Efficient Error Control in 3D Mesh Coding

D.C. Cernea[1,*], A. Munteanu[1], A. Alecu[1], J. Cornelis[1], P. Schelkens[1] and F. Morán Burgos[2]

[1]*Department of Electronics and Informatics – Interdisciplinary Institute for Broadband Technology,*
Vrije Universiteit Brussel (VUB), B-1050 Brussels, Belgium

[2]*Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, Spain*

[*]`cdcostin@etro.vub.ac.be`

Abstract—**Our recently proposed wavelet-based L-infinite-constrained coding approach for meshes ensures that the maximum error between the vertex positions in the original and decoded meshes is guaranteed to be lower than a given upper bound. Instantiations of both L-2 and L-infinite coding approaches are demonstrated for MESHGRID, which is a scalable 3D object encoding system, part of MPEG-4 AFX. In this survey paper, we compare the novel L-infinite distortion estimator against the L-2 distortion estimator which is typically employed in 3D mesh coding systems. In addition, we show that, under certain conditions, the L-infinite estimator can be exploited to approximate the Hausdorff distance in real-time implementations.**

Index Terms — Distortion metric, L-infinite, L-2, MAXAD, MSE, Hausdorff, 3D mesh coding

I. INTRODUCTION

The diversification of content and the increasing demand in mobility has led to a proliferation of heterogeneous terminals, with diverse capabilities. Efficient storage and transmission of digital data is therefore a critical problem, which can be solved by compressing the original data based on some predefined criteria.

There is a broad range of applications (e.g. in the medical area), where compact coding cannot come at the expense of information loss. A viable solution in this case is given by lossless coding, possibly coupled with multi-functionality support, such as scalability and progressive (lossy-to-lossless) reconstruction of the input data. Lossless coding is downsized however by the fairly low achievable lossless compression ratios. There are other applications, such as those in the field of remote sensing, where one can accept information loss in favor of higher compression ratios, provided that the distortions incurred are rigorously bounded. In such applications, lossy or near-lossless compression are suitable, but an appropriate distortion measure needs to be employed in order to accurately quantify and control the distortion incurred by the compression system.

The ideal distortion metric in lossy coding of meshes is the Hausdorff distance*,* as this metric expresses the maximum local error between the original and decoded meshes. However, to compute the Hausdorff distance, considerable processing power and memory space are needed, in particular for high resolution meshes. This becomes even more critical in scalable mesh coding systems, where, in order allocate rate, one needs to determine the Hausdorff distance for all possible decodings (i.e. at every spatial resolution and quality level) of the encoded object(s). This is computationally prohibitive, and as a consequence, the traditional approach is to replace the Hausdorff distance by the L-2 distortion, commonly known as the mean square error. The rate allocation is thus optimized for the L-2 distortion, while the Hausdorff distance is only used offline as means to evaluate compression performance and compare different coding systems. Such systems are obviously optimal in L-2 sense, and cannot claim any kind of optimality in Hausdorff-distance sense.

The L-2 distortion though, is an average distortion measure, giving a good approximation of the *global* error and an expression of the overall perceptual quality. One of its major drawbacks consists in the fact that it does not exploit "local" knowledge. Moreover, the local error behavior is lost, due to an averaging of the reconstruction error throughout the entire data. However, there are applications that require imposing a tight bound on the individual elements of the error signal, i.e. constraining the elements of the reconstruction error signal to be under some given thresholds. To address these issues, the L-infinite norm criterion has been proposed as a candidate for a perceptually meaningful norm in such applications, in that the L-infinite distortion provides a good approximation of the maximum *local* error.

The L-infinite distortion has been investigated in the context of coding of 3D graphics only recently [1]. With this respect, [1] proposes a scalable L-infinite mesh coding approach, simultaneously performing local error control and providing scalability in L-infinite sense. The L-infinite coding approach is instantiated by using the MESHGRID coding system [2]. Additionally, an L-infinite distortion estimator is proposed [1], expressing the L-infinite distortion in the spatial domain as a function of quantization errors occurring in the wavelet domain. Employing this estimator enables an optimized rate-allocation for given local-error bounds, without performing an actual decoding of the mesh. In addition, a fast algorithm for solving the rate-distortion optimization problem is designed, enabling a real-time implementation of the rate-allocation [1].

