Perceptually Optimized 3-D Transmission Over Wireless Networks
I Cheng, A Basu

To cite this version:
I Cheng, A Basu. Perceptually Optimized 3-D Transmission Over Wireless Networks. IEEE Transactions on Multimedia, Institute of Electrical and Electronics Engineers, 2007, 9, pp.386 - 396. 10.1109/tmm.2006.886291. hal-03224840

HAL Id: hal-03224840
https://hal.archives-ouvertes.fr/hal-03224840
Submitted on 12 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract—Many protocols optimized to transmissions over wireless networks have been proposed. However, one issue that has not been looked into is considering human perception in deciding a transmission strategy for three-dimensional (3-D) objects. Several factors, such as the number of vertices and the resolution of texture, can affect the display quality of 3-D objects. When the resources of a graphics system are not sufficient to render the ideal image, degradation is inevitable. It is therefore important to study how individual factors affect the overall quality, and how the degradation can be controlled given limited bandwidth resources and possibility of data loss. In this paper, the essential factors determining the display quality are reviewed. We provide an overview of our research on designing a 3-D perceptual quality metric integrating two important ones, resolution of texture and resolution of mesh, that control the transmission bandwidth requirements. A review of robust mesh transmission considering packet loss is presented, followed by a discussion of the difference in existing literature with our problem and approach. We then suggest alternative strategies for packet transmission of both 3-D texture and mesh. These strategies are then compared with respect to preserving 3-D perceptual quality under packet loss.

Index Terms—3-D transmission, packet loss, perceptual quality.

I. INTRODUCTION

A n important consideration in designing effective interactive online 3-D systems is to adaptively adjust the model representation, while preserving satisfactory quality as perceived by a viewer. While most research in the literature focuses on geometric compression [33] and use only synthetic texture or color, we address both geometry resolution and realistic texture resolution, and analyze how these factors affect the overall perceptual quality. Our analysis is based on experiments conducted on human observers. The perceptual quality metric derived from experiments allows the appropriate level of detail (LOD) to be selected given the computation and bandwidth constraints. Detailed surveys on simplification algorithms can be found in [26], [28], [47], [48], [53], [67], [70]. In early research terrain model and height fields [30], [40] were useful for flight simulation applications, which require an aerial view of the scene. Hierarchical approaches have been proposed in which regions are subdivided recursively forming a tree-like hierarchy such as R-Simp and BSP-Tree [15], [76]. Refinement methods in 3-D start with a minimal approximation on a set of selected points and apply multiple passes. In each pass, the set is split and the region is re-triangulated until the final high-resolution triangulation is reached. An early refinement technique can be traced back to Douglas’ algorithm on two-dimensional (2-D) curve simplification [35]. Fowler applied a hill-climbing technique to locate a candidate point to insert into the triangulation [39]. However, their approach may fail to find the global maximum within the mesh. Schmitt used a two-stage split-and-merge process [75]. Differing from the above techniques based on geometric metric, perceptually driven simplification methods are guided by human perception and quality preservation [29], [59]. Vertices are removed only if they are imperceptible and do not degrade the visual quality. Most perceptually driven techniques in the literature are designed for view-dependent visualization [54], [71], [85]. Many simplification techniques involve relocation of vertices and thus online transmission cannot be incremental [15], [43], [76], [83]. In the progressive meshes method, although the original mesh can be recovered exactly after all data are received, the edge collapse transformation creates new vertices and the \textit{vsplit} record stream increases network workload [47]. The adaptive real-time LOD technique also involves vertex relocation [88].

Perception of depth and realistic texture are the main factors to achieve realism and visual fidelity in the virtual world. In recent years, researchers started to incorporate color and texture into their mesh simplification models. When texture is mentioned in the literature, it often refers to synthetic or animated texture [82]. Synthetic texture or per pixel color stored in each vertex [29], [44], [77], [78] can be estimated or interpolated. For example, when walking through an animated scene, the next frame can be predicted based on available neighboring data [27]. Using interpolated or animated texture is a compromise in applications, which require fast interactive rendering. For applications requiring real life texture, interpolating color or estimating pattern between vertices is not acceptable. Photo-realistic texture maps are used in [91], but their effort is on recovering geometry from texture patches retrieved from multiple photographs, and not on generating LOD. A distance-based technique is applied to photo-textured terrain [56]; however, color interpolation between pixels is necessary in their technique to avoid blocky appearance of terrain texture. A tool called "Metro" was proposed in [31] for comparing a pair of simplified surfaces. However, the tool cannot be directly used to develop a perceptually optimized 3-D model transmission strategy. Interactive transmission of 3-D scenes was considered in [92]; however, the strategy considers rendering views on the server, rather than

IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 9, NO. 2, FEBRUARY 2007
view-independent texture and mesh transmission. Space optimized texture maps were discussed in [13]; however, issues relating to perceptual quality were not considered in this work. In related research, the issue of 3-D watermarking of meshes was considered in [42] and the approach was optimized based on subjective evaluations.

