Real-Time Interactive Virtual Tour on the World Wide Web (WWW)

Sanghyuk Yoon, Hai-jung Chen, Tom Hsu and Ilmi Yoon
San Francisco State University

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 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, these image-based approaches may require special cameras or effort to take panoramic views and provide only one fixed-point look-around and zooming in-out rather than ‘walk around’, that is a very important feature to provide immersive experience to virtual tourists. The Web-based Virtual Tour using Tour into the Picture employs pseudo 3D geometry with image-based rendering approach to provide viewers with immersive experience of walking around the virtual space with several snap shots of conventional photos.

1. Introduction

The more intensively World Wide Web is used in our daily life, the more strongly people desire for diverse, sophisticated and interactive media types on the WWW. Tourists may desire to plan their trips in advance by selecting tourist attractions by previewing through the WWW. Homebuyers also want to pre-select strong candidate houses for actual visit by inspecting the candidate houses, preferably not by a few still images, but by interactive 3D graphics. Therefore, interactive 3D graphics on the WWW (Web3D) such as Web-based Virtual Tour become more attractive and demanded.

There are two major approaches to provide Web3D (Interactive 3D Graphics on the WWW). One is traditional 3D graphics that transmits 3D geometry with textures to the client for rendering. VRML is an example and a standardized approach by W3C consortium. Even though VRML is a standardized of 3D on the WWW by W3C consortium, it was not as widely accepted and used as HTML does because of the limitations of VRML; VRML doesn’t support streaming or progressive transmission, VRML users complain that they get lost in 3D space due to the difficulty of handling user interfaces, graphics user interface like buttons, slides and other widgets can be added by using java script node, but usually results in unstability issues. VRML uses ASCII while binary supports more compact representation; 3D modeling tools export to VRML format, but usually more than 10 to 100 times bigger than the native format. These limitations naturally leaded to the development of many diverse Web3D formats including MPEG4, Java3D, X3D (the successor of VRML) [Aaron 99], and proprietary formats (e.g., MetaStream [Sree 99]) that solve many of the issues mentioned above. With these new techniques, however, photo-realistic virtual tour is yet challenging due to the nature of web application’s environment such as limited bandwidth and no guarantee of high computation power on the client side.

The other approach to provide Web3D is Image-
based rendering that has attractive advantages over traditional 3D rendering approach. Unlike traditional 3D approach, image-based rendering enables skipping the labor-intensive 3D modeling process of photo-realistic scenes (Fig. 1). As a consequence, the resulting model is much smaller and does not require high bandwidth nor 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 may require special cameras or effort to take panoramic views and provide only one fixed-point look-around and zooming in-out rather than ‘walk around’, 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 also 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, QuicktimeVR or Web-based Virtual Tour (WVT). This technique also requires special tool 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, but the navigation itself is still in 2D and limited to the illusion created by transitions between images, not from actual navigation inside the space where each image is taken.

Tour Into the Picture (TIP) was introduced by Horrry et al. [Horry 97]. By constructing a properly textured pseudo-3D geometry space from a single picture using the spidery mesh, a viewer can tour into the scene (Fig. 2a). 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 originally developed to produce 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.

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. Authors extended TIP by introducing the use of multi-resolution representation of the picture that the visual quality keeps nearly unchanged in the touring. In addition to [Li 2001], the web-based virtual tour needs to employ progressive transmission and effective compression of the multi-resolution image.
resolution images. Therefore, we are currently implementing wavelet based approach that supports compression, progressive transmission and multi resolution all together as used in JPEG 2000 standard [Antonini 92].

We extend the TIP to Web-based Virtual Tour (WVT) that runs on WWW browser as a plug-in. We extended WVT to handle multiple images to expand the virtual space since when navigating a scene generated only from one image, users feel strong desire to navigate further behind the given scene. Virtual spaces are smoothly expanded by adding more photos during the production of the virtual tour and during navigation by receiving images progressively according to the direction of viewers walking and viewing (Fig. 3a). The WVT also provides guided tour using a directional signals to keep viewers from going to undesirable directions, seeing seriously distorted images or being lost (Fig. 3b).

In summary, web-based Virtual Tour using TIP provides a simple (easy to create) and efficient way (no requirement for high computation nor high bandwidth) for ordinary users to create photorealistic virtual tour on the web using only a conventional hand-held camera with a few steps of user interactions (Fig. 1). The user interactions are effectively used to provide viewers with pleasant navigation such as guided tour and smooth transition between scenes.

In chapter 2, we explain TIP briefly and then explains the extensions of production component, WVT-Maker (used by content creators) and viewer component. WVT-Navigator (used by browsers) with focus on multi-image extension and guided tour. Results are presented at chapter 3 and then conclusion and future work are discussed at chapter 4.

