

# MESH-BASED MOTION-COMPENSATED INTERPOLATION FOR SIDE INFORMATION EXTRACTION IN DISTRIBUTED VIDEO CODING

*Denis Kubasov and Christine Guillemot*

IRISA / INRIA Rennes, Campus Universitaire de Beaulieu,  
Avenue du Général Leclerc, 35042 RENNES Cedex – France  
{denis.kubasov,christine.guillemot}@irisa.fr

## ABSTRACT<sup>1</sup>

This paper addresses the problem of side information generation in distributed video compression (DVC) schemes. Intermediate frames constructed by motion-compensated interpolation of key frames are used as side information to decode Wyner-Ziv frames. The limitations of block-based translational motion models call for new motion models. This article studies the benefits of mesh-based motion-compensated (MC) interpolation for side information extraction in DVC. A hybrid block-based and mesh-based solution addressing the problem of motion discontinuities and occlusions is also described. The increased correlation between the side information and the Wyner-Ziv encoded frames leads to performance gains which can reach up to 1 dB with respect to block-based solutions.

**Index Terms** — Video coding, Image motion analysis, Motion analysis, Motion compensation, Mesh generation.

## 1. INTRODUCTION

Distributed source coding (DSC) has been recently studied as a potential solution for compressing information in applications requiring simple encoders as well as error resilient signal compression. DSC finds its foundation in the seminal Slepian-Wolf [1] and Wyner-Ziv [2] theorems. The Slepian-Wolf theorem establishes rate bounds to the problem of lossless separate encoding and joint decoding of two binary correlated sources  $X$  and  $Y$ . The Wyner-Ziv theorem is the lossy counterpart to the problem of coding a Gaussian source  $X$  under a constraint of a given distortion  $D_X$  when the decoder has extra knowledge on  $X$  via a second correlated Gaussian source  $Y$ . These theorems establish asymptotic bounds but do not provide any practical solution. Most Slepian-Wolf and Wyner-Ziv

practical coding systems are based on channel coding principles. The statistical dependence between the two correlated sources  $X$  and  $Y$  is modeled as a virtual correlation channel analogous to binary symmetric channels or additive white Gaussian noise (AWGN) channels. The source  $Y$  (called the side information) is thus regarded as a noisy version of  $X$  (called the main signal).

Practical video compression schemes applying the DSC paradigm – referred to as Distributed Video Coding (DVC) – have been developed in a pixel [3] or transform-video representation domain [4,5]. Selected key frames (often every second frame of the video sequence) are coded using Intra coding. The intermediate frames are quantized and encoded using the Wyner-Ziv principles. The decoder first decodes the key frames. It then reconstructs intermediate frames by motion compensated temporal interpolation between adjacent key frames. Wyner-Ziv coding presents interesting properties as lending itself naturally to combat inter-frame loss propagation, hence to loss-resilient transmission. However, it still suffers from a performance gap with respect to classical motion-compensated predictive coding solutions. This can be partly explained by the limitations of block-based translational motion models with respect to motion capture and tracking between distant frames (here the Intra coded key frames). Improvements to block-based MC interpolation have been described in [6,7].

In this paper, the problem of side information extraction is addressed by considering mesh-based motion models. The DVC architecture considered is similar to the one described in [3]. However, here, the key-frames are encoded with the H.264 Intra mode instead of H.263. The mesh-based MC interpolation turns out to be very beneficial, showing PSNR performance gains between 0.5 and 1 dB when the key frame distortion is below a certain level. This can be explained by the fact that mesh-based motion models better represent affine transformations, especially in the case of large objects and of camera motion. The performance gap between the two approaches decreases when the key-frame quantization noise increases. The problem of motion estimation of small objects and of occlusions is then

---

<sup>1</sup> The work presented was developed within DISCOVER, a European Project (<http://www.discoverdvc.org>), funded under the European Commission IST FP6 program.

addressed by considering a hybrid block and mesh-based approach.

