LETTER

Efficient Motion Vector Re-Estimation Based on a Novel Cost Model for a H.264/AVC Transcoder

Soongi HONG[†a)], ***Nonmember***, **Yoonsik CHOE**[†b)], ***Member***, ***and*** **Yong-Goo KIM**[††c)], ***Nonmember***

SUMMARY In transcoding, it is well known that refinement of the motion vectors is critical to enhance the quality of transcoded video while significantly reducing transcoding complexity. This paper proposes a novel cost model to estimate the rate-distortion cost of motion vector composition in order to develop a reliable motion vector re-estimation method that has reasonable computation cost. Based on a statistical analysis of motion compensated prediction errors, we design a basic form of the proposed cost model as a function of distance from the optimal motion vector. Simulations with a transcoder employing the proposed cost model demonstrate a significant quality gain over representative video transcoding schemes with no complexity increase.

key words: *H.264/AVC, transcoding, motion vector re-estimation, motion vector cost estimator*

1. Introduction

The recent explosive deployment of smart devices poses a challenge for creating a seamless multimedia service across all such terminals, which have different computation and display capabilities. To meet the demands for high quality of experience (QoE) over such varying devices, dynamic adaptive streaming via HTTP has been standardized by 3GPP [1] and MPEG [2], [3], etc. These standards basically provide mechanisms to change service quality according to network conditions, and large scale service provisioning can be easily realized because they are based on the standard HTTP protocol. However, changes in service quality require a number of predefined media presentations (MPs); it is difficult for this number to be large because this MP corresponds to a copy of an encoded media stream at a certain quality level. Therefore, speedy and reliable transcoding has regained a lot of attention in order to support a higher quality of seamless multimedia service. In this paper we focus on H.264/AVC video transcoding which involves resolution changes for accommodating the variety of terminal displays.

Such transcoding involves resolution changes [4], [5], and it is well known that the composition and refinement of motion vectors (MVs) are very important because they directly affect the quality of transcoded video and can also significantly reduce the complexity of transcoding. In order to quickly find an appropriate MV for each prediction unit of a reduced resolution video, there have been various MV re-estimation (MVRE) approaches that combine the corresponding MVs of higher resolution.

In order to develop a reliable MVRE method without an excessive increase in computations, this paper proposes a novel cost model to estimate the rate-distortion (RD) cost of MV composition and employs H.264/AVC transcoder with it.

The remainder of this paper is organized as follows. Section 2 explains the proposed cost model for the MVRE of the H.264/AVC transcoder. Section 3 presents an MVRE method based on the proposed cost model. Section 4 provides the simulation results. Finally, Sect. 5 concludes this paper.

2. The Proposed Rate-Distortion Cost Model for Motion Vector Composition

To obtain a relationship between the optimal MV mv_i and the motion compensated prediction (MCP) error, D_i, D_i is plotted against distance l, $l = \pi(dmv) = \pi(mv - mv_i)$, where π is distance measure and dmv is the motion vector difference (MVD) between any MV, mv, and the optimal MV, mv_i. For confirmation of these relationships, we calculate the differential distortion $\Delta D_i/\Delta l$ using a JM encoder [6] with a fixed QP. Based on these experimental results, we assume that the differential distortion can be denoted by

$$\frac{dD_i}{dl} \simeq \frac{\Delta D_i}{\Delta l} = a\frac{c_i}{l}, \tag{1}$$

which depicts the proportional relation between the $\Delta D_i/\Delta l$ and c_i and the inverse-proportional relation between the $\Delta D_i/\Delta l$ and l.

Using (1) and assuming $D_i(l)$ to be integrable $l > 0$,

$$D_i(l) = ac_i \int \frac{1}{l}\, dl = \begin{cases} ac_i ln(l) + b & \text{if } l > 0, \\ c_i & \text{if } l = 0 \end{cases} \tag{2}$$

with a and b parameterizing the functional relationship between distance, l, and distortion, $D_i(l)$. These parameters come from simulations using a variety of test sequences and described in Table 1.

For bits spent on the MVs, R can be written as

Manuscript received August 13, 2015.
Manuscript revised October 18, 2015.
Manuscript publicized December 4, 2015.

