A Framework for Interactive Virtual Tours

Frank Edughom Ekar

Abstract—This paper introduces a framework for the creation, management and deployment of interactive virtual tours. A panoramic image acquisition unit is adapted to acquire panoramic images and video streams from a wide variety of sources. The acquired panoramic images and video streams are fed into a transform engine that performs any required transformations on the input data in a uniform and seamless manner. A package generator coupled with the transform engine synthesizes complete virtual tour packages by combining the transformed data with a broad range of multimedia resources to create navigable virtual tours. Perspective correction and related transformations on the data are carried out by a viewing engine allowing a plurality of viewers to independently and simultaneously interact with the virtual tours. User interaction with the virtual tour packages is enhanced by a control engine that facilitates bi-directional communication between various elements of the system.

Index Terms—Interactive Virtual Tour, Framework, Package Synthesis, Bi-directional Communication, User Interaction.

I. INTRODUCTION

The recent proliferation of affordable and powerful personal computing devices in the form of the ubiquitous personal computer and associated hardware and software, personal imaging devices in the form of digital still image and video cameras and associated tools and techniques, and accessible personal communications in the form of the Internet and related platforms has generated strong and growing interest in the creation and use of virtual tours.

A virtual tour is typically an environment containing representations of remote locations and permitting one or more users to interact with the environment in a manner that gives the user the perception of being right there (immersed) within the remote location. The representations could contain still images, video, audio, navigational databases or any other type of data that is required to facilitate meaningful user interaction. Virtual tours have found significant applications in areas such as medicine, remote surveillance, telepresence, forensics, defense, flight simulators, navigation support systems, tourism, and personal and professional websites.

A very important component of the data comprising a virtual tour is visual information in the form of still images, information is usually represented as environment maps in virtual tours. In this context, an environment map refers to a specification of visual information for a set of locations in a

scene such as that contained in a digitized image, video or computer-generated rendering of the scene.

When creating environment maps for use in virtual tours, it is desirable to utilize a system that encodes as much visual information as possible in a single image frame and that preferably does so cost-effectively and without any moving parts. The limited field of view offered by conventional cameras and lenses make them unsuitable for this task. Consequently, panoramic imaging systems capable of much wider fields of view are used to create environment maps for use in virtual tours. Panoramic imaging systems, however, are generally characterized by various types of distortions that need to be corrected before the environment maps generated using them can be viewed comfortably by a human observer. In particular, perspective distortions must be corrected. Furthermore, the value offered by a virtual tour system is closely related to the degree of interactivity permitted by the system.

Accordingly, this paper introduces a virtual tour system featuring a unique control unit or control engine that facilitates bi-directional communication between various elements of the system – permitting an unprecedented degree of interactivity.

The remainder of this paper is organized as follows: Section II is an overview of relevant literature and related work providing a review of panoramic imaging systems that approximate the ideal system for the capture of data for virtual tours. In Section III, the problem definition, methodology/approach in the form of an outline highlighting some key components of the virtual tour system introduced in this paper is presented. Results and brief discussions of underlying issues and future scope appear in Sections IV-VII. Section IV describes the image acquisition unit of the virtual tour system. A discussion of the transform unit and package generator appears in Section V. In Section VI, aspects of the viewing engine are described. Section VII highlights the features of the control unit while Section VIII concludes this paper.

II. LITERATURE REVIEW

Panoramas have been in use as early as the 18th century [1]. Some of the earliest panoramic imaging techniques involved the manual stitching of overlapping image segments acquired with conventional photographic devices. Creating panoramas in this fashion was tedious and resulted in visible seams between image segments. More recently, rotating cameras have been used to build panoramic mosaics [2][3]. Such systems are constrained to work effectively only in static environments. One remedy to this constraint involves using an orientation sensing device to sense the orientation of the user and then using the input to rotate a conventional camera to cover the field of view indicated by

Published on December 23, 2019.

Frank Edughom Ekar is with the Admiralty University of Nigeria, Nigeria.

