VIEWPOINT-BASED SEARCH AND BROWSE OF DIGITAL ARCHIVE CONTENT

R. Kadobayashi
3-5Hikaridai, Seika-cho, Soraku-gun, Kyoto619-0289, Japanrieko@nict.go.jp

KEYWORDS:3D, Archaeology, DataRetrieval, InformationSystems, SpatialInformation.

ABSTRACT

We propose new search and browse methods that use three-dimensional (3D) viewpoints as queries of digital archive content. The idea is that by linking 3D models and photographs via spatial information, the view point in particular, we can use them as a reference for each other when browsing photographs and/or information, or walking through the 3D virtual space. A 3D viewpoint-based approach is especially useful for searching collections of archaeological photographs that contain many different images of the same object. Our method is designed to enable users to retrieve images that contain the same object but show a different view and to browse groups of images taken from a similar viewpoint. In addition, the information attached to each image, such as description and URL to other resources, is automatically displayed so that the user can obtain knowledge about the objects in the current 3D scene. We also propose using 3D scenes to query by example, which means that users do not have the problem of trying to formulate appropriate queries. This combination gives users an easy way of accessing not only photographs but also archived information. A 3D viewpoint-based method can also be used when viewing 3D models or walking through 3D virtual spaces. To view particular scene/scenes, a user just needs to choose one or more photograph(s) in the digital archives. Then, the system automatically detects the view points of the photographs and uses them to render 3D scenes. This helps users to view the 3Dmodels/scenes easily. The prototype system that uses the data of an actual archaeological site will also be introduced.

1. INTRODUCTION

Recording cultural heritages in digital form is becoming more popular. The number of images of monuments and artefacts is increasing rapidly and the size of their 3D models is becoming larger. This causes the problem that users have difficulties in retrieving images and/or viewing 3D models as they want. As a result, there is a strong demand for an easy-to-use user interface for searching and browsing the content of digital archives.

Among many kinds of digital archives, our research focuses on archaeological digital archives. A notable feature of these archives that sets them apart from other kinds of digital archives is that they contain many different images of the same object. This is due to the nature of archaeological study, i.e., excavation. Excavation of a particular site often takes place over several years or even several decades. The appearance of uncovered objects changes as the excavation proceeds. This is one reason for the existence of different images of the same object. Another reason is that to record objects and the site in as much detail as possible, photographs are taken from as many different view points as possible. Recording the work done at the site is also important. As a result, there are many images of the site or of objects with researchers in a variety of situations, such as surveying and drawing. Images in which the main subject is the same but the appearance, viewpoint, or situation is different are shown in Figure1.

Figure1:Photographs of same subject showing different appearance (top right), viewpoint(bottom left), or situation (bottom right).

These photographs are from the photo archive of the Gemiler Island Site, which was a result of collaborative work with the Research Group of Byzantine Lycia (RGBL, 2005), a Japanese research group consisting of art historians, architectural historians, historians, archaeologists, and students (Kadobayashi, 2004). The group start edits research on an archaeological site of the Byzantine era on Gemiler Island in Turkey in 1991 (Tsuji, 1995), and they have taken more than 30,000 photographs over the decade. When creating the photo archive, which has more than 15,000 photographs, they added several keywords to each photograph. Due to the diversity of their backgrounds, they used different vocabulary. Therefore, in the photo archive, there are a lot of images in which the main subject is the same but the keywords are different or the main subject is slightly different, e.g., distant view and close up, but the keywords are the same. This variance makes it difficult even for the group members who are experts of the archive content to retrieve the desired images.

Keyword-based searching of images in archaeological digital archives is often unsatisfactory. Users may have difficulty combining several keywords to formulate a query that can distinguish between images that are very similar to each other. More importantly, a keyword-based search requires users to have some knowledge about the content so they can make an appropriate query. However, this assumption is too restrictive in actual use. Another retrieval method, called Content-Based Image Retrieval (CBIR) (Flickner et al., 1995; Smith et al., 1996; Assfalg et al., 2000), uses image features, such as color, shape, and texture, that are automatically extracted from images. However, due to the special characteristics of the images contained in archaeological digital archives, as mentioned above, CBIR is also unsuitable.

