Improved Method to Select the Lagrange Multiplier for Rate-Distortion Based Motion Estimation in Video Coding

José Luis González-de-Suso, Amaya Jiménez-Moreno, *Student Member, IEEE,* Eduardo Martínez-Enríquez, *Student Member, IEEE,* and Fernando Díaz-de-María, *Member, IEEE*

Abstract—The motion estimation (ME) process used in the H.264/AVC reference software is based on minimizing a cost function that involves two terms (distortion and rate) that are properly balanced through a Lagrangian parameter, usually denoted as λ_{motion}. In this paper we propose an algorithm to improve the conventional way of estimating λ_{motion} and, consequently, the ME process.

First, we show that the conventional estimation of λ_{motion} turns out to be significantly less accurate when ME-compromising events, which make the ME process to perform poorly, happen. Second, with the aim of improving the coding efficiency in these cases, an efficient algorithm is proposed that allows the encoder to choose between three different values of λ_{motion} for the Inter 16x16 partition size. To be more precise, for this partition size, the proposed algorithm allows the encoder to additionally test $\lambda_{motion} = 0$ and λ_{motion} arbitrarily large, which corresponds to minimum distortion and minimum rate solutions, respectively. By testing these two extreme values, the algorithm avoids to make large ME errors.

The experimental results on video segments exhibiting this type of ME-compromising events reveal an average rate reduction of 2.20% for the same coding quality with respect to the JM15.1 reference software of H.264/AVC. The algorithm has been also tested in comparison with a state-of-the-art algorithm called CALM (Context Adaptive Lagrange Multiplier). Additionally, two illustrative examples of the subjective performance improvement are provided.

I. INTRODUCTION

MODERN video compression standards such as H.264/AVC [1] achieve a high compression efficiency by allowing the encoder to choose among many coding options, such as several partition sizes (PS), motion vectors (MV), reference pictures (Ref), and quantization parameters (QP). Therefore, the selection of the optimal combination of coding options from a rate-distortion (R-D) perspective becomes a critical task. Generally, the video codec performs an optimization task called rate-distortion optimization (RDO), which aims to find the coding option that minimizes a distortion measure subject to a given rate restriction:

$$\min_{\boldsymbol{\theta}} \{D(\boldsymbol{\theta})\} \text{ subject to } R(\boldsymbol{\theta}) \leq R_c, \qquad (1)$$

where $\boldsymbol{\theta}$ is a combination of the different coding options (PS, MV, Ref, QP, etc.); $D(\boldsymbol{\theta})$ represents the distortion between the original and the reconstructed coding unit; $R(\boldsymbol{\theta})$ is the rate needed to encode it (the number of bits needed to encode headers, MVs, Ref indexes, and residual transform coefficients); and R_c the maximum rate allowed (the rate constraint).

Using a Lagrange formulation, this constraint optimization problem turns to be an unconstrained problem [2]:

$$\min_{\boldsymbol{\theta}} \{J_{mode}\}$$
$$\text{with } J_{mode}(\boldsymbol{\theta}) = D(\boldsymbol{\theta}) + \lambda_{mode} R(\boldsymbol{\theta}), \qquad (2)$$

where λ_{mode} is the Lagrange multiplier that weights the relative importance between $D(\boldsymbol{\theta})$ and $R(\boldsymbol{\theta})$. A given value of λ_{mode} yields a solution $\boldsymbol{\theta}^*(\lambda_{mode})$ that turns out to be an optimal solution to the original RDO problem (1) for a particular value of $R_c = R(\boldsymbol{\theta}^*)$.

This solution involves testing each combination of coding options $\boldsymbol{\theta}$ for each coding unit, so that the overall D along all the coding units is minimized subject to the rate constraint. Dynamic programming can solve this problem [3], but this kind of approach is not feasible in practice due to the fact that computational complexity grows exponentially with the number of coding units. Therefore, in practical implementations of the H.264/AVC standard, such as the Joint Model (JM) [4], many simplifications are made in order to obtain a more efficient solution. These simplifications are outlined next. Hereafter, we select the macroblock (MB) as coding unit.

First, the MB independence hypothesis is made so that each MB can be coded and optimized independently of the rest. Although this hypothesis is not true because the $\boldsymbol{\theta}^*$ chosen for an MB actually depends on those $\boldsymbol{\theta}^*$ chosen for previously coded MBs, is a necessary approximation to obtain a practical solution of the optimization process.

Furthermore, in order to avoid testing all the QP values allowed by the standard for a given λ_{mode}, a relationship between the Lagrange parameter λ_{mode} and the QP has been experimentally derived ([5], [6]):

$$\lambda_{mode} = C \cdot 2^{(QP-12)/3}, \qquad (3)$$

where C is a constant that depends on the slice type.

Concerning the ME-related optimization, which obtains solutions for MVs and Refs, the evaluation of D and R in (2) for every potential MV would not be feasible since each

evaluation involves DCT-like transform computation, quantization, entropy coding, and inverse processes for reconstruction. The solution to this high-complexity problem consists on simplifying the MV search by minimizing a low-complexity cost function instead of directly minimizing J_{mode}. Thus, the ME process is usually formulated as the minimization of a second Lagrangian cost function:

$$\min_{MV, Ref} \{J_{motion}\}$$
$$\text{with } J_{motion} = D_{motion}(MV, Ref) +$$
$$+\lambda_{motion} R_{motion}(MV, Ref), \quad (4)$$

where D_{motion} is the sum of absolute differences (SAD) between the original and predicted blocks given by MV and Ref; R_{motion} is an estimation of the number of bits needed to encode the motion-related information; and λ_{motion} is a Lagrange multiplier.

Once the set of near-optimal MVs and their corresponding Refs have been found by the ME process, they are used to obtain the optimal PS by minimizing J_{mode}, what is referred to as the mode decision (MD) process. This MD process in the H.264/AVC context refers to the selection of the partition size for either intra- or inter- prediction; furthermore, the inter-prediction can use one or two (bi-prediction) reference images.

Finally, considering that D in J_{mode} is calculated as a sum of squared differences (SSD) and D_{motion} in J_{motion} is computed as a SAD, an experimental relationship between λ_{motion} and λ_{mode} has been established [5]:

$$\lambda_{motion} = \sqrt{\lambda_{mode}}. \quad (5)$$

At this point, given a QP value set by a rate-control algorithm in order to meet certain target rate R_c, the Lagrange multipliers can be estimated as in Eqs. (3) and (5), and the optimal MV and Ref, and PS can be obtained by minimizing J_{motion} and J_{mode}, respectively. These considerations allow the system to obtain a near-optimal representation of the MB, $\widehat{\boldsymbol{\theta}}^*$, with a very significant reduction of the computational cost in comparison to the optimal solution.

The aim of this work is to find a coding option closer to the optimal than the one obtained by the JM reference model, without incurring in a significant increase of the computational cost. Specifically, we allow the encoder to choose between different values of λ_{motion}, which is equivalent to change the relation between λ_{motion} and λ_{mode} Lagrange multipliers.

Since the discrete version of Lagrangian optimization introduced by Everett [2], many works have focused on this approach. For example, this Lagrangian optimization was used in [7] in a source coding application. In [8] an algorithm was proposed to obtain the optimal λ_{mode} parameter, best basis (wavelet packet), and quantization step size in a wavelet packet-based coding. In [9] particular QPs were selected for each coding block to minimize the D under a R_c in a HDTV digital recording application. Nevertheless, since these approaches turned out to be expensive in computational cost terms, many works were proposed to avoid testing all the possible QP values by modelling R and D as a function of the QP ([10]–[13]). Furthermore, the value of the λ_{mode} parameter to be used in the optimization process can be obtained

analytically using these $R - D$ models as $\lambda_{mode} = -\frac{dD}{dR}$ ([14], [15]).

