Research on Fast Algorithm for Video ROI Detection

Pengyu Liu
School of Electronic Information & Control Engineering / Beijing University of Technology, Beijing, China
Email: liupengyu @bjut.edu.cn

Kebin Jia
School of Electronic Information & Control Engineering / Beijing University of Technology, Beijing, China
Email: kebinj@bjut.edu.cn

***Abstract*—It is a hot topic to combine visual perception principles in video encoding. A fast algorithm for region of interest (ROI) detection based on human visual system (HVS) and coding information is proposed in this paper. Low-level coding information is employed as the characteristics of visual perception analysis in order to avoid a large amount of computation for visual perception characteristics analysis processing. Then, video ROI detection is carried out in temporal domain and spatial domain respectively. Last, combine the two domains saliency detection results to mark the video ROI level. The simulation results validate that the proposed algorithm by means of fast motion vector (MV) noise filtering and translational MV checking procedure, then realize fast visual perception characteristics analysis and ROI detection effectively.**

***Index Terms*—region of interest (ROI), visual perception characteristics, temporal and spatial domain, motion vector (MV)**

I. INTRODUCTION

With the development of multimedia information processing and communication technologies, video encoding has broad application prospects in information processing, national defense, research & education, medical treatment, aerospace engineering and other fields, and become one of hotspots for multimedia information processing technology. In order to obtain video codec with low complexity, high quality and high compression ratio, the fast encoding technology based on the new generation of video coding standard H.264/AVC has been developing rapidly since it was jointly released by ITU-T and ISO/IEC[1].

Exiting research show that HVS is sensitive to the video scene perception and assign different visual importance to different regions [2][3]. One of the key steps based on HVS video encoding is the calculation of ROI according to the visual perception characteristics [4][5]. Up till now, many researches have carried out research work on ROI. Moving zone detection technique for pixel precision is able to detect the moving foreground area, but its complexity in calculation makes it not applicable in instant encoding [6]. Liu Y et al [7] came up with algorithms for video sequences with large scale global motion and complicated background texture separately, yet their detection performance is far from satisfaction. Liu Z et al,[8] and Babu et al [9] proposed to carry out moving zone detection in compressed domain, separately. However, such algorithms mainly employ MV information, therefore, making it even not applicable in effective detection on moving object in global motion zone.

Encoding technique based on HVS mentioned above faced a common problem to handle non-motion region with equal visual importance in visual encoding progress. However, human eyes have different attention and sensitivity for different regions. Therefore, it is a problem requiring further study to take into consideration of the effect of different spatial visual characteristics region on HVS when obtaining temporal visual characteristics region, in order to design low-complexity, instant and effective analysis and detection algorithm for video ROI.

A fast video ROI detection algorithm based on low-level encoding information and HVS is proposed in this paper. In order to simplify the visual perception analysis progress, this algorithm correlates the encoding information in video bit-stream with the visual perception characteristics together. The spatial and temporal saliency detection are carried out by means of MV, prediction modes and other auxiliary coding information, which save amount of computing time in feature extraction for ROI detection. Because of almost no additional computation is increased to the video codec, the ROI detection computation complexity is lower compared with other existing algorithms. The detection results are satisfied among the compared algorithms. So the proposed algorithm reaches the balance between the saliency detection accuracy and computational complexity.

Manuscript received April 1, 2013; revised May 1, 2013; accepted May 28, 2013.
Corresponding author: Pengyu Liu.
E-mail: liupengyu@bjut.edu.cn

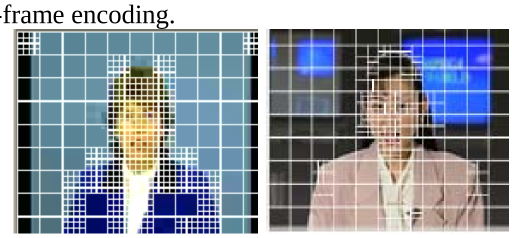
doi:10.4304/jsw.8.11.2871-2880

This paper is organized as follows: Section II analysis the relationship between coding information and the spatio-temporal characteristics of visual perception. Section Ⅲ provides an overview of the proposed framework and describes in detail each one of its subsystems. Section IV focuses on the evaluation of the proposed algorithm, and conclusions are given in the last part.

II. Spatial and Temporal Characteristics Analysis of Visual Perception

As we known, HVS has different perception sensitivity to visual scenes. Spatial contrast is the most basic visual processing mechanism in HVS.

HVS usually assigns higher importance to figures in foreground and moving object. For example, in H.264/AVC, I-frame encoding mainly select Intra 4×4 prediction mode. Foreground motion details are normally processed using P-frame with sub-block prediction mode collection Inter 8 set {8×8, 8×4, 4×8, 4×4}. As HVS is relatively non-sensitive to smooth background region, therefore, it is processed using Intra 16×16 prediction mode in I-frame encoding and macro block prediction mode collection Inter 16 set {Skip, 16×16, 16×8, 8×16} in P-frame encoding.