The article is organized as follows. In Section 2 the mesh-based motion estimation and interpolation is described. Section 3 addresses the problem of motion discontinuity and occlusions. The rate-distortion performances of the DVC scheme obtained with these motion models are described in Section 4. Conclusion and future directions are given in Section 5.

## 2. MESH-BASED MOTION COMPENSATION

### 2.1. Mesh construction

The side information extraction process comprises three steps: mesh construction for the original key frame, estimation of mesh position in the reference key frame, and bidirectional frame interpolation using obtained mesh-based motion model. A Delaunay triangulation algorithm is used to construct the 2D mesh. Here a regular triangle mesh is considered to avoid requiring a segmentation mask. In order to support hierarchical multi-resolution approach to motion estimation a set of nested regular Delaunay triangulations with different cell sizes is associated with the frame.

### 2.2. Mesh position estimation

The motion estimation aims at finding an optical flow which minimizes the displaced frame difference between two given frames. The mesh-based optical flow model describes a displacement  $\vec{u}(s) = (dx(s), dy(s))$  of every point  $s$  belonging to a certain triangle  $T$  through the displacements of its vertices (1).

$$\begin{cases} dx(s) = \sum_{j \in T} w_j(s) dx_j \\ dy(s) = \sum_{j \in T} w_j(s) dy_j \end{cases}, \quad (1)$$

where  $w_j(s)$  are the weights calculated as barycentric coordinates of  $s$  with respect to the vertex  $j$ , and where  $(dx_j, dy_j)$  is a motion vector of vertex  $j$ .

The optimal parameters  $\{(dx_j, dy_j)\}_{j=1}^N$  of the mesh-based motion model ( $N$  is a total number of vertices in the mesh) are found by applying a sum-of-squares minimization technique to the MSE of the displaced frame difference (2):

$$MSE = \sum_{s \in \Omega_1, (s+\vec{u}(s)) \in \Omega_2} [I_1(s) - I_2(s+\vec{u}(s))]^2, \quad (2)$$

where  $I_1(s)$  and  $I_2(s)$  are the original and reference key frames respectively, adjacent to the current group of

pictures (GOP), and where  $\Omega$  defines the support for the estimation. The vector  $\vec{u}(s) = (dx(s), dy(s))$  is computed according to (1).

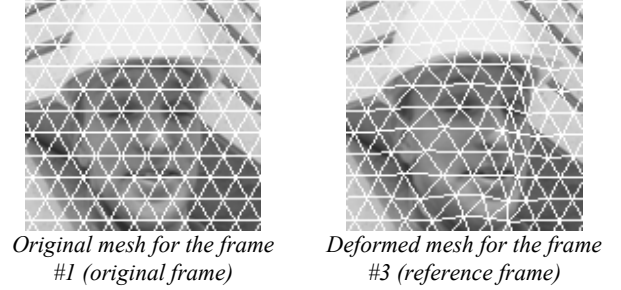


Figure 1 – Mesh deformation

Using a Newton-Gauss differential algorithm, the minimization problem turns out to be a problem of finding a solution for the following system of  $2N$  equations:

$$\begin{cases} \frac{\partial MSE}{\partial dx_j} = 0 \\ \frac{\partial MSE}{\partial dy_j} = 0 \end{cases}, j = \overline{1, N}, \quad (3)$$

