Morphable 3D-Mosaics: a Hybrid Framework for Photorealistic Walkthroughs of Large Natural Environments

Nikos Komodakis and Georgios Tziritas
Computer Science Department, University of Crete
E-mails: {komod, tziritas}@csd.uoc.gr

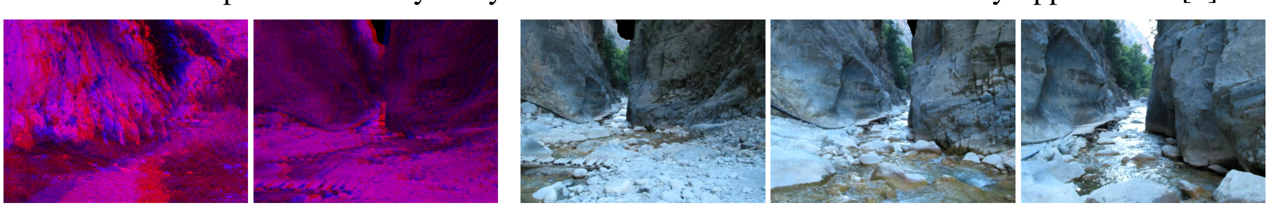
Abstract

This paper presents a hybrid (geometry- & image-based) framework suitable for providing photorealistic walkthroughs of large, complex outdoor scenes at interactive frame rates. To this end, based just on a sparse set of real stereoscopic views from the scene, a set of *morphable 3D-mosaics* is automatically constructed first, and then, during rendering, a continuous morphing between those 3D-mosaics that are nearby to the current viewpoint is taking place. The morphing is both photometric, as well as geometric, while we also ensure that it proceeds in a physically valid manner, thus remaining transparent to the user. The effectiveness of our framework has been demonstrated in the 3D visual reconstruction of the Samaria Gorge in Crete, which is one of the largest and most beautiful gorges in Europe.

I. Introduction

One research problem of computer graphics that has attracted a lot of attention recently is the creation of modeling and rendering systems capable to provide photorealistic & interactive walkthroughs of complex, real-world environments. Two are the main approaches proposed so far for this purpose. On one hand (according to the purely geometric approach), a full 3D geometric model of the environment is constructed first, and that model is then used during rendering for visualizing the input scene. On the other hand, image-based rendering (IBR) methods skip the geometric modeling part completely, and instead attempt to create novel views simply by properly resampling a given set of images from the scene. Although IBR techniques can attain higher level of photorealism (compared to geometric methods), they, however, require a large amount of input data and thus their usage is typically restricted to scenes of small size.

So, although a lot of research has been done regarding small scale scenes, up to now only few works have been presented that can deal efficiently with the visual reconstruction of large scale environments. To address this issue, the work presented in this paper describes a novel hybrid (geometry- & image-based) system, capable of visually reconstructing large-scale, complex outdoor environments. We note that the scene reconstruction provided by our system is always photorealistic. Moreover, our method is capable of visualizing this photorealistic reconstruction at interactive frame rates, i.e. in real time. To this end, one of the main contributions of this work is the proposal of a new compact data representation for a 3D scene, consisting of a set of morphable (both geometrically and photometrically) 3D models that are constructed automatically based just on a sparse set of real stereoscopic views. We also note that, due to the above mentioned properties, our system can be extremely useful in applications such as 3DTV, gaming etc., where one seeks to create a 3D depth impression of the observed scene. E.g., by being able to provide photorealistic walkthroughs of real world environments at interactive frame rates, our system can be used in virtual reality applications. In fact, this has been one of the starting motivations, and so our system was already used as part of a 3D virtual reality installation in the Natural History Museum of Crete. In this case, the ultimate goal was to provide a lifelike virtual tour of the Samaria Gorge (i.e. one of the largest and most magnificent gorges in Europe) to all visitors of the museum. To this end, the most beautiful spots along the gorge were selected, which were then visually reconstructed by our system. A pair of circular polarized LCD projectors has been used for the stereoscopic projection, and so the museum visitors were able to participate in the interactive virtual tour simply by wearing stereo glasses that were matched to the circular polarization of the projectors. Two sample stereoscopic views, as would have been rendered by the VR hardware, are shown in Figure 1(a). A corresponding video [1], containing an overview of our system as well as a short clip from a virtual tour into the Samaria Gorge, is also available at the following URL: `http://www.csd.uoc.gr/~komod/research/morphable_3d_mosaics/`. We also note that a brief description for a very early version of our framework has already appeared in [2].



(a) (b)

Fig. 1: (a) Two stereoscopic views as would have been rendered by the VR system (for illustration, these are shown in the form of red-blue images). **(b)** Also, a few non-stereoscopic views are shown, which were rendered by our system as the virtual camera traverses the so-called "Iron Gates" area, which is the most famous part of the Samaria gorge.