In this survey paper we give an overview of our L-infinite estimators recently introduced in [1] and investigate how the L-infinite distortion metric performs compared to the L-2 distortion metric, when both are instantiated for the MESHGRID coding system [2].

The paper is structured as follows. Section II makes a theoretical overview of the two distortion metrics, while section III compares them experimentally. Finally, section IV draws the conclusions of this work.

II. DISTORTION METRICS

Let us consider the generic situation wherein the scalable coding system decomposes the input 3D object into L different sources of information, each of these sources being progressively encoded. The sources can be independent spatial regions, if we consider a spatial partitioning of the input mesh, such as [3-5], or wavelet subbands, if we consider a wavelet-based coding approach, such as MESHGRID [2] and WSS [6, 7]. The classical approach in lossy compression consists in optimizing the compression scheme so as to maximize the overall compression ratio for a given reconstruction error. The quality of reconstruction can be measured by using a distortion measure $d(\mathbf{x}, \mathbf{y}) \geq 0$ that assigns a distortion or cost to the reproduction of the input $\mathbf{x} = (x_1, ..., x_k)$ by the output $\mathbf{y} = (y_1, ..., y_k)$.

A common distortion measure, which is easy to compute and has certain usefulness in analysis, is the (un-normalized) squared error, defined as:

$$d_2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|^2 \qquad (1)$$

Relation (1) can be written as $d_2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2^2$, where $\|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$ is the L-2 norm.

In a similar manner, we can define a more general L-p norm as:

$$\|\mathbf{x} - \mathbf{y}\|_p = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \qquad (2)$$

This allows for an extension of the distortion measure d to any power of the L-p norm: $d_{p,r}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p^r$. For $r = p$ we obtain an additive distortion, simply denoted as $d_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p^p$, commonly known as the pth power distortion (or L-p distortion). A very common distortion metric is the L-2 distortion, corresponding to $r = p = 2$.

It is important to highlight the terminology difference between the L-p norm, defined by (2), and the L-p distortion, given by the L-p norm to the power p. We point out that in the remainder of this work, unless it is clearly stated, we systematically refer to the L-p distortion instead of the L-p norm.

A variation of the L-p norm is the L-infinite norm, defined as $\|\mathbf{x} - \mathbf{y}\|_\infty = \max_i |x_i - y_i|$. The L-infinite distortion measure is given by:

$$d_\infty(\mathbf{x}, \mathbf{y}) = \max_{1 \leq i \leq n} |x_i - y_i| \qquad (3)$$

In general, the squared error distortion (or L-2 distortion) expressed by (1) is regarded as a useful indicator of perceptual quality. Its statistical average, which is commonly referred to as the Mean Square Error (MSE), is then regarded as giving a good approximation of the *global* error. An alternative to MSE is the L-infinite distortion, which, opposite to the L-2, provides a good expression of the maximum *local* error [8].

In order to perform an optimized rate allocation in a scalable coding system, one needs to know the actual distortion that occurs by decoding the input at a given set of rates, corresponding to specific decoding points in the bit-stream. Since measuring the distortion at all these points is prohibitive, one needs to estimate the distortion instead, as a function of rate and various coding parameters in the system. The way such estimates are determined is explained next.

Let us denote by D_{tot} the spatial-domain distortion in the reconstructed mesh, and by D_l the contribution in the total distortion of a given source $l, 1 \leq l \leq L$. In general, for additive distortion metrics, the spatial-domain distortion D_{tot} can be expressed as a linear combination of distortions $D_l(R_l)$ on source l, of the form:

$$D_{tot} = \sum_{l=1}^{L} q_l D_l(R_l) \qquad (4)$$

where R_l is the rate associated with source l, and the q_ls are weighting the different distortion contributions in the total distortion.

This additive distortion model is generic. In spatial-partitioning mesh-coding approaches, the sources are adjacent spatial regions (or mesh segments), hence $q_l = 1, \forall l$, and the sources are independent. In wavelet-domain approaches, such as MESHGRID, each source is a wavelet subband which is progressively encoded in a bitplane-by-bitplane manner, $D_l(R_l)$ is the source distortion-rate associated with subband l, and the weights q_l depend only on the distortion-metric type and wavelet filter-bank employed.

