A 4D Virtual/Augmented Reality Viewer Exploiting Unstructured Web-based Image Data

Anastasios Doulamis[1], Nikolaos Doulamis[1], Konstantinos Makantasis[2] and Michael Klein[3]
[1]*Computer Vision Photogrammetry Lab.., Technical University of Athens, 9th Heroon Polytechniou str., Athens, Greece*
[2]*Department of Production and Survey Engineering, Technical University of Crete, Polytechnioupolis, Crete, Greece*
[3]*7 Reasons Company, ltd, 4-6th Bäuerlegasse, Vienna, Austria*

Keywords: 4D Modelling, 3D Reconstruction, Content-based Filtering, Web-based Images, Virtual/Augmented Reality.

Abstract: Outdoor large-scale cultural sites are mostly sensitive to environmental, natural and human made factors, implying an imminent need for a spatio-temporal assessment to identify regions of potential cultural interest (material degradation, structuring, conservation). Thus, 4D modelling (3D plus the time) is ideally required for preservation and assessment of outdoor large scale cultural sites, which is currently implemented as a simple aggregation of 3D digital models at different time. However, it is difficult to implement temporal 3D modelling for many time instances using conventional capturing tools since we need high financial effort and computational complexity in acquiring a set of the most suitable image data. One way to address this, is to exploit the huge amount of images distributing over visual hosting repositories, such as flickr and picasa. These visual data, nevertheless, are loosely structured and thus no appropriate for 3D modelling. For this reason, a new content-based filtering mechanism should be implemented so as to rank (filter) images according to their contribution to the 3D reconstruction process and discards image outliers that can either confuse or delay the 3D reconstruction process. Then, we proceed to the implementation of a virtual/augmented reality which allows the cultural heritage actors to temporally assess cultural objects of interest and assists conservators to check how restoration methods affect an object or how materials decay through time. The proposed system has been developed and evaluated using real-life data and outdoor sites.

1 INTRODUCTION

Digitalizing cultural sites and objects and creating 3D digital models is an important task to preserve Cultural Heritage (CH). 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. This implies an imminent need for a spatio-temporal monitoring of those sites to identify regions of potential material degradation, and unstable structuring conditions. Thus, a time varying 3D model (i.e., 4D modelling-3D geometric dimensions plus the time) should be developed to assess spatial and temporal diversity of CH objects but again under a cost-effective framework able to be applied to large-scale sites.

One main difficulty of implementing temporal 3D modelling is the complexity, and the respective financial effort, in acquiring a set of images required for the 3D reconstruction. One way to address this, is to exploit the huge amount of images distributing over visual hosting repositories, such as flickr and picasa (Doulamis et al, 2013). However, the main functionality of the existing web-based visual repositories is to socially share multimedia content, instead of archiving the images in a way that allows efficient and precise 3D modelling of objects of interest. This unstructured organization of images, with respect to 3D reconstruction, imposes new tools and methods in the area of content based filtering; ranking (filtering) images according to their contribution to the 3D reconstruction process while at the same time discarding image outliers that can either confuse or delay the 3D reconstruction process

Content Based Image Retrieval (CBIR) methods can be considered as the first approaches for organizing multimedia content (Murthy et. al., 2010). The main goal of a CBIR method is to find a set of images whose contents are similar to, or even match with, a given query from within a large image

Doulamis A., Doulamis N., Makantasis K. and Klein M..
A 4D Virtual/Augmented Reality Viewer Exploiting Unstructured Web-based Image Data.
DOI: 10.5220/0005456806310639
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (MMS-ER3D-2015), pages 631-639
ISBN: 978-989-758-090-1

database. According to (Daras et. al., 2012) the combination of multiple 3-D object descriptors can achieve better retrieval accuracy than a single descriptor vector alone. Thus, research should focus not only on the investigation of the optimal descriptor but also on the appropriate combination of low-level descriptors as well as on the selection of the best features and matching metrics. In particular, in the SHREC'11 search and retrieval competition, the best performance is achieved by combining a) the Spectral Decomposition of the Geodesic Distance Matrix and b) the Scale Invariant Feature Transform for meshes (meshSIFT) (Smeets et. al., 2009).

Towards this direction, (Murthy et. al., 2010) introduces a two layer image retrieval algorithm that exploits hierarchical clustering based on colour features. In the same context, the retrieval system of (Chum et. al., 2007) returns all different views of an object upon a user's query. However, the main limitation of these works is that they require a query image to carry out the retrieval process which is not suitable in our case; better organization of images with respect to 3D reconstruction.