Some works take into account the actual video content in the R and D modeling ([15]–[19]). Specifically, in [16] a R-D model was proposed in which the λ_{mode} selection depends on a parameter derived from the number of zero-quantized transformed residual coefficients, which was dynamically adapted. This approach had the disadvantage of decoupling QP and λ_{mode}, making the rate-control process more difficult. In [15] an algorithm was proposed to accurately select the value of λ_{mode} by considering a Laplace distribution of the quantized residual and adapting the λ_{mode} value to the actual video sequence, so that the overall coding efficiency was improved. They modeled R and D as function of QP, some features of the input sequence, and frame type. Then, λ_{mode} was obtained by following the corresponding analytical model. Although the model is elegant, its implementation turned out to be very complex. Besides, some hypothesis have to be made to build the model that require to implement specific solutions that are also used when these hypothesis are not met. In [17] a λ_{mode} estimation process for 3-D wavelet-based scalable video coding is proposed. λ_{mode} was accurately estimated from the target bit rate, some characteristics of the sequence, the motion estimation algorithm, and the wavelet filter, but the algorithm is conceived for medium to high bit rates. In [18] novel R and D models were proposed in the context of the emerging High Efficiency Video Coding (HEVC) standard. However, the estimation estimation of λ_{mode} from these models has been left for future work. In [19] another alternative for modeling the R and D terms based on a mixture of Laplacian distributions was proposed, but again the estimation of λ_{mode} was not discussed in the paper.

Some works have been reported concerning the λ_{motion} parameter. In [20] a linear model was established for both R_{motion} and D_{motion} to obtain analytically the optimal λ_{motion} value, but the method did not provide a significant improvement in performance against the reference model. The CALM method presented in [6] adjusted λ_{motion} for each block based on its context, that is, based on the Lagrangian cost of its neighboring blocks. This approach has been implemented in the JM reference software [4] since the 10.2 version.

In this paper, we propose a new approach that aims to provide the encoder with more freedom to decide on the λ_{motion} value and, thus, on the optimal MV selection process. It is based on the fact that the optimal λ_{motion} is actually content-dependent. The proposed method is simple and effective, and achieves good results in terms of PSNR and bit rate, outperforming the algorithm proposed in [6].

The rest of the paper is organized as follows: section II explains the motivation of the work. Section III gives some background and insights about the problem. Section IV describes the proposed method in depth. Section V is devoted to experiments and results. Finally, the conclusions are summarized in section VI.

TABLE I
PROBABILITIES (%) OF SELECTING A λ_i^* LOWER, EQUAL, OR HIGHER THAN $\lambda_{motion} = \sqrt{\lambda_{mode}}$ FOR A SET OF STANDARD SEQUENCES

	$P(\lambda_i^* < \lambda_{motion})$	$P(\lambda_i^* = \lambda_{motion})$	$P(\lambda_i^* > \lambda_{motion})$
Akiyo	1.21	93.87	4.92
Carphone	5.26	70.67	24.07
Claire	1.35	92.48	6.17
Foreman	6.07	60.27	33.66
Highway	9.61	72.04	18.35
Salesman	2.27	88.38	9.35

Fig. 1. Conditional pdf of λ_i^* given $\lambda_{mode} = 5.397$ and $P(\lambda_i^*/\lambda_{mode} = 5.397)$.

II. MOTIVATION

The goal of this work is to raise the coding efficiency of the H.264/AVC JM reference software [4] through an improvement of the ME-related RDO process that does not incur in a significant computational cost. In particular, our proposal relies on the observation that the approximation of J_{mode} by J_{motion} turns out to be suboptimal when the ME process is compromised. To achieve a better solution for these cases, we provide the encoder with the possibility of choosing a MV different from the one that actually minimizes J_{motion}.

A. Accuracy of λ_{motion} estimation

To prove that we are building on a sound hypothesis, we start by investigating the cases in which the estimation of λ_{motion} given in (5) is not accurate. To that end, given a value of λ_{mode}, the encoder has been modified to test several values of λ_{motion}. Each value of λ_{motion} produces a candidate MV resulting from the optimization of J_{motion} and each candidate MV is tested on J_{mode}. In this manner, the decision on the best MV is made using the actual R and D values, instead of estimates. As a result, an optimal MV and, consequently, an optimal value of λ_{motion} are selected. Thus, in those cases in which $\lambda_{motion} = \sqrt{\lambda_{mode}}$ is the best solution, this approximation is proven to be accurate, and vice versa.

Specifically, 21 different values of λ_{motion} were tested. An F_i factor

$$F_i = i \times \Delta F, \text{ with } i \in [0, 1, ..., 20], \Delta F = 0.2, \quad (6)$$

was introduced into J_{motion} to deliberately alter the balance between R_{motion} and D_{motion}:

$$\widehat{J}_{motion} = D_{motion} + (F_i \times \lambda_{motion}) \times R_{motion}. \quad (7)$$

On the one hand, as can be easily inferred, $F_i = 0$ produces a MV that minimizes D_{motion} without any rate considerations. On the other, the higher F_i, the more the decision depends on R_{motion}, in detriment of distortion considerations. Hereafter, we will refer to λ_{motion} as the value estimated as $\sqrt{\lambda_{mode}}$, to λ_i as the product $F_i \times \lambda_{motion}$, and to λ_i^* as the optimal λ_i value (the one selected on J_{mode}).

Following this procedure, we gathered λ_i^* resulting from encoding each MB of several standard video sequences using an IP GOP pattern at 30 frames-per-second (fps) and four QP values (20, 24, 28, 32), with RDO enabled[1] . Since our main interest was to determine in which cases λ_{motion} is not
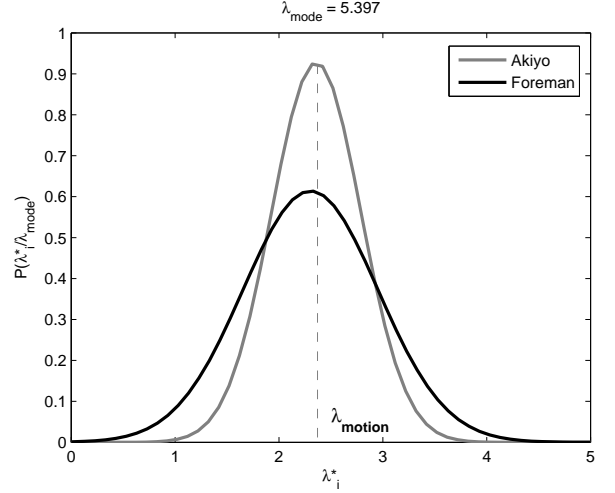
[1]Further details concerning the test conditions will be given later on Section V.

accurate, we grouped the resulting λ_i^* into three classes: lower, equal, or higher than λ_{motion}. The resulting probabilities are shown on Table I. $P(\lambda_i^* = \lambda_{motion})$ represents the probability of selecting λ_{motion} as the best coding option; $P(\lambda_i^* < \lambda_{motion})$ represents the probability of selecting $\lambda_i^* < \lambda_{motion}$, therefore giving more weight to the D_{motion} term; and $P(\lambda_i^* > \lambda_{motion})$ represents the probability of selecting $\lambda_i^* > \lambda_{motion}$, thus putting more emphasis on the R_{motion} term. According to the obtained results, choosing λ_{motion} as the optimal one is undoubtedly the most likely. Nevertheless, there is a significant probability of selecting a λ_i^* different from λ_{motion} (especially, in some sequences). Specifically, video sequences presenting size-changing objects (e.g., zoom, approaching objects), such as *Highway*, produce higher $P(\lambda_i^* < \lambda_{motion})$; sequences exhibiting high translational movements, such as *Foreman* or *Carphone*, show higher $P(\lambda_i^* > \lambda_{motion})$; and finally, in sequences showing low-motion content, such as *Akiyo* or *Claire*, λ_{motion} becomes optimal with high probability.

With the aim of illustrating these ideas with specific examples, Fig. 1 shows the conditional probability density function (pdf) of λ_i^* given λ_{mode}, $P(\lambda_i^*/\lambda_{mode})$, for *Akiyo* and *Foreman*. As long as the relation $\lambda_i^* = \sqrt{\lambda_{mode}}$ is accurate, the mean of $P(\lambda_i^*/\lambda_{mode})$ would tend to $\sqrt{\lambda_{mode}}$ and its variance would tend to zero. As can be observed, the variance is higher in *Foreman* than in *Akiyo*, and $P(\lambda_i^* = \lambda_{motion})$ is significantly lower.