In the case of an L-2 distortion metric, each D_l in expression (4) has the form given by (1). In wavelet-based coding approaches, the weights q_l in (4) depend on the gains of the wavelet filters. For orthonormal wavelet filters, $q_l = 1$, as the transform is unitary. However, for biorthogonal wavelet filters, one needs to account for the gains of the wavelet filters, and to re-scale the coefficients accordingly, such that the resulting wavelet transform is (approximately) unitary. The weights q_l can then be determined by calculating the L-2 norm of the low- and band-pass wavelet filters. It can be easily derived that for MESHGRID [2], the q_l factors depend also on the gains of the wavelet filters used in the wavelet decomposition of the reference-grid. For the particular wavelet-transform employed by MESHGRID [2], one can derive the total L-2 distortion in 3D as [1]:

$$D_{tot} = \left(\frac{105}{128} \right)^{3J} D_{LLL}^{(J)} + \sum_{j=1}^{J} \frac{1}{8} \left(\frac{105}{128} \right)^{3(j-1)} \left[D_{HHH}^{(j)} + \right.$$

$$\left. \frac{105}{64} \left(D_{HHL}^{(j)} + D_{HLH}^{(j)} + D_{LHH}^{(j)} \right) + \left(\frac{105}{64} \right)^2 \left(D_{HLL}^{(j)} + D_{LHL}^{(j)} + D_{LLH}^{(j)} \right) \right]$$

$$(5)$$

where $D_s^{(j)}$ is the L-2 distortion on subband s of the decomposition level j; in the naming of the subbands, L and H indicate a low-pass and band-pass filtering respectively, performed in a specific direction.

Concerning L-infinite coding, two possible approaches have been followed in [1] in the design of an L-infinite mesh coding system. First, a theoretical approach is followed for estimating the MAXAD, which is the Maximum Absolute Difference between the spatial-domain sample values in the original and reconstructed data. It is shown how, under worst-case scenario assumptions, the maximum quantization errors from the different wavelet subbands are linearly combined into a maximum spatial-domain reconstruction error. A generic formulation of the MAXAD is given in [1]. For the particular instantiation of the wavelet transform used by MESHGRID, the theoretical MAXAD estimator is given by:

$$D_{tot} \leq M_{tot} = \left(\frac{125}{64}\right)^J \frac{\Delta_{LLL}^{(J)}}{2} + \sum_{j=1}^{J}\left(\frac{125}{64}\right)^{j-1}\left[\frac{1}{2}\Delta_{HHH}^{(j)} + \right.$$
$$\left. + \frac{5}{8}\left(\Delta_{HHL}^{(j)} + \Delta_{HLH}^{(j)} + \Delta_{LHH}^{(j)}\right) + \frac{25}{32}\left(\Delta_{HLL}^{(j)} + \Delta_{LHL}^{(j)} + \Delta_{LLH}^{(j)}\right)\right]$$

(6)

where M_{tot} is the smallest upper bound of the MAXAD D_{tot}.

The theoretical L-infinite distortion estimator is data independent, since it is computed solely based on worst-case assumptions about the error contributions coming from the different wavelet subbands. Hence, the wavelet coefficients need not to be actually decoded, and no inverse wavelet transform needs to be performed. Therefore, this approach is very fast in computational terms. However, as shown later experimentally, this approach also overestimates the actual L-infinite distortion, since it relies on worst-case assumptions. Consequently, an improved approach is proposed in [1], in which the L-infinite distortion estimate is computed based on data-dependent statistical information. It is shown experimentally that the accuracy of this second approach is improved substantially, while the supplementary computational expenses are minimal. A statistical, data-dependant L-infinite estimator for an intra-band wavelet codec is presented in [1], which in case of MESHGRID has the following expression:

$$M_{tot} = \sqrt{2}\, erf^{-1}(P)\sqrt{\sum_l 2^{2b_l}\frac{3p_{b_l}^l+1}{3p_0^l+1}\left(\sigma_0^l\right)^2}$$

(7)

where $p_{b_l}^l$ is the probability that the wavelet coefficients in subband l fall in the deadzone when quantized with the embedded quantizer at level $b_l, b_l \geq 0$; σ_b^l denotes the standard deviation of the quantization errors occurring on the wavelet coefficients of subband l at level b, and $erf(\cdot)$ is the error function with the estimation probability $P \approx 1$, for which $erf^{-1}(P) \in \left[5/\sqrt{2}, 6/\sqrt{2}\right]$.

