4D Reconstruction of the Past

Anastasios Doulamis[b], Marinos Ioannides[a], Nikolaos Doulamis[a], Andreas Hadjiprocopis[a], Dieter Fritsch[c], Olivier Balet[d], Martine Julien[d], Eftychios Protopapadakis[b], Kostas Makantasis[b], Guenther Weinlinger[e], Paul S.Johnsons[e], Michael Klein[e], Dieter Fellner[f], Andre Stork[f] and Pedro Santos[f]

[a]Cyprus University of Technology, Digital Heritage Lab, Cyprus;
[b]Technical University of Crete, Greece;
[c]University of Stuttgart, Germany;
[d]DigiNext, France;
[e]7Reasons GmbH, Austria;
[f]Fraunhofer Institute for Computer Graphics Research IGD, Germany

ABSTRACT

One of the main characteristics of the Internet era we are living in, is the free and online availability of a huge amount of data. This data is of varied reliability and accuracy and exists in various forms and formats. Often, it is cross-referenced and linked to other data, forming a nexus of text, images, animation and audio enabled by hypertext and, recently, by the Web3.0 standard. Search engines can search text for keywords using algorithms of varied intelligence and with limited success. Searching images is a much more complex and computationally intensive task but some initial steps have already been made in this direction, mainly in face recognition. This paper aims to describe our proposed pipeline for integrating data available on Internet repositories and social media, such as photographs, animation and text to produce 3D models of archaeological monuments as well as enriching multimedia of cultural / archaeological interest with metadata and harvesting the end products to EUROPEANA. Our main goal is to enable historians, architects, archaeologists, urban planners and affiliated professionals to reconstruct views of historical monuments from thousands of images floating around the web.

Keywords: 4D Cultural Content Representation, Visual Content of Cultural Heritage, Visual Search, Augmentation and Interaction

1. INTRODUCTION

Cultural and creative industries refer to a range of economic activities which are concerned with the generation or exploitation of knowledge and information.[1] Nearly all commentators agree that cultural and creative industry has moved closer to the centre of the economic action mainly due to the recent advantages in information and communication technology. Surveys indicate that cultural and creative industry represents 4.5% of total European GDP and for 3.8% of the workforce; so, this sector of economy can constitute one of the main European engines for growth and job creation. The increasingly importance of creative industries to economic well-being, proponents suggesting that "*human creativity is the ultimate economic resource*," and that "*the industries of the twenty-first century will depend increasingly on the generation of knowledge through creativity and innovation*".[2]

Unlocking the potential of cultural and creative industries is an important factor in maintaining european competitiveness in the changing global environment; for Europe the rapid roll-out of new technologies and increased globalisation has meant a striking shift away from traditional manufacturing towards services and

Further author information: (Send correspondence to A.D & M.I.)
A.Doulamis: E-mail: adoulam@cs.ntua.gr, Telephone: +30 28210 37430
M.Ioannides: E-mail: marinos.ioannides@cut.ac.cy, Telephone: +357 25 00 2020

innovation. Factory floors are progressively being replaced by creative communities whose raw material is their ability to imagine, create and innovate. Therefore, we need to put in place the right conditions for creativity and innovation to flourish in a new entrepreneurial culture. Europe's creative industries offer a real potential to respond to these challenges thereby contributing to the Europe 2020 strategy and some of its flagship initiatives such as the Innovation Union, the Digital Agenda, tackling climate change, the Agenda for new skills and new jobs or an industrial policy for the globalisation era. In particular, creative industries are recognised as growth sectors in the aforementioned Commission Consultation Paper as well as in the Commission Staff Working Document "Challenges for EU support to innovation services – Fostering new markets and jobs through innovation" SEC(2009)/1195. Over the last few decades, society has amassed an enormous amount of digital information about the Earth and its inhabitants. However, these archives pale in comparison to the flood of data which is engulfing us. A new wave of technological innovation is allowing us to capture, store, process and display an unprecedented amount of geo-referenced information about our planet and a wide variety of environmental and cultural phenomena. While recognising this challenge, the former United States Vice President Gore presented the visionary concept of a Digital Earth in an address at the California Science Centre in January 1998. This event initiated one of the most ambitious and challenging programme ever proposed to the US scientific community. Digital Earth would be a virtual representation of our planet that enables a person to explore and interact with the vast amounts of natural and cultural information gathered about the Earth. Fifteen years after Mr Gore's conference, Google Earth, NASA WorldWind, and Microsoft Bing are some of the most well-known and widely adopted results of the Digital Earth vision. It is today possible to explore, at home, a virtual representation of our planet, flying through landscapes draped with aerial or satellite images, contemplating mountains, valleys and other natural sites. Representations of cities, and more detailed cultural heritage landmarks, are even accessible.