Simplification algorithms try to control the complexity of a mesh by developing various strategies for simplifying the LOD in different parts of a 3-D object. In order to easily control the simplification parameters on a 3-D object we will follow a simple model approximation strategy based on multi-resolution representation with photo-realistic texture and mesh. An example of geometric simplification is shown in Fig. 1, in which a Nutcracker toy model is simplified to various resolution levels (number of triangles is 1260 left, 950 middle, and 538 right).

One of the major drawbacks with most 3-D transmission algorithms is that they do not consider loss of data. Wireless communication necessitates addressing this issue. There are many wireless protocols that have been proposed in the last decade, including Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Indirect-TCP (I-TCP) [5], [89], Mobile TCP (M-TCP) [11], Fast-Retransmit Approach [25], SnooP Protocol [9], [10], Explicit Bad State Notification (EBSN) [7], Link-Level Retransmissions [4], New Reno [46], Selective Acknowledgments (SACK) [38], Detection of Out-of-Order and Response (DOOR) [87], Hierarchical Cache Design with New SnooP [45], TCP with Delayed Congestion Response (TCP-DCR) [12], and Wireless TCP (WTCP) [68]. Many of the proposed strategies are aimed at improving the shortcoming of TCP in invoking congestion control mechanisms for every packet loss. For wireless networks, where packet loss occurs as a result of unreliable links and route changes, the TCP strategy leads to further delays and degradation in transmission quality because packet re-transmission can cause further congestion and delays. Even though issues of multimedia transmission over wireless networks have received attention [36], [41], [86], relatively little work has been done addressing wireless 3-D transmission. In recent research, approaches for robust transmission of mesh over wireless networks [1]–[3], [22] have been outlined. However, these methods do not take joint texture and mesh transmission into account. Also, in [2], [22] it is assumed that some parts of the mesh can be transmitted without loss over a wireless network allowing progressive mesh transmission to give good results. However, this assumption implies implementing a special standard with a combination of UDP and TCP protocols, which in general cannot be guaranteed in an arbitrary wireless environment. Special models for packet loss probability have been developed by other researchers [52]. However, these models are usually associated with requirements such as retransmission. To keep our study applicable in an unrestricted ad hoc wireless environment, we simply assume packet-based transmission where a certain percentage of the packets may be lost. In this scenario, we compare how various types of 3-D transmission strategies fare, and how to take perceptual quality into account in designing a better strategy. We consider packet loss, rather than bit errors, over wireless networks using a UDP type protocol and try to avoid the problem of packet retransmission that can result in further congestion. In general, it would be interesting to look into FEC type strategies for increasing error resiliency in textmesh transmission for 3-D models. However, when packets are lost, as opposed to bits having errors, FEC coding may not be able to recover lost packets while requiring additional bandwidth for the error correcting bits.

The remainder of this paper is organized as follows: Section II reviews past work on perceptual quality evaluation and discusses how to relate bandwidth with texture and mesh reduction considering perceptual quality. Section III examines possible strategies for 3-D image transmission and analyzes which one is most suitable for optimizing perceptual quality under packet loss. Some experimental results are outlined in Section IV. Finally, conclusion and future work are summarized in Section V.

II. 3-D PERCEPTUAL QUALITY OPTIMIZATION

In the area of image compression, Mean Square Error (MSE) is commonly used as a quality predictor. However, past research has shown that MSE does not correlate well to perceived quality based on human evaluation [61]. Since this study, a number of new quality metrics based on the human visual system have been developed [32], [34], [55], [79], [80]. Limb originally looked at fitting an objective measure that closely estimated impairment ratings on five test pictures. A number of perception-driven rendering algorithms were developed to incorporate the Human Visual System (HVS) as a factor to compute global illumination so as to improve perceptual accuracy [8], [37]. A detailed overview of various issues in perceptually adaptive graphics can be found in [64].

