

一种有效的 MPEG 视频运动注意力区域提取方法

尤隽永, 刘贵忠, 李宏亮

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 提出了一种直接从 MPEG 压缩码流中提取运动注意力区域的方法, 其核心思想是根据运动注意力区域本身的特性, 利用位置、清晰度和运动矢量 3 部分信息以及相应的数学表达形式, 构造出一个运动注意力参数, 并在此基础上进行注意力区域的提取. 由于该方法同时考虑了运动注意力区域在空间和时间两方面的特性, 并且直接在压缩域中进行操作, 因此在提取准确度和运算效率上均有较大提高. 模拟结果表明, 与现有仅考虑运动信息的提取方法相比, 采用该方法提取的准确度高可 30% 以上, 基本上与人眼实际观测的结果相吻合; 与传统的像素域中运动对象提取方法相比, 运算时间可减少 1/2 左右, 而且提取出的目标区域更加符合基于内容的视频分析要求.

关键词: 运动注意力区域; 压缩码流; 运动对象提取

中图分类号: TN919.8 **文献标识码:** A **文章编号:** 0253 987X(2005)10 1135 04

Extract Method for Motion Attention Regions from MPEG Videos

You Junyong, Liu Guizhong, Li Hongliang

(School of Electronics and Information Engineering, Xi an Jiaotong University, Xi an 710049, China)

Abstract: To effectively analyze the real attention video motion objects, an efficient method to directly extract motion attention regions (MARs) from an MPEG coded stream was proposed. The essential idea of this method was as follows: Three properties of place, definition and motion vector in MARs were considered and then an MAR parameter was constructed to extract the attention regions by using these properties and their mathematical expressions. Because both of the space and time properties were taken into account and the frame compression operation was carried out directly, this method could greatly improve the extracting accuracy and computational efficiency. Simulation results indicate that the extracting accuracy increases by at least 30% compared to the existing methods, which basically accords with actual observation by eyes. In addition, the computing time decreases by about one half times compared to the traditional moving objects extract in pixel domain, and the extracted regions are more accordant with the video analysis demands based on contents.

Keywords: *motion attention region; coded stream; moving object extract*

运动对象的提取是视频分析的一个重要部分, 早期的工作集中在像素域内的操作^[1], 计算复杂度较大. 为了满足快速分析的要求, 通过分析运动矢量在空间和时间上的相关性以及变换系数等信息, 直接对 MPEG 压缩码流进行运动对象的提取和跟

踪^[2-4], 但用此类方法提取出的运动对象不一定全部是观众感应趣的区域. Itti 和 Koch 等人根据图像本身的特性(如密度、方向性以及色彩等方面的信息)构造了图像的视觉注意力模型^[5,6], 将图像中人眼所关注的区域提取出来. 基于此, Ma 等人考虑了运

动信息,在分析运动矢量的基础上提出了一种运动注意力模型^[7],通过考虑运动矢量的能量、空间和时间相关性进行运动注意力的分析,但该方法由于没有考虑区域本身的空间特性以及镜头的全局运动,因此常会造成误差。

本文充分考虑了 MPEG 压缩域内图像在空域和时域上的特性,通过预测残差的能量以及区域本身的空间和运动特性,提出了一种视频运动注意力区域的提取方法.实验表明,该方法可以有效地将视频内真正受关注的运动区域提取出来。

1 预测残差的能量及全局运动估计

在 MPEG 编码过程中,I 帧编码仅采用帧内压缩的方法,P 帧和 B 帧分别采用了不同类型的运动补偿方式,对补偿后的预测残差进行量化编码.如果对 P 帧和 B 帧的码流进行分析,即可获得预测残差的能量图,该图像正是当前帧与参考帧经过运动补偿后所得到的差值能量图。