The resolution of the first equation in (3) for a small increment of displacement  $\delta \vec{u}$ , leads to:

$$\begin{aligned} & \sum_{s \in \Omega_1, (s+\vec{u}(s)) \in \Omega_2} \sum_m \nabla_x I_2(s+\vec{u}) \cdot w_j(s) \cdot w_m(s) \cdot \\ & \cdot (\nabla_x I_2(s+\vec{u}) \delta dx_m + \nabla_y I_2(s+\vec{u}) \delta dy_m) = 0, \quad (4) \\ & = \sum_{s \in \Omega_1, (s+\vec{u}(s)) \in \Omega_2} -w_j(s) \cdot \nabla_x I_2(s+\vec{u}) \cdot \nabla_t I_1(s) \end{aligned}$$

where  $m$  is a node of the mesh. We finally obtain:

$$\begin{aligned} & \sum_{s \in \Omega_1, (s+\vec{u}(s)) \in \Omega_2} w_j(s) \nabla_x I_2(s+\vec{u}) \cdot \\ & \cdot \left[ \sum_m w_m(s) \nabla_s I_2(s+\vec{u}) \cdot \delta \vec{u}(m) + \nabla_t I_1(s) \right] = 0 \end{aligned}, \quad (5)$$

where  $\nabla_s$  and  $\nabla_t$  denote spatial and temporal gradients respectively. The system (5) is a huge sparse system (only few elements  $w_m(s)$  and  $w_j(s)$  are not null), so the smoothing of the motion field is needed. The minimization of (5) is carried out by a Levenberg-Marquard algorithm. To improve the convergence speed, a hierarchical multi-resolution approach is used, which also reduces the probability of falling into a local minimum. Figure 1 shows the deformed mesh obtained at the end of this step.

### 2.3. Frame interpolation

The motion model obtained at the previous stage corresponds to the motion between two adjacent key frames  $I_1$  and  $I_2$  (at time instants  $T_1$  and  $T_2$ ). According to the distributed video coding with side information paradigm, to

be able to decode a Wyner-Ziv frame  $I_{wz}$  at the time instant  $T_{wz}$  ( $T_1 < T_{wz} < T_2$ ) the decoder must have an approximation of it. This approximation is obtained using linear interpolation of frames  $I_1$  and  $I_2$  along motion trajectories at each pixel assuming that motion was uniform between  $T_1$  and  $T_2$ .

### 3. HYBRID INTERPOLATION

The mesh-based motion model provides improved interpolation accuracy compared with the block-based motion model when the motion field varies smoothly in a spatial domain. However, in real video sequences, scenes contain objects' boundaries and complicated motion along with smoothly varying motion. For such discontinuities in motion field block-based motion estimation and interpolation sometimes is more effective due to the locality of the optical flow optimization. Our experiments have shown that there are exist areas within the same frame where mesh-based interpolation is preferable to block-based interpolation, and vice versa (Fig. 2a and 2b).

This observation naturally leads to the following improvement. For every pixel of the interpolated frame we suggest the following locally adaptive interpolation formula, which takes into account local reconstruction error (MSE) in a small window for both methods:

$$Y(x, y) = \begin{cases} I_{comp}(x, y, \vec{b}) & \text{if } MSE(x, y, \vec{b}) < MSE(x, y, \vec{m}), \\ I_{comp}(x, y, \vec{m}) & \text{otherwise} \end{cases} \quad (6)$$

where

$$MSE(x, y, \vec{d}) = \sum_{(i,j) \in Window} \left[ I_1(x+i+b_x^B, y+j+b_y^B) - I_2(x+i+m_x^F, y+j+m_y^F) \right]^2,$$

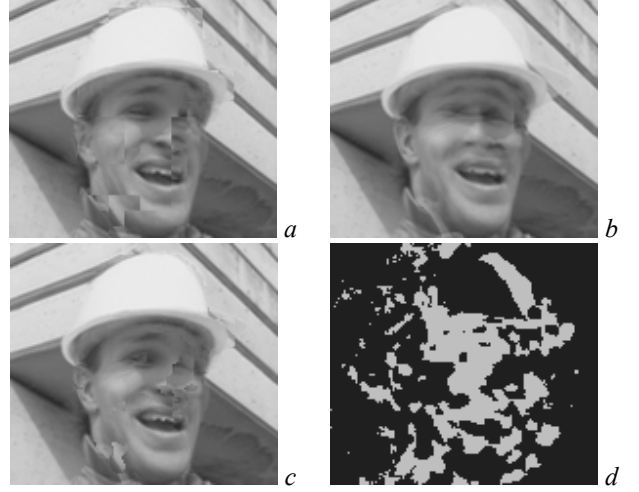
and  $\vec{b}$  and  $\vec{m}$  are motion vectors for pixel at position  $(x,y)$  (indices B and F stand for backward and forward constituents of vectors) estimated by block-based and mesh-based motion estimation respectively,