Various factors affecting perceptual quality including Geometry, Texture, Shading, Polish, Frame Rate, Distance, Visual Masking and Adaptation, and Foveation [14], [37], [49], [55], [61], [62], [73], [74], [81] have been reviewed in our past work [65]. We will assume that factors other than texture and geometry resolution are fixed during perceptual evaluations. We consider only these two factors since they dictate the bandwidth necessary for transmission.

In recent years, perceptually adaptive graphics [64] has received increasing attention in the graphics and visualization community. In EUROGRAPHICS 2000, a state-of-the-art report was presented on visual perception [60]. A group of researchers from computer graphics, psychology and other disciplines gathered in 2001, as a result of the joint effort between EUROGRAPHICS and SIGGRAPH, to discuss the importance of human perception when striving for realism in the virtual world [54], [71], [84], [85]. More effort has been expended on verifying geometric error estimation with perceptual evaluation experiments in order to achieve higher visual fidelity of 3-D display. Most perceptually driven techniques developed so far focus on view-dependent rendering. These techniques can be applied to dynamic scenes [63], [71], and can be used to

Fig. 1. Nutcracker toy model at various mesh resolution levels.
compute the relative resolutions between the region of interest and the periphery [6], [71]. In order to achieve higher visual quality, user-guided simplifications were also suggested [50], [69]. By contrast, our approach is view-independent, applied to relatively static 3-D objects and does not need user intervention when predicting visual quality.

A. Review of Perceptual Metric Used

Five 3-D objects (Doll, Nutcracker, Pot, Head and Dog) were used as stimuli in the experiments. These objects were acquired with the Zoomage 3-D scanner. Fig. 2 illustrates the scanning process, and Fig. 3 shows the texture, mesh, and canonical view of the Nutcracker object. The other objects (dog, doll, head, and pot) used in the experiments are shown in Fig. 4.

The participants (judges) were asked to compare the target stimulus with the two referential stimuli and assign it one of the following ratings: very poor (1), poor (2), fair (3), good (4), very good (5).

Fig. 5 illustrates two referential stimuli (left and right) and one target stimulus (center) in the experiment.

 Considering perceptual evaluations, we observed that:

- perceived quality varies linearly with texture resolution (Fig. 6, left);
- perceived quality varies following an exponential curve for geometry (Fig. 6, right). (We consider an exponential, rather than a high degree polynomial, curve in order to have only a few parameters to estimate. Also, with several parameters in a polynomial there is likely to be significant variations in the parameters’ values for small variations in the types of objects.)

Scaling the texture \((t)\) and geometry \((g)\) between 0 and 1, it can be shown that:

\[
Q(g, t) = \frac{1}{\frac{1}{m+(M-m)t} + \left(\frac{1}{m} - \frac{1}{m+(M-m)t}\right)(1-g)^c}
\]  

where \(m\) and \(M\) are, respectively, the minimum and maximum ratings, and \(c\) is a constant.

Details of the perceptual evaluations and metric derivation can be found in our prior work [65]. Important issues relating to the perceptual evaluation process, such as number of subjects, reliability of evaluations and factors influencing the evaluation process are described in [65] and are thus skipped here. Other research and approaches from our group on issues related to perceptual evaluations can be found in [16], [23]. Note that the quality value varies in the range of \(1/(m)\) to \(5(M)\), the range of values allowed in the perceptual ratings.

B. Relating Perceptual Metric to Bandwidth

Consider now that \(b\) is the estimated total bandwidth for the transmission time interval, \(T\) is the texture and \(G\) is the geometry file sizes, possibly compressed, at maximum resolution. We assume that as the texture (or geometry) is scaled by a factor \(t\) (or \(g\)) in both dimensions the corresponding file sizes get reduced to \(t^2T\) (or \(g^2G\)). This is equivalent to assuming that the compression method scales linearly based on the dimensions of texture or geometry; a simplification that needs to be modified in future work based on the scaling functions with respect to size of different texture and mesh compression methods that may be used. For \(b < T + G\), to utilize the bandwidth completely we must have:

\[
b = t^2T + g^2G.
\]  

Given \(b\) we can choose the relative proportion of texture and mesh to create a 3-D model in many different ways, as long as (2) is satisfied. The question is “What is the optimal choice maximizing perceptual quality?” Considering \(m = 1, M = 5, and
and values in the range can both be equal to 1 and quality can.

