A Visual Quality Prediction Model for 3D Texture

Irene Cheng and Pierre Boulanger
Department of Computing Science, Univ. of Alberta, Edmonton, CANADA,{lin,pierreb}@cs.ualberta.ca

Abstract

Online bandwidth limitations and fluctuations impose a major challenge in estimating the amount of data to transmit in a given time period. Over or under estimation of bandwidth can jeopardize the visual fidelity of the transmitted and related multimedia data. We propose a Visual Quality Prediction (VQP) model which supports an adaptive fragmented texture transmission approach taking bandwidth fluctuations into consideration, and adjusts the data size (and thus quality) of the next block of texture data to be transmitted. The transmission depends on a set of predictors used to optimize an overall best effort visual quality.

1. Introduction

Differing from previous approaches which focus mainly on 2D and grayscale images, in this paper we propose a VQP model which takes both 3D and 2D properties of color texture into consideration. Since high-resolution color texture images are much larger than the mesh data, we focus on supporting bandwidth adaptation using texture reduction, which can be associated with an efficient level-of-detail (LOD) algorithm. In order to achieve satisfactory interactivity, applications such as online games use synthetic texture which is often simple and easy to duplicate so that the transmitted data can be small. Our model is designed for applications such as displaying museum exhibits and medical images, where high-resolution real texture is required.

Visual quality models have been discussed in the literature [OHM*04], but to the best of our knowledge, none of them addresses visual fidelity in a systematic manner, taking bandwidth limitation and fluctuation, together with 3D and 2D texture properties into consideration. A perceptual approach was used, by approximating the Contrast Sensitivity Function (CSF), to simplify details in a scene [Red01]. The metric derived from the CSF was used to perform simplification operations [LH01]. Prior approaches were improved by accounting for textures and dynamic lighting [WLC*03]; their focus was on mesh and not texture simplification. Our view-independent approach analyzes the intrinsic property of the texture image, independent of viewing direction. An image fidelity assessor was discussed in [TPA98]; their technique accepts two grayscale images as input and outputs a distortion value, while our technique is applicable to color as well as grayscale images, and predicts the overall visual fidelity of 3D objects. Two visual fidelity algorithms for mesh simplification were discussed in [WFM01]. Visual difference predictor was used to select the appropriate global illumination algorithm [VMK*00]. The visibility of differences between two images was used to determine whether a particular area of a synthetic scene needed refinement [BM98]. Unlike our approach, these techniques are not designed to guide real texture reduction at multiple scales.

The rest of this paper is organized as follows: Section 2 explains how the VQP model can support adaptive fragmented texture transmission. Section 3 discusses the 3D and 2D properties associated with a texture image. Section 4 introduces the VQP computational model and discusses how to apply it in online visualization. Finally, Section 5 concludes the work and outlines future directions.

2. Bandwidth adaptation using fragmented texture

Bandwidth fluctuation is a major challenge in online visualization. Since an exact bandwidth is not available before transmission, the size and thus the quality of data to transmit have to be predicted. Over or under estimation often results in inefficient use of limited resources. The goal of an adaptive strategy is to absorb the fluctuations periodically. We proposed absorbing bandwidth fluctuation by dividing texture data into fragments and altering the quality of the fragments not yet transmitted based on current bandwidth [CB04b]. While absorbing the time surplus (or deficit), by increasing (or decreasing) the fragment data size, the challenge is how to assign different qualities to these fragments so that a best-effort overall quality can be obtained based on given limited resources.

One important concern is whether fragmentation will increase the total data transmitted. We used 5 images of different resolutions in our experiments. Each resolution has 4 versions containing the number of fragments $n = 2^2$, 4^2, 8^2 and 16^2 respectively. Each version was tested at JPEG quality 100%, 90%, 80%, 60%, 40%, 20% and 10%. It was concluded that, when $n \leq 256$ and each fragment size $\geq 128^2$ pixels, the sum of the fragmented files was smaller than the single non-fragmented file with the same JPEG quality (Fig. 2.1). When this is true at a given quality level, it is also true at lower quality levels. In order to control the data size of fragmented texture, it is therefore important to determine the display device resolution, and thus the texture resolution to be transmitted, in order to compute the number of fragments of optimal dimension.

