

RESEARCH

Open Access

Region-of-interest determination and bit-rate conversion for H.264 video transcoding

Shu-Fen Huang, Mei-Juan Chen*, Kuang-Han Tai and Mian-Shiuan Li

Abstract

This paper presents a video bit-rate transcoder for baseline profile in H.264/AVC standard to fit the available channel bandwidth for the client when transmitting video bit-streams via communication channels. To maintain visual quality for low bit-rate video efficiently, this study analyzes the **decoded information** in the transcoder and proposes a **Bayesian theorem-based region-of-interest (ROI)** determination algorithm. In addition, a **curve fitting** scheme is employed to find the **models of video bit-rate conversion**. The transcoded video will conform to the target bit-rate by re-quantization according to our proposed models. After integrating the ROI detection method and the bit-rate transcoding models, the ROI-based transcoder allocates more coding bits to ROI regions and reduces the complexity of the re-encoding procedure for non-ROI regions. Hence, it not only keeps the coding quality but improves the efficiency of the video transcoding for low target bit-rates and makes the real-time transcoding more practical. Experimental results show that the proposed framework gets significantly better visual quality.

Keywords: Video transcoding, Region-of-interest, Re-quantization, Motion vector

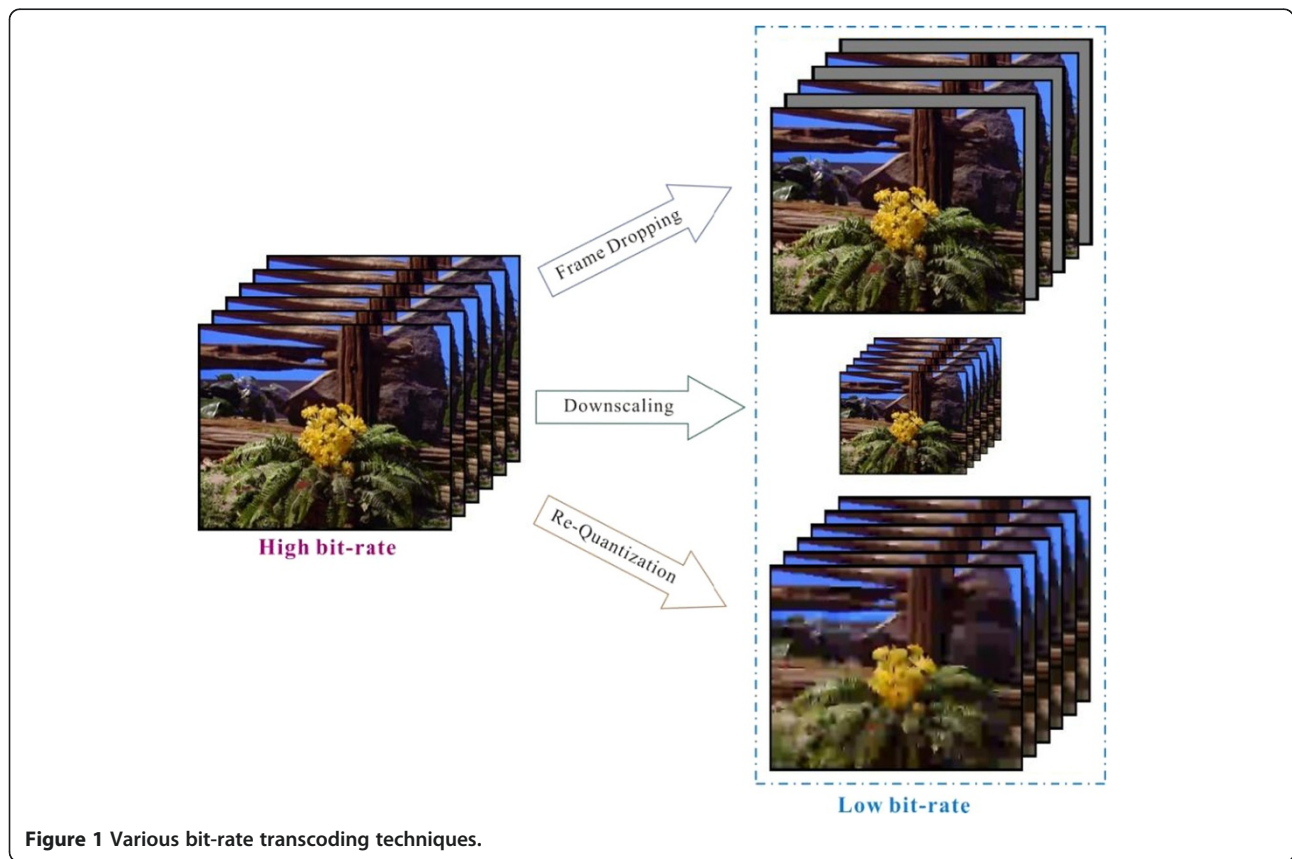
1. Introduction

Image and video have been widely used in our lives for many applications such as video conferencing, Internet protocol television, video phone and video surveillance, etc. To achieve these applications, an efficient video compression technology is required. H.264/AVC [1] video codec is one of the most popular video coding standards recently, and it overtakes MPEG-2 and other standards. After the coding process, video bit-streams are transmitted to a variety of communication channels and consumer devices. Therefore, a problem of heterogeneous communication channel can be encountered. Fortunately, the so-called transcoding technology, which can translate the data format, the bit-rate, the spatial resolution, the frame rate, and the coding standard, provides a way to solve the heterogeneous problems. The cascaded pixel-domain transcoder (CPDT) [2] can fully decode the bit-stream and re-encode it according to the required data format. To re-encode the video bit-streams efficiently, the transcoder is usually designed to make good use of the partially decoded information. The

work in [3] establishes the relation of residual and coding bits using Lagrangian optimization; therefore, the appropriate motion vector at the target bit-rate could be obtained. In [4], the Bayesian theorem and Markov chain are utilized to model the probability of modes for the transcoding between H.264/AVC and SVC. The work in [5] adjusts the discrete cosine transform coefficients to decrease the ones needed to be inverted and also to create a new method to measure the distortion in this transformation. Liu and Yoo [6] propose a fast mode mapping method to reduce the power consumption for MPEG-2 to H.264/AVC transcoder via mode type and macroblock (MB) activity. Among all the video transcoder systems, it is especially worthy to focus on bit-rate transcoding, also known as transrating, because of the limited transmission bandwidth. As shown in Figure 1, the well-known methods to achieve bit-rate transcoding include downsizing frame resolution, frame dropping, and re-quantization of transform coefficients. Tang et al. [7] analyze the corresponding rate-quantization (R-Q) models for the different spatial reduction ratios (SRR). Hence, they can evaluate the proper frame size to adjust the target bit-rate. In [8], a parabolic motion vector re-estimation algorithm is proposed to efficiently predict

* Correspondence: cmj@mail.ndhu.edu.tw

Department of Electrical Engineering, National Dong Hwa University, No. 1, Sec. 2, Da Hsueh Rd, Shoufeng, Hualien 97401, Taiwan



the motion vectors in downscaling video for downsizing transcoder. Kapotas and Skodras [9] drop video frames in a compressed domain to attain suitable frame rate when the bandwidth of the network is limited. The work in [10] considers temporal and spatial complexities for arbitrary frame rate transcoding. Yeh et al. [11] analyze visual complexity and temporal coherence to propose a frame selection algorithm for temporal video transcoder. To meet the restricted network bandwidth, Kwon et al. [12] estimate the new quantization step (Qstep) size by using only one control parameter, which employs a simple rate control algorithm.