[†]The authors are with Department of Electrical & Electronics Engineering, Yonsei University, Shinchon-dong, Seoul 120–749, South Korea.
[††]The author is with Department of Media Engineering, Korean German Institute of Technology, Sangam-dong, Seoul 121–270, South Korea.
a) E-mail: s82.hong@samsung.com
b) E-mail: yschoe@yonsei.ac.kr
c) E-mail: ygkim@kgit.ac.kr

DOI: 10.1587/transinf.2015EDL8181

Table 1 Model parameters of the proposed rate-distortion cost estimator for motion vector re-estimation

Model Param.	MCP Mode	QP22	QP27	QP32	QP27
a	P16x16	0.270	0.176	0.110	0.068
	P16x8	0.222	0.175	0.122	0.077
	SMB8x8	0.290	0.220	0.152	0.099
	SMB8x4	0.319	0.261	0.188	0.131
	SMB4x4	0.386	0.321	0.239	0.170
b	P16x16	47970.726	54673.469	70479.474	94136.002
	P16x8	21883.706	27772.913	37053.554	48359.039
	SMB8x8	13441.495	15997.214	20363.250	27043.093
	SMB8x4	7723.545	9752.059	12649.936	16757.108
	SMB4x4	4638.107	5744.370	7352.676	9550.192

$$R(mv) = \text{mvd2bitsTb}[mv - pmv_p], \tag{3}$$

where mv is the desired MV of current blocks, pmv_p is the predicted motion vector (PMV) extracted from neighborhood MBs of pre-coded video, and "mvd2bitsTb" is the conversion table for the motion vector difference (MVD), $mv - pmv_p$, to the bits which are a result of entropy coding.

Using (2) and (3), our proposed MV cost estimator is defined as

$$f_i(mv) = D_i(\pi(mv - mv_i)) + \lambda_{\text{motion}} R(mv), \tag{4}$$

where mv is the desired MV of current blocks, mv_i is the the optimal MV of the current block, and π is the distance measure. To use the MV cost estimator in (4), the distance measure π has to be defined properly. In this paper, π is mathematically induced from simple assumptions. In the H.264/AVC, an image is segmented and encoded into blocks of variable block-sizes from 4x4 to 16x16. Being small enough to compare the size of an image and properly estimated by MCP, we assume that these segmented image blocks have the monotonous MCP error. Since the MCP error increases with concentric circles in the this case, the distance measure has the equivalent and independent weights of horizontal and vertical direction. Thus, the distance measure can be defined as

$$l = \pi(mv - mv_i) = \pi(dmv_i) = |dmv_x| + |dmv_y|, \tag{5}$$

where dmv_x and dmv_y are the horizontal component and vertical component of the MVD dmv_i, respectively.

We evaluate the correctness of the proposed cost estimation model and the model parameters. First of all, we divide the motion vector domain centered around mv_i into four sub-region to examine the regional impacts. And then, we calculate the mean absolute percentage error (MAPE) between the estimated rate-distortion costs and the conventional H.264/AVC rate-distortion costs. The MAPE in each sub-region is calculated as

$$E_{R_d} = \frac{1}{N_{R_d}} \sum_{mv \in R_d} \left| \frac{f_i(mv) - j(mv)}{j(mv)} \right|, \tag{6}$$

where N_{R_d} is the cardinality of the sub-region set $R_d = \{mv : 4(d-1) \le l < 4d, d \in \{1,2,3,4\}\}$, $f_i(mv)$ is the estimated rate-distortion cost at mv, and $j(mv)$ is the calculating values from the H.264/AVC rate-distortion equation at mv.

Table 2 The mean absolute percent error of the proposed cost estimation model comparing with H.264/AVC at each sub-region and four QPs.

QP	R_1	R_2	R_3	R_4	Avg.
QP22	0.225	0.163	0.215	0.259	0.216
QP27	0.189	0.153	0.209	0.257	0.202
QP32	0.142	0.134	0.195	0.251	0.181
QP37	0.096	0.111	0.174	0.239	0.155
Avg.	0.163	0.140	0.198	0.252	0.188

Table 2 depicts the MAPE of the proposed cost estimator at each sub-region and four QPs. Generally speaking, the rate-distortion functions of high QP values tend to become simpler than that of low QP values. Thus, the MAPE is decreased with larger QP value. In aspect of regional impacts, the MAPEs of R_1 and R_2 are significantly less than that of other regions. Our proposed cost estimation model can work as a good guide for the motion vector re-estimation because the the average MAPE of all of sub-regions and QPs is only 0.188.