(a) I frame encoding (b) P frame encoding

Figure 1. Relationship between prediction modes with visual attention

As shown in Fig.1, it can be find that the high consistency between prediction mode decision results and visual attention. Therefore, the prediction mode can be regarded as spatial characteristics of visual perception analysis.

Besides spatial contrast, motion perception is one of the most important visual processing mechanisms. The visual information related to temporal motion would generate stronger response in HVS.

According to the relative relation between background and foreground, it can be discussed in the following two types.

- **Moving foreground under static background**

(a) Silent (b) Paris

Figure 2. Relationship between motion vectors with visual attention under static background

As shown in Fig.2, HVS pay more attention to the moving regions, such as head and shoulder area, face, arms and etc. Those regions are in the ROI with relative larger MV value. Non-ROI regions with relative smaller MV value or zero MV usually belong to the static background which could only arouse low attention of HVS.

As relatively high consistency exists between MV value and visual attention characteristic, MV can be regarded as temporal characteristics of visual perception analysis in this type.

- **Translational motion foreground under moving background**

In the practical application, due to the moving of the camera, amount of non-zero MV appeared in both the foreground and background at the same time. As shown in Fig.3, in Bus video sequence, the horizontal displacement of camera causes obvious MV along the road side and the parked car, while HVS tend to track the moving bus. The similar situation also occurred in the Coastguard video sequence. The horizontal displacement of camera causes obvious MV along the water surface and the shore. However, HVS is only interested in the moving vessels.

(a) Bus (b) Coastguard

Figure 3. Relationship between motion vectors with visual attention under translation background

Under these conditions, only on the basis of MV value does not agree with the visual attention. As a result, corresponding MV detection mechanism needs to be formulated as well in order to remove the interference caused by errors which are generated due to horizontal movement.

III. Detection Visual ROI

In the proposed visual ROI detection algorithm, the complexity of processing procedure will be controlled strictly. Otherwise it will influence the real-time performance of the saliency detection algorithm. The response intensity of temporal motion visual information is larger compared with that of spatial motion visual information caused by HVS [10]. In this paper, the temporal visual characteristics region detection is performed firstly, followed by detection in the spatial visual characteristics region.

In order to improve the temporal detection accuracy, the MV noise filtering and the attenuating translation MV interference error should be done.

A. Detection Temporal Visual Significance Region

In theory, there are no relative movements in the background region, so it should produce zero MV in static background. However in reality, non-zero MV noise would be generated randomly due to external change of illumination and internal change of video encoding parameters. As a result, MV detection mechanism should be done to filter out the interference of MV noise.

- **Main idea of filtering MV noise**

There exist strong motion continuity and relativity of the moving object in a video sequence. If v_s is generated in the current encoding block $s(x,y)$, there is a high probability that a MV with similar direction and value exists in the reference region C_{rr} of the corresponding position block in the previous frame. If no MV exists in C_{rr}, v_s should be treated as MV noise and would be filtered out.

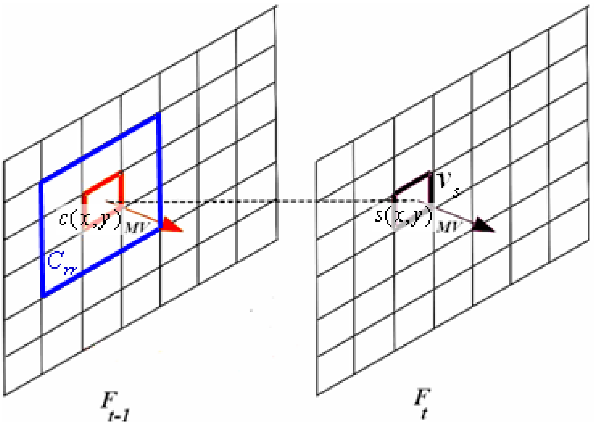


Figure 4. Schematic diagram of reference region and the current MV

According to the direction of v_s (horizontal, vertical and oblique motion), C_{rr} is determined as follows:

- **C_{rr} with horizontal motion**

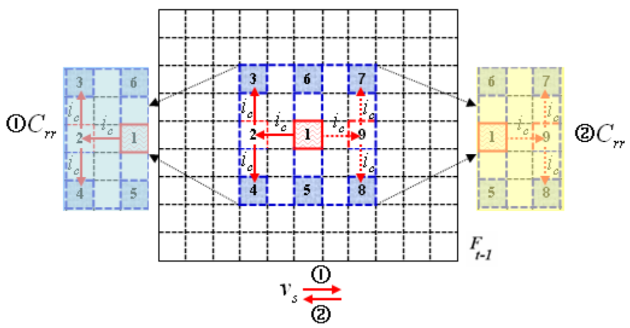


Figure 5. Schematic diagram of C_{rr} with horizontal motion

As shown in Fig.5, take horizontal motion toward the right direction of v_s as example, in the previous reference frame F_{t-1} find MB-1, which has the same position coordinate with the current encoding MB signed $c(x,y)$.

