http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (19): 5386-5390, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.5386.5390
© 2013 Asian Network for Scientific Information

A Coding Information-Based Video ROI Detection for Fast Layered Video Coding Method

Pengyu Liu and Kebin Jia
School of Electronic Information and Control Engineering, Beijing University of Technology,
100124, Beijing, China

Abstract: A fast and hierarchical video coding method using lower coding information is proposed in this study. Firstly, a visual region of interest (VROI) detection algorithm is given based on the interdependency between in the coding information (inter-frame prediction mode, motion vector, etc.) and visual perception characteristics After that, visual interested priority is set and the hierarchical video coding method is developed according to the setting results. At last, different visual perception characteristic saliencies are the key to constitute the low complexity video coding framework. The simulation results show that the proposed hierarchical video coding method effectively alleviates the contradiction between complexity and accuracy. Compared with H.264/AVC (JM17.0), it reduces nearly 80% video coding time approximately and maintains a good video image quality and improves video coding performance significantly.

Key words: Fast video coding, visual perception characteristics, coding information, visual region of interest

INTRODUCTION

How to achieve low complexity, high quality and high compression-ratio has become one of the hottest research areas in signal and information processing. Up to now, many scholars carried out a lot of work at fast video coding algorithm or visual perception analysis but few of them combine the two kinds of coding technique in a video coding framework to jointly optimize the performance of video coding (Narwaria *et al*., 2012; Yao *et al*., 2012).

Tsapatsoulis *et al*. (2007) detected the region of interest by color, brightness, direction and complexion but they ignored the motion visual characteristics. Wang *et al*. (2006) built a model of visual attention to extract region of interest by motion, brightness, face, text and other visual characteristics. Fang *et al*. (2012) proposed that the region of interest obtains method based on wavelet transform or in the compressed domain. Considering the region of interest, the above video coding methods based on Human Visual Systems (HVS) are lack of computing resource allocation optimization and the additional computational complexity which was caused by visual perception analysis is neglected also.

Kim *et al*. (2006) reduced the loss of rate-distortion performance under limited computing resource by controlling the motion estimation search points. Saponara *et al*. (2006) adjusted the numbers of reference frames, the prediction mode and the motion estimation search range according to the Sum of Difference (SAD).

Su *et al*. (2007) set the parameters of motion estimation and mode decision to achieve a self-adaptive computational complexity controller. This kind of algorithm ignores the differences of the perception in various video scenes that using the same coding algorithm for all encoding contents in video.

The study in this study found that there is important theoretical significance in using visual perception principle to optimize the computing resource allocation. The proposed method in this study, optimizes computing resource allocation more effectively by using visual perception principle and then proposes an efficient hierarchical video coding algorithm based on visual perception characteristics.

VISUAL PERCEPTION CHARACTERISTICS DETECTION

Detection of temporal visual saliency region: On the real condition, non-zero motion vector random noise in background will appear. The horizontal displacement of camera will bring out global motion vectors. It is necessary to develop appropriate motion vector detection to filter motion vector random noise interference and translational motion vector error.

Motion vector random noise filtering: The types of motion vectors are divided into three categories: horizontal motion, vertical motion and oblique motion. c(x, y) represents the encoded macro-block which has the same position in previous frame. Define:

Corresponding Author: Pengyu Liu, School of Electronic Information and Control Engineering, Beijing University of Technology, 100124, Beijing, China