In the applications of video processing, perceptual video coding [13] has obtained more and more attention due to human visual perception. Especially for the low bit-rate situation, region-of-interest (ROI)-based video coding can maintain better visual quality compared to the non-ROI video coding. According to the properties of the human visual system (HVS), the ROI areas are

usually defined by skin color, contrast, video content, objects, and motion information. Liu et al. [14] extract moving objects in the compressed domain by considering global motion compensation, correspondence matrix, and temporal tracking. Mak and Cham [15] identify the background motion and then analyze the decoded motion data to complete the video object segmentation. Chi et al. [16] determine ROI regions through a visual rhythm feature as user attention model. Liu et al. [17] investigate the event detection in a compressed domain automatically. After motion trajectories are extracted, they use prediction residuals to improve the robustness of the proposed video activity detection strategy. De Bruyne et al. [18] estimate the reliable motion information to enhance the ROI detection based on the forward/backward projection of motion vectors (MVs) and similarity correlations. In addition, Poppe et al. [19] present a

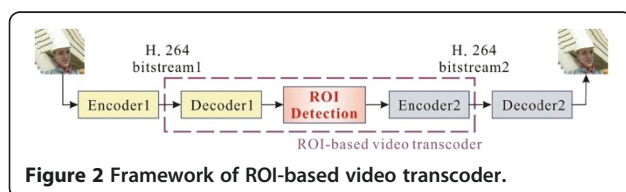


Table 1 Probability of motion intensity

MI (%)	0	$\leq 0.5 MI_{avg}$	$\leq MI_{avg}$	$\leq 1.5 MI_{avg}$	$> 1.5 MI_{avg}$
$P(MI)$	6.25	49.52	19.70	5.43	19.10
$P(MI ROI)$	7.45	2.95	11.04	14.12	64.44
$P(ROI MI)$	29.31	1.46	13.78	63.94	82.96
$P_{real}(ROI MI)$	29.29	1.47	13.78	63.95	82.98

$P(ROI) = 24.59\%$.

Table 2 Probability of skin color pixels for one MB

Number of skin color pixels (%)	0 to 51	52 to 102	103 to 153	154 to 204	205 to 256
$P(\text{skin})$	78.79	6.53	2.49	1.86	10.32
$P(\text{skin} \text{ROI})$	27.60	17.33	8.22	6.80	40.05
$P(\text{ROI} \text{skin})$	8.61	65.26	81.18	89.90	95.43
$P_{\text{real}}(\text{ROI} \text{skin})$	8.61	65.22	81.01	89.83	95.41

$P(\text{ROI}) = 24.59\%$.

fast moving object detection algorithm that relies on the structure of the encoded bit-stream at the syntax level. The complicated entropy decoding to get motion information is unnecessary; thus, their method achieves high execution speeds for video surveillance applications.

Combined with a video transcoder and the idea of ROI, a compressed domain transcoder from MPEG-2 to H.264/AVC is proposed in [20] to reduce the computation complexity. The ROI areas are re-encoded by a closed loop transcoder for high video quality. The regions without motion vectors are defined as non-ROI and re-encoded by an open loop transcoder to speed up the transcoding process. In multipoint video conferencing, Li et al. [21] present a spatial adaptation ROI system to solve the limited display sizes. Kwon and Lee [22] propose an ROI transcoder by setting different quantization parameters (QPs) to the coded MBs via the subjective importance of different slice distributions. The center of the whole frame is the most important, and the more outer the center is, the less important it is. It can be described by a quadratic function. Although the research topic is similar to our goal, the ROI region is not precise enough in [22], since only the center region is defined as the ROI area. It is therefore worthy to pay more attention to the H.264/AVC ROI transcoding to reduce the complexity of the transcoder especially at the low bit-rate situation.

The proposed ROI-based video transcoder is shown in Figure 2. In this paper, by utilizing the decoded information from the front-decoder in the transcoder, we take the motion intensity and skin color information into consideration and then propose a Bayesian theorem-based ROI detection method. After estimating the ROI macroblocks (MBs), the models of bit-rate conversion are proposed for bit-rate transcoding. Our proposed models can transcode video bit-stream to the target bit-rate by the re-quantization method and allocate more coding bits to more important areas. Furthermore, a closed-loop video transcoding system, which saves re-encoding time by reducing the processing of motion estimation in non-ROI MBs, is also proposed.

The rest of this paper is organized as follows. Proposed ROI determination method based on Bayesian theorem describes the details of our proposed ROI detection

algorithm. Proposed re-quantization models analyze the features of bit-rate conversion for baseline profile in H.264/AVC and the re-quantization models for I frames and P frames, respectively. The frameworks of our proposed ROI-based bit-rate transcoder are described in the frameworks of the proposed H.264/AVC bit-rate transcoder. Simulation results show the experimental results, and the conclusions are made finally.

2. Proposed ROI determination method based on Bayesian theorem

To automatically decide ROI, we employ the statistical technique to model the ROI MBs of common video content in [23]. First of all, user-defined visual attention maps are adopted to analyze the probability distribution of the

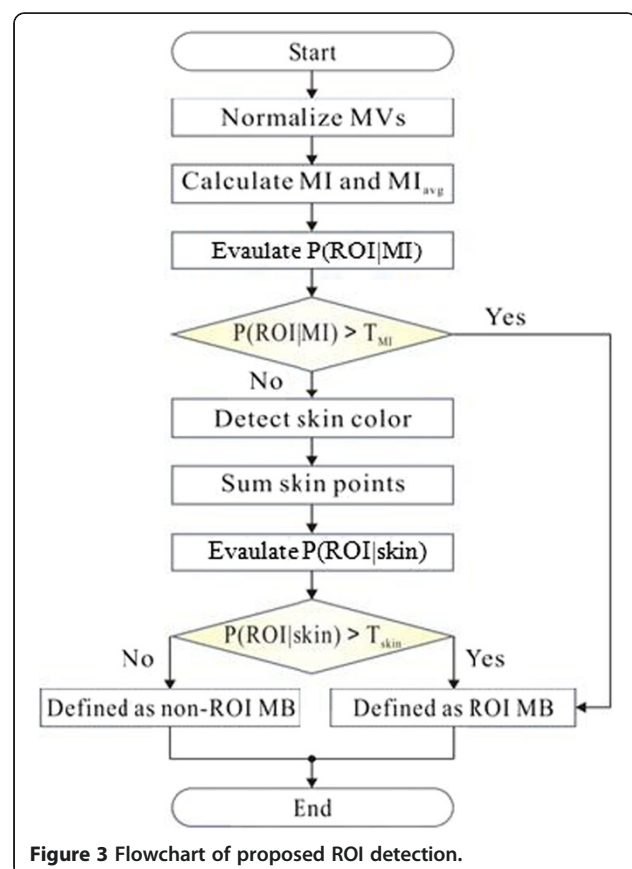


Figure 3 Flowchart of proposed ROI detection.