Firstly, take MB-1 as the starting point, then perform horizontal motion of i_c macro blocks in the opposite direction of v_s and get MB-2 signed $c(x, y-i_c)$.

Secondly, centered at MB-2, make vertical extension of i_c macro blocks both upwards and downwards to obtain two vertical vertices: MB-3 signed $c(x-i_c, y-i_c)$ and MB-4 signed $c(x+i_c, y-i_c)$.

Last, determine the rectangular reference region ① C_{rr} designated surrounded by four macro blocks which are MB-3, MB-4, MB-5 and MB-6.

If v_s is horizontal motion toward the left direction, use the same method to determine the rectangular reference region ② C_{rr} designated surrounded by four macro blocks which are MB-5, MB-6, MB-7 and MB-8.

In the above description, the position coordinates of MB-3 to MB-8 are given is Table Ⅰ.

TABLE I.
THE POSITION COORDINATES OF VERTEX MACRO-BLOCK WITH HORIZONTAL MOTION

macro blocks	position coordinates
MB-3	$c(x-i_c, y-i_c)$
MB-4	$c(x+i_c, y-i_c)$
MB-5	$c(x+i_c, y)$
MB-6	$c(x-i_c, y)$
MB-7	$c(x-i_c, y+i_c)$
MB-8	$c(x+i_c, y+i_c)$

In Table Ⅰ, i_c is defined as formula (1):

$$i_c = \left[\frac{|v_{sx}|}{w_s} + 1 \right] \quad (1)$$

$|v_{sx}|$ is the MV magnitude of the horizontal direction v_s. w_s is the width of the current coding block. $[\cdot]$ represent the round numbers calculation.

- **C_{rr} with vertical motion**

If v_s make the vertical downward motion, the processing steps to determine reference region ③ C_{rr} are as follows:

Firstly, take MB-1 as the starting point, then perform vertical motion of j_c macro blocks in the opposite direction of v_s and get MB-2 signed $c(x-j_c, y)$.

Secondly, centered at MB-2, make horizontal extension of j_c macro blocks both leftwards and rightwards to obtain two horizontal vertices: MB-3 signed $c(x-j_c, y-j_c)$ and MB-4 signed $c(x-j_c, y+j_c)$.

Last, determine the rectangular reference region ③ C_{rr} designated surrounded by four macro blocks which are MB-3, MB-4, MB-5 and MB-6 (as shown in Fig.6).

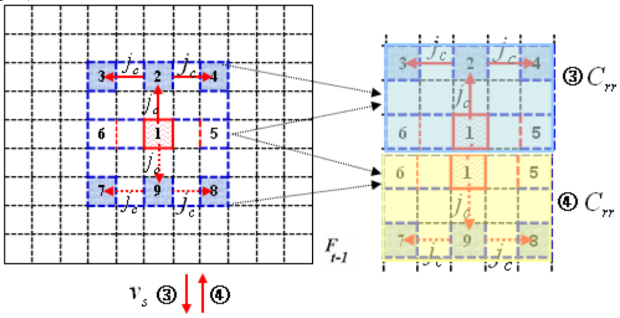


Figure 6. Schematic diagram of C_{rr} with vertical motion

If v_s make vertical upward motion, use the same method to determine the rectangular reference region ④ C_{rr} designated surrounded by four macro blocks which are MB-5, MB-6, MB-7 and MB-8.

In the above description, the position coordinates of MB-3 to MB-8 are given is TableⅡ.

TABLE II.
THE POSITION COORDINATES OF VERTEX MACRO-BLOCK WITH VERTICAL MOTION

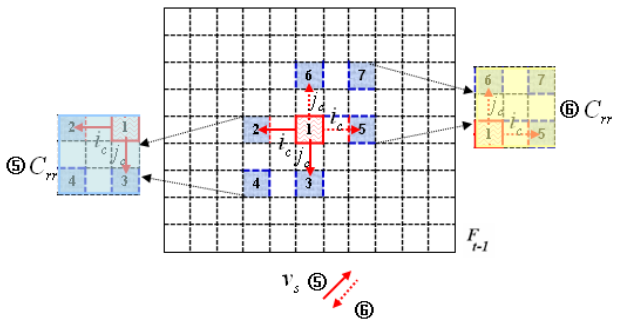
macro blocks	position coordinates
MB-3	$c(x-j_c, y-j_c)$
MB-4	$c(x-j_c, y+j_c)$
MB-5	$c(x, y+j_c)$
MB-6	$c(x, y-j_c)$
MB-7	$c(x+j_c, y-j_c)$
MB-8	$c(x+j_c, y+j_c)$

In Table Ⅱ, j_c is defined as formula (2):

$$j_c = \left[\frac{|v_s|}{h_s} + 1 \right] \tag{2}$$

$|v_{sy}|$ is the MV magnitude of the vertical direction v_s. h_s is the height of the current coding block. $[\cdot]$ represent the round numbers calculation.

