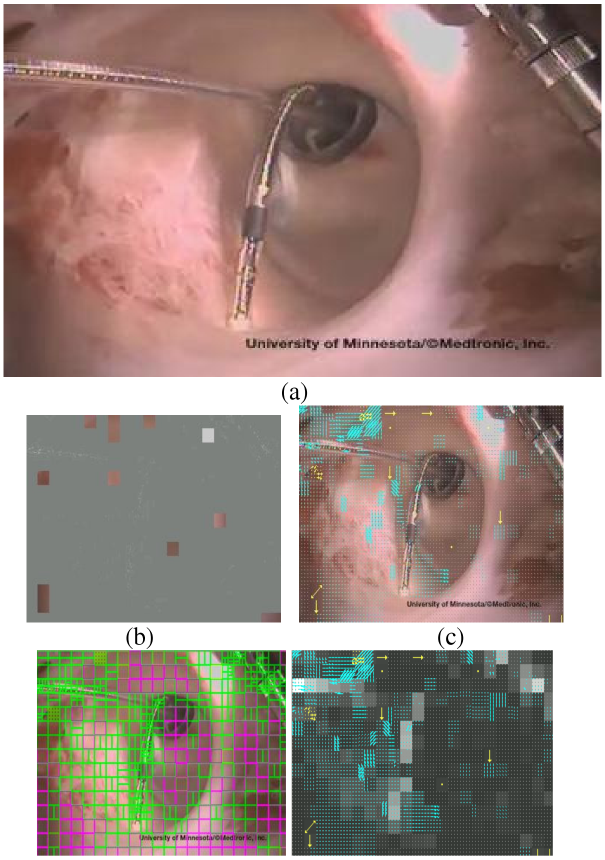


Fig. (16): Samples of “Rahsa” sequence.(a) Original Image (b) MB division. (c) Motion vector. (d) Classification result of intra- and inter-prediction. (e) Classification result of intra- and inter-predictio without background

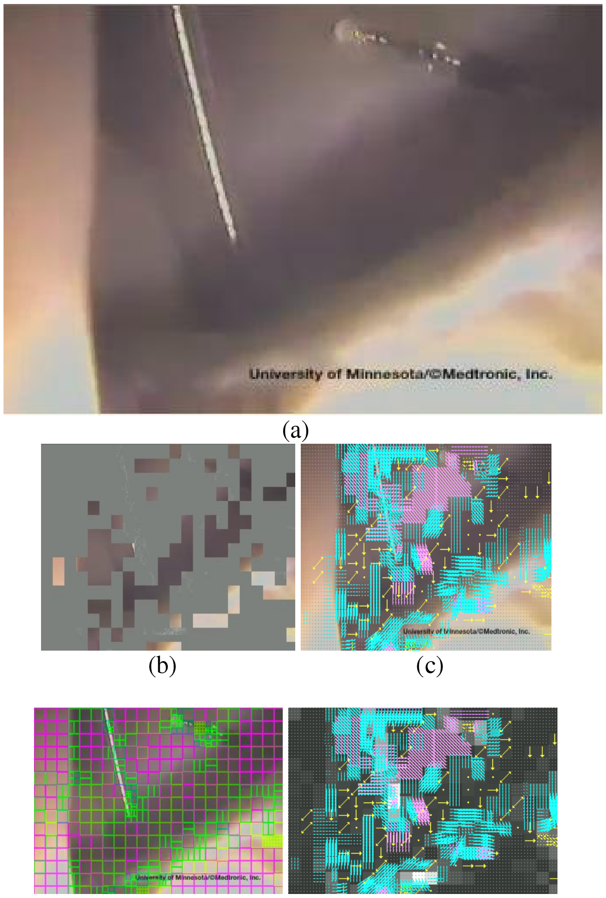


Fig. (17): Samples of “Ralse” sequence.(a) Original Image (b) MB division. (c) Motion vector. (d) Classification result of intra- and inter-prediction. (e) Classification result of intra- and inter-predictio without background

References

[1] F. Pan, X. Lin, S. Rahardja, K.P. Lim, Z.G. Li, D. Wu, S. Wu, "Fast mode decision algorithm for intra-prediction in H.264/AVC video coding", IEEE Trans. on Circ. and Sys. for Vid. Tech., Vol.15, No.7, pp.813-822, July 2005.

[2] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, "Overview of the H.264/AVC video coding standard", IEEE Trans. on Circ. and Sys. for Vid. Tech., Vol.13, No.13, No.7, pp.560-576, 2003.

[3] ece.ut.ac.ir/classpages/Multimedia/h264.ppt .

[4] Envivio http://www.envivio.com/products/h264.html.

[5] C. Kim, Q. Li, C.C.J. Kuo, "Fast intra-prediction model selection for H.264 codec", SPIE Int. Sym. ITCOM, Orlando, Florida, Sep. 2003.

[6] F. Pan, X. Lin., "Fast mode decision for intra-prediction", ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT 7th Meeting, Pattaya II, Thailand, March 2003.

[7] F. Pan, X. Lin, S. Rahardja, K.P. Lim, Z.G. Li, "A directional field based fast intra-mode decision algorithm for H.264 video coding", IEEE Int. Conf. on Mult. and Expo., Vol.2, pp.1147-1150, June 2004.

[8] J. Kim, J. Jeong, "Fast intra-mode decision in H.264 Video coding using simple directional masks", Proc. of SPIE, Vol.5960, pp.1071-1079, 2005.

[9] R. Garg, M. Jindal, M. Chauhan, "Statistics based fast intra- mode detection", Proc. of SPIE, Vol.5960, pp.2085-2091, 2005.

[10] B. Jeon, J. Lee, "Fast mode decision for H.264", ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT 10th meeting, Waikoloa, Hawaii, Dec. 2003.

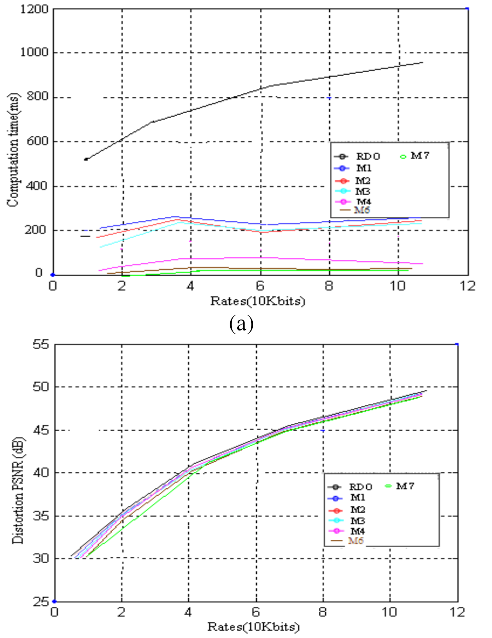
[11] C. Yang, L. PO, W. Lam, "A fast H.264 intra-prediction algorithm using macroblock properties", ICIP, pp.461-464, 2004.