3. Motion Vector Re-Estimation Based on the Proposed Cost Model

Let mv_i be the MV, md_i be the MCP mode, and c_i be the MCP error from the ith corresponding MB partition. Thus, the context sets extracted from pre-coded bitstream are a set of K corresponding MVs $V = \{mv_1, mv_2, \cdots, mv_K\}$, a set of K corresponding MCP errors $D = \{c_1, c_2, \cdots, c_K\}$, and a set of K corresponding MCP modes $M = \{md_1, md_2, \cdots, md_K\}$. For the proposed cost estimator, we have to define the corresponding region which is spatially related to the current MB of low resolution domain (transcoded bitstream) in high resolution domain (pre-coded bitstream) and the downsizing factor as described in Fig. 1. As there are at most K MVs, the estimated MV cost J can be obtained as

$$\begin{aligned} J(mv) &= \frac{1}{S} \sum_{i=1}^{K} \frac{s_i^I}{s_i^I + s_i^O} f_i(mv) \\ &= \frac{1}{S} \sum_{i=1}^{K} \eta_i D_i(\pi(mv - mv_i)) + \lambda_{\text{motion}} K R(mv), \\ \eta_i &= \frac{s_i^I}{s_i^I + s_i^O}, \quad S = \sum_{i=1}^{K} \eta_i, \end{aligned} \tag{7}$$

where $mv_i \in V$ is a MV from the ith corresponding MB partition of the pre-coded H.264/AVC video, $c_i \in D$ is a distortion of a corresponding MV mv_i, s_i^I is the size of MCP mode in the corresponding region, s_i^O is the size of MCP mode out of corresponding region, η_i is an area-weighted strength of ith corresponding MB mode, and $f_i(mv)$ is designated in (4). In the case of downsizing by an integer factor, η_i is always one because s_i^O is always zero which means that a whole area of MCP mode is included in a corresponding region. Note that if a corresponding partition is coded with skip mode in a precoded H.264/AVC video, then its MV is set to the one predicted from the MVs of previously coded partitions. Also, if a corresponding partition is intra-coded, it

will not be involved in estimating the new MV. For this reason, if all the corresponding partitions are intra-coded, the new MB partition will also be transcoded with intra mode.

As described in Fig. 1, the proposed algorithm can estimate the MV cost of current P16x16 MB and SMB modes if the corresponding regions are defined in accordance with various downsizing factors and inter modes. Thus, our proposed algorithm can support all of inter modes without a SKIP mode, which means that P16x16, P16x8, P8x16, P8x8, SMB8x8, SMB8x4, SMB4x8, and SMB4x4 are supported by our proposed algorithm.

Using (7), we can estimate the cost of any given MV mv. The search pattern, the set S of the MVs, needs to be defined for serving MVs to the proposed MV cost estimator. If the set S whose cardinality L is defined as $S = \{mv_s : s = 1, \cdots, L\}$, the estimated cost set E is represented as $E = \{J(mv_s) : s = 1, \cdots, L\}$ where $J(\cdot)$ is the proposed cost estimator in (7). Then, let E^{TO} be a totally ordered set of E, which is denoted by

$$E^{\mathrm{TO}} = \{J(mv_k) : J(mv_k) \le J(mv_{k+1}), mv_k \in S\}, \tag{8}$$

where $k = 1, \cdots, L$. Because the elements of E^{TO} are estimated costs of given MVs and totally ordered, we can establish a set C of the re-estimated MV candidates by getting the MVs associated with the lower index elements of the set E^{TO}. Therefore, a set C is defined as

$$C = \{mv_k : J(mv_k) \in E^{\mathrm{TO}}, k = 1, \cdots, N\}, \tag{9}$$

where $N \le L$. Consequently, a set C is a subdivision of S into non-overlapping, nonempty subsets and has MVs associated with the Nth minimum estimated costs of E^{TO}. It is known that the search pattern has important influence on speed and distortion performance in block motion estimation. Generally speaking, the large cardinality of set S can provide good coding efficiency, but they spend a lot of time testing the search points in the set S. So, to balance speed and distortion, we propose the corresponding MV centered local search strategy. For example, if the current MB has the three corresponding MVs $mv_i \in V$ in the pre-coded H.264/AVC video and the local search points $mv_r \in P$ with 1 integer-pixel resolution, the total number of search points (L) is 27 which is equal to the size of the product set $S = V \times P$. Then, if the four MVs are used for candidates of a re-estimated MV ($N = 4$), the MVs with the estimated cost $J(mv_k) \in E^{\mathrm{TO}}$ from first to fourth are picked as the candidates $mv_k \in C$ from the MVs of S. Finally, we select a re-estimated MV from the MV candidates and their ± 1 quarter-pixel neighborhoods by the Lagrangian cost function of H.264/AVC in the fractional MV refinement process. Because a target application of this letter is the motion vector re-estimation scheme, the inter mode selection method is same as that of the H.264/AVC JM reference software. The only difference between these two methods is that the MV estimation of H.264/AVC is replaced by the proposed cost estimation based motion vector re-estimation (CEMVR).