(e-mail: frankekparng@gmail.com).

the user's orientation [4]. Apart from the obvious handicap of not being able to permit more than one user to comfortably navigate the remote scene, the inherent delay between the transmission of the orientation signal and the positioning of the camera to coincide with the desired orientation severely limits the practical value of this system. Wide-angle lenses such as the fish-eye lens were later used to acquire a substantially larger field of view without the constraint that the environment be static. Such wide-angle lenses, however, introduce significant non-linear distortions, some of which are hard to model, into the images they generate. Aggarwal demonstrated techniques for the calibration of systems based on fish-eye lenses. Frank et al. [6] introduced robust methods for correcting distortions in arbitrary panoramic imaging systems.

The construction of arbitrary perspective-corrected views of panoramic images requires that the panoramic imaging system possess a unique effective viewpoint. Fouch et al. [7] built a dodecahedral mirror system with the faces of the mirrors directed at twelve separate cameras such that the entire assembly meets the unique effective viewpoint constraint. This system, while eliminating the need to have a static environment and to correct distortions in the images, is limited by its complexity and cost. More robust solutions to the problem involve the use of catadioptric systems consisting of reflecting and refracting elements. Greguss Pal [8] proposed a compact system composed of several reflecting and refracting components and satisfying the unique center of projection constraint. Other catadioptric systems relying primarily on their reflective elements (mirrors) include the hyperbolic mirror system used by Once et al. [9], with which they studied the accuracy of visualization versus warping time in telepresence applications. Catadioptric systems have also been applied to the problems of motion, localization and map building and in robot navigation [10-15]. Extensive studies on the properties of catadioptric systems have been carried out by Geyer et al. [16] and Drucker [17]. A review of the state of the art in computer-aided panoramic imaging has been published by Yagi [18].

Frank et al. [19] applied neural networks to the construction of arbitrary perspective-corrected views from panoramic images. Hase et al. studied the precision of 3D measurements on pairs of stereoscopic panoramic images captured simultaneously using two cameras equipped with catadioptric panoramic imaging systems. Contemporary panoramic imaging systems are incapable of generating a substantially spherical environment map in a single image frame - relying instead on the expensive, error-prone and labor-intensive stitching of multiple conventional or wide-angle image segments to do so. Although catadioptric systems such as that introduced by Greguss Pal [8] are capable of capturing a panorama in a single image frame, their vertical field of view is substantially less than 180 degrees - making the generation of a spherical environment map from the single image frame impractical. Furthermore, catadioptric systems typically utilize a plurality of reflecting surfaces that reflect light from the scene onto focusing devices (so-called relay lenses) that ultimately focus the image of the scene onto a photosensitive surface such as photographic film or a CCD or CMOS array. This structure reduces the quality of the images from such systems and causes astigmatism and other aberrations. Additionally, the compression of all the visual data in the scene onto a conventional single photosensitive surface further degrades the

quality of the images, especially around the edges. The most cost-effective contemporary solution for generating spherical environment maps involves the stitching or blending of two sequentially captured opposing fish-eye hemispheres. The use of more than one sequentially captured image segment introduces a temporal discontinuity in the resulting spherical environment map. Consequently, this technique is effective only in relatively static environments — severely limiting its practical value. An alternative based on fish-eye or similar wide-angle lenses uses two separate cameras to simultaneously capture the opposing hemispheres. This eliminates the temporal discontinuity but introduces spatial discontinuities since the two image segments no longer share a common effective viewpoint. In any case, the expensive, error-prone and labor-intensive stitching process is still required for the generation of substantially spherical environment maps. In addition to the aforementioned limitations, contemporary panoramic imaging systems are incapable of creating realistic 3-dimensional spherical environment maps with depth information from a single imaging device or camera.

III. PROBLEM DEFINITION/METHODOLOGY/APPROACH: SYSTEM OUTLINE

As illustrated in Fig. 1 which depicts a simple block diagram of the key components of the virtual tour system in this paper, a panoramic image acquisition unit is adapted to acquire panoramic images and video streams from a wide variety of sources.

Fig. 1: Simple block diagram of virtual tour system.