[12] B. Meng, O.C. AU, C. Wong, H. Lam, "Efficient intra- prediction algorithm in H.264", IEEE, pp.837-840, 2003.

[13] C. Tseng, H. Wang, J. Yang, "Improved and fast algorithms for intra 4×4 mode decision in H.264/AVC", IEEE, pp.2128-2131, 2004.

[14] C. Hsu, M. Ho, J. Hong, "An efficient algorithm for intra- prediction in H.264", IEEE, pp.35-36, 2006.

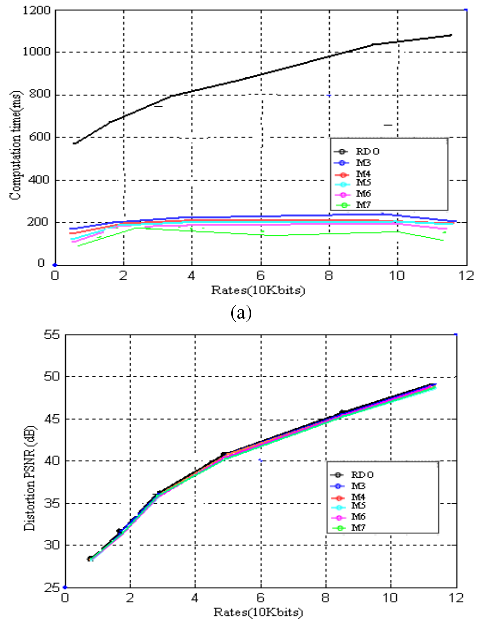
[15] Joint Video Team (JVT), Reference software JM7.1, http://bs//hhi.de/~suehring/tml/download/JM7.1.zip.



(a)

(b)

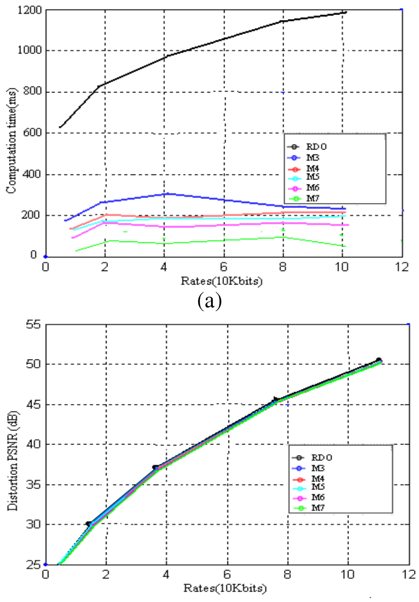
Fig. (12): “Ranna” sequence (IPPP seq.). (a) Computational complexity. (b) R-D performance



(a)

(b)

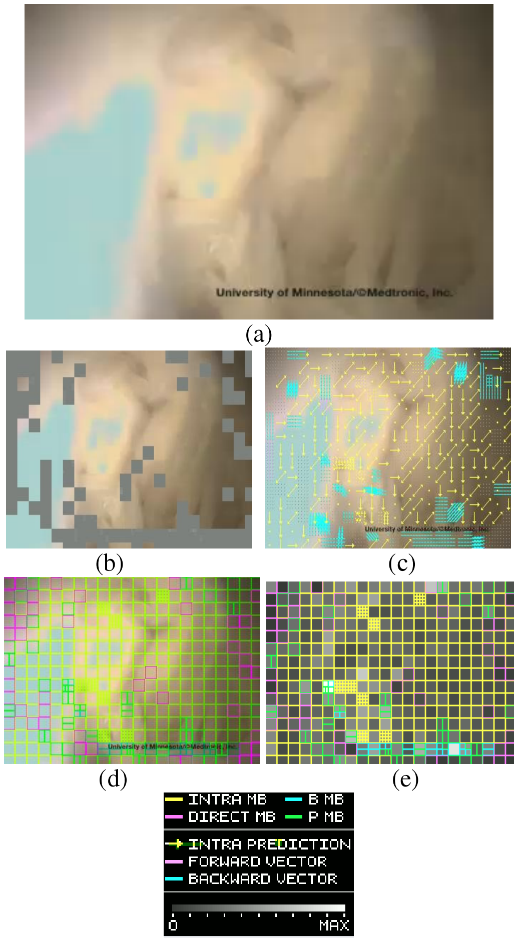
Fig. (13): “Rahsa” sequence (IPPP seq.). (a) Computational complexity. (b) R-D performance



(a)

(b)

Fig. (14): “Rals” sequence (IPPP seq.). (a) Computational complexity. (b) R-D performance



(a)

(b)

(c)

(d)

(e)

(f)

Fig. (15): Samples of “Ranna” sequence.(a) Original Image (b) MB division. (c) Motion vector. (d) Classification result of intra- and inter-prediction. (e) Classification result of intra- and inter-predictio without background. (f) Legend

Comparisons with the case of exhaustive search (RDO) were performed with respect to the change of average PSNR (ΔPSNR), the change of average data bits (ΔBit), and the change of average encoding time (ΔTime), respectively.

The PSNR is derived from

$$\mathrm{PSNR} = 10\log_{10}(255^2 / \mathrm{MSE}) \qquad (9)$$

Therefore, in the rest of this paper we used the overall PSNR value of all three components Y, U, and V using Eq. (9).

Δ time be defined as

$$\Delta\mathrm{Time}\,\% = \frac{T_{\mathrm{prop}} - T_{\mathrm{ref}}}{T_{\mathrm{ref}}} \times 100 \qquad (10)$$

Also, the bitrate increase is defined as

$$\Delta\,\mathrm{Bitrate}\,\% = \frac{\mathrm{bitrate}_{\mathrm{prop}} - \mathrm{bitrate}_{\mathrm{ref}}}{\mathrm{bitrate}_{\mathrm{ref}}} \times 100 \qquad (11)$$

A group of experiments were carried out on different sequences. The encoding bitrates, the PSNR values, and the time saving factor (as compared with the H.264 RDO method) for three test sequences with different quantization parameters are listed in Tables 5~7. Generally speaking, as can be seen from these tables, we have saved 40-50% of the total encoding time with 0.1~1.5% rate decrease in average, at the expense of only 0.015 distortion in average for non-ROI.

Figures (12), (13), and (14) show the examples of RD and the complexity curves of sequences “ranna” (class A), “rahsa” (Class B), and “rals” (Class C) for IPPP sequences. From these figures, one can see that the proposed fast decision scheme gives better rate-distortion performance while providing a speed-up. In these figures, the RDO, Pan’s method, improved Pan’s method, and four forms of fast proposed methods (Fast intra + Fast inter + Fast multi reference frames+ROI) are compared (see Table 4). Figures15~7 show examples of classification results and final mode decision for “ranna”, “rahsa” and “rals”’ sequences. In part (c) of these figures, the black blocks are the macroblocks with inter-prediction and part (d) shows the motion vector and intra-prediction mode for each block.

