VIDEO ENCODER OPTIMISATION TO ENHANCE MOTION REPRESENTATION IN THE COMPRESSED-VIDEO-DOMAIN

R. M.T. P. Rajakaruna[1*], W. A. C. Fernando[2] and J. Calic[2]

[1*] Corresponding Author, Department of Mechatronics, Faculty of Engineering, South Asian Institute of Technology and Medicine (SAITM), Sri Lanka, Email: thilini.r@saitm.edu.lk

[2] I-Lab Multimedia Communications Research, Centre for Vision, Speech and Signal Processing, University of Surrey, UK..

ABSTRACT

Compressed-domain content analysis enables fast content-based video applications. However, conventional encoder implementation, limited to optimising data compression, does not necessarily result in content representative compressed features. In this paper, we present a novel method for video encoder optimisation that enhances reliability of content representation in the compressed video. We propose a configurable encoder model to jointly optimise bit-rate, distortion and content representation of the encoded video. In encoder optimisation for motion estimation, we analyse the behaviour of motion representation and compression efficiency over a range of content complexities. A mathematical model to control the extent of noise in selected motion vectors is presented. We demonstrate that the compressed domain motion information can be enhanced without incurring a rate-distortion overhead. The computational overheads are marginal and can be eliminated by exploiting the inherently parallel nature of the proposed optimisation model.

Key words**:** Content Analysis, Compressed-domain, Encoder Optimisation, Video Coding

1. INTRODUCTION

Data embedded in the compressed video domain - encoding parameters such as motion vectors and transform coefficients - provide a low resolution representation of video content. In real-time scenarios, compressed-domain data can be exploited to fulfil the demand for fast video content analysis in a range of content-based applications. For instance in multimedia applications there is an increasing demand for content-based functionalities[1] such as event detection[2] and content driven search. Video content analysis is also increasingly used in domains such as security, system automation and human-computer-interaction.

However, effective video analysis in the compressed domain remains a challenge due to sparsity and noise in the compressed features[3]. This is a result of conventional encoder parameter selection[4], limited to optimising compression efficiency[5], which does not necessarily result in accurate content description in the compressed domain. Compression efficiency is critical for optimum use of bandwidth and storage resources. On the other hand, other aspects of video utilisation such as video content-based applications would benefit from enhanced accuracy of content representation in the compressed video stream.

Video coding methods to enable efficient content analysis have been proposed in literature in [6], [7] and [8]. Methods present in [6] and [7] each support only a given application. In [8] we proposed a content driven motion selection mechanism 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.

In this paper, we propose a configurable encoder model that enhances the reliability of content representation in the compressed video in standard-compliant video coding, while maintaining efficient video compression. In encoder optimisation for motion estimation, an objective function for Motion Description Error (E_{MV}), is combined with bit-rate and distortion constraints in a Lagrange [9] optimisation model. In order to analyse the behaviour of the encoder under different combinations of optimisation parameters and motion content levels, a motion calibrated synthetic data set covering different scene complexities and activity levels was developed as the learning data set. A mathematical model is developed to represent the sensitivity of E_{MV} to optimisation parameters. A fully configurable Rate-Distortion-Motion Description Error (RDE) optimisation criterion is proposed using the E_{MV} sensitivity model, to control the extent of noise in selected motion vectors and compression efficiency.

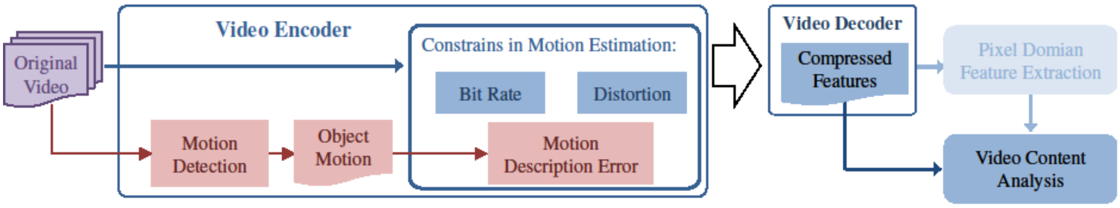


Figure 1: Overview of the proposed encoder implementation with Rate-Distortion-Motion Description Error (RDE) optimisation

An overview of background and the proposed RDE optimisation criterion for video coding are outlined in Section 3. Optimum solutions for the RDE solution space is discussed in Section 4 using a set of synthetic video sequences, while section 5 gives the conclusions.