$$I_{comp}(x, y, \vec{d}) = \frac{T_2 \cdot I_1(x+i+d_x^B, y+j+d_y^B)}{T_1+T_2} + \frac{T_1 \cdot I_2(x+i+d_x^F, y+j+d_y^F)}{T_1+T_2}$$

is a result of frame interpolation with motion vector  $\vec{d}$ .

To improve the hybrid interpolation performance in scenes with very high but uniform motion (fast camera pan) a small improvement has been done. Our observations have shown that when camera motion exceeds search range for the block-based motion estimation, the motion field produced by it tends to be random. At the same time MSE for many pixels even with erroneous motion vectors is not high enough to overcome MSE for the mesh-based interpolation at the same pixel with good motion vector due

to repetitive patterns existing within the frame. To resolve this problem we propose to apply a penalty to all pixels within each block which is based on the dissimilarity of the



Frame interpolated with block-based(a), mesh-based(b) and proposed hybrid(c) motion estimation. Switching mask for hybrid interpolation(d) – light pixel for block-based and dark pixels for mesh-based.

**Figure 2** – Hybrid MC interpolation result

motion field in the neighborhood of this block. More precisely, the formula (6) is changed to (7):

$$Y(x, y) = \begin{cases} I(\vec{b}) & \text{if } MSE(\vec{b}) \cdot f(Pen(\vec{b})) < MSE(\vec{m}), \\ I(\vec{m}) & \text{otherwise} \end{cases} \quad (7)$$

where  $Pen(\vec{b})$  is the average Euclidean distance between the vector  $\vec{b}$  of the block containing  $(x,y)$  and vectors for nine neighboring blocks, and  $f(d)$  is a monotonous function. Coordinates  $(x,y)$  are omitted on the right for compactness. In our experiments we calculated  $Pen(\vec{b})$  after bidirectional motion estimation step but before spatial motion smoothing step in block-based motion estimation algorithm [7].

Figure 2d shows the binary mask corresponding to the switching rule (7) (MSE is calculated in a window 5x5), and Figure 2c presents the final result of the proposed hybrid interpolation.

### 4. EXPERIMENTAL RESULTS

The R-D performances of the mesh and hybrid MC interpolation approaches are assessed using a transform-domain distributed video codec with a GOP length equal to 2. The block-based hierarchical full search motion estimation is based on the algorithm in [7], with block sizes of 16x16 and 8x8 (coarse and fine levels),  $\pm 32$  pixels for the search range, and  $\pm 4$  pixels for the search range of the bi-directional motion estimation.

