Chapter 1

Introduction

1.1 Overview

Over the years the word "Multimedia" has meant a variety of things to different people. In this book we will use multimedia to cover image, video and 3D (image and three-dimensional model) content; as well as consider the network transmission related to these types of content.

Our focus will be on "Online" multimedia applications and related technologies. The requirements and challenges for online applications differ significantly from offline or standalone applications. For example, in a standalone computer gaming environment the size of the images and 3D models being used may not be an issue of major concern as long as there is sufficient memory on the computer being used, while the time taken to display 3D objects would be a primary concern to enhance the speed of interactivity. For online gaming, however, the speed of communication among a collection of players located remotely from one another is an important factor in the design.

A standalone imaging system, that is designed to record diagnostic information for doctors, only needs to record data with a high enough quality to detect abnormal conditions. An online Tele-health application that is designed to help doctors make diagnostic decisions on patients located at a remote location would need to deal with several other issues, such as: (i) Efficient storage and representation of the multimedia information on patients so that fast retrieval and visualization is feasible; (ii) The network conditions that exist between the site recording the data (patient site) and the site visualizing the data (doctor site); (iii) Possibly allowing progressive update of a visual object through interaction with the doctor; (iv) Permitting modification or annotation on the object; (v) Supporting collaborative

diagnosis among multiple doctors located at different sites.

A key difference in several approaches discussed in this book, compared to usual algorithms, is taking perceptual quality into account. Traditional methods of evaluating "quality" of multimedia objects relied on numerical measures, such as signal-to-noise ratio, which treated visual data purely as a collection of numbers and computed an estimate of the difference between the original information recorded and the data being observed. From the viewpoint of a human observer, however, the only thing that counts is the perceived quality of one object relative to another; *i.e.*, does one object look "better" than another and if so "how much better." Special attention is given to designing a metric to estimate the perceptual quality of 3D objects; in this context, setting up an environment to measure "subjective" human evaluations and statistical methods for estimating the reliability of the subjective judgments are discussed.

The topics discussed roughly cover five categories: some networking issues in online multimedia, TexMesh simplification and view independent transmission, view dependant transmission and server side rendering, content and background creation, and creating simple online games. Under networking issues we consider transmission strategies based on actively monitoring network resources, for one or more servers, and accordingly adapting or scaling the data transmitted. The objective is to provide users service quality close to a level desired by them; an area often referred to as Quality-of-Service (QoS) based delivery. A survey of strategies for wireless multimedia communication is also covered in Chapter 3. Chapters 4 to 9 cover topics related to joint texture-mesh (TexMesh) simplification, adaptive online transmission based on available bandwidth, perceptual factors and packet loss in wireless or unreliable networks. Predictive methods and rate-distortion optimization for server side rendering are discussed in Chapters 10 and 11. A brief overview of techniques for content and background creation is given in Chapters 12 and 13. Finally, some examples of simple networked online games are provided in Chapters 14 and 15.

Different strategies for 3D scene representation can be considered for online applications, including 3D TV transmission [Kang *et al.* (2000)]. Strategies can assume fixed geometry and static texture to view-dependent rendering to image based rendering with or without geometry [Magnor *et al.* (2003)]. In this book we will discuss client side rendering, with static geometry and texture, and server side view-dependant rendering, considering perceptual quality and rate-distortion optimization. Issues related to image based rendering will not be considered primarily to limit the scope of

the book, but also to keep our focus more on applications with limited bandwidth.

In previous remote visualization transmission strategies, as shown by solid lines in Fig. 1.1 [Kang *et al.* (2000)], the client's request defines the processing required at the server. In contrast, the strategies described in this book involve equally important server decisions in order to generate the best possible geometry and texture rendering at the client. In Fig. 1.1 the new strategies are indicated by dotted lines, which involve both client and server policies in the remote visualization framework.

