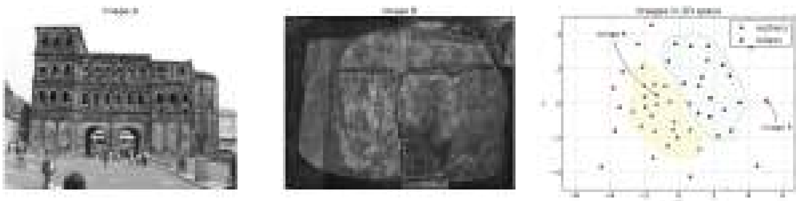

Fig. 2. Representation of images in a 2-dimensional space and depiction of image clusters and outliers (Figure created by the authors and cited in [14]).

Figure 3 presents the F1-score regarding partitioning performance using the two different proposed approaches as for outliers' removal (DBSCAN and CSP) along with the center-based clustering algorithm k-means and density-based algorithm Mean Shift. Both the proposed approaches outperform k-means and Mean-Shift with the CSP behaving better than the conventional DBSCAN, especially for a large number of image outliers. The results have been obtained by averaging F1 scores on cultural heritage objects of our "wild" image dataset.

Finally, Fig. 4 shows the gradual 3D reconstruction performance of a historic monument in Germany, the Porta Nigra. We can see that a rough 3D reconstruction can be achieved even when a very small number of images have been selected.

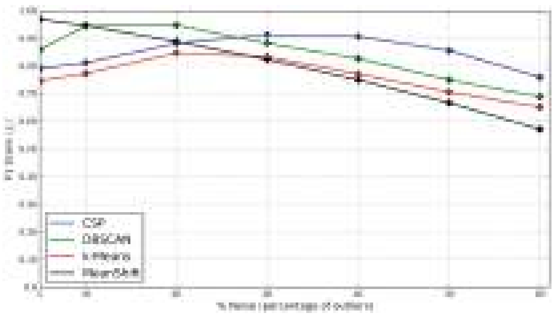

Fig. 3. The effect of a different number of clustering algorithms on outliers' removal performance (Figure created by the authors and cited in [14]).

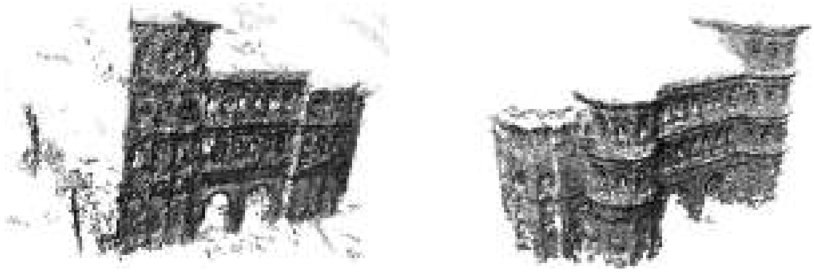

Fig. 4. 3D reconstruction of rear and front view sides of Porta Nigra. For this reconstruction 30 images were used that contained 20% of outliers (Figure created by the authors and cited in [14]).

4 The 4D Modelling Search Engine

4.1 Searching Unstructured Content

Our approach for cost-effective space and time reconstruction of tangible Cultural Heritage objects, involves an entire workflow of computer vision, photogrammetry, and 3D reconstruction, semantic enrichment, indexing and searching techniques. The 4Ds reconstruction approach presented in the current research work is innovative in the sense that it combines advanced techniques from Computer Vision, Photogrammetry, 3D representation and semantic representation in order to produce feasible and cost-effective 4D views of Digital Cultural Heritage content that exists 'in the wild' [4, 5].

The pipeline of the 4D Reconstruction method proposed in this paper consists of the following processes, which are shown in Fig. 5.

The Search engine. It is the entry point of the 4D Reconstruction workflow and is responsible for acquiring visual and textual content 'in the wild', i.e. from publicly available online sources, as well as for metadata capturing. The results of the search engine form the reconstruction data set that will capture the content of Cultural Heritage through space and time.

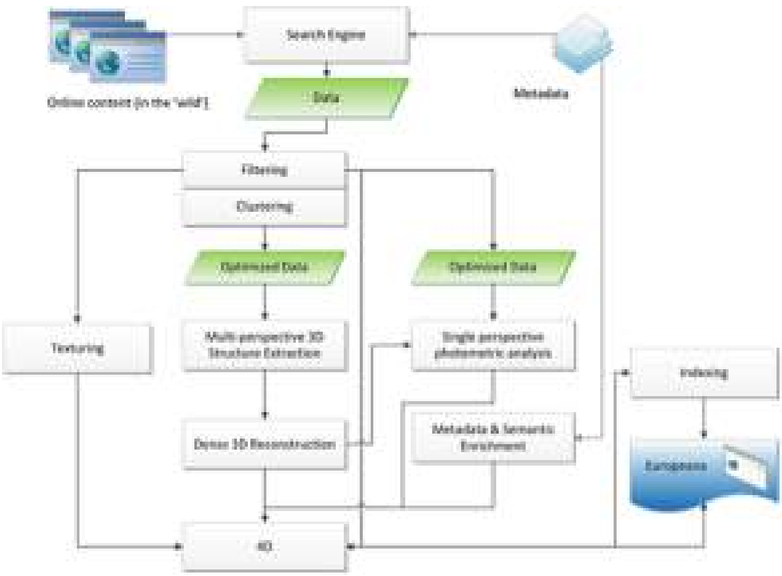

Fig. 5. The 4D reconstruction workflow of Digital Cultural Heritage (Figure created by the authors).

