A convex-optimization framework for frame-level optimal rate allocation in predictive video coding
Aniello Fiengo, Giovanni Chierchia, Marco Cagnazzo, Béatrice Pesquet-Popescu

To cite this version:
Aniello Fiengo, Giovanni Chierchia, Marco Cagnazzo, Béatrice Pesquet-Popescu. A convex-optimization framework for frame-level optimal rate allocation in predictive video coding. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2014, Florence, Italy. 2014, <10.1109/ICASSP.2014.6855023>. <hal-01202774>
A CONVEX-OPTIMIZATION FRAMEWORK FOR FRAME-LEVEL OPTIMAL RATE ALLOCATION IN PREDICTIVE VIDEO CODING

Aniello Fiengo1, Giovanni Chierchia1, Marco Cagnazzo1, and Béatrice Pesquet-Popescu1

1 Institut Mines-Télécom; Télécom ParisTech; CNRS LTCI, 75014 Paris, France

ABSTRACT

Optimal rate allocation is among the most challenging tasks to perform in the context of predictive video coding, because of the dependencies between frames induced by motion compensation. In this paper, we derive an analytical rate-distortion model that explicitly takes into account the dependencies between frames. The proposed approach allows us to formulate the frame-level optimal rate allocation as a convex optimization problem. Within this framework, we are able to achieve the exact solution in limited time (even for large-size problems), thanks to the flexibility offered by recent convex optimization techniques. Experiments on standard sequences demonstrate the interest of considering the proposed rate-distortion model and confirm that the optimal rate allocation ensures a better distribution of the total bit budget, with superior results (in the rate-distortion sense) with respect to the standard H.264/AVC rate control.

Index Terms— Video coding, rate distortion, convex optimization, resource allocation.

1. INTRODUCTION

Predictive coding is one of the best tools at hand to exploit temporal redundancies induced by motion. As a matter of fact, motion-compensated prediction plays a key role to reduce significantly the bit rate in state-of-the-art video codecs such as H.264/AVC [1] and H.265/HEVC [2]. Several techniques have been proposed in the literature to select the coding parameters that achieve an optimal trade-off between rate and distortion. Many conventional coding schemes tend to make these choices frame by frame [3, 4, 5, 6, 7, 8]. However, it is widely recognized that, from a rate-distortion standpoint, the optimal choice for a single frame may be potentially suboptimal for encoding the remaining frames, because of the chain of predictions created by motion compensation. Consequently, an optimization that jointly takes into consideration the dependencies between frames may yield a significant bit reduction [9, 10, 11, 12, 13].

In the last decades, a substantial research effort has been made to enlarge the optimization scope from single frames to groups of frames. A first class of techniques amounts at either considering long-term reference frames [14, 15] or segmenting video frames into objects to be coded separately [16, 17]. However, these approaches exploit only indirectly the temporal dependencies between frames. A more theoretical approach consists in embedding temporal dependencies in a rate allocation problem. In this context, one of the most relevant work dates back to Uz, Shapiro and Czigler [18], who provided an optimal strategy for performing frame-level bit allocation in presence of quantizer feedback. The point is that motion-compensated frames carry the quantization errors affecting the previous frames and, consequently, these errors propagate towards the prediction residuals of successive frames, resulting in multiple quantization errors when such residuals are coded. The basic result shown in [18] is that both easily-predicted frames and good predictors need to be allocated a higher bit rate, in order to limit the propagation of quantization errors. Similar insights were provided in [9] within an operational rate-distortion (R-D) framework, but this method requires to evaluate a set of operational R-D points for each frame, which makes the computational burden prohibitive in many applications, because of the involved multi-pass coding. Recent work [12] extended at pixel-level the rate allocation model in [18] and used it to drive an heuristic for selecting the optimal quantization parameters. Interestingly, the authors observed that a motion estimation approach which exploits temporal dependencies may lead to observable gains in the rate-distortion sense.

Related work. The optimal rate allocation is, in general, a non trivial problem. In the seminal work [18], to take into account the temporal dependency between a frame \( I_n \) and its predictor \( I_{n-1} \), the rate-distortion function was decoupled into two terms, the prediction error and the distortion of \( I_{n-1} \), leading to a recursive model that can be managed analytically. The same decoupling idea was recently rediscovered in [13], but instead of considering a recursive model, the allocation problem was formulated in terms of rate as a function of distortion and solved by resorting to a series of convex relaxations based on a first-order Taylor approximation. A similar error-splitting model was also employed in [19] within the distributed video coding (DVC) framework, in order to analyse the R-D performance of decoding strategies and propose new schemes for multi-view DVC.

Contributions. In this paper, we propose an efficient solution to exactly solve the frame-level bit allocation problem.
We use the same theoretical foundations as in [18], but

1. we extend the R-D model by letting the exponential decay vary at each frame, allowing us to better represent the intrinsics non-stationarities in a group of frames (in [18], it is explicitly required that the exponential decay needs to be the same for all the frames);

2. we provide an efficient solution based on proximal tools [20] in order to find the exact solution of the R-D problem (within the limit of the accuracy of the model), in limited time even for hundreds of frames.