7- Conclusion

In this paper, an adaptive intra- and inter-prediction mode decision algorithm for H.264/AVC standard was integrated using a hybrid scheme that is appropriate for efficient and accurate compression of medical images. The model uses lossless compression in the region of interest, and very high rate, lossy compression in the other regions. In the proposed algorithm, we decreased the encoding time by reducing the number of candidate modes in non-ROI.

In order to evaluate the impact of different parts of the proposed algorithm, they were added step by step and the related experimental results were shown. The experimental results showed that the proposed algorithm has reduced the number of RDO calculations and has improved the compression ratio with respect to the previous algorithms. As the experimental results show, the proposed algorithm can be used for challenging work of intra- and inter-prediction mode decision in the H.264/AVC for medical video encoders with low computational cost.

Table (5): Experimental results for IPPP type sequences, distortion comparison

Video	method	PSNR
Ranna	M1	-0.079
	M2	-0.2
	M3	-0.081
	M4	-0.27
	M5	-0.01
	M6	-0.080
	M7	-.0.015
Rahsa	M1	-0.071
	M2	-0.023
	M3	-0.060
	M4	-0.01
	M5	-0.13
	M6	-0.061
	M7	-0.014
Rals	M1	-0.083
	M2	-0.17
	M3	-0.065
	M4	-0.21
	M5	-0.204
	M6	-0.012
	M7	-0.014

Table (6): Experimental results for IPPP type sequences, rate comparison

Video	Method	Bit
Ranna	M1	1.540
	M2	1.004
	M3	1.210
	M4	0.890
	M5	0.735
	M6	0.789
	M7	0.12
Rahsa	M1	1.001
	M2	0.940
	M3	0.932
	M4	0.0942
	M5	0.954
	M6	0.941
	M7	0.27
Rals	M1	1.902
	M2	0.987
	M3	0.982
	M4	0.941
	M5	0.980
	M6	1.001
	M7	0.941

Table (7): Experimental results for IPPP type sequences, complexity comparison.

video	Method	Time
Ranna	M1	-35.42
	M2	-37.34
	M3	-39.25
	M4	-41.32
	M5	-42.23
	M6	-37.22
	M7	-40.0
Rahsa	M1	-38.24
	M2	-39.25
	M3	-41.50
	M4	-42.02
	M5	-40.34
	M6	-39.34
	M7	-38.02
Rals	M1	-35.26
	M2	-34.25
	M3	-34.25
	M4	-39.50
	M5	-40.45
	M6	-43.47
	M7	-38.1

2) ROI as high motion activities can be detected using conventional motion estimation of H.264. Figure (11) shows ROI detection based on motion activities.

After detection of ROI in a reference frame, the H.264 block matching algorithm based on motion compensation method is used for ROI tracking.

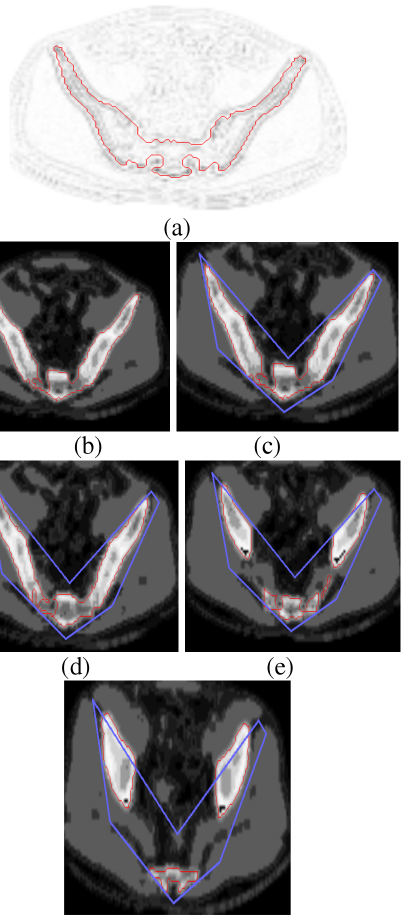
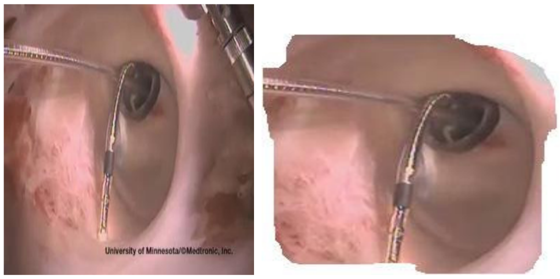

(a)
(b) (c)
(d) (e)
(f)

Fig. (10): Contrast segmentation (a) edge detection using Sobel operator, (b)-(d) ROI tracking in different frames



(a) (b)

Fig. (11): ROI detection based on motion. (a) Original image. (b) ROI region with high motion

5-2- Fast intra- and inter-prediction mode decision based on ROI.

Once the ROI is segmented in each frame, a hybrid intra- and inter- prediction mode decision method is used for image coding. To reduce the encoding computational complexity of intra prediction in H.264, the proposed algorithm selects both I4MB and I16MB prediction for ROI and only I16MB for non-ROI prediction. Also, to achieve fast inter prediction mode decision, background is predicted with coarse MBs (16×16 to 8×8) and ROI are coded with finer MBs(16×16 to 4×4).

6- Experimental results

Our proposed algorithm was implemented into JM7.1, provided by JVT according to the test conditions specified in VCEG-N81 document as listed in Table (3) [21]. Experiments were carried out on the recommended sequences with various quantization parameters for IPPP… type. For IPPP… experiments, the total number of frames was 300 for each sequence, and the period of I-frame was 100. The used test platform was Pentium IV-2.8 GHz with 256 Mbytes RAM. We compared the performance of our proposed algorithm (fast motion estimation + fast intra + fast inter + selective frame + two step adaptive search range decisions+ ROI) with other available approaches. To show the impact of different parts of the algorithm, these parts were added at different steps and the results were analyzed. Thus, the experiments were ordered in seven states as listed in Table (4).

