OFF-LINE MOTION DESCRIPTION FOR FAST VIDEO STREAM GENERATION IN MPEG-4 AVC/H.264

Yi Wang[*1], *Xiaoyan Sun*[2], *Feng Wu*[2], *Shipeng Li*[2], *Houqiang Li*[1], *Zhengkai Liu*[1]

[1]University of Science and Technology of China, Hefei, 230027, P.R. China
[2]Microsoft Research Asia, Beijing, 100080, P.R. China
[1]wy1979@mail.ustc.edu.cn, {lihq, zhengkai}@ustc.edu.cn, [2]{xysun, fengwu, spli}@microsoft.com

ABSTRACT

The rate-distortion optimal mode decision as well as motion estimation adopted in H.264 brings a big challenge to real-time encoding and transcoding duo to the high computation complexity. In this paper, we propose a hierarchical motion description model to present the motion data of each macroblock (MB) from coarsely to finely. A preprocessing approach is developed to estimate the motion data for each MB at each quality level with regard to its reference quality, its adjacent MBs and the target bit-rate. The resulting motion data can be coded and stored as metadata in a media file or a stream. Moreover, we propose a method to readily extract the specific motion data from the model for each MB at given bit-rates. Experimental results have shown the effectiveness of our proposed motion description model in terms of coding efficiency as well as fast bit-rate adaptation in comparison with that of H.264.

1. INTRODUCTION

Variable block size and rate-distortion (RD) optimization techniques are two key components that greatly improve the coding efficiency of MPEG-4 AVC/H.264 (H.264 in short) in a wide range of bit-rates [1][2]. H.264 defines seven MB partition modes from 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, to 4x4. Each MB is allowed to have different numbers of motion vectors (MVs) dependent on its partition mode.

The optimal partition mode as well as MVs for a macroblock is determined by the following Lagrange cost function [3][4],

$$J = D + \lambda \times R . \quad (1)$$

Here D is the distortion of the reconstructed frame. R is the bit-rate of coded motion data and quantized coefficients. λ is Lagrange multiplier, which is related to QP. It is a process with excessive computational complexity.

Furthermore, the selected optimal MVs and modes are strongly bound with the target bit-rate. If the motion data obtained at high bit-rates is directly applied to generate bit streams at low bit rates, the coding efficiency may drop more than 6dB. It also brings a big challenge to the existing fast transcoding techniques [5][6], since motion re-usage is no longer appropriate to transcode H.264 bit streams. The motion data has to be refined or re-estimated towards a given target bit-rate.

A question is rising here. Can we achieve an off-line motion description for each raw video sequence or generated stream to get rid of complicated motion estimation in encoding or transcoding? In this case, motion data at any target bit-rate can be directly extracted from the off-line motion description. With the off-line motion description, one can certainly expect that the encoding as well as the transcoding of one video content would become quite rapidly and efficiently. It will be of great benefit to video streaming, home video generation, video sharing, etc.

However, it is not an easy job to achieve a feasible off-line motion description since 1) the motion data at different bit-rates are quite different, and 2) motion data of a macroblock is strongly related to its reference quality and the motion data of its temporally as well as spatially neighboring macroblocks. In this paper, a motion alignment scheme is introduced in which a hierarchical model is presented to describe the motion data of each macroblock from coarsely to finely. In this model, the motion data of one macroblock is organized by partition modes rather than by quantization steps. A pre-processing approach is developed to obtain the MVs of every macroblock under all modes. Then, the resulting motion information is efficiently compressed utilizing the correlations among spatially neighboring macroblocks as well as partition modes to form the off-line motion description of the video content. Furthermore, we propose a RD equal-slope approach to readily extract a specific set of MVs from the description for efficient video representation at a given quantization step.

The rest of this paper is organized as follows. Section 2 describes our proposed motion model and develops an algorithm to estimate motion data. Section 3 discusses how to efficiently compress motion description. Section 4 proposes a method to extract a specific motion data from the motion description. Experiment results are reported in Section 5. Conclusion and discussion are presented in Section 6.

[*] This work is done during Wang's internship at MSR Asia.

2. MOTION DESCRIPTION

It can be observed that different optimal MVs as well as partition modes are selected to encode the same video content at different quantization parameters (QPs) in MPEG-4 AVC/H.264. A straightforward way to enable the off-line motion description is that the motion information generated at each QP setting is recorded in series. Hence, given a target QP, we can directly achieve the corresponding motion information from the motion description and provide the exact same performance as that of the H.264 method

But there are several drawbacks to this method. Firstly, the volume of the resulting motion data is huge and there exists many redundant MVs among the sets of motion data. However, if we try to directly reduce or merge some sets of MVs, some preliminary experiments have shown that it will significantly drop the coding performance because the motion data of each MB is strongly related to its temporally and spatially neighboring MBs. Secondly, such motion description method is not able to be used when the MBs in one frame are coded at different QPs, for example, to support constant bit-rate (CBR). In this case, the motion information would be assembled by extracting the MVs from different sets of recorded motion data. Generally, the resulting motion information is far from the optimal one due to which the performance is degraded greatly.