A major benefit of the L-infinite mesh coding approach is that the Hausdorff distance between the original and
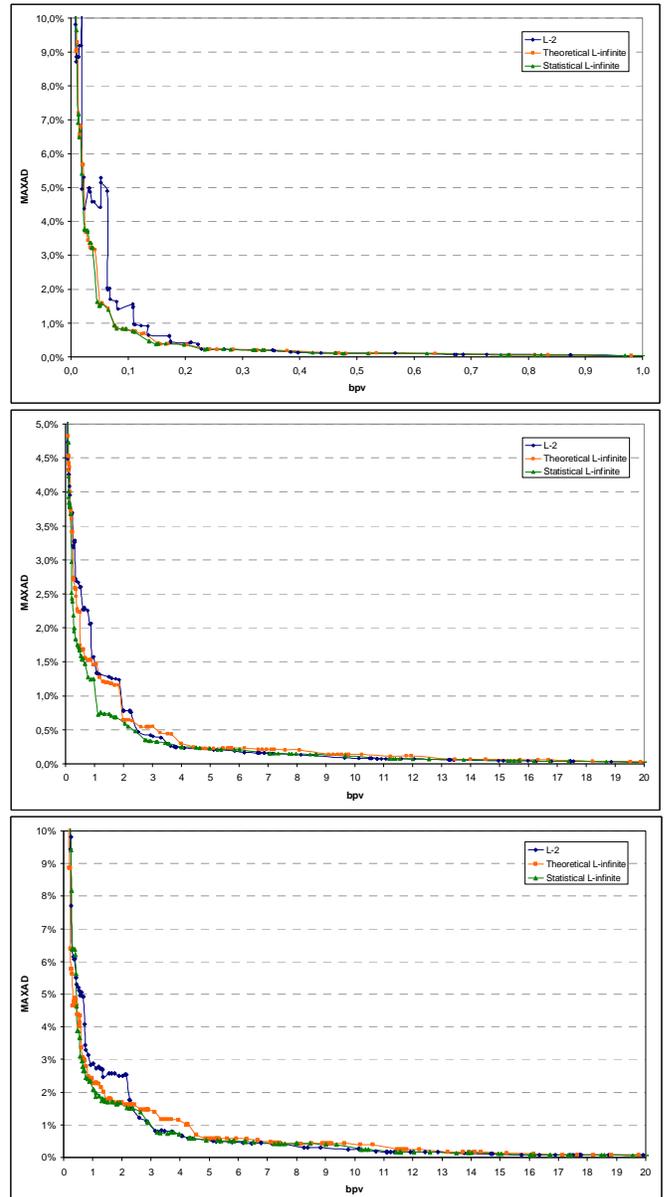

Figure 1: MAXAD versus rate for the *Mars* (top), *Heart* (middle) and *Melted Tangle Cube* (bottom) models for the L-2 and L-infinite coding systems.

reconstructed meshes at a given resolution is upper-bounded by the MAXAD [9]. Specifically:

$$h(A, \bar{A}) \leq MAXAD$$

(8)

where $h(A, \bar{A})$ is the Hausdorff distance and A and $\bar{A}$ are the original and decoded sets of vertices at a given rate and resolution. Note that, in the proof of (8), it is assumed that a bijective mapping exists between the vertices in A and $\bar{A}$. For more details, the interested reader is referred to [9].