Table (3): Experiment conditions

GOP	IIIII or IPPP
Codec	JM 7.1
MV Search Range	±16
Quantization Parameter	10, 16, 24,28,36,40
Number of References	5
Common Coding Option	Hadamard Transform, CABAC, RDO is enabled
Format	CIF and QCIF
Number of Frames	300

Table (4): Different methods used in our experiments

Category	Intra-prediction (I4MB,I16MB Chroma)	Inter-Prediction	Multi-Reference Algorithm	Early Termination
RDO	RDO	RDO	RDO	Yes
M1	Pan's Method	RDO	RDO	Yes
M2	Improved Pan's method(3.1)	RDO	RDO	Yes
M3	Proposed Alg. (3.2)	RDO	RDO	Yes
M4	Proposed Alg.(3.2)	Proposed Alg. (split/merge) (4.1)	RDO	Yes
M5	Proposed Alg.(3.2)	Proposed Alg. (split/merge) (4.1)	Proposed Alg.(4.2)	Yes
M6	Proposed Alg.(3.2)	Proposed Alg.(2-step Adaptive search range)(4.3)	Proposed Alg.(4.2)	Yes
M7	Proposed Alg.ROI(5.2)	Proposed Alg. ROI(5.2)	Proposed Alg.(4.2)	Yes

4-2- Analysis and complexity reduction of multiple reference frames motion estimation

In H.264/AVC, motion estimation is allowed to search multiple reference frames. Therefore, the required computation is highly increased, and it is in proportion to the number of searched reference frames.

Experimental results show some facts that are used to decide on the number of references frames [20].

1- For QP=20, QP=30 and QP=40, it can be seen that 65%, 79%, and 95%, of macroblocks need only one reference frame, respectively. Therefore, we should proceed the block matching process from the nearest reference frame to the farthest reference frame.

2- Another interesting point is that low bitrate cases are more likely to have the best reference frames close to the current frame than higher bitrate cases.

3- We can see that for QP=20, there are 59.84%, 05.00%, 04.88%, 28.11%, and 02.17% of the macroblocks selected as P16×16, P16×8, P8×16, P8×8, and intra, respectively, when only one previous frame is searched. For QP=30, there are 75.97%, 05.36%, 05.45%, 11.04%, and 02.18% of the macroblocks selected as P16×16, P16×8, P8×16, P8×8, and intra, respectively. For QP=40, the corresponding percentages are 89.34%, 03.21%, 03.07%, 01.69%, and 02.69%.

4- In H.264/AVC, the SKIP mode is a special case of P16×16. The percentages of SKIP macroblocks after searching one reference frame are 44.57%, 62.69%, and 79.14% for QP=20, 30, and 40, respectively.

This result shows that a large percent of MBs are coded as 16×16 or skip mode that use only one reference frame, while for large QP this fact is amplified. According to these observations, in the following, we have listed the steps for each macroblock to check whether it is necessary to search the next reference frame at the end of each reference frame loop.

1- After the prediction procedure, residues are transformed, quantized, and entropy coded. If we face the situation for which the transformed and quantized coefficients are very close to zero in the first reference frame, stop the matching process for the remaining frames.

2- Calculate the sum of absolute transform difference (SATD). If it is less than a threshold (THSATD), stop the searching process.

3- If the best reference frame is the previous frame and the best motion vector is the same as that of the SKIP mode or 16×16 mode and QP is larger than a threshold (THQP), the multiple reference frames loop will be early terminated. The determination of THQP is empirically obtained in [20]. In the proposed method 76%-96% of computations for searching unnecessary reference frames can be avoided.

Also, similar to intra-prediction we used an early termination technique based on early detection of zero blocks.

4-3- Adaptive search range decision algorithm (proposed algorithm)

In reference software JM [7], the fixed search range window size equal to 16 is used for motion estimation. This full search range is inefficient and increases the computational complexity. Therefore, a two-step adaptive search range decision algorithm is presented. In the first step, we have adopted the length of MV obtained in the first reference frame as the maximum size of the search window for the motion estimation in the second reference frame. And, the length of MV in the second reference frame is selected as the maximum size of the search window in the third reference frame, and so forth. This step of the proposed algorithm reduces the SW size for 4 reference frames, and for the first reference frame a full search is applied. In the second step, an adaptive search range decision method is presented that determines the search window according to the MVs of previous MBs. The primary search window size is selected as the average of MVs of the encoded MBs immediately on left, above, above the left, and above the right of current MB, computed as

$$SWR.x=(MVA.x+MVL.x+MVAR.x+MVAL.x)/4$$
$$SWR.y=(MVA.y+MVL.y+MVAR.y+MVAL.y)/4 \quad (9)$$

where SWR.x, MVAx, MVL.x, MVAR.x, and MVL.x denote the search window, motion vector above, motion vector left, motion vector above the right, and motion vector above the left, in the x axis, respectively. Similarly y stands for the y axis. By increasing the search window size around the computed SW, the speed of motion estimation can be improved. The increment of the SW size is stopped when the block matching has not improved with respect to the primary computed SW size.

5- Region of interest (proposed algorithm)

Region of interest (ROI) video coding provides higher quality in ROI, but poorer quality in the background, for a given total bit rate (TBR). ROI coding, where regions of greater diagnostic importance to the medical expert are coded with greater quality than the background, is a useful tool. In this paper, first the input images are segmented into the object of interest and background using a series of morphological operations and a chain code-based shape coding is used to code the ROI's shape information. After ROI segmentation, we propose a method for achieving high quality within a diagnostically important ROI using finer MBs sizes and minimum quantization parameter (QP=21). Also, non-ROI are compressed using coarse MBs sizes and maximum quantization parameters (QP=51).

5-1- ROI segmentation

In this paper we present two major contributions for ROI segmentation. 1) A new block based method that exploits the ROI using contrast difference in the image. In this case object to be segmented differs greatly in contrast from the back ground image. Changes in contrast can detected by operators that calculate the gradient of an image. This paper calculates the gradient of an image by sobel operator ,which creates abinary mask using a user-specified threshold value .we determine a threshold value using the graythresh function in matlab. The binary gradient mask is dilated using the vertical structuring element followed by the horizontal structuring element. The dilated gradient mask completes the outline of the ROI. But there are still holes in the interior of ROI, that are filled with imfill function. Two different regions are presented in figure (10) are detected by this method.

Table (2): Number of candidate modes

Block Size	RDO	Pan's Method	ProposedMethod (min)	Proposed Method (max)
4×4 (Y)	9	4	2	3
16×16 (Y)	4	2	1	2
8×8 (U/V)	4	3 or 2	1	2

3-2- Fast intra-prediction mode selection based on statistical properties of adjacent MBs and reference pixels (proposed algorithm)

The method proposed in this section analyzes the characteristic of reference pixels and uses the similarity between adjacent MBs, while the improved Pan's method analyzes the characteristic of the 4×4 block itself. As a result, the combination of these three methods can achieve better results. The proposed algorithm is as follow.

1- Find the maximum value of EDH. Denote the corresponding mode by M1.

2- If the modes for one of the top or left blocks are M1, then choose M1 as the best candidate mode for the current block. Go to step 7.

3- For 4×4 luma block, compute the mean of absolute difference (MAD) of its reference pixels. If this value is smaller than a predetermined threshold, select M1. Go to step 7.

This result is yielded from the fact that if the similarity of reference pixels of a block is high, the difference between different prediction modes will be very small. For this case, it is not necessary to check all 9 prediction modes, but only one prediction mode is enough [19].