These results show a correlation with motion content. In particular, λ_{motion} is not an accurate estimation for sequences such as *Foreman*, which exhibits random motion due to the hand-holding camera and the large head movements. In contrast, λ_{motion} turns out to be quite an accurate estimation for sequences such as *Akiyo*, which was captured with a static camera and shows small head movements.

In accordance with these results, which suggest that there seem to be certain correlation between the motion content and the accuracy of the $\lambda_{motion} - \lambda_{mode}$ relation, [6] identifies certain content-related events in which λ_{motion} needs to be

TABLE II
PROBABILITIES (%) OF SELECTING A λ_i^* LOWER, EQUAL, OR HIGHER
THAN $\lambda_{motion} = \sqrt{\lambda_{mode}}$ FOR A SELECTED SET OF ME-COMPROMISING
VIDEO SEGMENTS

	$P(\lambda_i^* < \lambda_{motion})$	$P(\lambda_i^* = \lambda_{motion})$	$P(\lambda_i^* > \lambda_{motion})$
Airshow (rotation)	2.53	79.69	17.78
Corvette (fade in)	5.74	36.05	58.21
Ice age (crossfade)	4.36	55.58	40.06
Nature (blurring)	2.66	85.58	11.76
Sintel (rapid mov.)	3.40	73.98	22.62

adjusted to improve $R - D$ performance. In particular, motion content described by high module and random-pointing MVs will better be coded by means of a modification on the $\lambda_{motion} - \lambda_{mode}$ relationship. Some video transitions that produce this behavior are found in [21], [22] and [23], naming the complex translational movements, rotations, fades or blurring as typical events that compromise the accuracy of λ_{motion}.

B. Analysis of ME-compromising events

To prove that the accuracy of the $\lambda_{motion} - \lambda_{mode}$ relation turns out to be significantly lower in those content-related events where the ME actually is not accurate, we have repeated the previous analysis focusing on selected video segments for which we know "a priori" that ME does not correctly work, such as non-translational events (fade transitions, rotation, blurring, etc.) or complex movements, as mentioned in [6].

To this purpose, we gathered λ_i^* resulting from encoding each MB of a set of selected video segments (using the same codec configuration of the previous section). The selected video segments are on-line available at [24]. The results are shown on Table II. When comparing these results to those of Table I, which referred to standard sequences, it becomes obvious that the probability of $\lambda_i^* = \lambda_{motion}$ is significantly lower for these selected ME-compromising segments. Furthermore, for the particular case of fade transitions, $P(\lambda_i^* > \lambda_{motion})$ is even higher than $P(\lambda_i^* = \lambda_{motion})$. This result is due to ME does not correctly work in ilumination changes and the fact that this transition affects the whole frame, so that every MB in the frame is affected by this bad ME.

Therefore, we hypothesize that the estimation of the Lagrangian parameter in J_{motion} can be improved for ME-compromising events. In other words, in these cases, the estimation of the Lagrangian parameter in J_{motion} should be adapted to produce a MV more similar to the one that would be obtained by evaluating J_{mode}.

III. BACKGROUND AND INSIGHTS

A. J_{motion} as a low-complexity alternative to J_{mode}

In this section the differences between J_{motion} and J_{mode} will be discussed in order to gain insight into the causes that may lead to poor performance of J_{motion}.

To find the optimal MV, the ME process should ideally evaluate J_{mode} for all the points in the search range. Given that this process is not computationally feasible, the ME process minimizes J_{motion} instead (Eq. (4)), which could be viewed

as a low-complexity estimation of J_{mode} and can be rewritten (from (4)) as follows:

$$J_{motion} = \sum_{(x,y) \in MB} \left| I(x,y) - \widehat{I}(x,y) \right| + \lambda_{motion} R_{motion}, \tag{8}$$

where x and y are the horizontal and vertical coordinates within the MB; $I(x,y)$ is the luminance of the pixel (x,y) on the original MB; $\widehat{I}(x,y)$ is the luminance of the pixel (x,y) on the predicted MB; and R_{motion} is an estimation of the amount of bits allocated to the MV-related information.

In other words, the goal of the ME process is to obtain, by minimizing J_{motion}, the same MV that would have been obtained by optimizing J_{mode}, which can be rewritten as follows:

$$J_{mode} = \sum_{(x,y) \in MB} \left(I(x,y) - \widetilde{I}(x,y) \right)^2 + \lambda_{mode} (R_{MV} + R_{coeffs+head}), \tag{9}$$

where $\widetilde{I}(x,y)$ is the luminance of the pixel (x,y) of the reconstructed MB; R_{MV} is the amount of bits allocated to the MV-related information; and $R_{coeffs+head}$ represents the bits allocated for transformed coefficients and headers.

The difference between the distortion terms in (8) and (9) comes from the SAD calculation and the use of $\widehat{I}(x,y)$ in J_{motion} instead of the SSD calculation and $\widetilde{I}(x,y)$ in J_{mode}. The difference between the rate terms is also clear: J_{motion} uses an estimation of the bits allocated to the MV-related information, while J_{mode} considers the actual rate including also the bits allocated to headers and DCT coefficients.

Thus, J_{motion} relies on low-complexity estimations of the R and D terms in J_{mode}. When these estimations produce significantly different errors, the balance between R_{motion} and D_{motion} moves from that of J_{mode}, making its minimization to, very likely, fail to produce the same MV than that of J_{mode}. In these cases, we could resort to adapt λ_{motion} to compensate for this unbalance.