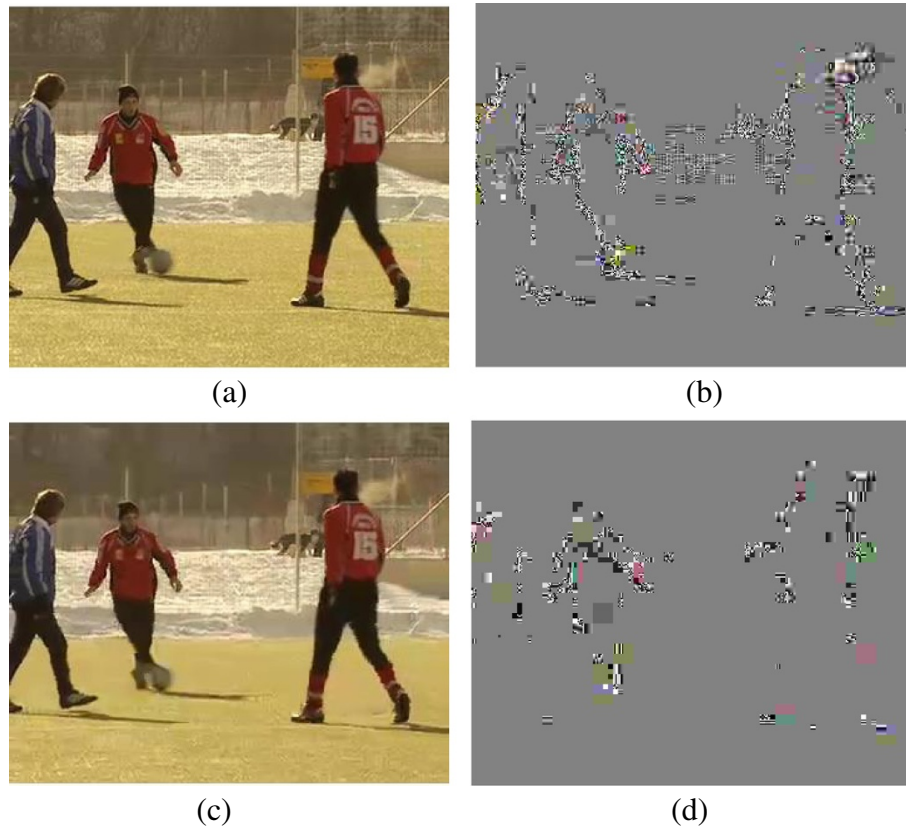


Figure 4 The coding results for the 10th frame of CIF Soccer sequence in different bit-rates. (a) Reconstructed frame in 700 Kbps (b) Residual in 700 Kbps (c) Reconstructed frame in 200 Kbps, and (d) Residual in 200 Kbps.

ROI. We test two sequences including Foreman and Soccer with common intermediate format (CIF 352×288) frame size and then analyze the relationship between the decoded data and ROI distribution. The auxiliary information includes motion intensity and skin color information. According to Bayesian theorem shown in Equations 1 and 2, we can update the ROI probability in the light of auxiliary data.

$$P(\text{ROI}|\text{MI}) = \frac{P(\text{MI}|\text{ROI})P(\text{ROI})}{P(\text{MI})} \quad (1)$$

$$P(\text{ROI}|\text{skin}) = \frac{P(\text{skin}|\text{ROI})P(\text{ROI})}{P(\text{skin})} \quad (2)$$

where $P(\text{ROI})$ means the prior probability of ROI before considering the auxiliary information; $P(\text{MI})$

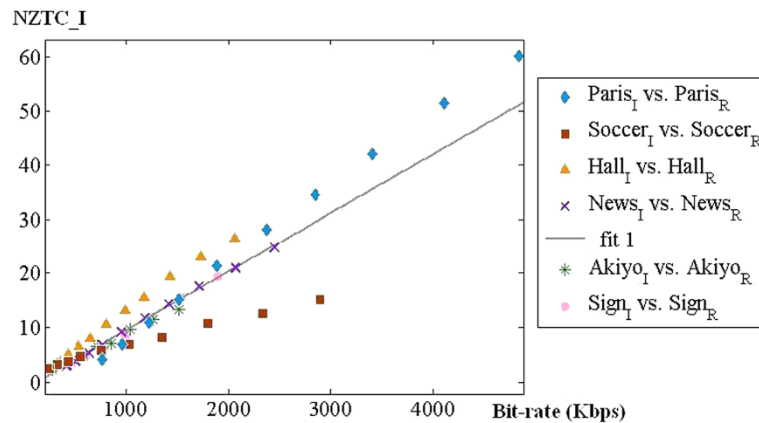


Figure 5 Relationship of bit-rate and NZTC for I frame.

Table 3 Coefficients of Equation 6 for the simulated sequences

	a_1	b_1	R^2
Akiyo	0.0089	0.1301	0.9960
Hall	0.0136	-0.6348	0.9968
News	0.0108	-1.3200	0.9985
Paris	0.0139	-5.7340	0.9986
Sign	0.0104	-0.9668	0.9860
Soccer	0.0047	1.9100	0.9948

Table 4 Coefficients of Equation 7 for the simulated sequences

	a_2	b_2	c_2	d_2	R^2
Akiyo	25.96	-0.21	28.94	-1.15E-02	0.9944
Hall	11.37	-0.14	39.20	-1.55E-02	0.9988
News	10.44	-0.18	39.46	-1.68E-02	0.9985
Paris	5.35	-0.15	42.55	-8.20E-03	0.9998
Sign	58.39	-0.75	37.10	-1.82E-02	0.9966
Soccer	23.39	-0.19	30.43	-1.30E-02	0.9984

and $P(\text{skin})$ indicate the marginal probabilities of motion intensity (MI) and skin color data, respectively. $P(\text{MI}|\text{ROI})$ and $P(\text{skin}|\text{ROI})$ represent the conditional probabilities of MI and skin color data based on ROI distribution, respectively; $P(\text{ROI}|\text{MI})$ and $P(\text{ROI}|\text{skin})$ represent the posterior probabilities of ROI given the auxiliary MI and skin color information, respectively.

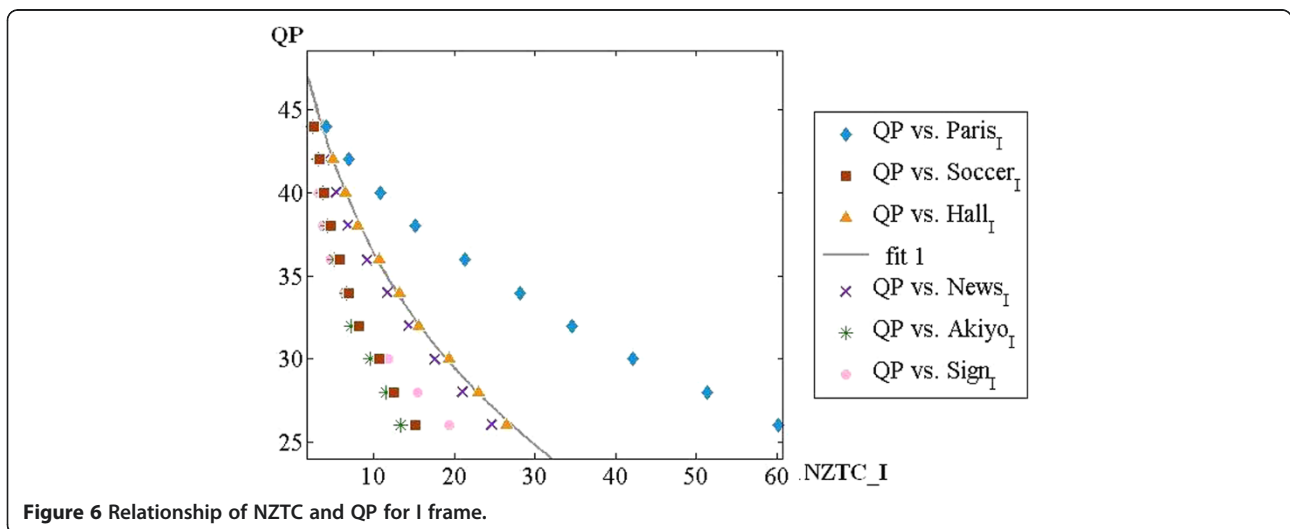
It is well-known that human eyes always pay more attention to the moving regions than the background. Because of the variable block sizes adopted in H.264/AVC, we perform motion vector normalization into 4×4 block size to ensure fairness and then calculate the motion intensity, $\text{MI}(i,j)$ for each MB located in (i,j) by Equation 3. To test the relationship between MI value and ROI, the average motion intensity, MI_{avg} , for the whole frame is defined in Equation 4:

$$\text{MI}(i,j) = \frac{1}{16} \sum_{r=0}^3 \sum_{s=0}^3 \sqrt{\text{mvx}(r,s)^2 + \text{mvy}(r,s)^2} \quad (3)$$

$$\text{MI}_{\text{avg}} = \frac{1}{h \times w} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \text{MI}(i,j) \quad (4)$$

where h and w denote the number of MBs in rows and columns of a frame, respectively; mvx and mvy are the components of motion vectors for horizontal and vertical directions, respectively. It is obvious that conditional probability increases with the larger value of motion intensity in Table 1.