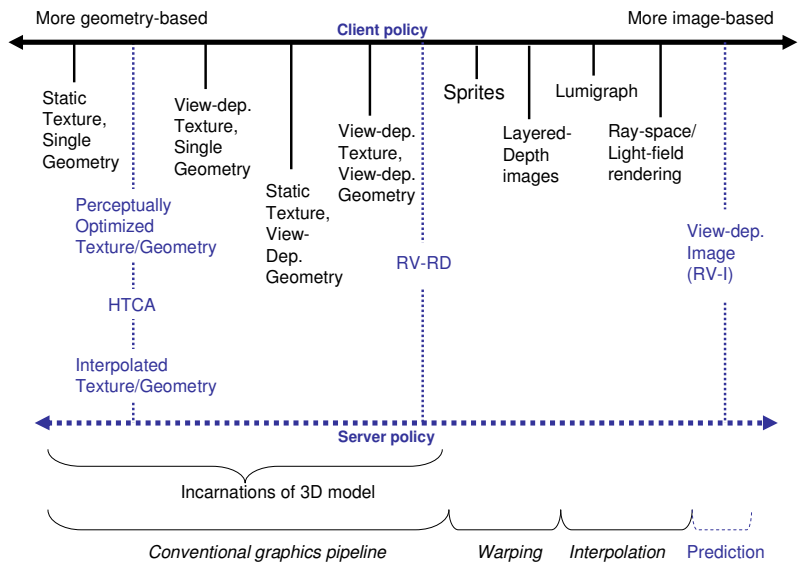

Fig. 1.1 Strategies for online 3D transmission.

The target audience for this book includes both students and researchers in multimedia, and online multimedia developers. It is expected that the readers will have a basic knowledge of JAVA or a related language, be quite familiar with the Internet, and have a keen interest in online applications. Keeping this broad group in mind several aspects of the implementations are discussed, including JAVA3D and data representations. Some of the source and object codes relating to implementations are also provided.

The implementation related discussions in this book are based on the JAVA^{TM} language. For 3D applications we will consider the JAVA3D language. An overview of the JAVA3D language is given in an online Appendix

for convenience. The description of the servers relies on the open source Apache code. The primary motivation behind choosing these software platforms is that they are available to all students, researchers and practitioners free of cost; also, there are various discussion forums where development and debugging information can be obtained as and when needed.

Another topic related to a 3D multimedia environment is the representation of 3D data. We limit the scope of the current manuscript to a collection of static 3D objects which can collectively constitute a dynamic scene. 3D objects are created using surface images (texture) and a triangulated set of vertices (mesh) representing a 3D model. Other 3D representations, such as voxels, will not be discussed.

1.2 Networking Issues in Online Multimedia

Strategies for transmission of multimedia information are dependant on ongoing monitoring of network resources. Users (clients) in an online environment often have minimum requirements on the performance of an application. Performance may be measured in terms of response time for a request, jitter in video, quality of 3D models, and so on.

Providing QoS guarantees in a deterministic fashion is virtually impossible in a dynamic network, such as the Internet, where the available bandwidth can fluctuate significantly from one period of time to the next. We therefore focus on statistical approaches to bandwidth monitoring and providing QoS guarantees. In Chapter 2 an algorithm for optimal bandwidth monitoring given a single server, communicating with a client, is discussed. What is unique about this monitoring approach is that the amount of bandwidth testing is dependant on the current statistical properties of the network. Prior approaches often relied on prediction based on historical data; the problem with such methods is that the behavior of the Internet can change considerably over time, thus estimate based on data collected over the last 15 minutes, for example, may not be a good indicator of the expected characteristics in the next 15 minutes. Thus, the approach presented for bandwidth monitoring uses a starting fraction of the current time interval, in which a visual object is to be transmitted, to estimate the bandwidth.

The performance of multimedia systems can often be improved if more than one server is used in an application. For example, to retrieve a 3D object the texture component could be retrieved from one server and the

mesh component from another server. Distributed retrieval, however, has many challenges, such as, distributed bandwidth monitoring, deciding how to divide a retrieval task among various servers, and combining the transmissions from various servers into one meaningful result. Some of the possible options for distributed bandwidth monitoring are presented in Chapter 2. Here again, statistical approaches are developed for the monitoring process. Two strategies for multi-server bandwidth monitoring are considered: one in which each server performs its share of a task independent of other servers, and another in which all the servers collaborate to complete a task. It is shown through statistical modeling coupled with simulation experiments that the collaborative approach is superior to the independent multi-server model.