2. RDE OPTIMISATION CRITERIA

A typical video sequence contains a wide range of content and motion, that require different coding options to achieve optimum encoding efficiency. This requirement for flexibility has been addressed in mainstream video coding standards, by allowing for many different coding modes and coding parameter selections to encode the video[4]. The operational control of the video encoder, that task to determine a set of coding parameters to represent the video sequence, forms the basis of the encoder optimisation problem.

In order to achieve optimum rate-distortion (RD) trade-off, a constrained optimisation function is typically used in encoder parameter selection. This is commonly implemented as a rate constrained cost function. If m denotes a candidate motion vector (MV) within the search region,

$$m(\lambda) = \arg\min\{D(m) + \lambda R(m)\} \quad (1)$$

In (1) the number of bits to encode \emph{m} is given by R(m) and D(m) denote the estimated distortion that would result from using m to displace the coding block. λ is the Lagrange parameter. The optimum value for λ in (1) is defined as a function of quantisation parameter (QP)[5].

The proposed RDE optimisation model address accuracy of candidate motion vectors along with rate-distortion efficiency, in encoder motion estimation process, as illustrated in Fig. 1 In order to accommodate different requirements of applications in expected precision and reliability in the content analysis and video compression efficiency, the optimisation presented in (1) is extended into an unconstrained function. If λ_R and λ_E denote the control parameters for R(m) and E_{MV} objective functions respectively, the extended unconstrained optimisation function is given by,

$$m(\lambda) = \arg\min\{D(m) + \lambda_R R(m) + \lambda_E E_{MV}(m)\} \quad (2)$$

where $\lambda_R > 0$ and $\lambda_E > 0$. A novel objective function, the motion description error, $E_{MV}(m)$ is formulated, defined for a given candidate motion vector m, within the search range. If m = (V_x, V_y), E_{MV} (m), is given by,

$$E_{MV}(m) = \left\| \left(\acute{V_x}, \acute{V_y}\right) - \left(V_x, V_y\right) \right\| \quad (3)$$

where $\left(\acute{V_x}, \acute{V_y}\right)$ is he reference MV that represent expected motion. E_{MV} is modelled independent of the existing rate, distortion objective functions, in order to retain the rate-constraint provided by the conventional rate-distortion optimisation. Additionally, this eliminates the inherent limitations of using translational motion model to represent motion.

3. NUMERICAL SOLUTIONS TO λ_R AND λ_E

Lagrange multiplier theory [9] states that for a given constraint value with specific values for Lagrange multipliers, there exists an equivalent unconstrained function that would yield the same solution. Therefore, the critical problem in the implementation is to select the optimum values for the control parameters, λ_R and λ_E , that would result in the optimum performance under the given circumstance.

In order to identify the optimum values for λ_R and λ_E under different circumstances, the behaviour of the proposed encoder implementation using RDE optimisation criterion needs to be analysed over a wide range of options. Therefore, solutions for the RDE space was obtained over a exhaustive range for λ_R and λ_E as well as quantisation parameter settings and different levels of scene content and motion. A set of synthetic sequences were used that contain known motion activity given in Fig. 2.

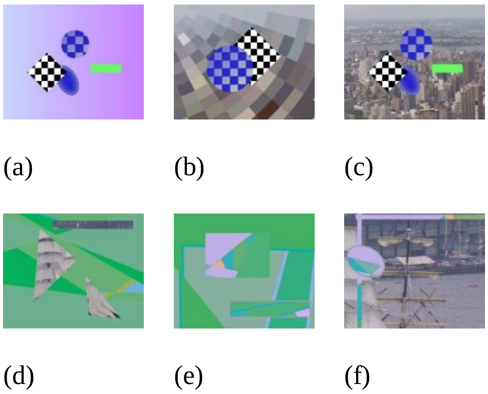


Figure 2: Frame # 0 of the synthetic learning data set (a)-(f)` $Syn_A$'-`Syn_F'

A. RDE Solutions for control parameters

Three performance measurements are considered in the evaluation to reflect the primary objective functions addressed in RDE model; Bit-rate, distortion (measured by average MSE and Y-PSNR) and average motion description error. Average motion description error, denoted by E_{MV}, represents the extent to which MVs selected by each encoder configuration has deviated from accurate motion information, which should reach zero for best performance in content analysis.