- **C_{rr} with oblique motion**



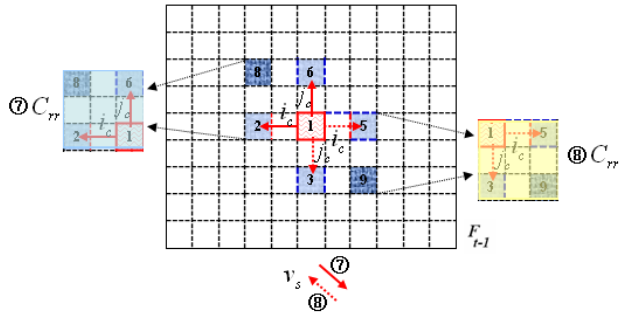


Figure 7. Schematic diagram of C_{rr} with oblique motion

If v_s make the oblique motion, determine the reference regions of ⑤ C_{rr}, ⑥ C_{rr}, ⑦ C_{rr} and ⑧ C_{rr} in the same way according to the different motion directions of v_s. C_{rr} is given in Fig.7.

The position coordinates of MB-1 to MB-9 are given in TableⅢ.

TABLE III.
THE POSITION COORDINATES OF VERTEX MACRO-BLOCK WITH OBLIQUE MOTION

macro blocks	position coordinates
MB-1	$c(x, y)$
MB-2	$c(x, y-i_c)$
MB-3	$c(x+j_c, y)$
MB-4	$c(x+j_c, y-i_c)$
MB-5	$c(x_c, y+i_c)$
MB-6	$c(x_c-j_c, y_c)$
MB-7	$c(x_c-j_c, y_c+i_c)$
MB-8	$c(x-j_c, y-i_c)$
MB-9	$c(x+j_c, y+i_c)$

In Table Ⅲ, i_c and j_c are defined as formula (1) and formula (2).

- **Calculate $\overline{V}_{rr}$ in C_{rr}**

$$\overline{V}_{rr} = \frac{\sum_{\in C_{rr}} \vec{v}_{rr}}{num_{C_{rr}}} \tag{3}$$

Here, $\vec{v}_{rr}$ is the MV of MB in C_{rr}, $num_{C_{rr}}$ is the summation times. (x, y) denotes the position coordinates of current encoding MB.

- **Filtering MV Noise**

The computation formula is given as follows.

Here, (x, y) denotes the position coordinates of current encoding block.

$$M_1(x,y)=\begin{cases}3, & if \quad |\overline{V}_{rr}|=0\\ 2, & else\,if \quad |v_s|\ge|\overline{V}_{rr}|\\ M_2(x,y), & else\,if \quad |v_s|<|\overline{V}_{rr}|\end{cases} \tag{4}$$

If $|\overline{V}_{rr}|=0$, consider v_s is caused by MV noise and should be filtered out. Set v_s ==0 and mark the current region as 3.

If $|v_s|\ge|\overline{V}_{rr}|$, there exist obvious motion characteristics in the current encoding block when compared with nearby macro blocks and current region belongs to dynamic foreground region which should be marked as 2.

Else, translational MV detection should be carried out in order to tell whether current encoding block belongs to background region or foreground translation region which has similar motion characteristics as its nearby macro block.

- **Checking translational MV**

Based on the theory of pixel region change detection [11] the formula of foreground translational region detection is as follows:

$$M_2(x,y)=\begin{cases}1, & if \quad SAD_{(x,y)}\ge\overline{S}_c\\ 0, & else\end{cases} \tag{5}$$

- **Calculate $SAD_{(x,y)}$ and $\overline{S}_c$**

$SAD_{(x,y)}$ is the sum of absolute difference of the current encoding block and its corresponding encoded block with the same position coordinates in previous frame. The value of $SAD_{(x,y)}$ can describe the variation degree of encoding blocks in two adjacent video frames. $SAD_{(x,y)}$ is defined as follows:

$$SAD_{(x,y)}=\sum_{i=1}^{M}\sum_{j=1}^{N}|s(i,j)-c(i,j)| \tag{6}$$

Here (x,y) stand for the position coordinates of the encoding block. $s(i,j)$ represents the pixel value of the current encoding block. $c(i,j)$ represents the pixel value of the corresponding block in previous frame. *M*, *N* is the partition dimension of current encoding block respectively. The value of $SAD_{(x,y)}$ is a representation of the change of an encoding block in two adjacent frame.

As there exists difference between different video sequence, different encoding parameters, especially quantization steps, would affect the code distortion and cause change in the value of $SAD_{(x,y)}$. In order to reduce the detection error caused by these uncertainties mentioned above, the proposed algorithm performs translational MV interference detection with a self-adaptive dynamic threshold $\overline{S}_c$, which is determined using the averaged SAD value of the all the encoding blocks in the background region in the previous frame.

$$\overline{S}_c=\frac{\sum_{x,y\in S_c}SAD_{(x,y)}}{Num} \tag{7}$$

S_c stands for the background region in the previous frame. $\sum_{x,y\in S_c}SAD_{(x,y)}$ stands for the summation of the $SAD_{(x,y)}$ values for all the encoding blocks enclosed in S_c. *Num* stands for the cumulative number.