The Filtering and Clustering engines are responsible for analyzing the visual content of 2D images and for applying advanced Computer Vision filtering and clustering algorithms that will optimize the acquired data set. Images found 'in the wild' and especially in web search engines, are not accurate. The optimization goals of those engines are twofold: (a) the acquisition of multiple-perspectives' 2D content and (b) the identification of different lightning conditions of the same cultural object. Filter-optimized and cluster-optimized data are directed towards the 3D Reconstruction and Photometric engines respectively [39, 40].

The 3D Structure Extraction engine employs algorithms for 3D structure extraction from 2D images based on multiple perspectives of the same cultural object. These perspectives have been identified by the clustering engine. The goal of the 3D Structure Extraction Engines is to produce dense clouds of 3D points that will be used in the dense 3D Reconstruction engine.

4.2 The Recommendation Engine

The main purpose of the recommendation engine is to further refine the retrievals obtained using various CBIR techniques as the ones presented in Sect. 4.1. The reason behind the proposed mechanism lies in the need of multiple uses of the same data sets in different applications (e.g. subsets of the same data can be used for 3D reconstruction, touristic promotion, books publication, etc.). An image filtering scheme for images of cultural interest has been developed [21]. The model utilize a semi-supervised approach for the creation of an appropriate distance learning metric, which is used for the filtering. User's feedback is involved only for a minor set of data, defined using optics algorithm and sparse modelling representative selection. Such approach facilitates the refinement of image data sets collected online from repositories, such Flickr, always under the scope of the end user needs.

The meta-filtering approach is based on a total ranking approach, for every available image x_j, described by the following equation:

$$r_j = \sum_{\substack{i=1 \\ i \neq j}}^{|P|} \frac{1}{\boldsymbol{w}_i^p d_{\boldsymbol{A}}(x_i, x_j)} + \sum_{\substack{i=1 \\ i \neq j}}^{|N|} \frac{1}{\boldsymbol{w}_i^n d_{\boldsymbol{A}}(x_i, x_j)} \tag{1}$$

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, $\boldsymbol{w}_i^p\left(\boldsymbol{w}_i^n\right)$ is a weight value for the importance of the i-th annotated positively (negatively) image, and $d_{\boldsymbol{A}}(x_i, x_j)$ is a distance metric defined both on user's annotated and the non-annotated images of the data set.

For any image points, we need to take the distance of them. Similar to the approach of [41], 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 S and a set of dissimilar pairwise constraints D. Both sets of constraints should be provided by the user as a relevance feedback in order to guide the problem to an acceptable solution. The problem formulation is stated as [41]:

$$\begin{aligned} & \min_A t + \gamma_s tr(\boldsymbol{A} \cdot \boldsymbol{S}) - \gamma_d tr(\boldsymbol{A} \cdot \boldsymbol{S}) \\ & s.t.\ \ tr\left(\boldsymbol{XLX}^T\boldsymbol{A}\right) \leq t \\ & \boldsymbol{A} \in S_+ \end{aligned} \tag{2}$$

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

The meta-filtering approach is based on three main phases. The first stage of the methodology aims at the detection of representative samples and its annotation as relevant or irrelevant by the user. The second stage involves the distance metric learning according to the user defined relevance sets. The final stage ranks the rest of the images using both similarity and dissimilarity rankings based on the previously stated distance metric.

The proposed methodology was applied in three cultural monuments. These monuments were Knossos, Porta Nigra and Fontana dei Quatro Fummi (see Fig. 6). For every monument, four cases of image filtering are employed. The filtering scenarios can briefly described as: (a) need for exterior images of the monument, (b) special attributes, (c) people around the monument and (d) various images (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

