RATE DISTORTION ANALYSIS IN A DISPARITY COMPENSATED SCHEME

Valentina Davidoiu[†], *Thomas Maugey*[†‡], *Beatrice Pesquet-Popescu*[†], *Pascal Frossard*[‡]

[†]*TÉLECOM ParisTech*, [‡]*École Polytechnique Fédérale de Lausanne*
{davidoiu,pesquet}@telecom-paristech.fr, {thomas.maugey,pascal.frossard}@epfl.ch

ABSTRACT

This paper addresses the problem of rate distortion analysis in the context of multi-view image coding, where images are predicted via disparity compensation based on depth map. We first present an analytical model for the variance of the residual error in a predicted frame when the prediction is done with the help of a compressed depth map. This residual variance model presents a convenient expression that separates the different error origins (reference frame quantization, depth map coding, and motion activity). We then validate the novel analytical model by testing separately its different underlying hypotheses. Finally, we illustrate an application of our analytical model in a simple bit allocation problem where the objective is to determine the optimal distribution of a global bit budget among reference frame, depth map and disparity-compensated frame. We observe that the optimal allocation given by the analytical model corresponds in practice to the best rate distribution for high bitrate, which confirms the potential of the proposed model in the design of rate-controlled multi-view coding algorithms.

***Index Terms*—** Video-plus-depth, motion vectors, bit-rate allocation

1. INTRODUCTION

A typical multi-view video sequence consists in a set of N temporally synchronized video streams coming from N cameras that capture a real world scene from different viewpoints. This multi-view video is widely used in 3DTV and free viewpoint TV (FTV) systems. Considering the data volumes associated with multi-view video systems which have to be encoded, decoded and rendered, efficient compression is crucial for the success of this technology. Recently, several algorithms have proposed to reduce redundancy between images through the estimation of depth information. The depth estimation can be used for view prediction with the help of Depth Image Based Rendering (DIBR) algorithms that are based on warping a camera view to another view. Given one view with its depth information, the technique theoretically has the ability to render a new synthetic view at a different position. However, DIBR algorithms are quite sensitive to depth discontinuities and disocclusions. The quality of the depth map has a strong influence on the image estimation and thus the compression efficiency of multi-view coding. Various methods have been studied to exploit the depth maps characteristics like 3-D motion estimation or new distortion metrics that take into consideration camera parameters and global video characteristics [1]. In addition, the statistical dependencies between the temporal and inter-view reference pictures have been exploited for improving the temporal and inter-view prediction structures for better compression [2]. But the distortion of the reconstructed views depends not only on the quality of the depth map, but also on the quality of the reference images and the characteristics of the image content.

In this paper, we propose a rate-distortion analysis of depth-based multi-view coding schemes, where we highlight the relative importance of the quality of the depth map, the reference frame and the residual error in the overall view reconstruction quality. Based on the model proposed in [3] for the distributed video coding framework and based on the work proposed in [4] for estimating the distortion of synthesized views (with no original frames), we propose an analytical rate-distortion model that describes the global behavior of the multi-view coding algorithm when the depth map is explicitly encoded. We validate the model with a series of experiments that emphasize the contributions of the different sources of errors in the global distortion. We finally illustrate the application of the novel rate-distortion model to a simple problem of rate allocation at encoder, which targets the optimal distribution of rate shares to the encoding of the reference frame, the depth map or the residual error. We confirm by coding experiments that the theoretical rate allocation corresponds to the optimal coding strategy in practice for high bitrates. This highlights the potential of our rate-distortion model in the design of efficient multi-view image encoders. However, for low bitrates, the model hypotheses are not completely verified and the rate prediction becomes less precise.

This paper is organized as follows. In Sec. 2 we introduce the model for the right frame prediction error distortion in a stereo coding system. Then, in Sec. 3, some experiments are performed in order to validate the proposed solution. In Sec. 4 we use this model to solve a rate allocation problem and finally, we conclude in Sec. 5.

2. PROPOSED DISTORTION MODEL

2.1. Hypotheses and calculation

An accurate rate distortion model plays an important role in multimedia compression and transmission due to its efficiency in computation and low complexity. Based on classical assumptions and on fundamental results in information theory, our model yields an interesting expression of frame estimation error which is separated in three independent terms: the error coming from the term related to the quantization of the reference frame, from the quantization of the disparity map and the one coming from the motion/disparity error.

