IMAGE-BASED TRANSFORMATION OF VIEWPOINT AND SCENE APPEARANCE

By

Steven Maxwell Seitz

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN – MADISON

1997

IMAGE-BASED TRANSFORMATION OF VIEWPOINT AND SCENE APPEARANCE

Steven Maxwell Seitz

Under the supervision of Professor Charles R. Dyer

At the University of Wisconsin–Madison

This thesis addresses the problem of synthesizing images of real scenes under three-dimensional transformations in viewpoint and appearance. It is demonstrated that a variety of three-dimensional scene transformations can be rendered on a video display device by applying simple transformations to a set of basis images of the scene. The virtue of these transformations is that they operate directly on images and recover only the scene information that is required in order to accomplish the desired effect. Consequently, they are applicable in situations where accurate three-dimensional models are difficult or impossible to obtain.

A central topic is the problem of *view synthesis*, i.e., rendering images of a real scene from different camera viewpoints by processing a set of basis images. Towards this end, two algorithms are described, called *view morphing* and *voxel coloring*, that warp and resample pixels in a set of basis images to produce new images that are physically-valid, i.e., they correspond to what a real camera would see from the specified viewpoints. Techniques for synthesizing other types of transformations, e.g., non-rigid shape and color transformations, are also discussed. The techniques are found to perform well on a wide variety of real and synthetic images.

Contributions include a uniqueness result, which proves that all views on the line segment between the two camera centers are uniquely determined from two uncalibrated views of a scene. Importantly, neither dense pixel correspondence nor camera information is needed, given a constraint that no occlusions occur between the input viewpoints. To treat the general case of many views, a novel *voxel coloring* framework is introduced that enables (1) the derivation of a set of *color invariants*, and (2) an efficient solution to the dense correspondence problem from widely-spaced views. This algorithm has several advantages, most notably its ability to easily handle occlusion and views that are arbitrarily far apart, and its usefulness for *panoramic* visualization of scenes.

IMAGE-BASED TRANSFORMATION OF VIEWPOINT AND SCENE APPEARANCE

By

Steven Maxwell Seitz

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN – MADISON

1997

Abstract

This thesis addresses the problem of synthesizing images of real scenes under three-dimensional transformations in viewpoint and appearance. Solving this problem enables interactive viewing of remote scenes on a computer, in which a user can move a virtual camera through the environment and virtually paint or sculpt objects in the scene. It is demonstrated that a variety of three-dimensional scene transformations can be rendered on a video display device by applying simple transformations to a set of basis images of the scene. The virtue of these transformations is that they operate directly on images and recover only the scene information that is required in order to accomplish the desired effect. Consequently, they are applicable in situations where accurate three-dimensional models are difficult or impossible to obtain.

A central topic is the problem of *view synthesis*, i.e., rendering images of a real scene from different camera viewpoints by processing a set of basis images. Towards this end, two algorithms are described that warp and resample pixels in a set of basis images to produce new images that are physically-valid, i.e., they correspond to what a real camera would see from the specified viewpoints. Techniques for synthesizing other types of transformations, e.g., non-rigid shape and color transformations, are also discussed. The techniques are found to perform well on a wide variety of real and synthetic images.

A basic question is uniqueness, i.e., for which views is the appearance of the scene uniquely determined from the information present in the basis views. An important contribution is a uniqueness result for the no-occlusion case, which proves that all views on the line segment between the two camera centers are uniquely determined from two uncalibrated views of a scene. Importantly, neither dense pixel correspondence nor camera information is needed. From this result, a *view morphing* algorithm is derived that produces high quality viewpoint and shape transformations from two uncalibrated images.

To treat the general case of many views, a novel *voxel coloring* framework is introduced that facilitates the analysis of ambiguities in correspondence and scene reconstruction. Using this framework, a new type of scene invariant, called *color invariant*, is derived, which provides intrinsic scene information useful for correspondence and view synthesis. Based on this result, an efficient voxel-based algorithm is introduced to compute reconstructions and dense correspondence from a set of basis views. This algorithm has several advantages, most notably its ability to easily handle occlusion and views that are arbitrarily far apart, and its usefulness for *panoramic* visualization of scenes. These factors also make the voxel coloring approach attractive as a means for obtaining high-quality three-dimensional reconstructions from photographs.

Acknowledgements

This thesis is dedicated to my parents, Gary and Sheila Seitz, who taught me that education is a lifelong pursuit and inspired me to pursue a Ph.D.

I wish to thank Chuck Dyer for his guidance and support throughout my graduate studies. His advice and infallible judgement on all matters academic and otherwise have been a tremendous influence and an invaluable resource for me. I am very fortunate to have had the opportunity to work with him during my time at U.W. Madison. I would also like to thank the other members of my thesis committee, Nicola Ferrier, Vladimir Lumelsky, Vadim Shapiro, and Jude Shavlik for their comments and insights on this work.

I am especially grateful to Larry Rowe for introducing me to scientific research as an undergraduate at U.C. Berkeley and for encouraging me to pursue graduate studies in computer science. Kyros Kutulakos contributed key ideas to this work and provided a strong influence both as peer and collaborator.

