Introduction
Aleš Leonardis, Franc Solina, Ruzena Bajcsy

In the past, the fields of computer vision and computer graphics have been considered as tackling the inverse problems. Traditionally, computer vision starts with input images and process them for the purpose of understanding geometric and physical properties of objects and scenes, and to build appropriate models. On the other hand, traditional computer graphics starts with geometric models and then generates, manipulates, and displays virtual representations in the form of images. In the recent years, we have seen great efforts towards the integration of computer vision and computer graphics techniques, in particular, in the areas of realistic modeling of objects and scenes, interactive computer graphics, and augmented reality.

Which are the main reasons that have brought the researchers of computer vision and computer graphics closer together? Computer graphics community has been active in inventing increasingly better, faster, and more complex methods of animation for creating virtual 3D synthetic environments. This paradigm has been successfully demonstrated in the constantly expanding series of complex digital animations. Yet, even though these animations have been successful (also commercially), it is clear that in order to create even more realistic virtual environments and on a much larger scale, the cost and time involvement have to be lowered. Such a leap in cost and quality is possible by incorporating sensor data of actual physical objects and environments, that can later be modified and extended with synthetic data. For example, it should be less costly to make a 3D model of a complex geometric object directly from images, than to have an animator construct such a model by hand.

Sensor data, which are crucial for an efficient creation of larger and more realistic virtual(ized) environments, need to be properly interpreted, processed, and modeled—and this is precisely what computer vision does. In other words, computer vision provides the tools needed
to transform the real world back into the virtual. The marriage of computer graphics and computer vision techniques is further facilitated by increasingly lower costs of hardware for image capture and processing.

Some of the most critical issues in this emerging field are:

- generation of highly realistic 3D graphical models from a collection of calibrated or non-calibrated images obtained with multiple cameras, stereo-rigs, or video cameras;
- building of dynamic deformation models using the motions captured from real video sequences;
- dynamic image-based rendering;
- role of computer vision in interactive computer graphics;
- tracking, ego-motion estimation, and registration techniques for harmonious integration of real worlds and computer generated objects;
- virtual studio techniques.

The key element for the successful merger of computer graphics and computer vision technique are appropriate models. The models provide the information describing the geometry, the dynamics, and many other attributes of objects and scenes that represent the prior knowledge and impose a set of constraints for analysis, and later for rendering the virtual environments.

The Book at a Glance

The book focuses on the integration of computer vision and computer graphics techniques in the areas of realistic modeling of objects and scenes, interactive computer graphics, and augmented reality. Each chapter of the book presents recent results within this emerging domain. The results encompass both theoretical formulations and derivations, as well as numerous examples of applications. These include urban and archeological site modeling, modeling dressed humans, medical visualization, figure and facial animation, real-time 3D-teleimmersion and telecollaboration, augmented reality as a new user interface concept, and augmented reality for underwater scene understanding. The chapters have been designed to serve as technical overviews with extensive references to related work. We hope that this will enable researchers working in individual fields to quickly get acquainted with the main common integration issues, while providing the experts with in-depth technical details.
The book starts with a chapter that describes a method to completely automatically recover a 3D scene structure from a sequence of images acquired by an unknown camera undergoing unknown motion. Zisserman et al. argue that, in contrast to previous approaches, which have used calibration objects or landmarks, their approach is far more general since no other information than images themselves is required. The automatic process can be thought of, at its simplest, as converting a camcoder to a sparse range sensor. Together with more graphical post-processing, such as triangulation and texture mapping, the system becomes a “VHS to VRML” converter. The authors demonstrate two applications of their method: the first is the construction of 3D graphical models of a piece-wise planar scene, and the second is the insertion of virtual objects into the original image sequence, which is of use for post-production in the film industry.

The next chapter by Cross and Zisserman describes a novel approach to reconstructing the complete surface of an object from multiple views, where the camera circumnavigates the object. To achieve the goal, the approach combines two sources of information, namely, the apparent contour and the imaged surface texture. The authors argue that the proposed approach has significant advantages over using either the contour or texture alone: in particular, the geometric constraints available are complementary, so that the deficiencies of one source can be overcome by the strengths of the other. In addition, the novelty lies also in an implementation which uses different surface representations as appropriate for accuracy and efficiency. Numerous examples of automatically generated texture-mapped graphical models demonstrate that the approach is successful for various objects and camera motions. The objects may contain concavities and have non-trivial topology.

