Abstract—Performance of real-time video processing applications such as surveillance systems, content-based search, is limited by the complexity of video content analysis in the pixel domain. A low complex alternative is to analyse the video in the compressed domain, where content features already available in the compressed video are directly used in the analysis. However, this is achieved at the expense of output precision and reliability, due to compression-efficiency driven feature selection at the encoder. Therefore, video applications could benefit from enhanced reliability of data embedded in the compressed video.

In this paper, we present a scalable optimization model that addresses the accuracy of content features in parallel with the conventional rate-distortion optimization criterion. We analyse and optimize rate-distortion performance of video encoder under content description accuracy constrain, using a motion calibrated synthetic data set containing a range of scene and motion complexity levels. Finally, using a natural video data set, we demonstrate that the proposed optimization framework can be used to enhance compressed feature accuracy without incurring a rate-distortion overhead.

Index Terms—Feature Extraction, Video Analysis, Video Coding, Vision Applications

I. INTRODUCTION

Conventionally, video encoder is optimized for the use of video in communication and entertainment applications, where the distortion resulting from lossy compression is minimized given the affordable compressed data rate [1]. However, video consumption has evolved dramatically over the past two decades to wide spread use in a range of applications such as surveillance, control systems, medical imaging, and navigation, which use video analysis to extract information [2]. Similarly, in multimedia applications there is an increasing demand for content-based functionalities [3] such as event detection, content driven search, and classification, which allow the user to access multimedia with greater flexibility. In real-time resource-limited scenarios, these applications can exploit information embedded in the compressed video to fulfill the demand for efficient video analysis, instead of decoding the video to access pixel-domain information [4]-[5]. However, information carried in the compressed domain, such as motion vectors, transform coefficients - referred to as compressed features - are often not reliable enough for content analysis. This is a result of conventional encoder parameter selection, limited to optimizing compression efficiency, which does not necessarily result in accurate content description in the compressed domain. Compression efficiency is critical for communication applications. On the other hand, other aspects of video utilization, such as video analysis and its applications, would benefit from enhanced quality of additional information present in the compressed video stream.

In order to optimize compressed domain feature accuracy in content description, along with rate distortion (RD) constrains, this paper presents a study on Rate-Distortion-Content Description (RDCD) behaviour of video encoding system. The conventional encoder decision criterion is extended with a novel objective, which gives a scalable framework to manage affordable content description error and RD performance.

Application-aware video coding methods have been proposed in literature in [6], [7] and [8]. Enforced H.264 block-size selection is used by kapatos et al. [6], to demarcate scene changes. This requires a modified decoder and change-aware applications. kas et al. [7] proposes background MV correction in the encoder to improve motion based video indexing. Methods present in [6] and [7] each support only a given application. In [8], we proposed a content driven motion selection mechanism, independent of compression constraints, to enhance performance of applications based on object motion. However, compression requirements should be considered in parallel with application requirements to provide a flexible solution. Rate distortion performance under constrain functions to control aspects of video encoding system such as complexity and power consumption has been proposed in literature [9] - [10]. However, to the best of our knowledge,
there has been no framework to optimize video encoder to facilitate video content-based video applications.

In this work, we analyse the RD behaviour under content description accuracy constrain and propose a scalable Rate-Distortion-Content Description (RDCD) framework, as illustrated in Fig. 1. We developed a motion calibrated synthetic data set covering different scene complexities and activity levels, to analyse the framework under known content motion. Using a natural video data set, we demonstrate the scalable implementation of the proposed framework.

The rest of the paper is organized as follows. Section II discusses the problem background and proposed RDCD framework is presented in Section III. Experimental results and the conclusion are given in Sections IV and V respectively.

II. SCOPE AND PROBLEM BACKGROUND

A typical hybrid video encoder consists of several coding tools; intra frame and motion compensated (inter) predictive coding, transform coding for prediction residual and entropy coding. Out of these, compressed domain content analysis uses the content features extracted for predictive coding, namely motion vectors and prediction residual data.

Existing coding standards, such as H.263 [11] and H.264/AVC [12], use a block based 2D translation model to represent motion across frames. Frame is divided into rectangular blocks, and each block is matched with candidate positions in a search region to select the best prediction resulting in a 2D motion vector (MV). Since a sequence of frames would contain a widely varying texture and motion levels, the selection of coding parameters such as MVs and block size divisions -referred to as coding mode- to accommodate these variations, form the encoder optimization problem. Typical encoder control process assumes motion estimation and block mode decision to be independent of each other for simplicity [13], and handle these in successive iterations. We limit the scope of this work to encoder optimization for motion estimation, and the results are presented for fixed block size.