Another feature of archaeological digital archives is that the size of the 3D models of an archaeological site tends to be larger than that of works of art, such as sculptures. For example, we have made a 3D model of Church III, a ruined Byzantine basilica on Gemiler Island and its floor plan measures approximately 30x15 m, which has more than 6,000,000 polygons (Figure2) (Kadobayashi et al., 2003). The more the data size increases, the less operational the model will be.

Figure 2. View of 3D model of Church III, a Byzantine basilica on Gemiler Island, Turkey. Number of polygons of this model is reduced to about 600,000, 1/10 compared to original.

Users often have to struggle to manipulate 3D models to view a particular part of the model. If instead they could specify the view point by using images that are easier to handle, this would greatly improve the user interface.

We, therefore, propose a viewpoint-based method (Kadobayashi et al., 2005a; Kadobayashi et al., 2005b) for searching and browsing the content of digital archives. The idea is that by linking the 3D models and photographs via spatial information, "viewpoint" in particular, we can use them as a reference for each other when browsing photographs or walking through the 3D virtual space. Intuitively, the proposed method compares the view point of a user of a digital archive who explores the virtual 3D space of an archaeological site with that of a photographer who took photographs of the site. Our method is designed to enable users to retrieve images that contain the same object but show a different view and to browse groups of images taken from a similar viewpoint.

Figure3:Conceptual example of viewpoint-based search with query by example interface. Selected 3D scenes as query (top) and best matched photographs (bottom). Note that difference between two photographs is hard to express by keyword sboth when giving keywords and searching photographs.

In addition, the information attached to each image, such as description and URL to other resources, is automatically displayed so that the user can obtain knowledge about the objects in the current 3D scene. This approach allows users not only to retrieve images easily and intuitively without specifying the correct keywords but also to access content in a digital archive.

As for the handling/walking-through of 3D models, our viewpoint-based method allows users to see a desired scene(s) by just choosing a photograph(s) from the archive. The system automatically renders the 3D scene based on the viewpoint of the selected photograph(s). Hence, users can explore the virtual 3D world more easily.

A 3D scene can be used for a query to search and browse photographs and information. Using photograph(s) as a key for specifying a particular 3D scene(s) to see the proposed viewpoint-based approach for browsing content does not require the user to have knowledge about the content or skill to draw example pictures. It enables users to enjoy quick and easy browsing of content.

The rest of this paper is organized as follows. Section 2 describes the proposed methods for effective and efficient content browsing and section 3 discusses the prototype system. Section 4 concludes this paper.

2. VIEWPOINT-BASED APPROACH

2.1 Basic idea

We can regard a photograph as a sliced image of a real 3D world accordino to the photographer's viewpoint while we can regard a 3D scene as a snapshot of a virtual 3D world according to the user's viewpoint. Once the viewpoint is defined, we can compare them and judge the similarity. Our idea can be summarized as "a viewpoint-based search exploiting the analogy between camera and computer graphics (CG) with a query-by-example interface."

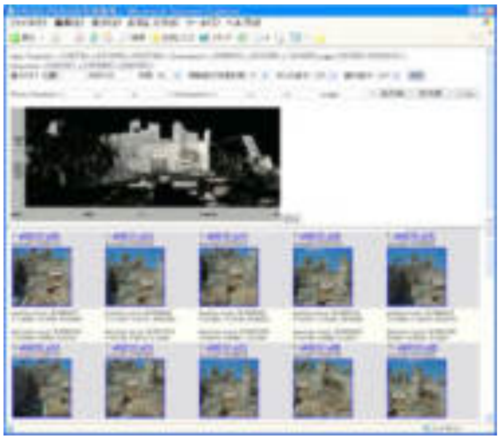

Figure4:User interface of viewpoint-based photograph search. 3D model is displayed in upper part. Lower part is for showing photographs listed in descending order of matching score.

First, it is based on the idea of using viewpoints as queries rather than using keywords like keyword-based search or image features like CBIR. Viewpoint is basically "a place from which you can see something" (Longman; 2003), but it implies the direction at which a person looks. Hence, in this study, we define viewpoint as a 3D vector starting from a viewer's eye to a target the viewer is looking at. This means viewpoint consists of position (x,y,z) and direction (x,y,z). A viewer could be a photographer or a user of a digital archive.

