Applications and Improvement of H.264 in Medical Video Compression

Hongtao Yu, Zhiping Lin, *Senior Member, IEEE*, and Feng Pan, *Senior Member, IEEE*

***Abstract*—This paper aims at applying H.264 in medical video compression applications and improving the H.264 rate control algorithm with better perceptual quality. First, H.264 is briefly reviewed and introduced to the area of medical video compression. Second, a new motion complexity (MC) measure is defined to express the complexity of motion contents in a video frame, and a new H.264 rate control scheme with the MC measure and perceptual bit allocation is proposed for medical video compression. Third, two sets of experiments are conducted: the comparison between MPEG-4 and H.264, and the comparison between JVT-H014 [1], which is the H.264 adopted rate control algorithm, and our proposed rate control scheme. The first set of experiments shows that compared with MPEG-4, H.264 can achieve a significant average peak signal-to-noise ratio (PSNR) gain of up to 4.35 dB for the test medical video sequences, and thus is much more effective when applied in medical video compression. The second set of experiments shows that compared with H014, the proposed rate control scheme can achieve better perceptual video quality, with an average PSNR gain of up to 0.19 dB for the test medical video sequences.**

***Index Terms*—H.264, medical video compression, motion complexity (MC), perceptual bit allocation, rate control.**

I. Introduction

In recent years, ongoing improvements in technology have had dramatic impacts on the medical healthcare field. From diagnosis to treatment, advanced technologies are playing an ever-increasing role and the area of medical video is no exception. Easy access to medical videos and three-dimensional (3-D) medical data sets, such as computed tomography (CT), magnetic resonance imaging (MRI), echocardiography, and so on, provides doctors with better capability of analyzing and diagnosing patients' conditions. Medical videos and 3-D medical data sets, however, require large amounts of memory storage and transmission bandwidth in telemedicine applications. This calls for video compression to reduce significantly the amount of data needed to represent a video.

Unlike in some other video compression applications where quality could be compromised with compression ratio, it is a very challenging task to compress medical videos as it requires both high quality and high compression ratio. On the one hand, for diagnostic purposes, it is essential that the compression process causes no tangible loss of detail and introduces no noticeable artifacts which could be otherwise misinterpreted as being pathological in nature. On the other hand, due to the limitation of storage and transmission and the huge amount of medical videos, high compression ratio is often required. As a result, although some progress in medical video compression has been made in the past decade, much remains to be done in this area.

In 1994, Tsai *et al.* [2] developed a compression scheme for angiogram video sequence based on a full frame discrete wavelet transform. This full frame design exploits the local characteristics of the compensated difference signals and achieves a higher compression gain. By adaptively searching the prediction error to identify the locations of the block artifacts and by modifying the error coding accordingly, it is possible to eliminate the artifacts from the final image. Gibson *et al.* [3] proposed a lossy wavelet-based approach for the compression of digital angiogram videos. Analysis of the high-frequency subbands of a wavelet decomposition of an angiogram video reveals significantly sized regions containing no diagnostically important information. A texture modeling approach is used to encode the high-frequency subband wavelet coefficients in such regions. This is only performed in regions which are considered diagnostically unimportant, with diagnostically important regions encoded as normal. In [4], a hybrid model has been discussed for the compression of CT sequences. The model uses lossless compression in the region of interest (ROI) with high quality, and lossy compression in other regions with very high compression ratio and reasonably good quality. The compression scheme is designed to automatically segment and utilize ROI in order to achieve a good balance between video quality and compression ratio. Note that all the aforementioned medical video compression methods have been developed using proprietary techniques and, hence, there lacks a common platform for medical practitioners to share/exchange compressed medical video data.

In order to disseminate compressed medical videos to average users efficiently and effectively, a number of efforts have been made to establish a common video compression standard for medical applications. The digital imaging and communications in medicine (DICOM) is the most commonly used standard [5], [25] which facilitates the distribution and viewing of medical images. It allows both lossless and lossy compression for various kinds of video/image sequences including echocardiography and CT. In order to improve the performance of transfer mechanisms for CT, a new family of objects, that is, the enhanced CT image module, has been added into the newest version of DICOM [6]. As is well known, the recommended lossy compression methods in DICOM are

Manuscript received February 1, 2005; revised July 29, 2005. Recommended by Guest Editor Y. Lian.

H. Yu and Z. Lin are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ezplin@ntu.edu.sg).

F. Pan is with the Institute for Infocomm Research, Singapore 119613.

Digital Object Identifier 10.1109/TCSI.2005.857869

1057-7122/$20.00 © 2005 IEEE

JPEG 2000 and MPEG-2. While investigation on MPEG-2 video compression scheme is still on going for the current DICOM standard, it is useful to explore the feasibility of using other video compression schemes as well. To address the needs of different applications, many video compression standards, such as ITU-T H.261, H.263, ISO/IEC MPEG-1, MPEG-2 and MPEG-4, have been ratified over the past decade. In [7], MPEG-4 has been used for the compression of ultrasound sequences. Experimental results showed that with support of object-based coding and compression efficiency, MPEG-4 is a good solution for compression of ultrasound sequences. Compression ratios of up to 90:1 have been obtained with objective image quality similar to H.263. Additionally, using MPEG-4 to encode different parts of an ultrasound sequence as separate objects would benefit telemedicine applications. In [8], MPEG-4 was also used for the compression of echocardiography sequences, which can greatly reduce the bit rate of echocardiography sequences for both archiving and transmission purposes. All the above work has shown that MPEG-4 can be effectively used in telemedicine applications.

A question arises naturally: Is MPEG-4 already the best video compression standard or is it possible to introduce a new standard which is superior to MPEG-4 for medical video compression? Recently the H.264 video coding standard has been very successful in many applications including two-way video communication, television broadcasting, video streaming over packet networks, digital media storage, and digital cinema, etc. To the best of our knowledge, no attempt has been made so far in applying H.264 in medical video compression. This is the main motivation for the research reported in this paper.

The objectives of this paper are twofold. First, we introduce H.264 to the area of medical video compression. Second, we propose a new perceptual bit allocation scheme for H.264 to further improve the perceptual quality of encoded medical videos. It is important to point out at this point that medical videos discussed in this paper include both traditional medical videos [sequences of two-dimensional (2-D) images taken over time] such as the "Echocardiography" sequence, and 3-D medical data sets (sequences of 2-D images acquired over space from 3-D objects) such as the "CT" sequence. We observe that neighboring frames in a 3-D medical data set exhibit certain degree of correlation similar to those in a traditional medical video and hence video compression techniques can be effectively applied to 3-D medical data sets as well. For this reason and for the convenience of exposition, we use "medical videos" for both traditional medical videos and 3-D medical data sets in this paper.

The rest of the paper is organized as follows. The next section briefly reviews the H.264 standard. Section III presents our proposed rate control scheme for H.264. Section IV presents experimental results on two typical medical videos comparing MPEG-4 and H.264 without rate control, and H.264 with the existing rate control algorithm (H014) [1] and with our proposed scheme, respectively. We show that for the given medical videos, there is a significant improvement in using H.264 compared with MPEG-4, and there is a consistent improvement in using H.264 with our proposed scheme compared with using H014. This paper concludes with Section V.