2.1 Motion Alignment

To limit the redundancy among MVs, motion alignment is firstly introduced into the motion description method in which a hierarchical model is presented to describe the MVs of each MB. In this model, motion data is organized by partition modes instead of QP steps. As shown in Figure 1, the motion data of a MB is categorized into seven groups according to the partition modes (Skip mode and Direct mode are converted into 16x16 mode to reduce the dependency among MBs). In one hand, there may be several sets of MVs in one group in favor of different quality references and neighboring MVs; On the other, one group may have NULL MVs which indicates that this mode has never been used in encoding.

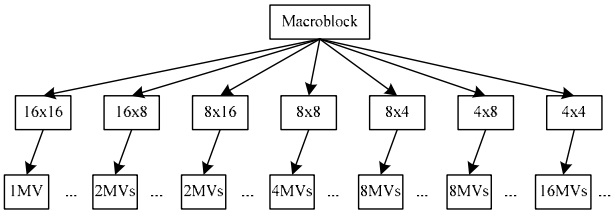

Figure 1. The hierarchical model for motion description.

Furthermore, it is often the case that the smaller the QP is, the smaller the partition mode is selected (here we regard 16x16 mode as larger partition model while 4x4 model as smaller partition mode). Accordingly, an important constraint is introduced to the hierarchical model, namely, when mode A is selected for one MB at a certain QP by equation (1), mode B with larger sub-block size will not be selected for this MB in case of QP increasing. Our later experiments will show that such a constraint has little effect on coding efficiency.

2.2 Motion Estimation

How to generate the set of MVs of each group is another tough problem of off-line motion description. It is because the optimal motion and mode of one MB are hard to be determined in H.264 without the information of the reference and the MVs of its adjacent MBs. However, the reference and the adjacent MVs are unavailable until real encoding, especially in the case of CBR.

To solve this problem, we propose a motion estimation method to obtain the MV groups that can result in the similar performance as that of the original H.264. It is described as follows. Firstly, we introduce two modules in H.264 encoder: multi-QP motion estimation and motion quantization, as shown in Figure 2. The QP ranges from QP_{min} (the smallest QP to be supported) to QP_{max} (the biggest QP). Multi-QP motion estimation is to determine the optimal modes as well as motion vectors corresponding to different QPs for each MB. Motion quantization is to generate the motion description based on our proposed hierarchical model. The two processes are performed at MB level. The frame buffer will hold all reconstructed references at different QPs.

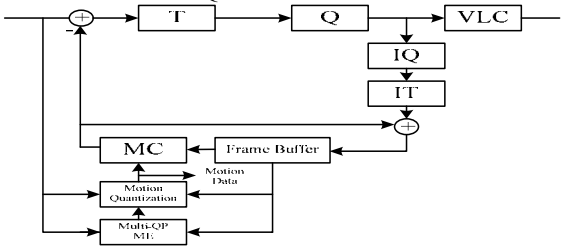

Figure 2: Motion estimation for motion description

The details of multi-QP motion estimation are described as follows.

Algorithm: Multi-QP Motion Estimation

For each MB in P frame, set QP = QP_{min}

Step 1: Update the reference frame buffer and the neighboring motion buffer according to QP.

Step 2: Decide optimal mode and motions at QP by motion estimation and mode decision. Notice that the candidate modes should have sub-block size no smaller than the previous selected mode for QP-1 in the case of QP> QP_{min};

Step 3: Set QP =QP+1 if QP< QP_{max}, and go to Step 1.

It can be observed that different sets of MVs may be selected under one partition mode for one MB. To further reduce the MV redundancy in one mode, motion quantization module is presented to reduce the number of sets of MVs as much as possible to facilitate the motion description. It is described as follows.

Algorithm: Motion Quantization

Given the sets of motion vectors under mode A

Step 1: Group all sets of motion vectors as a cluster.

Step 2: Find a set of motion vectors (MV_{best}) to minimize the criterion (2)

$$MV_{best} = \arg\min_{i \in S} \sum_{j \in G} (SAD(mv_i) + \lambda_j * R_{mv_i}) \qquad (2)$$

Here S is all sets of motion vectors and G is the QP collection corresponding to the mode. λ_j is Lagrange multiplier for QP_j.

Step 3: Remove the sets of MVs if their Euclidean distance to MV_{best} is smaller than a given threshold.