Expression (8) implies that if such a bijective mapping exists, then an L-infinite-constrained coding actually implies Hausdorff-distance-constrained coding [1]. Practically, one has the possibility to set a specific target bound on the L-infinite distortion, and due to (8), the Hausdorff distance will be guaranteed to be below that target bound.
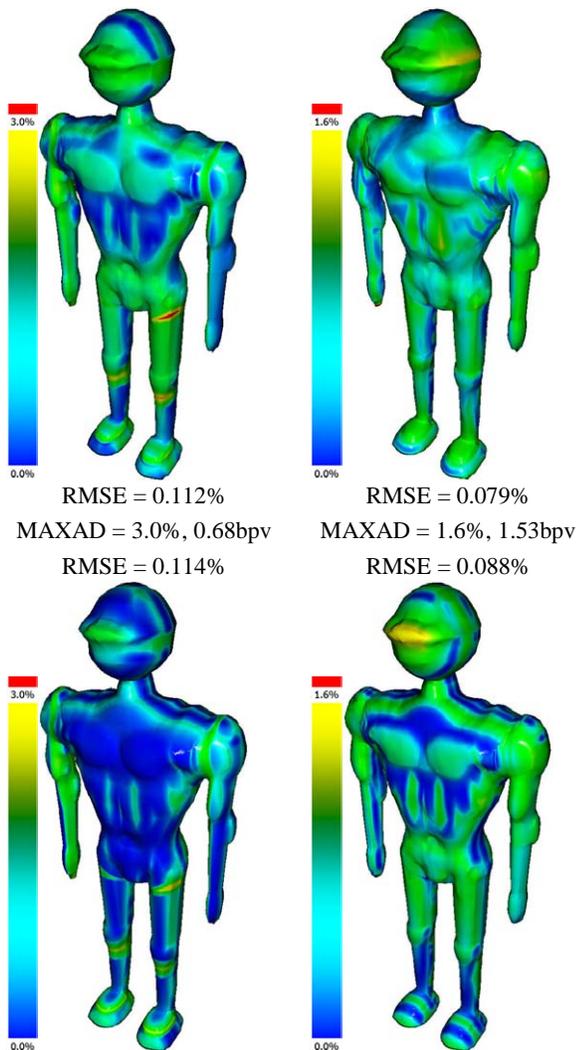
Figure 2: L-2 (top) versus statistical L-infinite (bottom) coding of the *Humanoid* model. The color shades are proportional to the local error and exceeding the MAXAD bound is indicated in red.

RMSE = 0.099%   RMSE = 0.113%
MAXAD = 2.0%, 1.11bpv

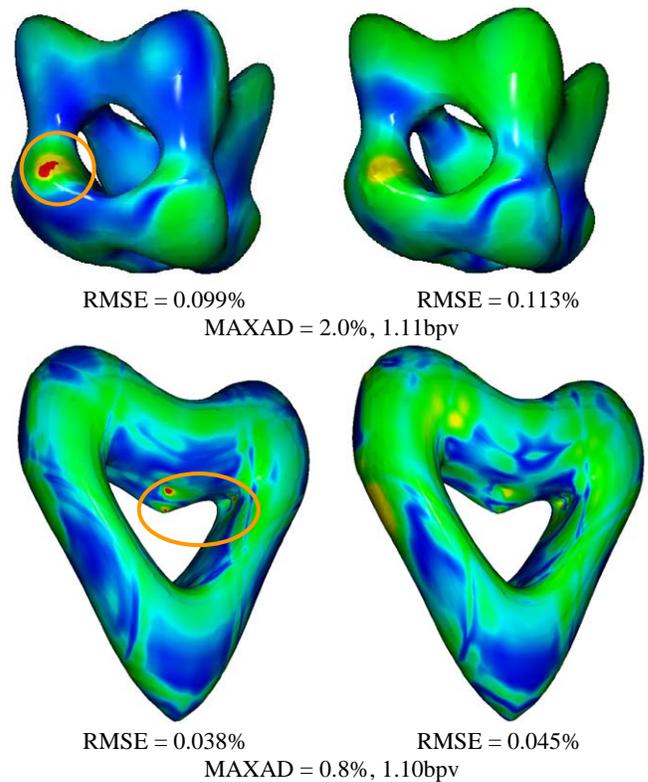RMSE = 0.038%   RMSE = 0.045%
MAXAD = 0.8%, 1.10bpv

Figure 3: L-2 (left) versus L-infinite (right) coding of the *Melted Tangle Cube* (top) and *Heart* (bottom) models. The ellipses highlight the areas where the vertex errors exceed the requested MAXAD bound.

It is important also to point out that the considered L-infinite mesh coding approaches achieve scalability in L-infinite sense [1], corresponding to a perfectly predictable L-infinite / Hausdorff-distance upper bound for every allowable truncation point. These features render the considered L-infinite-constrained coding as a unique and interesting alternative to all mesh coding techniques proposed so far in the literature.