一般将不存在局部运动的区域称为背景,用前景代表局部运动区域,所以全局运动参数应根据背景的运动信息进行估计,并要尽量避免前景运动信息参与进来.通常局部运动部分的残差能量较高,而背景部分的残差能量则常为 0,因此采用一般的背景提取方法就可把运动背景区域提取出来.Li 等人^[8]提出了一种快速鲁棒的全局运动估计算法,该方法仅利用预测残差的直流变换分量(DC 系数)来进行背景区域的估计,而且未对 B 帧进行估计.由于前景区域内也可能存在残差 DC 系数很小的部分,因此残差的 DC 系数往往并不能反映真实的背景信息.为了更加准确地提取背景区域,本文采用预测残差的能量作为区分背景与前景部分的参数.此外,还加入了 B 帧的全局运动估计,对 B 帧中的背景区域,利用其所有前向运动矢量(包括前向预测和双向预测中的前向运动矢量)进行运动估计.对于背景区域中具有前向预测运动矢量的像素点,采用二元线性回归算法及下式所示的 6 参数全局运动模型^[9]估计全局运动信息

$$\left. \begin{aligned} x &= a_0 + a_1 x + a_2 y + \\ y &= b_0 + b_1 x + b_2 y + \end{aligned} \right\} \quad (1)$$

式中: x 、 y 表示当前帧像素点坐标; x 、 y 表示参考帧像素点坐标; a_0 、 a_1 、 a_2 、 b_0 、 b_1 、 b_2 是全局运动参数。

2 运动注意力区域

前景中并非所有的区域都是人们所关注的部分,其中真正受关注的区域被称之为运动注意力区域.一般来说,前景中运动注意力区域和非注意力区域在空域和时域上存在 3 个方面的差别.通常摄像机在拍摄一段场景时,镜头总会跟随着注意力区域,并将其置于画面的中心位置处,另外往往会利用虚化其他区域的方法来突出注意力区域.因此,本文采用空域上的位置和清晰度信息作为判别运动注意力区域的前 2 个指标.此外,剔除掉全局运动之后,运动更剧烈的部分往往比运动缓慢的部分更能引起观众的注意,所以采用时域上宏块的运动矢量与全局运动矢量的差值作为反映该宏块是否属于运动注意力区域的第 3 个指标。

2.1 MPEG 码流中的 DCT 系数快速计算方法

MPEG 编码中,由于 I 帧没有运动估计和补偿过程,因而其中所有块的离散余弦变换(DCT)系数均可以从压缩码流中直接得到,而 P 帧和 B 帧使用了预测编码的方式,计算其 DCT 系数时需要先对压缩码流进行解码得到原始数据.为了达到快速分析的目的,Yeo^[10]提出了一种不需要完全解码,从 MPEG 码流中直接计算直流分量 DC 系数的方法,本文将推广到快速计算全部 DCT 系数上。

如图 1 所示,通过当前块 B_c 的运动矢量 M_c 可以得到其参考块 B_r 在参考帧中的位置.由于参考块和参考帧中原始块的位置往往并不重合,参考块通常位于 4 个原始块 B_1 、 B_2 、 B_3 和 B_4 之间,利用这 4 个原始块各自的 DCT 系数即可计算出参考块的 DC 系数^[10].另外,Yeo 还给出了利用 4 个原始块的 DC 系数近似计算参考块 DC 系数的表达式