Figure 1. Firenze automatically reconstructed by the V-City project

A crucial driving force for the development and economic growth of the creative industries is ICT technologies. Using innovative IT solutions in growing areas of the creative sector - such as advertising, digital media, gaming, including serious gaming, interactive design, cultural and educational services - opens up manifold competitive advantages for research, development and business. For this a lot of research have been conducted and funded in the recent years in the area of (i) digitisation, especially for tangible cultural (creative) objects, (ii) augmented reality and cultural experiences tools, (iii) information retrieval and use and (iv) virtual heritage and Web archiving.

The revolutionary expansion and developments of information and communication technologies in all areas of our life and the introduction of the Internet technology during the last 20 years is characterised by the transition from text, audio, and video to high-speed multimedia and 3D visualisation services such as 3D TV, -Cinema, -Games, -digital libraries which are interlinked to other on-line services like Google Search, Google Earth and Google Maps. Therefore, on-line digital libraries such as EUROPEANA and UNESCOs Memory of the World expanded their content to include 3D museums objects archaeological sites a d monuments. In such a context, digitisation, preservation and online availability of digitised cultural content has always been a top-level priority research.

Photogrammetry aims at reconstructing the 3D geometry of an object based solely on 2D photographs. This has been an active topic of research in the field of Surveying, Computer Science and Computer Vision for many years. Today, its research outcomes in combination with an enormous increase in the availability of computational resources at an affordable price have given rise to a number of techniques and software demonstrating the maturity of the technology. There is great potential for applications using these techniques in a wide range of fields, for example, terrain modelling, reconstruction of archaeological monuments or entire cities, the building of virtual worlds in computer games, etc.

Passive scanning methods based on aerial or satellite photography have been used for topographic modelling of natural open areas with great success for many years. Active scanning techniques, such as laser and acoustic methods, take advantage of newer technology advances, accurate geo-referencing and powerful computer machinery. They can produce very dense and accurate 3D point clouds,[3] especially useful for the automatic extraction of height/depth information. However, these techniques present major challenges when it comes to modelling cultivated, wooded or built-up areas from the air, as well as non-convex terrain features such as caves, overhang cliffs, bridges and other buildings. Incorporating texture information is also difficult. According to[3] the combination of images, laser scanning and existing GIS maps is considered to be one of the most robust and flexible approaches to automatically creating low resolution, photo-textured models, for example in.[4,5] However, the above methods rely on data captured from structured sources such as aerial photographs from survey aircrafts or street-level photographs from moving vehicles. The advantage is obvious; known / regulated sampling rate and geo-location.

In contrast, Structure from Motion (SfM) algorithms[6] use data from unstructured sources, e.g. photographs taken by off-the-shelf cameras at different times, by different people, from different positions, in order to recover the structure of a scene as a collection of 3D points as well as the camera position. The problem becomes more complicated when the source images or videos are harvested from the web as no assumption can be made about camera characteristics, illumination, resolution, geo-location, etc. However, this extra complication is well worth it given the wealth of information available online and for free over the web. Systems for urban reconstruction from video sources have been proposed.[7,8] The latter achieves 3D reconstruction in real time. However, the main advantage of using video as opposed to unordered images is the camera location relationship between successive video frames. Much more complex is the reconstruction from unordered and often irrelevant images. Such a system requires computationally intensive pre-processing and filtering of millions of images using algorithms for landmark identification, for example.[9,10] These algorithms mainly operate on a pairwise basis which although time-consuming, it can be massively parallelised, for example.[11] The first system to apply SfM algorithms to internet photo collections was.[12] This system was limited to processing only a few thousands of images. A more successful system[11] can perform dense modelling from Internet photo collections consisting of millions of images. The speedups obtained are due to some refinements e.g. filtering by early 2D appearance-based constraints, but mainly because of utilising graphics processors and parallelising the computation on multi-core computer architectures.

In recent years, the use of approaches based on 3D geometry has seen rapid progress in many different areas from digital factories of the future to car, flight- and surgical training simulators to 3D maps, 3D TV/Cinema and games but also in cultural heritage applications. 3D geometric modelling is another powerful tool to better understand, share and communicate cultural heritage: It allows compression of the huge amount of data, checking and validating the consistency of the separate 1D or 2D data interpretations as well as an effective archiving of the complete object. Furthermore, a consistent 3D geometric model is essential for the re-use and post-process computations that need an accurate and coherent geometry of cultural heritage objects (for example in the area of restoration). In general, if we want to interconnect the 3D data (from museums' objects, artifacts, archaeological sites or monuments) with other related 3D objects or with multimedia data, this poses different challenges but also provides new, exciting and innovative possibilities compared to more established data forms like texts, images or sound. Subsequently, the digitisation and 3D content creation process in the field of cultural heritage is confronted also with another challenge. Each digitised object (geometry) needs logical (semantic) information. Without this logical information, scanned 3D objects like monuments are nothing more than a structure of geometrical primitives (points or triangles). The importance of 4D urban environment modelling can be, further, emphasised by examining its two main components: (i) 3D content and (ii) time-varying content.(4D World: X+Y+Z+Time=4D); 3D content is of primary importance for our lives as we are living in 3D world and we perceive most of the events occurring in this world by depth information. The importance of 3D content can be highlighted by examining the work of J. Fodor about the modularity of mind.[13] In this work it is stated that our brain follows a hierarchical way of thinking starting from the simplest thought and ending at the most advanced one. Concerning on objects, or, in general, world, description the simplest "concept" coming to our mind is their 3D structure. Although, humans can perceive and understand material objects by their spatial dimensions, to perceive and understand the higher-level meaning of events, the dimension of time is a prerequisite.

