CREATION OF VIRTUAL WORLDS FROM 3D MODELS RETRIEVED FROM CONTENT AWARE NETWORKS BASED ON SKETCH AND IMAGE QUERIES

Theodoros Semertzidis[a,b], *Kevin McGuinness*[c], *Petros Daras*[b], *Lambros Makris*[b],
Noel O. E'Connor[c], *Michael G. Strintzis*[a,b]

[a]Information Processing Laboratory, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, GR54124, Thessaloniki, Greece

[b]Informatics and Telematics Institute, Centre for Research and Technology Hellas, GR57001, Thessaloniki, Greece

[c]CLARITY: Center for Sensor Web Technology Dublin City University, Glasnevin, Dublin 9, Ireland

ABSTRACT

The recent emergence of user generated content requires new content creation tools that will be both easy to learn and easy to use. These new tools should enable the user to construct new high-quality content with minimum effort; it is essential to allow existing multimedia content to be reused as building blocks when creating new content. In this work we present a new tool for automatically constructing virtual worlds with minimum user intervention. Users can create these worlds by drawing a simple sketch, or by using interactively segmented 2D objects from larger images. The system receives as a query the sketch or the segmented image, and uses it to find similar 3D models that are stored in a Content Centric Network. The user selects a suitable model from the retrieved models, and the system uses it to automatically construct a virtual 3D world.

1. INTRODUCTION

IPTV and social media TV are just around the corner and 3DTV has already arrived. Continuing this trend, user generated 3D virtual environments are predicted to be the new big thing in communications and infotainment platforms. Since the target audience is extremely broad, these places have to be constructed dynamically from each user based on the exact needs of the session. To gain significant user acceptance the technology has to be both useful and easy to use. The plethora of multimedia content already available on the web could conceivably act as ready-to-use building blocks for these new virtual environments.

The current Internet architecture has neither the structure nor intelligence to efficiently exploit the available multimedia content. Content Aware Network (CAN) architectures [1, 2, 3, 4] have emerged to address this deficiency, offering more efficient content indexing, search, and delivery. Van Jacobson et. al. [4] proposed a Content-Centric Networking (CCN) approach where the content is routed based on hierarchically named components. The CCN protocol is based on two packet types: Interest packets and Data packets. The consumer transmits an Interest packet, which is propagated in the network to nodes with available content. Data packets return to the consumer from the path that the Interest packets passed. Based on the CCN architecture, Daras et. al. [5] proposed an extension that introduced similarity-based multimedia content search in such networks. In search enabled CCN the user is able to retrieve content objects by name as well as by example.

Assemb'Live, 3Dxplorer, web.alive and Second Life are just some of the existing commercial software for 3D virtual worlds. Unfortunately, none of them feature an easy-to-use interface for immediate construction of 3D environments from existing multimedia content. In this work, we present a framework for constructing 3D virtual environments using a simple interface. The 3D models that are used as building blocks for the 3D environment are retrieved after a similarity search procedure that matches the 2D queries with similar 3D models in a CAN or a local 3D database. The query may be a sketch or an interactively segmented object from a 2D image.

The remainder of the paper is organized as follows: Section 2 presents the system architecture. Section 3 gives a detailed description of the interactive segmentation method that we used. Section 4 describes the sketch user interface while Section 5 stresses the similarity search procedure inside the CAN. Section 6 discusses the construction of the 3D scene. Finally, Section 7 draws the conclusions and gives insights for the future work.

2. SYSTEM ARCHITECTURE

The system consists of the virtual world creator (VWC) application, which is the main application that provides the graphical user interface, the searchGateway application, which acts as a gateway between TCP/IP, and the CCN network and the

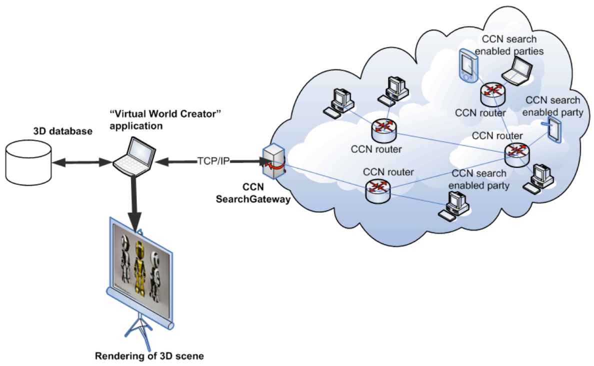

Fig. 1. The system architecture

(a) Source Image

(b) Interactively segmented object

(c) Segmented object with clear boundaries found using the eraser tool

Fig. 2. Extracting a query object using the Interactive Segmentation Tool

searchProxy application, which runs on each node of the Content Centric Network to perform the similarity search process (Figure 1).