Unsupervised content based organization has been introduced in (Kekre *et al.*, 2010) to create codebooks based on colour descriptors. Another approach exploits fuzzy Support Vectors Machines (fSVMs) to cluster visual information while visual encoding is carried out on the use of dominant colour descriptor (Min and Cheng, 2009). The main limitations of these approaches rely on the usage of global image features to encode visual content. Therefore, they are not suitable for an efficient content-based organization towards 3D reconstruction due to the fact that global descriptors fail to represent the different view instances of an object since they are not able to capture geometric characteristics.

Some other techniques filter out the retrievals using textual and/or geo-location information. The tools combine geo-information along with a hierarchical clustering method that exploits visual features to obtain dense groups (Arampatzis et. al, 2013) and (Papadopoulos et. al., 2010). Again, these methods are based on global descriptors, failing to represent image geometry.

A content-based filtering method suitable for organizing visual content located over distributed web-based image repositories for the purposes of 3D reconstruction have been proposed in (Makantasis et. al., 2014). The algorithm exploits local geometric properties and density based clustering methods to organize unstructured multimedia content in a way to accelerate 3D reconstruction, while keeping its performance as precise as possible. A semi-supervised approach is presented in (Protopapadakis et al, 2014)

While the work of (Makantasis et. al., 2014) introduces a method for efficiently ranking images according to their contribution to the 3D reconstruction process, it fails to incorporate the results into a virtual and augmented reality interface. The latter not only exploits the organization of the images in order to achieve 3D modelling but also permits the users to spatio-temporally assess the derived 3D models (Hadjiprocopis et. al, 2014). Only under such a framework, the content based organization methods of (Makantasis et. al., 2014) can be exploited in real-life conditions. The developed scheme assists conservators to check how restoration methods affect an object or how materials decay through time, archaeologists to better document an item, curators to properly display them, and the creative industries to disseminate cultural knowledge in a digestive way worldwide. Today, the construction of high fidelity 3D models is a time consuming task, with limited functionalities since it cannot capture time evolutions properties of an item (how it behaves in time) and scalable functionalities needed for different types of CH actors. In addition, the time required to get a precise model is often too high due to the manual effort needed to be interwoven in the reconstruction process. For this reason, 3D digitalization is mainly applied to individualized items of museums' collections, to indoor cultural assets where temporal variations is minimal and to famous sites/museums where adequate financial resources can be given to perform such a digitalization.

This paper is organized as follows. In Section 2, we discuss the algorithms used for modelling the retrieved images under a geometrically invariant framework. In this section, we also discuss the methods for transforming the image data into multidimensional key-points onto a manifold in a way that the distance between two points coincide with the visual similarity distance between the two images. Then, Section 3 presents the content based filtering algorithms used for structuring the retrieved data from distributed Web repositories for 3D reconstruction purposes. This section includes the algorithm used for removing the image outliers as well as the methods used for extracting the most appropriate images for 3D reconstruction. Section 4 discuss the 4D CH viewer as well the respective augmented reality functionalities that allow for the end-users to overlay, track and manipulate 3D reconstructed objects over 2D interfaces. Finally, Section 5 presents the experimental results along with 4D cultural heritage viewer representation, while Section 6 concludes the paper.

2 CONTENT MODELLING

2.1 Geometrically Invariant Modelling

Initially, the ORB (Oriented FAST and Rotated BRIEF) is used for locally representing the visual properties of an image (Rublee et. al., 2011). ORB extracts a set of image keypoints which are then used to describe geometric characteristics of an image under an invariant affine transformation way. Selection of ORB compare to other local descriptors, like SURF, SIFT, is that it gives the same performance as SIFT while being two orders of magnitude faster.

Then, the visual similarity between two images is computed. Visual similarity is measured over the correspondence points of the two images. The correspondence points are estimated by applying a nearest neighbour matching algorithm on the extracted key-points. In this paper, the local sensitive hashing approach and the hamming distance are adopted (Lv et. al., 2007) as far as the matching algorithm is concerned since the extracted key-points of ORB are described as a binary pattern. Let us denote as $k_i^{(A)}$ and $k_{j_i}^{(B)}$ two correspondent key points between two images A and B. Then, we form a set $M^{(A\to B)}$ that contains the correspondent key-points from the image A to B and a set $M^{(B\to A)}$ that contains the respective points from B. The intersection between those two sets defines the 2-way matching. Then Then, that the similarity between images A and B can be defined as

$$s_{i=A,j=B} = \left\| M^{(A,B)} \right\| / K \tag{1}$$

where $\left\| M^{(A,B)} \right\|$ refers to the cardinality of $M^{(A,B)}$ set and K the number of extracted key-points.