The subsequent chapter by Urban et al. also addresses the problem of scene reconstruction from multiple views. The authors, in particular, concentrate on consistent projective reconstruction which involves a set of more than four views. The method is based on concatenation of trifocal constraints and requires only linear estimates. The accuracy and stability of the method have been analyzed, and the projective reconstruction from seven real images has been successfully demonstrated.

Šára explores a bottom-up approach to precise and accurate 3D surface model reconstruction. The focus is on acquiring 3D models of natural objects for medical applications, augmented reality, and telepresence. The author proposes performing several successive steps in which more complex models are inferred from simpler models. The model at the lowest level consists of a set of unorganized points in 3D space obtained from polynocular stereo system which utilizes five fully-calibrated cam-
eras and an uncalibrated infrared texture projector. The intermediate-level model consists of local geometric primitives, called “fish-scales”. By linking together close and compatibly oriented fish-scales, a discrete pseudo-surface is obtained, which presents the high-level model. Throughout the chapter, the approach is demonstrated on a textured 3D geometric model reconstruction of a human face.

The contribution by Roth brings a systematic review of the problem of model building from sensor data, which stands at the interface between computer vision and computer graphics. The author first describes the basic steps in the model building pipeline; calibration, acquisition, registration, point creation, model creation, model compression, and texture creation. After which he systematically discusses open research questions that remain in each step and describes several overall research themes that he believes should further guide work in this area. Among the most important open problems in model building, the author lists: automation of the entire model building pipeline, incremental construction of the models, the role of active versus passive sensors, image-based versus model-based rendering, and environment modeling versus object modeling.

Skočaj and Leonardis address the problem of 3D reconstruction of objects of non-uniform reflectance using a structured light sensor. Namely, standard approaches using structured light sensors assume that the reflectance properties of the objects are uniform. The authors illustrate the need to devise an approach that overcomes this constraint, which means that objects consisting of both high and low reflective surfaces should reliably be reconstructed. They propose to systematically vary the illumination intensity of the light projector and to form high dynamic scale radiance maps. The authors report experiments on objects which have surfaces of very different reflectance properties, and demonstrate that range images obtained from high dynamic scale radiance maps are of much better quality than those obtained by the standard approach.

While the previous chapters have explored in particular the 3D reconstruction of objects and scenes, which can subsequently be used in computer graphics applications, this chapter, by Manning and Dyer, addresses the view-interpolation as a means of creating virtual views of scenes without explicit scene reconstruction. The authors present a technique, called “dynamic view morphing”, for view interpolation between two reference views of a dynamic scene captured at different times. The interpolations produced portray one possible physically-valid version of what transpired in the scene between the time points when the two reference views were taken. The presented method works with widely-
separated, uncalibrated cameras and sparse point correspondences, and does not involve finding the camera-to-camera transformation.

Tao and Huang address the problem of building deformation models of faces and facial expressions using the motions captured from real video sequences, which is an excellent example that demonstrates the close relationship between computer graphics and vision technologies. The authors propose an explanation-based facial motion tracking algorithm based on a piecewise Bézier volume deformation model (PBVD), which is a suitable model both for the synthesis and the analysis of facial images. With the PBVD model, which is linear and independent of the facial mesh structure, basic facial movements, or action units, are interactively defined. By changing the magnitudes of these action units, a variety of different animated facial images can be generated. The magnitudes of these action units can be computed (learned) from real video sequences using a model-based tracking algorithm. The authors present experimental results on PBVD-based animation, model-based tracking, and explanation-based tracking.

The subsequent chapter by Van Gool et al. brings together many different techniques for realistic object, scene, and event modeling from image data, to realize a system for visits to a virtual 3D archeological site. To model the landscape and buildings at the site, the authors propose a shape-from-video system that turns multiple, uncalibrated images into realistic 3D models. The texture that covers the 3D models of the landscape is synthesized from images of the natural surfaces which results in a compact, yet effective texture model. To model smaller pieces, like statues and ornaments, they use an active, one-shot range sensor which exploits the projections of a special pattern onto the object under observation to yield high resolution 3D models. Once the model of the site is built, one can navigate through this virtual environment accompanied by a virtual guide. The virtual guide responds through head movements and facial expressions. The authors also developed a technique for learning natural lip motions from observed 3D face dynamics which will be used to animate the virtual guide in the future versions of the system.

Jojić and Huang present another application which demonstrates how computer vision techniques can be exploited in conjunction with computer graphics for modeling dressed humans. They combine computer vision based approaches such as 3D reconstruction of a human body and analysis-by-synthesis of the behavior of cloth material with the computer graphics approaches for realistic rendering of complex objects. The experimental results presented in the chapter include building textured 3D models of humans from multiple images, dressing these models into vir-
tual garment, and joint estimation of cloth draping parameters and the underlying object’s geometry in range images of dressed objects.