II. A Brief Review of H.264

A. H.264 Overview

The newest international video coding standard, H.264/AVC [9], has been approved recently by ITU-T as Recommendation H.264 and by ISO/IEC as International Standard 14 496-10 (MPEG-4 part 10) Advanced Video Coding (AVC). H.264 aims at providing functionality similar to existing video coding standards such as H.263+ and MPEG-4 but with significantly better compression performance and improved support for reliable transmission. It is not a fundamentally different method, but rather a significant refinement of well-established methods. The elements common to all video coding standards are present in the current H.264/AVC recommendation. Some new techniques, such as spatial prediction in Intra coding, adaptive block size motion compensation, 4×4 integer transformation, multiple reference pictures (up to seven reference pictures) and content adaptive binary arithmetic coding (CABAC), are used in this standard.

The testing results [10]–[12] have shown that H.264 has achieved substantial superiority of video quality over that of H.263, MPEG-2, and MPEG-4. It has achieved up to 50% in bit rate saving compared to H.263 or MPEG-4 coding schemes. This means that H.264 offers significantly higher coding quality with the same bit rates [13]. Therefore, H.264 will be a serious contender for a variety of next generation multimedia applications.

B. Rate Distortion Optimization

One of the novel features of H.264 video coding is the use of seven different macroblock (MB) coding modes so that the temporal and spatial details in an MB are best presented. These coding modes are SKIP, INTER 16×16, INTER 16×8, INTER 8×16, INTER 8×8, INTRA 16×16, INTRA 4×4. In INTER 8×8 mode, each block can be further divided independently into 8×8, 8×4, 4×8, or 4×4 subpartitions. To select the best mode for each MB, all the MB modes are tried and the one that leads to the least rate-distortion (RD) cost is selected. This is to achieve the best tradeoff between the rate and distortion performance, and is called rate distortion optimization (RDO). RDO is achieved by selecting the mode with the least RD cost using Lagrangian multiplier. The procedure can be defined as follows [14]:

$$\min J_i : \quad J_i = D_i + \lambda_i \times R_i \tag{1}$$

where J_i is the RD cost of an MB coded using Mode i, R_i is the data bits it consumed, and D_i is the distortion it caused. The mode that has the minimum RD cost is selected as the optimum coding mode for this MB.

In most medical images and video sequences, there exist some regions that are of special interests to the doctors, and the other regions are of no diagnostic significance. Hence, the employment of RDO in H.264 makes it especially suitable for coding of medical images: we may decrease the Lagrangian multiplier λ_i in (1) to reduce distortion in the ROI, and increase λ_i to reduce bit consumption in regions of no medical interest. Therefore, it is possible to select the best coding mode so as to achieve a compromise between high compression performance and good quality for encoding videos by accurately detecting the ROI. If H.264 is applied in medical video compression, the best coding mode that represents the characteristics of medical video contents may be selected correctly and hence high performance achieved. Thus, H.264 is expected to be effective in medical video compression applications.

III. Perceptual Rate Control Scheme for H.264

Although H.264 is currently the most efficient video compression standard, its rate control algorithm is not very satisfactory and needs further improvement in order for it to achieve high coding efficiency in encoding medical video sequences. In [1], the target bit is estimated solely based on the buffer fullness, regardless of the frame's content. This may lead to a drastic drop in peak signal-to-noise ratio (PSNR), especially in the case of high motion scenes or scene changes. Jiang *et al.* [15] introduced the mean absolute difference (MAD) ratio as a measure of motion complexity (MC) to improve the video quality at scene changes. However, MAD ratio is not a good ways for representing the motion contents, as it can only represent the similarity between the current frame and its reference frame. Another problem of the adopted rate control algorithm in [1] is that MAD is selected to estimate quantization parameter (QP) and to decide coding modes. However, it is well understood that the minimum MAD does not translate into minimum perceptual distortion [16]. A better approach is to use the perceptual characteristics of the video contents. In [17], a video coding approach performing adaptive RDO guided by perceptual hints was proposed. The key idea is to adaptively adjust the Lagrange multipliers of the RDO coder control module based on visual attention analysis. In [18], a Lagrangian optimized rate control algorithm was proposed for the H.264 video encoder. The algorithm controls the bit rate by adjusting the Lagrangian multiplier adaptively for every picture and specifying the QP for every MB. The success of the above perceptual algorithms depends largely on the good estimation of visual features and thus a more accurate visual important map needs to be established.

In this section, we propose a perceptual bit allocation scheme for H.264 to address the above two problems of the existing rate control algorithm. We focus on accurately estimating target bit at scene changes and high motions, and updating the Lagrangian multiplier adaptively according to the perceptual characteristics of the video contents. Furthermore, the proposed rate control scheme includes frame level and MB level. At the frame level, we estimate the target bits by using our new defined MC measure that represents the amount of motion contents between two consecutive frames. At the MB level, we allocate bits perceptually by updating the Lagrangian multiplier according to the MB patterns.

A. MC and Bit Allocation

Rate control is a necessary part in H.264 encoder. The purpose of rate control is to produce high quality videos at a given bit rate by appropriately allocating bits to residual information, motion vectors, and some other parameters. In order to achieve consistent video quality, bits should be allocated according to

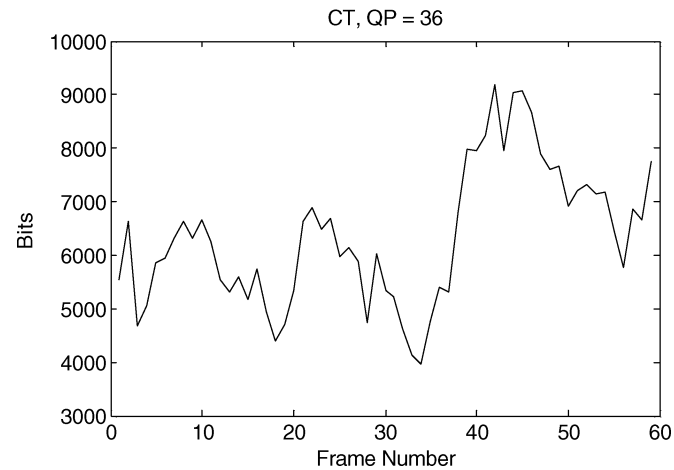


Fig. 1. Bit allocation of sequence "CT."

the MC. MC refers to the moving picture contents of two consecutive frames in a video sequence. If the objects in a frame are stationary, or a frame can be well matched to the previous frame with little difference, such a frame is a low motion frame. If there is a scene change between two consecutive frames, or a moving object is unable to find a good match in the previous frame, such a frame is considered as a high motion frame.