Mbits (total bandwidth over a 10

Mbits (total bandwidth over a 10

For a 3-D object optimum

texture and mesh and thus need to

be observed that the optimal

graph of (4) for varying $t$ for this case is shown in Fig. 7. It can

be pre-computed for a discrete number of $b$ values in the range

$[0, T + G]$ to allow fast selection of a perceptually optimized

model in an online application.

Example 1: Let $b = 100$ Mbits (total bandwidth over a 10 sec. interval, say)

Suppose that we have a 3-D model with overall texture size
(say a JPEG image size) equal to 20 Mbits and mesh size (e.g.,
a “.obj” file size) of 10 Mbits. Also, assume that this model is
similar to a class of objects that follow the perceptual quality
curve in (3). Thus, $T = 20$ Mbits and $G = 10$ Mbits.

For this example $t$, $g$ can both be equal to 1 and quality can be
equal to 5 (the maximum) in (3). That is, we can transmit the
entire model without the need for any tradeoff between texture
and mesh components.

Example 2: Suppose that we have the same 3-D model as in
Example 1, but that the bandwidth is much lower at 10 Mbits.
Thus, $b = 10$ Mbits, $T = 20$ Mbits and $G = 10$ Mbits.

In this case $t$ can vary in the range $[0, \sqrt{10}/20] = [0, 0.707]$ so that (2) can be satisfied. That is, we cannot transmit all
the texture and mesh and thus need to find the best compromise. The
graph of (4) for varying $t$ for this case is shown in Fig. 7. It can
be observed that the optimal $t$ is close to 0.54 in this example.

Example 3: Let $b = 12$ Mbits, $T = 20$ Mbits, $G = 10$ Mbits.

In this case $t$ can only vary in the range $[\sqrt{2}/20, \sqrt{10}/20] =
[0.316, 0.707]$ so that (2) can be satisfied. The graph of (4) for
varying $t$ for this case is shown in Fig. 8. The optimal value of $t$
is close to 0.6 for this example.

Let us consider some examples of the optimization.

In general, given $t, G$ and $T$ are parameters.

Example 1: Let $b = 100$ Mbits (total bandwidth over a 10 sec. interval, say)

Suppose that we have a 3-D model with overall texture size
(say a JPEG image size) equal to 20 Mbits and mesh size (e.g.,
a “.obj” file size) of 10 Mbits. Also, assume that this model is
similar to a class of objects that follow the perceptual quality
curve in (3). Thus, $T = 20$ Mbits and $G = 10$ Mbits.

For this example $t$, $g$ can both be equal to 1 and quality can be
equal to 5 (the maximum) in (3). That is, we can transmit the
entire model without the need for any tradeoff between texture
and mesh components.

Example 2: Suppose that we have the same 3-D model as in
Example 1, but that the bandwidth is much lower at 10 Mbits.
Thus, $b = 10$ Mbits, $T = 20$ Mbits and $G = 10$ Mbits.

In this case $t$ can vary in the range $[0, \sqrt{10}/20] = [0, 0.707]$ so that (2) can be satisfied. That is, we cannot transmit all
the texture and mesh and thus need to find the best compromise. The
graph of (4) for varying $t$ for this case is shown in Fig. 7. It can
be observed that the optimal $t$ is close to 0.54 in this example.

\[ Q(g, t) = \frac{1}{1 + 4t} + \left(1 - \frac{1}{1 + 4t}\right)(1 - g)^{2.7}. \] (3)

Maximizing (3) is equivalent to minimizing the inverse of this
equation; considering this and (2), optimizing quality reduces to
minimizing:

\[ Q_{b,G,T}(t) = \frac{1}{1 + 4t} + \left(1 - \frac{1}{1 + 4t}\right) \left(1 - \sqrt{\frac{b - t^2T}{G}}\right)^{2.7}, \] (4)

where $b$, $G$ and $T$ are parameters.

In this case $t$, $g$ can both be equal to 1 and quality can be
equal to 5 (the maximum) in (3). That is, we can transmit the
entire model without the need for any tradeoff between texture
and mesh components.

Example 2: Suppose that we have the same 3-D model as in
Example 1, but that the bandwidth is much lower at 10 Mbits.
Thus, $b = 10$ Mbits, $T = 20$ Mbits and $G = 10$ Mbits.