The acquired panoramic images and video streams are fed into a transform engine that performs any required transformations on the input data in a uniform and seamless manner. A package generator coupled with the transform engine synthesizes complete virtual tour packages by combining the transformed data with a broad range of multimedia resources to create navigable virtual tours. Perspective correction and related transformations on the data are carried out by a viewing engine allowing a plurality of viewers to independently and simultaneously interact with the virtual tours. User interaction with the virtual tour packages is enhanced by a control engine that facilitates bi-directional communication between various elements of the system.

IV. PANORAMIC IMAGE ACQUISITION UNIT: RESULTS AND DISCUSSION

Ideally, the panoramic image acquisition unit would be a versatile imaging system or device for creating representations of stimuli covering substantially all directions or only a subset of directions around a given reference or view point, comprising at least one grid of one or more focusing elements disposed on an N-dimensional and arbitrarily shaped surface, at least one grid of one or more sensor elements disposed on an N-dimensional and arbitrarily shaped surface, and optionally, at least one grid of one or more stimulus guide elements disposed on an N-dimensional and arbitrarily shaped surface, where N can be chosen to be 1, 2, 3, or any other suitable quantity. As explained in United States Patent Number 7567274 [21], a wide variety of panoramic imaging systems could be used to approximate the features of the versatile device. One such approximation is the catadioptric system outlined in Fig. 2. Additionally, the systems described in [2], [3], [4], [7] and [9] could be used to approximate the features of the versatile device.

Fig. 2: Image formation by a catadioptric system.

The system depicted in Fig. 2 is sometimes referred to as a Panoramic Annular Lens (PAL) and the images it produces as PAL images. The lateral field of view of the catadioptric system illustrated in Fig. 2 is the entire 360-degree range. The maximum angle above the horizon is denoted by

ϕ_A (less than 90 degrees in most practical cases) while the minimum angle below the horizon is denoted by

ϕ_B . Both ϕ_A and ϕ_B usually have values less than

90 degrees in practical catadioptric systems. Points in the real world object space seen at the same vertical angles above or below the horizon correspond to concentric circles in the image. Similarly, points on lines in the real world object space parallel to the optical axis of the catadioptric system are projected onto radial lines in the image. Consequently, catadioptric systems with the profile in Fig. 2 generate annular images similar to that shown in Fig. 3.

Fig. 3: Annular image generated by a catadioptric system.

V. TRANSFORM ENGINE AND PACKAGE GENERATOR: FURTHER RESULTS AND DISCUSSION

For simplicity, the virtual tour system is designed to transform input image data into spherical environment maps. The transform engine is typically implemented as a set of algorithms running on a computer but other implementations could be used. The intimate goal of the transform engine is the creation of a suitable representation of the spherical environment map from the panoramic image acquisition unit and the removal of any distortions that may be present in the data. The transform engine could, when required, transform the spherical environment maps into rectilinear panorama format by using a spherical to rectilinear coordinate transformation. Suppose we have as input data an image similar to the annular image depicted in Fig. 3.

Since the interior of the image contains no useful image data, it is often necessary (for example, for transmission over bandwidth-limited media like the Internet) to extract the useful 360-degree panorama from the image using a polar to rectilinear coordinate transformation. For example, for the point with rectilinear coordinates x and y on the 360-degree panorama, the corresponding point with rectilinear coordinates x' and y' on the source PAL image can be obtained by first calculating the corresponding polar coordinates θ' and r' and then using polar to rectilinear coordinate transformation to calculate the x' and y' coordinates as illustrated in Fig. 4, Fig. 5 and by Equations (1)-(4). R is the outer radius while r is the inner radius of the PAL image.

$$\theta' = \frac{2\pi}{P_w} x \quad (1)$$

$$r' = R - \left(\frac{R-r}{P_H} \right) y \quad (2)$$

$$x' = x_{center} + r' \cos \theta' \quad (3)$$

$$y' = y_{center} + r' \sin \theta' \quad (4)$$

, where P_w is the width and P_H is the height, respectively, of the 360-degree panorama, and x_{center} and y_{center} are the rectilinear coordinates of the center of the source PAL image as shown in Fig. 4. The resultant 360-degree panorama for the source annular image of Fig. 3 is shown in Fig. 6.

Fig. 4: Point on annular image expressed in both polar and rectilinear coordinates for the purpose of conversion into a 360-degree panorama.