A. Rate-constrained Motion Estimation

In a hybrid encoder, transform coding prediction residuals results in a lossy compression. Therefore, the objective of a typical encoder parameter selection is to achieve the optimum trade-off in resultant distortion and bit rate. Block matching techniques for motion estimation are typically based on minimizing matching error using a difference measurement such as Sum of Absolute Difference (SAD) or Mean Squared Error (MSE). In encoder optimization, in order to achieve optimum rate-distortion (RD) trade-off, this is typically extended as a rate constrained cost function. For example, in the reference software for H.263 [13] and H.264/AVC [14] Lagrangian optimization is used in encoder parameter selection, which forms an unconstrained optimization as given in (1). If \( m \) denotes a candidate vector within the search region, \( \lambda_M \) is the Lagrange parameter. The optimum value for \( \lambda_M \) is heuristically evaluated as a function of quantization parameter (QP). Quantization of residual transform coefficients results in a lossy compression, leading to increased distortion in the decoded video as the QP value increases. Therefore, for efficient RD performance, with increasing QP resources are allocated to encode motion information and residual data are manged by adjusting Lagrange parameter accordingly as given in (4) [13].

\[
\lambda_M = \sqrt{0.85 \times 2(QP - 12)^3/3}
\]  

While rate constrained motion selection achieves the stated objective of efficient compression, this does not necessarily result in reliable motion information. An example is given in Fig. 2, where the variation of RD cost (\( D + \lambda \frac{R}{M} \)) across the search region, for a block with actual motion (-2,-5). The motion search window has an offset of (-1,-1), resulting in an effective search region of [(-17,-17),(15,15)]. Minimum RD cost occurs at (-10,-5).

![Fig. 2. An example of RD Cost variation across the search region](image)

III. RATE-DISTORTION-CONTENT DESCRIPTION (RDCD) OPTIMIZATION MODEL

In order to enhance the accuracy of compressed-domain motion information, the motion selection process at the encoder needs to achieve a trade-off between MV reliability and RD performance requirements. Therefore, we extend the encoder motion selection criterion given in (1), with a novel objective function, Content Description Error, \( E_{MV}(m) \). The extended unconstrained optimization function is given by,

\[
m(\lambda) = \text{argmin}_m \{D_{SAD}(m) + \lambda_M R_M(m) + \lambda_E E_{MV}(m)\}
\]  

where, the control parameter, \( \lambda_E \), ensures the scalability of the proposed framework. In order to retain the rate constrain provided by the standard optimization function given in (1), we model the objective function for motion accuracy independent of the existing objective functions in the block matching...
process. Additionally, this would circumvent the inherent limitations of using translational motion model to represent motion. If \( m = (\hat{V}_x, \hat{V}_y) \) denotes a candidate vector within the search region, \( E_{MV}(m) \), is defined as a likelihood of candidate vector \( m \) given by,

\[
E_{MV}(m) = \| (\hat{V}_x, \hat{V}_y) - (V_x, V_y) \|
\]

(4)

where \( (V_x, V_y) \) is the reference MV that represent the expected motion. It is assumed that side information on frame motion is available to the encoder, calculated using pixel level data prior to encoding each frame. In [8], we have demonstrated that optical flow processed for block data structure can be used as an estimate of motion. We use a similar method in Section V, to evaluate the performance of the framework for natural video sequences.

In order to analyse RD behaviour of video encoding system under the motion accuracy constrain, we evaluate the RDCD framework for a set of sequences with different combinations of control parameters \( \lambda_E \) and \( \lambda_R \). The optimum Lagrange parameter, \( \lambda_M \), for rate constrained motion estimation given by (4) no longer holds for the extended optimization function. However, optimum \( \lambda_R \) will relate to QP in a similar monotonically increasing function, due to trade-off between rate and distortion. Hence the experimental values for \( \lambda_R \) for each QP was selected in the vicinity of \( \lambda_M \) as illustrated in Fig. 3. Since \( \lambda_E \) is an unknown quantity, we test the proposed encoder over an exhaustive range, with maximum value, \( \lambda_E^{max} \) selected as 500. \( \lambda_E^{max} \) reflect the scale of typical RD cost values observed in the conventional encoder.

A. Experimental data set

Measurement errors in reference data for \( E_{MV}(m) \), affect the performance of the proposed framework. This is particularly significant in the parameter-learning phase, where the reference information used to drive the objective function should be an exact representation of actual motion to ensure reliability of the process. Therefore, the framework should be evaluated using test data with known motion. In order to achieve this, we developed a synthetic data set with pre-calibrated motion, containing a range of scene and motion complexity levels as outline in Table I. Selected frames from the test data set are given in Fig. 4. All sequence are of CIF (352 X 288) resolution and 4:2:0 sub-sampling format. In the rest of the paper, we refer to these synthetic sequences as ‘Syn’. 

B. Modeling Sensitivity of Motion Description Error (\( E_{MV} \)) to Control Parameters \( \lambda_E \) and \( \lambda_R \)

Accuracy of MVs selected by RDCD framework is measured using \( E_{MV}(m(\lambda)) \) averaged over all blocks in a sequence of frames, in pixel units. \( m(\lambda) \) is the MV selected in (3). We denote this measurement, Average Motion Description Error by \( E_{MV} \). Our objective is to achieve minimum \( E_{MV} \), without compromising on RD performance.