III. EXPERIMENTAL RESULTS

The following experiments compare the L-2 and L-infinite distortion metrics presented so far, when employing the MESHGRID coding system. MESHGRID is a 3D object representation standardized by MPEG-4 AFX which consists of two basic components, including (*i*) a hierarchical reference-grid (RG), i.e. a 3D grid of points serving as a multi-resolution reference system for the object, and (*ii*) a hierarchical connectivity-wireframe (CW), describing the connectivity between vertices at each spatial resolution [2].

MESHGRID employs a wavelet-based coding technique to encode the RG in a scalable manner, whereas the CW is losslessly encoded at each spatial resolution using a 3D extension of chain-codes. For more details over MESHGRID one refers to [2].

In the first experiments, we plot the *actual* MAXAD versus rate for the L-2 and L-infinite codecs for two models and for a broad range of rates (Figure 1). We notice that all dots on the graphs in Figure 1 are decodable points, where the local error is clearly upper-bounded and guaranteed.

These results indicate that very large gaps in terms of MAXAD can occur for the L-2 codec, this phenomenon being completely un-controllable for this system. This shows an expectable result, i.e., an optimization with respect to the L-2 distortion lays no claim on minimizing the local error, in this sense having the potential of introducing large local error-spikes (i.e., large vertex-position errors), which otherwise are not present in an L-infinite-coding framework.

This phenomenon is illustrated visually in Figure 2-Figure 5. The experiments are performed on *Humanoid* at various MAXAD targets (Figure 2), and on four other models: *Melted Tangle Cube* (Figure 3 top), *Heart* (Figure 3 bottom), *Swiss Landscape* (Figure 4) and a smooth surface with sharp local features (*Mars* - Figure 5). The L-infinite codec decodes the object at specific MAXAD bounds, while the L-2 codec operates at the same rates as the L-infinite codec, minimizing the L-2 distortion for each rate. The target MAXAD values and the resulting bit-rates are given in the figures. Also, the

RMSE = 0.061%
MAXAD = 1.6%, 0.05bpv
RMSE = 0.074%
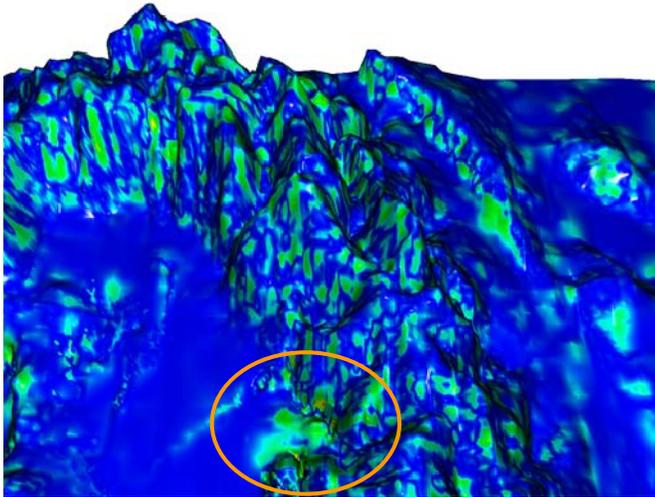
RMSE = 0.061%
MAXAD = 1.6%, 0.05bpv
RMSE = 0.074%