2. Web-based Virtual Tour (WVT) Using TIP (Tour Into The Picture) Technique

Web-based Virtual Tour consists of two parts. One is a WVT-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 WVT-Navigator for navigation of the built virtual space on the web.

WVT-Maker

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

1. Specify spidery mesh on each photo and save into a `.tip` file format. In this process, the user interactively specifies the spidery mesh consisting of one vanishing point, four radial lines, and two rectangles over the original picture (Fig. 2a). 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) are constructed with textures derived from the specified regions of the original image. New views can be generated by setting a new view point and rendering these 5 walls with associated texture maps at the view point. When implemented using OpenGL, texture map needs to

In chapter 2, we explain TIP briefly and then explains the extensions of production component, WVT-Maker (used by content creators) and viewer component. WVT-Navigator (used by browsers) with focus on multi-image extension and guided tour. Results are presented at chapter 3 and then conclusion and future work are discussed at chapter 4.

2. Web-based Virtual Tour (WVT) Using TIP (Tour Into The Picture) Technique

Web-based Virtual Tour consists of two parts. One is a WVT-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 WVT-Navigator for navigation of the built virtual space on the web.

WVT-Maker

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

1. Specify spidery mesh on each photo and save into a `.tip` file format. In this process, the user interactively specifies the spidery mesh consisting of one vanishing point, four radial lines, and two rectangles over the original picture (Fig. 2a). 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) are constructed with textures derived from the specified regions of the original image. New views can be generated by setting a new view point and rendering these 5 walls with associated texture maps at the view point. When implemented using OpenGL, texture map needs to

In summary, web-based Virtual Tour using TIP provides a simple (easy to create) and efficient way (no requirement for high computation nor high bandwidth) for ordinary users to create photorealistic virtual tour on the web using only a conventional hand-held camera with a few steps of user interactions (Fig. 1). The user interactions are effectively used to provide viewers with pleasant navigation such as guided tour and smooth transition between scenes.

In chapter 2, we explain TIP briefly and then explains the extensions of production component, WVT-Maker (used by content creators) and viewer component. WVT-Navigator (used by browsers) with focus on multi-image extension and guided tour. Results are presented at chapter 3 and then conclusion and future work are discussed at chapter 4.

2. Web-based Virtual Tour (WVT) Using TIP (Tour Into The Picture) Technique

Web-based Virtual Tour consists of two parts. One is a WVT-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 WVT-Navigator for navigation of the built virtual space on the web.

WVT-Maker

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

1. Specify spidery mesh on each photo and save into a `.tip` file format. In this process, the user interactively specifies the spidery mesh consisting of one vanishing point, four radial lines, and two rectangles over the original picture (Fig. 2a). 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) are constructed with textures derived from the specified regions of the original image. New views can be generated by setting a new view point and rendering these 5 walls with associated texture maps at the view point. When implemented using OpenGL, texture map needs to
handle double perspective since the original texture (taken directly from image) already include perspective component. This additional data 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. 2b).

2. Users also specify the relation to the nearby images, set an option of alpha blending and warping for smooth transition between multiple scenes. The assumption of the content creator of this Web-based Virtual Tour (WVT) is ordinary amateurs with conventional hand-held camera without special image mosaic or registration technique/tools. For example, the content creator walks around a place for the virtual tour taking several snapshots, constructs spidery mesh, saves into tip files as specified in the step 1, and fills in the tip files (scenes) into the scene map matrix where the relative relation between images are simply specified (Fig. 4).

3. After step 2, the content creator can run the WVT with default settings of transitions between multiple scenes. The default settings allow users to move forward if there a forward scene exists [from scene 1 to scene 2, or from scene 2 to scene 3] when viewer reaches two third of current scene’s depth. As discussed in [Li 2001], new views get blurred when viewer approaches to the real wall due to the fixed resolution of the original photo. The default setting lets the transition happen around at two third of the scene’s maximum depth to keep viewers seeing blurred images. This transition point (the depth where the transition happens) can be configured by simple user interaction. This user interaction is simple but result is intuitive and effective because the transition point depends on how the original images are taken and the way how the spidery mesh is created. In a transition to the next scene, the current view fades out and the next scene moves over to the center. This animation gives the user a sense of smooth motion and transition to the next scene as effectively used in [Tanaka 2002]. Therefore, without special effort of image mosaic or registration, virtual spaces are easily expanded by adding more snap-shots of photos.

