TWO OPTIMIZATIONS OF THE MPEG-4 FAMC STANDARD FOR ENHANCED COMPRESSION OF ANIMATED 3D MESHES

*K. Mamou, T. Zaharia, F. Prêteux**

ARTEMIS Department
Institut TELECOM
TELECOM & Management SudParis
9, Rue Charles Fourier,
91011 Evry Cedex - France

A. Kamoun, F. Payan, M. Antonini

Laboratoire I3S - UMR 6070 CNRS
Université de Nice - Sophia Antipolis
2000, Route des Lucioles
France

ABSTRACT

Recently, the MPEG-4 standard adopted a novel technology for compression of dynamic 3D meshes with constant connectivity and time-varying geometry, refered to as FAMC - Frame-based Animated Mesh Compression. In this paper, we propose two optimizations of the FAMC approach, aiming at improving the compression efficiency. The first one is based on a PCA (Principal Component Analysis) decomposition of the motion compensation error residuals. The second improves the bi-orthogonal (4-2) wavelet coding approach supported by the standard, by introducing an optimal bit allocation procedure, combined with an adapted quantization of wavelet coefficients. Experimental results show that both optimizations lead to significant gains in compression rate (about 20-30%) at low bitrates.

Index Terms— Mesh compression, animation compression, dynamic mesh compression, MPEG-4 AFX

1. INTRODUCTION

Today's hi-tech industries of video games, CGI films, special effects or CAD systems make use at a large scale of diverse animated 3D content, for a wide range of professional or general public applications. Within this highly motivating industrial framework, 3D animated content is most often represented as sequences of 3D key-frames, to be interpolated for ensuring the desired video framerates. The key-frame approaches obviously lead to highly complex representations. Efficiently storing, transmitting and rendering such representations becomes then a major challenge, as testifies the rich literature dedicated to this emerging research area (see [1] for an overview), which includes:(1) Local spatio-temporal predictive approaches [2]; (2) Principal Component Analysis (PCA)-based techniques [3, 4]; (3) Wavelet-based methods [5, 6]; and (4) Segmentation-based approaches [7].

*This work is partly supported by the HD3D project of the CapDigital competitiveness cluster.

Identifying the challenge of disposing of a normalized and efficient format for representing 3D animations, the MPEG-4 standard has recently adopted, as Amendment 2 of Part 16 - AFX (Animation Framework eXtension) the so-called FAMC (Frame-based Animated Mesh Compression) approach. FAMC outperforms both techniques previously considered in MPEG-4 and state-of-the art methods, while ensuring a wide range of functionalities, including compression efficiency, temporal, spatial and SNR scalability, and nearly-lossless compression.

The core of the FAMC approach is a skinning model, used for motion compensation purposes. Residual errors are encoded by using transform-based techniques, including DCT, bi-orthogonal integer-to-integer (4-2) wavelet transform [8], and LD transform [9]. FAMC provides a generic and modular architecture, which can easily accommodate several methods as well as further optimizations. In this paper, we specifically consider and investigate two different methods for optimizing the FAMC compression performances. The first one concern a PCA decomposition of the residual errors. The second optimizes the (4-2) wavelet performances, by introducing a novel bit allocation mechanism, combined with optimized quantization technique.

The rest of the paper is organized as follows. The FAMC encoder architecture is detailed in Section 2. The proposed optimizations are described in Section 3. Section 4 presents an experimental evaluation of the proposed optimization. Finally, Section 5 concludes the paper and opens perspectives of future work.

2. THE FAMC ARCHITECTURE

Figure 1 presents the FAMC encoder architecture.

The core of the FAMC approach is a skinning-based motion representation strategy which involves: 1) an optimal motion based segmentation approach [10], which aims at partitioning the mesh vertices into a set of clusters whose motion can be accurately described by a single affine 3D transform,

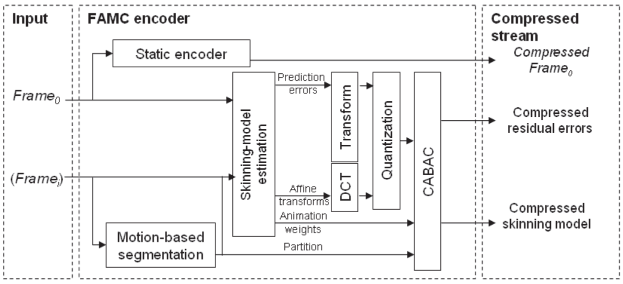

Fig. 1. Synopsis of the FAMC encoding algorithm.

and 2) a skinning module which expresses the motion of each mesh vertex as a weighted combination of the clusters motions.

The first frame of the sequence is compressed by using an arbitrary static encoder. The residual motion compensation errors are finally encoded with the help of a temporal/spatial transform.

