

一种基于梯度模型的 MPEG 压缩域的运动对象分割算法

孙涛¹⁾ 杨高波¹⁾ 刘理¹⁾ 张兆扬²⁾

¹⁾(湖南大学计算机与通信学院,长沙 410082) ²⁾(上海大学通信与信息工程学院,上海 200072)

摘要 视频数据大都是经过压缩域的形式存储和传输的,且直接在压缩域进行视频对象分割无需运动估计等复杂的计算,速度较快。本文提出了一种基于梯度模型的 MPEG 压缩域的运动对象分割算法。首先利用 DCT ($AC[1]$ 和 $AC[8]$)系数获得所有物体的边缘,然后综合在累积运动矢量基础上得到的边缘运动信息,从而获得感兴趣运动物体的边缘。仿真实验结果表明,它可以取得满意的分割质量。

关键词 压缩域 运动对象分割 梯度模型 DCT系数 运动矢量

中图法分类号:TP391 文献标识码:A 文章编号:1006-8961(2008)06-1109-06

A Moving Objects Segmentation Algorithm Based on Gradient Model in the Compressed Domain

SUN Tao¹⁾, YANG Gao-bo¹⁾, LIU Li¹⁾, ZHANG Zhao-yang²⁾

¹⁾(College of Computer and Communication, Hunan University, Changsha 410082)

²⁾(School of Communication and Information Engineering, Shanghai University, Shanghai 200072)

Abstract There are several advantages to perform video object segmentation in a compressed domain. The primary benefit is with much high processing speed. Moreover, the video data already existing in MPEG1 and 2 formats in many databases would have to be decompressed to employ the pixel-domain techniques. In this paper, a moving objects segmentation algorithm based on gradient model in compressed domain is proposed. Firstly, DCT coefficients ($AC[1]$ and $AC[8]$) are utilized to form the gradient image. Then we synthesize the edge motion message obtained from the accumulated motion vectors. Finally, desirable moving objects segmentation results are acquired.

Keywords compressed domain, moving objects segmentation, gradient model, DCT coefficients, motion vectors

1 引言

视频对象(VO)分割是当前视频信号处理领域的研究热点之一。它在基于对象的 MPEG-4 视频编码、智能视频监控系统、目标识别、视频数据库检索、基于视频对象的视频摘要等领域有广阔的应用前景。目前的研究大多是在空间域进行的,尽管可以得到满意的分割结果,但往往难以满

足实时性的要求。然而,由于视频数据量大,通常是以压缩后的视频流形式进行存储和传输。若还原到空间域再进行视频对象分割,则需要经过相对耗时的解码步骤,更加难以满足实时性的要求。因此,视频对象提取的研究热点近年来已经由像素域转向压缩域,即直接在压缩域中提取视频对象。

压缩域的视频对象分割通常利用 DCT 系数和宏块的运动矢量(MV)两类信息。DC 系数携带着

基金项目:国家自然科学基金项目(60572127);湖南省自然科学基金项目(05JJ30113)

收稿日期:2006-10-19;改回日期:2007-01-25

第一作者简介:孙涛(1979~),男,湖南大学通信与信息专业硕士研究生。主要研究方向为多媒体通信、视频对象分割。

E-mail:sun.tao2007@163.com

纹理信息,例如 DC 系数可以表示宏块的平均能量,因此 DC 系数可用来重建一帧图像的 DC 图^[1]。另一方面,运动矢量代表相邻帧间宏块的位置偏移,每一个 P 帧或 B 帧的运动矢量可以得到一个稀疏的运动场,也可以通过积累或内插得到一个稠密的运动场^[2]。Mezaris 和 Fatih 等人则综合利用了 DCT 系数和运动矢量信息,分别利用聚类^[3]和区域增长^[4]完成分割。

本文提出一种基于梯度模型的 MPEG 压缩域的运动物体的边缘分割算法。算法首先利用 DCT 系数(AC[1]和 AC[8])获得物体的边缘,然后利用迭代后向投影得到的可靠累积运动矢量求得边缘运动矢量,最后综合求它们的交集,从而获得运动物体的边缘。算法取得了满意的实验结果。