B. When J_{motion} does not work: an illustrative example

To illustrate the correlation between the ME-compromising situations and the lack of accuracy of the $\lambda_{motion} - \lambda_{mode}$ relation, in this section we develop an example that deals with a cross-fade transition. Fade transitions are characterized by general illumination changes that severely affect the performance of the block-matching-based ME process implemented on the reference software JM15.1, which is specifically designed for translational movements and is not able to cope with illumination changes. It should be noted that there are specific methods to deal with illumination changes, such as weighted prediction [21], but such solutions are out of the scope of this work. The selected example consists of a cross-fade happening between two consecutive frames (#253 and #254) of *Ice Age*. Fig. 2 shows the two considered frames. On this example, first, we used the reference software implementation and selected the MV by optimizing J_{motion}. We will refer to this approach as Reference Decision (RFD). The MVs obtained following the RFD approach (using the frame #253 as reference) have been superimposed on frame #254. As can be observed, some
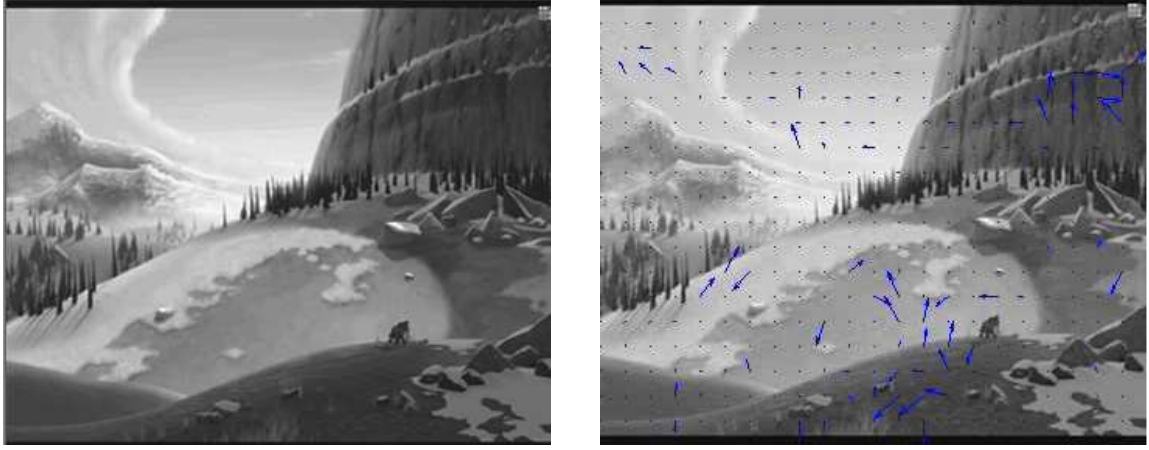
Fig. 2. Frames #253 (left) and #254 (right) of *Ice Age*. MVs are superimposed on the #254 frame.
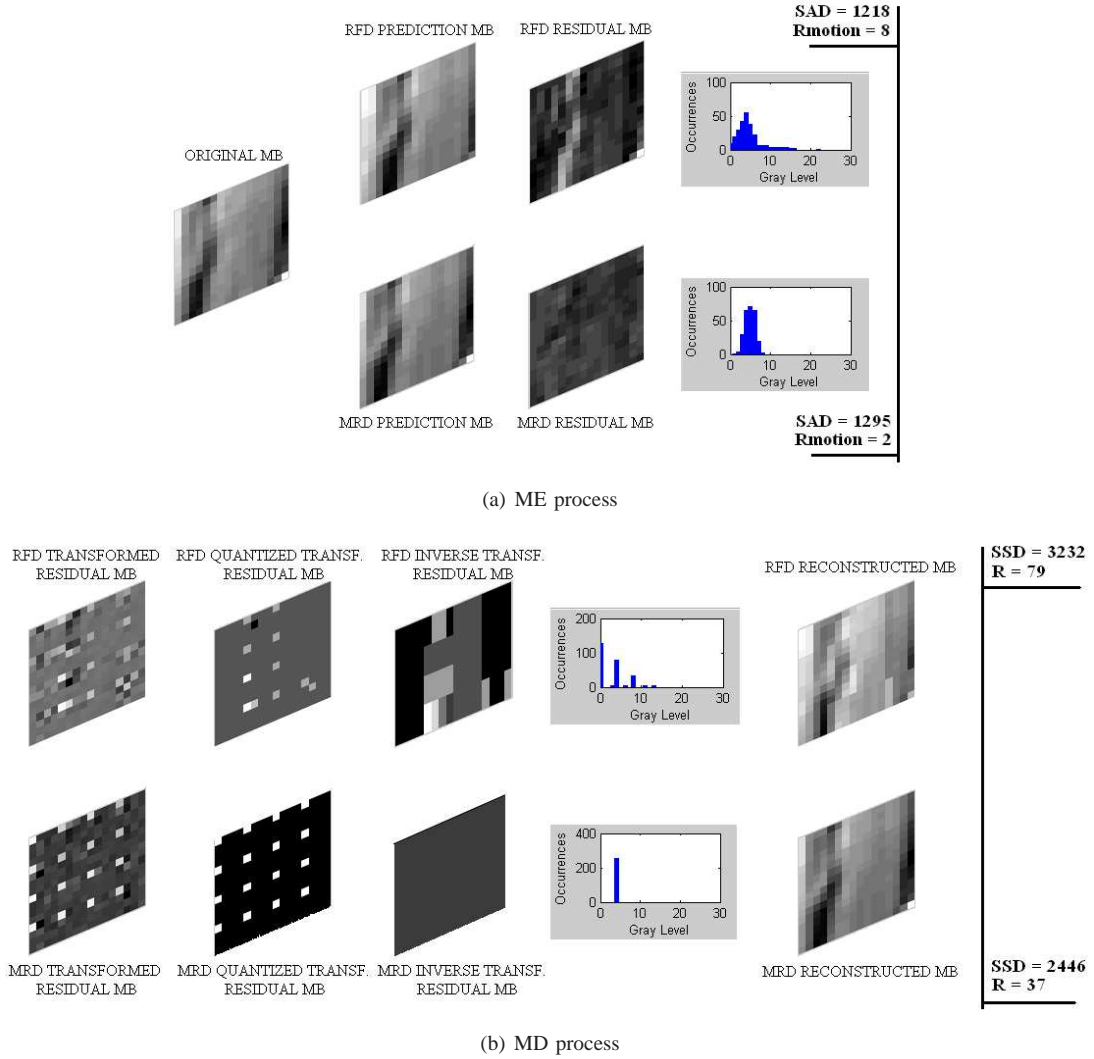

(a) ME process

(b) MD process

Fig. 3. Comparative illustration of RFD (top row) and MRD (bottom row) in the ME (a) and MD (b) processes.

large MVs appeared on regions where there is no actual movement. These MVs appeared due to illumination changes, which misled the ME process to find optimal locations different from the co-located, which, intuitively, seems to be the best option.

To avoid this error, we performed a different ME process. In particular, we used the modified cost function $\widehat{J}_{motion}$ (6), that allowed us to deliberately alter the balance between D_{motion} and R_{motion}. We employed an arbitrarily large F_i value in

order to select the MV that minimized R_{motion}, preventing the ME process from choosing unrealistic large MVs. We will refer to this second approach as Minimum Rate Decision (MRD).

In a fade transition, the co-located MB has very similar edges, but an average luminance different from that of the current MB. Consequently, the SAD may turn out to be high, implying that the ME process could select a MV pointing to an MB that, being different from the co-located, exhibits a more similar mean luminance. Fig. 3 aims to illustrate this idea with a visual example. It shows a parallel analysis of the two processes considered, RFD (top) and MRD (bottom). Fig. 3(a) focuses on the ME process (J_{motion} optimization), showing the predicted MBs, the residues, and their histograms. Fig. 3(b) focuses on the MD process (J_{mode} optimization), depicting the DCT coefficients (before and after the quantization process), the reconstructed residues, along with their histograms, and the reconstructed MBs.

As can be observed in Fig. 3(a), the variance in the luminance of the RFD residue is higher than that of MRD (see either the residual MBs or the corresponding histograms). Nevertheless, the SAD becomes higher for MRD because of the illumination change, being this higher SAD what makes RFD to make a wrong decision at the ME stage. To be more precise, the balance between D_{motion} and R_{motion} turns out to be dominated by D_{motion}: although R_{motion} for RFD is significantly higher than that of MRD, the MV proposed by RFD is chosen because of the SAD term.

Moving forward to Fig. 3(b), when the DCT coefficients are obtained, it becomes clear that the RFD transformed residual presents higher AC coefficients, implying that there is something more than a mean illumination change between the current and the selected reference MB. On the contrary, the MRD transformed residual matches well with an illumination change, as it mainly presents DC coefficients.

In the end, MRD turned out to be better in all senses. Regarding R, the solution of the RFD approach was less efficient since it had to encode some AC coefficients (not providing long runs to the entropy coding phase) and additional MV-related information. Concerning D, MRD modeled better the changes that had actually taken place between both frames and, consequently, the resulting SSD was lower than that of RFD, which failed to establish a proper balance between D_{motion} and R_{motion}.

Fig. 4 provides a graphical explanation from the Lagrange optimization theory point of view, using real $R-D$ data taken from the previous example. The ME process is illustrated in the left part of the figure. The two compared solutions are depicted in the $R_{motion}-D_{motion}$ space labelled as MV_{RFD} (MV associated with RFD) and MV_{MRD} (MV associated with MRD). The optimal solution for a given λ_i is the operating point in the $R_{motion}-D_{motion}$ space that is first "hit" by a plane wave of slope λ_i (dashed lines in Fig. 4) [3]. Therefore, as shown, MV_{RFD} becomes the best solution for $\lambda_i = \lambda_{motion}$, while MV_{MRD} provides the lowest-rate solution (λ_i arbitrarily large).

The MD process is shown in the right part of the figure. The two compared operating points, MV_{RFD} and MV_{MRD},

are depicted in the $R-D$ space, and the optimal solution for a given λ_{mode} is the one that is first "hit" by a plane wave of slope λ_{mode}. In this case, where the terms R and D are not estimations, MV_{MRD} becomes the optimal solution in this particular example.

In the same manner we have forced a R_{motion}-driven solution to correct a D_{motion}-biased solution, it is natural to think of the inverse case: sometimes λ_{motion} would produce a R_{motion}-biased solution that could be corrected by a D_{motion}-driven solution, which could be implemented just by using $F_i = 0$. We would refer to this alternative approach as Minimum Distortion Decision (MDD).

In summary, ME-compromising situations can lead to R_{motion} or D_{motion}-biased solutions, which will require the encoder to be able to select a different λ_{motion} value in order to make a more accurate decision in J_{motion}.

IV. PROPOSED ALGORITHM

This Section describes a computationally efficient method to find a more suitable value of λ_{motion}. First, we discuss how to reduce the number candidate values to make it feasible. And second, we propose a way to reduce the MD evaluations.