The virtual world creator tool integrates two different user input interfaces. The first is a sketchpad for the formation of sketch queries, while the second is an Interactive Segmentation Tool for extracting 2D queries from images. The main application (VWC) provides a dockable window for rendering the 3D virtual scene using an OpenGL-based renderer, which is the final output of the system. The VWC application also provides the user with the ability to select whether the search will be performed in the CAN or in a local multimedia database containing 3D models and their corresponding descriptors. With this feature the user is able to perform targeted search actions to specific databases that are locally available (e.g. a company owned database) or perform a more general search inside the Content Centric Network.

The other two applications are described in detail in [6]. We also briefly describe them in Section 5 since they are related to similarity search in CCN.

3. INTERACTIVE SEGMENTATION TOOL

To allow for an object in an image to be used as the basis for a 3D model, there must be some way for the user to express to the system the location and boundary of the relevant object. Scribble-based interactive segmentation [7] provides a fast and convenient way for a user to mark objects in images; the user simply marks sample object pixels and background pixels with the mouse, and the underlying algorithm updates the segmentation to reflect their input.

We used the popular simple interactive object extraction (SIOX) algorithm [8] to handle the interactive segmentation in our system. The algorithm operates by using pixels marked by the user to build a color model of the object and background regions, and then classifies the pixels in the image as either object or background based on their distance from this model. It has been integrated into the popular open-source imaging program GIMP as the Foreground Select Tool and is used by the Trace Bitmap feature of the Inkscape vector graphics illustrator. Below is a brief overview of the algorithm; a detailed description can be found in [8].

The algorithm assumes a feature space that correlates well with human perception of color distances with respect to the Euclidean metric. As such, the first step in the method is to transform the image color into the CIE-Lab space [9]. Once the image is in an appropriate color space, the next step is to generate a color signature [10] for the known object and background pixels indicated by the user markup. A modified k-d tree optimization algorithm [10] is used to efficiently generate the color signatures. Assuming known cluster sizes based on the perceived diversity on each color axis, the algorithm proceeds in two stages. In the first stage, given a single starting cluster containing the entire sample, the cluster is recursively bi-partitioned until each is within a prescribed size. Running the same k-d algorithm on the centroids found in the first stage then recombines the resulting clusters. Clusters containing less than 1% of the pixels are discarded. The final cluster centroids constitute the color signatures for the object and background regions.

Using the generated color signatures the unknown image pixels are then classified as foreground or background according to the minimum distance to any mode in the foreground or background color signatures. The result is a confidence matrix containing values between zero and one: zero denoting background, one denoting foreground. In the final stage of the algorithm, the confidence matrix is smoothed and regions disconnected from the largest object are removed. In addition to the scribble-driven interactive segmentation functionality provided by the SIOX algorithm, we also included a manual

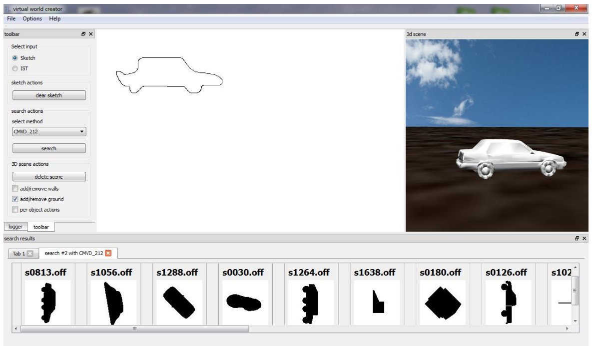

Fig. 3. The VWC application showing the sketch user interface in action. The 3D renderer is shown on the right and the results tab on the bottom pane.

eraser tool to allow users to quickly repair inaccuracies near the boundary of the object. Figure 2(a) and (b) show an image and the corresponding segmentation as found using the SIOX algorithm. Figure 2(c) shows the final segmentation after correction using the eraser tool.