Figure 5: L-2 (top) versus L-infinite (bottom) coding of the *Mars* model. The ellipses highlight the areas where the vertex errors exceed the requested MAXAD bound.
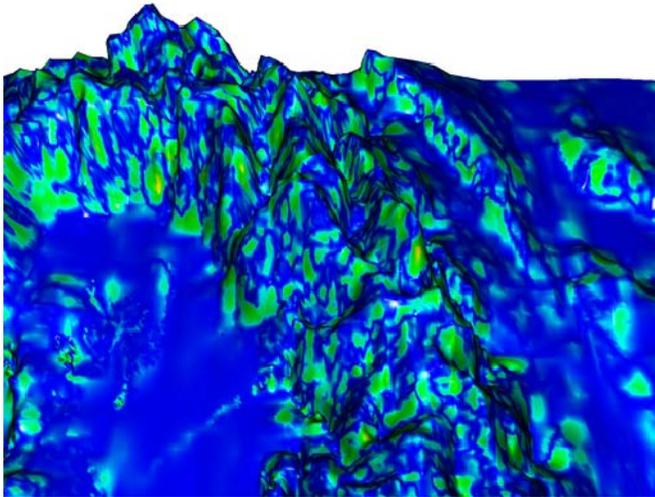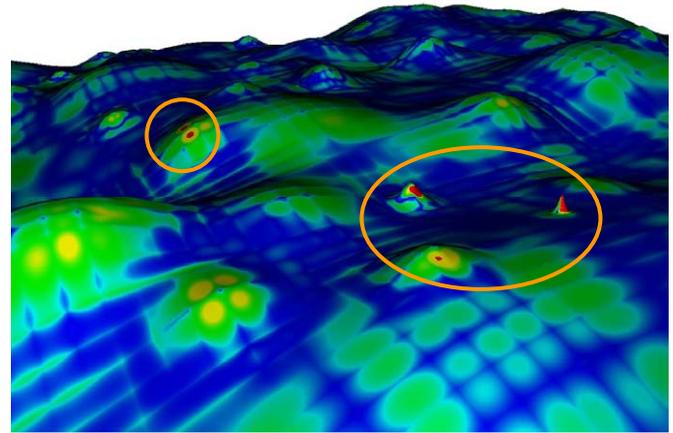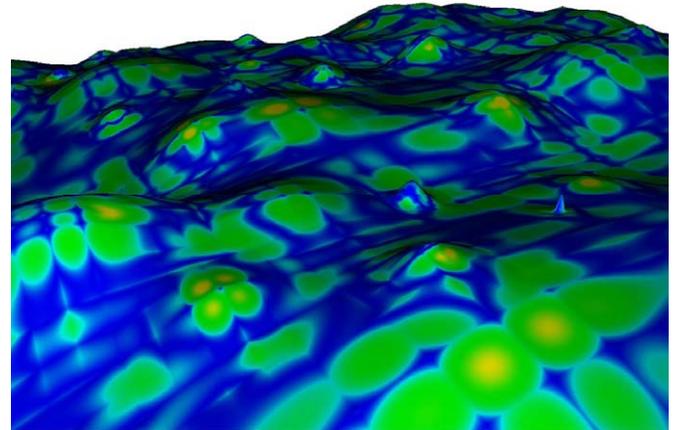

Figure 4: L-2 (top) versus L-infinite (bottom) coding of the *Swiss* model. The ellipse highlights the area where the vertex errors exceed the requested MAXAD bound.

root mean square error (RMSE) values at these bit-rates are given for each codec. The local errors are indicated using colors, the assigned color being proportional to the magnitude of the local error. In the considered color maps, exceeding the MAXAD bound is indicated in red. It is important to point out that the color range changes with the target MAXAD, so the color maps schemes are different for every rate. Therefore, the visual comparison should be performed only among the different estimators and not among rates.

These results demonstrate the local-error control performed by the L-infinite coding approach, which, as expected, never exceeds the imposed MAXAD bound. Visually though, the results are similar for the L-infinite and data-dependent L-2 estimators. This is also due to the particular nature of MESHGRID; indeed, for *Heart* for instance, for a MAXAD bound of 0.8% (Figure 3 bottom), the local error exceeds the MAXAD on 1121 reference-grid points (2.34%), going up to

4155 points (8.68%) for a MAXAD bound of 1%. Not all these errors are visible on the mesh, and this is due to the very specific nature of the MESHGRID system, for which the mesh vertices are connected only to a part of the reference-grid (RG) points [2]. Hence, errors on the reference-grid are translated to the mesh only for those RG points that are directly linked to mesh vertices. Nonetheless, it is important to remark that, although yielding very similar visual results to the L-infinite codec, the L-2 codec cannot claim *any* kind of local error control, at any rate and irrespective of the mesh type (smooth or sharp).

We notice also that in L-2 sense (i.e. in terms of RMSE), the L-2 codec version is systematically better than the L-infinite version. This *must be* the case, and it comes with no surprise. For any rate, the L-2 codec reaches the minimum L-2 distortion, because the rate allocation is optimized in L-2 sense. However, one can observe that the RMSE differences between the L-2 and L-infinite codec versions are small. Furthermore, despite of providing a smaller RMSE, the L-2

codec may be affected by large local errors, in particular at low rates.