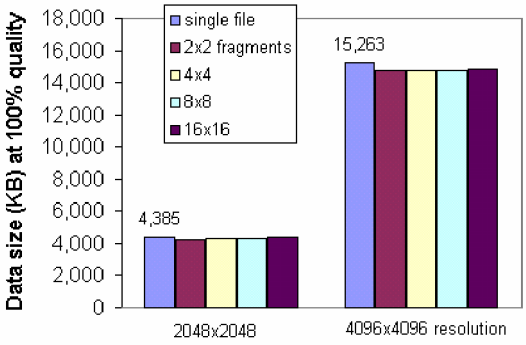


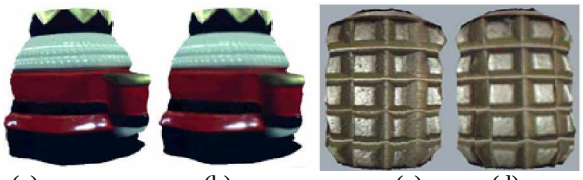
Fig. 2.1: *Experimental results show that the fragmentation approach does not increase the data size transmitted if fragments are divided into optimal dimensions.*

3. Texture reduction driven by 3D and 2D properties

In addition to the 2D properties of a texture image, there are other factors which can affect the visual quality of 3D objects. We classify these factors as geometry driven and texture driven visual predictors.

3.1 Geometry driven visual predictors

Past psychophysical experiments show that contrast induced by the 3D surface property is an important visual cue to predict the resulting quality [Nag84]. In order to represent three-dimensional real world objects on a two-dimensional display device, it is essential to impose the perception of depth and contrast on the human visual system (HVS). A rough surface requires more contrast then a flat surface in order to highlight the surface. The smoothness of a mesh surface is dictated by its underlying geometry, which can be estimated by the feature point distribution generated by a LOD technique [CB05][GH98] [Hop96][HH93]]. If the same texture quality is assigned to the entire surface without taking depth and contrast into consideration, a plain surface may have excessive quality, leaving insufficient bandwidth to more complex surfaces, thereby degrading the overall visual fidelity. An example of how feature point distribution can affect human perception is illustrated in Fig. 3.1. The grenade has vertical structures on the surface, and therefore the feature point distribution is higher than the back of the nutcracker, which is relatively flat. Note that even if the texture quality is reduced to half, there is no significant perceptual degradation on the shiny patch under the belt of the nutcracker. However, the grenade on the right (Fig.3.1d) shows noticeably lower perceptual quality.



(a) (b) (c) (d)

Fig. 3.1: *A snap shot of the nutcracker 3D model (a and b), and the military grenade model (c and d), with original texture quality (a and c), and half of the original texture quality (b and d).*

Feature point distribution is therefore identified as one of the geometry driven visual quality predictors for 3D texture. A high resolution texture image is divided into fragments of optimal dimensions. Higher quality is associated with higher feature point distribution to preserve the surface property and to allow better perception of depth and contrast. In online applications, texture data is made available to heterogeneous client displays of different resolutions. It is a waste of resources if excessive resolution is transmitted and cannot be displayed. As mentioned in Section 2, a solution is to request the display resolution from the client before transmission. Based on the display resolution, the texture of corresponding resolution is selected and fragmented into optimal dimension.

Although feature point distribution is a visual quality predictor, geometry driven texture reduction alone is not sufficient to predict the overall visual fidelity. We will discuss how texture masking and other 2D image properties affect the resulting quality.

3.2 Texture driven visual predictors

In addition to geometry, texture complexity also influences how the HVS perceives quality. Visual masking was suggested in the literature [FP97]. Psychophysical experiments showed that light intensity and contrast affect human perception [LH01] [Red01][WB01][MC95]. The light source can have constant and evenly spread illumination, but the reflective properties vary for different texture patterns. We use the range required for color quantization Z (RGB color model), the texture intensity component I (HSI color model), and the degree of visual masking M induced by the pattern complexity, as our texture driven visual predictors. We assess the performance of these predictors based on their impact on human perception.

The *ZIM* predictors

The texture size can be reduced, by lowering its quality, which means using a smaller quantization range instead of a true color range of 256 values for the red, green and blue components. By analyzing the relative sizes of the color range used to represent different texture patterns, we can predict the relative impact of reducing qualities on these patterns. For a texture with n pixels, Z is computed as follows:

$$Z = \sigma_R + \sigma_G + \sigma_B \qquad (1)$$

σ_R,σ_G and σ_B are the standard deviation for red (R), green (G) and blue (B) color values of the texture respectively.

I is computed as the average intensity (brightness) of the texture, using the standard formula to calculate the intensity of each pixel:

$$I = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{r_i + g_i + b_i}{3}\right) \qquad (2)$$

r_i, g_i and b_i are the red, green and blue color values of each pixel.