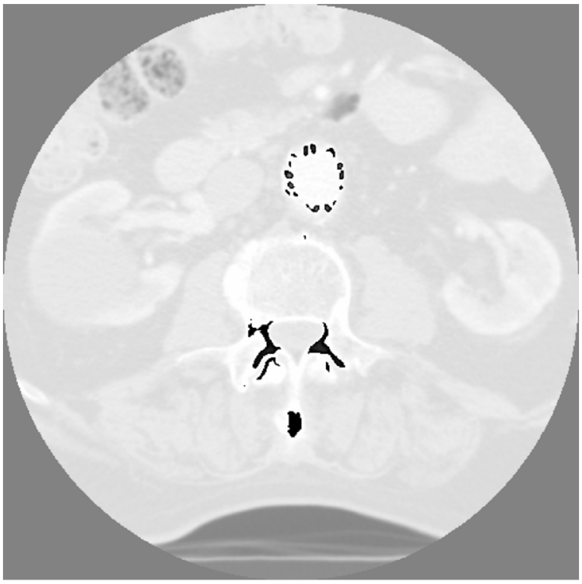
Fig. 1 shows the bit allocation of the sequence "CT" with fixed $\mathrm{QP} = 36$. From this figure, we can find that the bit allocation is in accordance with the complexity of frame's motion contents. For example, from the 43rd frame to the 44th frame, high motion happens at the black area. Fig. 2 shows the original images of the 43rd frame and the 44th frame. The number of bits allocated to the 44th frame is 14% more than that allocated to the 43rd frame. This shows that the number of allocated bits is a good measure to represent the MC of a frame.

However, it is difficult to obtain the actual bits consumed before entropy encoding. What we can do is only to estimate the target bit allocated to the current frame. Usually, there is a difference between the estimated bits and the actually consumed bits. The estimated target bit should represent the MC of current frame. Since there is a temporal correlation between consecutive frames, we can use the previous coded frames to estimate the current frame's MC, which can then be used to estimate the target bit. Based on the above discussion, we propose a new measure to represent the MC of a frame. In this measure, MC depends on the bits allocated to the previously encoded frames. Let $B_{P,i}$ be the predicted bits at frame i, B_j $(j = 1, 2, \cdots, i-1)$ be the actual allocated bits to the previously encoded frames, then MC at frame i is

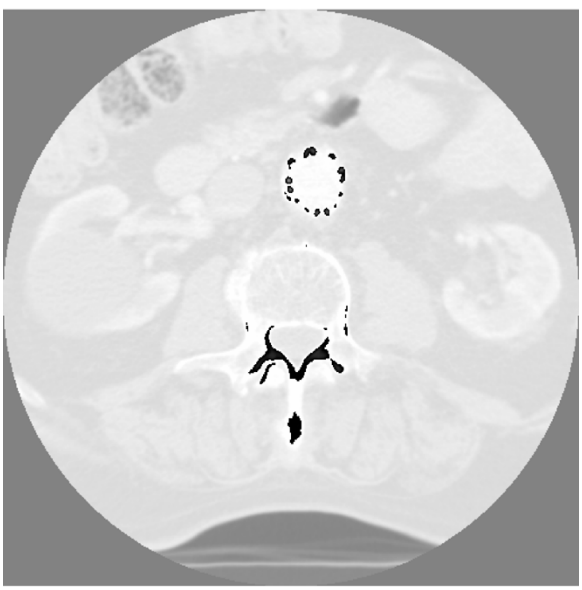
$$C_i = \frac{B_{P,i}}{\frac{1}{i-1}\sum_{j=1}^{i-1} B_j}, \qquad i = 2, 3, \cdots \tag{2}$$

where $B_{P,i}$ is the linear prediction of the actual previous frame's bits

$$B_{P,i} = \alpha_i B_{i-1}. \tag{3}$$

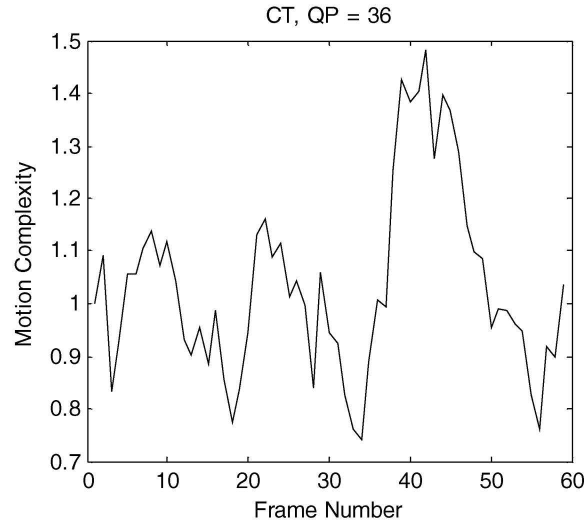


(a)

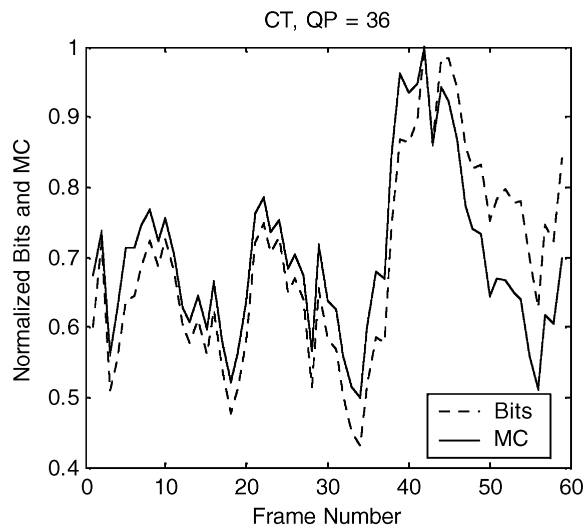


(b)

Fig. 2. Two consecutive frames with high motion in sequence "CT." (a) The 43rd frame. (b) The 44th frame.



(a)



(b)

Fig. 3. MC and bits of sequence "CT." (a) MC. (b) Normalized bits and MC.

It should be noticed that α_i needs to be updated after bits have been actually allocated. It depends upon the actual consumed bit by the current frame and the previous frame. That is

$$\alpha_{i+1} = \frac{B_i}{B_{i-1}} \times \alpha_i. \tag{4}$$

We illustrate the MC curve in Fig. 3(a). Comparing it with Fig. 1, we find that the shape of the MC curve is almost the same as that of the bit curve. This means that MC is also a good measure for the complexity of motion contents. To show the relationship between bits and MC clearly, we normalize them and combine the normalized results into one figure, that is, Fig. 3(b). From (2), it can be seen that the computation of MC is simple. It is only related to the actual bits of previous frames. Therefore, the proposed MC measure can accurately capture the complexity of current frame's motion contents, which helps accurately estimate the target bit in rate control.

B. Perceptual Mode Decision and Bit Allocation

From the description of H.264 mode decision and RDO in Section II, it is recognized that the selection of Lagrangian multiplier will influence the selection of coding modes. Fig. 4 illustrates the effect of updating the Lagrangian multiplier. In Fig. 4, each point on the RD curve represents a coding mode. Assume Mode 1 is the initial case where the Lagrangian multiplier is calculated according to (1). If we increase the Lagrangian multiplier, the coding mode will transfer to Mode 2 which means fewer bits are allocated and larger distortion is possible. If we decrease the Lagrangian multiplier, the coding mode will transfer to Mode 3 which means more bits are allocated and less distortion is allowed.