$$D(B_r) = \sum_{i=1}^4 \frac{w_i h_i}{64} D(B_i) \quad (2)$$

式中: $D(B_i)$ 表示原始块 B_i 的 DC 系数; w_i 、 h_i ($i=1,2,3,4$) 表示参考块在 4 个原始块中所占范围的大小。

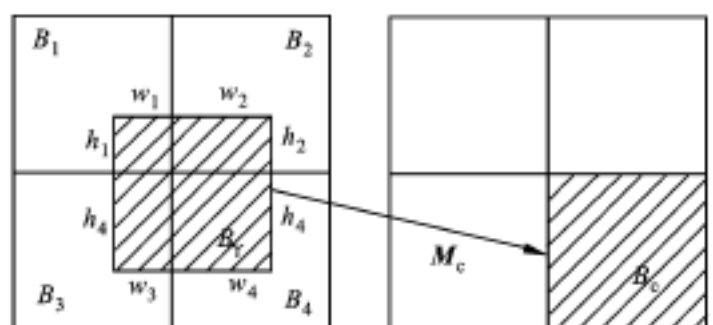


图 1 参考块、当前块和原始块示意图

将式(2)推广到计算参考块的交流变换分量(AC系数)上,即

$$A(B) = \sum_{i=1}^4 \frac{w_i h_i}{64} A(B_i) \quad (3)$$

式中: $A(B_i)$ 表示原始块 B_i 的 AC 系数。

如果当前块在 B 帧中而且采用的是双向预测模式,则分别计算其前后 2 个参考块的 AC 系数,然后取中值作为计算结果。记当前块和参考块之间的残差图像灰度为 B_d ,并用 $A(B_d)$ 表示残差块 B_d 的 AC 系数,由于它的 DCT 系数从压缩码流中可以直接得到,而且 DCT 是一种线性变换,则当前块的近似 AC 系数可由下式计算

$$A(B_c) = A(B_r) + A(B_d) \quad (4)$$

2.2 运动注意力区域的提取

根据前面的分析,运动注意力区域可由其位置、清晰度与运动矢量 3 部分参数决定,它们均以宏块中的子块作为处理单元,下面分别给出它们的表示形式。

假设图像中心点的坐标是 (x_0, y_0) ,当前子块 B_c 的中心点坐标为 (x_c, y_c) ,则 B_c 的位置参数为

$$P = 1 - \frac{((x_c - x_0)^2 + (y_c - y_0)^2)^{1/2}}{D_{\max}} \quad (5)$$

式中: D_{\max} 是图像中所有子块的中心点与图像中心点距离的最大值。

经过 DCT 变换之后,AC 系数通常表达了在一定方向和程度上的图像变化水平,可将其视为高频成分的表示。图像中某部分区域更加清晰是指其中细节成分更丰富,反映更多的图像信息,因此可用 AC 系数的绝对值之和来反映块的清晰度指标。如上文所述,当前子块的 AC 系数可按式(4)计算,数量为 63 个,计算所有 AC 系数绝对值之和的均值作为当前子块的清晰度参数,即

$$C = \frac{1}{63} \sum_{k=1}^{63} |A_k(B_c)| \quad (6)$$

另外,假设当前帧的全局运动矢量 $\mathbf{M}_M = (x_{GM}, y_{GM})$,当前子块 B_c 所在宏块的运动矢量 $\mathbf{M} = (x_c, y_c)$,以二者之间差分的模值作为 B_c 的运动矢量参数,即

$$M = \|\mathbf{M} - \mathbf{M}_M\| = ((x_c - x_{GM})^2 + (y_c - y_{GM})^2)^{1/2} \quad (7)$$

根据前面的分析,图像中子块距离图像中心位置越近,其内容越清晰,并且经过全局运动补偿之后,子块的运动越剧烈,则该子块属于运动注意力区域的可能性就越大,所以本文构造如下形式的子块运动

注意力参数 I_{MA} 来判断该子块是否属于注意力区域

$$I_{MA} = PCM \quad (8)$$

在提取参考帧中的运动注意力区域时,对帧内的每个子块利用式(8)计算其运动注意力参数,计算结果称为运动注意力参数图像。为了防止噪声点的影响,先对运动注意力参数图像采用邻域均值法进行平滑,再利用 Canny 算子对平滑后的图像提取边缘,所提取出的边缘点表示该点处对应的子块属于原始帧中的运动注意力区域并且处于该区域的边界。最后,取所有边缘点对应的运动注意力参数中的最小值作为提取注意力区域的阈值,对原始帧图像进行分割,将原始图像中运动注意力参数大于该阈值的子块提取出来,其结果就是运动注意力区域。

3 实验

为了验证本文提出的运动注意力区域提取方法,本文选用了几组典型的标准测试序列进行实验,它们是 Football、Table tennis、Coastguard、News1 和 Animals 序列。图 2 和图 3 分别给出了 Football 和 Table tennis 序列中 2 帧的运动注意力区域提取结果。