In this research, we address the aforementioned difficulties by proposing a complete approach able to search, process, model, 3D reconstruct, define metadata and build new production services and re-use of the cultural heritage content. In this respect, the rapidly evolving fields of computer graphics, vision and learning, 3D modelling and virtual reality, seems to hold a strong potential for proving effective solutions towards accurate and veridical spatial-temporal urban environment modelling. However, it is imperative to go beyond cataloguing, indexing, and keyword driven databases, to a paradigm where the computer at least partially understands the content of images. Going beyond the state of the art in scene understanding and 3D modelling will enable fundamental new methodologies to view and experience historical and/or temporally varying imagery. This way, we construct building time-varying 3D models that contain how the appearance, historical events around a place and structure have been evolved.

2. PREVIOUS WORK

A lot of research works and tools have been proposed in the literature for cultural heritage processing. However, a tool like dealing with 4D dimensional processing (3D objects plus the time) is a novel and unique concept. For 3D scene reconstruction and scene flow estimation our approach is based on computer vision and Photogrammetry techniques that require optical flow, as well as, disparity estimation. Both of these can be formulated as a challenge of finding corresponding points in different images. However, the correlation /matching engine is a very challenging task, because a scene's objects, generally, have different shape and appearance when seen from different points of view and at different times.

The techniques used to solve correlations are similar and can be categorised as energy-based, feature-based and area-based methods. Energy-based methods[14,15] minimise a cost function plus a regularisation term, in framework of Horn and Schunck,[16] to solve for the 2D displacements. Feature-based methods match features in different images. Selected features should have good discriminative properties and there should be high probability that the same point is selected in different images.[17] This kind of methods are using the concept of coarse-to-fine image warping, however, this down-sampling removes information that may be vital for establishing the correct matches. Characteristics works of feature-based methods are.[18,19] The authors of[18] propose an algorithm, based on HoG descriptor and the notion of "*data-driven uniqueness*" by which each image decides what is the best way to weight its constituent parts, is capable to handle cross-domain image matching. In[19] a SIFT flow algorithm is presented. This algorithm consists of matching densely sampled SIFT features between different images, while preserving spatial discontinuities. Considering that reliability of a detector to find the same physical interest points under different viewing conditions is its most valuable property, the authors of[19] present SURF. In[20] an approach that overcomes common assumption of feature-based methods limitations is presented. This approach computes a global image-based matching score between different images, making the matching process capable to handle projective distortions and partial occlusions. In[21] the Hough forests are introduced that provide an alternative way for the combination of machine learning and Hough transform. Hough forests can be regarded as task-adapted codebooks of local appearance that allow fast training and matching and extend Hough transform to new domains as object tracking and action recognition. In[22] curve correspondences, crease-edge silhouettes or class-specific internal texture edges, are used to match curves between different views of objects that belong to the same class, like dolphins or oak leaves. The method of[23] detects and matches repeated patterns to automatically geo-tagging photographs of man-made environments, by using affine-invariant descriptors[24] for 3D reconstruction and image retrieval.

3D model representation takes place by using computational geometry techniques. In the work of[25] feature-based and area-based methods are combined. The authors present an algorithm based on features found by Harris and Difference-of-Gaussians operators. The authors of[26] exploit Reeb graphs capability to preserve the topological properties of an object's shape. Finally, in[27] a procedural 3D building reconstruction method using shape grammars and detectors is presented.