One of the important characteristics of human vision system (HVS) is called edge masking, i.e., HVS is more sensitive to errors on a prominent edge in an image while less so to errors

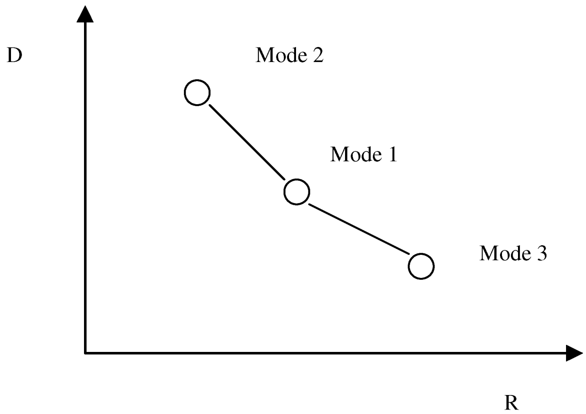


Fig. 4. Mode selection.

next to it. In other words, a dominant edge pattern will obscure the perception of other lower contrast variations in this block [19]. Hence, more bits should be allocated to the dominant edge pattern, while fewer bits should be allocated to the texture pattern. This observation can also be illustrated in Fig. 4. If the MB is texture pattern, fewer bits should be allocated to it and its coding mode is more likely Mode 2. On the other hand, if the MB is edge pattern, more bits should be allocated to it and its coding mode is more likely Mode 3. Therefore, we can consider the bits, the Lagrangian multiplier and the MB pattern simultaneously to allocate bits perceptually. That is, we update the Lagrangian multiplier according to the MB pattern, and this will change the mode selection, and correspondingly the allocated bits.

Based on the above analysis, a segmentation and mode decision method has been proposed to allocate bits perceptually. The steps of this method are listed as follows.

Step 1) Calculate the Sobel operators

$$\begin{aligned} G_x =& I_{x-1,y+1} + 2I_{x,y+1} + I_{x+1,y+1} - I_{x-1,y-1} \\ &- 2I_{x,y-1} - I_{x+1,y-1} \end{aligned} \quad (5)$$

$$\begin{aligned} G_y =& I_{x+1,y-1} + 2I_{x+1,y} + I_{x+1,y+1} - I_{x-1,y-1} \\ &- 2I_{x-1,y} - I_{x-1,y+1}. \end{aligned} \quad (6)$$

Step 2) Calculate the average squared gradients

$$G_{xx} = \sum_W G_x^2, \quad G_{yy} = \sum_W G_y^2, \quad G_{xy} = \sum_W G_x G_y. \quad (7)$$

Step 3) Calculate the coherence of the squared gradient [20].

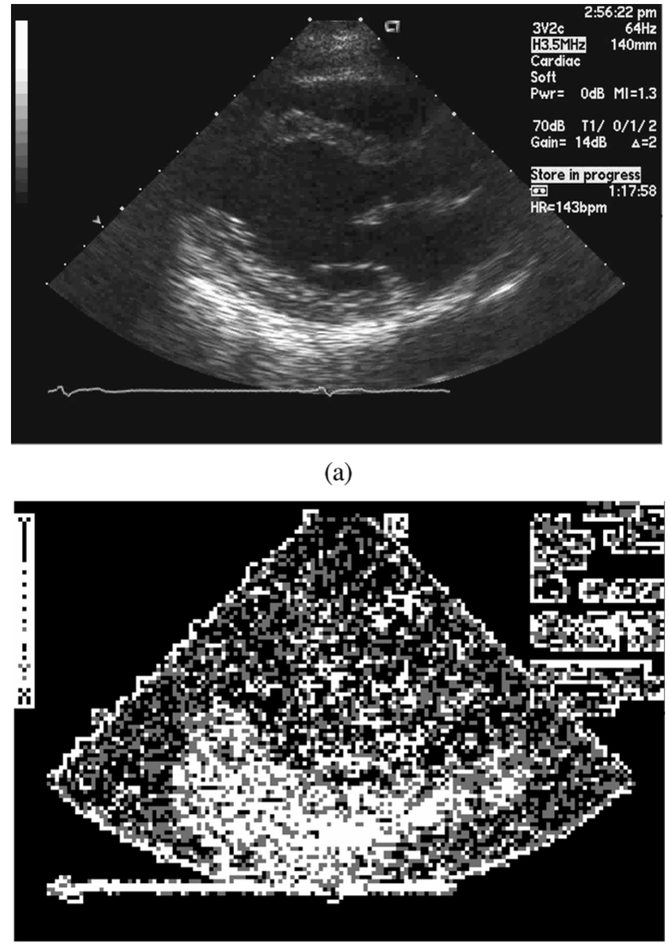
$$\text{Coh} = \sqrt{\frac{(G_{xx} - G_{yy})^2 + 4G_{xy}^2}{G_{xx} + G_{yy}}}. \quad (8)$$

Step 4) Determine the block patterns according to edge threshold T_{Edge} and texture threshold T_{Texture}

$$\text{Pattern} = \begin{cases} \text{Edge}, & \text{if Coh} > T_{\text{Edge}} \\ \text{Texture}, & \text{if } T_{\text{Texture}} < \text{Coh} \leq T_{\text{Edge}} \\ \text{Background}, & \text{others.} \end{cases} \quad (9)$$

Step 5) Update the Lagrangian multiplier

$$a_i = \begin{cases} 0.5, & \text{if Pattern} = \text{Edge} \\ 1.5, & \text{if Pattern} = \text{Texture} \\ 0.5, & \text{if Pattern} = \text{Backgrond} \end{cases} \quad (10)$$



(a)

(b)

Fig. 5. The 32nd frame of sequence "echocardiography." (a) Original image. (b) Segmentation image.

$$a = \sum_W a_i \quad (11)$$

$$\lambda' = a\lambda \quad (12)$$

where λ' is the updated Lagrangian multiplier.

The result of the above segmentation process can be shown with an example in Fig. 5. Fig. 5(a) is the original image of the 42nd frame of sequence "Echocardiography." Fig. 5(b) is its segmentation image. In Fig. 5(b), the black area is the background which is not important and can be greatly compressed. The white area is the ROI which has the most important cardiac information and should be compressed perceptually. It can be seen that most of the important edges and textures can be recognized in the segmentation image.

C. Rate Control Scheme

Based on complexity measure and perceptual mode decision, we propose a rate control scheme. This scheme includes two levels of rate control: frame-level rate control and MB-level rate control. At frame-level rate control, we estimate the target bit of the current frame according to MC. At MB-level rate control, we update Lagrangian multiplier according to the perceptual characteristics. The flow chart of the scheme is shown in Fig. 6.

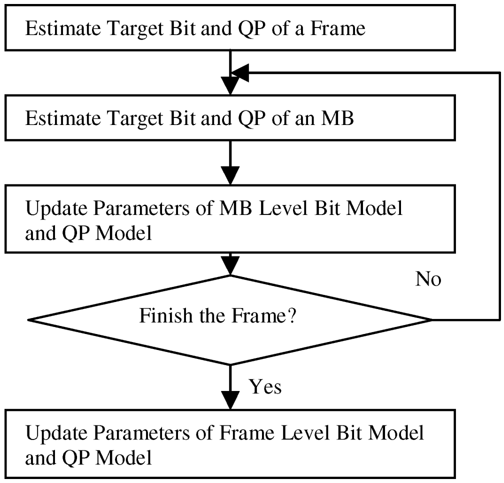


Fig. 6. Flowchart of the proposed rate control scheme.