We set the problem as illustrated in Fig. 1, obtaining in this way a simple expression for the variance of the frame estimation error. In this scheme we denote by R and L the images for the right and the left eyes, $\tilde{L}$ is the quantized version of L, $\mathbf{e_R}$ the right frame estimation error with its quantized version $\widetilde{\mathbf{e}}_\mathbf{R}$, and we also introduce the following notations: Q_L the quantization of the left image, Q_e the quantization of the prediction error, Q_D the quantization of the disparity map and DC the disparity compensation operation. In the following, $\mathbf{p} = (x, y)$ denotes the pixel in line x and column y of an image.

Our goal in this section is to obtain a mathematical expression

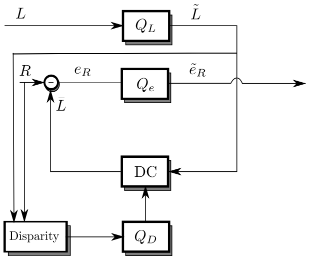


Fig. 1. Disparity compensated coding scheme.

for the variance of the frame error estimation $\mathbf{e_R}$, expressed as:

$$\sigma^2_{e_R} = \mathbf{E}[e_R(\mathbf{p})^2] = \mathbf{E}[(R(\mathbf{p}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}}))^2], \qquad (1)$$

where $d\mathbf{p}$ is the disparity vector associated with the position $\mathbf{p}$ in the left image L. This disparity vector is obtained thanks to the depth value, $Z(\mathbf{p})$, and the camera parameters (for registered cameras, the relation is: $d\mathbf{p} = fD/Z(\mathbf{p})$ where f is the focal length and D the distance between the views). $d\widetilde{\mathbf{p}}$ corresponds to the disparity value obtained with the coded depth map. Actually, this error can be expressed as:

$$\sigma^2_{e_R} = \mathbf{E}[(\underbrace{R(\mathbf{p}) - L(\mathbf{p} - d\mathbf{p})} + \underbrace{L(\mathbf{p} - d\mathbf{p}) - L(\mathbf{p} - d\widetilde{\mathbf{p}})} + \underbrace{L(\mathbf{p} - d\widetilde{\mathbf{p}}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}})})^2]. \qquad (2)$$

As highlighted in Eq. (2), the six terms can be grouped two by two. The first term corresponds to the disparity estimation error, while the second term is related to the coding of the disparity field when compensating the reference frame. The third term can be seen as the quantization of the reference frame. These three errors do not depend on the same quantity (quantization step, intra frames, quantization of the disparity map and type of the disparity). We thus assume the following hypothesis (that is tested in the next section):

Hypothesis 1 *The cross terms, ie. the terms containing two different errors in the development of Eq. (2), are supposed to be negligible. In other words, the three terms of Eq. (2) are decorelated.*

With this assumption, the following expression is obtained:

$$\sigma^2_{e_R} \cong \mathbf{E}[\underbrace{(R(\mathbf{p}) - L(\mathbf{p} - d\mathbf{p}))^2}_{\text{disparity estimation error } (M_{L,R})}] + \mathbf{E}[\underbrace{(L(\mathbf{p} - d\mathbf{p}) - L(\mathbf{p} - d\widetilde{\mathbf{p}}))^2}_{\text{Disparity coding error } (\sigma^2_{depth})}]$$

$$+ \mathbf{E}[\underbrace{(L(\mathbf{p} - d\widetilde{\mathbf{p}}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}}))^2}_{\text{Reference frame quantization error } (\sigma^2_{Q_L})}] \qquad (3)$$

$$\cong M_{L,R} + \sigma^2_{depth} + \sigma^2_{Q_L} \qquad (4)$$

where we introduced the following notations: $M_{L,R}$ is the variance of the disparity estimation error (with the non-quantized frames and the non-quantized disparity vector field), σ^2_{depth} the variance of the disparity coding error that represents the difference between the reference frame compensated with the disparity map and the reference frame compensated with the disparity map encoded. We assume that the third term, $\sigma^2_{Q_L}$, corresponds to the quantization error. In other words, we assume the following hypothesis:

Hypothesis 2 *The term* $\mathbf{E}[(L(\mathbf{p} - d\widetilde{\mathbf{p}}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}}))^2]$ *in equation (3) can be approximated by* $\mathbf{E}[(L(\mathbf{p}) - \widetilde{L}(\mathbf{p}))^2]$ *and then can be assimilated with the quantization error of the reference frame.*

In the next paragraph we study the behavior of these terms in different contexts.