A. Reduced set of λ_i values

The modified cost function $\widehat{J}_{motion}$ involving 21 different factors F_i was useful to set the motivation for this work, but becomes computationally impractical for coding purposes. Therefore, it is necessary to propose an alternative that allows us to take advantage of using a more suitable Lagrangian parameter in J_{motion} without incurring in a significant increase of the computational cost.

To this end, we decided to select a reduced set of three λ_i values: one higher than λ_{motion}, which would allow for compensating D_{motion}-biased solutions, one lower than λ_{motion}, which would allow for compensating R_{motion}-biased solutions, and λ_{motion}. In so doing, it seems reasonable to select the extremes, $\lambda_i = 0$ and λ_i arbitrarily large ($\lambda_i \to \infty$), since they would allow for avoiding the potentially largest errors. Interestingly, $\lambda_i = 0$ corresponds to the MDD discussed previously, and $\lambda_i \to \infty$ to the MRD.

Nevertheless, to study the suggested solution more in depth, we gathered λ_i values from encoding each MB of a set of video segments. For these experiments, we used an IP GOP pattern at 30 fps, four QP values (20, 24, 28, 32) and RDO enabled (both video segments and codec configuration are further described in Section V). The obtained results are shown in Table III, where the reference value, λ_{motion}, is labelled as $i = 5$, which corresponds to $F_5 = 0.2 \times 5 = 1$.

Regarding the $\lambda_i < \lambda_{motion}$ values, it seems reasonable to select $\lambda_i = 0$ since it clearly exhibits the highest probability among the λ_i values lower than λ_{motion}.

Now regarding the $\lambda_i > \lambda_{motion}$ values, the probabilities decrease with i, what suggests that taking $\lambda_i \to \infty$ could turn out unsuitable. Nevertheless, in this case, it should be noticed that the probability of selecting the predicted MV (MV_p) as optimal grows with i (because the increasing importance of the rate term). To be more precise, the probability of MV_p
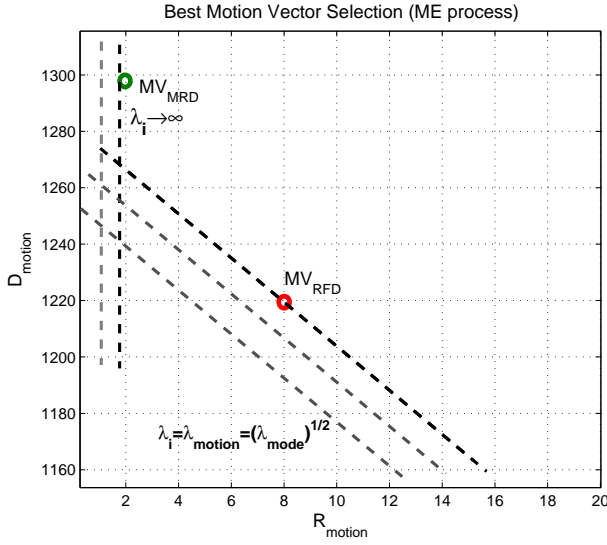
(a) Coding option selection in the ME process.
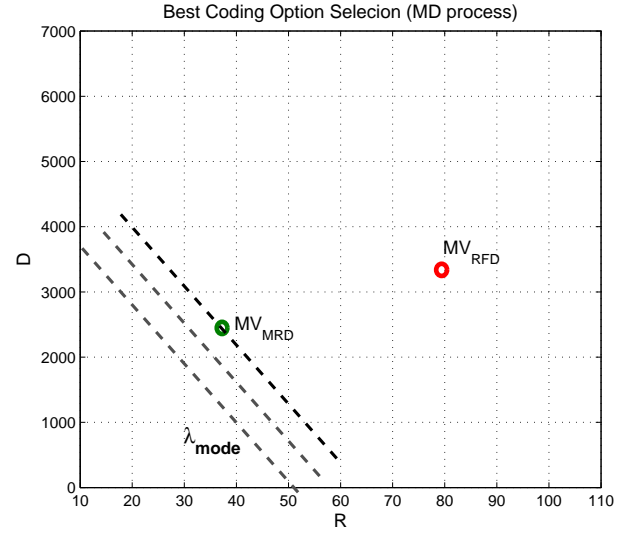
(b) Coding option selection in the MD process.

Fig. 4.   Graphical illustration of the optimal coding option selection.

TABLE III
PROBABILITY (%) OF SELECTING EACH λ_i VALUE

i	$P(\lambda_i^* = \lambda_i)$	i	$P(\lambda_i^* = \lambda_i)$
0	1.94	11	1.01
1	0.51	12	0.82
2	0.60	13	0.68
3	0.45	14	0.64
4	0.37	15	0.50
5	82.58	16	0.47
6	2.27	17	0.39
7	1.96	18	0.39
8	1.67	19	0.33
9	1.27	20	0.33
10	1.16		

to be the optimal for $\lambda_i > \lambda_{motion}$ is 69%. Taking this observation into account, we could compensate the D_{motion}-biased solutions with any λ_i value that leads to MV_p as optimal. So, we choose $\lambda_i \to \infty$.

As a conclusion, $\lambda_i = 0$ and $\lambda_i \to \infty$ (MDD and MRD, respectively) are good candidates to be evaluated in the ME process of each MB.

B. Reduced number of MD evaluations

As mentioned, the proposed algorithm aims to avoid large errors. Specifically, we suggest to consider three alternative MVs, resulting from $\lambda_i = 0$, $\lambda_i = \lambda_{motion}$, and $\lambda_i \to \infty$, where $\lambda_i = 0$ leads to the MV that minimizes the distortion (MDD) and $\lambda_i \to \infty$ to the one that minimizes the rate (MRD).

We must highlight that, during the ME process, the D_{motion} and R_{motion} terms are computed for each position in the search range. Therefore, only one ME pass is required to obtain the three MVs sought. Subsequently, these MVs should be tested on J_{mode} to obtain the optimal coding option.

To reduce the computational cost associated with the two additional J_{mode} evaluations, we propose to assess MDD and

MRD only for the 16x16 pixel PS, which is the most likely one [25]. Furthermore, to achieve higher computational savings, when the reference MV (derived by $\lambda_i = \lambda_{motion}$) turns out to be the same than that obtained by either MDD or MRD, only this reference MV is tested in the MD process (since the third option becomes very unlikely: MDD and MRD actually represent "opposite" solutions). As a result, as empirically shown in the next section, the proposed coding process does not incur in a significant increment of the computational cost with respect to the reference coding process.

C. Summary of the Algorithm

The complete algorithm is summarized in Algorithm 1.

Algorithm 1 Proposed coding process of an MB.

1: % **ME process** (J_{motion})
2: Obtain MV_{RFD}, MV_{MRD}, and MV_{MDD} for 16x16 PS.
3: Obtain MV_{RFD} for the remaining available modes.
4: % **MD process** (J_{mode})
5: **if** $MV_{RFD} \neq MV_{MRD}$ and $MV_{RFD} \neq MV_{MDD}$. **then**
6:     Test MV_{RFD}, MV_{MRD}, and MV_{MDD} for 16x16 on J_{mode}.
7:     Test MV_{RFD} for the remaining available modes on J_{mode}.
8: **else**
9:     Test MV_{RFD} for all the available modes on J_{mode}.
10: **end if**
11: Select optimal mode: min J_{mode}.
12: **return** Best mode.

V. EXPERIMENTS AND RESULTS

The proposed algorithm was implemented on H.264/AVC JM15.1 reference software [4]. The main test conditions were selected according to the recommendations of the JVT [26],

Parameter	Value
Profile IDC	Main
QP	20, 24, 28, 32
GOP	IP @ 30 fps.
ME algorithm	Fast Full Search
Search Range	±32
Reference Frames	3
Symbol Mode	CABAC
RD Optimizarion	ON

namely: main profile, ±32 pixel search range, CABAC, and RDO enabled. Moreover, we used an IP GOP pattern and four QP values (20, 24, 28 and 32). Table IV summarizes the encoder configuration.