As a final remark, it is important to notice that both L-infinite distortion estimators operate without actually decoding the models. This makes them very fast in assessing the actual distortion, being suitable for real-time applications [1]. This lies in contrast to a Hausdorff-distance computation, which can be performed only offline, in particular for a scalable codec or for high-resolution meshes.

IV. DISCUSSION AND CONCLUSIONS

This paper compares two distortion metrics, L-2 and L-infinite, when employed in 3D mesh coding. Two L-infinite distortion estimators are presented, expressing the L-infinite distortion in the spatial domain as a function of quantization errors produced in the wavelet domain. Based on these, the L-infinite oriented codec optimizes the rate allocation for which the L-infinite distortion is upper-bounded by a user-defined bound, and guaranteed to be below that bound. We find that this is an interesting and unique feature in the context of 3D object coding.

In summary, an L-infinite coding system (*i*) ensures a tight local error control, (*ii*) enables scalability in L-infinite sense, (*iii*) provides a close bound for the Hausdorff distance under certain conditions (i.e. when there exists a bijective mapping between vertices of original and reconstructed meshes at a given resolution), and (*iv*) allows for computing real-time estimates of the Hausdorff distance for a whole range of rates. In contrast, an L-2 system cannot claim any kind of local error control, at any rate and irrespective of the mesh type.

Based on the experimental results, we conclude that a data-dependent L-2 estimator is sufficient for applications for which geometry accuracy in not critical. However, L-infinite coding is the only available option for applications for which preserving geometry accuracy is compulsory. Examples of such applications include 3D CAD, 3D representation and coding of industrial devices, 3D topography, or medical applications. In such applications, the physical characteristics of the objects of interest, such as volume, shape or heights can be measured based on their mesh geometry. Altering the geometry via compression affects these physical characteristics; hence, controlling the local error on the mesh geometry by following an L-infinite mesh coding approach is of critical importance. Other potential applications include geometry-based watermarking of 3D models, such as [10, 11]. Following an L-infinite mesh coding approach offers the possibility to control the geometric perturbations generated by the watermark embedding process, and opens the door for applications that simultaneously provide compression and watermarking of 3D models.

REFERENCES

[1] A. Munteanu, D. C. Cernea, A. Alecu, J. Cornelis, and P. Schelkens, "Scalable L-infinite coding of Meshes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 3, pp. 513-528, May-June 2010.

[2] A. Salomie, A. Munteanu, A. Gavrilescu, G. Lafruit, P. Schelkens, R. Deklerck, and J. Cornelis, "MESHGRID – A Compact, Multi-Scalable and Animation-Friendly Surface Representation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14, no. 7, pp. 950-966, July 2004.

[3] S.-B. Park, C.-S. Kim, and S.-U. Lee, "Error Resilient 3-D Mesh Compression," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 885-895, October 2006.

[4] D. Tian and G. Al-Regib, "Multistreaming of 3-D Scenes With Optimized Transmission and Rendering Scalability," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 736-745, June 2007.

[5] Y. Zhidong, S. Kumar, and C.-C. J. Kuo, "Error-resilient coding of 3-D graphic models via adaptive mesh segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 7, pp. 860-873, July 2001.

[6] M. Lounsbery, T. D. Derose, and J. Warren, "Multiresolution analysis for surfaces of arbitrary topological type," *ACM Transactions on Graphics*, vol. 16, no. 1, pp. 34-73, 1997.

[7] A. Khodakovsky, P. Schröder, and W. Sweldens, "Progressive geometry compression," *SIGGRAPH 2000*, pp. 271-278, 2000.

[8] A. Alecu, A. Munteanu, J. Cornelis, and P. Schelkens, "Wavelet-based Scalable L-infinity-oriented Compression," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2499-2512, September 2006.

[9] D. C. Cernea, A. Munteanu, A. Alecu, J. Cornelis, and P. Schelkens, "Scalable joint source and channel coding of meshes," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 503-513, March 2008.

[10] O. Benedens, "Geometry-based watermarking of 3-D models," *IEEE Computer Graphics and Applications*, vol. 19, no. 1, pp. 46-55, January 1999.

[11] A. G. Bors, "Watermarking mesh-based representations of 3-D objects using local moments," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 687-701, March 2006.