Easy and Effective Virtual Tour on the World Wide Web

Ilmi Yoon¹, Andrew Kang¹, John Roberts¹, and Sanghyuk Yoon²

¹San Francisco State University
²Rocky Mountain Biological Laboratory

Web-based Virtual Tour has become a desirable and demanded application, yet challenging due to the nature of web application’s running environment such as limited bandwidth and no guarantee of high computation power on the client side. Image-based rendering approach has attractive advantages over traditional 3D rendering approach in such Web Applications. Traditional geometry-based approach, such as VRML, requires labor-intensive 3D modeling process, high bandwidth and computation power especially for photo-realistic virtual scenes. QuickTime VR and IPIX as examples of image-based approach, use panoramic photos and the virtual scenes that can be generated from photos directly skipping the modeling process. But, QuickTime VR and IPIX may not only require special cameras or effort to take panoramic views but also provide only one fixed-point navigation (look-around and zooming in-out) rather than ‘walk around’, that is a very important feature to provide immersive experience to virtual tourists. Easy and Effective Virtual Tour constructs virtual tour using several snap shots of conventional photos without special tools, build simple 3D space within each photo using spidery mesh, and expand the virtual spaces by connecting each other using simple user intervention to specify correspondence. The expanded virtual space provides virtual tourists with free navigation and immersive experience of walking around through the WWW.

1. Introduction

The more intensively World Wide Web is involved in our daily life (from searching information to doing business), the more sophisticated and interactive media types are desired and demanded. Tourists may plan their trips in advance by selecting tourist attractions by previewing through the WWW, or homebuyers search for houses on the WWW and want to narrow down the number of candidates for actual visit. When previewing through still image sceneries, viewers may frequently wish to jump into the scene and freely navigate inside to get more information if possible. When virtual tour can be easily constructed and browsed with similar cost/effort of taking and browsing still images, then, web-based virtual tour will be obviously preferred over still images.

The cost of web-based virtual tour, in users perspective, includes downloading time and response time during navigation (rendering speed). In developer’s perspective, the cost includes labor of construction as well as requirement of special tools or knowledge. In this paper, we present a technique, EEVT (Easy and Effective Virtual Tour) that lowers the cost of viewer’s navigation close to the level of browsing still images while viewers are allowed to freely navigate inside virtual scenes. Building EEVT also can be done easily by ordinary users using handheld camera without requiring special 3D knowledge, labor-intensive effort nor special tools.

2. Related Research

There have been two major approaches to provide web-based virtual tour, more generally, interactive 3D graphics on the WWW (Web3D). One is traditional 3D graphics that transmits 3D geometry with textures to the client for rendering. VRML/X3D, Java3D, and MPEG4 are examples of geometry based Web3Ds. All these geometry-based techniques suffer to support photo-realistic virtual tour because geometry-based approach requires labor-intensive effort for modeling, creates huge datasets, and requires intensive computation power.

The other approach to provide Web3D is Image-based rendering that has attractive advantages over traditional 3D rendering approach. Image-based rendering enables skipping the labor-
intensive 3D modeling process of photo-realistic scenes. As a consequence, the resulting model is much smaller and does not require high bandwidth or intensive computation power at the client side. QuickTime VR [Chen 95] and IPIX as well known examples, use panoramic photos. The virtual scenes generated from panoramic photos directly enables skipping the modeling process. But, these image-based approaches require special cameras or effort to take panoramic views. Most of all, QuickTime VR and IPIX provide only one fixed-point navigation (look-around and zooming in-out only) rather than ‘walk around (free navigation)’, that is a very important feature to provide immersive experience to virtual tourists.

Concentric Mosaics [Shum 99] provides a much richer user experience by allowing users to move freely in a circular region and observe significant parallax and lighting changes. Concentric Mosaics is created by constraining camera capture motion to planar concentric circle, composing slit images taken at different locations, and indexing all input image rays in 3 parameters. And then, new views are rendered by combining the appropriately captured rays. This method still presents vertical distortions, but depth correction can alleviate the effect. Concentric mosaic requires much bigger data sets and more intensive per pixel computation than IPIX or QuicktimeVR. This technique also requires special tools and setups (a number of cameras mounted on a rotating horizontal beam that is supported by tripod) to capture concentric mosaic. Therefore, it becomes a difficult approach to ordinary users with one hand-held conventional camera.