2 算法概要

本文算法的原理框图如图 1 所示。它主要由以下几个部分组成:利用从 MPEG 压缩流部分解码得到 DCT 系数求 I 帧的空间梯度;从 MPEG 压缩流部分解码得到的运动矢量迭代后向投影累积和从相邻 P 帧到 I 帧的映射;求 I 帧的边缘运动矢量;运动物体边缘分割;跟踪。其中,I 帧的空间梯度和 I 帧的边缘运动矢量是一个相互的作用过程,前者对后者的修正去掉了噪声运动矢量,后者对前者的修正去掉了背景物体边缘。

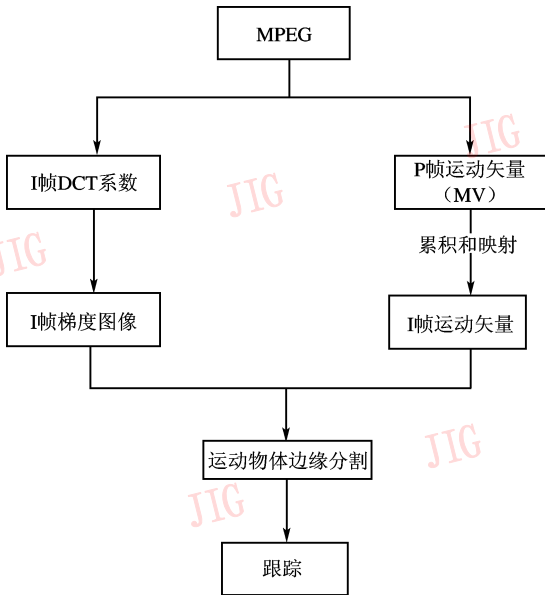


图 1 本文算法的原理框图

Fig. 1 Block diagram of the proposed approach

3 梯度图像的求取

MPEG1/2 视频标准采用了 DCT 变换作为去除空间冗余的必要前提步骤。其中,DC 系数体现了 8×8 块的平均能量,AC 系数体现了块的纹理分布。显然,经过了 DCT 变换的数据不再是块内某一点的像素值,不能通过 Sobel 算子等像素域求梯度的方法来直接获得物体的边界。

一个 8×8 块的 x 方向的平均梯度^[5]可计算如下:

$$\bar{f}_x(x,y) = - \sum_{x,y} w(x,y) \sum_{u,v} \frac{c(u)}{2} \frac{c(v)}{2} F(u,v) \times \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right) \quad (1)$$

式中, $w(x,y)$ 为权值函数,令 $w(x,y) = \sin\left(\frac{(2x+1)\pi}{16}\right)$ 化简可得:

$$\bar{f}_x(x,y) = - \frac{\pi}{2} F(1,0) = - \frac{\pi}{2} AC[1] \quad (2)$$

同理可令 $w(x,y) = \sin\left(\frac{(2y+1)\pi}{16}\right)$ 得:

$$\bar{f}_y(x,y) = - \frac{\pi}{2} F(0,1) = - \frac{\pi}{2} AC[8] \quad (3)$$

因此,DCT 系数的 AC[1](图 2),赋予其适当的系数 $\left(-\frac{\pi}{2}\right)$,可以作为块的水平方向的平均空间梯度。同理,AC[8] 赋予其适当的系数可以作为垂直方向的平均空间梯度。图 3 是 Forman 序列的第 5 帧经过 IDCT 检测得到的梯度图。其中, $\alpha = -\frac{\pi}{2}$ 。

显然,利用 AC[1]和 AC[8]就可以不经过相对耗时的 IDCT 变换步骤,只需部分解码 DCT 系数即可获得 I 帧图像的梯度图像。第 i 块的平均空间梯度记为

$$Grad_i = \left(-\frac{\pi}{2} AC[1]\right) + \left(-\frac{\pi}{2} AC[8]\right) \quad (4)$$