4. During the test run, content creators are not only allowed to specify the transition points (depth value for forward transition) but also the angles for left and right transitions between images [from scene 1 to scene 4, or from scene 1 to 5 or 5 to 6 in figure 4]. We can use these angle as a trigger for transition to the near by scenes and at the same time as a sign for blocking users orientation turning movement to keep users from seeing distorted images or blank image when near-by scene is not available. As shown in the figure 5, the permissible angle to turn the orientation increases when user walks into the scene. Therefore, the WVT-maker lets user set the minimum permissible angle to turn at the starting point and the maximum permissible angle at right before the forward transition point. Usually viewer should not turn the orientation at the starting point, so minimum is zero and creator needs to set only maximum in case the default setting is not satisfactory. WVT-navigator computes the permissible angle at any depth using linear interpolation between two angles.

![Fig. 5 – Viewer’s turning orientation can be controlled by WVT-Navigator by the value entered by content creator during production. Permissible angle (α < β) for viewer’s turning orientation increases as viewer walks into the scene.](image)

This user interaction works effectively since the content creator’s understanding of the whole scenes makes most intuitive transition and blocking decision more than any other computation for conventional snap shots. In addition, alpha blending can be set on/off to make the transition effect even smoother.

**WVT-Navigator**

The WVT Navigator is the 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 WVT-Maker. This WVT-Navigator program also provides smart streaming of the next scenes according to viewer’s moving direction or orientation while the first image is being navigated (Fig. 6). As shown at figure 4b, 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 WVT-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 is 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.

![Download Queue Diagram](image)

**Fig. 6 – (a)(b)(c) shows the download queue when viewer's position moves from camera 1 to camera 2 or camera 3. Green cell represents the scene is downloaded and pink cell means the scene is currently being downloaded.**

When virtual scene is constructed by connecting more than several images, then the virtual space start to build complex structure. One of majority complaints from the Web3D applications such as VRML with complex structures is that users easily get lost in 3D space. 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 WVT currently provides guided tour using a direction signals to keep viewers from going to undesirable directions or being lost (Fig. 2b).

### 3. IMPLEMENTATION AND RESULTS

We implemented WVT-Maker and WVT-Navigator. WVT-Navigator, IE and NS plug-in for Windows version is available at [http://tlaloc.sfsu.edu/~yoon/WVT](http://tlaloc.sfsu.edu/~yoon/WVT). The overall procedure for production and publishing is (1) content creator takes photos, (2) create spidery mesh using WVT-Maker for each photo, (3) save into each tip files, and (4) create a wvt file that associates all the tip files and user’s configurations [transition depth and angle]. All these files can be published as if they are html files. (5) It only needs to update wvt file format in the mime.type in the web server, so browser will recognize the format to launch the plugin. All computation for smart downloading is done in the client side without any support from the server side that substantially reduces the server side workload.

The SFSU campus virtual tour has 7 photos (Fig.7) in 800x600 resolution. We experimented WVT-Navigator running on Pentium III 1.0 GHz, 256 Mbyte main memory on the LAN or cable modem (1M bps) environment and the WVT-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 most effectively by running the Plug-in.

The current version of plug-in demonstrates all the functionality explained in chapter 2. A few ongoing implementations will improve issues of (1) current version only works with NS and IE 5.5 (Service Pck version 1) and lower versions due to recent changes of the support for Plug-ins compatibility from IE. (2) Currently viewers can

![Original Photos](image)

**Fig. 7 – original photos taken for constructing SFSU campus virtual tour**
control viewers’ position and orientation using keyboard (w:forward, s:backward, a:left, d:right, q:up, z:down, and r:reset) at WVT-Navigator. Mouse control, buttons and option control (moving speed, alpha blending on/off, and etc) will be added in near future for better user interface.

We also observed that image quality quickly degrades when approaching towards the rear wall. The problem can be solved by employing multi-resolution images as discussed in [Li 2001]. In addition to [Li 2001], the web-based virtual tour needs to employ progressive transmission and effective compression of the multi resolution images. For example, 800x600 images can be transmitted and rendered in much lower resolution (200x150), and then progressively refined to the highest resolution while being navigated. We are currently implementing wavelet based approach that supports compression, progressive transmission and multi resolution all together as used in JPEG 2000 standard [Antonini 92].

4. CONCLUSION and On-going Research

Web-based Virtual Tour 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. It provides

Fig. 8 – New views rendered at WVT-Navigator. New views on each column are derived from the same column in figure 7.
guided tour and easily expands the virtual space by adding more photos. The WVT plugin and a demonstration are available at http://tlaloc.sfsu.edu/~yoon/WVT.

Currently, we are extending the current WVT Maker that lets content creators create the dimension of virtual space matrix and drag and drop photos into the cells of the matrix. And then WVT Maker automatically creates virtual walls using default settings. WVT Maker also automatically creates transition point between virtual scenes. With all these default setting and automatic creations, very large virtual tour can be easily created as a test run. The test run can be easily modified to be the final tour by allowing users to modify each setting as desired.

References