Fig. 5: 360-degree panorama generated from annular image.

The point with coordinates x and y on the 360-degree panorama corresponds to the point with coordinates x' and y' on the source annular image as illustrated in Fig. 4.

Fig. 6: Resultant 360-degree panorama for the source annular image of Fig. 3.

The transform engine also implements algorithms for rectifying any distortions introduced into the spherical environment maps by the use of non-ideal focusing units, use of photosensitive and/or focusing element grids of insufficient density (number of elements), a geometry that differs from a spherical geometry or the misalignment of the photosensitive and focusing surfaces before the data is made available for further processing. The transform engine also implements means of accepting ready-made environment maps in spherical, cylindrical, rectilinear or other format and performs any coordinate transformations needed to convert same to spherical environment maps in any desired output format. The major function of the package generator is the creation of interactive virtual tour packages from one or several spherical environment maps created by the transform engine. The package generator and the transform engine can be combined into a single virtual tour authoring unit, implemented in software, and providing a means of storing, retrieving and editing the virtual tour packages. The virtual tour package can be arranged as a series of interconnected or linked spherical environment maps with information describing how the individual spherical environment maps are interconnected or linked and additional information specifying multimedia content (audio, video, hotspots and so on) to be rendered in response to the activation of interactively defined regions of the spherical environment maps.

VI. VIEWING ENGINE: FURTHER RESULTS AND DISCUSSION

Once the virtual tour package has been created, it can be exported to the viewing engine. Spherical environment maps not requiring transformation could be exported directly into the viewing engine from the spherical image/video acquisition unit. The viewing engine can be implemented in the same computer software as the virtual tour authoring unit or separately as a stand-alone software or hardware component.

Fig. 7: Representation of 360-degree panorama as spherical environment map and selection of a region of interest on the resultant spherical environment map.

Fig. 8: Projection of a point on a spherical environment map onto a perspective-corrected point on the object plane.

As explained in Section V, input images are first re-projected onto a spherical surface to create a spherical environment map. The projection or re-projection of a panoramic image onto the surface of a sphere Fig. 5 is illustrated conceptually in Fig. 7. Portions of the sphere for which panoramic image data is not available (for example, when the panorama is acquired using a catadioptric system such as that depicted in Fig. 2 with a vertical field of view that is less than π radians) can be replaced with user-supplied data or simply filled with a uniform color. The viewing engine provides a means for the user to interactively select a particular spherical environment map from the virtual tour package and specify viewing parameters such as a lateral angle θ , a vertical (elevation or azimuth) angle φ and a magnification coefficient. These view parameters together with a view window onto which perspective-corrected views of the spherical environment map are rendered specify a region of interest on the spherical environment map that is projected onto a perspective-corrected object plane as shown in Fig. 7 and Fig. 8 in which the spherical environment map is represented in 3-dimensional X-Y-Z space. The

magnification coefficient is directly related to the angular span of the width of the perspective-corrected view window as measured from the center, O , of the sphere. When the viewing engine is implemented in computer software, the viewing parameters can be entered using the mouse, keyboard, joystick or any other suitable input device.

The perspective transformation is achieved by projecting the selected region of interest on the surface of the sphere onto the perspective-corrected object plane. Equations (5) through (12) describe the transformations required to create the perspective-corrected view. In equations (5) through (12) as in Fig. 7 and Fig. 8, u and v are the coordinates of a point on the perspective-corrected U-V object plane (the point $P(u, v)$ in Figure 7), θ and φ are the corresponding lateral and azimuth angles on the spherical environment map for the viewing parameters

θ_0 and φ_0 , and the magnification coefficient-related angular span λ ; d is the distance of the center of the object plane from the center, O , of the sphere while W and H are, respectively, the width and height of the perspective-corrected view window on the U-V object plane.