4. SKETCH USER INTERFACE

The sketch user interface (Figure 3) is used to get from the user not only the object to search for, but also the orientation, the position, and the scale inside the 3D virtual world. Since the search has to be conducted one object at a time, the user must trigger a search process each time a new object is sketched in the sketchpad. The segmentation procedure that we use makes the assumption that the user does not draw overlapping objects. Taking this into account a history of the sketched images is kept between each search. To isolate the new object from the sketched scene, the current sketch image is automatically subtracted from the previous one. The next step is the calculation of the bounding square. For performance reasons, if the bounding square is larger that 100×100 pixels, the image is scaled to 100×100 pixels. Finally, the 100×100 pixel sketch image is fed in the CMVD [11] descriptor extractor and a descriptor vector containing low-level features is computed. This vector is the actual query of the drawn object.

5. SIMILARITY SEARCH IN CONTENT CENTRIC NETWORK

Although content aware network architectures usually present significant advantages over previous networking architectures, they do not solve the similar content search problem, which is critical for the evolution of the Future Media Internet.

The CCN architecture description [4] and implementation [12] permit retrieval of content, provided that the consumer knows the name of the desired content object, or a prefix of this name. Our work in [5] was to propose a similarity search protocol to extend CCN architecture and in [6] to present a first attempt of using this protocol in a CCN to find 3D models from which to build a scene.

Search enabled CCN networks have two basic components as seen in Figure 1: a) the CCN Gateway which is responsible for the interconnection of the CCN with the TCP/IP network, and b) the CCN party which is a node of the CCN network that acts as consumer and producer of data. The CCN Search Gateway computer should have the searchGateway application and a CCND [12] daemon that is the actual CCN router. On the other hand, for the CCN party, the applications needed are: the CCND router, the searchProxy application that implements the search protocol and conducts the searches for similar objects, and a file proxy application. The file proxy application is available as a demo application within the CCNx project distribution, implementing a CCN repository [12]. Each searchProxy instance is responsible for indexing the content that is available on the partys local repository, and for replying to search queries if similar content exists in its index. CMVD [11] descriptors are extracted for every 3D object in the local repository that is to be indexed. A k-d tree indexing structure is used to perform fast nearest neighbor search on this index. Figures 3, 4 present search results retrieved from a sketch and a segmented image query respectively.

6. CONSTRUCTION OF THE 3D VIRTUAL SCENE

The orientation, position and scaling of the 3D models in the 3D virtual scene are the basic problems that we have to overcome in order to have an automated construction of the 3D environment minimum user intervention.

6.1. Sketch Input

In the case of sketch user input these three parameters are extracted from the sketchpad. The position of the 3D model in the 3D environment has to be inferred from the position of the object on the 2D plane. In our approach we assume that all the objects in the scene are attached to the ground and so the Y coordinate of the 3D world is always $Y = 0$. Thus by using as a reference point the lower-most point of the sketched object, we map the X-Y coordinates of the 2D object to the X-Z coordinates of the 3D world. A heuristic stretch factor is necessary to map the height of the sketch image to the required depth in the 3D world. The user interface provides the user with the ability to change the stretch factor and experience a deeper or narrower 3D scene after the insertion of the objects in the scene. In addition, taking into account that the 3D models are normalized at the unit sphere for the extraction of the CMVD descriptors [11], the ratio of the 2D sketches is used to apply the scale in the 3D world objects.

The orientation of 3D models is still an open research challenge. In this work we use a simple solution that works in most cases. We assume that if we had a large unbiased

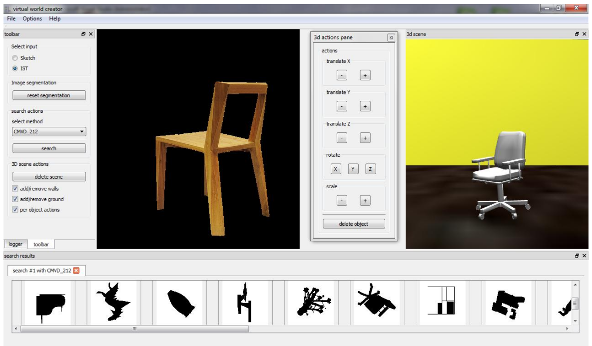

Fig. 4. Inserting and manipulating a 3D model in the scene.