- **Temporal visual characteristics region detection**

According to the calculation mentioned above, current encoding frame can be sorted into temporal visual characteristics regions with different significance based on the MV v_s of current encoding block and its MV relativity with adjacent blocks.

$$M(x,y)=\begin{cases}3, & if \quad |\overline{V}_{rr}|=0\\ 2, & else \quad if \quad |v_s|\ge|\overline{V}_{rr}|\\ 1, & else \quad if\ SAD_{(x,y)}\ge\overline{S}_c\\ 0, & else\end{cases} \tag{8}$$

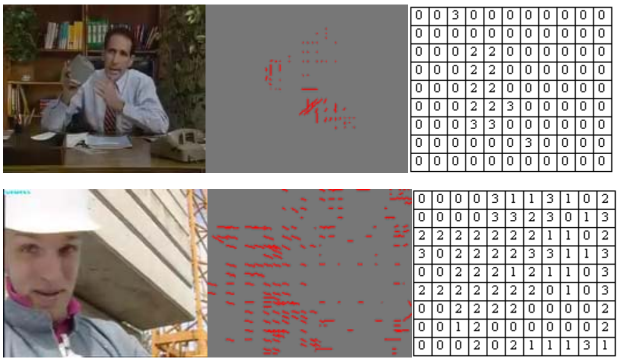
$M(x,y)=3$, means the current MV is MV noise.

$M(x,y)=2$, means the current encoding block is belonged to the foreground dynamic region.

$M(x,y)=1$, means the current encoding block is belonged to the foreground translational region.

$M(x,y)=0$, means the current encoding block is in the background region.

After filtering out the MV noise of blocks in region 3, they actually should belong to background region, which is marked as 0.



(a) Video frame (b) MV distribution (c) Detection results

Figure 8. Schematic diagram of temporal saliency detection results based on MV

According to the calculation procedure mentioned above, the current encoding frame can be sorted into temporal visual characteristics regions with different significance based on the low-level encoding information v_s of current encoding block and its MV relativity with adjacent blocks in C_{rr}.

Fig. 8 shows some temporal visual significance region detection results by MV filtering and translational MV checking processing. We can find that the macro blocks signed 1 or 2 are almost in the ROI.

B. Detection Spatial Visual Significance Region

HVS is also sensitive to the change in spatial domain, in order to improve the visual perception of the analysis results, it's needed to detect spatial visual significance region, analysis the correlation between prediction mode and spatial visual features. In this proposed algorithm, H.264/AVC standard is taken, and discuss the correlation between prediction mode and visual spatial attention.

- **Prediction modes in H.264/AVC**

Fig.9 shows all the encoding modes of H.264/AVC [12].

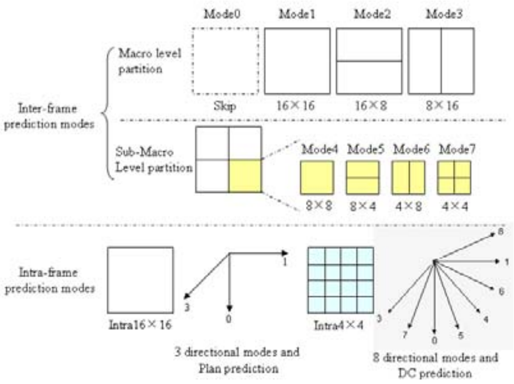


Figure 9. The intra and inter encoding modes in H.264/AVC

It can be find that H.264/AVC specifies 7 different block sizes for inter frames. The size of a block can be 16×16, 16×8, 8×16 or 8×8, and each 8×8 block can be further broken down to sub-macro-blocks of size 8×8, 8×4, 4×8, or 4×4. For each macro-block, in P-frame coding, H.264/AVC tries all possible modes in the following order: Skip, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4, Intra4×4 and Intra16×16. For intra modes, Intra4×4 needs to check 9 interpolation directions and Intra16×16 needs to check 4 interpolation directions [13].

- **Spatial visual characteristics region detection**

According to the analysis above, current encoding frame can be sorted into spatial visual characteristics region with different significance according to prediction mode.

$$S(x,y)=\begin{cases}2, & Mode_P \in \{Intra\} \\ 1, & Mode_P \in \{Inter8\} \ \ or \ \ Mode_I \in Intra4\times 4 \\ 0, & Mode_P \in \{Inter16\} \ \ or \ \ Mode_I \in Intra16\times 16\end{cases} \tag{9}$$

$\mathrm{mod}\,e_P$ and $\mathrm{mod}\,e_I$ stand for the prediction mode of current encoding block in P frame and I frame respectively.

If $Mode_P \in \{Intra\}$, it means intra prediction mode is selected for $\mathrm{mod}\,e_P$, the spatial visual characteristics significance is the highest. Current encoding block belongs to sensitive region and should mark $S(x,y)=2$.

If $Mode_P \in \{Inter8\}$ or $Mode_I \in Intra4\times 4$, it means current block is rich in spatial details with relatively high spatial visual characteristics significance. And current block belongs to attention region and should mark $S(x,y)=1$.