另外,为了更加充分地检验运动注意力区域提取的正确性,邀请了 10 位观察者对提取结果进行主观性评价。首先对每个测试序列的前 200 个 P 帧和 B 帧进行运动注意力区域提取,然后请观察者对各帧中的注意力区域进行观测并给出评价。评价结果分为 5 级:1 级表示提取区域完全不符合观察者对于运动注意力区域的判断,即符合程度为 0;2 级表示符合程度低于 50%;3 级表示符合程度在 50% 左右;4 级表示符合程度超过 50%;5 级表示完全符合。最后,将 10 位观察者的评价结果分别进行平均。表 1 给出了 10 位观察者对几种测试序列运动注意力区域提取结果的主观性判断,其中 R_c 表示评价结果为 4 级和 5 级的帧数所占比例。表 1 结果表明,本文提出的提取方法可以非常有效地提取出视频序列中真正受关注的运动区域。

表 1 几种典型测试序列的运动注意力区域提取的主观性评价结果

测试序列	各等级所占帧数均值/帧					R_c / %
	1级	2级	3级	4级	5级	
Football	0.3	5.6	10.7	22.5	160.9	91.7
Table Tennis	0	2.8	9.6	18.2	169.4	93.8
Coastguard	0.7	8.9	15.8	23.5	151.1	87.3
News1	1.4	10.6	16.4	30.5	141.1	85.8
Animals	1.1	8.3	18.2	27.6	144.8	86.2