Pseudo-3D photo collage is a simpler approach to create virtual walkthrough experience using multiple images [Tanaka 2002]. The basic idea comes from an artistic representation. "photo collage" on 2D canvases, that is, a general method of scanning and arranging original photos. Authors extend the idea to pseudo 3D by using the following interfaces; users apply their own photos and specify relations (spatial-hyperlinks) between them by specifying rectangle in one photo and the other rectangle in the other photo. These rectangle areas are now connected by spatial-hyperlink, a scene transits to the next when users click on one of the linked photos (the rectangle area), in a transition to the next scene, the current key photo fades out and the next key photo moves over to the center. This animation gives the user a sense of motion to the next position even though scene itself is composed of 2D still images. Pseudo-3D is simple and effective tool for ordinary users to make much use of their photos in a new spatio-temporal style, received strong positive feedback from many users, and many of its applications are presented at SIGGRAPH 2003 Web graphics and presentation. But the Pseudo-3D photo collage remains as connections of 2D images and limited to the illusion created by transitions between images, and does not provide free navigation inside the space by the viewers.

Tour Into the Picture (TIP), introduced by Horry et al. [Horry 97] constructs a properly textured pseudo-3D geometry space from a single picture using the spidery mesh and then allow a viewer to tour into the scene (Fig. 1). TIP is simple, but the result is impressive. A user can feel a plain 2D picture becomes a 3D scene. It can be toured into and viewed from different viewpoints. However, TIP is developed to produce off-line animations with well-controlled camera trajectory. When viewers get close to the walls and turn their orientation more than certain degree, then images may be seriously distorted due to the texture warping of the simple geometry (Fig. 2).

Tour Into the Picture Revisited [Li 2001] observed a problem of TIP’s original approach; the visual quality drops drastically when the viewpoint tours into the scene. It contradicted the real world experience where visual quality increases when viewer gets closer. Authors extended TIP by introducing the use of multi-resolution representation of the picture that the visual quality keeps nearly unchanged in the touring.
We make use of TIP to build Easy and Effective Virtual Tour (EEVT) on the WWW since TIP is naturally featured with benefits to Web3D applications such as (1) easy to create, (2) nearly no extra data needed to represent 3D not more than original photo, and (3) low computation requirement for rendering. However, navigating within a single scene created by a TIP makes virtual tourists feel confined and instigate strong desire to navigate further behind the given scene. Therefore, expanding virtual space by connecting multiple scenes became an essential task.

Seamless image registration or mosaics has been active research area within computer vision for more than a decade [Sawhney 1999][Kang 2004]. QuickTime VR or IPIX also makes use of this multi image alignment to create panoramic images as preprocessing and then allow viewers to see a portion of the panoramic image through a window. In EEVT, each scene constructs its own 3D space and users may freely navigate inside each scene. Unlike the panoramic view mosaic, the transition may happen at any position within each scene. Therefore, instead connecting photos during development as preprocessing, EEVT captures correspondence between scenes during the development and then uses the correspondence during actual navigation as a trigger to transition to the next scene as well as registration cue to align both scenes. This allows EEVT to expand the virtual space easily by making the smooth transition between scenes.

In summary, Easy and Effective Virtual Tour (EEVT) provides a simple (easy to create) and efficient way (no requirement for high computation nor high bandwidth) for ordinary users to create photo-realistic virtual tour on the web using only a conventional hand-held camera with a few steps of user interactions. Virtual spaces are smoothly expanded by adding more photos during the production of the virtual tour and during navigation by aligning scenes smoothly as well as receiving images progressively according to the direction of virtual tourists walking and viewing direction for pleasant navigation.

3. Easy and Effective Virtual Tour (EEVT)

Web-based Virtual Tour consists of two parts. One is a EEVT-Maker that allows content developers to build virtual space with several snap shots of conventional photos taken from hand-held cameras and the other is EEVT-Navigator for navigation of the built virtual space on the web.

3.1. EEVT-Maker

The web-based virtual tour is constructed using EEVT-Maker by following the steps below.

A. Photo connection

We assume the virtual tour content developer (in short, developer) took photos while walking around the location of interest, upload them into an image folder. EEVT-Maker reads in all the images in the folder, allowing users to pick and past into the Photo Connection Canvas (Fig. 3). Within the canvas, developer can choose two photos and assign the relation between the photos remembering the way photos were taken. Earlier implementation used matrix format where developer specify the size of connection matrix and then cut and paste photos into each cell. It was a simple connection, but limits each scene from being connected not more than 8 photos in static directions of forward, backward, left, right, forward left, forward right, backward left, and backward right. Current implementation uses
arrows to connect photos without limitations of number of photos to get connected. Approximation of relative positions where the photos are taken can be configured and then embedded into the system, simulating usual human walking. In addition, it can be easily extended to connect photos with non-planar relation such as up-forward, up-left, up-forward-left, down-forward, down-left, or down-right, etc. This relation between photos is used to guide the actual transition with default value or additional user input for more accurate alignment as needed in step C.