Using the aforementioned process for all N retrieved images, we conclude to an NxN symmetric matrix **S** whose elements $s_{i,j} \in [0\ 1]^T$. Values close to zero indicate no relation between the two images, while values near one a high relationship degree. The visual dissimilarity matrix **D** is defined as the logarithm of $\boldsymbol{S}$, $\mathbf{D} = [d_{i,j}] = -\log(\mathbf{S})$.

2.2 Multi-dimensional Manifolds

By exploiting the similarity matrix **D**, we can represent the N retrieved images as single points onto a multidimensional manifold. In particular, let us define as $\mathbf{x}^{(i)} \in R^{\mu}$ the coordinates of the i-th image onto the μ-dimensional space. Then, using the multi-dimensional scaling method (cMDS) [see (Cox and Cox, 2000)], we relate the coordinates of the μ-dimensional image points with the similarity distance matrix **D**.

In particular, images that are visually similar would be mapped onto points of the multidimensional manifold, which are located "closely enough" in the subspace. On the contrary, image outliers will be spread far away. The multi-dimensional manifold is constructed in a way so that the distance between two image points of the μ-dimensional space coincides with the respective similarity distance $d_{i,j} = -\log(s_{i,j})$ or $\left\| \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \right\| = d_{i,j}\ \forall i,j$.

In the sequel, we exploit the concepts of cMDS (Cox and Cox, 2000) to establish a connection between the space of the distances and the space of Gram matrix $\mathbf{B} = \mathbf{X} \cdot \mathbf{X}^T$ (Cayton, 2006). **X** is the matrix that contains all N image coordinates $\mathbf{x}^{(i)}$.

More specifically, matrix **D** is an Euclidean distance matrix if and only if $\mathbf{B} = -1/2 * \mathbf{H} \cdot \mathbf{D} \cdot \mathbf{H}$, where $\mathbf{H} = \mathbf{I} - 1/N * \mathbf{1}\mathbf{1}^T$ a positive semi-definite matrix with **I** the unity one and **1** a vector of all ones elements. Furthermore, this **B** will be the Gram matrix for a mean centred configuration with interpoint distances given by **D.** Assuming non Euclidean spaces, matrix **B** as described above will not be positive semi-definite, and thus it will not be a Gram matrix. To handle such cases, cMDS projects the Gram matrix **B** onto the cone of positive semi-definite matrices by setting its negative eigenvalues to zero.

Having estimated the Gram matrix **B**, we are able to get matrix **X** by spectrally decomposing **B** into $\mathbf{U} \cdot \mathbf{V} \cdot \mathbf{U}^T$ and then $\mathbf{X} = \mathbf{U} \cdot \mathbf{V}^{1/2}$. Let us now denote as q_i i=1,2,…N the eigenvectors of **B** and as λ_i the respective eigenvalues. Then matrix **U** is the square NxN matrix whose i-th column is the eigenvector q_i of **B** and **V** is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. Finally the dimension μ of the multidimensional space is equal to the multiplicity of non-zero eigenvalues of matrix **B**.

3 SELECTION OF THE MOST REPRESENTATIVE IMAGES FOR 3D RECONSTRUCTION

In this section, we describe the tools applied to (i) remove image outliers and (ii) select a set of representative images enabling an as much as possible precise 3D reconstruction.

3.1 Outliers Removal

Since in our system 3D reconstruction takes place from a set of loosely structured imagery data located over distributed web-based repositories, we expect that several images that fit a specific category will be in fact outliers. A density-based visual clustering algorithm is adopted in this paper to remove the outliers. As we have no prior knowledge regarding outliers density parameters (e.g., the area of the outliers' region and how dense it is), we apply a procedure to automatically estimate them as in (Makantasis et. al, 2014).

Then, a modified version of the DBSCAN, density based clustering algorithm, named Core Sample Partitioning (CSP), is selected to remove the outliers. The conventional version of DBSCAN creates a compact subset by including all points of the multi-dimensional manifold that are either density reachable or density connected to a core sample. Two points are density reachable if they are either direct reachable, that is, they belong to the same density area with respect to a similarity distance, or there exists a chain of points that some of them are directly reachable.