3. MODEL VALIDATION

In this section we present some experiments in order to validate the hypotheses of the model introduced in Sec. 2. The 4 multi-view video sequences in our tests have a spatial resolution reduced to 512×384 and have been all rectified. They could be considered as representative since they are different (indoor/outdoor, low/high motion, etc.). For the experiments, the disparity maps $d\mathbf{p}$ are created using a dense estimation method while their coding is done by a block-based segmentation and H.264/AVC compatible coding as presented in [5]. The reference left frame is encoded as an intra frame, using H.264/AVC, while the disparity map is encoded using JSVM 9.15 [6].

3.1. Decorrelation between the quantization error and the disparity estimation term

Hypothesis 1 is the key assumption of our model because it leads to a separation between the error coming from the term related to the quantization of the reference frame, from the quantization of the disparity map and from the disparity error. The following expressions represent the cross terms:

$$\sigma_{M_{L,R},depth} = \mathbf{E}[(R(\mathbf{p}) - L(\mathbf{p} - d\mathbf{p}))(L(\mathbf{p} - d\mathbf{p}) - L(\mathbf{p} - d\widetilde{\mathbf{p}}))]$$
$$\sigma_{M_{L,R},Q_L} = \mathbf{E}[(R(\mathbf{p}) - L(\mathbf{p} - d\mathbf{p}))(L(\mathbf{p} - d\widetilde{\mathbf{p}}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}}))]$$
$$\sigma_{depth,Q_L} = \mathbf{E}[(L(\mathbf{p} - d\widetilde{\mathbf{p}}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}}))(L(\mathbf{p} - d\widetilde{\mathbf{p}}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}}))].$$

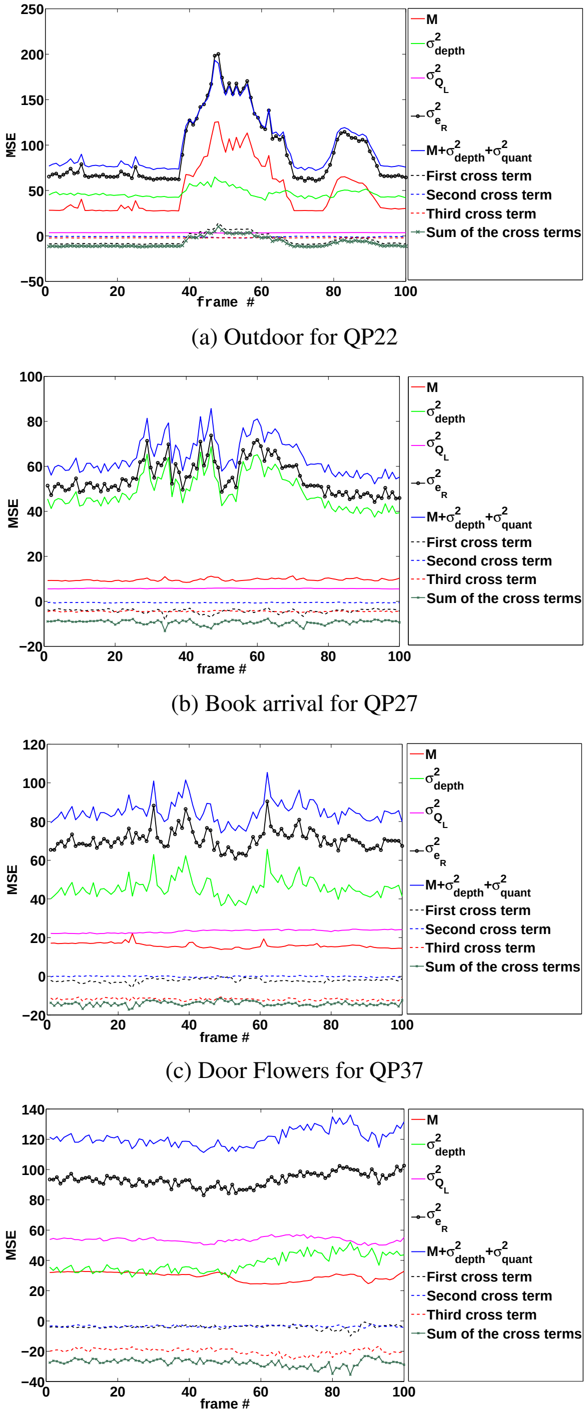
When developping all the terms in Eq. (2) and taking into account Hypothesis 2, we get:

$$\sigma^2_{e_R} = M_{L,R} + \sigma^2_{depth} + \sigma^2_{Q_L} + 2(\sigma_{M_{L,R},depth} + \sigma_{M_{L,R},Q_L} + \sigma_{depth,Q_L})$$