3. THE SCALABLE 4D ACQUISITION FRAMEWORK

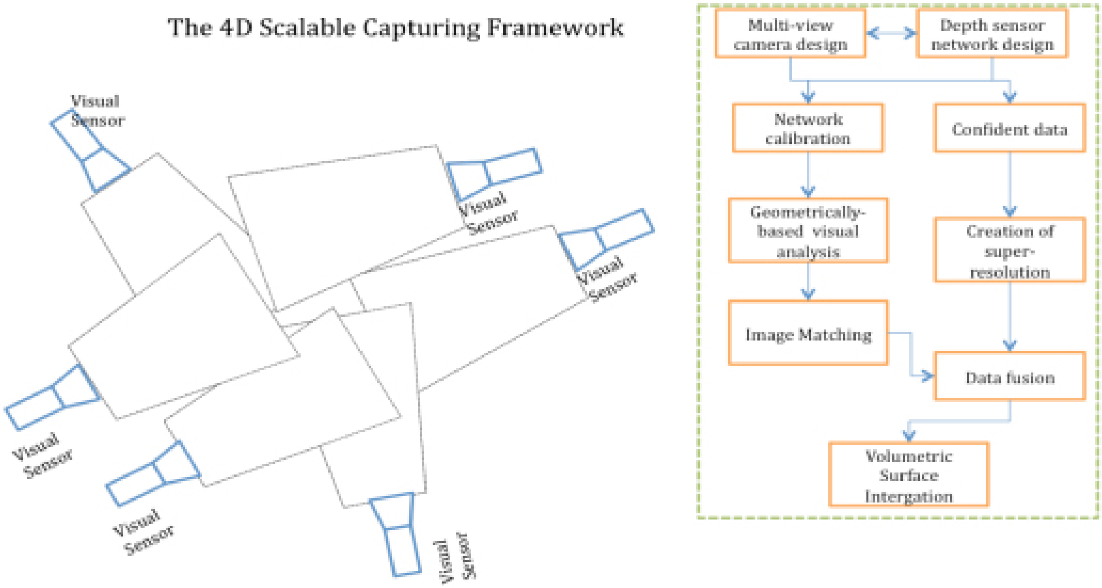

Figure 2. Architectural components of the proposed 4D scalable image capturing.

This section deals with technologies able to capture and virtually 3D/4D reconstruct human acting, such as performing arts under an affordable, cost-effective and accurate digitisation framework. To do this, we introduce an innovative technological framework able to combine advanced technologies in the area of Photogrammetry and Computer Vision. In this paper, a scalable capturing framework is adopted that combines 3D modelling technology coming by a set of multi-view stereo camera architecture and a low cost depth sensors (KinectTM). The balance of using multi-view stereo imaging and depth sensors are obtained in terms of accuracy and cost-effectiveness. We examine the drawbacks and the advantages of each technology. We also examine possible ways of fusing the information from different sensors in order to increase the accuracy, or reduce the number of cameras necessary for data capturing. In other words, the proposed architecture balances between the information provided by high resolution cameras plus dense image matching techniques and with low cost depth sensors able to generate depth information in real time.

As regards the multi-view imaging architecture, the 3D/4D information is extracted by applying dense image matching techniques. For each stereo model observing common content a correspondence is determined for each pixel individually. By using these correspondences between all stereo models, all 3D/4D Points can be triangulated based on their viewing rays at once. This leads to a very dense and accurate point cloud. In order to acquire time, this step is performed for each frame in time for all synchronised cameras – leading to 3D point clouds for each time stamp. However, this is a computational intensive process that increases the total cost of digitisation. To do this, we have included 3D/4D modelling tools appropriately designed for time varying shapes. Such methodologies exploit motion information as well as tracking methods. Subsequently, a volumetric integration of this depth information not only enables the extraction of a volumetric representation, but also to fill small gaps and reduce noise.

As regards the use of low cost depth sensors, the paper investigates methodologies for increasing sensor resolution through the use of super-resolution methods. Super resolution is accomplished by exploiting information regarding depth maps acquired by a set of overlapping KinectTM sensors of different orientations. In addition, we investigate ways of defining the level of confident of the data. This is performed in close collaboration with the results of the multi-view camera architecture. Thus, information from the one system is propagated to the other for improving the 3D reconstruction results and reducing the cost of capturing.

4. VISUAL SEARCH ENGINE

In Figure 3, we present a schematic diagram of our search engine used for visual content retrieval appropriate for cultural heritage objects. The engine is applied on the web and thus it searches objects in the "wild" visual content. Then, the results are improved by exploiting metadata from EUROPEANA and/or Memory of the World.