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_0 & \sin \varphi_0 \\ 0 & -\sin \varphi_0 & \cos \varphi_0 \end{bmatrix} \quad (5)$$

$$B = \begin{bmatrix} \cos \theta_0 & 0 & -\sin \theta_0 \\ 0 & 1 & 0 \\ \sin \theta_0 & 0 & \cos \theta_0 \end{bmatrix} \quad (6)$$

$$C = AB \quad (7)$$

$$d = \frac{W}{2 \tan \frac{\lambda}{2}} \quad (8)$$

$$D = [u \quad v \quad d] \quad (9)$$

$$E = DC = [\mathcal{E}_1 \quad \mathcal{E}_2 \quad \mathcal{E}_3] \quad (10)$$

$$\theta = a \tan 2(\mathcal{E}_1, \mathcal{E}_3) \quad (11)$$

$$\varphi = a \tan 2\left(\mathcal{E}_2, \sqrt{\mathcal{E}_1^2 + \mathcal{E}_3^2}\right) \quad (12)$$

In equations (11) and (12), $\text{atan2}(\text{argument1}, \text{argument2})$ is a trigonometric function that gives the arctangent of $\text{argument1}/\text{argument2}$. Once the

θ and φ angles have been computed using equations

(5) through (12), the image/video data contained in the spherical environment map at the locations corresponding to the θ and φ angles is then copied

onto the locations indicated by the u and v coordinates on the perspective-corrected view window on the U-V object plane - possibly using interpolation to account for non-integral values of the parameters and to improve the quality of the resultant image/video. Compared with the approaches used by the contemporary systems, equations (5) through (12) provide a faster means of constructing perspective-corrected views from spherical environment maps while providing full pan, tilt and zoom controls. Creating a look-up table containing a subset of points from the perspective-corrected window, using equations (5) through (12) on the points in the look-up table only and then using bilinear or other suitable forms of interpolation to correct all other points not in the look-up table leads to faster albeit less accurate perspective correction. It is also possible to use a table of pre-computed trigonometric values to speed up the calculations without any appreciable loss in image/video quality. The perspective-corrected view generated by the viewing engine is then displayed on a suitable display device such as a computer monitor or head-mounted display. Sample results are shown in Fig. 9 and Fig. 10.

Fig. 9: Portion of 360-degree panoramic image with visible perspective distortion.

Fig. 10: Image of Fig. 9 after perspective correction.

VII. CONTROL UNIT: FURTHER RESULTS AND DISCUSSION

As a means of permitting a higher level of interactivity with the virtual tour package, the control engine is connected operatively to the viewing engine and provides additional means of interaction with the virtual tour package. In particular, the control engine permits any spherical environment map to be selected from the virtual tour package and sends signals to the viewing engine that cause the viewing engine to permit the interactive navigation of the selected spherical environment map.

Fig. 11: One possible implementation of the control engine.

The control engine also indicates which spherical environment map is currently selected and what portion of the selected spherical environment map is currently displayed in the perspective-corrected view window. Furthermore, the control engine permits the interactive selection of any portion of any spherical environment map in the virtual tour package for viewing and transmits the corresponding viewing parameters to the viewing engine- causing the viewing engine to render the indicated view. The control engine and the viewing engine communicate bi-directionally. The control engine could represent spherical environment maps in preview form. That is, it could provide enough information for adequate identification of any spherical environment map without necessarily storing the entire spherical environment map itself. Accordingly, the control engine is preferably implemented as a set of software components displaying in 3- or 2-dimensional space previews (such as thumbnail images) of the spherical environment maps in the virtual tour package. This is illustrated conceptually in Fig. 11 in which P1, P2, P3,...Pn are preview elements. The preview data for the control engine is created by the package generator or the virtual tour authoring unit comprising both the transform engine and the package generator. More specifically, the control engine could be implemented as a set of software components each displaying a thumbnail image or video of a spherical environment map or a floor plan or map (with active regions defined and possibly incorporating global positioning system (GPS) and other useful

data) and representing the scene contained in the virtual tour package. Input (via the mouse, keyboard, or other suitable means) from the user can be used to select any spherical environment map for viewing by the viewing engine and then signals from the viewing engine can be used to indicate the selected spherical environment map and what portion of the selected spherical environment map is currently viewed. Fig. 12 illustrates.

Fig. 12: Interactive virtual tour with map and preview thumbnails.