However, there are still some ROI regions with lower motion. To segment ROIs accurately, we find other features more than just motion information. The human face also attracts the user's attention, and the skin color detection is used as pre-processing to efficiently extract the face region. We implement the method in [24] to determine skin samples by using Cb and Cr values directly. Every pixel will be labeled as skin color or not depending on whether it meets the upper and lower quadratic functions in the CbCr plane or not. Then, we discuss about the relationship between ROI and the skin color points in one MB. This additional auxiliary information can



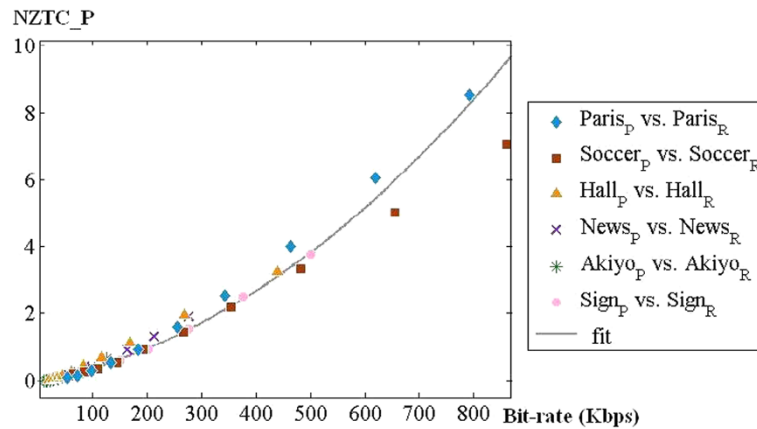


Figure 7 Relationship of bit-rate and NZTC for P frame.

effectively solve the problem of ROIs with lower motion activity. We summarize the real and estimated probabilities of ROI by Equations 1 and 2 given the motion intensity and skin color information in Tables 1 and 2, respectively. The estimated probabilities by Equations 1 and 2 are close to the corresponding real probabilities. The flowchart of the proposed ROI detection method is shown in Figure 3. Both the motion intensity and skin color information are considered for segmenting ROI regions.

3. Proposed re-quantization models

3.1. Analysis of video bit-rate conversion

In H.264/AVC standard, the total coding bits required for one frame contain the bits used to encode header information, motion data, and video residual as shown in Equation 5:

$$R_{\text{total}} = R_{\text{header}} + R_{\text{motion}} + R_{\text{residual}} \quad (5)$$

Header information contains the coding mode, reference frame number, transform size, etc. Compared to the whole frame coding bits, the bits used to code the header information are negligible. To improve the coding performance, H.264/AVC provides a motion vector prediction (MVP) strategy. After conducting motion

estimation, the real motion vector must be subtracted by the predicted motion vector, and then only the motion vector difference is encoded. Owing to the MVP method, the bits used to code the motion data are few. Consequently, **the bits used to code the video residual are the principal data in the video bit-stream.** As shown in Figure 4, the higher the video bit-rate, the more residual data are retained.

3.2. Proposed re-quantization models for I and P frames

The H.264/AVC encoder performs intra/interpredictions to get the residual and then transforms this data into frequency domain as transform coefficients. Besides, QP relates to quantization operation and also influences residual data. Therefore, we take the relationship of the bit-rate, the number of non-zero transform coefficients (NZTC), and the QP into consideration to find the bit-rate conversion models.

We test six sequences in CIF (352 × 288) resolution with various characteristics including Akiyo, Hall Monitor, News, Paris, Sign Irene, and Soccer. The total number of coding frames is 100, and the group of pictures (GOP) is 15 with 1 I picture followed by 14 P pictures. Curve fitting is utilized to model the relationship

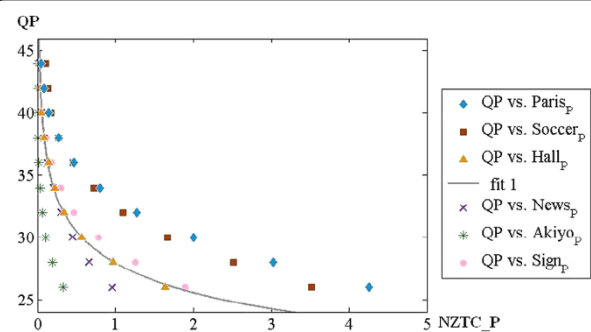


Figure 8 Relationship of NZTC and QP for P frame.

Table 5 Coefficients of Equation 8 for the simulated sequences

	a_3	b_3	c_3	R^2
Akiyo	3.18E-05	0.0013	-0.0237	0.9993
Hall	1.11E-06	0.0073	-0.1490	0.9992
News	1.13E-05	0.0042	-0.0910	0.9994
Paris	6.01E-06	0.0065	-0.3842	0.9994
Sign	9.34E-06	0.0030	-0.0325	0.9984
Soccer	2.99E-06	0.0060	-0.3073	0.9992

Table 6 Coefficients of Equation 9 for the simulated sequences

between bit-rate, NZTC and QP for I frame and P frame, respectively. The accuracy of the fitting result is evaluated via R^2 , also known as the coefficient of determination in regression analysis.

Figure 5 demonstrates the relationship of video bit-rate and NZTC for I frames. It is obvious that all of these curves have similar trends, and we can utilize Equation 6 to define the linear model. The coefficients of the corresponding video sequences are shown in Table 3.

Figure 6 exhibits the relationship of NZTC and QP data for I frames. Equation 7 illustrates the mathematical model, and Table 4 shows the coefficients of the simulated sequences. Thus, after estimating NZTC, we can get the QP value for re-quantization to reduce video bitrate by Equation 7:

3.2.2. P frame models

similar and can be fitted accurately by Equation 8. The coefficients for different video sequences are shown in Table 5.

Figure 8 shows the relationship of NZTC and QP data for P frames. Equation 9 expounds the re-quantization model so that we can accomplish the video bit-rate transcoding effectively. Table 6 shows the coefficients of our simulated sequences.