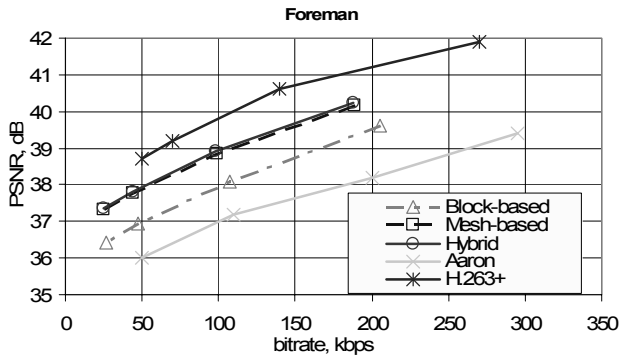


Figure 3 – R-D performance for the Foreman sequence

The R-D performances obtained with the different approaches for the luminance component of the wyner-ziv frames of the Foreman QCIF-30Hz sequence are shown in Fig. 3. The Wyner-Ziv frame rate is 15 frames per second. The Wyner-Ziv frames have been quantized at different quantization levels (corresponding to different numbers of bitplanes)  $\{1,3,5,7\}$ , to obtain the four rate-distortion points, while the key frames have been compressed at the same high quality for all R-D points (the h.264 quantization parameter is equal to 6). The R-D performance of another state-of-the-art transform-domain Wyner-Ziv codec from [5] with GOP=2 is also included as benchmarking (denoted as Aaron), as well as the R-D performance for the P-frames of the H.263+ codec in the I-P-I-P mode.

Mesh-based motion estimation provides noticeably better results than standard block-based motion estimation; the improvement is up to 1dB. It is also worth noticing that due to the side information of better quality, i.e. to increased correlation between the side information and the Wyner-Ziv encoded information, the number of parity bits requests at the decoder is reduced. However, the performance gap between the mesh and block based solutions decreases when the key frames are encoded at lower rates (that is quantized more coarsely) (see Fig. 4). The mesh-based approach is less robust to the key-frame quantization noise. The global setting of the key-frame parameters and side information extraction tools, inherent to the rate control strategy turns out to be a critical issue to be addressed. Note however that, in the experiments, the parameters used for the Laplacian correlation model for both approaches are those computed with side information generated by the block-based MC interpolation approach. Further improvement should be expected by adapting the correlation model parameters to the mesh generated side information.

## 5. CONCLUSION

In this paper, we have studied the performance of DVC with mesh-based motion-compensated interpolation. A hybrid

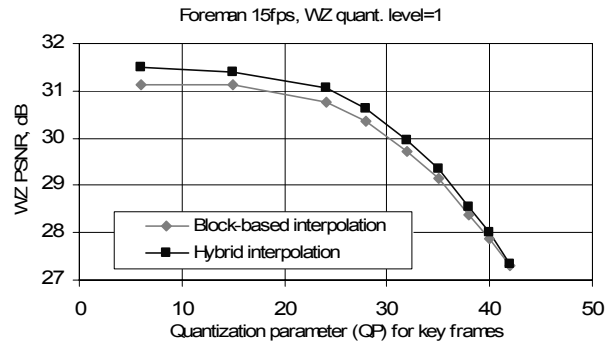


Figure 4 – Hybrid and block interpolation performance influenced by the key frame quantization parameter (QP).

motion estimation and interpolation approach combining block-based and mesh-based models is also described. Future efforts will be dedicated to improving the hybrid interpolation criterion, on estimating the parameters of the Laplacian correlation model. The potential advantage of mesh-based solutions for varying the GOP size will also be investigated.

## Acknowledgements

We are thankful to the IST and the Discover software development team for having provided the block-based DVC codec.

## REFERENCES

- [1] J. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Trans. on Information Theory*, Vol. 19, No. 4, July 1973.
- [2] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", *IEEE Trans. on Information Theory*, Vol. 22, No. 1, January 1976.
- [3] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv Coding of Motion Video", *Proc. 36th Asilomar Conference on Signals, Systems and Computer*, Pacific Grove, USA, Nov. 2002.
- [4] R. Purit and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles", *Proc. Allerton Conference on Communication, Control and Computing*, Oct. 2002.
- [5] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv Codec for Video", *Proc. SPIE Conference on Visual Communication and Image Processing*, Jan. 2004.
- [6] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, "Distributed Video Coding", *Proc. IEEE, Special issue on advances in video coding and delivery*, vol. 93, No. 1, pp. 71-83, Jan. 2005.
- [7] J. Ascenso, C. Brites and F. Pereira, "Improving Frame Interpolation With Spatial Motion Smoothing For Pixel Domain Distributed Video Coding", *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak Republic, July 2005.