If $Mode_P \in \{Inter16\}$ or $Mode_I \in Intra16\times 16$, it means gentle spatial change exists in the current block. With low spatial visual characteristics significance, it belongs to non-significant region and should mark $S(x,y)=0$.

C. Setting Visual Significance Region

According to the results of temporal and spatial visual characteristics region detection, the video ROI detection according to coding information and HVS can be expressed as formula (7).

$$ROI(x,y)=\begin{cases}3, & ((M(x,y)=2 or M(x,y)=1) \ \| \ (S(x,y)=1)) or (S(x,y)=2) \\ 2, & (M(x,y)=2 or M(x,y)=1) \ \| \ (S(x,y)=0) \\ 1, & (M(x,y)=0) \ \| \ (S(x,y)=1) \\ 0, & (M(x,y)=0) \ \| \ (S(x,y)=0)\end{cases} \tag{10}$$

In this algorithm, according to the calculation number of $M(x,y)$ and $S(x,y)$, the priority level of ROI is divided into 64grades, from high to low is 3~0.

The flow chart of setting region of interest is shown in Fig. 10.

IV Simulation Results

In this paper, two existing algorithms are used to do the comparison experiment. The experimental environment is set as Table Ⅳ.

TABLE IV.
EXPERIMENTAL ENVIRONMENT

Computer hardware	P4 @ 1.6 GHz and 2G RAM		
Experimental software platform	JM17.0, Visual C++ , Windows 2003		
	Encoding frames	Frame rate	Gop
	100	30f/s	IPPP
	Entropy encoding type	QP	Search range
	CAVLC	28, 32, 36	±16 pixels
	Reference frames number	Hadamard transform	RDO
	5	On	On

7 typical test sequences with different scenes, motion and flatness are selected separately.

According to visual perception characteristics analysis and significance region detection mentioned above, the marking results for regions of interest matching visual selective attention mechanism are obtained which was shown in Table Ⅴ.