B. Single Image Spidery Mesh Setup

Spidery mesh consists of one vanishing point, four radial lines, and two rectangles over the original picture (Fig. 1a). The four radial lines radiate from the vanishing point. Each edge of the inner rectangle is parallel to one edge of the outer rectangle. The inner rectangle is used to specify the rear window in the 3D space. As a result, a scene of 5 walls (rear, floor, ceiling, left and right walls) is constructed with textures derived from the specified regions of the original image. For developer’s convenience, EEVT uses default setting of spidery mesh for each photo and then spidery mesh for each image is automatically generated. Developer can always check and change the default setting interactively. This spidery mesh creates a 3D space where viewer can navigate within (Fig 1b). This spidery mesh adds only about 1 Kbyte to the original image, but the result is impressive allowing users to navigate in 3D space made from one image (Fig. 2).

However as shown in figure 2, the viewers’ movement needs to be controlled to keep viewers from seeing severe distortion or partial image (due to the lack of information) as noted in figure.

C. Smooth Transition between Scenes

QuickTime VR or IPX uses tripod to take panoramic views, align and seamlessly connect photos. Sometimes photos are taken with lots of overlaps and then automatically/semi-automatically restore camera parameters to connect photos [ref]. In either case, connected photos result in cylindrical panoramic views and viewers can only rotate around or zoom in & out from a fixed point.

In EEVT, each scene constructs its own 3D space and users may freely navigate inside each scene. Unlike the panoramic view mosaic, aligning each image as preprocessing doesn’t make any sense since transition between images may happen at any position and direction within each scene. Therefore, instead connecting photos during development as preprocessing, EEVT captures correspondence between scenes during the development and then uses the correspondence during actual navigation as a trigger to transition to the next scene as well as registration cue to align both scenes. This allows EEVT to expand the virtual space easily by making the smooth transition between scenes.

The major objective of this EEVT is easy-of-use in its creation and ordinary amateurs with only conventional hand-held camera should be able to create virtual tours without special image mosaic or registration knowledge. We only expect developers will take photos with certain amount of overlaps, so EEVT-maker can use them as clues of correspondence, so that the transition become unnoticeably smooth without special registration process.

The correspondence setting can be done easily by marking same object in two different pictures as shown in figure ???. To make it simple, users draw a bounding box around each object. To closer the bounding box fits to the object, the smoother the transition become. Setting correspondence is done same way for forward-backward relation or rotation relation for users perspective. But internal computation is slightly different.
In case of forward-backward relation, the forward correspondence checking while navigating 1.bmp and 2.bmp is performed. The correspondence setting at 1.bmp and 2.bmp ensures that the corresponding object is marked at both images and then used to trigger the transition (size of two boxes get similar) and position properly.

Figure 4 – Correspondence setting for forward & backward relation. Corresponding object is marked at both images and then used to trigger the transition (size of two boxes get similar) and position properly.

In case of rotation, the correspondence matching while transition is performed. The rotation correspondence setting at L1.bmp and 1.bmp ensures that the corresponding object is marked at both images. The transition happens when a new view gets outside of view volume of 1.bmp. Then L1.bmp is displayed to match the corresponding object to be shown relatively same position of the scene of 1.bmp.

Figure 5 – Correspondence setting for rotation. Corresponding object is marked at both images. The transition happens when a new view gets outside of view volume of 1.bmp. Then L1.bmp is displayed to match the corresponding object to be shown relatively same position of the scene of 1.bmp.
In case of forward-backward relation, correspondence boxes are used to trigger the transition from scene to scene. In figure 4, the yellow box is smaller than red box at their initial positions. But then viewer navigates by walking inside the scene 1, then yellow box becomes bigger. The graph in figure 4 shows the view volume of each view (blue for initial view of scene 1, red for initial view of scene 2, green for current view). When viewer navigates towards inside of scene 1 (green view volume), the size of two boxes become similar, and it means time to make transition to the next scene (red view volume). When the scene 2 is replacing scene 1, the initial position is translated to match the corresponding objects. The colored boxes are not visible during navigation and used only internally for computation.

In case of rotation, the view volume is checked if the new view is going outside of current scene. If new views going outside of current scene, then EEVT finds the next scene and matches the next scene to the current scene by first, scaling to make the size similar, second, panning the scene to match x axis of two objects, and then translating y axis.

EEVT-maker also has options such as alpha-blending, so scenes can be blended to each other during transition with special effect. This special effect can give viewers a sense of smooth motion and transition to the next scene as effectively used in Pseudo-3D photo collage [Tanaka 2002]. With all these effort, virtual spaces are easily expanded by adding additional snap-shots of photos without special effort of image mosaic or registration.

Developer can leverage default values for automatic processing when more than handful pictures are used to create large virtual tour. Instead of creating individual spidery mesh, EEVT-Maker automatically generates spidery mesh for each picture using default values. The EEVT allow users to navigate the virtual tour as test run and then modify settings interactively as needed.