A photograph's viewpoint is defined using the viewpoint of the camera that was used by the photographer to take the image rather than the photographer's viewpoint since it is not easy to accurately detect the position and direction of the photographer in the real world. The camera's viewpoint can be calculated from extrinsic camera parameters, Rotation (Rx,Ry,Rz) and Translation (Tx,Ty,Tz), which are obtained by applying a self-calibration method (Tsai, 1986). The view point is given to the image as metadata and stored in the digital archive. On the other hand, a user's viewpoint of a particular 3D scene is substituted for a viewpoint of a virtual camera, which is used to render the 3D scene. The virtual camera's viewpoint is a 3D vector starting at the camera position and pointing at the camera target.

The second idea behind our method is the use of a query-by-example interface when formulating queries. It is not intuitive to specify the 3D coordinates of the viewpoint, such as "position (12.345, 2.468, 9.753) and direction (-1.111, 2.222, -3.333). "Instead, 3D scenes that the user encounters on a virtual 3D tour of the archaeological site are used to form queries to search for photographs while photographs the user chooses are used to form queries for showing a particular 3D scene on the screen. Note that our method is not CBIR, although it uses the "query by example "technique, which is common in CBIR systems. The specified 3D scene is not used as a query but the view point of the scene is used as a query.

Finally, the analogy between the camera angle and the fields of view of the computer graphics is exploited. To obtain a similar appearance between photographs and 3D scenes, the field of view of the virtual world is set according to the focal length of the camera by which the photographs were taken.

2.2 Viewpoint-based photo search

Intuitively, the viewpoint-based method works like this. If the 3D scene, such as that shown at the top left of Figure 3, is specified by a user during a walk through the virtual 3D world, as shown in Figure 2, the system would rank the image shown at the bottom left of Figure 3 higher than the image shown at the bottom right of Figure 3. On the contrary, if the user specifies the 3D scene shown at the top right of Figure 3, the system would rank the image shown at the bottom right of Figure 3 higher than the image shown at the bottom left of Figure 3. Both images shown at the bottom of Figure 3 cover the same part of Church III , i.e., apse and bema, but the right image covers a wider area that includes the nave and southwall. The user may not know if these similar but different photographs exist in the archive. Even if the user knows such a pair exists, the user may not know how to distinguish the two images with keywords.

Furthermore, in the former example, the image, as shown at the right-hand side of the bottom of Figure 1, would be ranked higher as well as the images shown at the top of Figure 1. The image shown at the left-hand side of the bottom would be ranked lower because the viewpoint of this photograph is most different from the rest. We assume that grouping the images that have a similar viewpoint would help users browse many images effectively and efficiently.

The search is done like this. When a user wants to see photographs or obtain information, he/she first views the 3D model and walks through it until he/she comes across an interesting scene. After the user clicks the search button, the system detects the viewpoint of the 3D scene, uses it as a query, and then searches photographs that have a similar viewpoint. Note that this is not a content-based image retrieval. Unlike CBIR, the viewpoint-based search is robust against differences in color, shape, and texture betweeni mages.

The search algorithm is as follows. First, theinner product of the position vectors of the virtual camera rendering the 3D scene and that of the photographs is calculated and then multiplied by a weight set by the user. This is the position score. The orientation score is calculated similarly. The matching score is the sum of the position and orientation scores. The search result is a list of thumbnail photographs ranked in descending order according to the matching score.

The main advantages of our 3D viewpoint-based photo search are summarized as follows. It is very easy and intuitive for users to search and browse photographs because the query is automatically formed by the system when a user specifies the 3D scene of interest, and this does not depend on the users'level of knowledge about the content or their skill in drawing example images. In addition, our method retrieves images of the same object with a different appearance. This helps users search and browse photographs that were taken at different stages of the excavation process.

2.3 Viewpoint-based information browsing

Usually photographs in a digital archive have several keywords, an explanation, and URLs to related web pages, although that is rare for 3D models. This means users of a digital archive do not obtain pertinent information about a particular part of the 3D model easily just by viewing the 3D model in a traditional way.

To overcome this problem, information given to a photograph is regarded as information about a particular part of the 3D model if their viewpoints, as shown in a display, are similar. The system automatically provides a user with information given to photographs, which are the result of searches when the user browses the photographs. This has a great advantage because the user is not required to enter the technical term to specify the particular part of interest.