注: R_c 表示评价结果为 4 级和 5 级的帧数所占比例。

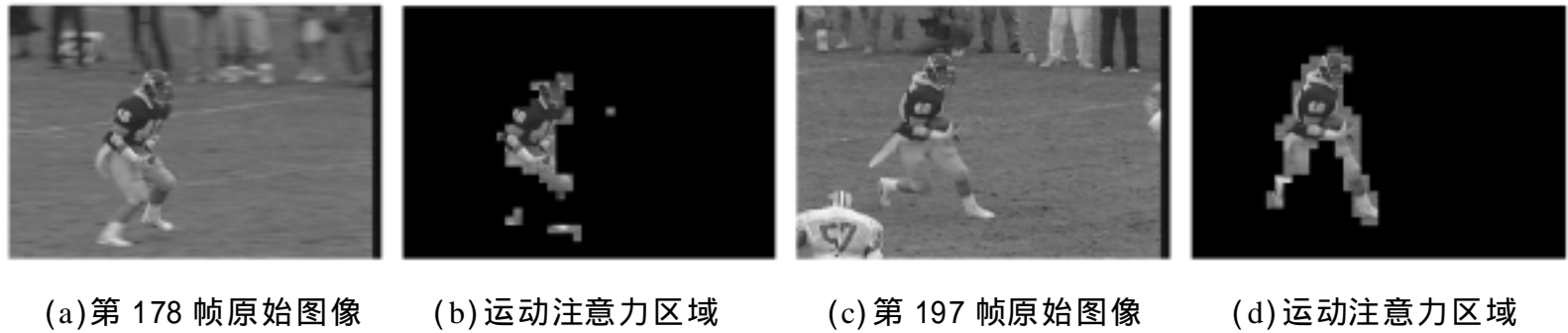


图 2 Football 序列中 2 帧运动注意力区域提取结果

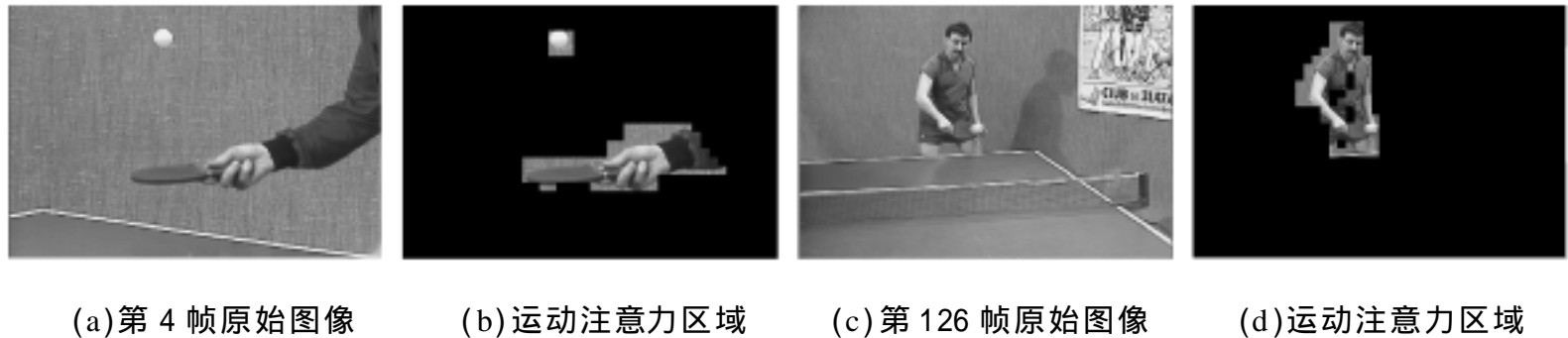


图 3 Table tennis 序列中 2 帧运动注意力区域提取结果

4 结 论

本文在深入分析 MPEG 码流的运动及空间特性的基础上,提出了一种新颖有效的运动注意力区域提取方法.首先,利用预测残差能量值的差异提取出全局运动背景部分,并在此基础上对 P 帧和 B 帧进行全局运动估计,有效地避免了因前景的局部运动带来的误差,可以准确地反映出视频序列中镜头的运动信息.其次,充分考虑运动注意力区域的空间位置和清晰度信息,并给出其表示方法,同时考虑注意力区域的运动矢量特征最终构造出一个运动注意力参数,在此基础上进行运动注意力区域的提取.实验表明,这种方法可以准确地提取出视频中真正受关注的运动区域.比起现有分析运动注意力的模型,本文方法更加全面地考虑了运动注意力区域的空间及运动特性,构造的参数也更加合理.在此基础上,下一步可以进行基于运动信息的视频摘要和检索以及快速浏览等方面的工作.

参考文献:

- [1] Jeon T, Ngan K N . Automatic segmentation of moving objects for video object plane generation [J] . IEEE Trans on Circuits and Systems for Video Technology, 1998, 8(5): 525 538 .
- [2] Nakajima Y, Yoneyama A, Yanaginara H, et al . Moving object detection from MPEG coded data [A] . SPIE Visual Communications and Image Processing, San Jose, USA, 1998 .
- [3] Ji S, Park H W . Moving objects segmentation in DCT-based compressed video [J] . Electronics Letters, 2000, 36(21): 1 769 1 770 .
- [4] Yoneyama A, Nakajima Y, Yanagihara H, et al . Moving object detection and identification from MPEG coded data [A] . IEEE ICIP1999, Kobe, Japan, 1999 .
- [5] Itti L, Koch C . Computational modeling of visual attention [J] . Nature Reviews Neuroscience, 2001, 2 (3): 194 203 .
- [6] Itti L, Koch C, Niebur E . A model of saliency-based visual attention for rapid scene analysis [J] . IEEE Trans on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1 254 1 259 .
- [7] Ma Yufei, Zhang Hongjiang . A model of motion attention for video skimming [A] . IEEE ICIP2002, New York, USA, 2002 .
- [8] Li Hongliang, Liu Guizhong, Li Yongli . A robust, efficient, and fast global motion estimation method from MPEG compressed video [A] . Third IEEE Pacific Rim Conference on Multimedia, Hsinchu, Taiwan, 2002 .
- [9] Dufaux F, Konrad J . Efficient, robust and fast global motion estimation for video coding [J] . IEEE Trans on Image Process, 2000, 9(3): 497 501 .
- [10] Yeo B L, Liu B . Rapid scene analysis on compressed video [J] . IEEE Trans on Circuits and Systems for Video Technology, 1995, 5(6): 533 544 .

(编辑 刘 杨)