2.1. Skinning-based motion modeling

The skinning principle consists of expressing the motion of each vertex as a weighted linear combination of the patch affine motions associated with the considered partition.

Thus, the predicted position $\widehat{\chi}_t^v$ of a vertex v at frame t is given by:

$$\widehat{\chi}_t^v = \sum_{k=1}^{K} w_v^k \; A_t^k \; \chi_1^v, \qquad (1)$$

where : 1) $\chi_t^v = (x_t^v, y_t^v, z_t^v, 1)^t$ describes the vector position of vertex v at frame t in homogeneous coordinates, 2) V, F and K stands for the number of vertices, frames and clusters, respectively, 3) w_k^v is a real-valued coefficient, so-called animation weight, which controls the influence of the patch k on the motion of the considered vertex v, and 4) A_t^k denotes the 3D affine transform associated to the cluster π_k at frame t, defined as:

$$A_t^k = \arg\min_A \left(\sum_{p \in \pi_k} \|A\chi_1^p - \chi_t^p\|^2 \right). \qquad (2)$$

The estimation of the optimal skinning parameters (affine transforms and animation weights) is described in [11].

The prediction residual errors are defined as:

$$\forall \; v \in \{1, ..., V\}, \quad e_t^v = \chi_t^v - \widehat{\chi}_t^v. \qquad (3)$$

2.2. Encoding of the residual errors

Several approaches can be used for encoding the residual errors. Let us first summarize the techniques currently supported by the FAMC standard.

FAMC currently the following transforms: 1) Discrete Cosine Transform (DCT), dedicated for applications where the compression efficiency is privileged, 2) Integer to integer lifting-based (4-2) bi-orthogonal wavelet transform [8], which makes it possible to control the peak error and to achieve nearly-lossless compression while ensuring temporal scalability, and 3) Layered-based decomposition (LD) [9], which ensures temporal and spatial scalability.

In addition to these methods, in this paper we propose two further optimizations, described in the next section.

3. OPTIMIZED ENCODING OF RESIDUAL ERRORS

3.1. PCA-based compression

In this section we propose to encode directly the prediction residual errors associated to each cluster π_k of the skinning partition by applying a PCA-based strategy. The choice of applying PCA to each cluster separately makes it possible to reduce the computational complexity and to better capture the local behaviors of the predictions errors. The principle is to consider the errors associated to the each cluster as a set of vectors defined in $R^{3 \times V_k}$ (V_k being the number of vertices belonging to cluster π_k) that are projected onto a reduced sub-set of PCA eigenvectors, corresponding to the most significant eigenvalues. Let us note that both projection coefficients and eigenvectors need to be transmitted to the decoder. As suggested in [4], we apply an adaptive quantization procedure guided by an optimization strategy aiming at minimizing the distortions under a fixed bitrate constraint.

3.2. Optimized wavelet-based compression

In parallel, we also propose to encode the prediction residual errors by using the wavelet-based coder proposed in [5]. The main feature of this coder is the optimized quantization of the subbands of wavelet coefficients obtained after a temporal wavelet filtering of the prediction errors. The optimized quantizers are obtained by introducing a model-based bit allocation process. This bit allocation process tends to optimize the trade-off between bitrate and quality of the reconstructed data.

Here is the outline of this coder: i) a temporal wavelet filtering first transforms the prediction residual errors in a multiresolution structure: one low frequency 3D data and several subbands of high frequency 3D details (the wavelet coefficients); ii) These data are encoded using scalar quantizers SQ, which depend on the optimal quantization steps computed during the allocation process; iii) the allocation process aims to dispatch the bit budget across the different subbands of the multilevel structure according to their influence on the quality of the reconstructed data for one specific user-given bitrate; iv) the quantized data are finally entropy encoded to produce the bitstream of prediction errors.

3.2.1. *Description of the bit allocation*

The general purpose of this bit allocation process is to compute the set of optimal quantizers $\{q^*\}$, which minimizes the reconstructed mean square error D_T for one specific user-given target bitrate R_{target}. The solutions $\{q^*\}$ are obtained by solving the problem

$$(\mathcal{P}) \begin{cases} \text{minimize} & D_T(\{q\}) \\ \text{with constraint} & R_T(\{q\}) = R_{target}, \end{cases} \tag{4}$$

with R_T the total bitrate. By using a lagrangian approach, the constrained allocation problem $\mathcal{P}$ can be solved by minimizing the criterion

$$J_\lambda(\{q\}) = D_T(\{q\}) + \lambda(R_T(\{q\}) - R_{target}), \tag{5}$$

with λ the lagrangian operator.