4- If the mean of absolute difference of horizontal references (MADH) is less than a threshold and M1 is a member of the set {mode 0, mode 3, mode 7}, select M1. Go to step 7.

5- Also, if the mean of absolute difference of vertical references (MADV) is less than a threshold and M1 is a member the set of {mode 1, mode 8}, select M1. Go to step 7.

It is obvious that if the similarity of horizontal reference pixels of a block is high, the difference between prediction results obtained with prediction modes 0, 3 and 7 will be very small. Also, if the similarity of vertical reference pixels of a block is high, the similarity between modes 1 and 8 is high.

6- If all above conditions are unsatisfied, use the improved Pan's method (explained in Section 3.1).

7- Terminate.

As such, in the worst case only three different 4×4 intra-mode costs will be evaluated. Also, for I16MB and 8×8 chroma blocks the improved Pan's method is used.

To increase the speed of the algorithm, we used the early termination of RDO calculations for all proposed algorithms as in [1].

4- Proposed fast inter-predicton mode decision method

The motivated facts about the inter-prediction are summarized below.

I. The block with high motion activities, instead of high textural details, can be better coded using smaller block sizes, while the block with less motion activities can be more efficiently encoded using larger block sizes.

II. It is observed that in natural image sequences, when the video objects move, the different parts of the video objects move in a similar manner. Then, homogenous regions are encoded using 16×16 block sizes while non-homogenous regions are encoded using smaller block sizes.

III. Video background is not homogeneous, but because of temporal redundancy, it is coded using 16×16 block sizes.

4-1- Fast inter-prediction mode decision using split and merge procedure(proposed algorithm)

The split procedure partitions the MB into variable size blocks using a quad-tree approach. In this method, a macroblock is divided into equal-size quarters. Then, using the similarities of motion vectors of adjacent blocks we will show how to merge the sub-blocks for quarter divisions. The proposed algorithm is summarized as follows.

1. Subtract the current frame from its previous frame and for any 16×16 MB compute

$$\begin{aligned} &N_i = \text{number of pixels belong to the set MBi} \\ &N_{im} = \text{number of nonzero pixels in difference MB} \end{aligned} \qquad (9)$$

2- If N_{im}/N_i is smaller than or equal to a predetermined threshold, choose direct mode as the final macroblock type.

3- Otherwise, split the block into four 8×8 blocks and conduct a new iteration of block matching for each of these four descending blocks.

4- If motion vectors of 8×8 sub-block are equal or three sub-block MV are the same and the forth unequal MV only differ by one quarter-pixel distance, choose mode 1(16×16). Terminate.

5- If MV0=MV1 and MV2=MV3 (see Figure 9), choose 8×16. Terminate.

6- If MV0=MV2 and MV1=MV3, choose 16×8. Terminate.

7- Repeat Steps 2 to 4 for each 8×8 blocks, except that sub-blocks are 4×4.

8- Terminate.

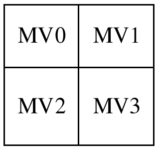


Fig. (9): MB division

The mode selection methodology employed in this paper is as follows. For I4MB, I16MB, and P the proposed algorithms are used to extract the best mode among the related category and at last RDO is used to extract the final mode.

II. The prediction modes of each block are correlated with those of neighboring 4×4 luminance blocks.
III. Normally, pixels along the direction of local edges have similar values. Therefore, a good prediction can be achieved by predicting the pixels using the neighboring pixels that lie in the same edge directions.
IV The optimal mode (found by a full-search) and other "good" (second or third best) modes are most likely to have similar directions.
V. The directional features of 4×4 blocks can be preserved roughly after down-sampling.
VI. Experimental results show that the reference pixels of a 4×4 luma block are likely to be similar to each other [22].

Based on these observations, we have proposed a fast intra-prediction mode selection algorithm. In this section, some new ideas are combined with the fast mode selection algorithm introduced in [1, 6, 7] to improve their efficiency.

3-1- Improved Pan's method for fast decision of I4MB

Pan et al. presented a fast mode selection for intra-prediction method in [1] in which the average edge direction of a given block is measured. The Sobel operators are first used to obtain the directional vector of each pixel in a block by

$$\overline{D}_{i,j} = \{dx_{i,j}, dy_{i,j}\} \tag{3}$$

where the Sobel operator are defined by

$$\begin{aligned} dx_{i,j} &= P_{i-1,j+1} + 2\times P_{i,j+1} + P_{i+1,j+1} - P_{i-1,j-1} - 2\times P_{i,j-1} - P_{i+1,j-1} \\ dy_{i,j} &= P_{i+1,j-1} + 2\times P_{i+1,j} + P_{i+1,j+1} - P_{i-1,j-1} - 2\times P_{i-1,j} - P_{i-1,j+1} \end{aligned} \tag{4}$$

The amplitude and angle of each edge vector can be calculated using

$$\mathrm{Amp}(\overline{D}_{i,j}) = |dx_{i,j}| + |dy_{i,j}| \tag{5}$$

and

$$\mathrm{Ang}(\overline{D}_{i,j}) = \frac{180^\circ}{\pi} \times \arctan\left(\frac{dy_{i,j}}{dx_{i,j}}\right) \tag{6}$$

where Ang(.) is fitted into one of the 8 modes.

Then, the edge direction histogram (EDH) of the block is found (that indicates the number of pixels with similar edge directions). Therefore, the cell *k* with the maximum amount indicates that there is a strong edge along that direction in the block and thus is assigned as the dominant block direction. Figure 8 shows the EDH of the image shown in Figure (7).

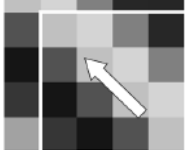


Fig. (7): An example of 4×4 edge patterns and their dominant direction

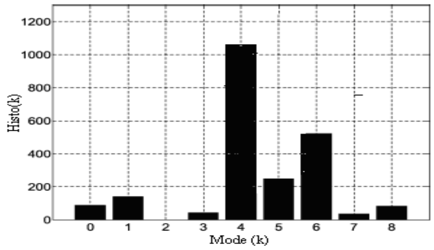


Fig. (8): Edge direction histogram of Fig. (7)

In the Pan's method, for I4MB there are 4 modes (1 DC, 1 from maximum amplitude of EDH, and its 2 neighbors) with 2 modes (1 DC mode and 1 directional) for each 16×16 luma block and 8×8 chroma block. Here, to improve the Pan's method, we will eliminate the DC mode from the candidates if the direction of the block is obvious, and otherwise, we choose only DC mode. To check whether the DC of the block is clear or not, the diff value, given in Eq. (7), is computed to check whether it is smaller than a predetermined threshold or not, using

$$\begin{aligned} \mathrm{diff} &= \sum_{i=0}^{i=15} |avg - p_i| \\ avg &= (\sum_{i=0}^{15} p_i + 8) >> 4 \end{aligned} \tag{7}$$