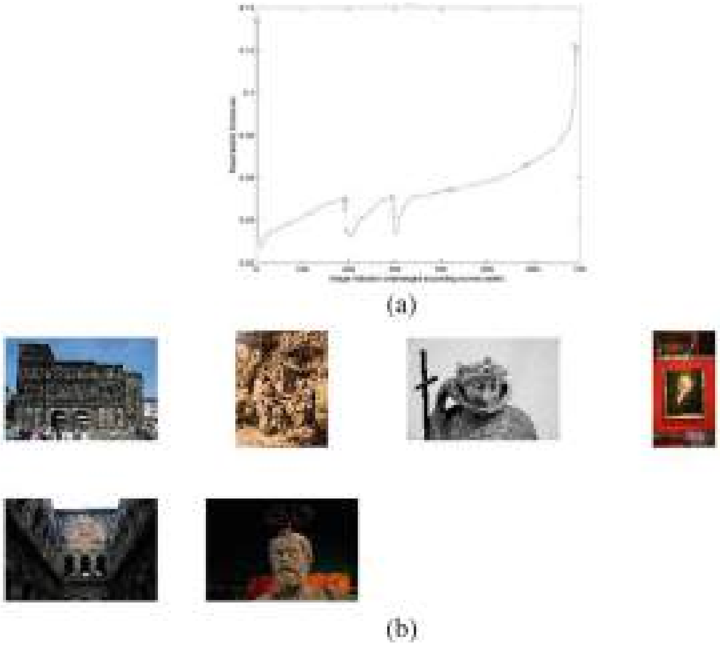

Fig. 6. (a) Optics algorithm results on the original data set of Porta Nigra. Points with 'o' mark the separation between sub-clusters. (b) The corresponding images of the marked points (Figure created by the authors and cited in [21]).

pairwise constraints shown in Eq. (2). In every case the ratio was 6 relevant to 18 irrelevant. Leading to user feedback of 24 images in total.

Finally, in Fig. 7, we present the final marking and adaptation after the feedback from the users. The results are obtained using both likes and dislikes of the users on images regarding their preferences and information needs. A predefined number of images with the higher values are retrieved as relevant. Dashed line indicates minimum value among these images.

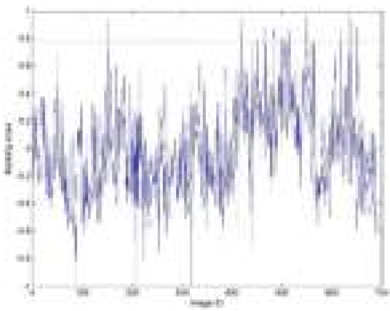

Fig. 7. Final ranking score using both accordance and discordance to the user's selection. A predefined number of images with the higher values are retrieved as relevant. Dashed line indicates minimum value among these images.

5 3D Reconstruction of the History City of Calw in Germany

The testbed area selected is the historical center of the city of Calw, a frequently visited area located 35 km southwest of Stuttgart (Germany), close to the northern part of the Black Forest (see Fig. 8). The old city counts with a broad history and rich cultural heritage. Among the 15th and 17th centuries, Calw was a commercial center of wood dissemination and processing, cloth production and salt trade, becoming an urban pole which was more important than Stuttgart, the real capital of the Federal State Baden-Württemberg, which is the third most extended and populated region (of sixteen) of Germany.

Fig. 8. (a) Location of the historical center of Calw relative to Stuttgart and (b) aerial perspective of this area.

The rich cultural heritage of Calw explains the existence of huge amounts of images, sketches, plans, drawings and maps. The Calw City Archive has several collections with high historical value, like the sketches, plans and drawings dating back to the 10th and 11th centuries, and some photos captured in the middle of the 19th century. In total, more than 3,500 photos between the years 1860 and 2014 have been offered in addition to many plans of the old town, sketches and paintings dating back to the early medieval ages (from 1,100 onwards). Furthermore, the city is still continuously visited and photographed by tourists, allowing to meet thousands of images in Open Access Image Repositories (OAIR) accessible online such as on Flickr, Picasa or Panoramio [28–30].

For acquisition of data, Leica ScanStation P20 (TLS system) and a photo camera Nikon D800E, both owned by the Institute für Photogrammetrie (IFP), were used. Additionally, several aerial images supplied by the Landesamt für Geoinformation und Landentwicklung of Baden-Württemberg (LGL-BW) were processed to obtain a complementary photogrammetric point cloud for the roof buildings (see Fig. 3). Regarding to the software was used: (a) Leica Cyclone for managing, registration, and modeling point cloud; (b) Adobe Photoshop for texture preparation, and (c) Trimble SketchUp for texture wrapping onto the 3D models.

This model scans 360° in horizontal and 270 in vertical direction around the fixed scan station. The maximum measuring range is of 120 m, with a frequency up to 1 million points per second. The Leica Scan Station P20 can register large datasets, including maps of intensity values and RGB images. The most representative technical parameters of this TLS model can be found on its product specifications.

For image capturing, the photo camera used was the Ricoh GXR, a digital compact camera of the Japanese Ricoh Company. This model has interchangeable units, each housing a lens, sensor and image processing engine. Each unit has features optimized for different purposes and environments. The lens used was a Carl Zeiss Lens Biogon 2.8/21 ZM. The images were captured with infinite focal length and using the function mode dial A (aperture priority), which leads to adjust the shutter speed for getting an optimal time exposure of the images.

Two types of images were used: (a) close-range, and (b) aerial images. Both imagery data were used for covering some deficiencies shown in the TLS point cloud and for obtaining additional information such as related to textures. Close-range images were captured from different positions at ground level.

Additionally, six aerial images of the historical center of Calw city were used for reconstructing the roof of the church, where TLS data could not be obtained. These images were supplied by the Landesamt fur Geoinformation und Landentwicklung of Baden-Württemberg (LGL-BW) and were captured with a ground sampling distance of 10 cm.