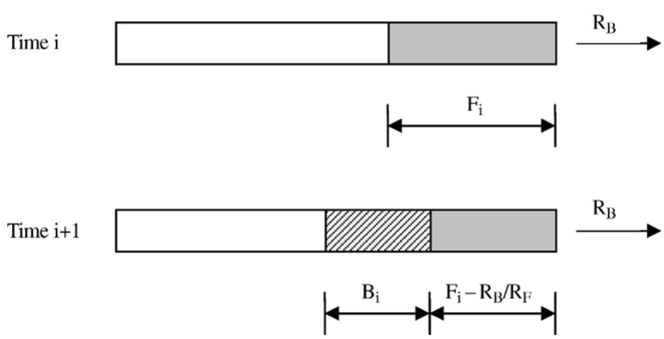


Fig. 7. Buffer model.

At the encoder, all the encoded bits are put into an output buffer for transmission. The output buffer fullness should be considered in target bit estimation so that the buffer is maintained neither overflow nor underflow. The output buffer can be illustrated with a fluid model as in Fig. 7.

In this model, R_B is the bit rate (supposed to be a constant), R_F is the frame rate, F_i is the buffer fullness at time i, B_i is the bits of frame i. Then at time $i+1$, the buffer fullness is

$$F_{i+1} = F_i + B_i - \frac{R_B}{R_F}. \tag{13}$$

In order to avoid buffer overflow, B_i should be determined according to the following constraint

$$B_i \le F_S - F_i + \frac{R_B}{R_F} \tag{14}$$

where F_S is the buffer size. This buffer model will be used in estimating target bit.

At the frame-level rate control, target bit estimation includes the total number of bits of a group of pictures (GOP), the remaining bits of a GOP, and the target bit of the current frame. The total number of bits of a GOP is determined by the following equation:

$$B = N \times \frac{R_B}{R_F} \tag{15}$$

where B is the total number of bits of a GOP, N is the number of frames in a GOP. The remaining bits of a GOP when encoding frame i is

$$B_{R,i} = \begin{cases} B, & i = 1 \\ B_{R,i-1} - B_{i-1}, & i = 2, 3, \cdots, N. \end{cases} \tag{16}$$

The bits consumed by the first frame, that is, the I frame, will not be estimated but is obtained according to the initial QP. The bit allocation of other frames will be estimated according to their MC, remaining bits of the GOP, buffer size, buffer fullness, bit rate, and frame rate. The target bit of the ith frame is

$$B_{T,i} = \frac{C_i \times B_{R,i} + \beta_i \times \left(F_s - F_i + \frac{R_B}{R_F}\right)}{N - i + 1}, \quad i = 2, 3, \cdots N \tag{17}$$

where C_i is the MC given by (2), β_i $(0 \le \beta_i \le 1)$ is a coefficient to control buffer fullness. In order to avoid buffer overflow, $B_{T,i}$ should be determined according to the following constraint:

$$B_{T,i} \le F_S - F_i + \frac{R_B}{R_F}. \tag{18}$$

Here, MC is employed to estimate target bit. From (17) it can be seen that the estimated target bit is proportional to MC. Thus, this scheme ensures that more bits will be allocated to high motion frames, and vice versa.

At the MB level rate control, we estimate QP and calculate the original Lagrangian multiplier. Then we use (5)–(12) to update the Lagrangian multiplier.

IV. Experimental Results

The evaluation of experimental results includes two parts: the comparison between MPEG-4 and H.264 without rate control, and the comparison between H.264 with existing rate control (that is, H014) and our proposed scheme. MPEG-4 is implemented with MoMuSys codec [21], and H.264 is implemented with H.264 reference model version JM6.1e [22]. In order to determine the target bit rate, we first encode the test sequences by using a fixed QP in H.264 without rate control. After this has been done, the actual consumed number of bits is obtained and used as the target bit rate in the subsequent experiments. Considering different network conditions, the QP's used in the experiment are 28, 32, 36, and 40, respectively. Thus, there are four types of target bit rate for each test sequence.

Table I lists the test video sequences used in our experiments. The conditions of the experiments in H.264 are listed in Table II. In the experiments, we only consider the features enabled in the Main Profile of H.264.

A. Comparison Between MPEG-4 and H.264

The test conditions of MPEG-4 are set to match H.264 as much as possible. For example, the frame type is IPPP, the frame rate is 30 fps, and the search range is 32, etc. Table III shows the results achieved from MPEG-4 and H.264 without rate control. For each sequence, four bit rates from high to low are selected as the target bit rates. To see the difference between MPEG-4 and H.264, we show PSNR gain and bit rate saving in the table.

TABLE I
TEST SEQUENCES

Sequence	Format	Size	Frame Number
CT	YUV	512×512	60
Echocardiography	YUV	640×480	60

TABLE II
TEST CONDITIONS IN H.264

Frame Type	IPPP
Frame Rate	30 fps
Slice Mode	OFF
RDO	ON
Rate Control	OFF
Hardmard	OFF
Search Range	32
Restrict Search Range	No restriction
Symbol Mode	CABAC
Partition Mode	No data partition
Out File Mode	Annex B

TABLE III
RESULTS ACHIEVED USING MPEG-4 AND H.264

Sequence	Target Bit Rate (kbps)	PSNR (dB)		PSNR Gain (dB)	Bit Rate (Kbps)		Bit Rate Saving (%)
		MPEG-4	H.264		MPEG-4	H.264	
CT	438.19	36.12	43.07	6.95	450.93	438.19	2.83
	285.03	35.05	40.55	5.50	303.56	285.03	6.10
	197.55	35.10	37.73	2.63	214.89	197.55	8.07
	145.10	35.07	34.91	-0.16	162.38	145.10	10.64
Echocardiography	2266.14	35.79	38.09	2.30	2266.51	2266.14	0.02
	1317.29	32.94	35.15	2.21	1292.83	1317.29	-1.89
	667.08	30.98	32.44	1.46	703.83	667.08	5.22
	308.26	30.41	29.81	-0.60	327.52	308.26	5.88

PSNR gain is the PSNR difference in decibels (dB) between H.264 and MPEG-4. Bit rate saving is the bit rate difference in percentage between H.264 and MPEG-4.

It is obvious from Table III that H.264 performs much better than MPEG-4 when applied in the test medical video sequences. For sequence "CT," H.264 obtains positive PSNR gain and bit rate saving in most cases. Even a significantly high PSNR gain of 6.95 dB is earned at high bit rate. For sequence "Echocardiography," H.264 also obtains positive PSNR gain and bit rate saving in most cases.

To see the overall performance comparison for the test sequences, we employ average PSNR difference and average bit rate difference as performance measure [23]. These measures are often used to compare RD performance between two different methods. The results are shown in Table IV.