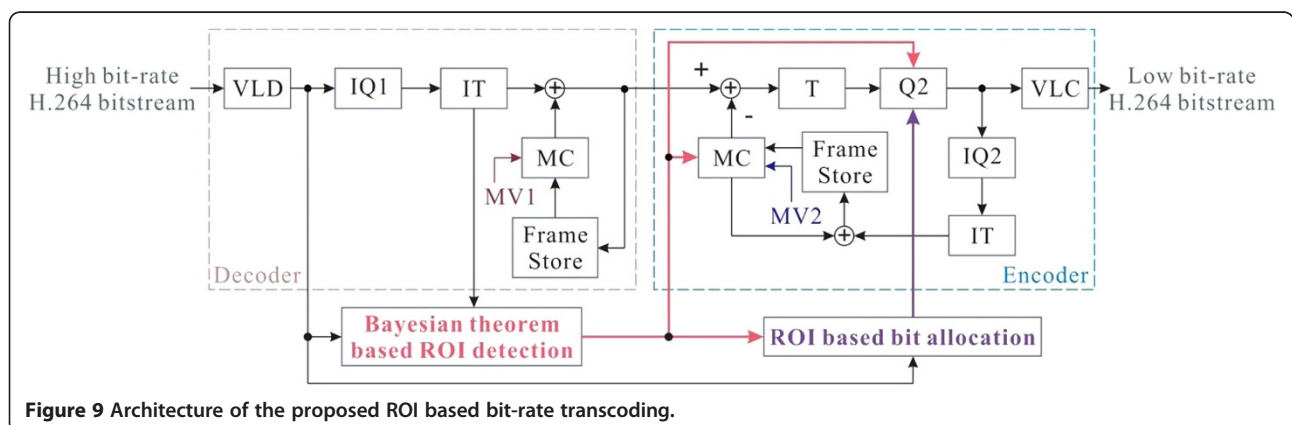
4. Frameworks of the proposed H.264/AVC bit-rate transcoder

Equations 10 to 13 are proposed to achieve the whole processing of video bit-rate transcoding. In I frames, we can get the original number of non-zero transform coefficients ($NZTC_{I,ori}$), original bit-rate ($R_{I,ori}$), and original quantization parameter ($QP_{I,ori}$) from the front decoder. To meet the target bit-rate ($R_{I,tar}$), we propose Equation 11 to estimate target QP for re-quantization. The $NZTC_{I,tar}$ indicates the target number of non-zero transform coefficients shown in Equation 10. $NZTC_{P,tar}$, $NZTC_{P,ori}$, $R_{P,tar}$, $R_{P,ori}$, $QP_{P,tar}$ and $QP_{P,ori}$ are for P frames.

$$Q_{P_{1,\text{tar}}} = \frac{a_2 \exp(b_2 \text{NZTC}_{1,\text{tar}}) + c_2 \exp(d_2 \text{NZTC}_{1,\text{tar}})}{a_2 \exp(b_2 \text{NZTC}_{1,\text{ori}}) + c_2 \exp(d_2 \text{NZTC}_{1,\text{ori}})} Q_{P_{1,\text{ori}}} \quad (11)$$

$$\text{NZTC}_{\text{P,tar}} = \frac{a_3 R_{\text{P,tar}}^2 + b_3 R_{\text{P,tar}} + c_3}{a_3 R_{\text{P,ori}}^2 + b_3 R_{\text{P,ori}} + c_3} \text{NZTC}_{\text{P,ori}} \quad (12)$$

$$\text{QP}_{\text{P,tar}} = \frac{a_4 \text{NZTC}_{\text{P,tar}}^{b_4}}{a_4 \text{NZTC}_{\text{P,ori}}^{b_4}} \text{QP}_{\text{P,ori}} \quad (13)$$



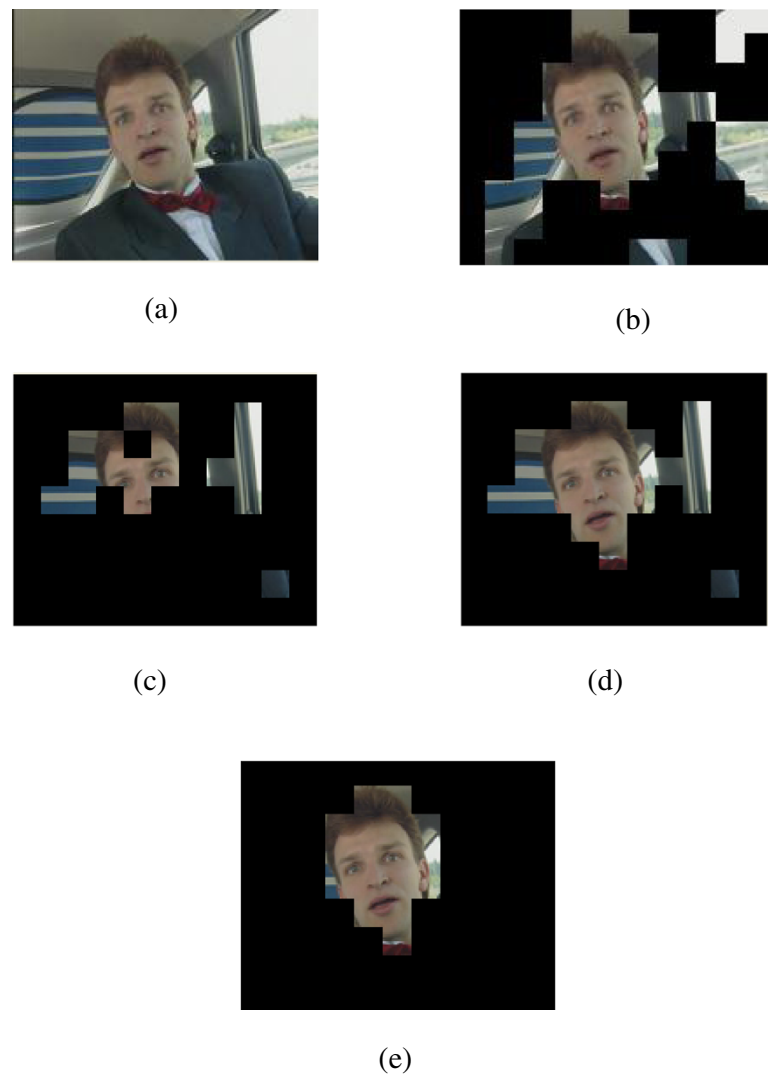


Figure 10 Decoded and segmentation results of the 49th frame in Carphone sequence by various conditions. (a) Decoded result of the 49th frame in Carphone Sequence (b) Bruyne’s method [18] (c) Using only motion information by setting 0.75 to T_{ML} (d) Setting 0.75 and 0.85 to T_{ML} and T_{skin} , respectively and (e) Setting 0.85 and 0.85 to T_{ML} and T_{skin} , respectively.

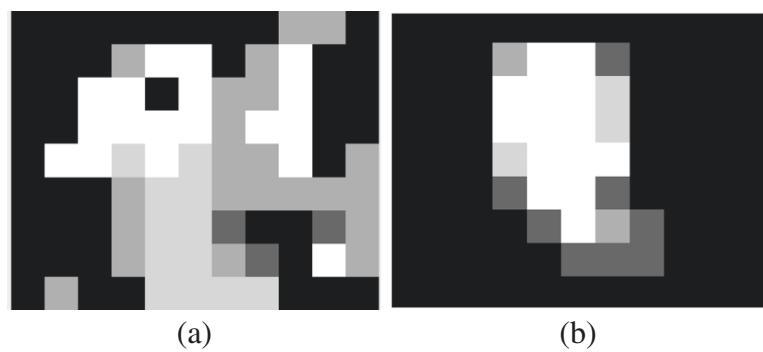







Figure 11 MI and skin data. (a) MI data (b) Skin color data of the 49th frame in Carphone sequence.

Table 7 Definition of gray color in Figures 8 and 9

Color	MI data	Number of skin points
	$MI = 0$	0 to 51
	$0 \leq MI \leq 0.5 MI_{avg}$	52 to 102
	$0.5 MI_{avg} \leq MI \leq MI_{avg}$	103 to 153
	$MI_{avg} \leq MI \leq 1.5 MI_{avg}$	154 to 204
	$MI > 1.5 MI_{avg}$	205 to 256

Firstly, we get the original NZTC and the original bit-rate from the front decoder, and then, Equations 10 and 12 are adopted to figure the target NZTC data in the corresponding target bit-rate for I pictures and P pictures, respectively. Second, the target QP required for each MB in the rear encoder are calculated by Equations 11 and 13 for I pictures and P pictures, respectively. Therefore, the high bit-rate bit-stream can be transcoded to a low bit-rate via re-quantization method. Figure 9 refers to the architecture of our proposed bit-rate transcoding.