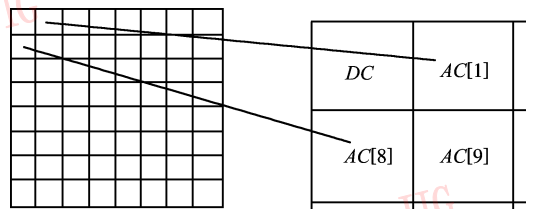


图 2 8×8 块中 AC[1]和 AC[8]的位置

Fig. 2 The positions of AC[1] and AC[8] in a 8×8 block

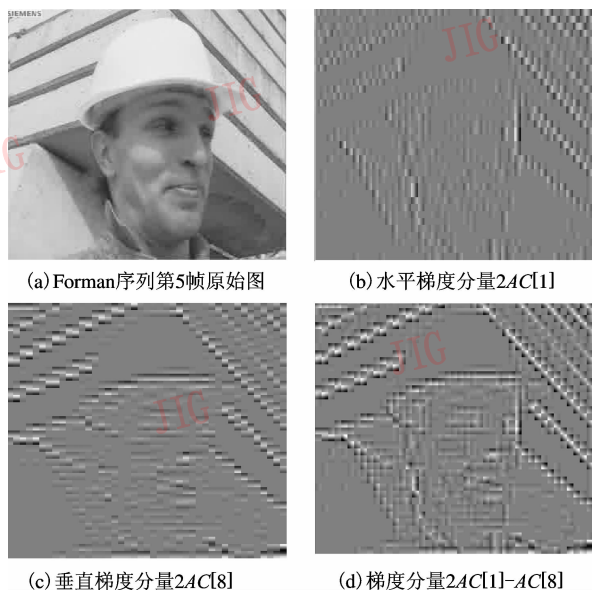


图 3 Forman 序列第 5 帧的梯度图

Fig. 3 The gradient image of Forman sequence(fifth frame)

当满足下列不等式时,为物体的边缘块

$$Grad_i > T_1 \tag{5}$$

其中,阈值 T_1 的选取取决于具体的视频序列。一般来说,如果取得的过低,会增加运算量;反之,会漏掉物体的边缘块。

通过这一步骤,可以获得 8×8 块精度的物体边缘信息。当然,它既包括需要提取的运动物体边缘,也包括背景物体的边缘信息。因此,要得到运动物体的边缘,还需要进一步的处理,去除背景物体的边缘。

4 由运动矢量获得运动物体边缘信息

4.1 运动矢量的处理

对于 MPEG 视频流,P 帧宏块主要有 3 种类型:前向预测(forward prediction)、跳过未编码(skipped)和帧内编码(intra coded)。其中,跳过未编码和帧内编码的宏块是没有运动矢量的,为了得到 P 帧的完整运动矢量场,需要进行以下处理:

跳过未编码的宏块:根据 MPEG-2 标准^[6],它的运动矢量与前一个块的运动相同。

帧内编码(intra coded)的宏块:取其 8 邻域运动矢量的均值。

MPEG1/2 编码标准中有 3 种编码类型:I 帧、P 帧和 B 帧。P 帧和 B 帧的运动矢量是用来进行运动补偿的,即 P 帧或 B 帧的运动矢量是以编码压缩为目的的,由于同时存在着量化噪声等因素影响,它们

往往不能反映物体的真实运动。还需要对运动矢量进行累积以获得可靠的运动矢量场。

MPEG1/2 采用的是过去的某图像帧作为当前帧的预测帧的方法来消除连续帧之间的时域冗余的。如格式为 $IP_0P_1P_2P_3P_4$ 来讲, P_0 是基于 I 帧预测的, P_1 是基于 P_0 帧预测的。本文鉴于 MPEG1/2 这种时域模式,利用当前帧之后若干帧的运动矢量场,进行迭代后向投影来获得更加可靠的累积运动矢量场。

文献[4]中提出一种运动矢量从 P 帧映射到 I 帧的方法,本文把这种方法扩展用于运动矢量的迭代后向投影积累,以求得到可靠的累积运动矢量场。图 4 描述了当前帧(P_{i+1})对其参考帧(P_i)的运动矢量的投影积累方法。