The M predictor has a counter effect on the Z and I predictors, and is defined based on two observations: (i) The HVS can discriminate better on brighter surfaces. (ii) Irregular, dense and mixed color patterns on the surface tend to lower the discriminating capability, while the HVS can discriminate better on a plain color surface [FPS*96][FP97]. We define M based on these two observations.

$$M = (1 - (\omega_I (1 - I) + \omega_\zeta \zeta)) \qquad (3)$$

ω_I and ω_ζ are the weights for I and ζ respectively. ζ is the gradient count, computed as follows: The count for two adjacent pixels is either zero or one. When the difference between two adjacent color values, either the red, green or blue component, exceeds a predefined threshold the count is one. Let us define η_{east} and η_{south} as the east and south neighbors of a pixel. Given a threshold Γ and a counter starting from zero, we start from the top left pixel in the texture. If the difference of the red, green or blue value between the pixel and $\eta_{east} \geq \Gamma$, the counter is incremented by one. If the difference between the pixel and $\eta_{south} \geq \Gamma$, the counter is incremented again. The process is repeated for all the $(n-1)*(n-1)$ pixels of the image. Increasing light intensity has a counter effect on ζ. Surfaces with higher M values (lower masking effect) are assigned higher qualities.

The *ZIM* predictors are normalized in the range [0,1]. We analyzed the performance of the *ZIM* predictors using 24 different texture patterns (Fig. 3.2). Starting from the bottom left and moving towards the top right, we notice that the texture patterns change from plain to relatively complex. It is observed that patterns (b) and (c) have the same *ZIM* values, although (b) has a higher Z value (more affected by color quantization) than (c). This is because texture (c) has higher I and M values (brighter and less visual masking effect). Both texture (i) and (j) have irregular patterns, but (i) is darker and can mask better, and thus has a slightly lower *ZIM* value than (j). The sharper edges

and brighter background in texture (s) result in a higher *ZIM* value compared with the edges in (l), (n), (p) and (q). Texture (n) and (p) are extracted from the grenade surface (Fig. 3.1). (p) is slightly brighter than (n) and thus the *ZIM* value is higher. Fig. 3.3 highlights the ζ effect of each texture pattern.

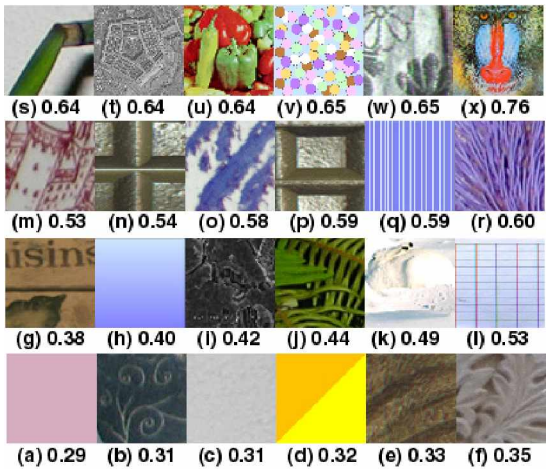


Fig. 3.2: *24 color texture patterns with corresponding ZIM values used in the experiments.*

Fig. 3.3: *Highlight of the* ζ effect by assigning a grayscale value 200 to pixels which have gradient count one, and a value 0 to other pixels.

4. The VQP computation model and application in online visualization

Visual quality is an important consideration when representing high resolution real texture in online 3D applications. In a constrained environment, there is a tradeoff when distributing limited resources, e.g. bandwidth, among different multimedia data. Earlier perceptual experimental results show that after mesh data has reached a certain minimum resolution, allocating resources to increase texture resolution is more beneficial to the overall visual fidelity, than sharing the remaining resources between texture and mesh data [PCB05][RRP00]. This concept is used when transmitting museum data [KTL*04]. In this paper, we introduce a VQP computation model, taking both geometry and texture properties into account, to predict the overall resulting quality.