4.2. Training procedure for model parameters

In fact, the features of various video sequences are distinctive especially for the different applications. Accordingly, the training procedure of the input sequence for its model parameters is necessary before utilizing these bit-rate conversion models. We set I frame for the first coding frame and P frame for the second frame. QP data is set to 25, 30, 35, and 40 for pre-processing coding. A curve fitting scheme is employed to find model parameters. In spite of the training procedure making a little overhead, this promotes accurate bit-rate conversion models.

4.3. ROI-based video bit-rate transcoder

For low bit-rate transmission of video bit-streams, the general re-quantization method diminishes the visual quality of the entire frame significantly. To solve this problem, more coding bits should be allocated in the areas with more users' interest. Hence, an automatic ROI segmentation algorithm is needed. Combined with the ROI segmentation mentioned in section 2, we propose a ROI-based bit-rate transcoder to maintain video quality in ROI MBs.

Equations 14 to 17 are proposed to allocate the limited coding bits according to the importance of each MB. Firstly, Equation 14 calculates the estimated total coding bits for the current frame after transcoding. R_{ori} , R_{tar}

Table 8 Bit-rate (Kbps) comparison

Sequence	Target bit-rate	CPDT	Proposed ROI-based transcoder
Akiyo	220	220.15	235.17
Foreman	900	900.30	899.91
News	390	390.10	401.92

Table 9 Total coding time (T) and time saving (TS) comparison

Sequence	Target bit-rate	CPDT (sec)	Proposed ROI-based transcoder	
			T (s)	TS (%)
Akiyo	220	401.03	200.71	49.95
Foreman	900	405.02	160.02	60.49
News	390	403.05	211.54	47.52

and Bit_{ori} represent the original bit-rate, target bit-rate, and original bits for the current frame, respectively. Then, the coding bits are apportioned to MB located in (i, j) by Equation 15; w_{total} means the weighting sum within the current frame, and $w(i, j)$ is defined as the weighting value for MB located in (i, j) . We define the weighting value as 1 if this MB is in the ROI region; otherwise, the weighting value is set to the average of conditional probabilities. After allocating coding bits for each MB by Equation 15, we multiply the $Bit(i, j)$ with $(w \times h \times 30)$ as the bit-rate for Equations 6 and 8 during the re-quantization process. Then, we adopt Equations 7 and 9 to evaluate target QP for re-quantization in the rear encoder and then transcode the video as target bit-rate. In order to reduce the coding complexity of the transcoding processing, we reuse the decoded modes for the non-ROI MBs and employ the full modes in the ROI MBs to maintain the visual quality. In addition, the H.264/AVC coding standard supports SKIP mode to improve compression efficiency, and we retain this advantage for the non-ROI areas. In short, the re-encoding time for our proposed ROI-based transcoder gets speeded up by ignoring the motion estimation and mode decision for non-ROI MBs.

$$Bit_{total} = \frac{R_{tar}}{R_{ori}} Bit_{ori} \quad (14)$$

$$Bit(i, j) = \frac{w(i, j)}{w_{total}} Bit_{total} \quad (15)$$

$$w_{total} = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} w(i, j) \quad (16)$$

$$w(i, j) = \begin{cases} 1, & \text{if } MB(i, j) \in ROI \\ \left(P_{i,j}(ROI|MI) + P_{i,j}(ROI|skin) \right) / 2, & \text{otherwise} \end{cases} \quad (17)$$

Figure 9 illustrates the application of the proposed ROI determination method. It can be used in the ROI-based video transcoder to automatically extract ROI maps. The video quality in the ROI region can be better than that of the non-ROI regions for setting the different coding parameters in the encoder in Figure 9.

Table 10 PSNR (dB) comparison for ROI areas

Sequence	Target bit-rate	Frame number	CPDT	Proposed ROI-based transcoder
Akiyo	220	32nd	35.71	37.06
Foreman	900	72nd	42.45	45.81
News	390	65th	35.58	37.78

5. Simulation results

5.1. Setup of simulation

The proposed video bit-rate transcoder is implemented on the JM15.1 [25] software. The computer used for our simulation has 3 GHz CPU and 3 GB RAM. The total coding frame number is set to 100 for each sequence, and the GOP is 15 with 1 I frame followed by 14 P frames. Pre-encoded bit-streams are encoded by 30 frames per second and 1 megabits per second (Mbps). The search range is set to ± 32 , and the number of reference frames is set to one for motion estimation.

The 8×8 transform size is disabled. We adopt the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [26] for the quality evaluation. The SSIM value used to measure the structural distortion is between -1 and 1 , and it is a popular method owing to its high correlation with HVS.

5.2. Simulation results of ROI detection

The decoded 49th frame in the Carphone sequence and the ROI detection result by [18] are shown in Figure 10a, b. Figure 10c demonstrates the ROI extraction without skin color data. This method could not extract the human face accurately. Figure 10d improves the segmentation results according to the skin color information. The different thresholds for T_{MI} and T_{skin} are shown in Figure 10e.

Figure 11a, b illustrates MI data and skin data, respectively. The definitions of gray color are shown in Table 7. From Figure 11a, we can observe that the motion information could be inconsistent in the whole frame.

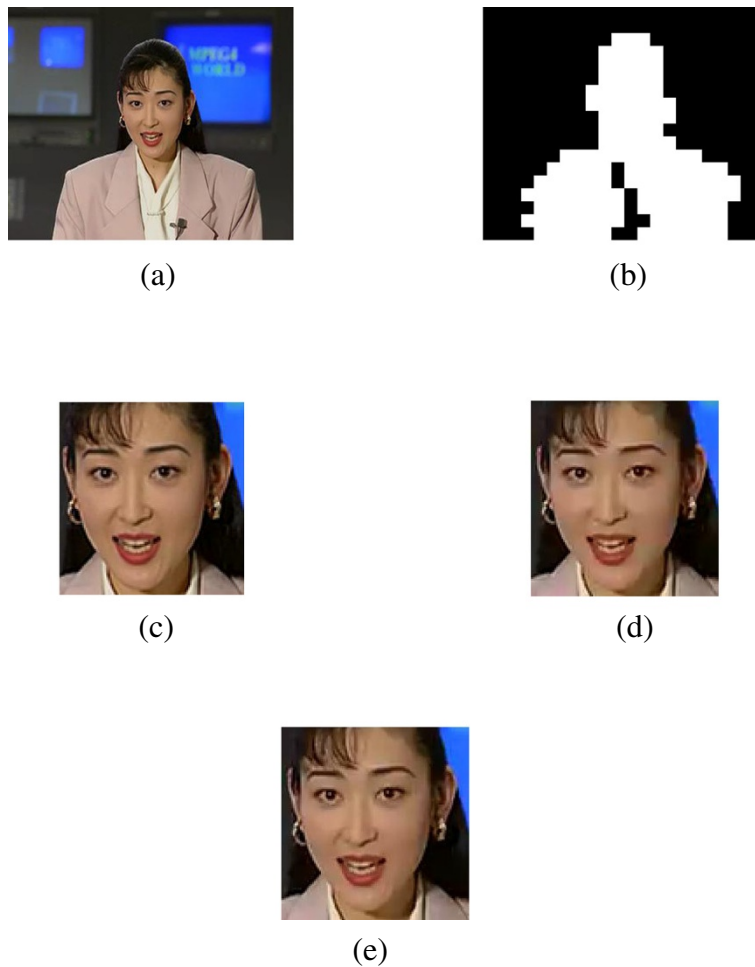


Figure 12 The subjective comparison for the 32nd frame of Akiyo sequence with bit-rate of 220 Kbps. (a) Pre-encoded frame. The pre-encoded bitstream with bit-rate 1Mbps **(b)** ROI map detected by [23] **(c)** Partial frame of pre-encoded sequence **(d)** Partial frame of CPDT method. $PSNR_{ROI}$: 35.71dB, $SSIM_{ROI}$: 0.94, and **(e)** Partial frame of ROI-based transcoder. $PSNR_{ROI}$: 37.06 dB, $SSIM_{ROI}$: 0.96.