2.4 Viewpoint-based 3D scene rendering

A viewpoint-based approach can be used for rendering the 3D scene too. Usually, it is difficult to control the 3D model to obtain the desired scene. This will be worse if the size of the model increases. The idea is to use a photograph as an example for rendering the 3D scene. It is very intuitive and easy for users because they just need to select a photograph(s) in the digital archive. Then, the system automatically sets the view point of the 3D space according to the viewpoint of the photograph(s) and renders the scene without their operation on 3D models. This is not the image-based rendering method that uses several images to render a 3D scene. Our method uses the images as an intuitive interface to specify the viewpoint, as described in Section 2.1.

3. PROTOTYPE SYSTEM

We have implemented a prototipe system that contains 302 photographs of Church III as test data. These photographs were taken at the Gemiler Island site with a digital camera in 2002 when we carried out 3D digitizing of two basilicas, called Church II and Church III. The system interface is web-based, and a Cortona VRML Client (Cortona; 2005) is used to tour the 3D virtual space.

First, we needed to obtain extrinsic camera parameters, position (X,Y,Z), and orientation (ω,φ,κ), of each photograph. To do so, we developed camera calibration software. The software can show both an RGB image produced by a laser scanner and a photograph so that the user matches the feature points that appear in both images. After specifying more than six corresponding features in both the 3D image and photograph, the software executes a calibration and then outputs the camera parameters into a file. The final result is obtained by improving the extrinsic camera parameters obtained by solving linear equations for camera calibration (Faugeras, 1993) with non linear optimization. The error between reference points projected onto the image plane using extrinsic and intrinsic camera parameters obtained by calibration beforehand and corresponding points in two dimensions was used by the function minimized by non linear optimization. The Newton method was used for non linear optimization. Then, using the extrinsic camera parameter, the photograph's viewpoint vector was calculated.

Figure 5: Results of viewpoint-based search. Scenes used inquery are outside southwall, inside southwall, apse from west, and apse from east (froma bove). In each case, top ten retrieved images shown in lower part include part specified in scene.

We made a light 3D model of Church III in VRML format by reducing the number of polygons from 6,300,000 to approximately 600,000. Inspite of the drastic reduction, the model well preserves the features of the parts of Church III and lets users grasp the whole shape at a glance. However, the model may not be suitable for close-up observation.

3.1 Viewpoint-based photo search

The user interface of the prototype system is shown in Figure 4. In the upper part, there are buttons and dialog boxes for user interactions, and a 3D model viewer by which the user can walk through the 3D space. The lower part is an image viewer where retrieved images are shown according to the matching score in descending order. The matching score compares the vectors of the camera position and direction as described in Section2.2.

The search procedure is very simple. The user views the 3D scene displayed in the viewer and walks through the 3D space until he/she comes across an interesting scene and clicks the "search" button. Note that the number of photographs displayed in the image viewer is determined not by the threshold of the matching score but by the parameters that the user can select from several options, such as 40, 100, or "all." The 3D scene is very intuitive, even for those who do not have any knowledge about the site or monument. Thus, the user can easily search and browse photographs of interest.

Figure 6. Screen shot of viewpoint-based information browsing. Five images with best match are automatically shown in a "collection viewer" at right with keywords and description if a user specifies automatic display of information be fo research. Accumulated information from five highest-ranked images is automatically shown in a window (bottom) if a user specifies automatic displays.

An example of retrieval results is shown in Figure 5. As seen in the upper two cases, all of the top teni mages are very similar to the 3D scene. On the other hand, in the lower two cases, the apse displayed in the 3D model is the same, but the viewpoints are different. The 3D scene of the third example is a view of the apse from the west while that of the fourth example is a view from the east. Note that the users do not need to specify the direction explicitly with keywords. All they have to do is rotate the viewpoint of the 3D scene. Then, they can browse images that match the 3D scene.

It should be noted that the 3D model does not require many details, and it could be a simple model with a small number of polygons. The main objective of the use of the model is to formulate a query rather than to observe it's shape closely. As mentioned before, it is difficult to control 3D models with a large number of polygons and/or many textures. One possible option could be to use different 3D models in viewpoint-based photo and information searches and in viewpoint-based 3D scene rendering.

3.2 Viewpoint-based information browsing

A screen shot of view point-based information browsing is shown in Figure 7. At the top, there are buttons and dialog boxes for user interactions, and a 3D model viewer lets the user walk through the 3D space. On the right is a collection viewer showing the most relevant images with information attached to each image. Below is a summary viewer only showing information attached to the most relevant images selected. These viewers not only help users to search and browse photographs but also to access information stored in the digital archive. A window listing thumbnail images ranked according to the matching score is also available but is not shown in the figure.