The paper is organized as follows. Section 2 illustrates the proposed R-D model and discusses the differences with [13]. Section 3 describes the bit allocation problem and the algorithm to solve it. Section 4 provides model validation and illustrates the performance of the proposed rate allocation method. The conclusions are given in Section 5.

2. RATE-DISTORTION MODEL

Accurate Rate-Distortion (R-D) modelling plays a fundamental role in optimal bit allocation. Due to the different characteristics of frames, as well as and the sophisticated compression techniques employed in coding algorithms, analytic R-D modelling is still an open problem. According to the classical R-D theory [21], it is possible to express the relationship between R and D for a frame compressed at high bit rate as

\[ D = \alpha \sigma^2 2^{-2R} \]  

(1)

where \( \alpha \) and \( \sigma^2 \) are, respectively, the p.d.f. shape factor and the variance of the residual DCT coefficients, and \( D \) is measured as the Mean Square Error between the original frame and the reconstructed frame. As shown in [22], such type of analytic formula does not work well for hybrid video coders. We therefore present an alternative R-D model based on the work [18]. To do so, we consider a group of frames \( I = (I_0, \ldots, I_{N-1}) \) of size \( N > 0 \) and we denote by \( r = (r_0, \ldots, r_{N-1}) \) the rates used to encode such frames. We further assume that frame \( I_0 \) is intra encoded (I-frame) and, for \( n \geq 1 \), frame \( I_n \) is predicted from frame \( I_{n-1} \) and the prediction residual is encoded spatially (P-frames).

**I-frame.** For frame \( I_0 \), we follow the model proposed in [18] and employ the following R-D function:

\[ D_0 = \alpha_0 M_0 2^{-\beta_0 r_0}, \]  

(2)

where \( M_0 = \sigma_0^2 \) is the variance of \( I_0 \) and \( (\alpha_0, \beta_0) \) are model parameters estimated as explained in Section 4. It is worth noting the difference with the R-D function proposed in [13], which reads (after a simple inversion):

\[ \hat{D}_0 = \frac{a_0 \cdot G}{r_0 - c_0 \cdot G} - b_0. \]  

(3)

where \( (a_0, b_0, c_0) \) are model parameters and \( G \) is the average gradient of a frame.

**P-frame.** For frames \( I_n \), with \( n \geq 1 \), we choose a model that is very close to [18, 19, 13], but with a different parametrization, which consists in expressing the R-D function as:

\[ D_n = \alpha_n (M_n + D_{n-1}) 2^{-\beta_n r_n}, \]  

(4)

where \( M_n \) is the innovation of frame \( I_n \), i.e. the residual that would result if \( I_n \) was predicted (through a prediction function \( f_n \)) from a non-quantized reference frame \( I_{n-1} \)

\[ M_n = \mathbb{E} \{ [I_n - f_n(I_{n-1})]^2 \}, \]  

(5)

and \( (\alpha_n, \beta_n) \) are model parameters estimated in Section 4. The distortion of \( I_n \) is hence decoupled in two terms: the error due to the motion-compensated prediction of \( I_n \) and the distortion due to the quantization of the reference frame \( I_{n-1} \).

3. RATE ALLOCATION ALGORITHM

Optimal rate allocation consists in finding the vector of rates that minimizes the global distortion while keeping the total rate under a given budget \( \eta > 0 \),

\[ \min_{r \in [0, \infty]^N} \sum_{n=0}^{N-1} D_n(r_n, \ldots, r_0) \quad \text{s.t.} \quad \sum_{n=0}^{N-1} r_n \leq \eta. \]  

(6)

The R-D function of frame \( I_0 \) is given in Eq. (2), while the one for the \( n \)-th P-frame is given in Eq. (4). Therefore, we can demonstrate by the induction principle that the operational R-D function \( D_n \) actually depends on all the frames involved in the chain of predictions leading to \( I_n \), yielding

\[ D_n(r_n, \ldots, r_0) = \sum_{\ell=0}^{n} \alpha^{(n, \ell)} M_\ell 2^{-\sum_{j=\ell}^{n} \beta_j r_j}, \]  

(7)

where \( \alpha^{(n, \ell)} = \prod_{j=\ell}^{n} \alpha_j \). Note that Eq. (7) reduces to Eq. (2) when \( n = 0 \). To gain some insight into the solution of Problem (6), we introduce a vector \( u = (u_{n\ell})_{0 \leq n \leq N-1, 0 \leq \ell \leq n} \) defined as

\[ u_{n\ell} = \sum_{j=\ell}^{n} \beta_j r_j, \]  

(8)

which allows us to express the global distortion as a separable sum of exponentials

\[ F(u) = \sum_{n=0}^{N-1} \sum_{\ell=0}^{n} \alpha^{(n, \ell)} M_\ell 2^{-u_{n\ell}}. \]  

(9)

Therefore, Problem (6) can be reformulated as follows

\[ \min_{r \in \mathbb{R}^N} F(Lr) \quad \text{s.t.} \quad r \in C, \]  