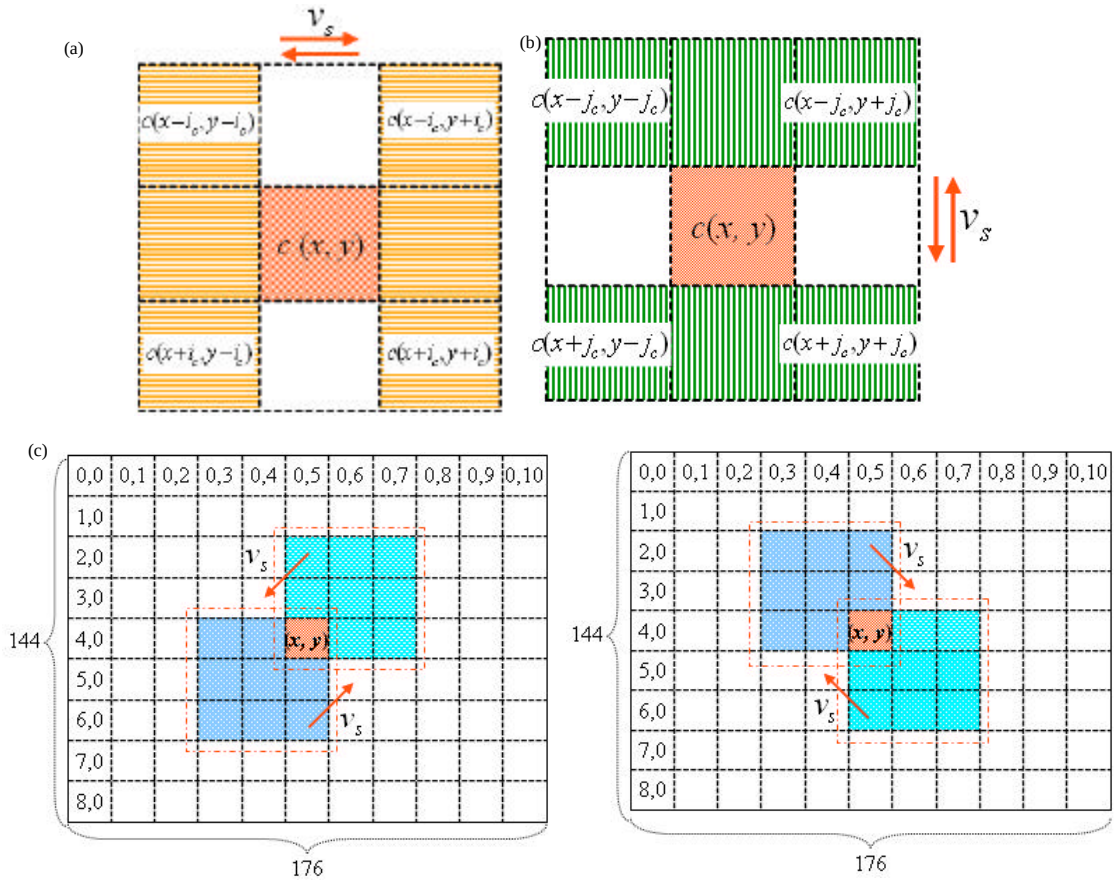

Fig. 1(a-c): Schematic diagram of reference region based on the motion vector direction (a) Horizontal movement reference region, (b) Vertical movement reference region and (c) Oblique movement reference region

$$i_c = \text{int}(|v_s| / w_s + 1),\ j_c = \text{int}(|v_s| / h_s + 1)$$

$|v_s|$ represents the amplitude of motion vector, w_s and h_s represent the width and height of the encoded macro-block, respectively.

The reference region (C_{rr}) is defined as shown in Fig. 1.

Horizontal movement C_{rr}: Take c(x, y) as the initial search point, move i_c macro-blocks horizontally opposite to $|v_s|$, then vertically upward and downward extend i_c macro-blocks, therefore, obtain the vertical coordinate pmv = MV_{pred_time}, C(x+i_c, y-i_c) or c(x-i_c, y+i_c), c(x+i_c, y+i_c).

Vertical movement C_{rr}: Take c(x, y) as the initial search point, move j_c macro-blocks vertically opposite to $|v_s|$, then horizontally leftward and rightward extend j_c macro-blocks, obtain the horizontal coordinate c(x-j_c, y-j_c), c(x-j_c, y+j_c) or c(x+j_c, y-j_c), c(x+j_c, y+j_c).

Oblique movement C_{rr}: Take c(x, y) as the initial search point, move i_c macro-blocks opposite to the horizontal component of $|v_s|$ and move j_c macro-blocks opposite to the vertical component of $|v_s|$, then get a rectangular region consisted of c(x, y), c(x-i_c, y), c(x, y-j_c) and c(x-i_c, y-j_c).