In this case $t$ can vary in the range $[0, \sqrt{10}/20] = [0, 0.707]$ so that (2) can be satisfied. That is, we cannot transmit all
the texture and mesh and thus need to find the best compromise. The
graph of (4) for varying $t$ for this case is shown in Fig. 7. It can
be observed that the optimal $t$ is close to 0.54 in this example.

Example 3: Let $b = 12$ Mbits, $T = 20$ Mbits, $G = 10$ Mbits.

In this case $t$ can only vary in the range $[\sqrt{2}/20, \sqrt{10}/20] =
[0.316, 0.707]$ so that (2) can be satisfied. The graph of (4) for
varying $t$ for this case is shown in Fig. 8. The optimal value of $t$
is close to 0.6 for this example.

In general, given $T$ and $G$ for a 3-D object optimum $t$ can
be pre-computed for a discrete number of $b$ values in the range
$[0, T + G]$ to allow fast selection of a perceptually optimized
model in an online application.
III. PERCEPTUALLY OPTIMIZED TRANSMISSION

To simplify the model of wireless transmission, we assume that data is sent in packets of equal size and there is a possibility that a certain proportion of these packets may be lost. Various protocols [4], [25] suggest re-transmission approaches in case of packet loss; however, re-transmission is not conducive to time bound real-time applications, such as 3-D visualization for online games. We consider several possible strategies for packet construction in wireless 3-D transmission, and then analyze the pros and cons of each.

Strategy A

Packets are formed by breaking up the 3-D image into fragments, where a fragment contains data from a connected (for simplicity a rectangular) region.

Limitations of Strategy A

This strategy is simple to implement, however, missing packets can create unacceptable voids in parts of objects.

Strategy B

Progressive transmission of 3-D data in packets; i.e., initial packets can store base layers (as in JPEG2000 image or Progressive Mesh [47], [51] transmission) and later packets can contain detailed model information.

Limitations of Strategy B

This strategy follows excellent research on simplification and can be made compatible with recent image and video coding standards [66]. The main drawback lies in the necessity to receive packets at the base and lower levels of a hierarchy before packets at higher levels can become useful. A packet lost at the base layer, for example, would make packets received from subsequent layers of little use.

Strategy C

Robust Progressive transmission of 3-D data in packets, by transmitting multiple copies of the base layer packets.

Limitations of Strategy C

This approach reduces the possibility of missing data in more important layers. For example, if the probability of packet loss is 10%, then if duplicate copies of all base layer packets are transmitted the chances of missing data at the base layer becomes $0.1^2$, i.e., 1%. The weakness of the method lies in the need to send redundant data, thereby increasing bandwidth requirements, and the lack of quality in the case where an original as well as its duplicate packet gets lost. Also, base layer packets need to be received before other packets, which cannot necessarily be guaranteed in an ad hoc wireless network.

Strategy D

3-D Partial Information Transmission (3PIT): In this approach we break up the texture and mesh into packets by subsampling into overlapping but nonidentical components. At the client site the overall texture and mesh are reconstructed based on interpolation from the received packets. An implementation of this approach is given in the following algorithm:

SERVER SITE:

$T$: original texture;

$M$: original mesh, in a regular form allowing easy subsampling;

Construct $T_1, T_2, \ldots , T_n$ by regular, nonidentical subsampling of $T$;

(Comment: For example, given a $100 \times 100$ pixel texture $T$; we can construct $T_1, T_2, \ldots , T_{16}$ by defining $T_1$ as $T(0 + 4i, 0 + 4j), i, j = 0, \ldots , 24; T_2$ as $T(0 + 4i, 1 + 4j), i, j = 0, \ldots , 24; \ldots , T_{16}$ as $T(3 + 4i, 3 + 4j), i, j = 0, \ldots , 24).$

Construct $M_1, M_2, \ldots , M_n$ by regular, nonidentical subsampling of $M$;

Form packets $P_1, P_2, \ldots , P_n$ where $P_i = T_i + M_i; i = 1, \ldots , n$, with header and subsampling information added to each packet;

Transmit $n$ packets to a client on request, possibly in a randomized order;

CLIENT SITE:

Request server to transmit a 3-D object;

Receive packets from server;

Uncompress mesh and texture data stored in this packet;

Set up initial display based on first packet received and interpolation information stored in header;

Update display based on next packet received;

Limitations of Strategy D

One of the shortcomings of this approach is that the texture and mesh data receive equal importance; i.e., the same fraction of each is transmitted in a packet. The perceptual quality analysis in the last section shows that for optimizing perceptual quality the relative importance of texture and mesh can vary depending on the available bandwidth; this issue is not taken into account in Strategy D.