In order to identify the impact of the extension to encoder optimisation with E_{MV} objective function, JVT implementation of the state-of-the-art video coding standard, H.264/AVC, JM reference encoder version 15.1[5] was used as the reference encoder in the simulations.

4. EXPERIMENTAL RESULTS

The RDE solutions for the proposed encoder are compared against that of the JM reference encoder. Fig. 3 presents the RDE solutions for `Syn_B' sequence over the tested range of control parameters denoted in green, and the RDE results for JM reference encoder at identical QP values,denoted in black.

In addition, two scenarios were considered outside the defined scope of the RDE optimisation criterion:

- λ_R = 0 to demonstrate the need for R(m) in the optimisation function
- λ_E = 0 to compare performance with the RD optimisation criterion across the tested range of λ_R values.

Fig.3(a) illustrate that, at λ_R = 0, although E_{MV} is significantly low, encoder RD performance is affected by the optimisation criterion. On the other hand, RDE performance for both JM reference and λ_E = 0 demonstrate high E_{MV} values while a sub set of solutions for the proposed encoder optimisation demonstrate optimal solutions in all three objectives. RD solutions are compared in Fig. 3(b). It can be noted that a subset of solutions of the proposed optimisation function, reach RD performance similar to the reference encoder, while some solutions demonstrate an RD gain over both reference encoder and λ_R= 0 scenario. This can be attributed to the reference motion information used in biasing the E_{MV} objective function. The reference motion information provide an additional constraint to the encoder optimisation, thus improving the possibility to avoid local minima in the optimisation.

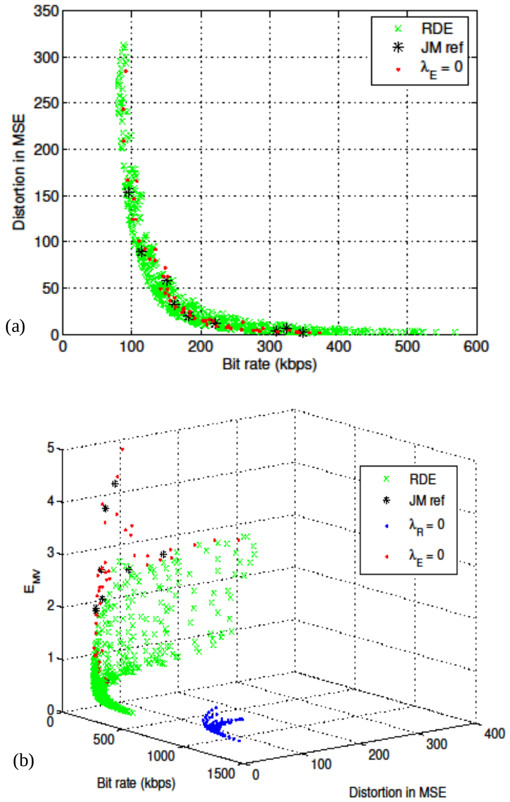
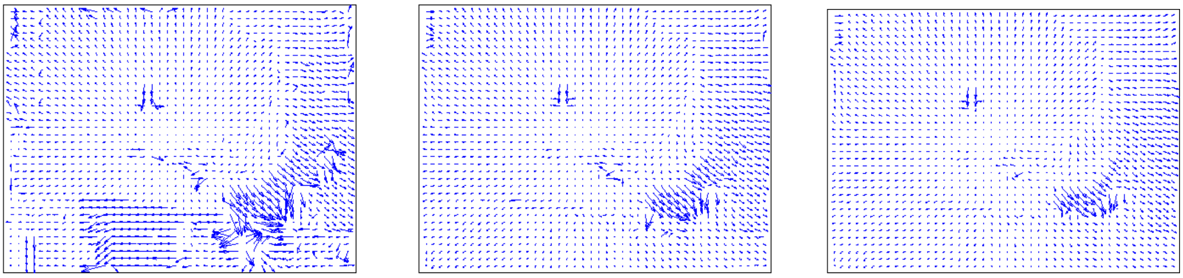