Several experiments have been done in order to verify the validity of Hypothesis 1. For several sequences and for several QP, we estimate the intensity of the cross terms and their influence on the difference between the real distortion $\sigma^2_{e_R}$ and the approximation of Eq. (3).

The curves in Fig. 2 show the evolution of $\sigma^2_{e_R}$ (in circle marked solid black line) and of the approximation $M_{L,R} + \sigma^2_{depth} + \sigma^2_{Q_L}$ (in solid blue line) for all the video sequences at different QPs. These experiments also indicate the evolution of the $M_{L,R}$ (in solid red line), σ^2_{depth}(in solid green line), $\sigma^2_{Q_L}$(in solid magenta line), and of the cross terms: $\sigma_{M_{L,R},depth}$ (in dashed black line), $\sigma_{M_{L,R},Q_L}$ (in dashed blue line) and σ_{depth,Q_L}(in dashed red line). The sum of these cross terms (in cross marked solid green line) and the sum of all the terms which represent the development of the Eq. (2), without the decorrelation assumption, are also displayed.

For videos such as "Book Arrival", "Door Flowers", "Leaving Laptop" having low disparity, we have drawn the following conclusions regarding the influence of the cross terms on the model. The first term $\sigma_{M_{L,R},depth}$ is the most important among the cross terms, and it depends on the content of the sequence. Moreover, its evolution explains the difference between $M_{L,R} + \sigma^2_{depth} + \sigma^2_{Q_L}$(in plain blue line) and $\sigma^2_{e_R}$ (in circle marked solid black line). Concerning the two other terms, we remark that the second term $\sigma_{M_{L,R},Q_L}$ is almost constant wrt the quantization steps, and the third cross term σ_{depth,Q_L} decreases with the increasing value of QP. At high bitrates we have a good estimation of the distortion error, since the term



(a) Outdoor for QP22

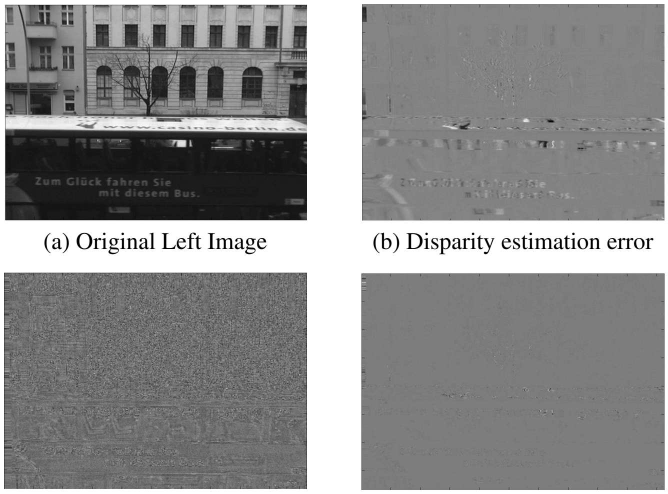
(b) Book arrival for QP27

(c) Door Flowers for QP37

(d) Leaving Laptop for QP42

Fig. 2. Evolution of the errors measured on different video sequences at different quantization steps.

with the biggest importance is $M_{L,R}$. For multi-view sequences like "Outdoor", one can observe a different behavior for the three cross terms, because a larger disparity is present. The first cross term $\sigma_{M_{L,R},depth}$ behaves like in the sequence "Indoor", being the most important quantity which influences the sum for the three cross terms, only that it looses its importance because the second term $\sigma_{M_{L,R},Q_L}$ has an important role for the frames where the inter-view variation is very significant. The third cross term σ_{depth,Q_L} decreases with the increasing value for QP. According to Fig. 3, we can see that the disparity estimation error and the reference quantization error are *decorrelated* (then $\sigma_{M_{L,R},Q_L}$ is low), and we can also observe that our model estimates very well the frame estimation error if the sequence has high disparity. For this case a significant quantity is the first cross term, $\sigma_{M_{L,R},depth}$. In "Indoor" sequences like "Book arrival", "Door flowers" and "Leaving laptop" where there is no significant disparity, we can observe that the most important term is σ^2_{depth}. The novelty of the proposed model is precisely to take into account this quantity.