Strategy E (Method Adopted)

3-D Perceptually Optimized Partial Information Transmission (3POPI T): This approach extends 3PI T by taking perceptual quality into account. The algorithm modifies Strategy D by a bandwidth estimation step followed by perceptually optimized packet creation. Details are described below:

SERVER SITE:

$T$, $M$: as for Strategy D;

Receive bandwidth estimate ($B_e$) and estimated loss proportion ($L$) from requesting client;

Compute server transmitting bandwidth: $B_s = B_e/(1 - L)$;
Compute optimum texture and geometry scaling factors $t_e$ & $g_e$ following procedure for minimizing (4) in the last section, considering bandwidth to be $B_e$;

Compute scaled texture ($T_s$) and mesh ($G_s$), assuming transmitting bandwidth $B_s$ based on factors $t_e$ & $g_e$;

(Comment: Specifically $T_s = \left( \frac{1}{g_e^2} \right) T$ and $G_s = \left( \frac{1}{g_e^2} \right) G$; with texture and mesh possibly being interpolated to higher than the current maximum size in case the scaling factors are greater than 1.) Construct $T_{s1}, T_{s2}, \ldots, T_{sn}$ by regular, nonidentical subsampling of $T_s$;

Construct $M_{s1}, M_{s2}, \ldots, M_{sn}$ by regular, nonidentical subsampling of $M_s$;

Form packets $P_1, P_2, \ldots, P_n$, where $P_i = T_{si} + M_{si}$; $i = 1, \ldots, n$, with header and subsampling information added to each packet;

(Comment: Number of packets $n$ is chosen based on prior decision on packet size.)

Transmit $n$ packets to a client, possibly in a randomized order;

CLIENT SITE:

Request server to transmit a 3-D object;

Receive packets from server for bandwidth estimation;

Estimate bandwidth ($B_e$) based on number of packets received [90] in a certain time interval and estimate loss proportion ($L$);

Receive packets from server containing partial data on the 3-D object;

Uncompress mesh and texture data stored in this packet;

Set up initial display based on first packet received and interpolation information stored in header;

Update display based on next packet received;

Comments on Strategy E

On first observation it may appear that this strategy does not take packet loss proportion ($L$) into account in the transmission strategy. However, in reality this is not the case. Without any packet loss, the transmission bandwidth ($B_e$) would be used to compute the optimum texture and mesh scaling factors. When packets are lost the remaining packets may not be perceptually optimal for the effective bandwidth after packet loss. We thus form packets that are optimal at a lower bandwidth ($B_e$).

Our algorithms are intentionally designed without the addition of redundant packets, since there is no way to guarantee that an original as well as its corresponding redundant packets are not lost. Also, addition of redundant packets incrases bandwidth requirement thereby lowering performance with packet loss compared to lossless transmission at the effective bandwidth.

We can consider that perceptually adaptive redundancy is added into the algorithm in Strategy E based on the estimate of packet loss. However, it should be noted that we do not need to transmit duplicate packets based on acknowledgements and that the trade-off between texture and mesh is taken into account.

One of the drawbacks of Strategy E is the need to estimate bandwidth and packet loss ratio. This estimation based transmission may not be practical where feedback from client to a server is not reliable, or for multicasting over heterogeneous networks with varying packet loss and bandwidths. This issue needs to be addressed in future research.

Notes on the Implementation of Strategy E

Given a texture (image) and mesh (structure) file we need a program to create subsampled files for each type of data. These subsampled texture and mesh files, which could be named by the pixel or mesh locations selected in different blocks, then needs to be put together in packets along with other header information. The header needs to include information that allows the relative locations of subsampled files included in that packet to be identified. To allow a packet to be decoded on its own, without additional information from other packets, it is necessary to include certain global information (like the size of the overall texture and mesh) in every packet. In order to speed up processing and visualization speed it is useful to compute many subsamples beforehand and store in the server. Similarly, interpolation using information from packets can be speeded up by using look-up tables that store the weights for the interpolating method used, rather than computing these weights every time. Look-up table based interpolation has been used in the past for real-time foveated videoconferencing [14].