The improved Pan's method is proposed as follows:
1- Find the maximum value of the edge directional histogram H. Denote the corresponding mode by M1.
2- If diff > T, carry out the RDO procedure for 3 modes at the most (M1 and its two neighbors).
3- Else, if diff <T, carry out the RDO procedure for two candidate modes at the most. The DC with maximum EDH (M1).
4- For I16MB, based on the same observation as above, and after down-sampling by a factor of 2, if diff1 > T1, consider only the primary prediction mode decided by edge direction histogram as a candidate for the best prediction mode. The diff1 in this case is computed by

$$\begin{aligned} \mathrm{diff1} &= \sum_{i=0}^{i=64} |avg - p_i| \\ avg &= (\sum_{i=0}^{64} p_i + 32) >> 6 \end{aligned} \tag{8}$$

5- If diff1 < T1, choose the maximum prediction mode and the DC mode. Extract the maximum prediction mode as I4MB but with DC and only 3 directions of intra16.
6- For 8×8 chroma block, and after down-sampling by a factor of 2, use the same procedure as I16MB but by using Eq. (7).

Pan's method can reduce RDO calculation from 592 to 132 times. The number of candidate modes and the RDO calculation in the worst and the best cases are listed in Table (2). This table summarizes the number of candidates selected for RDO calculation based on EDH. As can be seen from this table, the encoder with the fast mode decision algorithm needs to perform only 33 or 100 RDO calculations, which are much less than that of Pan's method (132) and current H.264 video coding (592).

A separate motion vector is required for each partition or sub-partition. Each motion vector must be coded and transmitted; in addition, the choice of partition (s) must be encoded in the compressed bitstream. Choosing a large partition size (e.g., 16×16, 16×8, 8×16) means that a small number of bits are required to indicate the choice of motion vector (s) and the type of partition; however, the motion compensated residual may contain a significant amount of energy in frame areas with high details. Choosing a small partition size (e.g., 8×4, 4×4, etc.) may give a lower energy residual after motion compensation, but requires a larger number of bits to signal the motion vectors and the choice of partition (s). The choice of partition size therefore has a significant impact on compression (a small partition size may be beneficial for detailed areas).

H.264/AVC as an enhanced reference picture selection as H.263++ enables efficient coding by allowing an encoder to select (for motion compensation purposes) among a large number of pictures that have been decoded and stored in the decoder. This paper describes a new method to achieve fast inter prediction mode decision that is based on coarse MBs (16×16 to 8×8) of background region and finer MBs (16×16 to 4×4) for ROI.

2-3- Quantization parameter and rate control in H.264

A rate control algorithm dynamically adjusts encoder parameters to achieve a target bitrate(TBR). In particular, the quantization parameter QP regulates the TBR. When QP is very small, almost all that detail is retained. As QP is increased, some of that detail is aggregated so that the bit rate drops – but at the price of some increase in distortion and some loss of quality. In H.264/AVC residuals are transformed into the spatial frequency domain by an integer transform, named H transform, that approximates the familiar Discrete Cosine Transform (DCT). The Quantization Parameter determines the step size for associating the transformed coefficients with a finite set of steps. Large values of QP represent big steps that crudely approximate the spatial transform, so that most of the signal can be captured by only a few coefficients. Small values of QP more accurately approximate the block's spatial frequency spectrum, but at the cost of more bits. In this paper, to regulate the TBR, the MBs of the background region are encoded with the maximum quantization parameter allowed by H.264 (QP=51), in order to maximize the number of null coefficients.

2-4- RDO procedure

For I-frames, all MBs are predicted as Intra and for P-frames all MBs are predicted as intra or inter. H.264/AVC encoder encodes the best mode using all mode combinations of luma and chroma and chooses the one that gives the best RDO performance. The best prediction mode among all possible intra-/inter-predictive modes is achieved by minimizing Equ. (1),

$$\begin{aligned} &J(s,c,MODE|QP,\lambda_{mode}) \\ &= SSD(s,c,MODE|QP)+\lambda_{mode}R(s,c,MODE|QP) \end{aligned} \tag{1}$$

where QP is the macroblock quantization parameter, λ =0.85×2(QP−12)/3 is the Lagrangian multiplier, and MODE indicates different prediction modes of a macro block. R (.) represents the rate, i.e., the number of bits associated with chosen MODE.

where SSD is defined as:

$$\begin{aligned} &SSD(s,c,MODE|QP)= \\ &\sum_{x=1,y=1}(S_y(x,y)-c_y(x,y,MODE|QP))^2 \\ &+\sum_{x=1,y=1}(S_v(x,y)-C_v(x,y,MODE|QP))^2 \\ &+\sum_{x=1,y=1}(S_u(x,y)-C_u(x,y,MODE|QP))^2 \end{aligned} \tag{2}$$

According to the RDO procedure of intra-prediction in H.264/AVC, the number of mode combinations for luma and chroma blocks in a macroblock is N8× (16×N4 + N16), where N8, N4, and N16, denote the number of modes for 8×8 chroma blocks, and 4×4 and 16×16 luma blocks, respectively [5]. Also, according to the RDO procedure of inter-prediction, for M block modes, N reference frames, and ±W search range, $M.N.(2W+1)^2$ positions should be tested for a single reference frame and a single block mode. This makes the complexity of the encoder extremely high. In order to reduce the encoding complexity with little RD performance degradation, fast intra- and inter-prediction mode decision methods are proposed.

3- Proposed fast intra-prediction mode decision methods

This section presents a new fast intra-prediction algorithm. The proposed method is based on several facts that we have observed from the statistics of different sequences as follows:

I. Figure (6) shows the total number of 4×4 and 16×16 intra-coded macroblocks at different quantization parameters (QPs). As can be seen from this figure, fast detection of 4×4 intra-prediction mode can significantly improve the encoding speed at low QPs, while 16×16 intra-prediction can improve the speed at large QPs.

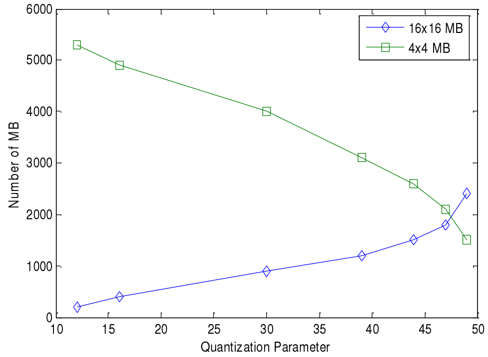


Fig. (6): Number of 4×4 and 16×16 intra-coded macroblocks for different quantization parameters

This paper aims at applying H.264 in medical video compression and improving the previous two-step adaptive method using diagnostically region of interest. The proposed encoding scheme is suitable for medical applications, where higher resolution and higher quality ROI is a useful functionality for detail analysis. We have verified the different parts of the proposed algorithm step by step by implementing it on the JM7.1 [15] reference software and have compared its performance with previous and other available fast algorithms. Experimental results on medical videos show that the proposed method with a high compression performance reduces the encoding cost up to 40% with a negligible loss in the reconstructed video quality.