The method of detect motion vector random noise is defined as Eq. 1:

$$M_1(x, y) = \begin{cases} 3, & \text{if } |\bar{V}| = 0 \\ 2, & \text{else if } |v_s| \geq |\bar{V}_{rr}| \\ M_2(x, y) & \text{else if } |v_s| < |\bar{V}_{rr}| \end{cases} \quad (1)$$

In Eq. 1, (x, y) represents the coordinates of the current macro-block. $\bar{V}_{rr}$ represents the mean motion vector in C_{rr}.

If $|\bar{V}_{rr}| = 0$, v_s is set to 0 and $M_1(x, y)$ is set to 3 which means v_s is caused by motion vector random noise.

If $|v_s| \geq |\bar{V}_{rr}|$, $M_1(x, y)$ is set to 2 which means that the current macro-block has more saliency motion characteristics compared with neighbored macro-blocks and it belongs to foreground dynamic region.

Otherwise, $M_1(x, y)$ is set to $M_2(x, y)$ and then the motion vector is going to be detected whether is translational or not. The translational motion vector detection can distinguish the macro-block belonging to background region or foreground translational region which has the similar motion characteristics in neighbored macro-blocks.

Translational motion vector detection:

$$M_2(x, y) = \begin{cases} 1, \text{ if } SAD_{(x, y)} = \sum_{i=0, j=0}^{M, N} |S(i, j) - c(i, j)| \geq \bar{S}_c \\ 0, \quad \text{else} \end{cases} \quad (2)$$

In Eq. 2, (x, y) represents the coordinates of the current macro-block, s(i, j) represents the pixel of the current macro-block, c(i, j) represents the pixel of the corresponding macro-block in previous frame, M and N represent the pixels number in the horizontal or vertical direction of the current macro-block, respectively.

In order to reduce the detection error, this study sets up an adaptive dynamic threshold $\bar{S}_c$ to detect the translational motion vector interference. $\bar{S}_c$ is the mean SAD of all macro-blocks which are considered in background at previous frame:

$$\bar{S}_c = \frac{\sum_{x, y \in S_c} SAD_{(x, y)}}{Num} \quad (3)$$

In Eq. 3, S_c represents the background region in previous frame:

$$\sum_{x, y \in S_c} SAD_{(x, y)}$$

represents the sum of the SAD in S_c, Num represents the summation times.

Temporal visual saliency region marking:

$$M_1(x, y) = \begin{cases} 3, & \text{if } |\bar{V}_{rr}| = 0 \\ 2, & \text{else if } |v_s| \geq |\bar{V}_{rr}| \\ 1, & \text{else if } SAD_{(x, y)} \geq \bar{S}_c \\ 0, & \text{else} \end{cases} \quad (4)$$

In Eq. 4, M(x, y) = 3 represents motion vector random noise and after motion vector random noise filtering M(x, y) is set to 0. M(x, y) = 2 represents foreground dynamic region. M(x, y) = 1 represents foreground translational region.

M(x, y) = 0 represents background region.

Detection of spatial visual characteristic region: In this study, prediction mode decision is regarded as the spatial characteristics of visual perception analysis:

$$S(x, y) = \begin{cases} 2, & \text{mod } e_p \in (\text{Intra16} \times 16, \text{Intra4} \times 4) \\ 1, & (\text{mod } e_p \in \text{Intra18}) \text{ or } (\text{mod } e_I \in \text{Intra4} \times 4) \\ 0, & (\text{mod } e_p \in \text{Intra16}) \text{ or } (\text{mod } e_I \in \text{Intra16} \times 16) \end{cases} \quad (5)$$

In Eq. 5, mod e_p represents the predicted mode of the current macro-block in frame P. mod e_I represents the predicted mode of the current macro-block in frame I. Some researches have proved that mode decision is accordant well to visual attention. The macro-blocks choose sub-block prediction modes in intra-frame or inter-frame coding with high probability and attended highly by human eyes when spatial visual characteristic varies intensely or abundant image contents include more moving details. The macro-blocks have been chosen by macro-block prediction mode in intra-frame or inter-frame coding with high probability and attended lowly by human eyes when spatial visual characteristic varies slowly or abundant image contents include smooth movements(Liu *et al*., 2010; Liu *et al*., 2011).