IV. EXPERIMENTAL RESULTS

We show some preliminary implementations towards deploying 3POPIT over a lossy wireless network. Two programs are shown: (i) Combining and interpolating based on various texture and mesh subsamples and (ii) Comparison of perceptually optimized versus nonoptimized transmission. Note that our approach is consistent with recommendations in MPEG-4 [66], with the novelty lying in perceptual optimization depending on available bandwidth and packet loss. Also, JAVA3D based implementation and MPEG-4 compatibility makes platform independent [58] deployment possible.

A. Combining and Interpolating 3-D Models Based on Subsampled Packets

Fig. 9 shows the effect of receiving and combining 1, 2, 4 and 8 of 16 subsamples of the standard Lena texture. The interpolation strategy used was based on weighting depending on distances of up to four of the closest neighbors of a missing pixel. We also observed that a fixed structure of packet loss, e.g., first boxes checked in first and third rows & fourth boxes checked in second and fourth rows on interface in Fig. 9 top right corner, produced noticeable distortions in image reconstructed after interpolation; by contrast, random set of packets lost often produced better results.

Fig. 10 shows the effect of receiving and combining 2, 4 and 8 of 16 subsamples of the nutcracker mesh. Note that results may
Fig. 9. Interpolating and reconstructing Lena image; top row shows online interface with original image (left), transmitted packets displayed with missing pixels (right), and interpolated image (middle). Middle row shows reconstructed images when 1 (left), 4 (middle), and 8 (right), of 16 packets are received. Bottom row shows close up of right part of the hat in the image when 1 (left), 4 (middle), and 8 (right) of 16 packets are received.

Fig. 10. Interpolating and reconstructing mesh of nutcracker model when 2 (left), 4 (middle), and 8 (right) of 16 packets are received.

vary from one execution to another for a random percentage of packet loss.

Fig. 11 shows the effect of optimized versus nonoptimized transmission on perceptual quality. Two versions (top) and (bottom) of the same model are shown, with the mesh on the left and the texture mapped on the right. Although the texture and mesh together for the top and bottom models use nearly the same bandwidth, 125 and 134 Kb, respectively, the top one is favored by most viewers based on perceptual experiments.

B. Comparison of Results With Perceptual Optimization

We now show some results with a user interface that allows 1 to 16 out of 16 packets to be selected, or a random percentage of packets to be lost. The packets received are indicated by check marks in the square boxes on the right. Selecting the repair texture box ensures that missing texture pixels are interpolated, otherwise the un-interpolated texture is mapped.

Fig. 11. Two representations of the Nutcracker texture + mesh models: Left has lower quality mesh, requiring 125 Kb total bandwidth, and higher perceptual quality; Right has higher quality mesh, and lower quality texture requiring a total bandwidth of 134 Kb, but has lower perceptual quality.

Fig. 12. Actual texture + mesh (texmesh) model after 2 of 16 packets are received.

Fig. 13 shows interpolated texmesh models when the transmission is optimized for LOW packet loss: after, respectively, (a) 2, (b) 4, (c) 8, and (d) 12 of 16 packets are received. Observe that the perceptual quality continues to improve as more packets are received.

Fig. 14 shows interpolated texmesh models when the transmission is optimized for HIGH packet loss: after, respectively, (a) 2, (b) 4, (c) 8, and (d) 12 of 16 packets are received. Note that the difference in perceptual quality between (b) and (d), when few and most packets are received, respectively, is not significant in this case.

Fig. 15 compares the texture-mapped models and the mesh when transmission is optimized for high and low packet loss.
In this example, most of the packets are received. Thus, it is expected that the transmission optimized for low packet loss will have better perceptual quality. Observe that the rendered model at the top left looks clearer and structurally very similar to the shape at the top right. The bottom row shows the meshes for the two representations. The mesh on the right is denser because the representation is optimized for higher packet loss, thus it allows for redundancy in structural information in case most of the packets are lost. However, given the same overall (texture + mesh) storage, redundancy in mesh makes the texture of lower quality, resulting in lower perceived quality compared to the representation in the left column.

Fig. 16 compares the two representations discussed in Fig. 15 in case of high packet loss. In this case the lack of redundancy in the mesh representation on the left makes the structure look unacceptable for the shape at the top left.