Behaviour of \( E_{MV} \) over varying control parameters, for different QP values are illustrated in Fig. 5. It is observed that \( E_{MV} \) generally increases with QP and frame texture and motion complexity. We limit the scope of this work to motion
accuracy analysis for video content variation at constant QP. Therefore, we consider QP 30 in the rest of the analysis. Using the experimental data, motion vector accuracy, $E_{MV}$ is modelled as a function of $\lambda_R$ and $\lambda_E$ at constant QP, according to (5).

$$E_{MV} = \alpha \lambda_E^\beta \lambda_R^\gamma + \epsilon$$

Therefore at a given motion accuracy requirement, $\hat{E}_{MV}$, the control parameter $\hat{\lambda}_E$ is defined by,

$$\hat{\lambda}_E = \sqrt{\frac{\hat{E}_{MV} - \epsilon}{\alpha \hat{\lambda}_R^\gamma}}$$

Fig. 6 and 7 illustrate the estimated $E_{MV}$ compared with experimental results, for test sequences 'SynD' and 'SynF' respectively. The estimations for $\alpha$, $\beta$, $\gamma$ and $\epsilon$ coefficients for the synthetic data set is given in Table II. Coefficients of each sequence vary with scene texture and motion levels.

<table>
<thead>
<tr>
<th>Test Sequence</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SynA</td>
<td>1.56</td>
<td>-0.96</td>
<td>0.90</td>
<td>0.19</td>
</tr>
<tr>
<td>SynB</td>
<td>0.27</td>
<td>-0.61</td>
<td>0.87</td>
<td>0.36</td>
</tr>
<tr>
<td>SynC</td>
<td>0.73</td>
<td>-0.51</td>
<td>0.53</td>
<td>0.56</td>
</tr>
<tr>
<td>SynD</td>
<td>0.80</td>
<td>-0.77</td>
<td>0.65</td>
<td>0.18</td>
</tr>
<tr>
<td>SynE</td>
<td>2.75</td>
<td>-0.80</td>
<td>0.61</td>
<td>0.46</td>
</tr>
<tr>
<td>SynF</td>
<td>1.07</td>
<td>-0.59</td>
<td>0.36</td>
<td>0.13</td>
</tr>
</tbody>
</table>
C. Motion-Accuracy Constrained R-D behaviour

In order to understand the RD behaviour of the system under content description accuracy constrain, we evaluated the framework with a range of control parameters under different quantization levels. Fig. 8 illustrate an example of the solutions obtained in the RDCD space, and the corresponding Pareto optimal solutions, identified within the solutions, is given in Fig. 9. Other sequences also demonstrate similar behaviour.

Using the solutions selected to be Pareto optimal, i.e. non-dominating solutions, the feasible space for control parameters $\lambda_E$ and $\lambda_R$ was identified for each sequence, and a generalized representation is given in Fig. 10.

IV. EXPERIMENTAL RESULTS

We validate the proposed RDCD framework for both natural and synthetic test sequences based on two criteria. Firstly, Rate-distortion performance and motion description error are compared with that of a typical RD optimized encoder. Secondly, scalability of the framework is demonstrated by evaluating the performance for target $E_{MV}$ values for given $\lambda_R$. Video sequences in the test set are of CIF (352 X 288) resolution and 4:2:0 sub sampling format. Since we analyse the framework under constant quantization, the experiments were carried out at QP 30.

A. Rate-Distortion-Motion Accuracy performance

The performance of the RDCD framework is compared with that of the JM reference software for state-of-the-art video coding standard H.264/AVC [14]. $\lambda_M$ for the reference system was selected according to (4). A quantitative comparison of performance is given in Tables III and IV for selected control parameter values for synthetic and natural sequences, respectively. Results illustrate that the proposed framework can achieve improved motion accuracy with minimal or no rate-distortion overheads. Highlighted results indicate solutions, where the proposed framework significantly outperforms the reference system.

B. Scalability of RDCD framework

The motion accuracy scalability is critical for practical implementation of the proposed framework. Table V outline the accuracy control parameter, $\lambda_E$, estimated for a given $\lambda_R$ and a target $E_{MV}$, along with the achieved $E_{MV}$. Results illustrated that the RDCD framework can be tailored to reach a given performance threshold according to the requirements of the application.
V. CONCLUSIONS

In order to facilitate efficient video analysis in resource-limited systems, this paper proposes a scalable encoder optimization framework that enhances content description in the compressed domain, in parallel with rate-distortion optimization criterion. Rate distortion behaviour of the video encoding system was analysed and optimized under content description accuracy constrain, using a motion calibrated synthetic data set containing a range of texture and motion complexities. Sensitivity of the extended optimization function to control parameters is studied using experimental data, to identify the optimal solution space for control parameters. The practical implementation of the proposed framework is presented using a set of natural videos. The results demonstrate that the proposed Rate-Distortion-Content Description framework can be used to enhance content description in compressed domain without compromising on rate distortion performance.

The control parameter selection criterion presented in this paper need to be extended to support different quantization levels for fully scalable implementation of the framework. Additionally, a better resource allocation could be achieved by dynamically controlling the optimization process. These aspects will be addressed in the future.

**REFERENCES**