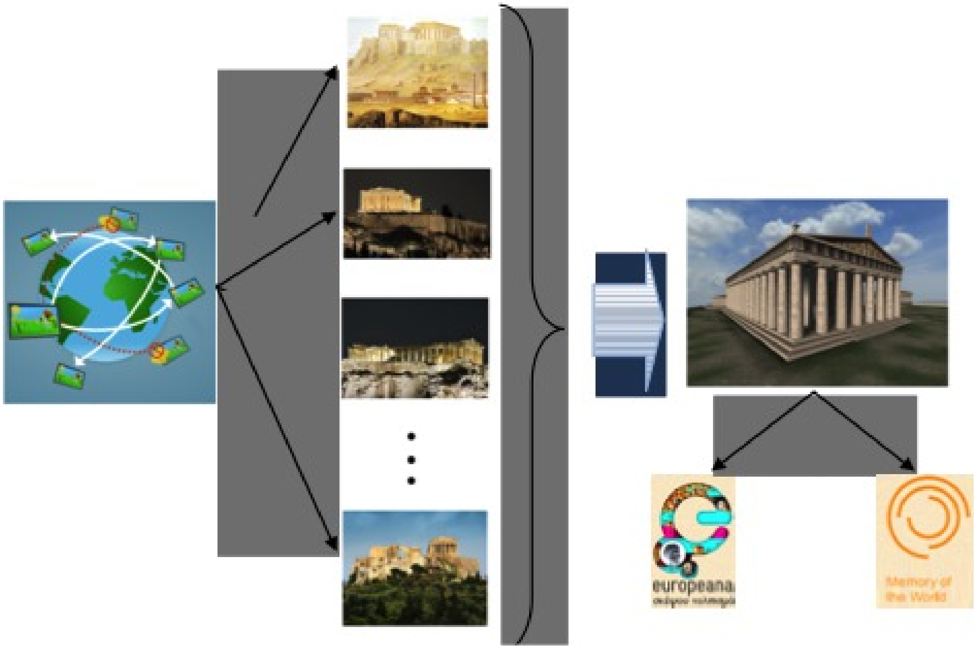

Figure 3. The search engine approach adopted in this paper

In this section, we propose visual computing methodologies with the purpose of encoding the 4D computational geometry of the detected moving objects, retrieve 3D/4D data based on a content oriented and abstract framework and finally fit and retrieved deformable models into the animation properties of the real captured data. Towards this direction, initially, we apply segmentation algorithms able to isolate the foreground objects from the background. This step is critical since it allows the human objects to be separated with the background content. The foreground objects are dynamically updated, processed and enriched by virtual retrieved surrogates. On the contrary, background content is updated by the information provided by the depth sensor network. A set of deformable models is used to describe the human movement. In order to allow a cost effective authoring interface as regards human acting, initially, a voxel 3D skeletonisation algorithm is performed. In this way, we are able to reduce the amount of information needed to describe the human object; therefore we reduce the cost of enrichment the captures 3D models with additional 3D/4D data and also the cost of manipulated the created physical or virtual objects with synthetic animation variables to simulate human like emotional behaviours. 3D skeletons are medial axes transform and encode mostly motion information. 3D skeletons separate motion estimation from shape and thus, they allow a more accurate 3D model updating through time.

Working with large collections of 3D models requires fast content-based retrieval techniques, especially since public collections are often insufficiently annotated. In that case a keyword based search alone is not promising. While research on example-based retrieval where users provide a full model as the query has recently found a lot of interest in the community, its practical application is difficult, since a good example is often not at hand. For this research, a more abstract form of content-based retrieval algorithms have been proposed in the context of this research, such as sketch based shape retrieval, which seems to be much more cost-effective, fast and reduce the respective resources in the generation of efficient creative processes.

The derived 3D skeletons are tracked in time using motion capture and tracking computer vision methodologies. Tracking process is assist through the selection of appropriate geometrically enriched data. By taking into consideration motion tracking, which is performed on the 3D skeletons to increase accuracy and computational efficiency, as well as the set of deformable models appropriate extracted by the content-based retrieval algorithms, we are able to automatically update and fit either the detected captured models or the retrieved deformable 3D/4D data to several different emotional & behavioural animation states, resulting in cost-effective and simultaneously efficient creative processes.

Despite the efficiency of the aforementioned methodology, possible errors (in the 3D skeletonisation and tracking process) generate erroneous virtual 3D reconstructions. To address this difficulty, we enhance the results of the 3D modelling of moving objects using information derived from the depth sensor network. In this case, we are able to exploit the depth information to improve 3D modelling accuracy.

5. VISUALISATION AND AUGMENTED REALITY

The main purpose of this research is to provide the technological framework for enriching the content with additional overlays and virtual objects using smart and cost effective augmented reality methods. The research exploits the outcome of the content based retrieval module (see section 4). This way, the content creators are able to retrieve content of interest in a cost effective manner (through for example the usage of a Sketch or drawing). In the following, the authoring environment tool has the algorithms to transform and scale the retrieved 3D or 4D content in a way to be synchronised with the current scene.

This paper researches on efficient techniques so that the changes in the viewer's position can be properly reflected in the rendered graphics. Tracking method are usually based on a two-stage process; a learning or feature extraction stage, and a tracking stage. During the learning stage, some key features relatively invariant to the environment are extracted such that it is mostly a bottom-up process. It can enable reinitialisation of the tracking system to recover when it fails by using automatic recognition. Approaches based on particle filters and Kalman models fail to provide accurate results and on the other to conclude to a computational and cost effective solution. For this reason, in the context of this paper, we incorporate semi-supervised learning algorithms in the area of tracking; Particularly, we allow a minimum user interaction (obtained by the intuitive interactive interface) and then, we implement a classification based tracking strategy to achieve an efficient object tracking and synchronisation.

Personalised Viewing: Since we support a collaborative, interactive interface where multi-users of different properties (e.g., different professions) are incorporated in the authoring process, we need to develop tools for personalised viewing; this means that a part of a scene is presented to the end users according to their particularly interest. In this way, we increase the level of interaction of the content creators, since the system automatically focuses on parts of the scene that are interested by the user.