It can be seen from Table IV that H.264 performs much better than MPEG-4 in average PSNR difference and bit rate difference. This result is in accordance with what has been described in the JVT report [10]. This shows that H.264 is a more efficient solution than MPEG-4 for medical video compression.

TABLE IV
AVERAGE PSNR DIFFERENCE AND BIT RATE DIFFERENCE BETWEEN H.264 AND MPEG-4

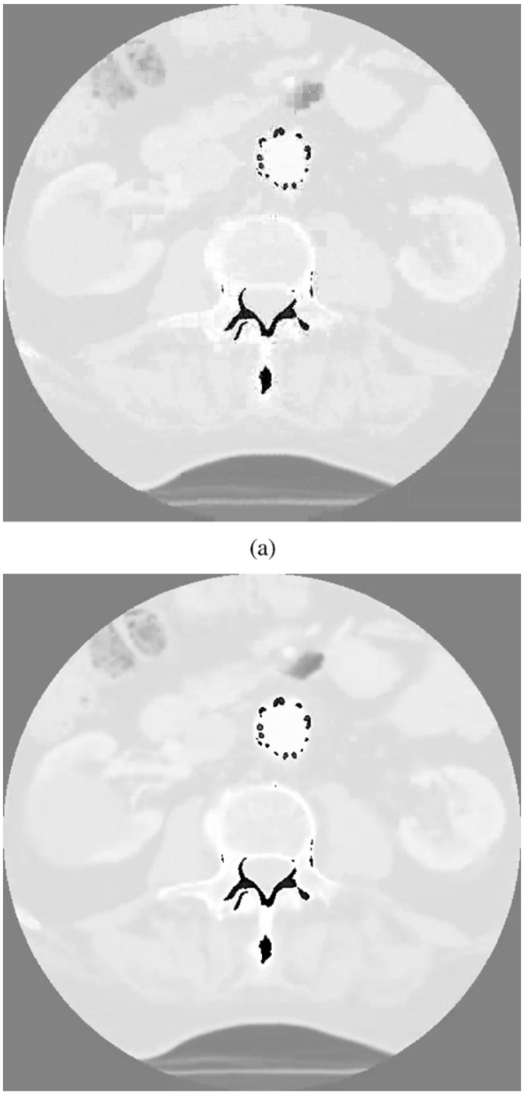
Sequence	Average PSNR Difference over Full Range of Bit Rate (dB)	Average Bit Rate Difference over Full Range of PSNR (%)
CT	4.35	< -100
Echocardiography	1.51	-59.71

Fig. 8 gives the reconstructed images of the 44th frame of sequence "CT" at the bit rate of 438.19 kilobits per second (kbps), where the original image is given in Fig. 2(b). Fig. 9 gives the reconstructed images of the 32nd frame of sequence "Echocardiography" at the bit rate of 667.08 kbps, where the original image is given in Fig. 5(a). From these figures we can see that the reconstructed images of H.264 are better than those of MPEG-4. For example, there are some blurs in Fig. 8(a) while the same parts are clear in Fig. 8(b).

B. Comparison Between H.264 Rate Control Algorithm and the Proposed Scheme

1) Objective Quality Evaluation: Table V shows the results achieved by the H014 rate control algorithm and the proposed rate control scheme. To see the difference between H014 and the proposed scheme, we use PSNR gain and bit rate saving as performance measures in the table. PSNR gain is the PSNR difference in decibels between the proposed scheme and H014. Bit rate saving is the bit rate difference in percentage between the proposed scheme and H014.

As observed from Table V, the proposed scheme is able to improve both the PSNR and bit rate in most cases. As QP increases, or in other words the target bit rate decreases, the PSNR gain of the proposed scheme compared with H014 increases. This shows the proposed scheme performs better in the low bit rate network environment. In particular, the proposed scheme performs better for high motion sequence "CT." PSNR gain is positively obtained for all the target bit rates for sequence "CT." Although sometimes the bit rate saving is negative, for example at 285.03 kbps, the proposed scheme still gets bit rate saving in most cases and is thus overall better in bit rate saving. For sequence "Echocardiography," the proposed scheme obtains positive bit rate saving in all the cases and positive PSNR gain in most cases. To see the performance of sequence "Echocardiography" at very high bit rate, we have also done experiments in the case of $\mathrm{QP} = 24$ and target $\mathrm{Bit\ rate} = 3402.04$ kbps. In this case, the PSNR and the bit rate of H014 are 41.29 dB and 3403.36 kbps, respectively, while the PSNR and the bit rate of the proposed scheme are 41.26 dB and 3403.38 kbps, respectively, implying that the PSNR gain and the bit rate saving of the proposed scheme are -0.03 dB and 0.00%, respectively. This shows that at very high bit rate for sequence "echocardiography," H014 performs slightly better than the proposed scheme. Despite of this, the proposed scheme is still overall better than H014.



(a)

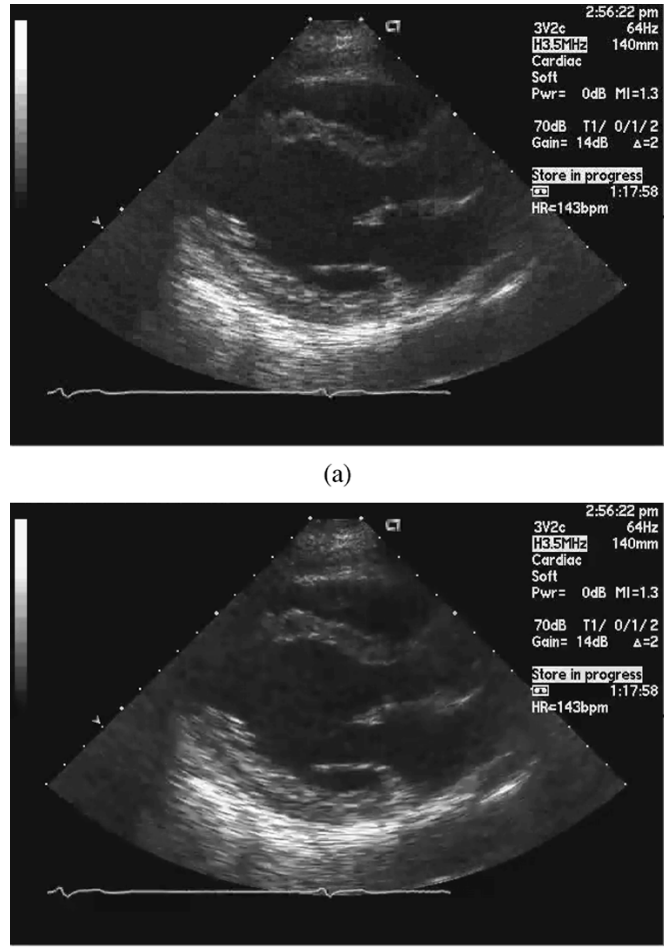
(b)

Fig. 8. The reconstructed images of the 44th frame of sequence "CT" at bit rate of 438.19 kbps. (a) MPEG-4. (b) H.264.

Table VI is the results of average PSNR difference and bit rate difference between the proposed rate control scheme and H014. As can be seen from Table VI, the average performance of the proposed scheme is better than H014. Comparing with Table V, although sometimes the PSNR gain and bit rate saving do not perform well at some target bit rates, the average PSNR difference and bit rate difference are still good. This shows the proposed scheme outperforms the H.264 rate control algorithm H014.