The main drawback of the conventional DBSCAN algorithm in the context of our paper is that, although it minimizes the probability of excluding relevant images, it also includes some of the outliers in the target subspace. On the contrary, CSP exploits the notion of directly density-reachability, creating a set that minimizes the probability of an image outlier to belong to the partitioned subset of relevant images. If we suppose that a large enough set of images for an object is available, the proposed modified DBSCAN approach selects images for the 3D reconstruction process that yield low computational complexity, while its precision performance remains almost the same.

3.2 Representative Images for 3D Reconstruction

The modified CSP algorithm discards image outliers since it detects sparse image samples spreading far away from the dense subspace of the “relevant images” onto the multi-dimensional manifold. This way, we exclude the visual space into two areas; the one of the outliers and the one of the relevant images. Having detected the relevant set, we need then to proceed to the extraction of a small set of representative images of the object of interest that can maximize geometric depiction. This is achieved in this paper using a spectral clustering method. The advantages of the spectral clustering algorithm is that it treats clustering as a graph partitioning problem without making specific assumptions on the form of the created clusters (Bach and Jordan, 2003).

Spectral clustering optimally solves a graph partitioning problem so that (i) elements within a cluster present the maximum coherence, while (ii) elements across data presents the minimum coherence. This is achieved by estimating the similarity degree among the partition vertices. That is, this similarity across vertices belonging to the same cluster is expected to increase while, on the contrary, the similarity among the vertices of different partitions is expected to decrease. To avoid convergence of the algorithm to trivial solutions in which a cluster consists only of one element, normalization factors are imposed in the minimization process.

Since a graph can be straightforward represented in a matrix representation, we can formulate the aforementioned twofold minimization process in a matrix form as

$$\hat{\mathbf{a}}_r : \min \sum_{r=1}^{M} \frac{\mathbf{a}_r^T \cdot (\mathbf{L}-\mathbf{E}) \cdot \mathbf{a}_r}{\mathbf{a}_r^T \cdot \mathbf{L} \cdot \mathbf{a}_r} \quad (2)$$

where $\mathbf{a}_r = [\ldots a_r^u \ldots]^T$ is an index vector, whose the u-th entry equals to unity whether the respective u-th image is assigned to the r-th partition C_r, and zero otherwise.

We also denote as $\mathbf{E}$ the adjacent matrix of the graph G. Therefore, we have that matrix $\mathbf{E}$ is expressed as $\mathbf{E} = [w_{ij}]$ where the elements of w_{ij} are modelled as $w_{ij} = d_{ij} = -\log(s_{ij})$. We recall that $s_{i,j}$ is the similarity matching between the images i and j respectively.

Let us also denote as $\mathbf{Z}$ the degree matrix of the graph G as a diagonal matrix $\mathbf{Z} = diag(\ldots z_i \ldots)$ with

$z_i = \sum_{j \in C} w_{ij}$. Then, we can define the Laplacian matrix of the graph G as **L**=**Z**-**E**.

In order to solve the aforementioned optimization problem, we need to relax the index vector of $\mathbf{a}_r = [\ldots a_r^u \ldots]^T$ to take continuous values instead of binary ones. This means that we assume that each image is possible to be assigned to all potential clusters but of different degree of membership. The relaxed continuous solution of $\mathbf{a}_r = [\ldots a_r^u \ldots]^T$ is obtained by exploiting the Ky-Fan theorem (Fan, 1951). Then, a rounding process is adopted to discretize the continuous solution into a binary one.

The approach adopted in this paper treats the rows of the continuous matrix as M-dimensional feature vectors. Each row indicates the degree of "fitness" (the association degree) of the corresponding image to each of the M available clusters. It has been shown in (Bach and Jordan, 2003) that the aforementioned approximation provides the minimum Frobenius distance between the continuous and the discrete solution. Thus, this is the closest approximate solution to the continuous optimum provided by the Ky-Fan theorem. The well-known k-means algorithm is applied to optimally estimate the most suitable clusters. The number of clusters is set equal to the parameter M. We recall that M is the number of estimated partitions of graph G. When we increase the value of variable M, we lead to the selection of a greater number of images used for 3D reconstruction and thus for a more precise performance. This increment, however, also yields to an increase of the computational complexity as well. The estimation of an appropriate value for parameter M takes place in regard to application scenarios. Different scenarios suggest different constraints in regard to devices' computational power and available memory as well as the desired re-construction time.

Then, we extract per each cluster the image point located closer to the its centre as the most representative one.