Perceptual and Real-time Rendering: Rendering process is very expensive and not cost effective. Perpetual rendering transforms rendering into a cost-effective process, since rendering is implemented in a selective way according to the current content creators interest. Perceptual is required for achieving a cost effective perceptually-based selective rendering algorithms are adopted in order to optimise the 3D world and achieve fast real-time interaction. In order to economise on rendering computation, selective rendering guides high level of detail to specific regions of a synthetic scene and lower quality to the remaining scene, without compromising information transmitted. Such decisions go beyond current task-based approaches and will be guided by context based information.

6. THE VISUAL AUTHORING INTERFACE

This section describes the visual interface that will enable non-programmers to make use of the technological components developed of our research for authoring purposes. For that aim, we propose to implement an advanced user interface to author interactive characters using a visual programming approach that consists in providing graphical bricks that the user can assemble, control or parameterise without taking care of the software implementation details.

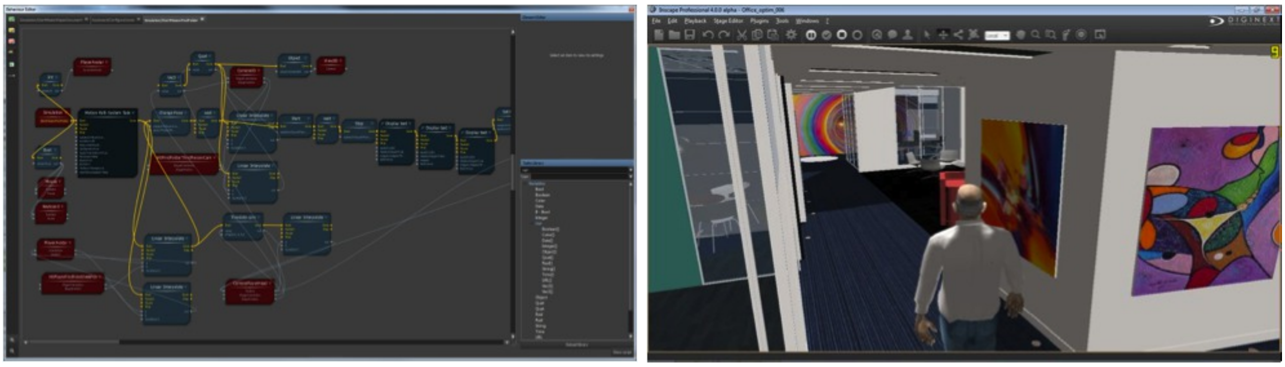

Figure 4. Visual Authoring Interactive Content

The presented research approach extends the state-of-the art visual authoring concept based on the Vertigo rendering engine developed for the industrial sector, and can only produce content for that engine. This represents a serious limitation especially when considering the “universal” aspect of our research that should not limit authors to create interactive characters for one single proprietary engine. For that purpose, we abstract our previous rendering engine which does not only represent serious technical challenges but also ambitious research objectives as this would mean decoupling the visual programming interface from the features/capacities/constraints of the targeted engine and implementing a common model for character simulation. For that purpose, we consider four of the most commonly used engines (Unity, Ogre, OpenSceneGraph, and CryEngine), we then abstract data structures related to the character rendering and simulation features they have in common. We also implement an extension mechanism that enable the support of specific features proper to some engines.

Then, we research on new visual programming paradigm and related visual interfaces required to control the components delivered by other research components. Users will be able to link in the same visual area a set of different behaviour objects inspired from data flow programming, event-based programming, finite-state machines, timeliness and imperative programming. The objective is to propose, for each level of the creation pipeline, consistent and coherent interfaces and modus operandi without needing to learn how to use a dedicated tool.

Finally, this task aims to support the collaborative authoring of digital characters involving multiple authors in its different phases. This is done by implementing the required functionalities to interoperate with the collaborative authoring server. This also requires implementing the publishing capabilities that will enable the export of the produced character, including their behaviour, on the server, in a form suitable for the targeted rendering/animation engine using, as often as possible, existing standards (e.g. collada, 3DS, LUA scripts for the animations).

7. INTEROPERABLE METADATA DESCRIPTION FORMAT OF CREATIVE HUMAN ACTIVITIES

Systems specifications are surveyed in such a way to derive an interoperable description framework based on which we are able to design and define the Creative Human Activities Metadata Interface (we call it CHAMI in the rest of this paper). CHAMI aims at specifying a set of metadata that are necessary for representing the rich information of creative human actions. Another research aspect in the introduction of an interoperable metadata format able to specify the structure and the relations between the extracted metadata. Thus, CHAMI specifies not only the appropriate metadata for representing human creativity but also the structure and semantics of relations between its components. Figure 5 presents an indicative description framework of the basic metadata elements used. As is observed, the metadata are discriminated into four main categories; the low level feature metadata, the contextual and environmental metadata, the socio-cultural factors as well as the emotional attributes.