(a) Original Left Image

(b) Disparity estimation error

(c) Reference quantization error

(d) The first crossed term

Fig. 3. (a) frame 50 of "Outdoor", (b) difference between the right frame and left frame compensated with the disparit, (c) difference between the reference frame compensated with the disparity encoded and the reference frame encoded compensated with the disparity, and (d)the first crossed term.

3.2. Approximation for quantization distortion

In Hypothesis 2, it is assumed that the error between the compensated reference frame and the compensated quantized reference frame can be assimilated to the simple quantization error of the reference frame, since the average is performed over the frame (and a compensation is just a block displacement).

QP	22	27	31	37	42
Outdoor	0.7925	0.1053	0.0460	0.0723	0.0436
Book Arrival	0.1884	0.8376	1.0408	0.8885	0.3452
Door Flowers	0.8593	1.0702	1.1324	0.8973	0.2631
Leaving Laptop	0.7620	0.7998	0.7541	0.0516	0.8970
Average	0.6505	0.7032	0.7433	0.4774	0.3872

Table 1. Table1. Error (%) between the two quantities $\mathbf{E}[L(\mathbf{p}) - \widetilde{L}(\mathbf{p})]$ and $\mathbf{E}[L(\mathbf{p} - d\widetilde{\mathbf{p}}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}})]$ for 4 multi-view video sequences (512x384, 100 frames) at different quantization steps.

To confirm this hypothesis a series of tests have been performed, in which we have evaluated the behavior for these two values. After testing different video sequences at different quantization steps, we calculate the difference between $\mathbf{E}[(L(\mathbf{p}) - \widetilde{L}(\mathbf{p}))^2]$ and $\mathbf{E}[(L(\mathbf{p} - d\widetilde{\mathbf{p}}) - \widetilde{L}(\mathbf{p} - d\widetilde{\mathbf{p}}))^2]$, and normalize it with respect to the value of the quantization error of the reference frame. Table 1 presents the error between the two quantities, proving that the two distortions are very similar for different quantization steps.

4. RATE ALLOCATION EXAMPLE

In this section, we illustrate the application of the proposed model for a rate allocation problem. More precisely, we want to determine theoretically how to share a total bitrate R between the depth map, the residual and the reference frame rates. This allocation has to be optimal, in the sense that it minimizes the total distortion:

$$D_{Total} = D_{right} + D_{left},$$

which is the sum of the distortion of the right frame and the left frame. For this purpose, we aim at minimizing the unconstrained criterion

$$J = D_{Total} + \lambda(R_{depth} + R_{e_R} + R_L - R).$$

The expression of the total distortion is obtained by first adopting the classical rate-distortion (RD) model for a source X, which states that the distortion D_X is equal to $\mu_X \sigma_X^2 2^{-2R_X}$, where μ_X is a constant depending on the distribution that we estimate based on the theoretical work of Fraysse *et al.* [7], and R_X is the source rate. The variance σ_X^2 corresponds to the entire frame variance for the left frame, and the error variance for the right frame. This error variance is replaced by the error expression proposed previously (Eq.(4)). We analytically calculate the expressions of $\frac{\partial J}{R_L}$, $\frac{\partial J}{R_{e_R}}$ and $\frac{\partial J}{R_{depth}}$ and set them to zero for computing the optimum of the objective function. Firstly, we find that R_{depth} is the solution of:

$$R_{depth} + \frac{1}{2}\log_2\left(4\alpha_R\alpha_L\sigma_L^2(\sigma_{depth}^2 + M_{L,R})\right) - \frac{1}{2}\log_2\left((2\sigma_{depth}^2 + \frac{\mathrm{d}}{\mathrm{d}R_{depth}}\sigma_{depth}^2 + 2M_{L,R})^2\right) - R = 0.$$