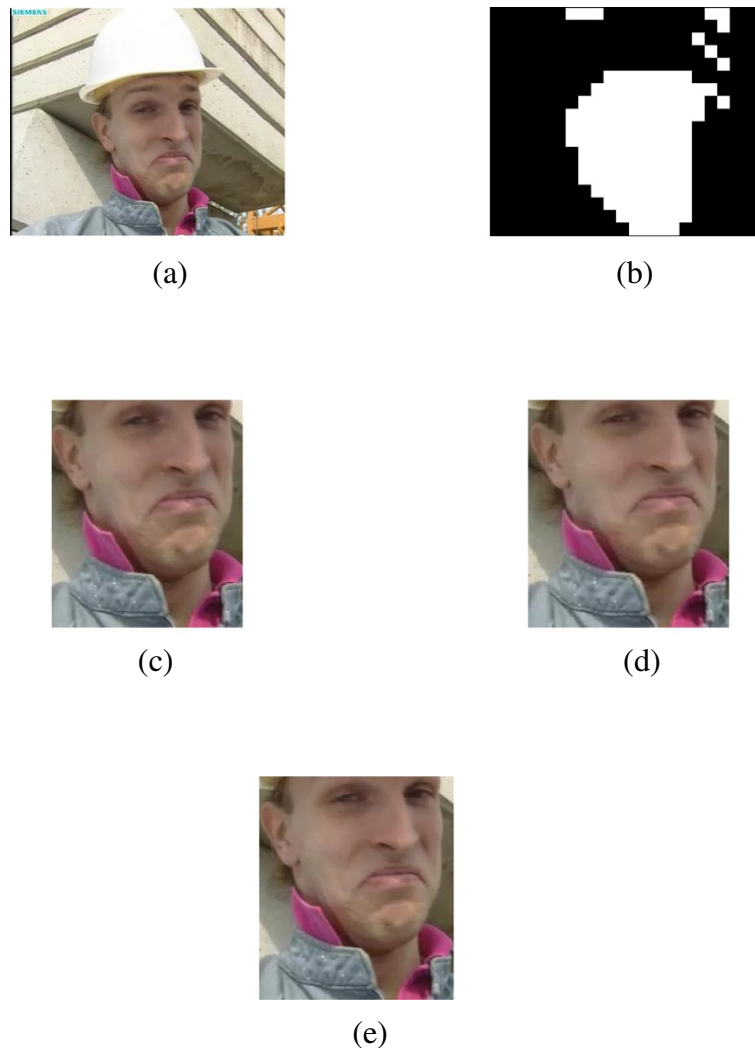


Figure 13 The subjective comparison for the 72nd frame of Foreman sequence with bit-rate 900 Kbps. (a) Pre-encoded frame. The pre-encoded bit-stream with bit-rate 1Mbps (b) ROI map detected by [23] (c) Partial frame of pre-encoded sequence (d) Partial frame of CPDT method, $PSNR_{ROI}$: 42.45dB, $SSIM_{ROI}$: 0.97, and (e) Partial frame of ROI-based transcoder. $PSNR_{ROI}$: 45.81dB, $SSIM_{ROI}$: 0.98.

Therefore, the ROI determination result will be more precise with both MI information and skin data.

5.3. Simulation results of the proposed ROI-based bit-rate transcoding

We test three sequences including Akiyo, Foreman, and News in CIF (352×288) format. Tables 8, 9, and 10 express the experimental results of the bit-rate data, coding time, and ROI PSNR comparison, respectively. We transcode simulated sequences from 1 Mbps to 220 Kbps, 900 Kbps, and 390 Kbps. The CPDT and proposed ROI-based transcoder represent the results produced by the cascaded pixel domain transcoder and the proposed ROI-based transcoder, respectively. From Tables 8, 9, and 10, it is obvious that the CPDT method controls the bit-rate

accurately but consumes much coding time because of fully decoding and encoding the video sequences. The proposed ROI-based transcoder reduces re-encoding time significantly and saves computational time at least by 47.52%. The proposed method also provides better visual quality in ROI regions ranging from 1.35 to 3.36 dB.

Figures 12, 13, and 14 demonstrate the comparison results of the subjective qualities for Akiyo, Foreman, and News sequences, respectively. We can observe that our ROI-based transcoder outperforms the CPDT method in both PSNR and SSIM measurements and the subjective quality. The proposed ROI-based scheme takes advantage of the human visual system and also reduces the computational complexity of the H.264 video transcoder dramatically.

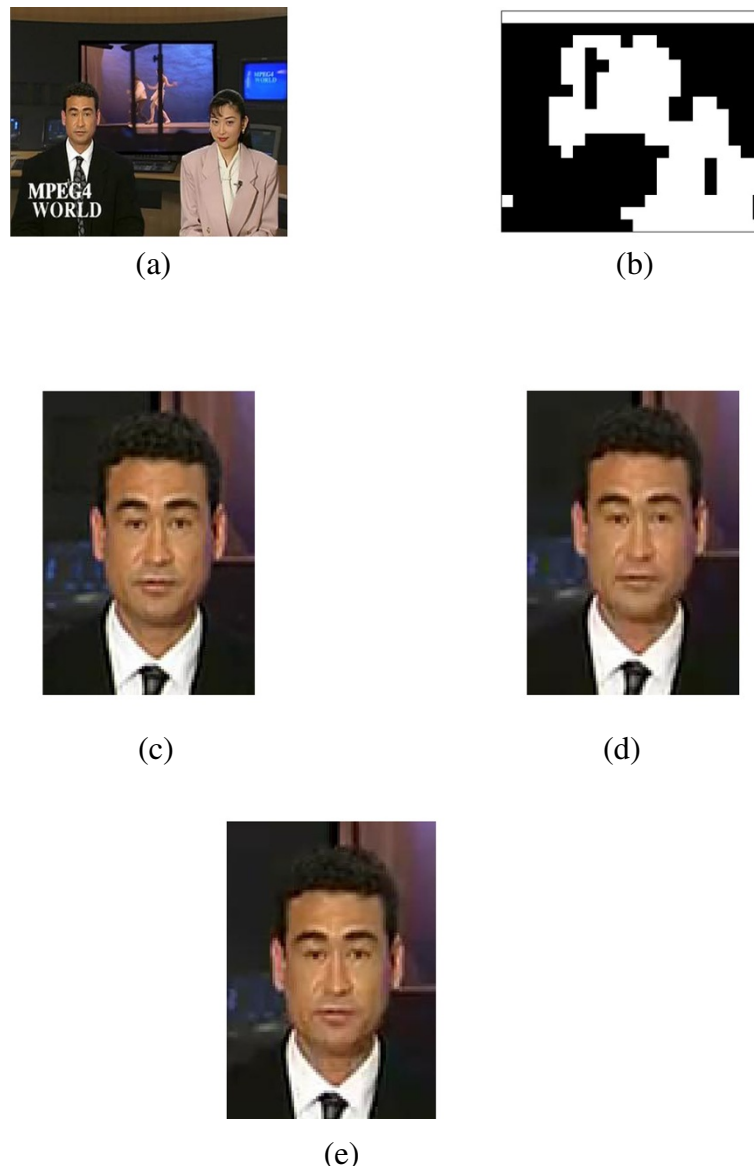


Figure 14 The subjective comparison for the 65th frame of News sequence with bit-rate of 390 Kbps. (a) Pre-encoded frame. The pre-encoded bit-stream with bit-rate 1Mbps (b) ROI map detected by [23] (c) Partial frame of pre-encoded sequence (d) Partial frame of CPDT method. $PSNR_{ROI}$: 35.58dB, $SSIM_{ROI}$: 0.95, and (e) Partial frame of ROI-based transcoder. $PSNR_{ROI}$: 37.78dB, $SSIM_{ROI}$: 0.97.