4 VIRTUAL AND AUGMENTED REALITY INTERFACE

Virtual Reality (VR) is a computer simulation of a real or imaginary system that enables a user to perform operations on the simulated module and shows the effects in real time. Real-time interactive applications, which respond immediately to user input, allow the user to fly and walk through various scenes and inspect different objects of interest in the reconstructed scene. Augmented Reality (AR) is understood as an implementation of the virtual reality into the real world. Therefore it is also known as "mixed reality". This method is enhanced with additional 3D graphics superimposed on user's field of view usually through the use of a portable media device. Furthermore AR also brings about an interactive experience, but aims to supplement the real world, rather than creating an entirely artificial environment.

The implementation of the proposed content-based filtering method into a virtual reality framework would be very useful for cultural heritage actors. It allows for easy distribution and publishing of the content from everywhere and for everybody. Virtual and augmented reality tools enrich the current digital counterparts to be straightforward applicable for using under different circumstances such as remotely and virtually navigating an archaeological site (déjà-vu), retrieve similar CH objects in terms of conservation methods used, style, operational used, evaluate the effect of different capturing technologies, promote mixing of tangible with the intangible content (humans' behaviours on items functional use), etc.

The visualization of the 3D objects plus the time is an essential part, as it is what the user perceives. In order to appreciate the richness of these 4D objects, we need a performant multipurpose VR viewer. The VR viewer is an integration of two levels of sub-viewers corresponding to the dimensions: the 3D sub-viewer and the 4D sub-viewer. The 3D VR sub-viewer ensures a true 3D visualization of the static objects. For the 4D viewer, the 4th dimension is the time and its visualization is typically based on animation. The problem is how to balance the time used to render the scene and how much is stored in memory in form of pre-computed animations.

The implementation into the real world is based on various initiation processes. The first one is image recognition which allows the user to project virtual objects into the real environment based on the position of the image target. As the name implies, image targets are images that can be detected and tracked. Unlike traditional markers, data matrix codes and QR codes, image targets do not need special black and white regions or codes to be recognized. There are sophisticated algorithms to detect and track the features that are naturally found in the image itself. The system recognizes the image target by comparing these natural features against a known target resource database. Once the image target is detected, it tracks the image as long as it is at least partially in the camera's field of view. Feasible targets can be photos, game boards, magazine pages, book covers, product packaging, posters, greeting cards as well as architectural environments. In our case, photos were used for image tracking.

Once an image is tracked, there is also a possibility of extended tracking, where the target remains visible, even if the tracker is out of sight. This provides more stability for the augmented object, keeping it in sight even if only partial tracker recognition is given.

The second tracking method is the positioning of the objects based on GPS information which currently has an accuracy of 10 meters. This is a more inaccurate solution and limits the viewing of the object to a predefined site.

For a more sophisticated handling of the objects augmented reality can be combined with 3D real time game engines, resulting in an interactive augmented reality system. This way, human senses are more involved and the immersion into the presented virtual content is consequently much greater.

We have implemented all the techniques under the framework of a European Union project and is served as an interface to the public, and includes some of the above mentioned features. Virtual content can be presented in above described approaches in different ways. One is referencing the objects on geographical position presented on referenced maps of different types (historical, hydrological, geological, etc.). The other is enhancement of the already existing media.

The objects can be viewed, interacted or further information about them can be accessed. The information can consist of related multi-medial content like objects, images, texts, links, videos and audio. A live demonstration is available on the youtube and on the authors' web site.

5 EXPERIMENTAL RESULTS

The research aims to analyse, design, develop and validate an innovative system for the rapid and cost-effective 4D (time varying 3 dimensional space modelling) model reconstruction from images selected from the wild (being captured and stored in image repositories for non-professional 3D reconstruction use) and support the aim of digital libraries Europeana and UNESCO Memory of the World to build a sense of a shard European cultural history and identity. A 4D dimensional research is also presented in (Kyriakaki, 2014).

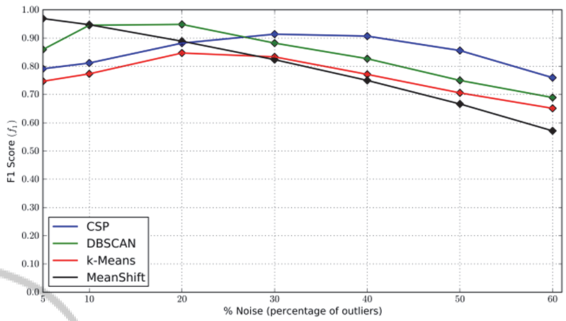

Figure 1: *F1* Score versus noise outliers using different density-based clustering methods.