In the course of graduate studies, I have benefited from interactions with many people. I would especially like to thank Brian Roddy, Aaron Seitz, Kathy Suchenski, Sanjay Tiwari, and John Watrous for many illuminating discussions. Advice and insights on this work were provided by several current and former members of the vision group at U.W. Madison, including Russel Manning, Brian Morgan, Andrew Prock, Dan Reznik, Brent Seales, and Steve Walden. I also thank the following people who graciously permitted their images to be used in this thesis, in order of appearance: John Watrous, Su-hui Chiang, Shubu Mukherjee, Dionisios Pnevmatikatos, Alain Kagi, Elton Glaser, Babak Falsafi, and Doug Burger.

Most of all, I am deeply grateful to Sandhya Subramanian for her endless encouragement and enthusiasm, and for making these last several years truly enjoyable.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Humans perceive the three-dimensional world by means of its projection onto our retinal fields. By mimicing this projection process, the eye can be fooled into perceiving depth and three-dimensionality in computer-generated imagery. Interactive computer graphics systems exploit this trick to enable a user to visualize and transform artificial objects and scenes in three dimensions via 2D display and input devices. In this thesis, we consider the problem of designing computer algorithms to visualize and manipulate the appearance of *real* objects and scenes that exist physically or have existed in the past. Unlike artificial scenes which must be painstakingly described in minute detail, real scenes have an intrinsic structure that can be visually captured in photographs and drawings.

At present, making use of this structure in computer graphics applications requires reconstructing the 3D structure of the scene from the images. This reconstruction problem, despite being well-studied in the computer vision literature, is not entirely solved and existing techniques are prone to a variety of errors and instabilities (cf. Chapter 6). Even relatively minor errors in shape can cause distracting artifacts when the model is texture-mapped and reprojected. Rather than attempt to extract a 3D scene description from the photographs, an alternative approach is to model the effects of scene transformations in the *image domain*, by determining how the image should change in response to a particular scene transformation. As shown in Fig. 1.1, applying a scene transformation and rendering the resulting scene yields a corresponding image transformation. Similarly, image transformations can *imply* scene transformations. By modifying the image, it should therefore be possible to mimic physical changes to the real 3D scene. We use the phrase *image-based scene transformation* to describe the synthesis of 3D scene, camera, or illumination transformations via modifications to one

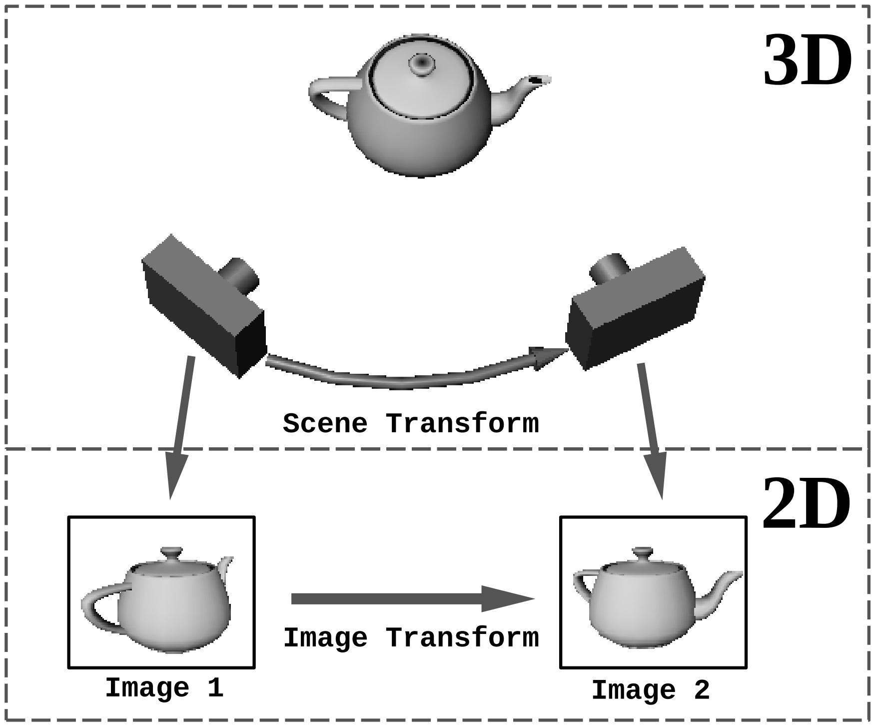

Figure 1.1: Image-based Scene Transformation. A scene transformation such as a change in camera viewpoint (top) causes the image to change (bottom). Similarly, a well-designed image transformation can *imply* a 3D transformation in the scene. We refer to this latter technique of synthesizing 3D effects via 2D image operations as *image-based scene transformation*.

or more basis (input) images.

An advantage of this image-based approach is that precise modeling and reconstruction of shape, illumination, texture, and other scene quantities from images is avoided. Rather, we recover only the image information that is required to create the desired 3D effect. Consequently, image-based transformations can be simpler and more efficient than methods that require 3D reconstructions. More importantly, we will argue that image-based transformations are more *powerful* in that they enable visualization and manipulation of scenes that are difficult or impossible to reconstruct.