The different data sources (laser and image data) are integrated into a unique coordinate system in a procedure called registration. This procedure is performed into two steps: firstly the registration of the successive TLS stations and, finally, the registration are included with the photogrammetric point clouds. TLS scan stations are merged and registered into a unique point cloud referred into a local coordinate system. However, this point cloud shows some gaps such in the roof areas. Thus, a complete reconstruction of the building is not feasible due to missing data. A detailed description of the point clouds is shown in Fig. 9.

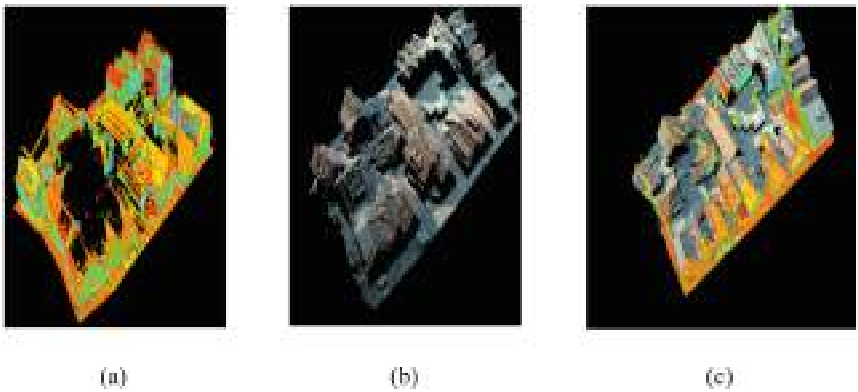

Fig. 9. Point cloud obtained by means of (a) TLS and (b) aerial photogrammetry, and (c) both merged and registered (Figure created by the authors and cited in [29, 30]).

Roof area can be reconstructed using aerial images. For this, a point cloud derived from these aerial images is created. SURE software, developed by the IFP, is used to extract geospatial information from aerial images with a remarkable level of overlap. Software performance is based on the algorithm called Semi Global Matching. Two steps are carried out: (a) determining the image orientation manually or automatically, and followed with a Bundle Block Adjustment and (b) applying Dense Image Matching algorithm to retrieve geospatial information from each single overlapped pixel, based on the first step.

A system of constraints is established by means of equivalent tie-points in the different point clouds, which are partially overlapped. The Iterative Closest Point (ICP) algorithm, implemented in the used software, performs rigid registration for at least two given laser point clouds which already partially aligned to merge. Basically, two steps are carried out: (a) to create a pairing between the same scanned points from different scan stations, and (b) to compute rigid registration for matching these points and minimizing the distances in between (see Fig. 10). Some images of the final 3D model reconstructed are exposed in Figs. 11 and 12. From both figures, we can see the high detailed model and the accuracy achieved. More specifically, Fig. 11 presents the 3D model of the protestant church of St. Peter and Paul and Fig. 12 a block of some characteristic buildings of the town.

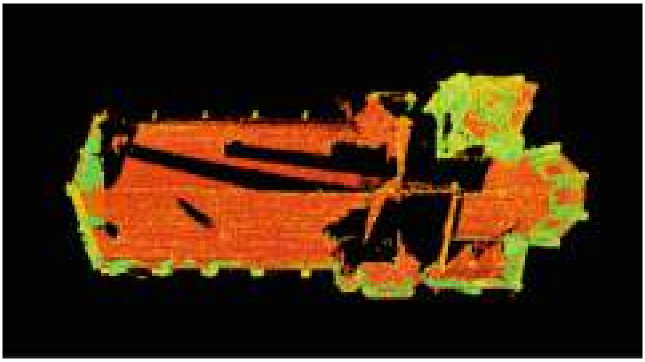

Fig. 10. Top overview of the obtained TLS point cloud. Internal black gaps correspond to missing data (Figures created by the authors and cited in [29, 30]).

6 The 4D Viewer

6.1 Modelling for 3D Realtime and VR Environments Production Pipeline

The results of a computer generated 3D model through SfM and/or dense matching is bound to the quality and quantity of the source images available. It reflects the time epoch for which all images were taken. Missing features of a model as well as anomalies in the resulting 3D model can be substituted through a camera matched, manual modelling approach. This technique also allows the 3D reconstructions using historic photos. However, the results obtained might be subjective and are depending on the operator's personal perception and his skills in 3D modelling. Nevertheless a