5.1 Evaluation in Detecting Image Outliers

Using expert's assessment, we have initially annotated a large collection of 31,000 images into two categories; (i) the one of "relevant image set" and (ii) the one of image outliers. Fig 1 presents the *F1* Score for two density-based image partitioning approaches, the conventional DBSCAN and the CSP method, regarding outliers' removal. For the evaluation, we range the noise, the percentage of image outliers, in the created datasets from 5% to 60%. We observe that for a small number of image outliers (less than 30%) the conventional DBSCAN algorithm yields better performance. As the noise increases, however, a common case for our web-based loosely structured visual content, the performance of the new CSP density-based algorithm outperforms the conventional DBSCAN approach. This is due to the fact that the CSP partitioning approach is more prone to false negatives, while DBSCAN is more prone to false positives.

In the same figure, we have also compared the results with two other methods, the K-Means and MeanShift approach. For both cases and for high values of noise, the adopted CSP method yields better performance.

5.2 Evaluation in Image Ranking with Respect to 3D Reconstruction

Initially, an annotation set is created using experts' assessment. Let us denote as C_n a set that contains the n most appropriate images for 3D reconstruction, i.e., images that correspond to different geometric views of the object. In the sequel, the spectral based ranked images are extracted by setting graph partitions to be equal to *n/5*, *2n/5*, *3n/5*, *4n/5* and *n*, resulting in a maximum reconstruction accuracy of 20%, 40%, 60%, 80% and 100% respectively. Fig. 2 represents the results regarding reconstruction accuracy with respect to the number of selected representatives. In the same figure, we have compared the results two other approaches; the k-means and the min cut graph partitioning method. As is observed, our method outperforms both the two compared ones, in the sense that it extracts images contributing more to the 3D reconstruction process.

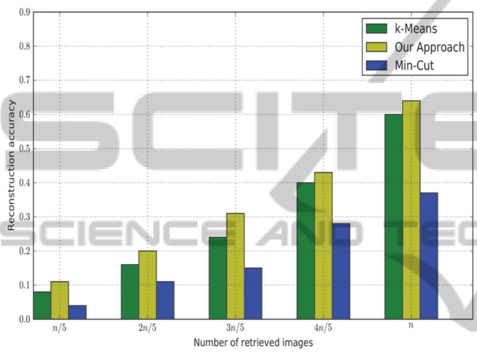

Figure 2: Reconstruction accuracy in regard to the number of selected representatives and comparisons with other methods.

Fig. 3 depicts an example of a 3D reconstruction from the Archangelos Michael Church. The reconstruction has been obtained using images. In this figure, we illustrate from left to right the point clouds, 3D meshes and the textured 3D mesh.

5.3 Virtual Reality Interface

The Virtual Reality (VR) viewer is an attempt to create an end-user software which manages the presentation of the various data-sets (imagery, 3D scans, 3D models, etc.). In the case study the 3D model of Archangelos Michael church with different information levels is presented.

The viewer allows the user to access different information through various functions. It is designed with a user friendly interface for an intuitive working experience. They can be viewed from all sides, as well as from out- and inside. With the slice-tool the model can be cut and inspected deeper.

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. Sliding over the

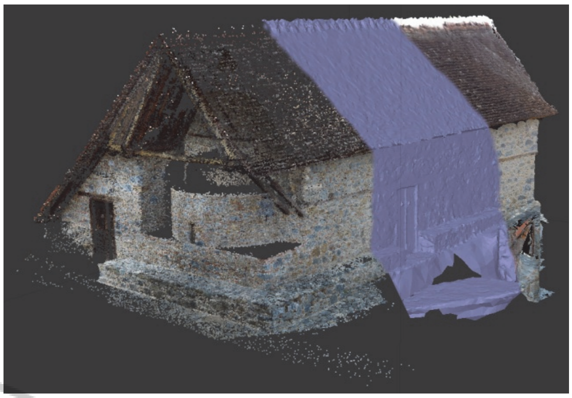

Figure 3: 3D reconstruction using the images obtained from ranking algorithm. The object of interest is the Archangelos Michael church; left to right: point cloud, 3D mesh, textured 3D mesh.

images the user can navigate in time and across different parts of the model (Fig. 4).