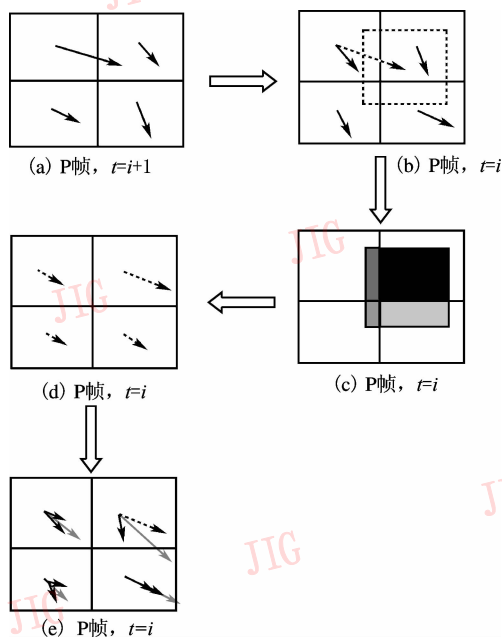


图 4 P_{i+1} 对 P_i 的运动矢量的投影

Fig. 4 The motion vector project of current frame P_{i+1} relative to reference frame P_i

4.2 运动矢量从 P 帧到 I 帧的映射

对于 I 帧和其后相邻的 P 帧来说,P 帧宏块的运动矢量是指它相对于 I 帧中对应宏块的位置偏移,而 I 帧是没有运动矢量的。因此,需要把邻近 P 帧的运动矢量映射到 I 帧。首先,把 P 帧宏块的运动矢量直接赋给该宏块所覆盖的 4 个 8×8 块,然后把所有 8×8 块按上述投影的方法映射到 I 帧,也就是按 I 帧的块占相邻 P 帧在 I 帧相对应的匹配块的百分比多少来求 I 帧的块的运动矢量。这种方法与文献[5]中采用的直接把 P 帧的运动矢量映射到 I

帧的相同位置的方法相比较,能比较精确地得到 I 帧的块精度的运动矢量场。此外,在文献[3]中提出了一种取 I 帧前后邻近的 P 帧运动矢量平均值的方法,但它不适用于第一个 GOPs 的 I 帧。

通过上述处理,可以得到 I 帧的能较真实地反映物体运动的运动矢量场。当然,噪声是不可能全部消除的。

4.3 检测边缘运动矢量

运动矢量大致可分为:前景运动矢量、背景运动矢量、边缘运动矢量和噪声运动矢量^[7]。其中,边缘运动矢量暗示着运动物体的边界。不管背景运动大于前景运动(如 dancer 序列),还是小于前景运动或背景不动(如 tennis 序列前半部分),边缘运动矢量和噪声运动矢量都发生了强度或方向上的突变。令第 i 个运动矢量记为 MV_i , 其 8 邻域运动矢量平均值记为 $MV_{avg}(i)$, 当满足下列不等式时,为边缘运动矢量或噪声运动矢量。

$$\begin{aligned} |MV_i - MV_{avg}(i)| > T_2 \\ \text{或者} |MV_i - MV_{avg}(i)| < T_2 \end{aligned} \quad (6)$$

其中,阈值 T_2 的选取取决于具体的视频序列。

同时应用不等式(5)和式(6)就可以去掉背景物体边缘和噪声运动矢量,剩余的即可被认为是运动物体边缘。

5 跟踪

得到 I 帧运动物体的边界分割(VOP)后,需要进行对象跟踪。为了实现视频对象(VO)的分割,采用 Favalli 提出的算法^[8]对运动物体进行跟踪。该算法首先在编码端标出要跟踪的物体,然后在压缩域利用要跟踪的物体所包含的宏块的运动矢量进行跟踪。其具体在压缩域的跟踪算法是:根据运动矢量计算出当前帧(t 时刻)组成物体的宏块在后一帧($t+1$ 时刻)的位置,重叠面积超过 25%的宏块记为物体在后一帧所包含的宏块。上述算法存在的问题是:被跟踪的物体会渐渐地“膨胀”。针对这个问题,本文通过在每个 GOPs 开始的 I 帧对运动物体重新分割来消除误差的积累和扩散。