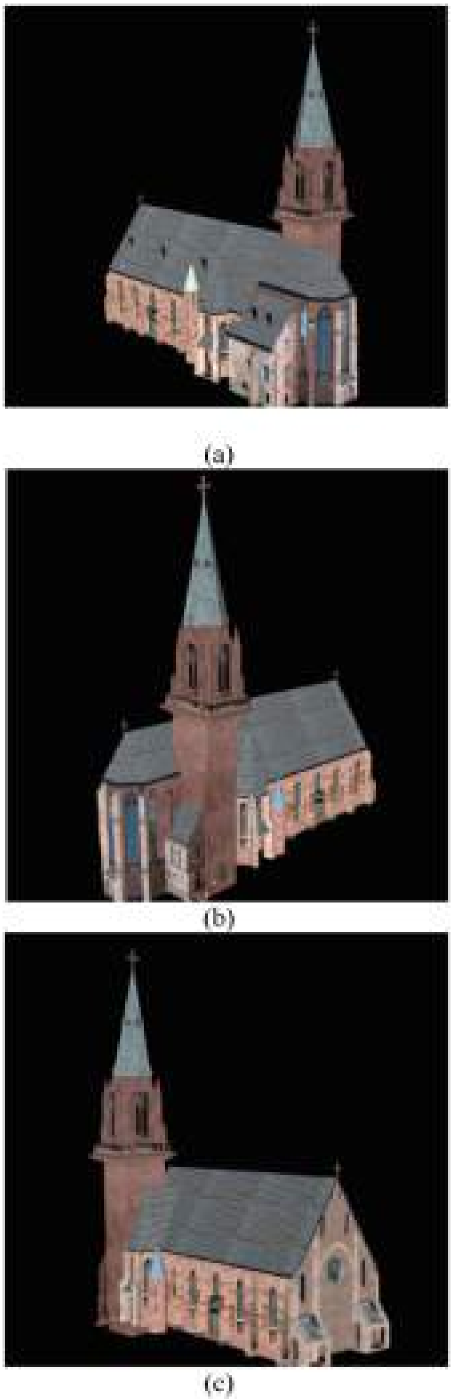

Fig. 11. Different viewpoints on the 3D photorealistic model of the church (Figure created by the authors).

Fig. 12. 3D building model obtained in a certain area surveyed.

comparison between a SfM calculated structure and the para-metrical model can lead to an adequate accuracy level. As once defined by Paul Debevec (1996) [42], an expert of computer vision, his interpretation of photogrammetry is as follows: "A method for interactively recovering 3D models and camera positions from photographs." [42].

The city of Calw in Germany has been chosen as a test site for 3D reconstruction where different approaches were applied to model the central square of the settlement in different time settings (see Sect. 5). According to the project plan not only the central square of the town, but also the somewhat larger historical settlement core was modeled.

The results will be made accessible in an interactive virtual 3D realtime environment, powered by a game engine allowing the user to walk or fly-through the 3D model of the settlement and retrieve additional information on places and buildings on his demand. Further the switch into different time periods of the environment or a building will be possible depending on the accuracy and quantity of the source material which is necessary to model out the complete set of structures [31]. In Fig. 13, we show some of the 3D reconstruction results.

In order to automate the process of image rectification one could also use the calculated camera positions generated by a SfM processing and therefore bypass the manual alignment of the vanishing points as described above, eliminating subjective factors in the rectification process. This however requires more than one view to allow for the simultaneous image data processing in a bundle block adjustment.

Fig. 13. Historical image and the resulting 3D model of the City Hall of Calw (Figure created by the authors and cited in [29]).

To illustrate this process, we set up a test scene consisting a 3D model of the Hermann-Hesse Birthplace building, which was previously constructed through a manual camera match approach, from which several rendered images from different viewpoints were taken to provide a source for the SfM model generation. Any SfM sofware, such as VisualSfM, will deliver the camera positions, orientations and a sparse 3D model, which can be reimported to a 3D modelling program (e.g. Autodesk 3Ds Max). Inside the 3D modelling program one of the reconstructed camera views was chosen to verify the matching of the perspective with the background image, which served as a template for a newly constructed geometry, representing the outlines of the object (house). Since the constructed geometry, matched seamlessly with the vanishing points of the projected background image it can be confirmed that the camera match calculated from the SfM procedure is adequate for this purpose. An example of the camera setup is shown in Fig. 14.

Fig. 14. Set-up with the calculated camera position and the 3D Model of Hermann Hesses Birthplace (Figure created by the authors and cited in [29]).

Within the photogrammetric and GIS community some standard data formats have established. For semantic building and city models the OGC standard CityGML is well-known. However, the computer graphics community deals with completely different data formats, such as Collada or 3Ds max. Suitable data conversion therefore have to be performed to fuse information given by modelling exports with photogrammetric processing steps. Figure 15 exemplarily shows the Hermann Hesse birthplace, which has been selected for testing purposes and has been simplified as input for the reconstruction algorithm. Top left depicts the textured model of the building, top right the corresponding point cloud. The reconstructed façade is shown on the bottom left and the suggested completion of the model in CityGML is depicted on the bottom right.