Figure 3: (a)RDE solutions (b)RD solutions for varying control parameters for the `Syn_B' sequence, compared with JM reference

Fig.4 illustrates motion estimated at two control parameter settings, λ_E = 50 λ_R = 7 and λ_E = 150 λ_R = 11 at QP 30 for `Table tennis', compared with that of the JM reference encoder, with the corresponding E_{MV} values. In these three instances, the encoder demonstrate similar RD performance, while E_{MV} varies considerably. The extent of noise in the estimated motion gradually decreases at higher control parameters, as expected in the RDE implementation. For better precision in E_{MV}, λ_E can be dynamically selected for varying scene complexity levels.



(a) (b) (c)

Figure 4: A capture of motion estimated for `Table tennis' sequence with (b)λ_E = 50 λ_R = 7 and (c) λ_E = 150 λ_R = 11 using the RDE optimisation, compared with (a)JM reference encoder output at similar RD performance (bit rate 600.44 kbps & PSNR 34.42 dB) at QP 30.

5. CONCLUSION

In order to facilitate accurate motion description in the compressed video domain this paper presents an unconstrained encoder optimisation model that jointly optimise bit rate, distortion and motion representation. Optimisation criteria is formulated using Lagrange theory, with two control parameters to configure the encoder according to required compression efficiency and the content representation.

It was observed that the RD performance of the RDE model is comparable with the reference encoder for a set of control parameters, at improved motion representation. A mathematical model was derived for the sensitivity of Motion Description Error, E_{MV}. The practical implementation is presented using natural video sequences. The results demonstrate that the proposed RDE optimisation criteria can be used to enhance motion representation in compressed-domain without compromising on RD performance. Additionally, by implementing the reference data estimation in parallel with the cost function for RDE optimisation, overall encoder complexity can be minimised significantly.

6. REFERENCES

[1]. N. Dimitrova, Hong-Jiang Zhang, B. Shahraray, I. Sezan, T. Huang, and A. Zakhor, "Applications of video-content analysis and retrieval," Multimedia, IEEE Trans. on, vol. 9, no. 3, pp. 42 – 55, jul-sep 2002.

[2]. Jin S. H. and Ro Y. M., "Video event filtering in consumer domain," Broadcasting, IEEE Trans. on, pp. 755 –762, 2007.

[3]. M.T. Coimbra and M. Davies, "Approximating optical flow within the mpeg-2 compressed domain," Circuits and Systems for Video Technology, IEEE Trans. on, pp. 103 – 107, 2005.

[4]. T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G.J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," Circuits and Systems for Video Technology, IEEE Trans. on, pp. 688 – 703, 2003.

[5]. G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," Signal Proc. Magazine, IEEE, pp. 74 –90, 1998.

[6]. S.K. Kapotas and A.N. Skodras, "A new data hiding scheme for scene change detection in h.264 encoded video sequences," in Multimedia and Expo (ICME), 2008 IEEE Int. Conf. on, 2008, pp. 277–280.

[7]. C. Kas and H. Nicolas, "Joint global motion estimation and coding for scalable h.264/svc high-definition video streams," in Content-Based Multimedia Indexing, Int. Workshop on, 2008, pp. 433 –438.

[8]. R.M.T.P. Rajakaruna, W.A.C. Fernando, and J. Calic, "Facilitating motion-based vision applications by combined video analysis and coding," in Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE Int. Conf. on, 2010, pp. 1102 –1105.

[9]. H Everett, "Generalized lagrange multiplier method for solving problems of optimum allocation of resources," Operations Research, pp. 399–417, 1963.

[10]. D. Marpe, T. Wiegand, and G.J. Sullivan, "The h.264/mpeg4 advanced video coding standard and its applications," Communications Magazine, IEEE, pp. 134 –143, 2006.

[11]. G. Sullivan, T. Wiegand, and K. Lim, "Joint model reference encoding methods and decoding concealment methods," Document JVT-IO49, San Diego, 2002.

[12]. R.M.T.P. Rajakaruna, W.A.C. Fernando, and J. Calic, "Application-aware video coding architecture using camera and object motion models," in Industrial and Information Systems, 2011 IEEE Int. Conf. on, 2011