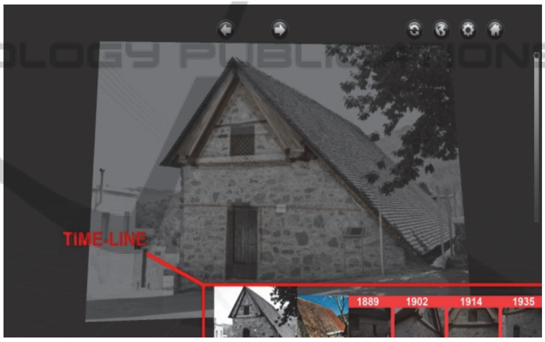

Figure 4: The Time-line feature in the VR Viewer.

The system also includes an information box in which meta-data and in case of additional reconstructions also para-data of the site together with its context (period, special morphology, references, additional images) are presented as well as links to architectural, historical, artistic, social, religious or political interest information (Fig. 5).

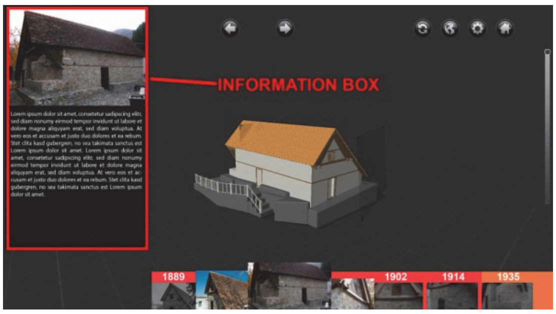

Figure 5: The Information Box feature in the VR viewer.

The slice tool enables the user to view

intersection of the reconstructed model with the help of a slicing plane (Fig. 6). This provides an insight to the interior construction of the model. With this tool also the inner image overlays are accessed.

Figure 6: The Slice Tool in the VR Viewer.

In terms of augmented reality models can be for example placed on a picture. In our test we made a perspective related integration of the model so the user can view the 3D model and therefore also the non-visible parts from the 2D picture (Fig. 15).

The augmented content can be shown on the original site or on recognized images in various medias to enrich existing print- (books, articles, maps, posters) and online-media.

5.4 Augmented Reality Interface

In terms of augmented reality models can be for example placed on a picture. In our test we made a perspective related integration of the model so the user can view the 3D model and therefore also the non-visible parts from the 2D picture (Figure 7).

The augmented content can be shown on the original site or on recognized images in various medias to enrich existing print- (books, articles, maps, posters) and online-media. Another example of augmented reality system is shown in Fig. 8. A live demonstration of the AR viewer and its functionalities is available on the youtube and on the authors' web site.

Figure 7: Augmented Reality – alignment of the model over a physical image.

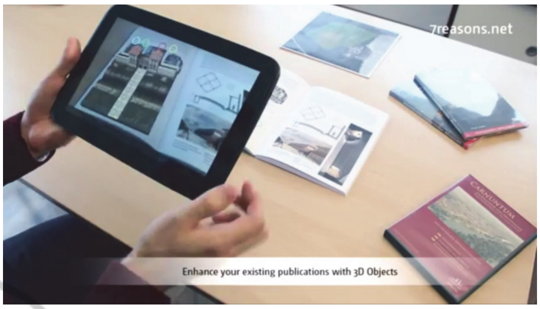

Figure 8: Another example of Augmented Reality.

6 CONCLUSIONS

In this paper, a new timely evolved 3D reconstruction process is proposed using as inputs photos derived from web-based loosely structured visual repositories, like flickr and picasa. The method allows a 4D representation of outdoor cultural sites (3D geometry plus the time). To fulfil our objectives, initially we apply a content-based filtering algorithm able to remove outliers that may confuse the reconstruction process on the use of density-based clustering. Then, we proceed with the extraction of a set of representative images that can described as much as possible the different geometric views of an object. This is accomplished using spectral graph partitioning. Finally, a virtual/augmented reality interface is proposed to assist cultural heritage actors to evaluate defects, restoration processes and other phenomena on cultural objects of interest.

Experimental results on real-life web-based visual data indicate the outperformance of the proposed content-based filtering method than other conventional schemes. The evaluation has been carried out using objective criteria as of *F1* score and reconstruction accuracy. In addition, the virtual/augmented reality viewer is demonstrated against actual cultural sites while its features are also available on youtube network.

ACKNOWLEDGEMENT

The research leading to these results has been supported by European Union funds and National funds (GSRT) from Greece and EU under the project JASON: Joint synergistic and integrated use of "eArth obServation, navigatiOn and commuNication technologies for enhanced border security funded under the cooperation framework.