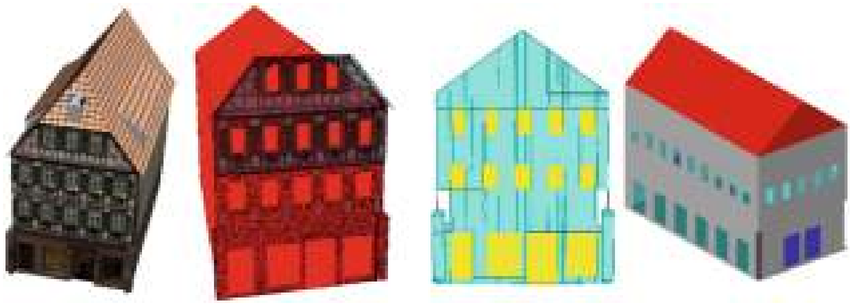

Fig. 15. From Left to Right: Textured model of the Hermann Hesse birthplace, simplified CityGML model with an overlying point cloud, reconstructed primitives, bottom right: grammar based completion (Figure created by the authors).

6.2 The 4D Viewer

The main role of a 4D viewer is to depict not only the 3D geometry of a CH asset but also its evolution in time. For this reason, time parameters have been added in the viewer. The user can also manage different angles of a monument and different cameras views. Such extensibility has a huge impact in the field of cultural heritage especially for city views and city momentums where big data manipulation over the web can be easily controlled by any end user using parameters to scroll through (i) time and (ii) 3D object manipulation [43]. This is also very important in searching CH content taking into account time constraints. For instance, image a monument under different weather conditions or its impact on different climate phenomena and material decay factors.

On the bottom of the screen a slider is provided which allows the user to move through different time periods accompanied with the images of the monument in the current period [44]. Sliding over the images the user can navigate in time and across different parts of the model (Fig. 16). The slice tool enables the user to view intersection of the reconstructed model with the help of a slicing plane (Fig. 17). This provides an insight to the interior construction of the model.

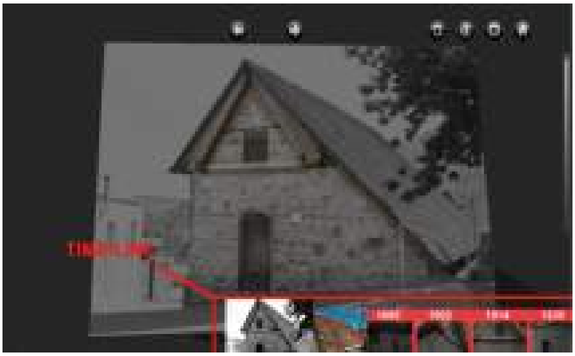

Fig. 16. The time-line feature in the VR viewer (Figure created by the authors and cited in [44]).

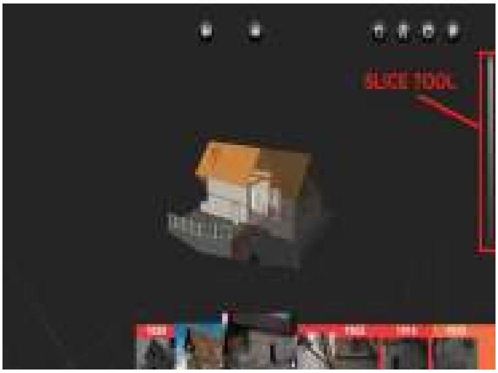

Fig. 17. The slice tool in the VR viewer (Figure created by the authors and cited in [44]).

7 Conclusions

Time dimension is a crucial part of a precise 3D reconstruction. This is especially critical for outdoor CH monuments that undergo natural decay due to environmental phenomena and disasters. 3D modelling a CH object in time independently is a process which is time consuming, tedious and of great effort. This is mainly due to the fact that most of the geometric structures of a monument remains intact through time. To overcome this issues, in this chapter we investigate innovative aspects regarding 4D modelling of cultural heritage items that is, 3D geometrically reconstruction plus the time.

Our research focuses on unstructured visual items collected from distributed multimedia repositories located over the web or from visual content uploaded on the short tweet posts. For the first case, we implement a clustering algorithm enabling the removal of the outliers. For the second case, we develop an event detection scheme from the tweets messages and then we apply an image clustering to find the most representative data to trigger a computationally efficient 3D reconstruction through time. A search engine is developed accompanying with a recommendation system.

3D modelling is evaluated on a historic German city, that is, of Calw. The results derived are of high resolution and reconstruction accuracy. The 3D models are embedded on a 4D viewer the key functionality of which is the time slide. This allows for the users to navigate through time on content of interest.

Acknowledgement. This work is supported by the FP7 project: Four Dimensional Cultural Heritage World (4D CH World) funded by European Union under the grant agreement 324523 and the H0202 project: Transforming Intangible Folkloric Performing Arts into Tangible Choreographic Digital Objects (TERPSICHORE) funded by RISE European Union programme under the grant agreement 691218.