II. Related work

An example of another hybrid (i.e. geometry- and image-based) system is the work of Debevec et al. [3]. However, this approach is mostly suitable for architectural type scenes, and also requires that a basic geometric model of the whole scene has already been recovered by the user interactively. In [4], an image-based technique is proposed by which an end-user can create walkthroughs from a sequence of photographs, but the resulting renderings are not of high quality and may contain considerable artifacts and distortion. Also, in "plenoptic modeling" [5] a warp operation is introduced that maps panoramic images (along with disparity) to any desired view. However, this operation is not very suitable for use in modern 3D graphics hardware. Lightfield [6] and Lumigraph [7] are two popular image-based rendering methods, but they require a large number of input images and so they are mainly used for small scale scenes. To address this issue, work on unstructured/sparse lumigraphs has been proposed by various authors. One such example is the work of Buehler et al. [8]. However, in that work, a fixed geometric proxy (which supposedly describes the global geometry of the scene at any time instance) is being assumed, an assumption that is not adequate for the case of 3D data coming from a sequence of sparse stereoscopic views. This is in contrast to our work, where view-dependent geometry is being used due to the continuous geometric morphing that is taking place. Another example of a sparse lumigraph is the work of Schirmacher et al. [9]. Although they allow the use of multiple depth maps, any possible inconsistencies between them are not taken into account during rendering. This is again in contrast to our work, where an "optical flow" between wide-baseline images is estimated to deal with this issue. Furthermore, this estimation of "optical flow" between wide baseline images significantly reduces the required number of views. For these reasons, if any of the above two approaches were to be applied to large-scale scenes, like those handled in our case, many more images (than ours) would then be needed. In addition, due to our rendering path, which can be highly optimized in modern graphics hardware, we can achieve very high frame rates during rendering, whereas the corresponding frame rates listed, e.g., in [9] are much lower due to an expensive barycentric coordinate computation that needs to take place in this case. In the "Interactive visual tours" approach [10], video (from multiple cameras) is being recorded as one moves along predefined paths inside a real world environment, and then image based rendering techniques are used for replaying the tour and allowing the user to move along those paths. This way, virtual walkthroughs of large scenes can be generated. However, a specialized acquisition system is needed in this case, whereas our method requires using just off the shelf digital cameras. Finally, in the "sea of images" approach [11], a set of omnidirectional

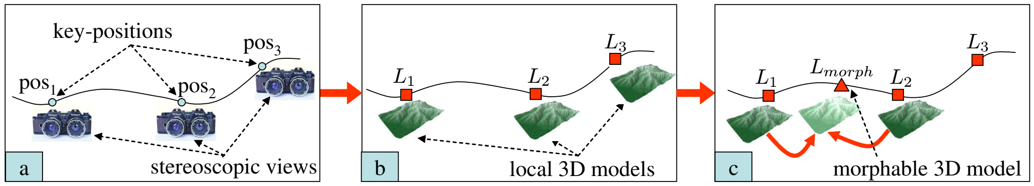


Fig. 2: (a) A sparse set of stereoscopic views is captured at key-positions along the path **(b)** One local 3D model is constructed out of each stereoscopic view **(c)** As the user traverses the path, a morphable 3D model is displayed during rendering. This way, a continuous morphing between successive local models takes place at any time, with this morphing being both photometric as well as geometric.

images are captured for creating interactive walkthroughs of large, indoor environments. However, this set of images needs to be very dense with an image spacing of about 1.5 inches.

III. Overview of the morphable 3D-mosaics framework

As already mentioned, our system is capable of providing photorealistic walkthroughs of large scale scenes at interactive frame rates. So assuming, for simplicity, that during the walkthrough the user motion takes place along a predefined path of the environment[1], the input to our system is then a sparse set of stereoscopic views captured at certain locations (called *"key-positions"* hereafter) along that path (see Figure 2a). Considering initially the case where only one view per key-position exists, a series of *local 3D models* are then constructed, one for each stereoscopic view, with these local models capturing the photometric and geometric properties of the scene at a local level, and containing just an approximate representation of the scene's geometry (see Figure 2b). Then, instead of trying to create a global 3D model out of all these local models (a task that can prove to be extremely difficult in many cases, requiring a very accurate registration between local models), we rather follow a different approach. The key idea is that during the transition between any two successive key-positions (say pos_1, pos_2) along the path, a *"morphable 3D-model"* L_{morph} is displayed by the rendering process (see Figure 2c). At position pos_1 this model coincides with the local model L_1 at that position, while as we are approaching pos_2 it is gradually transformed into the next local model L_2, coinciding with that upon reaching key-position pos_2. Therefore, during the rendering process, and as the user traverses the predefined path, a continuous morphing between successive local 3D models takes place all the time. It is important to note that this morphing between local models is both photometric as well as geometric. Moreover, we ensure that it always proceeds in a physically-valid way, thus remaining transparent to the user of the system. For this purpose, algorithms capable of extracting both 2D correspondences between wide-baseline images, as well as 3D correspondences between local geometric models are proposed and used.

[1]Of course, we note that our system can be readily extended to handle the case of stereoscopic views that have been captured not just along one path but throughout the scene.