In the current implementation, we apply equal weights to the predictors. *ZIM* together with the feature point distribution predictor, generated by a LOD technique, are substituted into Equation (4), to predict the overall texture quality of the resulting 3D object,

$$\wp = \sum_{k=1}^{a} \omega_k \rho_k = \sum_{g=1}^{b} \omega_g \rho_g + \sum_{t=1}^{c} \omega_t \rho_t \qquad (4)$$

with $a = b + c$. (5)

Notation

$\wp$ — The overall visual quality predictor.
ρ_k — A visual quality predictor.
ω_k — The weight applied to ρ_k.
ρ_g — A geometry driven visual predictor.
ω_g — The weight applied to ρ_g.
ρ_t — A texture driven visual predictor.
ω_t — The weight applied to ρ_t.
a — Total number of predictors.
b — Number of geometry driven predictors.
c — Number of texture driven predictors

The model can be extended to accommodate more geometry and texture driven visual quality predictors:

4.1 Online visualization

We used texture fragments to absorb bandwidth fluctuations. Instead of applying a uniform quality, each fragment is assigned a different quality based on the associated visual quality predictor $\wp$. The VQP model is designed to support the fragmentation approach. A smaller area of the texture has a more uniform pattern, while the entire texture image taken as a whole is likely to contain surfaces of diversified patterns, making the $\wp$ value more of a global average. In the current implementation, we use the JPEG quality scale 0% to 100% because JPEG images are widely supported on the web and in JAVA applets. However, the VQP concept can be applied to other compression schemes, such as JPEG2000. We used the Intel JPEG encoder and decoder in our experiments. Each LOD generated is assigned a default quality scale Q_i based on viewing distance [CB04a]. Given a visual quality predictor $\wp$, the data size S_i corresponding to a quality scale Q_i is given by:

$$S_i = A(100\wp)^{BQ_i} \qquad (6)$$

Q_i is normalized in the range [0, 1]. A and B are constants for a given $\wp$. Note that the data size S_i is maximum when $Q_i = 1$ and minimum when $Q_i = 0$. We choose an exponential function for the mapping because of the exponential characteristic of the JPEG quality *vs.* file size curve (Fig. 4.1).

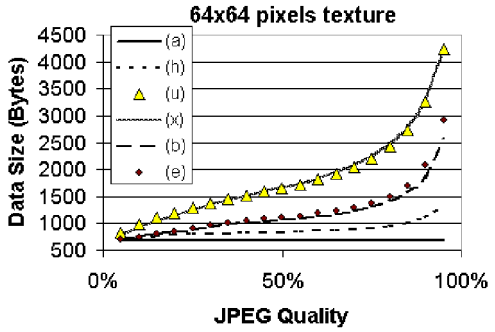


Fig. 4.1: *JPEG file size relates to quality following an exponential function.*

During online transmission, the estimated data size of the next fragment to be transmitted is computed based on the current bandwidth. By selecting several (quality, data size) pairs, constants A and B in Equation (6) can be derived by the curve of the best-fit method. For example, texture (x) has A = 825.45 and B = 0.34. In the quality range [5%, 100%], texture (x) has the

best-fit curve with correlation coefficient of 0.98. To compute constants A and B, we denote $R = (100\%)^B$ and establish the equation:

$$S_i = AR^{Q_i} \quad (7)$$

Constants A and R are solved by the regression method using Equation (8).

$$\log S_i = \log A + Q_i \log R \qquad (8)$$

Constant B is solved by combining Equations (6) and (7), and taking the logarithm on both sides:

$$B = \frac{\log R}{\log(100\%)} \qquad (9)$$

From Equation (8) and (9):

$$Q_i = \frac{\log S_i - \log A}{B \log(100\%)} \quad (10)$$

Equation (10) can then be used to compute the quality based on a given data size. An alternate way to obtain Q_i and S_i is to generate a lookup table during preprocessing and store the (quality, data size) pairs. The most matching value is selected during runtime. Fig. 4.2 illustrates the result of incorporating *ZIM* in the variable quality approach. The texture data is divided into 4 fragments. Suppose the available bandwidth can support 4492 bytes of data. Applying a fixed quality to every fragment, each one can have 30% quality (Fig. 4.2 I). If the variable quality approach is applied, texture (x), (q), (j) and (c) can have 44%, 34%, 27% and 26% quality respectively maintaining a total data size of 4492 bytes (Fig. 4.2 III). The qualities of texture fragment (c) and (j) are lower than 30% but there is no significant degradation in the image. Texture (x) is upgraded from 30% to 44%. Note that a satisfactory quality of the eyes and nose of the baboon is maintained, closer to the original image (Fig. 4.2 II).