References

1. Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C.: Methods for 3D digitization of cultural heritage. J. Cult. Herit. **8**(1), 93–98 (2007)
2. Bond, A., Langstaff, L., Baxter, R., Kofoed, H.G.W.J., Lisitzin, K., Lundström, S.: Dealing with the cultural heritage aspect of environmental impact assessment in Europe. Impact Assess. Project Apprais. **22**(1), 37–45 (2004)
3. Adger, W.N., Barnett, J., Brown, K., Marshall, N., O'Brien, K.: Cultural dimensions of climate change impacts and adaptation. Nat. Climate Chang. **3**(2), 1666 (2012)
4. Kyriakaki, G., Doulamis, A., Doulamis, N., Ioannides, M., Makantasis, K., Protopapadakis, E., Hadjiprocopis, A., Wenzel, K., Fritsch, D., Klein, M., Weinlinger, G.: 4D reconstruction of tangible cultural heritage objects from web-retrieved images. Int. J. Herit. Digit. Era **3**(2), 431–452 (2014)
5. Ioannides, M., Hadjiprocopis, A., Doulamis, N., Doulamis, A., Protopapadakis, E., Makantasis, K., Santos, P., Fellner, D., Stork, A., Balet, O., Julien, M.J., Weinlinger, G., Johnson, P.S., Klein, M., Fritsch, D.: Reconstruction using multi-images available under open access. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Strasbourg, (1), pp. 169–174, 2–6 September 2013
6. Gomes, L., Pereira Bellon, O.R., Silva, L.: 3D reconstruction methods for digital preservation of cultural heritage: a survey. Pattern Recogn. Lett. **50**, 3–14 (2014)
7. Ene, D., Radvan, R.: Digital model for cultural heritage conservation status evaluation. In: 4th International Conference on Developments in eSystems Engineering, DeSE, pp. 627–631 (2011)
8. Kitamura, K.: Common software for digital exhibition of Japanese cultural heritage in literature. In: International Conference on Culture and Computing, Culture and Computing, 6680347, pp. 137–138 (2013)
9. Bjerke, B., Karlsson, M.: Social Entrepreneurship: To Act as if and Make a Difference. Edward Elgar Publishing, Cheltenham (2013)
10. Doulamis, A., Soile, S., Doulamis, N., Chrisouli, C., Grammalidis, N., Dimitropoulos, K., Manesis, C., Potsiou, C., Ioannidis, C.: Selective 4D modelling framework for spatial-temporal land information management system. In: Proceedings of SPIE - The International Society for Optical Engineering, Paphos, Cyprus, vol. 9535 (2015)
11. Doulamis, A., Doulamis, N., Ioannidis, C., Chrysouli, C., Grammalidis, N., Dimitropoulos, K., Potsiou, C., Stathopouloua, E.K., Ioannides, M.: 5D modelling: an efficient approach for creating spatiotemporal predictive 3D maps of large-scale cultural resources. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Taipei, Taiwan, vol. 2(5), p. 61 (2015)

12. Doulamis, N., Doulamis, A., Varvarigou, T.: Adaptive algorithms for interactive multimedia. IEEE Multimedia Mag. **10**(4), 38–47 (2003)
13. Avrithis, Y.S., Doulamis, N.D., Doulamis, A.D., Kollias, S.D.: Efficient content representation in MPEG video databases. In: Proceedings of IEEE International Workshop on Content based Access Interactive Video Libraries in Conjunction with the IEEE Computer Vision and Pattern Recognition Conference (CVPR), Santa Barbara, CA, USA, pp. 91–95, June 1998
14. Makantasis, K., Doulamis, A., Doulamis, N., Ioannides, M.: In the wild image retrieval and clustering for 3D cultural heritage landmarks reconstruction. Multimedia Tools Appl. **75**(7), 3593–3629 (2016)
15. Doulamis, A., Ioannides, M., Doulamis, N., Hadjiprocopis, A., Fritsch, D., Balet, O., Julien, M., Protopapadakis, E., Makantasis, K., Weinlinger, G., Johnsons, P.S., Klein, M., Fellner, D., Stork, A., Santos, P.: 4D reconstruction of the past. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 8795, p. 87950J (2013)
16. Ntalianis, K., Doulamis, N.: An automatic event-complementing human life summarization scheme based on a social computing method over social media content. Multimedia Tools Appl. **75**(22), 15123–15149 (2016)
17. Doulamis, N.D., Doulamis, A.D., Kokkinos, P., Varvarigos, E.: Event detection in Twitter microblogging. IEEE Trans. Cybern. **46**(12), 2810–2824 (2016)
18. Ntalianis, K., Mastorakis, N., Doulamis, A., Tomaras, P.: Social media video collection summarization based on social graph user interactions. In: Recent Advances In Knowledge Engineering And Systems Science, pp. 162–167 (2013)
19. Makantasis, K., Doulamis, A., Doulamis, N.: A non-parametric unsupervised approach for content based image retrieval and clustering. In: Proceedings of the 4th ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream in Conjunction with ACM Multimedia, Barcelona, pp. 33–40, October 2013
20. Doulamis, A., Katsaros, G.: 3D modelling of cultural heritage objects from photos posted over the Twitter. In: IEEE International Conference on Imaging Systems and Techniques (IST), Chania, Crete, October 2016
21. Protopapadakis, E., Doulamis, A., Matsatsinis, N.: 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.) EuroMed 2014. LNCS, vol. 8740, pp. 102–110. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13695-0_10
22. Yiakoumettis, C., Doulamis, N., Miaoulis, G., Ghazanfarpour, D.: Active learning of user's preferences estimation towards a personalized 3D navigation of geo-referenced scenes. GeoInformatica **18**(1), 27–62 (2014)
23. Doulamis, N., Yiakoumettis, C., Miaoulis, G., Protopapadakis, E.: A constraint inductive learning-spectral clustering methodology for personalized 3D navigation. In: Bebis, G., et al. (eds.) ISVC 2013. LNCS, vol. 8034, pp. 108–117. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41939-3_11
24. Doulamis, N., Yiakoumettis, C., Miaoulis, G.: Personalised 3D navigation and understanding of geo-referenced scenes. In: IEEE 14th International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2013, Spain, 4–7 June 2013
25. Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S.: Relevance feedback: a power tool for interactive content-based image retrieval. IEEE Trans. Circuits Syst. Video Technol. **8**(5), 644–655 (1998)
26. Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. Read. Inf. Retr. **24**(5), 355–363 (1997)