The next chapter by Leberl et al. describes urban site modeling. The authors argue that while 3D models of buildings have long been produced based on photogrammetric technology, the focus today has shifted towards producing 3D computer models of urban areas on a large scale, with perhaps half a million buildings of one metropolitan area. Thus, they discuss various issues related to the creation of such large data sets, with verified accuracy and detail, in a reasonable time and at a moderate cost.

Clapworthy et al. have identified a number of relatively-unconnected areas where computer graphics aspects i.e., visualization and animation, are influenced by the use of computer vision techniques. These techniques include 3D scene and object reconstruction, motion capture, and segmentation. The authors present various examples from diverse areas such as medical visualization, biomechanics, figure animation, and robot teleoperation.

In his chapter, Wojdala presents an exciting new technology, called “Virtual Studio”, whose main goal is to combine two separate images or video streams. A typical example is to merge a foreground, filmed with a camera, with a computer-generated background, so well, that the composite looks as if it was shot together, in one environment. To answer the basic question, namely, how realistic virtual sets can appear, the author discusses various techniques involved into this very interdisciplinary technology. These techniques are related to computer graphics, graphics hardware, chroma keying, lighting, video camera parameters, camera tracking, and interaction between real and virtual worlds. While some problems still remain, the current level of visual realism has reached the point where more and more broadcasters are using this technology, even for complex, live-to-air productions.

The contribution by Daniilidis et al. presents an application which demonstrates the confluence of computer vision, computer graphics, and communication. In particular, they implemented a new medium for telecollaboration, which was realized in the form of two tele-cubicles connected at two Internet nodes. At each telecubicle a stereo-rig is used to provide an accurate dense 3D reconstruction of a person in action. The two real dynamic worlds are transmitted over the network and visualized stereoscopically. The full 3D information facilitates interaction with any virtual object. The remote communication and the dynamic nature of telecollaboration offers the challenge of optimal representation for graphics and vision. Thus, the authors treat the issues of limited bandwidth, latency, and processing power with a tunable 3D represen-
tation, where the user can select the trade-off between delay and 3D resolution by tuning the spatial resolution, the size of the working volume, and the uncertainty of reconstruction.

The last four chapters deal with Augmented Reality—a technology by which a user’s view of the real world is augmented with additional information from a computer model. In her chapter, Klinker gives an overview of this new technology which shows great promise as a new user interface concept for many applications. Namely, users can work with and examine real 3D objects, while receiving additional information about these objects or the task at hand. Yet, augmented reality applications require fast and accurate solutions to several very complex problems, such as tracking the user and the real object, handling the occlusions and reflections, as well as the motion of the virtual user. Klinker discusses computer vision based solutions which are currently considered to be among the most promising approaches towards solving these issues.

Simon et al. focus on one of the most crucial problems in augmented reality, namely, how to achieve a harmonious integration of real world and computer generated objects. They propose a robust and accurate registration method which performs pose computation over the sequence of images in a completely autonomous manner. The accuracy of the pose computation is achieved by combining model-image correspondences of tracked curves in an image and 2D keypoint correspondences matched in the consecutive frames. The authors demonstrate the seamless integration of the real and virtual worlds by integrating a virtual car into a real-world video sequence.

In the next chapter by Gagalowicz and Gérard, the authors also tackle the problem of tracking in images. The approach is model-based and proceeds as a two-step process. After the interactive calibration session, the geometric model of the object is automatically rendered with texture. Then, a 3D predictor gives the position of the object model in the next image and the fine tuning of the position is obtained by minimizing the error between the textured model and the real image of the object. The robustness of the approach has been verified by creating an augmented reality sequence.

In the last chapter, a specific application of augmented reality is presented by Murino and Fusiello. They describe how to integrate visual and acoustic 3D data to enhance the perception of an underwater environment during teleoperation tasks.

The chapters in this book present only a selection of some of the representative approaches that have emerged in the recent years. In fact, the problems have mainly been observed from the point-of-view of computer vision researchers. A similar book, but more from the point-
of-view of computer graphics researchers, would nicely complement this volume and shed additional light on this exciting new research domain. Many challenges are waiting in this interdisciplinary research field. One of the ultimate challenges is to bring these exciting ideas, in the form of easy to use tools, into the hands of all computer users.