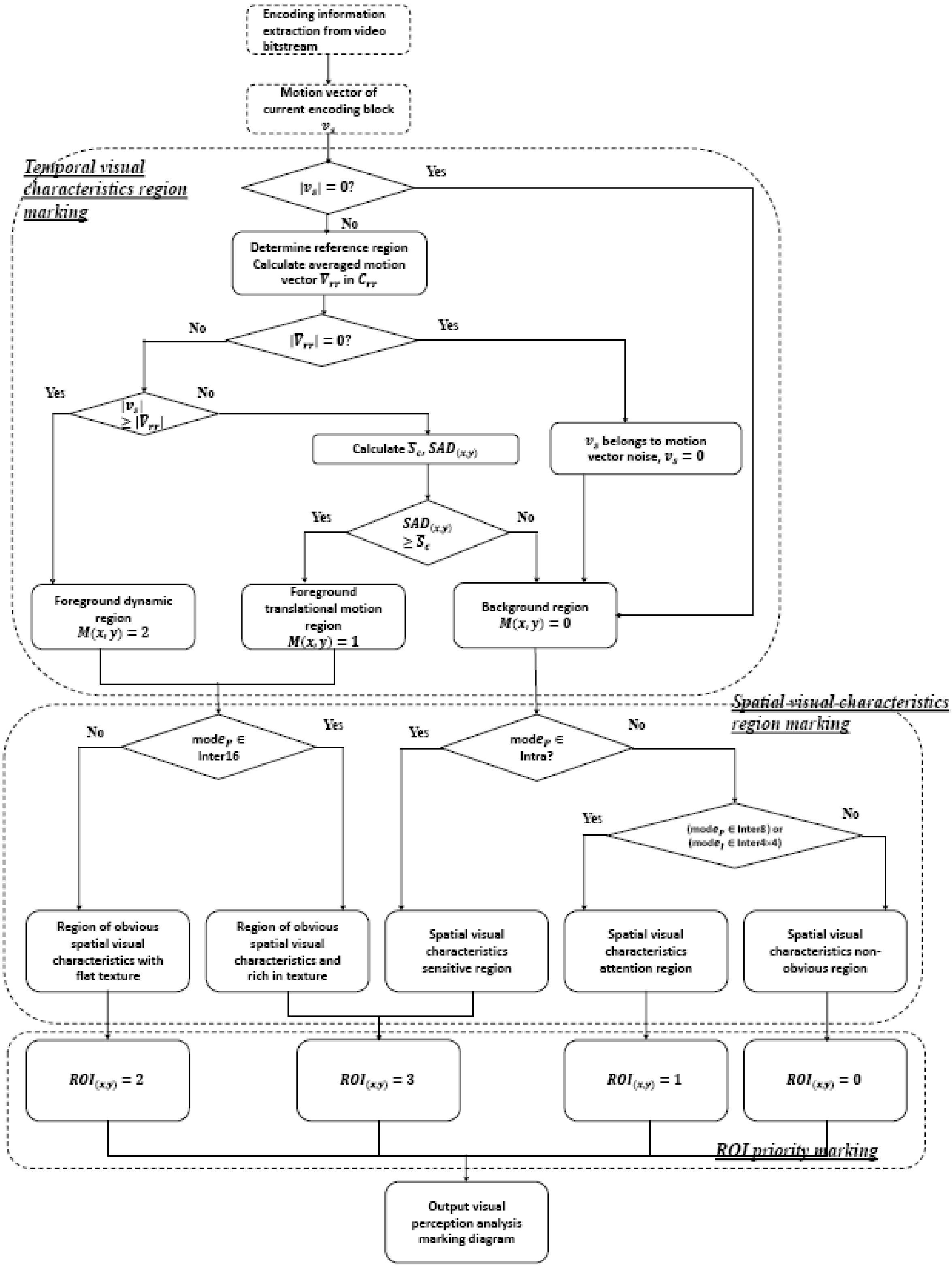


Figure 10. Flow chart of the region of interest setting algorithm based on encoding information and visual perception characteristics analysis

TABLE V.
SETTING RESULTS FOR REGION OF INTEREST BY USING THE PROPOSED ALGORITHM

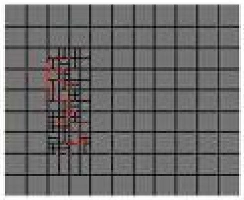
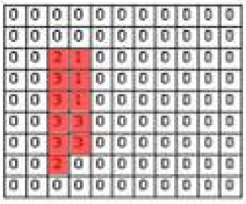
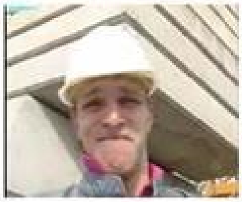
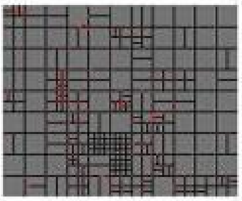
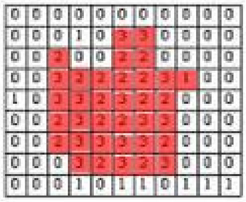
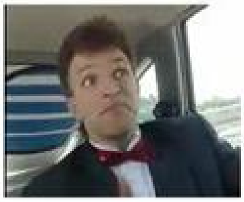
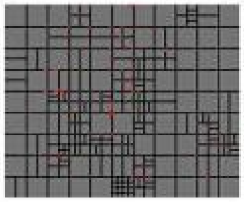
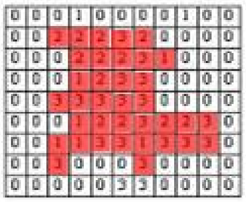
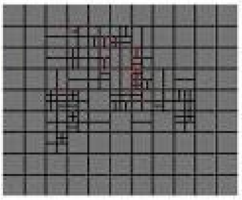
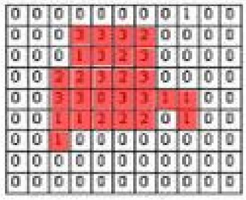
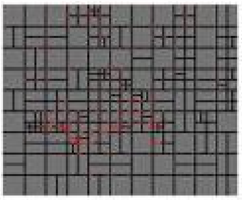
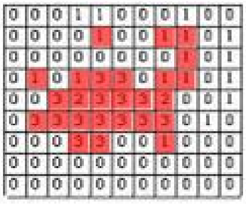
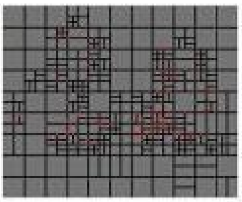
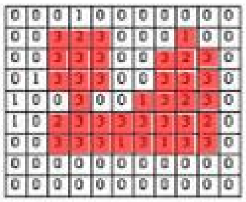
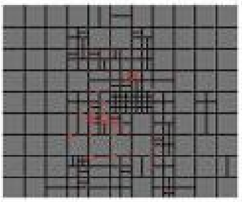
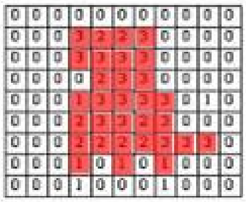
Video Sequences	Coding Information (prediction mode and MV)	Setting ROI
		
		
		
		
		
		
		

TABLE VI.
PROPOSED ALGORITHM PERFORMANCE TESTING RESULTS

Video Sequences	Compared algorithms	QP	△Computing Time (%)	△Bit-Rate (%)	△PSNR- Y (dB)
Hall	reference [7]	28	-13.75	-1.57	+0.12
	reference [8]		-5.49	-0.12	+0.04
Foreman	reference [7]	28	-12.88	-2.97	+1.21
	reference [8]		-10.71	-1.64	+2.01
Car-phone	reference [7]	32	-27.88	-2.41	+1.14
	reference [8]		-7.98	-0.69	+2.24
News	reference [7]	32	-18.74	-1.39	+0.08
	reference [8]		-6.21	-0.87	+0.01
Coastguard	reference [7]	32	-36.01	-1.87	+2.49
	reference [8]		-5.66	-1.21	+2.21
Paris	reference [7]	36	-19.77	-1.15	+2.04
	reference [8]		-3.78	-0.93	+1.21
Silent	reference [7]	36	-12.45	-0.84	+0.10
	reference [8]		-2.64	+0.06	-0.11

"+" sign in the table means increase, "-" sign means decrease.