D. Navigator Guidance

When virtual scene is constructed by connecting more than handful images, then the virtual space become complex structure to navigators’ impression. One of frequent complaints from the Web3D applications such as VRML with complex structures is that users easily get lost in 3D space. Users can come back to original position using reset, but it would be better to give good guidance about navigators trajectory or current location. Therefore, proper user guidance such as mini maps becomes an important feature for pleasant navigation as easily seen in many 3D games like Quake. The EEVT-maker can create a simple map automatically based on the relation between photos and show the map along with view direction and location as well as trajectory during navigation. The map can be toggled in & out as needed by user.

3.2. EEVT-Navigator

The EEVT Navigator is the web-based viewer program, that is, IE & NS plug-in that allows viewers to navigate the virtual scene using their browsers. It provides walk-in and look-around navigation at each scene and smooth transition from one scene to the others as designed at EEVT-Maker. By making use of scene relations specified at EEVT-Maker, the EEVT-Navigator program can provide smart streaming that download the first scene and allow viewers to start navigation while downloading the next scenes according to viewer’s moving direction or orientation (Fig. 6). If a viewer is in scene 1, then scene 2, 5, and 8 are being downloaded while the viewer is navigating the scene 1. When the viewer moves to scene 2, then EEVT-Navigator checks the availability of the new scene. If not available yet, then users forward movement is blocked till it gets available. If the viewer moves to a new scene and currently scene 8 is being downloaded, then scene 3, 9, and 6 are added to the download queue with higher priority than scene 5 (Fig. 6b). If the viewer turns the direction to the right at scene 8 instead scene 2,
then scene 9 and 10 gets higher priority to be downloaded. This enables minimizing the waiting time for entering a new scene.

To give proper feedback to navigator that certain direction is being blocked due to the availability depends on streaming as well as degree of visual distortion, EEVT-viewer shows directional signal as shown at left upper corner at examples in figure 8.

4. RESULTS

We implemented EEVT-Maker and EEVT-Navigator using OpenGL. EEVT-Navigator was initially developed as NS plug-in for Windows version following the Netscape Plug-in Specification [Netscape] and available at http://tlaloc.sfsu.edu/~yoon/WVT. MicroSoft dropped the support for NS Plug-in spec starting from Internet Explorer version 5.5 Service pack 2. To support both IE and NS, we developed IE plug-in using ActiveX and is available at same URL. Unfortunately, IE plug-in is not as compatible to NS plug-in as we expected, so porting from NS plug-in required time-consuming effort. Due to the overhead of maintaining the two separate versions, not to mention supporting different platform (Unix, PC, Mac, Linux, etc), we decided to make use of Java Web Start and JOGL (Java OpenGL Binding) that will run cross over any platform and browsers. We are currently porting into JOGL and Java Web Start version with features explained in section 2, step 3 & 4 and the JOGL version will be available soon.

The SFSU campus virtual tour has 7 photos (Fig.7) in 800x600 resolution. We experimented EEVT-Navigator running on Pentium III 1.0 GHz, 256 Mbyte main memory on the LAN or cable modem (1M bps) environment and the EEVT-Navigator enabled smooth real time navigation (> 5fps) of the whole virtual scenes with the relatively large image size (800x600). New views are captured and presented in figure 8. The results can be demonstrated by running the plugin. However, some of recent approach such as smooth transition using correspondence and map are experimented in EEVT-Maker and being implemented at JOGL version, not at IE or NS version. Also current NS and IE Plug-ins allow viewers to control viewers’ position and orientation using only keyboard (w:forward, s:backward, a:left, d:right, q:up, z:down, and r:reset). Mouse control, buttons and option control (moving speed, alpha blending on/off, and etc) will be added at the JOGL version.

These day, digital camera can capture much higher resolution than 800x600, so we plan to make use of higher resolution by supporting progressive transmission. We are currently implementing wavelet-based approach in JOGL and Java Web Start version that supports compression, progressive transmission and multi resolution all together as used similarly in JPEG 2000 standard [Antonini 92].

5. CONCLUSION

We developed an easy to use (for content developer) and pleasant (for virtual tourists) multiple scene connection to expand virtual space as freely and many as content developers want. EEVT easily generates real time photo-realistic virtual tour from a few clicks of photos, and provides realistic ‘walk-around’ visualization to the users by adding little overhead (less than 1k for each image) for storage and transmission to the original photos. The EEVT also provides guided tour using directional signals to keep viewers from going to undesirable directions, seeing seriously distorted images or being lost. The EEVT plugin and a demonstration are available at http://tlaloc.sfsu.edu/~yoon/WVT.

References


Fig. 7 – original photos taken for constructing SFSU campus virtual tour

Fig. 8 – New views rendered at WVT-Navigator. New views on each column are derived from the same column in figure 7.