Since scalar quantizers are used, the $3D$ data are splitted in three $1D$ (*coordinate*) subbands, and then encoded separately with three different scalar quantizers. So, the reconstructed mean square error can be defined by

$$D_T(\{q\}) = \sum_{i=0}^{N} w_i \sum_{j=1}^{3} D_{i,j}(q_{i,j}), \tag{6}$$

where $\{w_i\}$ are weights due to the wavelet non-orthogonality (i is the resolution level), $D_{i,j}$ the mean square error relative to the coordinate subband i, j ($j = 1$ for the x-coordinates, $j = 2$ for the y-coordinates, and $j = 3$ for the z-coordinates), and $q_{i,j}$ the associated quantization step. The numerical values of the weights $\{w_i\}$ depends on the wavelet used, and are given in [5]. In this paper, we use only the lifting-based (4,2) bi-orthogonal wavelet coefficients. In parallel, the total bitrate R_T can be developed in

$$R_T(\{q\}) = \sum_{i=0}^{N} \sum_{j=1}^{3} a_{i,j} R_{i,j}(q_{i,j}), \tag{7}$$

with $D_{i,j}$ the bitrate relative to the coordinate subband i, j, and $\{a_{i,j}\}$ the coefficients relative to the subsampling.
The optimal quantization steps $\{q^*\}$ are finally obtained by solving the following system:

$$\begin{cases} \frac{\partial J_\lambda(\{q\})}{\partial q} = 0 \\ \frac{\partial J_\lambda(\{q\})}{\partial \lambda} = 0 \end{cases}. \tag{8}$$

By merging (6) and (7) in (5), the system (8) becomes a system of $(3N + 3)$ equations with $(3N + 4)$ unknowns (the set $\{q_{i,j}\}$ and λ):

$$\frac{\frac{\partial D_{i,j}(q_{i,j})}{\partial q_{i,j}}}{\frac{\partial R_{i,j}(q_{i,j})}{\partial q_{i,j}}} = -\lambda \frac{a_{i,j}}{w_i} \tag{9a}$$

$$\sum_{i=0}^{N} \sum_{j \in J_i} a_{i,j} R_{i,j}(q_{i,j}) = R_{target}. \tag{9b}$$

The solutions of (9a) can be obtained by inverting the equations. Unfortunately, this stage is impossible given the complexity of the equations. To overcome this problem, we use an iterative algorithm depending on λ [12].

3.2.2. *Overall Algorithm*

The optimal solutions of the system (9) for the given bitrate R_{target} are computed thanks to the following overall algorithm:

1. λ is given. For each set (i, j), compute $q_{i,j}$ that verifies (9a);
2. while (9b) is not verified, calculate a new λ by dichotomy and return to step (1);
3. stop.

The computation of the quantization steps $\{q_{i,j}\}$ as solutions of (9a) can be done according to different methods. We observe that the probability density functions of the processed data can be modeled by a Generalized Gaussian Distribution (GGD). Hence, we can also use the model-based algorithm presented in [5] to compute the solutions of (9a). This leads to a fast and low-complexity algorithm. The interested reader should refer to [12] for more details and explanations.

4. EXPERIMENTAL RESULTS

Experiments have been carried out on the MPEG-4 test data set, including about 30 animation sequences with various sizes, shapes, and motions.

Figure 2 compares the rate-distorsion curves for the standardized integer to integer lifting/DCT-based FAMC versions to the optimized encoders PCA/optimized lifting-based FAMC we propose. The bitrates are expressed in KBits/s. The distortions here are expressed as the Da error [13] between initial and reconstructed meshes.

Let us note that in all cases and for all models the PCA-based FAMC encoder offers the best performances. Here, the KL transform involved ensures the optimality of the decomposition with regard to the the L^2 norm. However, these high compression performances are obtained at the cost of a higher computational complexity in $O(min(V_{k0}, F)^3)$ (π_{k0} being the cluster with the maximal number of vertices).

The proposed optimization of the lifting-based FAMC provides significant gains at low bitrates ($e.g.$ up to 40% of gain for "Chicken" and 20% for "Cow" w.r.t the FAMC lifting-based approach). The optimized lifting-based version offers competitive performances when compared to the standardized DCT-based FAMC while offering a lower complexity ($O(F \times V)$ instead of $O(F \times log_2(F) \times V)$).

5. CONCLUSION AND FUTURE WORK

In this paper, we have proposed two optimizations of the emerging MPEG-4 FAMC (Frame-based Animated Mesh Compression) standard. Both optimizations concern the encoding of the FAMC motion compensation residual errors. The first one, based on a PCA decomposition, leads to the highest compression rates, outperforming both wavelet-based and DCT-based approaches. The price to pay is related to the higher computational complexity.