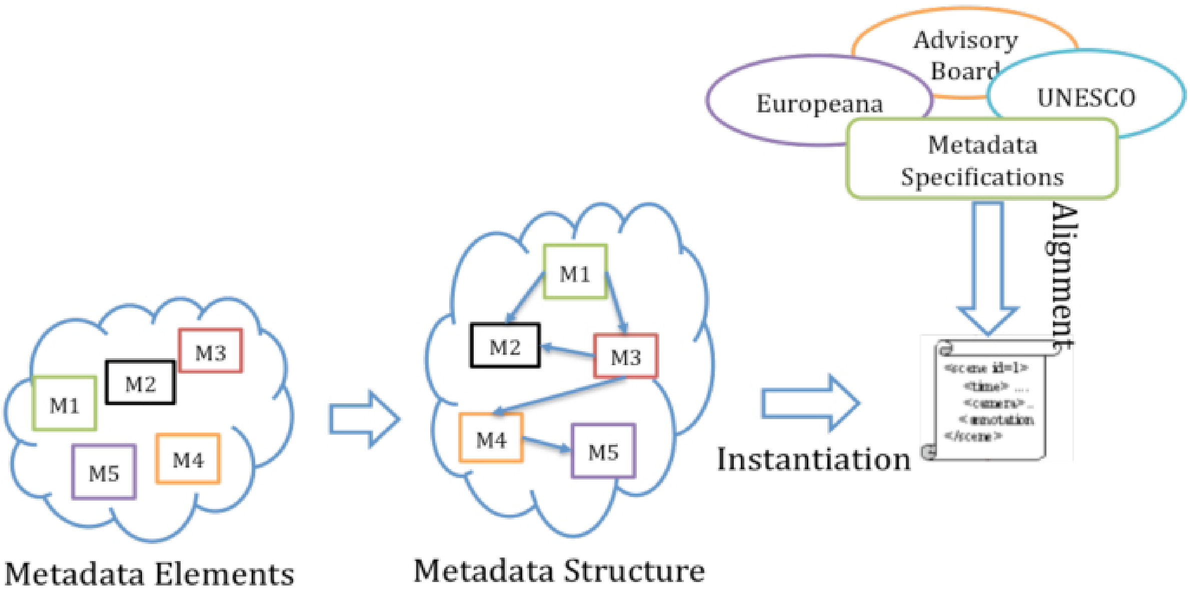

Figure 5. Description of the metadata format and structure used in the framework of this paper

We adopt a similar research framework that exists in the MPEG-7 standard for defining the associations between several metadata elements. In this way, we are able to introduce more complex description scheme that models the complexity of human creativity especially in cases where several people acting and interacting together. In this way, we are able to build the storytelling, for example for a performing art and associate it with contextual aspect according to the requirements of the design.

8. CONCLUSION

In this paper we describe a new approach which enables us to reconstruct cultural heritage content on 4 dimensions, i.e., the 3D spatial reconstruction plus the time. The paper adopts new schemes for 4D visual content reconstruction under a scalable way, visual search at the wild and tools for visualisation and augmented reality. We also describe a visual authoring interface and algorithms for interoperable description of the content and the exploitation of proper metadata. The latter is affiliated with the EUROPEANA framework.

Repurposing of data generated through academic research is essential, not only for the commercial benefit of organisations producing digital media, but also for the promotion of interest in cultural heritage among the wider population. Education and tourism are intrinsically linked fields, through engagement with visitors and tourists the enhanced experiences possible through the deployment of digital models and 4D presentations are also capable of educating those people in a non-intrusive way. The creation of digital models for this purpose is not however, unproblematic. The interpretative process involved in generating reconstruction is a subjective one, documentation of this process and the creation of paradata in accordance with the London Charter and Seville Principles on the creation and use of digital cultural-historical and archaeological data is of paramount importance, even if the story presented to the public is a simple and straightforward one. The layering of these forms of information into the "product" accessible to different audiences in a dynamic form, is something which the use of digital media permits easily.

While the creation of the models themselves is often a time-consuming process, the deployment of the data for a range of presentational objectives can be accomplished relatively easily. This flexibility allows for the use of modelled objects in digital presentations, for museum exhibits, brochures, static-media displays, guidebooks etc. The reproducibility of these images represents an investment in the future-proofing of educational or tourist products. Although static-media will almost certainly remain a part of the output from such projects, the essential dynamism of the underlying models, and the recording of paradata relating to their creation, enables them to be easily updated and enhanced as new information or revised interpretations become available.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7-PEOPLE 2007-2013 under REA grant agreement IAPP2012 n^o 324523.

REFERENCES

[1] Hesmondhalgh, D., [*The Cultural Industries*], SAGE Publications (2007).

[2] Florida, R., [*The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*], Basic Books (2002).

[3] Zlatanova, S., "Working group ii acquisition position paper: Data collection and 3d reconstruction," in [*Advances in 3D Geoinformation Systems*], Oosterom, P., Zlatanova, S., Penninga, F., and Fendel, E., eds., *Lecture Notes in Geoinformation and Cartography*, 425–428, Springer Berlin Heidelberg (2008).