A transformation of particular interest is viewpoint (i.e., camera panning, zooming, and translation). Cast as an image transformation, a change in camera viewpoint can be

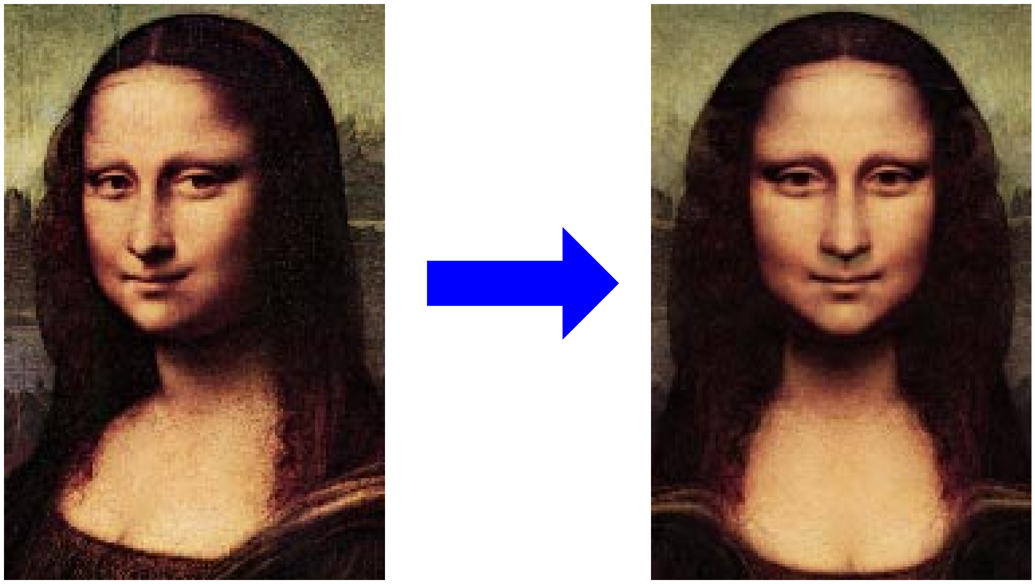

Figure 1.2: Image transformation of photographics, drawings, and paintings enables 3D visualization of historic objects and scenes. Here the Mona Lisa appears to rotate three-dimensionally by use of the view morphing technique described in Chapter 3.

modeled as a mapping, or *warp*, between pixels in one or more basis views and pixels in a new image, representing a synthetic view of the same scene. By categorizing the set of all such mappings, we can begin to devise mechanisms for *synthesizing* changes in viewpoint by warping images. This capability is critical for interactive graphics applications like visualization, virtual reality, flight simulation, and games, in which viewpoint changes continuously, subject to user-control. It also enables manipulation of scenes whose three-dimensional structure is unknown and unrecoverable. For instance, a present-day tourist could virtually visit a historic building, long ago destroyed, by warping a set of photographs from a time when the building was still standing. The user could view the building in 3D by moving a virtual camera through the environment. The illusion of a real three-dimensional space would be maintained by continually warping the images to correspond to the user's changing perspective. The same approach can be applied to drawings and paintings. For instance, Fig. 1.2 depicts a 3D rotation of Leonardo Da Vinci's *Mona Lisa*, generated using the image-based technique described in Chapter 3.

In this thesis we propose the use of image-based scene transformations as a general framework for visualizing and manipulating real 3D scenes on a computer display, and we describe algorithms to enable common transformations such as rigid and non-rigid 3D motions, viewpoint changes, and editing operations. These algorithms have the key advantage that they operate on a set of basis images and do not require 3D models of the scene. To further motivate the use of image transformations, we begin with examples of some simple image transformations that are capable of producing realistic 3D effects.

1.1 Examples of Image-Based Scene Transformations

In the early part of this century, Walt Disney and other pioneers developed the *cel animation* techniques that have been used since to generate cartoons and films like the *Three Little Pigs* [Wal33], *Fantasia* [Wal40a], and *Aladdin* [Wal92]. Cel animators employed a number of methods for conveying realistic motions, using paint and transparent sheets of celluloid as the basic medium. The success and continued popularity of their efforts attests to the efficacy of 2D techniques for producing realistic visual effects.

Whereas traditional animation focused primarily on 2D effects, many of which featured artistic distortions and exaggerations [Las87], certain 3D effects (i.e., *scene transformations*) were also employed. These included simulations of 3D camera translations by moving a small window over a background image. Translating the window conveys a frontoparallel motion of the camera, and scaling the window emulates forward or backward motion. While sufficient to give an impression of camera movement, this technique does not model depth effects like parallax and occlusion and therefore is somewhat unrealistic. A more compelling effect can be achieved by designing the background image with a specific camera path in mind [WFH+97]. This technique was used in Walt Disney's 1940 animated film, *Pinocchio* [Wal40b], in which a window was moved in a fixed path along a special background image. While the background appears warped when viewed in its entirety, it appears quite natural through the sliding window and the resulting animation gives a realistic 3D effect. Wood et al. [WFH+97] have recently developed computer algorithms for partially automating the creation of these warped