6. Conclusions

In this paper, a ROI-based video transcoder is proposed. Firstly, the Bayesian theorem is applied to automatically determine the ROI region for video transcoding. The region with moving foreground objects is segmented by decoding pre-encoded motion information from the input of the video transcoder. The skin color information is used to further find out the region of human attention. Second, a model-based video bit-rate transcoder is presented for baseline profile in H.264/AVC to meet the available communication bandwidth. We analyze the total coding bit-rate, NZTC, and QP data in the H.264/

AVC encoder to get our proposed models and reallocate the bits of each macroblock in consideration of ROI for bit-rate conversion. By virtue of the estimated target QP by the proposed I frame and P frame models, the simulated results show that the transcoded bit-stream conforms to the target bit-rate via the re-quantization method. Furthermore, this paper also proposes a ROI-based video transcoder to maintain visual quality in the region of user's interest for low bit-rate transmission. More coding bits are allocated in ROI MBs, and full coding modes are employed to ensure visual quality. We utilize the SKIP and the decoded mode for motion

estimation to reduce coding complexity on non-ROI MBs. Experimental results show that the proposed video bit-rate transcoder provides efficient video transcoding, accurate performance, and better video quality. In the future, the models of the total coding bit-rate, NZTC, and QP data for B frames will be investigated for the applications of high profile H.264/AVC video standard.

Competing interests

The authors declare that they have no competing interests.

Received: 30 October 2012 Accepted: 3 May 2013

Published: 29 May 2013

References

1. SK Kwon, A Tamhankar, KR Rao, Overview of H.264/MPEG-f part 10. *J. Visual Commun. Image Rep.* **17**, 186–216 (2006)
2. J Xin, CW Lin, MT Sun, Digital video transcoding. *Proceedings of the IEEE* **93**(1), 84–97 (2005)
3. CH Yeh, SJ Fan Jiang, TC Chen, MJ Chen, Motion vector composition through Lagrangian optimization for arbitrary frame-size video transcoding. *Optical Engineering* **51**(4), 047401 (2012)
4. CH Yeh, WY Tseng, ST Wu, Mode decision acceleration for H.264/AVC to SVC temporal video transcoding. *EURASIP J. Adv. Signal Process* **204** (2012)
5. J Xin, A Vetro, H Sun, Y Su, Efficient MPEG-2 to H.264/AVC transcoding of intra-coded video. *EURASIP J. Adv. Signal Process* **2007**, 075310 (2007)
6. X Liu, KY Yoo, Fast interframe mode decision algorithm based on mode mapping and MB activity for MPEG-2 to H.264/AVC transcoding. *J. Visual Commun. Image Rep* **21**, 155–166 (2010)
7. Q Tang, H Mansour, P Nasiopoulos, R Ward, Bit-rate estimation for bit-rate reduction H.264/AVC video transcoding in wireless networks, in *Proceedings of International Symposium on Wireless Pervasive Computing* (Santorini, 7–9 May 2008), pp. 464–467
8. CH Yeh, YH Chen, MC Chi, MJ Chen, Parabolic motion-vector re-estimation algorithm for compressed video downscaling. *J. Signal Process. Syst.* **61**(3), 375–386 (2010)
9. SK Kapotas, AN Skodras, Bit rate transcoding of H.264 encoded movies by dropping frames in the compressed domain. *IEEE Trans. Consum. Electron.* **56**(3), 1593–1601 (2010)
10. CT Hsu, CH Yeh, CY Chen, MJ Chen, Arbitrary frame rate transcoding through temporal and spatial complexity. *IEEE Trans. Broadcast.* **55**(4), 767–775 (2009)
11. CH Yeh, SJ Fan Jiang, CY Lin, MJ Chen, Temporal video transcoding based on frame complexity analysis for mobile video communication. *IEEE Trans. Broadcast* **59**(1), 38–46 (2013)
12. SK Kwon, SW Kin, JH Lee, JM Lee, An adaptive transcoding method for H.264 video coding. *IJCSNS* **8**(12), 154–160 (2008)
13. Z Chen, W Lin, KN Ngan, Perceptual video coding: challenges and approaches, in *Proceedings of the IEEE International Conference on Multimedia Expo, Suntec City, 19–23 Jul 2010*, pp. 784–789
14. Z Liu, Y Lu, Z Zhang, Real-time spatiotemporal segmentation of video objects in the H.264 compressed domain. *J. Visual Commun. Image Rep.* **18**, 275–290 (2007)
15. CM Mak, WK Cham, Real-time video object segmentation in H.264 compressed domain. *IET Image Processing.* **3**(5), 272–285 (2009)
16. MC Chi, CH Yeh, MJ Chen, Robust region-of-interest determination based on user attention model through visual rhythm analysis. *IEEE Trans. Circuits Syst. Video Technol* **19**(7), 1025–1038 (2009)
17. H Liu, MT Sun, RC Wu, SS Yu, Automatic video activity detection using compressed domain motion trajectories for H.264 videos. *J. Visual Commun. Image Rep.* **22**, 432–439 (2011)
18. S De Bruyne, C Poppe, S Verstockt, P Lambert, R Van de Walle, Estimating motion reliability to improve moving object detection in the H.264/AVC domain, in *Proceedings of IEEE International Conference on Multimedia and Expo, New York, 28 Jun-3 July 2009*, pp. 330–333
19. C Poppe, S De Bruyne, T Paridaens, P Lambert, R Van de Walle, Moving object detection in the H.264/AVC compressed domain for video surveillance applications. *J. Visual Commun. Image Rep* **20**, 428–437 (2009)
20. R Xie, S Yu, Region-of-interest based video transcoding from MPEG-2 to H.264 in the compressed domain. *Optical Engineering* **47**(9), 097001–1–097001–7 (2008)
21. H Li, Y Wang, CW Chen, An attention-information based spatial adaptation framework for browsing videos via mobile devices. *EURASIP J. Adv. Signal Process* **2007**, 1–12 (2007)
22. SK Kwon, HY Lee, A transcoding method for improving the subjective picture quality. *IJCSNS* **9**(1), 135–138 (2009)
23. SF Huang, MJ Chen, MS Li, Region-of-interest segmentation based on Bayesian theorem for H.264 video transcoding, in *Proceedings of Visual Communications and Image Processing Conference* (Tainan, 6–9 Nov 2011)
24. MC Chi, MJ Chen, CH Yeh, JA Jhu, Region-of-interest video coding based on rate and distortion variations for H.263+. *Signal Process. Image Commun.* **23**(2), 127–142 (2008)
25. Joint Video Team software JM, <http://iphome.hhi.de/suehring/ttml/download/>
26. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: form error visibility to structural similarity. *IEEE Trans. Image Processing* **13**(4), 600–612 (2004)

doi:10.1186/1687-6180-2013-112

Cite this article as: Huang et al.: Region-of-interest determination and bit-rate conversion for H.264 video transcoding. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:112.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com