[4] Pu, S., Vosselman, G., and Vi, C., "Automatic extraction of building features from terrestrial laser scanning," in [*International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*], 33–39 (2006).

[5] Schwalbe, E., "3d building model generation from airborne laser scanner data using 2d gis data and orthogonal point cloud projections," in [*Proceedings of ISPRS WG III/3, III/4, V/3 Workshop on Laser Scanning*], 12–14 (2005).

[6] Hartley, R. I. and Zisserman, A., [*Multiple View Geometry in Computer Vision*], Cambridge University Press, second ed. (2004).

[7] Cornelis, N., Cornelis, K., and Van Gool, L., "Fast compact city modeling for navigation pre-visualization," in [*Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*], *CVPR '06*, 1339–1344, IEEE Computer Society, Washington, DC, USA (2006).

[8] Pollefeys, M., Nistér, D., Frahm, J. M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S. J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G., and Towles, H., "Detailed real-time urban 3d reconstruction from video," *Int. J. Comput. Vision* **78**, 143–167 (July 2008).

[9] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**, 91–110 (2004).

[10] Li, X., Wu, C., Zach, C., Lazebnik, S., and Frahm, J.-M., "Modeling and recognition of landmark image collections using iconic scene graphs," in [*Computer Vision, ECCV 2008*], Forsyth, D., Torr, P., and Zisserman, A., eds., *Lecture Notes in Computer Science* **5302**, 427–440, Springer Berlin Heidelberg (2008).

[11] Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., and Szeliski, R., "Building rome in a day," *Communications of the ACM* **54**(10), 105–112 (2011).

[12] Snavely, N., Seitz, S. M., and Szeliski, R., "Photo tourism: exploring photo collections in 3d," *ACM Trans. Graph.* **25**, 835–846 (July 2006).

[13] Fodor, J., [*The Modularity of Mind: An Essay on Faculty Psycholgy*], Bradford Bks, A BRADFORD BOOK (1983).

[14] Alvarez, L., "Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-Space Based Approach," *Journal of Visual Communication and Image Representation* **13**, 3–21 (Mar. 2002).

[15] Papenberg, N., Bruhn, A., Brox, T., Didas, S., and Weickert, J., "Highly accurate optic flow computation with theoretically justified warping," *INTERNATIONAL JOURNAL OF COMPUTER VISION* **67**, 141–158 (2006).

[16] Horn, B. K. P. and Schunck, B. G., "Determining optical flow," *ARTIFICAL INTELLIGENCE* **17**, 185–203 (1981).

[17] Tuytelaars, T. and Mikolajczyk, K., "Local invariant feature detectors: a survey," *Found. Trends. Comput. Graph. Vis.* **3**, 177–280 (July 2008).

[18] Shrivastava, A., Malisiewicz, T., Gupta, A., and Efros, A. A., "Data-driven visual similarity for cross-domain image matching," *ACM Trans. Graph.* **30**, 154:1–154:10 (Dec. 2011).

[19] Liu, C., Yuen, J., Torralba, A., Sivic, J., and Freeman, W. T., "Sift flow: Dense correspondence across different scenes," in [*Proceedings of the 10th European Conference on Computer Vision: Part III*], *ECCV '08*, 28–42, Springer-Verlag, Berlin, Heidelberg (2008).

[20] Pons, J.-P., Keriven, R., and Faugeras, O., "Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score," *Int. J. Comput. Vision* **72**, 179–193 (Apr. 2007).

[21] Gall, J., Yao, A., Razavi, N., Van Gool, L., and Lempitsky, V., "Hough forests for object detection, tracking, and action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 2188–2202 (Nov. 2011).

[22] Prasad, M., Fitzgibbon, A., Zisserman, A., and Van Gool, L., "Finding nemo: Deformable object class modelling using curve matching," in [*Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*], 1720–1727 (2010).

[23] Tuytelaars, T. and Gool, L. V., "Wide baseline stereo matching based on local, affinely invariant regions," in [*In Proc. BMVC*], 412–425 (2000).

[24] Pritchett, P. and Zisserman, A., "Matching and Reconstruction from Widely Separated Views," in [*3D Structure from Multiple Images of Large-Scale Environments, Lecture Notes in Computer Science 1506*], Koch, R. and Gool, L., eds., 78–92, Springer-Verlag (June 1998).

[25] Furukawa, Y. and Ponce, J., "Accurate, dense, and robust multiview stereopsis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**(8), 1362–1376 (2010).

[26] Gong, F. and Kang, C., "3d mesh skeleton extraction based on feature points," in [*Proceedings of the 2009 International Conference on Computer Engineering and Technology - Volume 01*], *ICCET '09*, 326–329, IEEE Computer Society, Washington, DC, USA (2009).

[27] Mathias, M., Martinovic, A., Weissenberg, J., and Gool, L. V., "Procedural 3d building reconstruction using shape grammars and detectors," in [*Proceedings of the 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*], *3DIMPVT '11*, 304–311, IEEE Computer Society, Washington, DC, USA (2011).