It can be seen that the detection results have good consistency with HVS.

Statistical data as shown in Table Ⅵ reveals that using the proposed algorithm in this paper, the calculation time for region of interest detection can be saved up to 20.21% and 6.07% compared with reference [7] and [8], respectively, which enhance the timeliness of calculation. Strict control of output code rate is achieved with 1.74% and 0.77% reduction. PSNR-Y △ (dB) is improved by 1.03 dB and 1.09dB, especially. Compared with current methods available, the effectiveness of the algorithm presented here is reflected in result segmentation, data rate control and calculation expense.

V CONCLUSION

The paper researched the relevance between video encoding information and human visual system, and proposed a region of interest detection algorithm based on visual perception characteristics analysis and encoding information. The simulation results show that the proposed algorithm can effectively filter out the MV noise, weaken the interference of translational motion vector and get rid of visual redundancy, realize the detection of visual perception characteristics analysis and significance region detection fast and effectively. The proposed algorithm can be applied to video standards, such as HEVC, H.264/AVC and so on. And it can also be applied in other algorithms involving image/video analysis and management based on visual region of interest.

ACKNOWLEDGMENT

The research work is supported by the National Key Technology R&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. Thanks for the help.

REFERENCES

[1] Liu Yunxia, Li Zhitang, Ma Xiaojing, Reversible Data Hiding Scheme Based On H.264/AVC without Distortion Drift, *Journal of Software, vol. 7, n 5, May 2012, pp. 1059-1065.*

[2] Itti L, Koch C, Niebur E, A model of saliency-based visual attention of rapid scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 20 n 11, November 1998, pp.1254-1259.*

[3] Jiang Ming, Ma Zhaofeng, Niu Xinxin, Yang Yixian, An HVS-based JPEG image watermarking method for copyright protection, *International Journal of Digital Content Technology and its Applications, October 2011, vol. 5, n. 10, pp. 11-19.*

[4] Yao Wei, Chau Lap-Pui, Rahardja Susanto, Joint rate allocation for statistical multiplexing in video broadcast applications, *IEEE Transactions on Broadcasting, March 2012, vol. 58, n 3, pp. 417-427.*

[5] Guang Hu, Shengzhi Yuan, Research on target detection and automatic extraction of region of interest in infrared serial images, *Journal of Software, vol. 6, n 2, February, 2011, pp.225-232.*

[6] Song H, Kuo C-C J, A region-based H.263+ codec and its Vrate control for low VBR video, *IEEE Transactions on Multimedia, vol.6, n 3, June 2004, pp. 489-500.*

[7] Liu Y, Li Z G, Soh Y C, Conversation video communication of H.264/AVC with region-of-interest concern, *IEEE International Conference on Image Processing, October 8-11, 2006, Atlanta, GA, United states, pp.3129-3132.*

[8] Liu Z, Zhang Z, Shen L, Moving object segmentation in the H.264 compressed domain, *Optical Engineering, vol.46, n 1, January 2007, pp. 170031-170035.*

[9] Babu R V, Ramakrishnan K R, Srinivasan S H, Video object segmentation: a compressed domain approach, *IEEE Transactions on Circuits and Systems for Video Technology, vol.14, n 4, April 2004, pp. 462-474.*

[10] Itti L, Baldi P, A principled approach to detecting surprising events in video, *IEEE International Conference on Computer Vision and Pattern Recognition, June 20-25, 2005, San Diego, CA, United states, pp.631-637.*

[11] Radke R J, Andra S, Al-Kofahi O, Image change detection algorithms: a systematic survey, *IEEE Transaction on Image Processing, vol.14, n 3, March 2005, pp. 294-307.*

[12] Zhang Hao, Deng Yuetang, Liu Zhenye, Zhao Yuan, Zhan Haiyan, Complexity measurement and fault detection techniques for H. 264 optimized functions, *Journal of Software, vol. 7, n 8, August 2012, pp. 1845-1854.*

[13] Byeungwoo Jeon, Jeyun Lee, Fast Mode decision for H.264, *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG*, 2003.

Pengyu Liu received the MS degree in Information and Communication Engineering from Jilin University in 2004, and the Ph.D degree in electronic circuit and system from Beijing University of Technology in 2011. She is currently a lecturer, supervisor of postgraduate in Beijing University of Technology. Her research interests are in the areas of multimedia information processing, video coding, and information systems.

Kebin Jia, Professor, Ph.D. supervisor, was born in China, Anyang city of Henan Province in 1962, graduated from University of Science and Technology of China in 1998 and received a doctor's degree in signal and information processing. He is currently the dean of Electronic Information and Control Engineering, Beijing University of Technology. His major research field is signal information processing and multimedia information systems.