After the motion quantization, we achieve the off-line motion description by providing the motion vectors under each mode at MB level. At the same time, we can also record the value of λ when the mode is selected. λ is a option parameter in the proposed motion description. If it exists, we can significantly reduce the complexity of the motion extraction.

3. CODING OF THE MOTION DESCRIPTION

In this section we will discuss on how to represent the motion description efficiently. Our statistic analysis has shown that the motion vectors of different modes for one MB have stronger correlation than that of the neighboring MBs. Therefore, we code MVs of a MB according to the sub-block size from big to small. A predicted motion vector, MVP, is formed by taking into account the correlation of the adjacent macroblocks as well as modes.

1) If the mode of a MB is the first one without NULL MVs, the first set of MVs in this mode is predicted from the median of MVs from the available left, up, up-right MBs. The other sets of MVs in this mode are predicted from the previous coded set.

2) The sets of MVs in the following modes are predicted from the last sets of MVs in the previous coded mode.

The obtained residues of MV are coded by Exp-Golomb entropy coding. If λ exists, it is also compressed with the prediction from the previous coded λ by Exp-Golomb entropy coding.

4. MOTION DATA EXTRACTION

As the mode and motion information are well organized in the hierarchical model, we need to develop a method to efficiently extract the desired mode and set of motion vectors for each macroblock according to the target QP.

If only mode and motion information are included in the description, we have to check all the recorded modes and motion vectors to select the best pair for one MB according to equation (1). Whereas, it should be noted that since the MVs in the description are accurate enough, the distortion *D* and the rate *R* in equation (1) can be computed by prediction residue and motion residue rather than by reconstructed reference. This makes the selection process much simple without loss in performance.

In case assistant information λ exists, it can help accelerate the motion extraction. As we know, λ indicates the slope of a RD curve. Given a QP, we can find the nearest λ corresponding to the QP. Then the mode corresponding to the λ will be selected as the optimal one. If more than one candidate sets of motion vectors are reserved for this mode, we can select the best one using the method mentioned above. Motion vectors for other modes can be ignored. Thus, the complexity of the motion extraction can be reduced significantly with λ provided.

5. EXPERIMENTAL RESULTS

In this section we present several experiments to evaluate the efficiency of our proposed motion description algorithm. The proposed scheme is tested on the reference software JM61e of H.264. Four standard sequences, foreman, mobile, bus, and news (QCIF, 30HZ, 90 frames), are used for all simulations.

5.1 Size of motion description

We compared the sizes of raw video sequence with that of the proposed motion description. The results are shown in Table 1. For each frame, the overhead bit spent on the motion description is less than 4% of total bits. It can be readily stored as user data in a media file or a stream.

Table 1. Comparison between bits of raw video sequence and motion description

sequence	raw video sequence (bits/frame)	motion description (bits/frame)	
		One Mode One Motion	One Mode Multi-motion
Foreman	304128	7355(2.4%)	8874(2.9%)
Bus	304128	8817(2.9%)	11595(3.8%)
Mobile	304128	7658(2.5%)	9337(3.1%)
News	304128	2505(0.8%)	3007(1.0%)

5.2 Motion description in fast AVC/H.264 encoding

We apply the motion description in fast MPEG-4 AVC/H.264 encoding to verify its accuracy. Three cases are tested. Case 1 and Case 2 adopt our proposed motion description and motion extraction algorithms. Case 3 is the original MPEG-4 AVC/H.264 scheme. In detail, for one MB,

- Case 1: Only one set of MVs is reserved in one mode (OMOM)
- Case 2: More than one set of MVs can be recorded for one mode (OMMM)
- Case 3: RD optimal MVs and mode are selected by H.264 scheme.

The comparison results are shown in Figure 3. Case 2 performs slightly better than Case 1 due to more available motion information. In fact, both of Case 1 and Case 2 with proposed motion description achieve very similar performance compared with the original MPEG-4

AVC/H.264. This proves that the proposed motion description can effectively represent the motion property of video sequences at different bit-rates.

5.3 Motion description in fast AVC/H.264 transcoding

Furthermore, we apply the motion description in transcoding scenario. The efficiency of our method is tested in comparison with two traditional transcoding methods: Full-ME and Motion Re-usage.

- Case 1-2: the same as those in Experiment 5.2
- Case 3: Full-ME
- Case 4: Motion Re-usage

The input high quality video stream is generated by H.264 encoder with QP equal to 8. The decoded sequence is referred to as original sequences in the testing. The motion vectors in the input bit streams are re-used in Case 4. The results of Case 1-3 are achieved as same as those in Experiment 5.2.