6 试验结果

6.1 试验结果分析

通过两个典型的 MPEG-2 的标准测试序列来验

证本文算法:Table tennis 序列(SIF, 352×240)和 dancer 序列(CIF, 352×288)。它们代表不同运动复杂性的序列,在 MPEG-2 编码时,编码格式为 IPPPP。

序列 Table tennis 的前半部分的背景基本上是静止的,有两个前景运动物体。图 5 给出了第 0 帧的分割过程和结果。从图 5(b)中,可以看到有大量的非运动物体的边缘块,在边缘运动矢量的强度阈值取 2,方向阈值取 20° 的情况下,去除非运动物体的边缘块,得到最终分割结果,如图 5(d)所示。

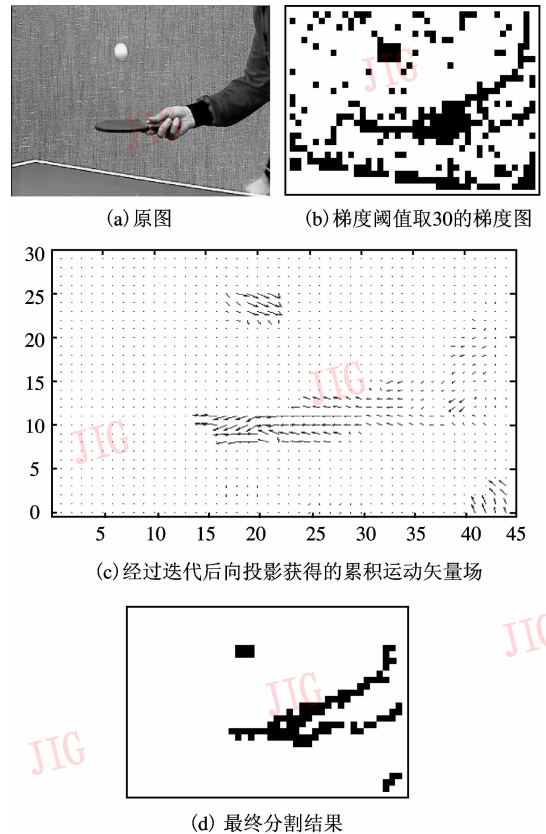


图 5 Table tennis 序列的分割结果

Fig. 5 The segmentation results of Table tennis sequence

序列 dancer 具有剧烈的背景运动。同样也有两个前景运动物体。图 6 给出了第 30 帧的分割过程和结果。从图 6(b)中可以看到存在着背景运动物体的边缘块。在边缘运动矢量的强度阈值取 5,方向阈值取 30° 的情况下,去除背景运动物体的边缘块,得到最终分割结果,如图 6(d)所示。