HIERARCHICAL VIDEO CODING SCHEME

Depend on the previous researches for fast mode decision algorithm and fast motion estimation algorithm, the computing resource will be optimized by intra prediction mode decision, inter prediction mode decision, Motion estimation search range and numbers of references. The hierarchical video coding scheme proposed here is developed based on the visual perception characteristic analysis results according to the foregoing paragraphs.

Priority setting for visual region of interest: Based on the abundant video content and human visual selective attention principle, video sequences usually have temporal and spatial characteristics. The priority setting for visual region of interest is defined as Eq. 6:

$$ROI(x, y) = \begin{cases} 3, & ((M(x, y) = 2 \text{ or } M(x, y) = 1) \parallel (S(x, y) = 1)) \text{ or } (S(x, y) = 2) \\ 2, & (M(x, y) = 2 \text{ or } M(x, y) = 1) \parallel (S(x, y) = 0) \\ 1, & (M(x, y) = 0) \parallel (S(x, y) = 1) \\ 0, & (M(x, y) = 0) \parallel (S(x, y) = 0) \end{cases} \quad (6)$$

In Eq .6, ROI(x, y) represents the priority setting for visual region of interest, M(x, y) represents the salient degree of temporal visual characteristic, S(x, y) represents the salient degree of spatial visual characteristic, (x, y) represents the coordinates of the current macro-block.

Table 1: Hierarchical video coding algorithm based on visual perception characteristics

Coding scheme		Intra predition mode decision (Liu *et al.*, 2010)		Inter prediction mode decision (Liu *et al.*, 2011)		ME search range (Liu and Jia, 2011)	No. of reference frames
		Intra 16×16	Intra 4×4	Inter 16	Inter 8		
Frame P	ROI (x, y) = 3	Intra 16×16	Intra4×4	-	Inter 8	Layer 2, 3, 4	5
	ROI (x, y) = 2	-	-	Inter 16	-	Layer 1, 2, 3	3
	ROI (x, y) = 1	-	-	-	Inter 8	Layer 1, 2	2
	ROI (x, y) = 0	-	-	Inter 16	-	Layer 1	1
Frame I	ROI (x, y) = 1	-	Intra4×4	-	-	-	-
	ROI (x, y) = 0	Intra 16×16	-	-	-	-	-

-: No corresponding coding mode been selected

Settings for the resource allocation optimization: The hierarchical video coding algorithm based on visual perception characteristics is proposed as shown in Table 1.

EXPERIMENTAL RESULTS AND ANALYSIS

The experimental parameters are set as follows:

- **PC hardware configuration:** Pentium 4, 2 G RAM, 1.6 GHz frequency
- **Experimental software version:** JM17.0, Visual C^{++} compiler, Windows 2003 operating system
- **Video sequence formats:** QCIF; encoded frames: 100; Frame rate: 30f/s; GOP structure: IPPP; entropy coding type: CAVLC; QP: 28, 32, 36; motion estimation search range: ±16 pixels; the most number of reference frames: 5; Hadamard transform: On; rate-distortion optimization (RDO): On

In Table 2, the symbol "+" means enhancement or increase; symbol "-" means decrement or decrease. PSNR-Y means the peak signal-to-noise ratio of luminance and it also represents the quality of the reconstructed video image. PSNR-Y means the difference of the PSNR-Y. ROI-PSNR-Y means the non-zero region of the PSNR-Y in visual perception characteristics mark.

The simulation statistic results show that the computational complexity of the proposed method is lower compared with the H.264/AVC. The layered video coding algorithm reduces about 78.883% coding time on average under various QP (28, 32 and 36). In terms of bit rate control, the bit rate increases 1.647% on average, the PSNR-Y reduces 0.188dB on average. In non-zero region with visual perception characteristics which is the human visual attention region, the PSNR-Y reduces 0.153 dB on average. Compared with the human visual non-region of interest, the hierarchical video coding scheme gives the priority to ensure the quality of the visual perception characteristics saliency region. The proposed method inherits the advantages of low bit rate and high quality in H.264/AVC and maintains a good reconstructed video image quality.