How can we scale multimedia data based on monitored bandwidth to provide clients a statistical QoS guarantee? Some answers to this question are given in Chapter 2. It should be noted that statistical QoS guarantee is fundamentally different from traditional deterministic QoS guarantees. In a statistical approach, QoS is met with a chosen level of confidence, *e.g.*, 95% or 75% confidence. A 95% confidence level means that statistically 95 times out of 100 the approach is expected to satisfy a specified QoS requirement. Both simulations and real network tests are presented to validate the theoretical models behind single and multi-server optimal bandwidth monitoring and statistical QoS based transmission.

In Chapter 6 an alternative approach to joint texture and mesh (TexMesh) transmission considering network bandwidth and perceptual quality is presented. The strategy allows for adaptive control of remaining portion of the multimedia data to be transmitted.

1.3 Joint Texture-Mesh Simplification and View Independent Transmission

Methods for coding various types of visual information can differ significantly depending on the type of information. Images (or texture of 3D objects) can be transformed to the frequency domain for JPEG compression, or the wavelet domain for JPEG2000 compression. Video can be coded following the MPEG standard which consists of a collection of I-frames (Intra-coded static images compressed by themselves) and motion encoded frames (P-frames for Predictive coded frames & B-frames for bi-directionally coded frames); or the MPEG-4 standard which is more versatile and allows for

coding of 3D meshes, background and foreground, and Regions-of-Interest (ROIs). Various strategies exist for mesh coding, with several adopting strategies for mesh simplification based on viewpoint, distance to observer, and perceptual quality.

Perceptual quality and ROI can play a key role in reducing bandwidth requirements. One of the major limitations in developing online multimedia systems is maintaining reasonably good visualization quality while communicating over links with limited and often fluctuating bandwidth. To work under these restrictions one can consider the following approach: (i) Transmit a coarse visual object quickly; (ii) Take the viewers interest into account for updating the coarse model; (iii) Use perceptual quality studies to device strategy for update of mesh vs. texture, *i.e.*, what proportion of the bandwidth should be taken for mesh transmission with the remaining bandwidth being used for texture transmission. After a brief overview of literature on 3D simplification in Chapter 4, a newer approach based on scale space filtering is discussed in Chapter 5. We consider photorealistic textured meshes, incorporating online adaptation based on available bandwidth in Chapter 6 and perceptual issues in Chapter 7. Chapter 8 accounts for tradeoffs between texture and mesh from a perceptual perspective. Strategies for 3D perceptual quality optimization under packet loss over wireless or unreliable networks are outlined in Chapter 9.

1.4 View Dependent Transmission

View independent transmission is necessary when users need to quickly interact with an object and change viewpoints fast. However, these strategies require resources at the client site for interpolation and reconstruction from partial data. When clients are slow and bandwidth is quite limited, view dependent transmission can be a better option. In Chapter 10, some alternatives for remote 3D visualization through image transmission are proposed. Predictive compression algorithms are used to reduce bandwidth requirement and overhead at the receiver. Copyright issues are also resolved because the full 3D model is not transmitted to the client. Prediction schemes can be image-based or model-based. A simple image-based prediction scheme contains redundancy between different views of the same model. To suppress redundant data, some geometry information is used to predict the new views. Pixels that cannot be predicted because of occlusion, different lighting conditions, quantization issues, etc. can exist in

images. To overcome this problem, warping of two different views is combined to predict the new frame; or Layered Depth Images are taken from a single viewpoint but with multiple pixels along each line of sight are used to predict artifacts. Image-based schemes also need to handle holes filling in zooming operations. A different approach is to use model-based prediction, in which a simplified version of the mesh is stored at the client for rendering. In this case, the view rendered from the simplified mesh is subtracted from the one rendered from the complete mesh, and the difference is transmitted to the client.