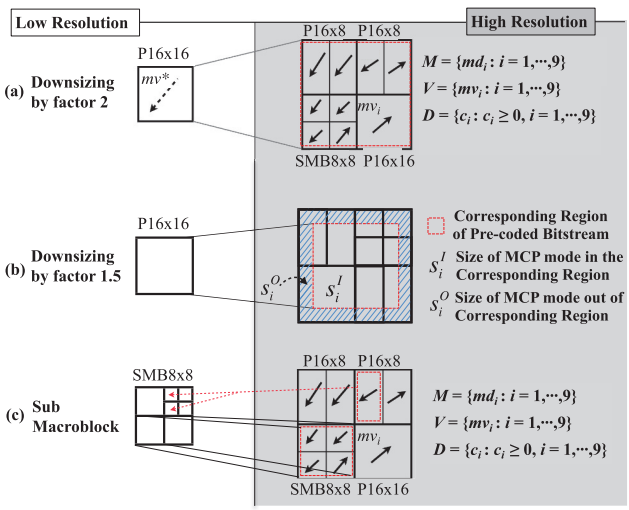

Fig. 1 Mapping schemes of corresponding region for a current macroblock depending on the downsizing factors and inter modes of (a) P16x16 with downsizing factor 2, (b) P16x16 with downsizing factor 1.5, and (c) SMB with downsizing factor 2.

4. Simulation Results

We conducted a sets of experiments to evaluate and compare the performance of the proposed CEMVR method against other existing methods for transcoding pre-coded H.264/AVC videos to downsized H.264/AVC videos. The detailed information about the test sequences and encoding conditions are listed in Table 3. The sequences were encoded in the H.264/AVC format with IPPP GOP structure, and then transcoded to new sequences at a 4-to-1 reduced frame size using the cascaded H.264/AVC Full-Decoding and Full-Encoding (FDFE) scheme, the existing H.264/AVC transcoding schemes (AWMVM [4] and KMCMV [5]), and the proposed H.264/AVC transcoding scheme (CEMVR). Full-search fractional (half- and quarter-pixel accuracy) ME with ± 32 as the search range and the infinite GOP size are used in the proposed, KMCMV, AWMVM and FDFE techniques. For objective comparison, the BD-PSNR and BD-RATE [7] are computed with respect to the sequentially processed video by decoding of pre-encoded bitstream and applying a downsampling filter to it, where the BD-PSNR and BD-RATE are the percentage of increasing PSNR at

Table 3 Test sequences and simulation conditions

Sequences	Resolution	Motion Type	FPS	No. Frames
SpeedBag			30	570
ElephantsDream		Regular	24	1569
BigBuckBunny		Motion	24	1431
BluSky	FHD		25	217
Factory	(1920×1080)		30	1339
PedestrianArea		Irregular	25	375
SintelTrailer		Motion	24	1253
Tractor			25	690

H.264/AVC Encoding Conditions				
Main Profile	1 Ref. Frame	Full Search Mode	32 MV Search Range	RDO On

Table 4 BD-Bitrates obtained by transcoding H.264/AVC precoded sequences

Pre ↓ Trans	Sequence	BD-Bitrate (%)		
		AWMVM	KMCMV	Proposed CEMVR
1920x1080 ↓ 960x540	SpeedBag	13.50	13.21	3.94
	ElephantsDream	10.99	10.54	3.98
	BigBuckBunny	7.40	7.17	0.98
	BluSky	7.34	6.76	1.38
	Factory	50.17	48.41	23.41
	PedestrianArea	27.26	26.02	7.89
	SintelTrailer	23.65	23.61	5.38
	Tractor	31.30	29.65	10.46
1080p Regular Motion		9.81	9.42	2.57
1080p Irregular Motion		33.10	31.92	11.79
Overall Average		21.45	20.67	7.18

the same bitrate and increasing bitrate at the same PSNR, respectively. The simulation was conducted on a PC with an Intel Core 2 Quad Q6600 2.4GHz, and 12GB of DDR3 RAM. Note that all the techniques are implemented based upon the H.264/AVC recommendations and JM17.2 reference software.