3.3 Viewpoint-based 3D scene rendering

The prototype system also has a function for rendering a 3D scene based on selected photograph(s). The system regards the specified image as a reference and uses the viewpoint of the virtual camera to render the 3D scene. Thus, the user can easily specify the desired scene to walk through.
When the system is initiated, the model is shown in the 3D model viewer and the photographs are listed below. A user can choose photographs by marking the check box. In the case shown in Figure 7, a user chose three images. Then, the user clicked the button "moving," which lets the system start changing the viewpoint of the virtual camera. As the user chose three photographs, the system changes the viewpoint three times. To achieve a natural scene-change effect, the system pauses two seconds at a scene before changing the viewpoint. Finally, the 3D model viewer shows the 3D model according to the viewpoint of the last selected photograph. The list window automatically changes the list view of photographs and shows the last selected photograph in the top row of the window. From Figure 7, it can be easily seen that a 3D scene, similar to the photograph that the user selected, is rendered in the model viewer.

Figure 7. Viewpoint-based 3D scene rendering. When the system is started, photographs are listed according to their file name as shown in the left most snapshot. User chose three images (marked with rectangle) from the list and the system automatically changes view of 3D model three times. Scene change is performed lik eanimation.

4. CONCLUSION

There is no doubt that more digital archives of our cultural heritage that include 3D CG models and photographs will appear, and hence, there is a need for an easy-to-use interface for searching and browsing content. We believe that our viewpoint-based methods of searching and browsing photographs and information by 3D scene and rendering 3D scenes by using photographs will comprise a powerful and easy-to-use interface for cultural heritage archives.
We are trying to analyze the photographs in our photo archive and apply the method to them so that we can demonstrate its usability. We are thinking of integrating a keyword-based search as well as improving the prototype system.

ACKNOWLEDGMENTS

The author would like to express her gratitude to Research Group of Byzantine Lycia for providing materials.

REFERENCES

Assfalg,J., and Para,P.,2000.Querying by Photographs: AVR Metaphor for Image Retrieval. IEEE Multi Media,Vol.7,No.1,pp.52–59.

Cortona VRML client.http://www.parallelgraphics.com/products/cortona/(accessed 24 May 2005)

Faugeras,O.D.,1993.Three-Dimensional ComputerVision: A Geometrical Viewpoint, MIT Press,Cambridge,MA,1993.

Flickner,M.,Sawhney,H.,Niblack,W.,Ashley,J.,Huang,Q.,Dom, B.,Gorkani,M.,Hafner,J.,Lee,D.,Petkovic,D.,Steele,D.,andYanker,P.,1995.Query by Image and Video Content: The QBIC System, IEEE Computer,Vol.28,No.9,pp.23–32.

Kadobayashi,R.,2004.Creating 3D Digital Archivi of Bizantine Ruins in Turkey. In:Proceedings of the Electronic Imaging and the Visual Arts(EVA2004Florence),pp.238–243.

Kadobayashi,R.,andFurukawa,R.,2005a.Combined use of 2D images and 3D models for retrieving and browsing digital archive contents.In:Proceedings of SPIE-IS&T Electronic Imaging,Vol.5665,pp.134-143.

Kadobayashi,R.,Furukawa,R.,Kawai,Y.,Kanjo,D.,andYoshimoto,J.N.,2003.Integrated Presentation System for 3D Models and Image Database for Byzantine Ruins In: Proceedings of the ISPRS Workshopon Vision Techniques for Digital Architectural and Archaeological Archives(IS-PRSXXXIV-5/W12),pp.187–192.

Kadobayashi,R.,andTanaka,K.,2005b 3D Viewpoint-based Photo Search and Information Browsing. In: Proceedings of ACMSIGIR 2005 (to appear).

Longman Dictionary of Contemporary English 4 with CD, Pearson ESL,(March2003)

Research Group for Byzantine Lycia: http://www.jttk.zaq.ne.jp/sfuku239/lycia/ (accessed 24 May 2005)

Smith,J.R.,andChang,S.F.,1996.Visual SEEk: a fully automated content-based image query system. In: Proceedings of ACM International Conference on Multimedia,pp.87–93.

Tsai,R.,1986.An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,pp.364–374.

Tsuji,T.,ed.,1995.The Survey of Early Byzantine Sites in oludeniz Area(Lycia,Turkey).Osaka University,Vol.35.