During transcoding, the number of SAD computation for a MB encoding in Full-ME is 7742 (Search range is 16.); while this number is averagely reduced to 2 with our proposed motion description. Motion re-usage does not perform SAD computation. As shown in Figure 4, the proposed scheme provides similar performance with FULL-ME at lower computation complexity. Motion re-usage is the simplest method among the four schemes, but its coding efficiency degrades greatly especially at low bit-rates.

6. CONCLUSION AND DISCUSSION

In this paper, we propose an off-line motion description organized in a hierarchical model for fast video stream generation. A pre-encoding algorithm is developed to obtain the modes and motion vectors of the motion description. Motion alignment as well as motion quantization is introduced in the mutli-QP motion estimation process. Furthermore, we propose an RD equal-slope approach to readily extract a specific motion data from the description. Experimental results have shown the effectiveness of our proposed method in terms of coding efficiency and computation complexity.

The generation of our proposed motion description is computational intensity, because the sequences have to be encoded at several QPs (the number is 21 in our experiment). But it can be performed off-line. In addition, as long as the description is ready, it can be used to produce the H.264 video stream at any bit-rate without the complicated motion estimation process.

Definitely, our proposed motion description can be used to empower the video adaptation for applications such as video streaming, IPTV, video sharing etc. For example, in the case of video streaming, the motion description could be produced in advance and stored as user data in server. Whenever a H.264 transcoding or encoding is needed, the server can easily extract a set of proper motion vectors from motion description and apply them into stream generation. Due to the omitting of motion estimation, bit-streams which meet the bandwidth can be real-time generated with high coding efficiency comparable with that of H.264.

7. ACKNOWLEDGEMENT

The work of Yi Wang, Houqiang Li, and Zhengkai Liu are supported by NSFC under contract No. 60333020 and open fund of MOE-Microsoft Key Laboratory of Multimedia Computing and Communication under contract No. 05071803.

8. REFERENCES

[1] "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H. 264/ISO/IEC 14496-10 AVC," in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-GO50, 2003.

[2] T.Wiegand, G.J.Sullivan, G.jontegaard, and A. Luthra: "Overview of the H.264/AVC Video Coding Standard", IEEE Trans. Circuits and Syst. for Video Technol., vod. 13, no. 7, pp. 688-703, Jul 2003.

[3] G.J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compressin", IEEE Signal Processing Magazine, vol. 15, no. 6, pp. 74-90, Nov. 1998.

[4] T.Wiegand and B. Girod, "Lagrangian Multiplier Selection in Hybrid Video Coder Control," in Proc. ICIP 2001, Thessaloniki, Greece, Oct. 2001.

[5] J. Xin, C.–W. Lin, and M.–T. Sun, "Digital Video Transcoding", Proceedings of the IEEE, Vol. 93, No. 1, Jan 2005

[6] Sung-Eun Kim, Jong-Ki Han, Jae-Gon Kim, "Efficient Motion Estimation Algorithm for MPEG-4 to H.264 Transcoder", in Proc. ICIP 2005, Genoa, Italy, Sep.2005

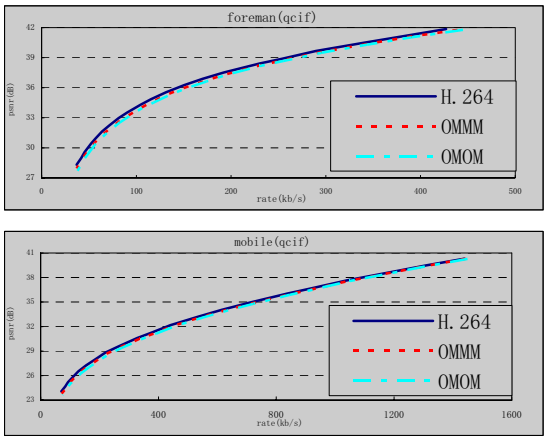

Figure.3. Performance comparison in encoding

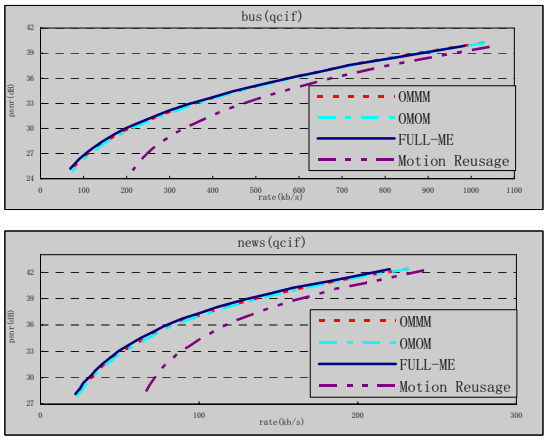

Figure.4 Performance comparison in transcoding