To assess the proposed algorithm in terms of R-D performance, we have used the average bit rate differences (ΔBR) and the average PSNR differences ($\Delta PSNR$), as described in [27]. Moreover, to evaluate the computational complexity of the proposed algorithm, the time increment (TI) was calculated as follows:

$$TI = \frac{Time(method) - Time(JM15.1)}{Time(JM15.1)} \times 100(\%). \quad (10)$$

The proposed algorithm has been tested with respect to the H.264/AVC reference software and with respect to an state-of-the-art algorithm called CALM [6], which suggests a context adaptive adjustment of λ_{motion} to improve coding efficiency. The comparative assessment has been performed on a varied set of video segments exhibiting ME-compromising events to show the improved performance of the proposed algorithm in these cases.

Since the proposed algorithm aims to improve the ME process, we have first tested this improvement avoiding potential interference from spatial prediction tools (Intra modes in Inter frames). Then, we have tested the coding performance adding the spatial prediction tools (this will be referred to as overall coding performance). Additionally, we have computed an upper performance bound resulting from assessing the complete set of λ_i values. Subsequently, we have analysed in detail the contribution of each part of the proposed algorithm, namely: the MDD and MRD. Finally, two illustrative examples of the improved subjective quality achieved by the proposed algorithm are also provided.

A. Evaluation of the ME performance

The proposed method aims to improve the performance of ME process by avoiding bad choices of the MV. Therefore, our first experimental evaluation is directed to assess the actual improvement of the ME performance. To this end, we have disabled the use of *Intra modes in Inter Frames* since this coding tool can mask failures of the ME process.

Table V shows the obtained results. For each of the considered sequences, the mean values of TI, ΔBR, and $\Delta PSNR$ across the four QP values are shown. Additionally, the last row of the table shows the mean values for all the sequences. These results reveal that the proposed algorithm clearly improve

the JM15.1 coding performance under the same experimental setup. Specifically, the proposed algorithm obtains an average ΔBR reduction of 9.27% for the same coding quality with respect to the reference software. Alternatively, these improvements can be seen in terms of PSNR, where the proposed algorithm achieves an average gain of 0.52 dB.

It is important to highlight that a higher gain is obtained on video segments where fade transitions take place, as on *Mobisode* or *Corvette*, where ΔBR reductions of 21.18% and 32.60% are obtained, respectively. This is due to the fact that the optimal value of λ_{motion} in these cases is different from the reference one with high probability, as it was shown on Section II-B (see Table II).

Comparing with CALM algorithm, the proposed method produces a better coding quality with quite similar complexity increment. It should be noted, however, that CALM works better in the low-complexity RDO scenario (RDO off).

Regarding the computational complexity, a good compromise have been achieved as TI reaches an average value of 3.07% comparing with the reference software and a 1.74% comparing with CALM, while providing very significant improvements in terms of R-D performance. Moreover, looking at the individual video segments, the highest value of TI incurred by the proposed method is close to 4%, while the worst case for CALM is close to 14%.

Although codec configurations using GOP patterns with B frames have not been assessed on this study, other techniques designed to cope with illumination changes [21] have shown similar coding improvements for GOP patterns with P or B frames. Therefore, although the experiments with B frames are left for future work, similar results are expected.

B. Evaluation of the overall coding performance

To evaluate the overall coding performance, *Intra modes in Inter Frames* coding option was enabled on the JM15.1 reference software. It is expected that the use of the Intra mode coding tool will compensate for some ME failures and, consequently, the performance improvement achieved by the proposed algorithm will be lower than the one obtained in section V-A.

The obtained results are shown on Table VI. An average 2.20% ΔBR reduction was achieved in comparison with the reference software. Alternatively, in PSNR terms, an improvement of 0.12 dB was obtained. On the one hand, the best results continue to appear in fade sequences like *Ice Age* and *Corvette* for the same reasons (now softened by the use of the Intra modes). On the other, the performance improvements becomes less relevant in zoom-type transitions, where the results tend to be similar to those of the reference. In summary, it can be concluded that, despite the use of the Intra mode coding tool overcomes some of the problems associated with ME-compromised events, allowing more freedom for the selection of λ_{motion} still provides significant R-D improvements in exchange of a low increment of computational complexity.

As can be seen, for the experimental protocol used in this paper, CALM does not provide any average improvement with respect to the reference software, likely because it was conceived for RDO-disabled operation.

TABLE V

PERFORMANCE EVALUATION OF THE PROPOSED ALGORITHM RELATIVE TO JM15.1 WITH INTRA CODING IN INTER FRAMES DISABLED. COMPARATIVE RESULTS REGARDING CALM [6] ARE ALSO PROVIDED

| Sequence	Size	Effect	# Coded Frames	Proposed method			CALM [6]		
				TI %	ΔBR %	ΔPSNR dB	TI %	ΔBR %	ΔPSNR dB
Ice Age	CIF	*crossfade*	42	1.87	−9.34	0.64	0.16	0.40	0.02
Ice Age	CIF	*crossfade*	13	2.89	−11.24	0.76	0.19	0.50	0.02
Nature	CIF	*blurring*	100	2.08	−1.67	0.08	0.00	0.01	0.00
Airshow	SD	*rotation*	150	3.33	−6.64	0.40	0.20	0.56	0.00
Corvette	SD	*fade in*	8	2.96	−32.60	1.95	0.63	−0.75	0.06
Corvette	SD	*zoom in*	50	4.03	−0.40	0.01	1.68	−0.01	0.00
Corvette	SD	*zoom out*	5	2.73	−0.58	0.03	14.06	−0.10	0.01
Mobisode	SD	*crossfade*	20	2.34	−21.18	0.82	−1.15	0.84	0.01
Controlled Burn	HD	*crossfade*	10	3.16	−15.86	0.82	−0.38	−1.77	0.10
Dinner	HD	*blurring*	62	3.98	−4.22	0.22	0.11	−0.73	0.02
Dinner	HD	*zoom out*	100	4.10	−1.05	0.05	−0.19	−0.24	0.01
Sintel	HD	*rapid mov.*	73	3.37	−6.43	0.41	0.64	−0.21	0.01
Average				**3.07**	**−9.27**	**0.52**	**1.33**	**−0.13**	**0.02**
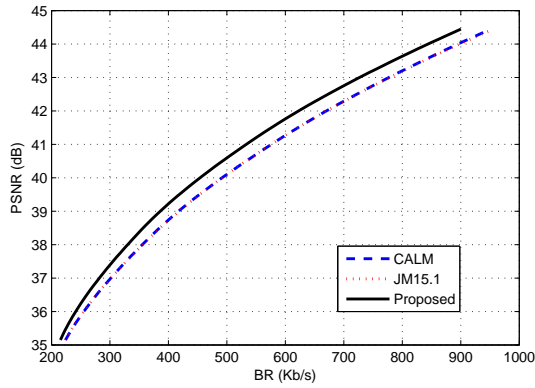
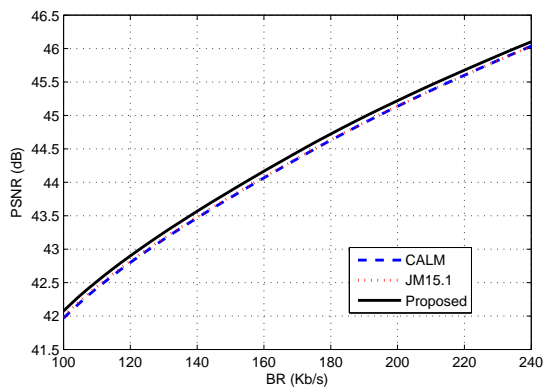Fig. 5. PSNR-BR curves achieved by the compared algorithms for *Ice Age*.

Fig. 6. PSNR-BR curves achieved by the compared algorithms for *Nature*.

Figs. 5 and 6 show the PSNR-BR curves for two selected sequences: *Ice Age* and *Nature*, respectively. As it can be observed, the proposed algorithm clearly outperforms both JM15.1 and CALM.

Finally, note that the computational cost is almost the same in the proposed algorithm than in the reference software. Specifically, using the proposed algorithm implies an increment of 2.99% of TI with respect to the reference.

C. An upper performance bound

An extended version of the algorithm that assess 40 different λ_{motion} values has also been tested with the aim of providing a upper performance bound. The procedure described in Section II-A was used for $i \in [0, 1, \cdots, 40]$ in (6). Table VII shows comparative results between the proposed method and this upper performance bound.