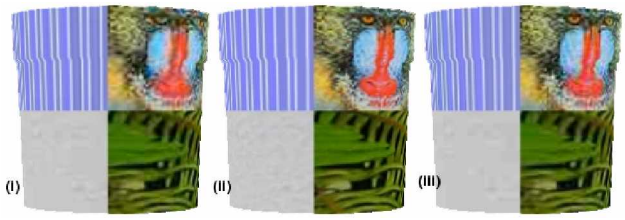


Fig. 4.2: *An example of variable quality assignment based on ZIM: Texture (x) with quality at (II) 100%, (I) 30% and (III) 44% are mapped onto a pot 3D object.*

5. Conclusion and future work

In this paper, we introduced a VQP model, which supports an adaptive fragmented texture transmission approach for 3D objects. Differing from other visual discrimination models in the literature, we incorporate geometry driven as well as texture driven visual predictors taking bandwidth fluctuation into account to best predict the quality of the resulting 3D object. Although the JPEG compression is used in the current implementation, our model can be applied to other compression schemes. The model is extensible to incorporate more predictors if required. In future work, we will assign different weights to different predictors and compare the results. We will also compare the performance of our model with other visual discrimination models for 3D objects.

References

[CB04a] I. Cheng and P. Boulanger, "Perception of scale with distance in 3D visualization," SIGGRAPH 2004 Poster.

[CB04b] I. Cheng and P. Boulanger, "Adaptive online transmission of 3D TexMesh using scale-space analysis," 3DPVT Sept. 2004, Greece.

[CB05] I. Cheng and P. Boulanger, "Feature Extraction on 3D TexMesh using Scale-Space Analysis and Perceptual Evaluation," IEEE Transaction on Circuits and Systems for Video Technology Special Issue 2005.

[BM98] M. Bolin and G. Meyer, "A Perceptually based adaptive sampling algorithm", ACM Siggraph 1998.

[FP97] J. Ferwerda and S. Pattanaik, "A model of visual masking for computer graphics," ACM Siggraph 1997.

[FPS*96] J. Ferwerda, S. Pattanaik, P. Shirley and D. Greenberg, "A model of visual adaptation for realistic image synthesis", ACM Siggraph 1996.

[GH98] M. Garland and P. Heckbert, "Simplifying surfaces with color and texture using quadric error metrics", IEEE Visualization 1998.

[HH93] P. Hinker and C. Hansen, "Geometric optimization," Proc. Visualization, 1993.

[Hop96] H.Hoppe, "Progressive meshes," Proceedings of Siggraph 1996.

[KTL*04] D. Koller, M. Turizin, M. Levoy, M. Tarini, G. Croccia, P. Cignoni and R. Scopigno, "Protected interactive 3D graphics via remote rendering", Siggraph 2004.

[LH01] D. Luebke and B. Hallen, "Perceptually driven simplification for interactive rendering", 12th Eurographics Workshop on Rendering Techniques 2001.

[MC95] A. McNamara and A. Chalmers, "Comparing real & synthetic scenes using human judgements of lightness", 6th Eurographics Workshop on Rendering 1995.

[Nag84] S. Nagata, "How to reinforce perception of depth in single two-dimensional pictures," Proc. Of the SID, Vol. 25/3, 1984.

[OHM*04] C. O'Sullivan, S. Hwlett, R. McDonnell, Y. Morvan and K. O'Conor, "Perceptually adaptive graphics", Eurographics 2004.

[PCB05] Y. Pan, I. Cheng, and A. Basu, "Perceptual quality metric for qualitative 3D scene evaluation," IEEE Transactions on Multimedia, April 2005, Vol.7, No.2.

[Red01] M. Reddy, "Perceptually optimized 3D graphics," IEEE Applied Perception, September/October 2001.

[RRP00] H. Rushmeier, B. Rogowitz and C. Piatko, "Perceptual issues in substituting texture for geometry," Proceeding of SPIE 2000 Vol. 3935.

[TPA98] C. Taylor, Z. Pizlo and J. Allebach, "Perceptually relevant image fidelity," Proc. Of IS&T/SPIE Int'l Symposium on Electronic Imaging Science and Technology 1998.

[VMK*00] V. Volevich, K. Myszkowski, A. Khodulev and E. Kopylov, "Using the visual differences predictor to improve performance of progressive global illumination computation", ACM Transaction on Graphics, Vol. 19, No. 1, April 2000.

[WB01] Z. Wang and A. Bovik, "A human visual system-based objective video distortion measurement system," Int'l Conf. On Multimedia Processing and Systems, Aug. 2001

[WFM01] B. Watson, A. Friedman and A. Mcgaffey, "Measuring and predicting visual fidelity", Siggraph 2001.

[WLC*03] N. Williams, D. Luebke, J. Cohen, M. Kelly and B. Schubert "Perceptually guided simplification of lit, textured meshes," ACM Symp. on Int. 3D Graphics, 2003.