We suppose that σ_{depth}^2 only depends on R_{depth}. Fig 4 shows the behavior of σ_{depth}^2 when R_{depth} is varying. The relationship between these quantities varies for every sequence and every temporal instant even if one can observe in each case that the evolution looks monotonic and convex. The previous equation can then be rapidly solved with a Newton method in order to find the optimal value R_{depth}^*. Then, we are able to write the following relationship between R_L and R_{e_R}:

$$R_L - R_{e_R} = \log_2\left(-\frac{2\alpha_L\sigma_L^2}{2\alpha_R(\sigma_{depth}^2 + M_{L,R})}\right)$$
$$R_L + R_{e_R} = R - R_{depth}^*$$

where σ_{depth}^2 is fixed at a value corresponding to R_{depth}^*. We deduce the optimal values for the rates. In practice, if we know the behavior of $\sigma_{depth}^2(R_{depth})$, we are able to elaborate a rate allocation algorithm based on the proposed model (the other parameters as $M_{L,R}$ and σ_L^2 are directly calculated online). In order to test the reliability of the proposed rate allocation solution, we perform the following test on "Outdoor" sequence (camera 1 and 3). For a fixed total rate, we experimentally determine the optimal allocation (a full search with a percentage step of 5% for each rate), and compare it to the theoretical one. For a total bitrate of $R = 2.33$ bpp, we find that the optimal experimental rate repartition was: $R_L = 0.60R$, $R_{e_R} = 0.25R$ and $R_{depth} = 0.15R$, while the proposed solution gives us: $R_L = 0.64R$, $R_{e_R} = 0.22R$ and $R_{depth} = 0.14R$ which is an acceptable prediction and show the potential of our solution. However, for lower bitrates we obtain a less acceptable rate prediction, since at low bitrate ($R = 0.94$ bpp), we predict a repartition:$R_L = 0.70R$, $R_{e_R} = 0.25R$ and $R_{depth} = 0.05R$ while the real optimal repartition is $R_L = 0.60R$, $R_{e_R} = 0.20R$ and $R_{depth} = 0.20R$. This shows that there is still some work to do to obtain an accurate rate prediction at low bitrates.

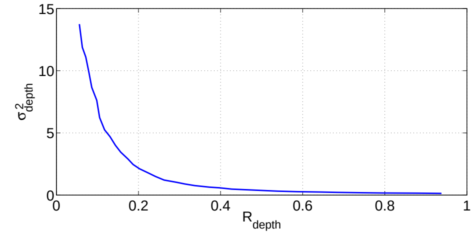


Fig. 4. σ_{depth}^2 as a function of R_{depth} for "Outdoor", 512×384

5. CONCLUSION

In this paper, a new distortion model has been proposed. At high bitrate, experimental results show that the estimation of the frame estimation error is good, and for video with significant disparity the proposed model shows a very good estimation of the frame estimation error. We also propose one application of this model to a rate allocation problem, and we observe that, for high bitrates, our model-based allocation solution gives a good prediction of the optimal rates of the reference image, the depth map and the residual. The solution presents however some limits for low bitrates.

6. REFERENCES

[1] Woo-Shik Kim, Antonio Ortega, PoLin Lai, Dong Tian, and Cristina Gomila, "Depth map distortion analysis for view rendering and depth coding," in *ICIP'09: Proceedings of the 16th IEEE international conference on Image processing*, 2009, pp. 721–724.

[2] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient compression of multi-view depth data based on mvc," in *Proc. 3DTV Conference*, 7–9 May 2007, pp. 1–4.

[3] T. Maugey and B. Pesquet-Popescu, "Side information estimation and new symmetric schemes for multi-view distributed video coding," *J. on Visu. Commun. and Image Repr.*, vol. 19, no. 8, pp. 589–599, Dec. 2008, Special issue: Resource-Aware Adaptive Video Streaming.

[4] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model," *J. on Visu. Commun. and Image Repr.*, vol. 24, pp. 666–681, 2009.

[5] Ismaël Daribo, Wided Miled, and Béatrice Pesquet-Popescu, "Joint depth-motion dense estimation for multiview video coding," *Journal of Visual Communication and Image Representation*, vol. 21, no. 5-6, pp. 487 – 497, 2010, Special issue on Multi-camera Imaging, Coding and Innovative Display.

[6] *JSVM 9.15 software package,*, CVS server for the JSVM software at : http://iphome.hhi.de/.

[7] A. Fraysse, B. Pesquet-Popescu, and J.C Pesquet, "On the uniform quantization of a class of sparse source," *IEEE Trans. on Inform. Theory*, vol. 55, pp. 3243–3263, Jul. 2009.