As can be observed, although the upper performance bound clearly exceeds the results of the proposed method, the room for improvement is quite moderate. The upper performance bound is not higher (as could be expected) because it has been obtained by taking locally optimal decisions (at MB level), which could not be globally optimal.

These results allow us to conclude that, although there is some room for improvement, the proposed solution provides an excellent balance between performance and computational cost: it achieves a 2.20% bit rate reduction vs. a 3.21% of the upper bound without incurring in a significant increment of the computational cost.

D. Evaluation of the MRD and MDD contributions

An analysis on the individual contributions of both MRD and MDD has been done to check their relative influence on the global performance. Table VIII shows the overall coding performance of both MRD and MDD with respect to the reference software. In the first case only MV_{MRD} is considered together with MV_{RFD}. In the second, it is MV_{MDD} the only additional MV considered. As can be seen, both strategies provides quite similar ΔBR reductions. Interestingly, it is worth mentioning that the TI generated by MRD is low in comparison to that of MDD. This is due to the fact that the probability of MV_{RFD} being the same than MV_{MRD} is higher than for MV_{MDD}.

It is also interesting to notice that, in some particular cases, working just with MRD or MDD outperforms the complete algorithm. The reason is that the decisions made are locally optimal (for the current MB), but given that they affect the encoding of neighboring MBs, sometimes they could be globally sub-optimal.

TABLE VI

PERFORMANCE EVALUATION OF THE PROPOSED ALGORITHM RELATIVE TO JM15.1 WITH INTRA CODING IN INTER FRAMES ENABLED. COMPARATIVE RESULTS REGARDING CALM [6] ARE ALSO PROVIDED

Sequence	Size	Effect	# Coded Frames	Proposed method			CALM [6]		
				TI %	ΔBR %	ΔPSNR dB	TI %	ΔBR %	ΔPSNR dB
Ice Age	CIF	*crossfade*	42	−0.55	−7.57	0.50	−0.65	−0.47	0.02
Ice Age	CIF	*crossfade*	13	1.21	−4.98	0.32	−1.25	−0.24	0.02
Nature	CIF	*blurring*	100	1.84	−1.81	0.09	0.51	0.26	−0.01
Airshow	SD	*rotation*	150	3.22	−0.85	0.04	0.47	−0.03	0.01
Corvette	SD	*fade in*	8	5.17	−6.21	0.28	1.24	−0.15	0.01
Corvette	SD	*zoom in*	50	3.79	−0.14	0.00	−0.30	0.00	−0.01
Corvette	SD	*zoom out*	5	5.12	−0.56	0.03	2.21	−0.04	0.00
Mobisode	SD	*crossfade*	20	3.53	−2.70	0.06	−0.54	0.76	−0.03
Controlled Burn	HD	*crossfade*	10	2.29	−1.18	0.04	−0.88	0.03	0.00
Dinner	HD	*blurring*	62	3.33	0.19	0.00	0.07	0.04	0.00
Dinner	HD	*zoom out*	100	3.55	−0.30	0.01	0.75	0.00	0.00
Sintel	HD	*rapid mov.*	73	3.32	−0.31	0.02	1.23	−0.13	0.00
Average				**2.99**	**−2.20**	**0.12**	**0.24**	**0.00**	**0.00**

TABLE VII

PERFORMANCE EVALUATION OF THE PROPOSED ALGORITHM WITH RESPECT TO AN EMPIRICAL UPPER BOUND. RESULTS IN BOTH CASES ARE RELATIVE TO JM15.1 WITH INTRA CODING IN INTER FRAMES ENABLED.

Sequence	Size	Effect	# Coded Frames	Proposed method			Upper Bound		
				TI %	ΔBR %	ΔPSNR dB	TI %	ΔBR %	ΔPSNR dB
Ice Age	CIF	*crossfade*	42	−0.55	−7.57	0.50	1924	−6.79	0.48
Ice Age	CIF	*crossfade*	13	1.21	−4.98	0.32	2022	−4.42	0.29
Nature	CIF	*blurring*	100	1.84	−1.81	0.09	1748	−2.58	0.12
Airshow	SD	*rotation*	150	3.22	−0.85	0.04	1927	−2.34	0.11
Corvette	SD	*fade in*	8	5.17	−6.21	0.28	2130	−5.70	0.24
Corvette	SD	*zoom in*	50	3.79	−0.14	0.00	1977	−2.11	0.09
Corvette	SD	*zoom out*	5	5.12	−0.56	0.03	1908	−2.26	0.13
Mobisode	SD	*crossfade*	20	3.53	−2.70	0.06	2106	−4.61	0.14
Controlled Burn	HD	*crossfade*	10	2.29	−1.18	0.04	2089	−1.81	0.06
Dinner	HD	*blurring*	62	3.33	0.19	0.00	1984	−1.41	0.06
Dinner	HD	*zoom out*	100	3.55	−0.30	0.01	2064	−3.49	0.15
Sintel	HD	*rapid mov.*	73	3.32	−0.31	0.02	2036	−0.97	0.05
Average				**2.99**	**−2.20**	**0.12**	**1993**	**−3.21**	**0.16**

TABLE VIII

INDEPENDENT PERFORMANCE EVALUATION OF MRD AND MDD.

Sequence	Size	Effect	# Coded Frames	MRD			MDD		
				TI %	ΔBR %	ΔPSNR dB	TI %	ΔBR %	ΔPSNR dB
Ice Age	CIF	*crossfade*	42	−0.46	−6.97	0.48	3.00	−7.14	0.49
Ice Age	CIF	*crossfade*	13	−0.04	−4.00	0.26	2.88	−4.12	0.27
Nature	CIF	*blurring*	100	0.81	−1.86	0.09	4.28	−2.16	0.09
Airshow	SD	*rotation*	150	2.07	−0.23	0.01	3.92	−0.28	0.01
Corvette	SD	*fade in*	8	3.04	−6.61	0.29	4.15	−6.49	0.29
Corvette	SD	*zoom in*	50	3.36	0.52	−0.03	3.85	0.46	−0.02
Corvette	SD	*zoom out*	5	1.75	0.15	0.00	3.50	0.17	0.00
Mobisode	SD	*crossfade*	20	−1.42	−2.89	0.07	1.18	−2.96	0.08
Controlled Burn	HD	*crossfade*	10	−0.97	−1.15	0.04	0.39	−1.28	0.04
Dinner	HD	*blurring*	62	0.66	0.17	0.00	0.32	−0.02	0.01
Dinner	HD	*zoom out*	100	2.00	0.50	−0.03	1.44	0.48	−0.02
Sintel	HD	*rapid mov.*	73	3.01	−0.24	0.01	6.33	−0.16	0.01
Average				**1.15**	**−1.88**	**0.10**	**2.74**	**−1.96**	**0.10**

E. Subjective quality evaluation

Although the objective R-D results shown in previous subsections support our claim, a brief comment about subjective quality is in order. To this purpose, we show in Figs. 7 and 8 two examples of reconstructed frames, for two different video segments, obtained with the reference software JM15.1 and the proposed method.

In the first example, one selected frame of the first of the *Ice Age* video segments (specifically, frame # 20) has been encoded, and we have comparatively shown the corresponding reconstructed versions of that frame. To make this comparison as fair as possible, we have adjusted the QP value so that the number of bits produced by this frame would be almost the same in both cases; in particular, it takes up to 8.3 Kb when encoded by the reference software and 8.2 Kb by the proposed method.

Fig. 7 shows three versions of a selected area of the mentioned frame in the *Ice Age* video segment: (a) original; (b) reconstructed by the reference software; and (c) reconstructed by the proposed method. As can be inferred when comparing Figs. 7(b) and 7(c), a higher subjective quality is achieved by the proposed method in comparison with the reference software. Specifically, when looking carefully at the region showing the snowy peak of the mountain, a lot of details are lost in the frame reconstructed by the reference software, while several of them are preserved in the version reconstructed by the proposed method. Another example can be found in the low part of the figures, where two characters (at small size) can be observed: in the reconstructed frame by the reference software one of this characters is missing, while it still appears in the frame reconstructed by the proposed method.