The second optimization is related to the bi-orthogonal (4-2) wavelet transform supported by the standard and is based on an optimal bit-allocation procedure, which leads to an adapted quantization of each wavelet sub-band. The proposed approach significantly improves (20-30% of average gain in terms of bitrates) the standard lifting-based technology.

Future work will concern the study of an enhanced motion compensation approach, as well as the inclusion of the proposed methods within the MPEG-4 FAMC standard.

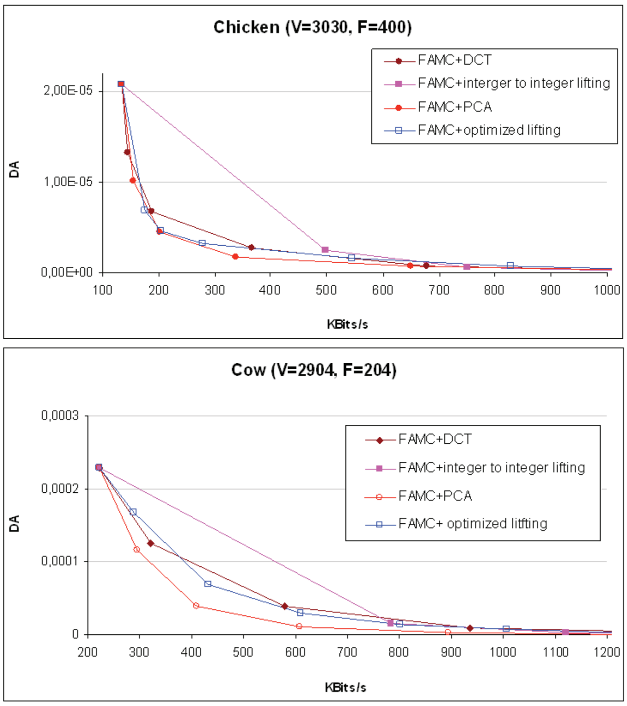

Fig. 2. FAMC vs. PCA and lifting optimized FAMC encoders: ”Chicken” and ”Cow” sequences.

6. REFERENCES

[1] K. Mamou, T. Zaharia, and F. Prêteux, “A preliminary evaluation of 3D mesh animation coding techniques,” in *Proc. of SPIE*, San Diego, USA, 2005, pp. 44–55.

[2] E. S. Jang, J. D. K. Kim, S. Y. Jung, M. J. Han, S. O. Woo, and S. J. Lee, “Interpolator data compression for MPEG-4 animation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 7, pp. 989–1008, July 2004.

[3] Z. Karni and C. Gotsman, “Compression of soft-body animation sequences,” in *Computers & Graphics 28, 1*, 2004, pp. 25–34.

[4] J. Heu, J.-H. Yang, C.-S. Kim, and S.-U. Lee, “R-D optimized compression of 3-D mesh sequences based on principal component analysis,” in *CM Southeast Regional Conference archive*, 2006, pp. 68–73.

[5] F. Payan and M. Antonini, “Temporal wavelet-based geometry coder for 3D animations,” *Elsevier Computer & Graphics*, vol. 31, no. 1, pp. 78–88, 2005.

[6] Y. Boulfani-Cuisinaud and M. Antonini, “Motion-based geometry compensation for dwt compression of 3D mesh sequence,” in *IEEE International Conference in Image Processing (CD-ROM)*, Texas, USA, 2007.

[7] G. Collins and A. Hilton, “A rigid transform basis for animation compression and level of detail,” in *Vision, Video, and Graphics*, Jul 2005, pp. 21–28.

[8] R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, “Wavelet transforms that map integers to integers,” *Applied and Computational Harmonic Analysis*, vol. 5, no. 3, pp. 332–369, 1998.

[9] N. Stefanoski, P. Klie, X. Liu, and J. Ostermann, “Scalable linear predictive coding of time-consistent 3D mesh sequences,” in *Proc. of 3DTV-CON*, May 2007.

[10] K. Mamou, T. Zaharia, and F. Prêteux, “Multi-chart geometry video: A compact representation for 3D animations,” in *Proc. of 3DPVT*, 2006, pp. 711–718.

[11] K. Mamou, T. Zaharia, and F. Prêteux, “A skinning approach for dynamic 3D mesh compression,” *Comput. Animat. Virtual Worlds*, vol. 17, no. 3–4, pp. 337–346, 2006.

[12] F. Payan and M. Antonini, “An efficient bit allocation for compressing normal meshes with an error-driven quantization,” *Computer Aided Geometric Design, Special Issue on Geometric Mesh Processing*, vol. 22, pp. 466–486, July 2005.

[13] M. Preda, “3D graphics compression core experiments description,” *ISO/IEC JTC 1/SC 29/WG 11 N8499*, 2006.