Our system can be also extended to handle the existence of multiple stereoscopic views per key position of the path, which are all related by a pure rotation of the stereoscopic camera. In that case, there will also be multiple local models per key-position. Therefore, before applying the morphing procedure, a *3D-mosaic* per key-position needs to be constructed as well. Each 3D-mosaic will simply comprise the multiple local models at the corresponding key-position and will itself be a bigger local model covering a wider field of view. Morphing can then proceed in the same way as before with the only difference being that these 3D-mosaics will be the new local 3D models to be used during the stage of morphing (in place of the smaller individual ones). So, in this case, a *morphable 3D mosaic* (instead of a morphable 3D model) will be transformed during the morphing.

Besides the proposal of a novel hybrid representation for a 3D scene, our system also includes new algorithms and techniques as part of its image-based modeling and rendering pipeline:

- More specifically, in the context of our photometric morphing procedure, a robust method for obtaining a dense field of 2D correspondences between a pair of wide-baseline images is proposed. In this case, the problem is that, due to the wide baseline, objects in the two images may appear at different scales. Therefore, simple similarity measures (e.g. correlation), that are typically used in stereo matching, are not appropriate anymore. To deal with this issue, we first reduce this task to a discrete energy minimization problem and then, to account for the existence of a wide baseline, the change of scale between corresponding local patches of the two images is also taken into account during the matching process.
- As part of our geometric morphing procedure, 3D correspondences between local geometric models needs to be established as well. A new approach is thus proposed for that purpose. Our method is not computationally demanding and is based just on solving a standard partial differential equation.
- Furthermore, in the context of the 3D mosaics construction, a technique for combining local 3D models (related to each other approximately by a 3D rotation) is presented, which is again based on solving a standard partial differential equation. Our method is robust enough so that it can cope with errors in the geometry of the local 3D models and always ensures that a consistent 3D mosaic is generated. To this end, geometric rectifications are applied to each one of the local 3D models during their merging.
- Finally, as part of our rendering pipeline, we propose the use of OpenGL pixel and vertex shaders so that both the photometric as well as the geometric morphing can be performed entirely on the GPU. This way, a highly optimized rendering path is obtained, which drastically

reduces the rendering time and allows very high frame rates to be achieved.

All of these algorithms are nicely integrated into a single framework, so that a complete, as well as powerful, image based modeling and rendering system is obtained in the end. Regarding the advantages of this system, the following points can then be made: **1)** To start with, no global 3D model of the environment needs to be assembled, a process which can be extremely cumbersome and error-prone, especially for large scale scenes where many local models need to be registered to each other. **2)** Furthermore, the rendering of all these local models is very inefficient and offers no scalability, thus leading to very low frame rates for large scale scenes. On the contrary, our method is scalable to large scale environments, as only one "morphable 3D-model" is displayed at any time. **3)** On top of that, our method also makes use of a rendering path which is highly optimized in modern 3D graphics hardware. **4)** In addition, by making use of an image-based data representation, our framework is capable of fully reproducing the photorealistic richness of the scene. **5)** Data acquisition is very easy (e.g. collecting the stereoscopic images for a path over 100 meters long took us only about 20 minutes), and requires no special or expensive equipment (but just a pair of digital cameras and a tripod). **6)** Finally, our framework makes up an end-to-end system, thus providing an almost automated processing of the input data, which are just a sparse set of stereoscopic images from the scene.

Also, as mentioned in the introduction, our system has already been successfully applied to the visual reconstruction of the Samaria gorge, and related results will be presented that demonstrate our method's effectiveness and power.

REFERENCES

[1] N. Komodakis and G. Tziritas, "Morphable 3d-mosaics: a framework for the visual reconstruction of large natural scenes," *in the video proceedings of CVPR 2006*.

[2] N. Komodakis, G. Pagonis, and G. Tziritas, "Interactive walkthroughs using morphable 3d-mosaics," in *3DPVT 2004*.

[3] P. E. Debevec, C. J. Taylor, and J. Malik, "Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach," in *SIGGRAPH 1996*.

[4] S. Fleishman, B. Chen, A. Kaufman, and D. Cohen-Or, "Navigating through sparse views," in *VRST 1999*.

[5] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering approach," in *SIGGRAPH 1995*.

[6] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH 1996*.

[7] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," in *SIGGRAPH 1996*.

[8] C. Buehler, M. Bosse, L. McMillan, S. J. Gortler, and M. F. Cohen, "Unstructured lumigraph rendering." in *SIGGRAPH 2001*.

[9] H. Schirmacher, M. Li, and H.-P. Seidel, "On-the-fly processing of generalized Lumigraphs," in *Eurographics 2001*.

[10] M. Uyttendaele, A. Criminisi, S. B. Kang, S. Winder, R. Hartley, and R. Szeliski, "High-quality image-based interactive exploration of real-world environments," *IEEE Computer Graphics & Applications*, 2004.

[11] D. G. Aliaga, T. Funkhouser, D. Yanovsky, and I. Carlbom, "Sea of images," in *VIS 2002*.