In the second example, one selected frame of the *Mobisode* video segment (specifically, frame # 17) was used. Again, the QP value was adjusted to obtain almost the same number of bits with the reference software and with our proposal; specifically, 52.5 and 51.1 Kb respectively. In Fig. 8 three versions of a selected area are shown (original (a), reconstructed by the reference software (b), and reconstructed by the proposed method (c)). As it can be observed in Figs. 8(b) and 8(c), in the region showing the bars of the stairs our method achieves better defined edges than the reference software. Moreover, this improvement can be also observed in the shaded peak of the suit in the right part of the figures.

This higher subjective quality can be explained by the fact that the proposed method improves the ME process, coding some MBs with higher efficiency. In such a way, for the same amount of bits than in the reference coding process, we are able to code with higher quality, preserving more details in the coded sequence.

VI. CONCLUSIONS AND FURTHER WORK

In this paper we have proposed an algorithm to improve the conventional way of estimating λ_{motion} and, consequently, the ME process in RDO-based video codecs. Specifically, an algorithm has been proposed that allows the encoder to choose between three different values of λ_{motion}. Actually, this choice has been limited to the Inter 16x16 partition size to avoid incurring in a significant increase of the computational cost. For this partition size, the proposed algorithm allows the encoder to additionally test $\lambda_{motion} = 0$ and $\lambda_{motion} \rightarrow \infty$, which corresponds to minimum distortion and minimum rate solutions, respectively. By testing these two extreme values, the algorithm avoids to make large ME errors in ME-compromising events, which refer to a wide set of content-related events that make the ME process to perform poorly; for example: complex or non translational movement, edited transitions such as fades, blurring, etc.

The proposed algorithm has been extensively tested with respect to the H.264/AVC reference software and a state-of-the-art algorithm called CALM [6], which suggests a context adaptive adjustment of λ_{motion} to improve coding efficiency. Furthermore, the comparative assessment has been
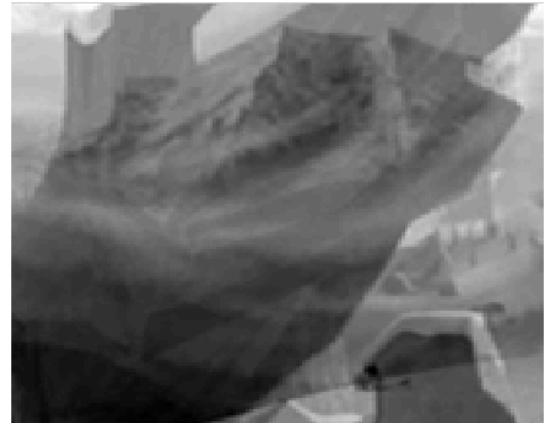
(a)

(b)

(c)

Fig. 7. Illustrative example of the achieved subjective quality. (a) selected part of the original frame selected from the *Ice Age* sequence; (b) reconstructed frame with the reference software; and (c) reconstructed frame with the proposed method.

performed on a varied set of video segments exhibiting ME-compromising events to show the performance of the proposed

(a)

(b)

(c)

Fig. 8. Illustrative example of the achieved subjective quality. (a) selected part of the original frame selected from the *Mobisode* sequence; (b) reconstructed frame with the reference software; and (c) reconstructed frame with the proposed method.

algorithm in these cases.

The experimental results allowed us to conclude that the

proposed algorithm substantially improves the performance of the ME process (when intra modes in inter MB are disabled), achieving average bit rate reductions of 9.27% with respect to the reference software, while the CALM algorithm achieved a bit rate reduction of 0.13%. When considering the overall coding efficiency, the performance improvement was lower because the Intra modes used in Inter frames actually compensated ME errors; nevertheless, the performance improvement was still significant: an average bit rate reduction of 2.20% with respect to the reference software; while CALM does not achieve any improvement.

Furthermore, we have experimentally tested the effectiveness of each of the two additional λ_{motion} values, concluding that both are equally important.

Finally, two illustrative examples of the improved subjective quality achieved by the proposed algorithm have been also provided.

Regarding future lines of research, we suggest the evaluation of the proposed algorithm with GOP patterns using B frames. Moreover, considering that there is still a moderate room for improvement, we propose the design of a data-driven regressor that would allow us to estimate more precisely the optimal λ_{motion} parameter, so that the algorithm would be able to use a near-optimal λ_{motion} value instead of just testing the extreme cases.

REFERENCES

[1] JVT, "Advanced Video Coding (AVC), ITU-T Rec.H264 and ISO/IEC 14496-10 (MPEG-4 Part 10)," 3 ed.

[2] H. Everett III, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operations research*, vol. 11, no. 3, pp. 399–417, 1963.

[3] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *Signal Processing Magazine, IEEE*, vol. 15, no. 6, pp. 23 –50, Nov. 1998.

[4] "JVT H.264/AVC reference software v15.1." [Online]. Available: http://iphome.hhi.de/suehring/tml/download/old_tml/

[5] G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *Signal Processing Magazine, IEEE*, vol. 15, no. 6, pp. 74 –90, Nov. 1998.

[6] J. Zhang, X. Yi, N. Ling, and W. Shang, "Context Adaptive Lagrange Multiplier (CALM) for Rate-Distortion Optimal Motion Estimation in Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 6, pp. 820 –828, june 2010.

[7] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers [speech coding]," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 9, pp. 1445 –1453, Sep. 1988.

[8] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *Image Processing, IEEE Transactions on*, vol. 2, no. 2, pp. 160 –175, apr 1993.

[9] S.-W. Wu and A. Gersho, "Rate-constrained optimal block-adaptive coding for digital tape recording of HDTV," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 1, no. 1, pp. 100 –112, march 1991.

[10] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 8, no. 4, pp. 446 –459, aug 1998.

[11] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application. I. Fundamental theory," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 7, no. 2, pp. 287 –298, apr 1997.

[12] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 6, no. 1, pp. 12 –20, feb 1996.

[13] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 7, no. 1, pp. 246 –250, feb 1997.

[14] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *Image Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 423 –438, april 2005.

[15] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace Distribution Based Lagrangian Rate Distortion Optimization for Hybrid Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 19, no. 2, pp. 193 –205, feb 2009.

[16] L. Chen and I. Garbacea, "Adaptive Lambda Estimation in Lagrangian Rate-Distortion Optimization for Video Coding," in *Visual Communications and Image Processing 2006*, vol. 6077, no. 1.   SPIE, 2006, p. 60772B.

[17] F. Yang, S. Wan, and E. Izquierdo, "Lagrange multiplier selection for 3-d wavelet based scalable video coding," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 2, 2007, pp. II – 309–II – 312.

[18] S. Ma, J. Si, and S. Wang, "A study on the rate distortion modeling for high efficiency video coding," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 181–184.

[19] B. Lee and M. Kim, "Modeling rates and distortions based on a mixture of laplacian distributions for inter-predicted residues in quadtree coding of hevc," *Signal Processing Letters, IEEE*, vol. 18, no. 10, pp. 571–574, 2011.

[20] P. Sangi, J. Heikkila, and O. Silven, "Selection of the Lagrange multiplier for block-based motion estimation criteria," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 3, may 2004, pp. iii – 325–8 vol.3.

[21] J. Boyce, "Weighted Prediction in the H.264/MPEG AVC Video Coding Standard," in *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, vol. 3, may 2004, pp. III – 789–92 Vol.3.

[22] M. Budagavi, "Video Compression using Blur Compensation," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2, sept. 2005, pp. II – 882–5.

[23] K. Kamikura, H. Watanabe, H. Jozawa, H. Kotera, and S. Ichinose, "Global Brightness-Variation Compensation for Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 8, no. 8, pp. 988 –1000, dec 1998.

[24] [Online]. Available: http://www.tsc.uc3m.es/~jlmolinero/Sequences.zip

[25] E. Martinez-Enriquez, A. Jimenez-Moreno, and F. Diaz-de Maria, "An adaptive algorithm for fast inter mode decision in the H. 264/AVC video coding standard," *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 2, pp. 826–834, 2010.

[26] G.Sullivan, "Recommended Simulation Common Conditions for H.26L coding efficiency Experiments on Low-Resolution Progressive-scan source material," ITU-T, September 2001.

[27] G.Bjontegaard, "Calculation of Average PSNR Differences Between RD-Curves," ITU-T, VCEG-M33, April 2001.