VIII. CONCLUSION AND FUTURE SCOPE

This paper introduced a system for creating and interactively navigating virtual tours using a wide variety of panoramic image sources that could be transformed into suitable spherical environment maps. Furthermore, a unique control unit or control engine that facilitates bi-directional communication between various elements of the system - permitting an unprecedented degree of interactivity was introduced. Present and future applications of the virtual tours include medicine, remote surveillance, telepresence, forensics, defense, flight simulators, navigation support systems, tourism, personal and professional websites.

REFERENCES

- [1] R. Benosman, "Panoramic imaging from 1767 to the present," *Proc. International Conference on Advanced Robotics, Workshop*, 1, 9-10(2001).
- [2] R. Szeliski, "Video mosaics for virtual environments," *Computer Graphics and Applications*, 16(3), 23-30(1996).
- [3] H. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment," *Proc. Fifth European Conference on Computer Vision*, 2, 103-119 (1998).
- [4] T. Boulton, "Remote reality demonstration," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 966-967 (1998).
- [5] S. Shah, and J. Aggarwal, "Intrinsic parameter calibration procedure for a (high-distortion) fish-eye lens camera with distortion model and accuracy estimation," *Pattern Recognition*, 29(11), 1775-1788 (1996).
- [6] F. Ekpar, H. Hase, and M. Yoneda, "Correcting Distortions in Panoramic Images using Constructive Neural Networks," *International Journal of Neural Systems*, 13(4), 239-250 (2003).
- [7] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Steam, and H. Fouch, "The office of the future: a unified approach to image-based modeling and spatially immersive displays," *ACM SIGGRAPH*, 179-188 (1998).
- [8] P. Greguss, "PAL-optic based instruments for space research and robotics," *Laser and Optoelectronics*, 28, 43-49(1996).
- [9] Y. Once, K. Yamazawa, H. Takemura, and N. Yokoya, "Telepresence by real-time view-dependent image generation from omni-directional video streams," *Computer Vision and Image Understanding*, 71(2), 154-165(1998).
- [10] C. Fermuller, Y. Aloimonos, and T. Brodsky, "New eyes for building models from video," *Computational Geometry: Theory and Applications*, 15, 3-23 (2000).
- [11] Y. Yagi, S. Kawato, and S. Tsuji, "Real-time omni directional image sensor (copis) for vision-guided navigation," *IEEE Transactions on Robotics and Automation*, 10(1), 11-22 (1994).
- [12] Y. Yagi, S. Nishizawa, and S. Tsuji, "Map-based navigation for a mobile robot with omni-directional image sensor," *IEEE Transactions on Robotics and Automation*, 11(5), 634-648(1995).
- [13] J. Zheng, and S. Tsuji, "Panoramic representation for route recognition by a mobile robot," *International Journal of Computer Vision*, 9(1), 55-76(1992).
- [14] J. Chahi, and M. Srinivasan, "Reflective surfaces for panoramic imaging," *Applied Optics*, 36, 8275-8285 (1997).
- [15] A. Hicks, and R. Bajcsy, "Reflective surfaces as computational sensors," *Image and Vision Computing*, 19(11), 773-777(2001).
- [16] C. Geyer, and K. Daniilidis, "Catadioptric projective geometry," *Proc. International Conference on Advanced Robotics*, 1, 17-30, (2001).
- [17] D. Drucker, and P. Locke, "A natural classification of curves and surfaces with reflection properties," *Mathematics Magazine*, 69, 249-256(1996).
- [18] Y. Yagi, "Omni-directional sensing and its application," *IEICE Transactions on Information & Systems*, E82-D(3), 568-579(1999).
- [19] F. Ekpar, H. Hase, and M. Yoneda, "Constructing arbitrary perspective-corrected views from panoramic images using neural networks," *Proc. 7th International Conference on Neural Information Processing*, 1, 156-160 (2000).
- [20] H. Hase, H. Kawai, F. Ekpar, M. Yoneda, and J. Kato, "On the Precision of 3D Measurement by Stereo PAL Images," *Transactions of the IEICE*, J86-A (4), 474-481 (2003).
- [21] F. Ekpar, "Method and apparatus for creating interactive virtual tours," United States Patent Number 7567274, July 28, 2009.