Although transmitting images instead of the full model speeds up rendering time and reduces workload on the client side, little has been explored regarding how to extract the client's requested views from a 3D model effectively. In practice, video-rate is allocated for current implementations that require considerable bandwidth. Furthermore, the predictive approach can only deliver approximation and the rendered quality is significantly constrained by the network. In fact, the server has to replicate every step the client has to perform. A rate distortion theoretical approach (RV-RD) is proposed in Chapter 11, in which both server and client adopt cooperating policies in order to offer the user the best possible rendering with the available bandwidth. The client interactively determines the view of interest. Based on the client's request and available bandwidth, the server progressively transmits 3D data and texture. The received data is stored and then rendered based on the client's capacity. RV-RD is a very flexible system: at one extreme, if the server's policy ignores the viewpoint required at client side, it reduces to the transmission of a compressed representation of the whole scene. At the opposite, if the server does not transmit any 3D data, RV-RD works as a simple image transmission on demand system.

1.5 Content and Background Creation

Creating 3D content and background is a challenging problem limiting the widespread use of 3D multimedia. For 3D model creation there are active (using structured light and laser) approaches and passive approaches that do not need special lighting or other intervention. To address some of the complexities of 3D model creation, Chapter 12 reviews 3D reconstruction by passive methods. A high-level conceptual outline of the field is given, referring to the literature for technical details. The chapter discusses the framework for calibration, and first reviews major mono-modal

methods: shape from silhouette, shape from shading, shape from self-shadow, shape from focus and shape from stereo. Subsequently, a general framework for multimodal methods, focusing on the deformable model techniques is outlined; in this context a specific example considering the joint use of silhouette and stereo data is presented.

Rather than constructing 3D scene geometry at considerable scanning or reconstruction cost, an alternative is to use panoramic imaging (possibly from multiple viewpoints) to create backgrounds. Chapter 13 discusses remote visualization of panoramic images following progressive compression standards. Implementations show that such an approach allows effective remote display and zoom into panoramic images for large size visualization windows. A solution is proposed following JPEG2000, which is intrinsically more efficient than JPEG and that it is scalable in image resolution and quality.

1.6 Implementing Simple Online Games

To motivate the utility of the theoretical models and associated implementations a few application scenarios are considered. In Chapter 14 we discuss how a simple two-player 3D game can be designed. This game, called "Castles," can be considered as a more complex and 3D version of the "Go" board game. The objective here is to have two players compete for space using 3D building blocks. In order to have a realistic demonstration even issues of balancing and gravitational effects of structures need to be taken into account. For an online game that does not require real-time imaging input data most of the texture and models can be pre-stored on client computers, and only the control, scene composition and interaction information needs to be transmitted and updated on an ongoing basis. The framework in Chapter 14 is extended to a networked online environment in Chapter 15, allowing multiple players to login using an online interface.

The design and implementation of a slightly more complex online game called "Siege" is discussed in Chapter 16. The idea behind Siege is to protect your tower from attacks by other players while at the same time using your own siege weapons to destroy others' towers. Siege is based on a 1983 board game called Crossbows and Catapults by Base Toys. This game gives readers an idea on how to design weapons and control their functionalities.

1.7 Online Appendices

Some of the material considered helpful in developing interactive online multimedia is supported through online appendices located at www.cs.ualberta.ca/∼anup/3Dbook. At present, a review of JAVA3D and description of Remote Machine Invocation (RMI) based collaborative online 3D editing are available on this website. However, we plan to add more online implementation tutorials at this site over the next few years.

References

Kang, S. B., Szeliski, R. and Anandan, P. (2000). Geometry-image representation tradeoff for rendering, in *IEEE Int. Conference on Image Processing*.

Magnor, M., Ramanathan, P. and Girod, B. (2003). Multi-view coding for image based rendering using 3D scene geometry, *IEEE Trans. on CSVT*, pp. 1092–1106.