图 7 和图 8 为本文算法对 Table tennis 和 dancer 序列的运动对象的分割结果。可见,本文算法能够较好地实现运动对象的分割。

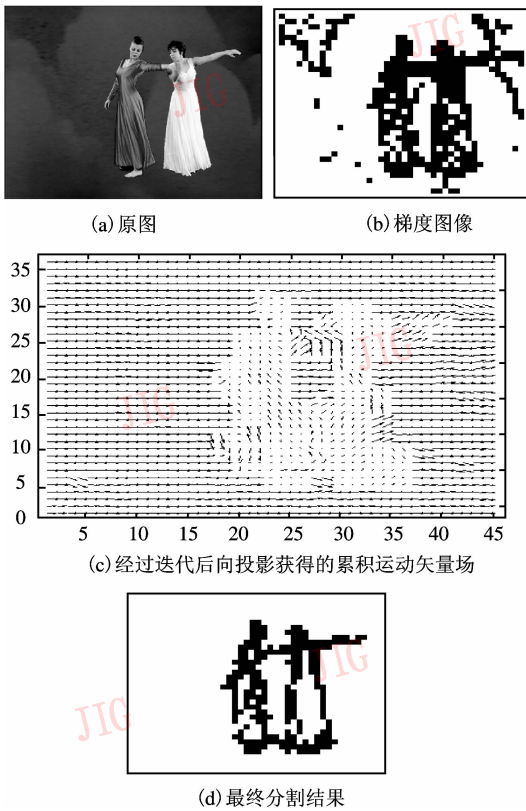


图 6 Dancer 序列的分割结果

Fig. 6 The segmentation results of dancer sequence

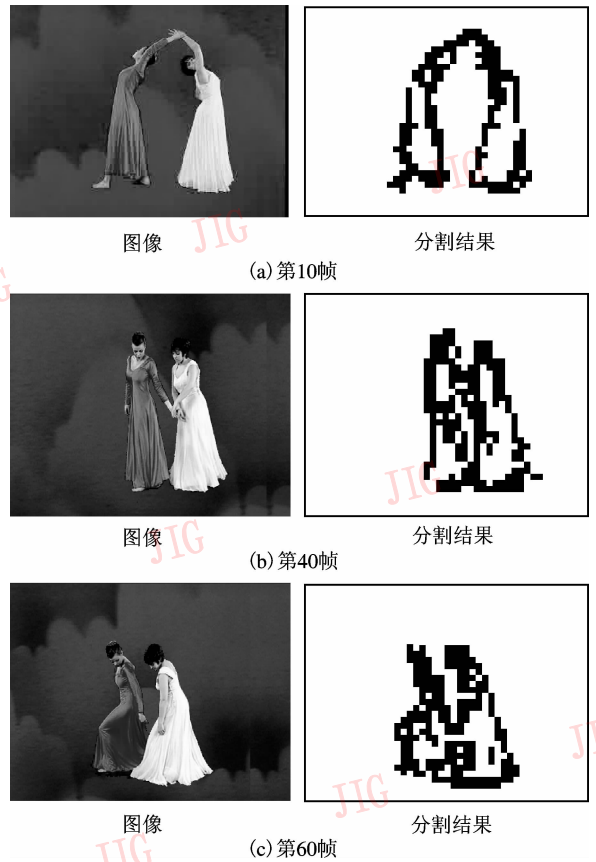


图 8 Dancer 序列的分割结果

Fig. 8 The final segmentation results of dancer sequence

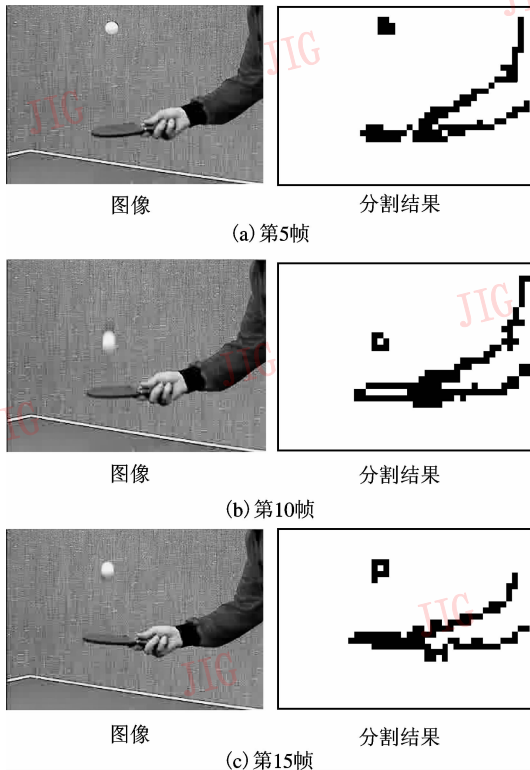


图 7 Table tennis 序列的分割结果

Fig. 7 The final segmentation results of Table tennis sequence

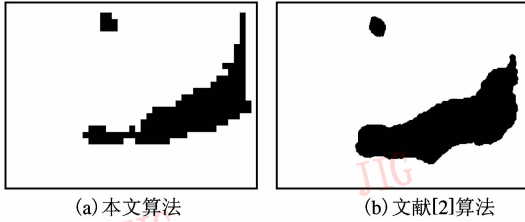
6.2 对比试验

由于已有文献很少涉及 dancer 序列的压缩域分割。所以本文仅对常见的 Table tennis 序列进行对比试验。为了便于和已有文献试验结果进行对比,首先在 Table tennis 序列的最终分割边界基础上进行填充(图 9、图 10)。