database of 3D objects and their 2D view descriptors, and we query this database with a sketch of a 2D view, then the most similar results would be not only the most similar objects but also the exact views of these similar objects. Moreover, by using the rotation matrix that we used for the extraction of the 2D views of each model, we could orient the 3D models in the 3D world with respect to the virtual camera coordinates. Since the descriptor extraction technique that we use extracts descriptors for 18 views of each 3D model, the application is able to orient the 3D model in question in 18 different angles in the scene [11]. For finer tuning of the orientation of the 3D model, the user interface of the VWC application allows for manual in place rotation of the model.

6.2. Segmented image input

In the case that the user selects to use the Interactive Segmentation Tool to extract a 2D object for his/her query the valuable information such as scale, and position, that are available in the sketch input, are lost. For this reason the GUI of the VWC application provide the functionality (middle of Figure 4) to place the object in a specific place in the 3D scene and scale and rotate it appropriately.

7. CONCLUSIONS AND FUTURE WORK

In this paper we presented a tool for automatic creation of 3D virtual environments from sketch or interactively segmented images through the retrieval of similar 3D models available in a Content Centric Network or local DBs.

The similarity search supporting CCN enables the users to retrieve multimedia content already available in the network without knowing any information about the content. The application provides two different user interfaces that are intuitive to use and provide a greatly enhanced user experience for multimedia content search while it helps users to express in detail their thoughts and ideas.

In the future we plan to develop a module for 3D rendering of the virtual world targeting a 3D autostereoscopic display. The choice of SIOX algorithm for interactive segmentation was primarily one of convenience, and may not be optimal. In particular, the interactive graph cuts algorithm has been shown to generally outperform SIOX for a large variety of natural scenes [7]. We plan to test this algorithm with our system in the future.

8. ACKNOWLEDGMENTS

This work was supported by the EU FP7 project 3DLife, ICT-247688.

9. REFERENCES

[1] N. Niebert, S. Baucke, I. El-Khayal, M. Johnsson, B. Ohlman, H. Abramowicz, K. Wuenstel, H. Woesner, J. Quittek, and L.M Correia, "The way 4ward to the creation of a future internet," in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2008, pp. 1–5.

[2] G. Bouabene, C. Jelger, C. Tschudin, S. Schmid, A. Keller, and M. May, "The autonomic network architecture (ANA)," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 1, pp. 4–14, Jan 2010.

[3] Teemu Koponen, Mohit Chawla, Byung-Gon Chun, Andrey Ermolinskiy, Kye Hyun Kim, Scott Shenker, and Ion Stoica, "A data-oriented (and beyond) network architecture," in *Proceedings of ACM SIGCOMM 2007*, Aug 2007.

[4] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, 2009, CoNEXT '09, pp. 1–12.

[5] P. Daras, T. Semertzidis, L. Makris, and M. G. Strintzis, "Similarity content search in content centric networks," in *Proceedings of the ACM international conference on Multimedia*, 2010, MM '10, pp. 775–778.

[6] T. Semertzidis, P. Daras, P. Moore, L. Makris, and M. G. Strintzis, "Automatic creation of 3d environments from a single sketch using content centric networks," *IEEE communications Magazine*, 2011, (to appear).

[7] K. McGuinness and N. E. O'Connor, "A comparative evaluation of interactive segmentation algorithms," *Pattern Recognition*, vol. 43, pp. 434–444, February 2010.

[8] G. Friedland, K. Jantz, and R. Rojas, "SIOX: simple interactive object extraction in still images," in *Seventh IEEE International Symposium on Multimedia*, Dec 2005, p. 253260.

[9] *Color Science: Concepts and Methods, Quantitative Data and Formulae*, Wiley, New York, 1982.

[10] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, pp. 99–121, November 2000.

[11] P. Daras and A. Axenopoulos, "A compact multi-view descriptor for 3d object retrieval," in *Seventh International Workshop on Content-Based Multimedia Indexing*, June 2009, pp. 115–119.

[12] "Project CCNx," `http://www.ccnx.org/`.