In order to consider arbitrary meshes, we need to consider mesh coding and connectivity. Current 3-D mesh coding techniques mainly focus on coding efficiency, i.e., compression ratio, by transmitting incremental data. This approach is good without packet loss but is vulnerable to channel errors for irregular meshes. Fig. 17 shows an example of error sensitivity of the Edgebreaker 3-D mesh coding method [57], [72]. With one error character in the connectivity stream, the decoded mesh can change significantly and can be impossible to reconstruct.

In Fig. 18, 0%, 30%, 50%, 60% and 80% randomly selected packet loss was again imposed on a Queen mesh and texture. However, the lost geometry was interpolated based on neighboring vertices and valence or connectivity information which is constant for most vertices in a regular mesh. It can be seen that
smoothness on the object surface begins to deteriorate at about 60% packet loss. Visual degradation becomes more obvious at 80% packet loss.

The benefit of adding perceptual optimization during packet loss can be seen in Fig. 19. The model on the right is perceived to be closer to the original, though both have 80% loss. Details of strategies for arbitrary 3-D model transmission under packet loss will be discussed in future work.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we reviewed factors controlling 3-D image degradation and outlined an approach for estimating perceptual quality considering variations in mesh and texture resolutions. A theoretical framework for determining the relative importance of texture versus mesh was presented. An approach to optimizing perceptual quality under packet loss was then outlined. Experimental results validate our approach.

We are currently working on implementing our approach on wireless handheld devices which have recently become much more powerful in processing power with much larger RAMs as well. Also, the preliminary implementation is not optimized for fast computation of interpolated values. The most computationally efficient approach for the interpolation would be to pre-determine neighbors and coefficients for interpolation, given partial packet transmission, and store various look-up tables. Using lookup tables, however, requires larger RAMs for handheld devices.

The packet loss model used is rather simple and is meant to illustrate the feasibility of our method. More realistic models incorporating burst error models in wireless networks need to be considered in future work. Also, we do not consider packet size and header length. Some preliminary work by our group incorporating these two parameters will be presented in a conference [24]; however, substantial work still needs to be done to test the influence of these factors over real wireless networks.

Our initial approach is based on a simple multi-resolution approach to mesh and texture reduction. A more advanced and systematic method could be based on joint texture-mesh simplification following a scale-space analysis [17]. We will investigate this direction in future research.

ACKNOWLEDGMENT

The authors gratefully acknowledge the assistance of M. Bates and L. Ying in the implementations.

REFERENCES


Irene Cheng is the Chief Scientist of the Online Multimediaucation Project sponsored by iCORE and Castle Rock Research and the 4D Medical Imaging project sponsored by Alberta Science and Research in the Department of Computing Science at the University of Alberta, Edmonton, AB, Canada. She has over ten years of R&D experience in industry, including Lloyds Bank and Zoomage, working on various information technology and 3-D vision systems. One of her unique contributions is to incorporate human perception—Just-Noticeable-Difference (JND)—with scale-space analysis, systematically based on psychophysical methodology, to improve 3-D simplification and transmission techniques. By applying the JND approach, automatic selection of scale based on a given object dimension is possible. Her research includes multimedia modeling, visualization and transmission, perceptual issues, and multimedia database structure. Her current research also covers optimizing transmission over unreliable networks taking packet loss into consideration. She has over 40 publications in international conferences and journals. Dr. Cheng has received numerous Scholarships and Fellowship from NSERC and iCORE among others.

Anup Basu received the B.S. degree in mathematics/statistics and the M.S. degree in computer science from the Indian Statistical Institute. He received the Ph.D. degree in computer science from the University of Maryland, College Park.

He was with Tata Consultancy Services, New Delhi, India, and Biostatistics Division, Strong Memorial Hospital, Rochester, NY. He is currently a Professor at the University of Alberta, Edmonton, AB, Canada. He is also an iCORE—Castle Rock Research—SUN Industrial Research Chair in Multimedia. He has published over 130 papers, patents, books and book chapters, in leading conferences and journals. He has also helped startup several technology companies. His current research interest includes Computer Vision, Graphics, 3-D Multimedia and Multimedia Communications over Heterogeneous Networks. He pioneered the use of foveation in image and video compression and stereo visualization in 1992–1993. He also introduced several novel panoramic image sensors in the mid-1990s. These research directions have been subsequently pursued by many researchers in leading institutes around the world.