目前大部分压缩域视频对象分割算法采用主观对比的方法评比分割质量。分别与文献[2]和文献[3]中提出的算法进行了分割实验对比。其中,文献[2]通过空间查值获得稠密运动矢量场,最后采用期望最大(EM)算法完成最终分割。该算法虽然能达到像素级精度,但其缺点也很明显,计算量大且分割出的运动物体过于“肥大”。文献[3]则利用一个反复拒绝的过程分出前景与背景,然后通过对该块的跟踪和聚类完成前景运动物体的分割。该算法能处理多种全局运动下的运动物体分割,但对于小的运动物体并不敏感。图 9 是对 Table tennis 序列的第 5 帧分别采用本文算法和文献[2]提出的算法进行的分割试验,图 10 是对第 10 帧分别采用本文

算法和文献[3]提出的算法进行的分割试验。

从图9、图10中可以看出,本文压缩域运动对象分割算法比文献[2]和文献[3]中提出的算法效果要好。

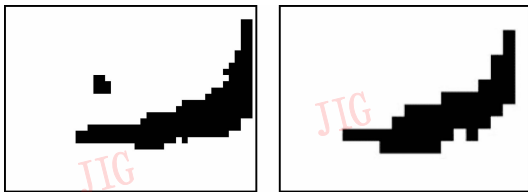


(a) 本文算法

(b) 文献[2]算法

图9 Table tennis 序列第5帧分割结果

Fig. 9 Results of segmentation for table tennis sequence, frames 5



(a) 本文算法

(b) 文献[3]算法

图10 Table tennis 序列第10帧分割结果

Fig. 10 Results of segmentation for table tennis sequence, frames 10

7 结 论

本文提出一种基于梯度模型的压缩域的运动物体的分割算法。算法首先利用 DCT 系数 ($AC[1]$ 和 $AC[8]$) 获得物体的块精度的边缘。再综合经过迭代后向投影获得的累积运动信息基础上得到的边缘运动信息来获得我们感兴趣

的运动物体的边缘。该算法由于不需要考虑背景运动的快慢或静止,适应性好。实验结果表明,本文算法能够在各种背景下自动地实现视频对象的准确分割。

参考文献 (References)

- 1 Zeng Wei, Gao Wen, Zhao Debin. Automatic moving object extraction in MPEG video [A]. In: Proceedings of the IEEE International Symposium on Circuits and Systems [C], Bangkok, Thailand, 2003, 2: 524 ~ 527.
- 2 Babu R V, Ramakrishnan K R, Compressed domain motion segmentation for video object extraction [A]. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C], Florida, USA, 2002: 3788 ~ 3791.
- 3 Mezaris V, Kompatsiaris I, Boulgouris N V, et al. Real-Time compressed-domain spatio-temporal segmentation and ontologies for video indexing and retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(5): 606 ~ 621.
- 4 Fatih Porikli. Real-time video object segmentation for MPEG encoded video sequences [A], In: SPIE Conference on Real-Time Imaging [C], New York, USA, 2004, 5297: 195 ~ 203.
- 5 Miguel T C, Michael D, Approximating optical flow within the MPEG-2 compressed domain [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(1): 103 ~ 107.
- 6 Song Junehwa, Yeo Boon Lock. Fast extraction of spatially reduced image sequences from MPEG-2 compressed video [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(7): 1100 ~ 1114.
- 7 Zeng Wei, Du Jun, Gao Wen, et al. Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model [J]. Real-Time Imaging, 2005, 11: 290 ~ 299.
- 8 Favalli L, Mecocci A, Moschetti F. Object tracking for retrieval applications in MPEG-2 [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(3): 427 ~ 432.