Table 4 shows BD-Bitrates. A positive value in the table indicates an increase in bit-rates for the BD-Bitrates compared to the FDFE method. Obviously the BD-Bitrate values in the table are almost positive because the FDFE method has the most superior performance compared to other methods in terms of compression efficiency. The performance of the proposed CEMVR (i.e., with four candidates and 1 quarter-pixel refinement) technique in terms of BD-Bitrate (see Table 4) outperforms the AWMVM (with 1 quarter-pixel refinement) and KMCMV (with four candidates and 1 quarter-pixel refinement) techniques due to the accuracy of the cost estimation model for MV selection (see Sect. 2) and MVRE procedure based on multiple candidate MVs (see Sect. 3). We recommend the proposed CEMVR technique for speeding up the entire encoding time with enhanced rate-distortion performance compared to other conventional techniques. The KMCMV, which has the multiple candidate MVs, demonstrated a meaningful gain (1.18% BD-Bitrate reduction) over the AWMVM for 1080p irregular motion test sequences, but the overall gain (0.78% BD-Bitrate reduction) is marginal since it cannot fully use the existing information of a pre-encoded bitstream to find candidate MVs, as discussed in Sect. 2. In contrast with the KMCMV, the proposed CEMVR achieves a significant gain over the AWMVM not only for 1080p irregular motion test sequences but also for overall test sequences. The gains of BD-Bitrate reduction are 21.31% and 14.27%, respectively. Moreover, compared with KMCMV, the proposed CEMVR still captures a lot of BD-Bitrate reduction which is 13.49% on the average of all test sequences. Specifically, a BD-Bitrate reduction of the proposed CEMVR about the factory sequence is far superior to that of the KMCMV, which is up to 25%. This is the experimental evidence of getting significantly better RD performance in test sequences with irregular motion as discussed in Sect. 2 because a factory sequence has many objects that move in random or complex ways.

From the literature, we have found that the AWMVM proposed by Tan *et al.* [4] is the best one considered both accuracy and complexity. The AWMVM reduces on average 77% of the computational time but suffers RD performance degradation for irregular motion video sequences and fast regular motion video sequences. We have compared the computational complexity of our algorithm with this well-known MVRE algorithm. We have also compared our result with the recent algorithm (KMCMV).The proposed CEMVR technique reduces 71% ∼ 88% of the computational complexity on average compared to the FDFE, whereas the AWMVM and KMCMV techniques reduce 71% ∼ 88% and 64% ∼ 85% compared to the FDFE, respectively. The significant gain of BD-rates is from the efforts of the proposed cost model because coding gain is preserved in any combinations of the local search range and multiple candidates compared with two existing schemes. The complexity reduction is depend on the effort of the local search range. The number of multiple candidates is the fine tuning factor for a balance of coding gain and complexity reduction. We decided that the local search range and the number of multiple candidates are set at 1 and 4.

5. Conclusions

In this paper, we proposed a novel MV cost estimator for re-estimating the MVs in a H.264/AVC transcoder. Since our proposed cost estimator can fully utilize the existing information of a pre-coded bitstream, we can estimate the cost of any given MV and select reliable MVs as candidate MVs by comparing their estimated cost in the MVRE process. As a result, compared to representative known schemes, the transcoder employing the MVRE algorithm by the proposed cost estimator leads to a significant quality gain without a critical complexity increment, especially for complex motion videos. Furthermore, our proposed cost estimator can easily be applied to any type of MVRE scheme if the bitstream is pre-coded.

References

[1] "3GPP; Technical Specification Group Services and System Aspects; Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)," no.26.247, 2010.
[2] "Dynamic adaptive streaming over http (dash)," FCD 23001-6, ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Daegu, Jan. 2011.
[3] I. Sodagar, "The mpeg-dash standard for multimedia streaming over the internet," MultiMedia, IEEE, vol.18, no.4, pp.62–67, April 2011.
[4] Y.-P. Tan and H. Sun, "Fast motion re-estimation for arbitrary downsizing video transcoding using h.264/avc standard," IEEE Transactions on Consumer Electronics, vol.50, no.3, pp.887–894, 2004.
[5] K. Kim, S. Hong, and Y. Choe, "Efficient motion re-estimation method based on k-means clustering for spatial resolution reduction transcoding," Picture Coding Symposium (PCS), 2012, pp.221–224, May 2012.
[6] K. Suehring, "H.264/avc reference software."
[7] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16/Q.6 VCEG Doc., VCEG-M33, April 2001.