The remaining parts of the paper are organized as follows. The next section briefly reviews the H.264 standard. Section 3 and 4 present the proposed fast intra- and inter-prediction mode decision algorithms, respectively. Section 5 describes the proposed ROI based system in detail. Experimental results are given in Section 6 and finally Section 7 concludes the paper.

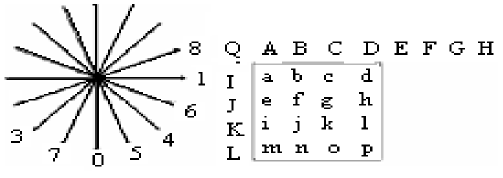
2- H.264/AVC overview

In common with earlier standards, H.264/AVC does not define the encoder, but defines the syntax of an encoded video bitstream together with the method of decoding the bitstream [16, 17]. The codec combines intra-picture prediction with inter-picture prediction to exploit the spatial and temporal redundancy, respectively. The proposed algorithms in this work are based on coarse MBs for intra- and inter-prediction of the non-ROI and finer MBs of ROI. Also, the MBs of non-ROI, are encoded with the maximum quantization scale allowed by H.264 (QP=51) in order to maximize the number of null coefficients. Therefore Intra- and Inter- prediction mode decision methods and quantization parameters are the key factors of proposed algorithms that are described in the following subsections.

2-1- Intra-prediction mode decision

Intra-prediction is based on the observation that adjacent macroblocks tend to have similar properties. Prediction may be formed for each 4×4 luma block (I4MB), 16×16 luma MB (I16MB), and 8×8 chroma block.

For prediction of 4×4 luminance blocks, the 9 directional modes consist of a DC prediction (Mode 2) and 8 directional modes (labeled 0, 1, 3, 4, 5, 6, 7, and 8) as shown in Figure 3(a). In Figure 3(b), the block (values of pixels “a” to “p”) is to be predicted using A to Q pixel values.



(a) (b)

Fig. (3): (a) Intra-prediction modes for 4×4 luminance blocks. (b) Labeling of prediction samples

The DC prediction (mode 2) is useful for those blocks with little or no local activities, the other modes (1-8) may only be used if all required prediction samples are available. For regions with less spatial details (i.e., flat regions), H.264/AVC supports 16×16 intra-coding; in which one of four prediction modes (DC, vertical, horiz–ontal, and planar) is chosen for prediction of the entire luminance components of the macroblock as shown in Figure 4 [18].

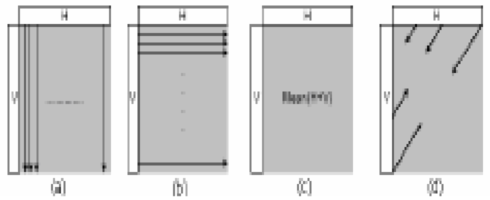


Fig. (4): Intra 16×16 prediction modes. (a) Mode 0 (vertical). (b) Mode 1 (Horizontal). (c) Mode 2(DC). (d) Mode 3 (plane)

H.264/AVC supports four chroma prediction modes for 8×8 chrominance blocks, similar to that of the I16MB prediction, except that the order of mode numbers is different: DC (Mode 0), horizontal (Mode 1), vertical (Mode 2), and plane (Mode 3). This paper selects I4MB and I16MB for ROI and only I16MB for non-ROI to reduce the computational complexity of intra prediction in H.264.

2-2- Inter-prediction mode decision

Inter-prediction is based on using motion estimation and compensation to take advantage of temporal redundancies that exist between successive frames. The important differences from earlier standards include the support for a range of block sizes (down to 4×4), multiple reference frames, and fine sub-pixel motion vectors (1/4 pixel in the luma component).

H.264/AVC supports motion compensation block sizes ranging from 16×16 to 4×4 luminance samples with many options between the two. The luminance component of each macroblock (16×16 samples) may be split up in 4 ways as 16×16, 16×8, 8×16 or 8×8. If the 8×8 mode is chosen, each of the four 8×8 macroblock partitions within the macroblock may be split in a further 4 ways as 8×8, 8×4, 4×8 or 4×4. The four macroblock type sizes and four macroblock subtype sizes are shown in Figure (5). These partitions and sub-partitions give rise to a large number of possible combinations within each macroblock.

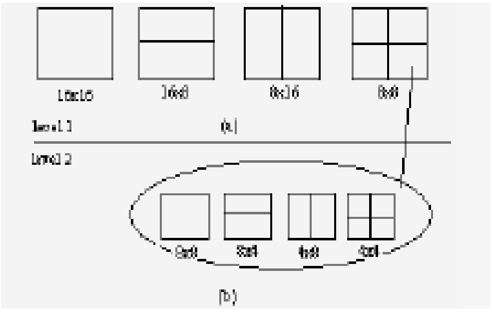


Fig. (5): Variable block size in H.264, (a) sizes for a MB type, (b) sizes for a sub MB type

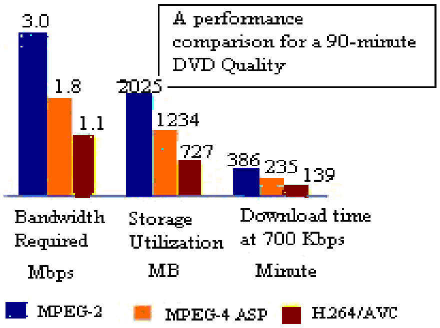


Fig. (1): Performance comparison of different video coding standards

To achieve an outstanding coding performance, H.264–/AVC employs several powerful coding techniques such as directional prediction of intra-coded blocks, interprediction with variable block-size motion compensation, multi-reference frame motion estimation, motion vectors with quarter-pel accuracy, in-loop deblocking filter, 4×4 integer transform, and the forth. According to these new features, the encoder computational complexity has extremely increased compared to previous standards. However, theses functionalities has made H.264/AVC difficult for real-time telemedicine video systems. Thus, until now, the reduction of its complexity is a challenging task. The improvement in coding performance comes mainly from the intra- and inter-prediction part. H.264/AVC employs the Lagrange RDO method to achieve the best coding mode of intra-and inter-prediction with highest coding efficiency. The RDO technique requires a lot of comput–ations since it tests the encoding process with all possible coding modes of intra- and inter-coding, and calculates their RD costs to choose the mode having the minimum required bitrate. The reference software of H.264/AVC, JM [15], adopts a full search for both motion estimation and intra-prediction. The run-time percentage of each function is shown in Figure 2 [20]. As shown in this figure, motion estimation is the most computationally intensive part. The instruction profile of the reference software shows that real-time encoding of CIF 30Hz video (baseline options, search range [-16.75, +16.75], five reference frames) requires 314,994 million instructions per second and memory access of 471,299 MBytes/sec.