REFERENCES

Arampatzis, A., Zagoris, K., and Chatzichristofis, S. A., 2013. Dynamic two-stage image retrieval from large multimedia databases. *In. Process. Manag., vol. 49, no. 1, pp. 274–285.*

Bach, F. R., and Jordan, M. I., 2003. Learning Spectral Clustering. *Computer Science Division, University of California at Berkeley, Berkeley*, California.

Cayton, L, 2006. Algorithms for manifold learning. University of California, *San Diego, Tech. Rep. CS2008-0923.*

Chum, O., Philbin, J., Sivic, J., Isard, M., and Zisserman, A., 2007. Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval. *In IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, pp. 1–8.*

Cox, T., Cox, M., 2000. Multidimensional Scaling. *Chapman & Hall/CRC, Second Edition.*

Daras, P., Axenopoulos, A., Litos, G., 2012. Investigating the Effects of Multiple Factors towards more Accurate 3D Object Retrieval. *IEEE Transactions on Multimedia, Vol. 14, No. 2, Page(s): 374 – 388.*

Doulamis, A., Ioannides, M., Doulamis, N., Hadjiprocopis, A., Fritsch, D., Balet, O., Julien, M., Protopapadakis, E., Makantasis, K., Weinlinger, G., S.Johnsons, P., Klein, M., Fellner, D., Stork, A., Santos, P. 2013. 4D reconstruction of the past. *Proceedings of SPIE - The International Society for Optical Engineering,* 8795, art. no. 87950.

Fan, K., 1951. Maximum Properties and Inequalities for the Eigenvalues of Completely Continuous Operators. *Proc. Natl. Acad. Sci. U. S. A.*, vol. 37, no. 11, pp. 760–766.

Hadjiprocopis, A., Ioannides, M., Wenzel, K., Rothermel, M., Johnsons, P.S., Fritsch, D., Doulamis, A., Protopapadakis, E., Kyriakaki, G., Makantasis, K., Weinlinger, G., Klein, M., Fellner, D., Stork, A., Santos, P. 4D reconstruction of the past: The image retrieval and 3D model construction pipeline. 2014 (2014) *Proceedings of SPIE - The International Society for Optical Engineering,* 9229, art. no. 922916, .

Kekre, D. H. B., Sarode, T. K., Thepade S. D., and Vaishali V., 2010. Improved texture feature based image retrieval using Kekre's fast codebook generation algorithm. *In Int. Conference on Contours of Computing Technology, Thinkquest~2010, S. J. Pise, Ed. pp. 143–149.*

Kyriakaki, G., Doulamis, A., Doulamis, N., Ioannides, M., Makantasis, K., Protopapadakis, E., Hadjiprocopis, A., Wenzel, K., Fritsch, D., Klein, M., Weinlinger, G., 2014. 4D Reconstruction of Tangible Cultural Heritage Objects from Web-Retrieved Images. *International Journal of Cultural Heritage in Digital Era.*

Lv, Q., Josephson, W., Wang, Z., Charikar, M., and Li., K., 2007. Multi-probe LSH: Efficient Indexing for High-dimensional Similarity Search. *In Proceedings of the 33rd International Confe-rence on Very Large Data Bases, Vienna, Austria, pp. 950–961.*

Makantasis, K., Doulamis, A., Doulamis, N., Ioannides, M., 2014. In the wild image retrieval and clustering for 3D cultural heritage landmarks reconstruction. *Multimedia tools and Applications, Springer Press.*

Min, R, and Cheng, H. D., 2009. Effective image retrieval using dominant color descriptor and fuzzy support vector machine. *Pattern Recognit., vol. 42, no. 1, pp. 147–157.*

Murthy, V. S., Kumar, S., and Rao P. S., 2010. Content Based Image Retrieval using Hierarchical and K-Means Clustering Techniques. *Int. J. Eng. Sci. Technol., vol. 2.*

Papadopoulos, S. Zigkolis, K., Kompatsiaris Y.,, and Vakali, A., 2010. Cluster-based Landmark and Event Detection on Tagged Photo Collections. *IEEE Multimedia.*

Protopapadakis, E., Doulamis, A., Matsatsinis, N. Semi-supervised image meta-filtering in cultural heritage applications. 2014. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* 8740, pp. 102-110.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. *In IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571.*

Smeets, D., Fabry, T., Hermans, J., Vandermeulen, D., and Suetens, P., 2009. Isometric deformation modelling for object recognition. *In Proc.CAIP'09, 2009, pp. 757–765.*