Table 2: Performance of the proposed algorithm compared with H.264/AVC standard

QP	Time (%)	Bit rate (%)	PSNR-Y (dB)	ROI-PSNR-Y (dB)
28	-78.55	+1.93	-0.232	-0.188
32	-78.88	+1.74	-0.191	-0.155
36	-79.22	+1.27	-0.141	-0.115
Average	-78.883	+1.647	-0.188	-0.153

The experimental results also proved the feasibility of the low complexity visual perception analysis method based on the coding information. A large number of experimental results show that the proposed hierarchical video coding scheme based on visual perceptual analysis can accelerate the coding speed under the condition of maintaining good subjective video image quality. The consistency between visual perception characteristic saliency degree and HVS indicates the rationality of the hierarchical video coding algorithm based on visual perception characteristics.

CONCLUSION

This study presents an efficient hierarchical video coding scheme. In order to achieve high coding performance, the proposed method used the video stream information (such as prediction mode, motion vector and etc.) to extract visual region of interest and set the priority. The ROI detection technique for video used here, avoid the large amount of computational complexity of the traditional ROI detection algorithms successfully and the ROI detection results are consistent with human visual perception characteristics. On the other hand, the video encoder select different coding algorithms based-on the ROI priority setting results. The above technologies achieve a hierarchical video coding method and improve coding performance effectively. Experimental results show that the proposed algorithm can maintain good video image quality and coding efficiency, moreover improve the H.264/AVC computational resource allocation. It lays the foundation for following study of fast video coding algorithm in HEVC.

ACKNOWLEDGMENT

The research study is supported by the National Key Technology R and D Program of China with Grant No. 2011BAC12B03, the National Natural Young Science Foundation of China with Grant No. 61100131and Beijing City Board of education Project with Grant No. KM201110005007.

REFERENCES

Fang, Y.M., W.S. Lin, Z.Z. Chen, C.M. Tsai and C.W. Lin, 2012. Video saliency detection in the compressed domain. Proceedings of the 20th ACM International Conference on Multimedia, October 29-November 2, 2012, Nara, Japan, pp: 697-700.

Kim, C., J. Xin, A. Vetro and C.C.J. Kuo, 2006. Complexity scalable motion estimation for H.264/AVC. Proc. SPIE, 6007: 109-120.

Liu, P.Y. and K.B. Jia, 2011. Research and optimization of low-complexity motion estimation algorithm based on visual perception. J. Inform. Hid. Multimedia Signal Process., 2: 217-226.

Liu, P.Y., X. He and K.B. Jia, 2011. A fast H.264 inter-frame prediction algorithm for special mode. J. Binggong Xuebao, 32: 439-444.

Liu, P.Y., X. He, K.B. Jia and J. Xie, 2010. Fast intra-frame prediction algorithm based on characteristic of macro-block for H.264/AVC standard. J. Beijing Univ. Technol., 36: 158-162.

Narwaria, M., W. Lin and A. Liu, 2012. Low-complexity video quality assessment using temporal quality variations. IEEE Trans. Multimedia, 14: 525-535.

Saponara, S., M. Casula, F. Rovati, D. Alfonso and L. Fanucci, 2006. Dynamic control of motion estimation search parameters for low complex H.264 video coding. IEEE Trans. Consum. Electron., 52: 232-239.

Su, L., Y. Lu, F. Wu, A. Li and W. Gao, 2007. Real-time video coding under power constraint based on H.264 codec. Proc. SPIE, Vol. 6508. 10.1117/12.703686

Tsapatsoulis, N., C. Pattichis and K. Rapantzikos, 2007. Biologically inspired region of interest selection for lowbit-rate video coding. Proceedings of the IEEE International Conference on Image Processing, September 16-October 19, 2007, San Antonio, TX., pp: 333-336.

Wang, Y., H. Li, X. Fan and C.W. Chen, 2006. An attention based spatial adaptation scheme for h.264 videos on mobiles. Int. J. Pattern Recong. Artif. Intell., 20: 565-584.

Yao, W., L.P. Chau and S. Rahardja, 2012. Joint rate allocation for statistical multiplexing in video broadcast applications. IEEE Trans. Broadcast., 58: 417-427.