Fig. 10 shows the PSNR results frame by frame. Fig. 10(a) is for sequence "CT" at the target bit rate of 285.03 kbps, while Fig. 10(b) is for sequence "Echocardiography" at the target bit rate of 667.08 kbps. It can be seen from the figure that the proposed scheme achieves higher PSNR than H014 at most frames. This shows again that the proposed scheme performs better than H014 in terms of PSNR.

2) Subjective Quality Evaluation: For subjective quality evaluation, we use the double stimulus continuous quality scale (DSCQS) test method which is described in ITU-R BT.500-10



(a)

(b)

Fig. 9. The reconstructed images of the 32nd frame of sequence "echocardiography" at bit rate of 667.08 kbps. (a) MPEG-4. (b) H.264.

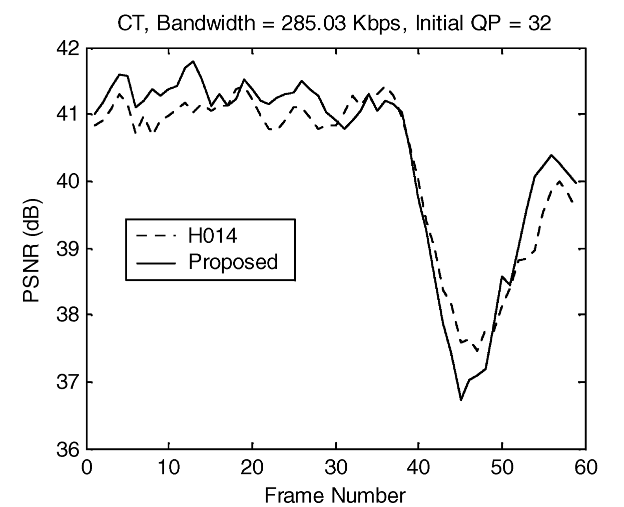
TABLE V
RESULTS OF H014 AND THE PROPOSED SCHEME

Sequence	Target Bit Rate (kbps)	PSNR (dB)		PSNR Gain (dB)	Bit Rate (Kbps)		Bit Rate Saving (%)
		H014	Proposed		H014	Proposed	
CT	438.19	42.89	42.97	0.08	439.45	439.80	-0.08
	285.03	40.28	40.43	0.15	286.37	286.92	-0.19
	197.55	37.43	37.69	0.26	198.00	197.91	0.05
	145.10	35.06	35.33	0.27	146.73	145.79	0.64
Echocardiography	2266.14	38.17	38.16	-0.01	2268.22	2265.24	0.13
	1317.29	35.20	35.27	0.07	1319.49	1318.66	0.06
	667.08	32.43	32.56	0.13	671.47	670.46	0.15
	308.26	29.84	30.00	0.16	311.73	308.45	1.05

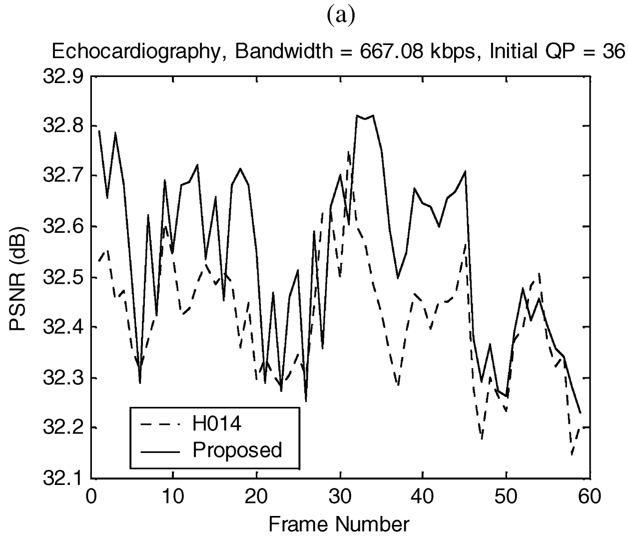
[24]. The mean opinion score (MOS) scales for observers to vote for the quality after viewing are excellent, good, fair, poor and bad. Correspondingly, the scores are 5, 4, 3, 2, and 1. Five observers were involved in the experiments, including a medical doctor. All of them have some knowledge of image processing. Table VII lists the evaluation results. The higher the score gain is, the better the perceptual quality is achieved. From the table, we can see that the subjective score of the

TABLE VI
AVERAGE PSNR DIFFERENCE AND BIT RATE DIFFERENCE BETWEEN THE PROPOSED RATE CONTROL SCHEME AND H014

Sequence	Average PSNR Difference over Complete Range of Bit Rate (dB)	Average Bit Rate Difference over Complete Range of PSNR (%)
CT	0.19	-2.65
Echocardiography	0.11	-2.66



(a)



(b)

Fig. 10. PSNR versus frame number. (a) CT, Bit rate $= 285.03$ kbps. (b) Echocardiography, Bit rate $= 667.08$ kbps.

proposed scheme is better than that of H014. This shows that the proposed scheme can achieve better perceptual quality than H014. It is interesting to see that positive score gains are achieved for all the low bit rate cases for both sequences, agreeing with the objective quality evaluation results in 1).

TABLE VII
SUBJECTIVE SCORE OF H014 AND THE PROPOSED RATE CONTROL SCHEME

Sequence	QP	Target Bit Rate (kbps)	Score		Score Gain
			H014	Proposed	
CT	28	438.19	4.20	4.20	0.00
	32	285.03	3.60	3.80	0.20
	36	197.55	3.60	4.00	0.40
	40	145.10	3.60	3.80	0.20
Echocardiography	28	2266.14	4.20	4.20	0.00
	32	1317.29	3.80	4.20	0.40
	36	667.08	4.00	4.20	0.20
	40	308.26	3.60	3.80	0.20

V. CONCLUSION

In this paper, we have introduced H.264 for compression of medical videos and 3-D medical data sets. Experimental results have shown that for medical video compression, H.264 performs much better than MPEG-4 both in PSNR gain and bit rate saving. This means that H.264 is more effective than MPEG-4, and hence is a good alternative solution for medical video applications. We have also proposed a new MC measure to represent the complexity of a frame's motion contents, and a perceptual mode decision algorithm by updating Lagrangian multiplier according to the perceptual characteristics of video contents, so that more bits are allocated to edge patterns. Based on the MC and perceptual mode decision, we have presented a new rate control scheme for H.264. In this scheme, MC is used to estimate a frame's target bit so that the bit allocation is in accordance with the complexity of frame's motion contents, and perceptual mode decision is employed to allocate the MB's bits perceptually. Experimental results have shown that our proposed rate control scheme outperforms H.264 rate control for medical video compression.

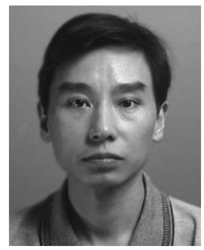
ACKNOWLEDGMENT

The authors would like to thank Dr. R. San Tan from Singapore National Heart Centre for his valuable comments on medical videos and his kind assistance in the evaluation of the reconstructed medical videos, and the anonymous reviewers for helpful comments and suggestions.