(10)
where $L: \mathbb{R}^N \rightarrow \mathbb{R}^{N(N+1)/2}$ is the linear operator that maps the vector $r \in \mathbb{R}^N$ into the vector $u \in \mathbb{R}^{N(N+1)/2}$ defined in (8), and $C$ is the nonempty closed convex set defined as

$$C = \{ r \in [0, +\infty]^N \mid \sum_{n=0}^{N-1} r_n \leq \eta \}. \quad (11)$$

Among the many approaches proposed in the literature to solve this class of problems, we do not transform the constrained problem in Eq. (6) to a Lagrangian formulation, but rather manage the bit budget as a hard constraint. We resort here to proximal algorithms [20, 23], which can handle a wide class of convex optimization problems involving non-smooth penalizations and hard constraints. In particular, we employ the primal-dual M+LFBF algorithm recently proposed in [24], which guarantees the convergence (under weak conditions) in a reasonable time even for large-scale problems, offers robustness to numerical errors and its structure makes it suitable for parallel implementations.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed algorithm, we selected eight video sequences composed of 84 frames at resolution 352x288. We encoded all the sequences with a GOP structure IPP...P of size $N = 12$, a frame rate of 30 pictures per second and CABAC as entropy coder. For the motion estimation, we set the maximum search range to ±16. Moreover, we estimated the model parameters $(\alpha_n, \beta_n, M_n)_{0 \leq n \leq N-1}$ by encoding the sequences in H.264/AVC with the rate control disabled, in order to manually fix seven different quantization parameters: 10, 12, 14, ..., 22. For each frame, we recorded the values of $D$ and $R$ produced at the encoder output and, after a logarithmic transformation of $D$, we estimated the model parameters by resorting to a linear regression. We then encoded all the sequences with the rate control of H.264/AVC enabled, setting seven different target bit-rates ranging from 2.2 to 4.0 Mb/s and reported the results in the following.

Model validation. We start our experimental analysis by comparing the proposed R-D model with the one in [13]. The validation was performed on the basis of the $R^2$ metric [25], which was designed to quantitatively measure the degree of deviation from a given model (the closer the value of $R^2$ is to 1, the more accurate the model). Table 1 shows the $R^2$ values associated to the R-D function of I-frames given in Eq. (1)-(3). For all the sequences, the proposed model (2) shows superior fitting performance, giving $R^2$ values very close to 1 and higher than (1) and (3). We skip the comparison between the R-D functions associated to P-frames, as we got $R^2$ values very close to each other (the R-D models are very similar).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>$R^2$ with (2)</th>
<th>$R^2$ with (3)</th>
<th>$R^2$ with (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hall</td>
<td>0.997</td>
<td>0.455</td>
<td>0.975</td>
</tr>
<tr>
<td>foreman</td>
<td>0.995</td>
<td>0.912</td>
<td>0.908</td>
</tr>
<tr>
<td>football</td>
<td>0.985</td>
<td>0.849</td>
<td>0.762</td>
</tr>
<tr>
<td>container</td>
<td>0.985</td>
<td>0.774</td>
<td>0.824</td>
</tr>
<tr>
<td>coastguard</td>
<td>0.984</td>
<td>0.596</td>
<td>0.395</td>
</tr>
</tbody>
</table>

Comparison with H.264/AVC. To assess the validity of our rate allocation method, we compared it with the standard rate control algorithm of H.264/AVC [26]. For the sequences “foreman” and “hall”, Figure 1 reports the rates, as a function of frame numbers, which are allocated by the proposed method (red line) and by H.264/AVC (blue line), while Figure 2 displays the corresponding distortions. As we can see from Figure 1, the allocation by the proposed method is quite different than the reference, especially for I-frames. Also, our distribution of rates is more uniform within the GOP. This is an important result, because often the H.264/AVC encoder shows a sub-optimal greedy behaviour: it allocates the largest part of the bit budget to the first frames of the GOP and hence it rapidly runs out of bits for the remaining GOP frames, causing an increase of the global distortion. Moreover, we collected in Figure 3 the average distortion as function of the average rate, and in Table 3 the corresponding PSNR increments (due to space limitations, we can only report a small subset of indicators). These results show that the distortion achieved with the proposed method is always lower than the one obtained by the standard rate-control algorithm.

5. CONCLUSIONS

We have proposed a new algorithm to exactly solve the frame-level rate allocation problem arising in predictive video coding. The obtained results demonstrate that the analytical R-D model presented in Section 2 allows us to accurately describe the temporal dependencies in a group of frames. Furthermore, our experiments indicate that the optimal rate allocation, when supported by an accurate R-D model, attains better results (in the R-D sense) than the standard rate control in H.264/AVC. The higher performance of our approach is related to its ability to see beyond the first frames of the GOP and to keep the rate budget for the successive frames when necessary. This is in contrast with the greedy behaviour of H.264/AVC rate controller, which tends to allocate the largest part of the bit budget to the first frames of the GOP.
Fig. 1: Rates vs frame numbers.

Fig. 2: Distortion vs frame numbers.

Fig. 3: Comparison between actual and estimated distortion.


6. REFERENCES