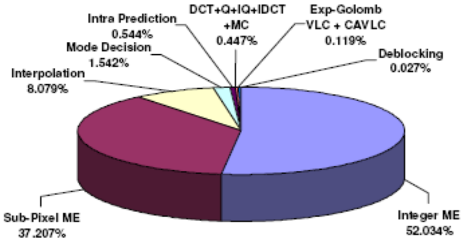


Fig. (2): Run-time percentages of functional blocks in H.264/AVC baseline encoder [20]

Also, H.264/AVC offers a rich set of prediction patterns for intra-prediction (i.e., nine prediction modes for 4×4 luma blocks and four prediction modes for 16×16 luma blocks). The intra-prediction mode decision is very complex and the number of computing RD cost values for luma and chroma of a macroblock is 592 [5]. Thus, the computational burden of these types of brute force-searching algorithm is far more demanding than any existing video coding algorithm.

The challenging topic of fast mode selection for intra- and inter-prediction mode decision for medical video is considered in this paper. To reduce the computational complexity, many algorithms have been proposed. For fast inter-prediction mode decision, the early termination technique [23] reduces the number of potential prediction modes. In [24], a classification method is proposed. Recently fast inter-mode selection algorithms were proposed in [25, 26, 27] to alleviate the encoder complexity due to the motion estimation and inter-mode decisions. In [28], a fast inter-prediction mode decision based on pre-encoding process is presented.

Fast intra-mode decision algorithms using edge detection histogram and local edge detection are proposed in [1, 6, 7]. However, their preprocessing stages still consume a coding time to detect the edge direction and to classify it into a limited direction. Also, there exist fast algorithms to select the optimal intra-prediction mode using simple directional masks in [8] with saving time of 70%, and statistical-based methods in [9] with saving time of 45%. Another fast intra-mode decision scheme is proposed in [10], where the encoding speed is approximately 30% faster than that of the RDO method. A new fast intra-prediction algorithm based on macroblock properties (FIPAMP) is presented in [11]. This algorithm achieves 10% to 40% of computation reduction while maintaining similar PSNR and bitrate performance of H.264/AVC codes. In [12], an efficient intra-prediction (EIP) algorithm based on early termination, selective computation of highly probable modes, and partial computation of the cost function is presented. Also, an improved cost func–tion to improve the coding performance is proposed in [13]. Also, in [14], a fast algorithm based on local edge information is obtained by calculating edge feature parameters. In [31], an ROI-DCT algorithm that uses more DCT coefficients in ROI, was proposed. Also, many schemes have been proposed [32]-[35], most of which use some common shapes or fixed shapes for ROI. In [29], the authors proposed a new approach for both inter-and intra-mode size selection. For intra-prediction mode decision, we improved Pan’s method [1, 6, 7], and for inter-prediction mode decision, the split/merge procedure was used. Also, an effective method for accelerating the multiple reference frames ME without significant loss of video quality was proposed that was based on analyzing the available information obtained from previously processed frames. In [30], the authors proposed a two-step adaptive search range decision algorithm to improve the proposed algorithm in [29]. In the first step, the proposed adaptive algorithm determines the search range for current MB according to MVs of previous MBs. In the second step, the length of MV obtained in the current reference frame is considered as the maximum size of the search window for motion estimation in the previous reference frame.

Fast Intra and Inter Prediction Mode Decision of H.264/AVC for Medical Image Compression Based on Region of Interest

Mehdi Jafari[(1)] - **Homayoun Mahdavi-Nasab**[(2)] - **Shohreh Kasaei**[(3)]
(1) Assistant Prof. - Department of Electrical Engineering, Kerman Branch, Islamic Azad University
(2) Assistant Prof. - Department of Electrical Engineering, Najafabad Branch, Islamic Azad University
(3) Prof. - Department of Computer Engineering, Sharif University of Technology

Recive: Summer 2009 Accept: Summer 2010

Abstract: This paper aims at applying H.264 in medical video compression applications and improving the H.264 Compression performance with better perceptual quality and low coding complexity. In order to achieve higher compression of medical video, while maintaining high image quality in the region of interest, with low coding complexity, here we propose a new model using H.264/AVC that uses lossless compression in the region of interest, and very high rate, lossy compression in other regions. This paper proposes a new method to achieve fast intra and inter prediction mode decision that is based on coarse macroblocks for intra and inter prediction mode decision of the background region and finer macroblocks for region of interest. Also the macroblocks of the background region are encoded with the maximum quantization parameter allowed by H.264/AVC in order to maximize the number of null coefficients. Experimental results show that the proposed algorithm achieves a higher compression rate on medical videos with a higher quality of region of interest with low coding complexity when compared to our previous algorithm and other standard algorithms reported in the literature.

Index Terms: H.264/AVC, intra- and inter-prediction, medical video compression, adaptive search range decision, Region of interest.

1- Introduction

In recent years, easy access to medical videos, such as computed tomography (CT), magnetic resonance imaging (MRI), echocardiography, positron emission tomography (PET), and so on, provides doctors with better capability of analyzing and diagnosing patients' conditions. However, the growth rate of medical video data is likely to surpass the rate of decrease in storage and bandwidth costs. As recent medical video applications are growing rapidly, medical video compression requires higher performance as well as new features. In medical video, it is a very challenging task to compress videos with a high quality, high compression ratio and low computational complexity for real-time applications. Today different 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 image/video sequences including echocardiography and CT. The recommended compression methods in DICOM are JPEG2000 and MPEG-2. To address the needs of different applications, both the international telecommunication union (ITU) and the international organization for standardization (ISO) have released standards for still image and video compression such as H.261, H.263, MPEG-1, MPEG-2 and MPEG-4. In [7] and [8] MPEG-4 has been used for the compression of ultrasound and echocardiography sequen–ces, respectively, for both archiving and transmission purposes.

The newest video coding standard is developed by the joint of video teams of ISO/IEC MPEG and ITU_T VCEG as the international standard 14496-10 (MPEG-4 part 10) advanced video coding (AVC) [1, 2]. Today H.264/AVC has been very successful in many applications including digital media storage, video streaming, TV and so on. The main motivation for the research reported in this paper is improvement and applying H.264 in medical video compression. H.264/AVC has gained more and more attention; mainly due to its high coding efficiency, minor increase in decoder complexity compared to existing standards, adaptation to delay constraints, error robustness, and network friendliness [1, 2]. Table 1[3] and figure 1[4] show the performance comparisons using MPEG-2, MPEG-4 (ASP), and H.264/AVC.

Table (1): Average bit-rate reduction compared to prior coding schemes

Standards	MPEG-4/ASP	H.263/HLP	MPEG-2
H.264/AVC	38.62%	48.80%	64.46%
MPEG-4/ASP	----	16.65%	42.95%
H.263/HLP	------	-----	30.615%