REFERENCES

[1] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu, and Y. Lu, "Adaptive rate control with HRD consideration," in *Joint Video Team of ISO/IEC MPEG & ITU-T VCEG 8th Meeting*, Geneva, Switzerland, May 2003.
[2] M. J. Tsai, J. D. Villasenor, and B. K. T. Ho, "Coronary angiogram video compression," in *Proceedings of 1994 IEEE Nuclear Science Symp. Medical Imaging Conf.*, Norfolk, VA, Oct. 1994, paper no. 8M65.

[3] D. Gibson, G. Tsibidis, M. Spann, and S. Woolley, "Angiogram video compression using a wavelet-based texture modeling approach," in *Proc. 2001 British Conf. Medical Image Understanding and Analysis*, Birmingham, U.K., Jul. 2001.
[4] S. B. Gokturk, C. Tomasi, B. Girod, and C. Beaulieu, "Medical image compression based on region of interest, with application to colon CT images," in *Proc. 23rd Annual Int. Conf. IEEE Engineering in Medicine and Biology Society*, Istanbul, Turkey, Oct. 2001.
[5] NEMA, Digital Imaging and Communications in Medicine (DICOM) Part 1: Introduction and Overview, National Electrical Manufactures Association, 2004.
[6] D. Clunie and B. J. Erickson, "The new enhanced multiframe CT and MR DICOM objects," in *Proc. 2005 SCAR (Society for Computer Applications in Radiology) Annual Meeting*, Orlando, FL, Jun. 2005.
[7] C. Lau, J. E. Cabral, A. H. Rambhia, and Y. M. Kim, "MPEG-4 coding of ultrasound sequences," in *Proc. SPIE*, vol. 3976, San Diego, CA, Feb. 2000, pp. 573–579.
[8] P. Barbier, M. Alimento, G. Berna, D. Cavoretto, F. Celeste, M. Muratori, and M. D. Guazzi, "Clinical validation of different echocardiographic motion pictures expert group-4 algorithms and compression levels for telemedicine," in *Proc. Medinfo 2004*, San Francisco, CA, Sep. 2004.
[9] *Information Technology—Coding of Audio-Visual Objects—Part 10: Advanced Video Coding*, ISO/IEC FDIS 14496-10, 2003.
[10] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[11] A. Joch, F. Kossentini, and P. Nasiopoulos, "A performance analysis of the ITU-T draft H.26L video coding standard," in *Proc. 12th Int. Packetvideo Workshop*, Pittsburgh, PA, Apr. 2002.
[12] S. Saponara, C. Blanch, K. Denolf, and J. Bormans, "The JVT advanced video coding standard: complexity and performance analysis on a tool-by-tool basis," in *Proc. 13th Int. Packetvideo Workshop*, Nantes, France, Apr. 2003.
[13] T. Hallbach, "H.264 video compression standard," in *Proceedings of 6th Nordic Signal Processing Symposium*, Bergen, Norway, Oct. 2003.
[14] T. Wiegand, H. Schwarz, A. Joch, F. Lossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.
[15] M. Q. Jiang, X. Q. Yi, and N. Ling, "Improved frame-layer rate control for H.264 using MAD ratio," in *Proc. 2004 IEEE Int. Symp. Circuits Syst.*, vol. III, Vancouver, BC, Canada, May 2004, pp. 813–816.
[16] S. Winkler, "A perceptual distortion metric for digital color video," in *Proc. SPIE*, vol. 3644, San Jose, CA, Jan. 1999, pp. 175–184.
[17] C. J. Tsai, C. W. Tang, C. H. Chen, and Y. H. Yu, "Adaptive rate-distortion optimization using perceptual hints," in *Proc. 2004 IEEE Int. Conf. Multimedia and Expo*, Taipei, China, Jun. 2004, pp. 667–670.
[18] M. M. Ghandi and M. Ghanbari, "A Lagrangian optimization rate control algorithm for the H.264/AVC encoder," in *Proc. 2004 IEEE Int. Conf. Image Processing*, Singapore, Oct. 2004, pp. 123–126.
[19] F. Pan, "Adaptive image compression using local pattern information," *Pattern Recognit. Lett.*, vol. 23, no. 14, pp. 1837–1845, Dec. 2002.
[20] A. M. Bazen and S. H. Gerez, "Systematic methods for the computation of the directional fields and singular points of fingerprints," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 905–919, Jul. 2002.
[21] Reference Model MoMuSys, MPEG. [Online]. Available: http://www.ganesh.org/~pmeerw/MPEG4_MoMuSys-FDIS-V1.0-990 812-bugfixed.tar.gz
[22] Reference Model JM6.1e, JVT. [Online]. Available: ftp://ftp.imtc-files.org
[23] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *Proc. ITU-T Q.6/SG16 VCEG 13th Meeting*, Austin, TX, Apr. 2001, Doc. VCEG-M33.
[24] "Methodology for the Subjective Assessment of the Quality of Television Pictures,", ITU-R Recomm. BT.500-10, Mar. 2000.
[25] Digital Imaging and Communications in Medicine (DICOM) Part 3: Information Object Definitions, 2004.



Hongtao Yu received the B.Eng. and M.Eng. degrees from Huazhong University of Science and Technology and Nanyang Technological University (NTU), Singapore, respectively.

He is now with NTU. His research areas are digital image processing, digital image/video compression, multimedia system, and telecommunication network management.

Zhiping Lin (SM'99) received the Ph.D. degree from the University of Cambridge, Cambridge, U.K., in 1987.

He was with the University of Calgary, Calgary, Canada, from 1987 to 1988, with Shantou University, China, from 1988 to 1993, and with DSO National Laboratories, Singapore, from 1993 to 1999. Since February 1999, he has been an Associate Professor with Nanyang Technological University (NTU), Singapore. He is also the Program Director of Bio-Signal Processing, Center for Signal Processing, NTU. His research interests include multidimensional systems and signal processing, array signal processing, and biomedical signal processing.

Dr. Lin is a coeditor of *Multidimensional Systems and Signal Processing* and an Associate Editor of *Circuits, Systems and Signal Processing*.

Feng Pan (M'97–SM'01) received the B.Sc., M.Sc., and Ph.D. degrees in communication and electronic engineering from Zhejiang University, China, in 1983, 1986, and 1989, respectively.

Since then, he has been teaching and researching in a number of universities in China, U.K., Ireland, and Singapore. He is now the Assistant Director of Media Division, Institute for Infocomm Research, Singapore. His research areas are digital image processing, digital image/video compression, digital television broadcasting, and pattern recognition. He has published more than 60 technical papers in the above area, and conducted many short courses for companies in Singapore and Southeast Asia.

Dr. Pan was the General Chairman of the 7th International Symposium on Consumer Electronics, Sydney, Australia, December 3–5, 2003, and has been serving as the member of organizing committee for a number of international conferences. He was the Chapter Chairman of IEEE Consumer Electronics, Singapore, from 2002 to 2004. He is currently the Associate Editor of the *International Journal of Innovational Computing and Information Control*.