27. Doulamis, N., Doulamis, A.: Evaluation of relevance feedback schemes in content-based in retrieval systems. Sig. Process. Image Commun. **21**(4), 334–357 (2006)
28. Fritsch, D., Klein, M.: 3D preservation of buildings – reconstructing the past. Multimedia Tools Appl. 1–18 (2017). article in press
29. Voulodimos, A., Doulamis, N., Fritsch, D., Makantasis, K., Doulamis, A., Klein, M.: Four-dimensional reconstruction of cultural heritage sites based on photogrammetry and clustering. J. Electron. Imaging **26**(1), 011013 (2017)
30. Balsa-Barreiro, J., Fritsch, D.: Generation of 3D/4D photorealistic building models. The testbed area for ***4D Cultural Heritage World*** project: the historical center of Calw (Germany). In: Bebis, G., et al. (eds.) ISVC 2015. LNCS, vol. 9474, pp. 361–372. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-27857-5_33
31. Doulamis, A., Doulamis, N., Makantasis, K., Klein, M.: A 4D virtual/augmented reality viewer exploiting unstructured web-based image data. In: 10th International Conference on Computer Vision, Theory and Applications, VISAPP, Berlin, March 2015
32. Doulamis, N., Yiakoumettis, C., Miaoulis, G.: On-line spectral learning in exploring 3D large scale geo-referred scenes. In: Ioannides, M., Fritsch, D., Leissner, J., Davies, R., Remondino, F., Caffo, R. (eds.) EuroMed 2012. LNCS, vol. 7616, pp. 109–118. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34234-9_11
33. Maltezos, E., Doulamis, A., Ioannidis, C.: Improving the visualisation of 3D textured models via shadow detection and removal. In: 9th IEEE International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), Athens, Greece, September 2017
34. Kyriakaki, G., Doulamis, N.: Metadata framework for long-term preservation of digital cultural experiences: the 'Viopolis' case. In: Proceedings of the 11th International Conference on Applications of Electrical and Computer Engineering, Athens, Greece, May 2013
35. Aizawa, A.: The feature quantity: an information theoretic perspective of TFIDF-like measures. In: ACM SIGIR 2000, pp. 104–111 (2000)
36. Xue, X.B., Zhou, Z.H.: Distributional features for text categorization. IEEE Trans. Knowl. Data Eng. **21**(3), 428–442 (2009)
37. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571 (2011)
38. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: Density-based spatial clustering of applications with noise. In: International Conference on Knowledge Discovery and Data Mining, vol. 240 (1996)
39. Abdel-Wahab, M., Wenzel, K., Fritsch, D.: Automated and accurate orientation of large unordered image datasets for close-range cultural heritage data recording. Photogramm. Fernerkund. Geoinf. **2012**(6), 679–689 (2012)
40. Ntalianis, K., Doulamis, A., Doulamis, N., Tsapatsoulis, N.: Video abstraction in social media: augmenting Facebook's Edgerank algorithm in video content presentation. In: IEEE International Conference in Image Processing (ICIP), Australia (2013)
41. Hoi, S.C., Liu, W., Chang, S.F.: Semi-supervised distance metric learning for collaborative image retrieval and clustering. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **6**(3), 18 (2010)
42. Debevec, P.E., Taylor, J.C., Malik, J.: Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In: Proceedings of the 23rd ACM Annual Conference on Computer Graphics and Interactive Techniques. ACM (1996)

43. Sardis, E., Anagnostopoulos, V., Doulamis, N., Varvarigou, T., Kyriakaki, G.: From 2D to 3D and 4D using timelines in mobile cultural experiences. In: Euromed 2014, Limassol, Cyprus (2014)
44. Doulamis, N., Doulamis, A., Ioannidis, C., Klein, M., Ioannides, M.: Modelling of static and moving objects: digitizing tangible and intangible cultural heritage. In: Ioannides, M., Magnenat-Thalmann, N., Papagiannakis, G. (eds.) Mixed Reality and Gamification for Cultural Heritage, pp. 567–589. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-49607-8_23