

# 一种从视频压缩码流中精确提取运动对象的快速算法<sup>1</sup>

王占辉 刘贵忠 刘 龙 刘洁瑜

(西安交通大学电子与信息工程学院信通系 西安 710049)

**摘 要:** 针对目前大部分视频对象分割方法相当复杂而且计算量大的问题, 提出了一种在压缩域粗分割, 在空域精细分割的方法。该方法利用压缩域中运动向量进行聚类, 得到运动对象的初始分割, 将分割模板通过运动参数映射到参考帧 I 帧, 解码初始分割区域进行 Canny 边缘检测和边缘跟踪, 即可得到精确的对象轮廓。该方法使得处理的数据量保持最小, 节约了处理时间并得到了像素级精度的分割对象。

**关键词:** 压缩域, 运动对象, Canny 算子, 运动向量

**中图分类号:** TP391.41 **文献标识码:** A **文章编号:** 1009-5896(2004)07-1157-06

## A Fast and Accurate Moving Object Extraction Method from Compressed Video Stream

Wang Zhan-hui Liu Gui-zhong Liu Long Liu Jie-yu

(School of Electron. & Info. Eng., Xi'an Jiaotong Univ., Xi'an 710049, China)

**Abstract** In the light of most current segmentation algorithms are of high complexity and huge computation, one algorithm of coarse segmentation in compressed domain and refined segmentation in spatial domain is put forward. The initial coarse segmentation masks from the motion vectors are obtained by applying Estimation Maximum(EM) algorithm. These blocks in the masks can be decompressed to obtain the origin image and the actual edges of the objects can be extracted by applying Canny edge detection and edge tracking only in the segmented regions. By using the proposed algorithm, the amount of data needed to be processed is kept in necessarily minimal, saving the computation time as well as gaining the pixel-wise edges of the segmented objects.

**Key words** Compressed domain, Moving object, Canny operator, Motion vector

### 1 引言

图像和视频数据库要操纵大量的可视化信息, 就需要一种有效、高效的搜索和检索算法。这些可视化信息通常以压缩格式存储。因此要提取感兴趣对象的特征, 不仅要分割对象、提取特征, 而且需要提前解码相应的图像或视频序列。解码过程是相对耗时的过程, 所以研究利用压缩域的信息进行分割和提取对象的方法是有益的。

大多数的分割算法是基于像素域(空域)的<sup>[1-3]</sup>。这些方法得到了比较精确的结果, 但是最大的不足就是需要大量的时间处理视频序列。大多数成功的像素域分割系统需要大约 20s 或更长的时间才能处理 1 帧<sup>[4]</sup>, 而且像素域的分割方法往往产生过分割。随着 H.263, MPEG 系列标准的制定和修改, 基于压缩域的视频对象分割成为研究的热点, 出现了一些压缩域分割方法<sup>[5-7]</sup>。这些方法存在的局限性是只能得到粗略的对象轮廓。其优点是避免了消耗大量的时间进行解码压缩码流。

发展一种适用于多种视频序列的通用的自动分割算法, 就现在的语义分割是不现实的, 因此我们将问题缩小为运动对象提取。综合分析空域和压缩域运动对象分割系统, 结合两种系统

<sup>1</sup> 2003-04-27 收到, 2003-09-19 改回

国家自然科学基金(No.60272072)、国家教育部高等学校博士点基金(No.2000069828)及跨世纪优秀人才计划(2002 年度)资助课题

的优点, 本文提出了一个新的运动对象分割系统. 利用运动信息进行视频对象的初始分割, 再部分解码初始分割的对象区域, 利用 Canny 算子提取对象的精确边缘. 从试验结果看, 系统有良好的分割结果, 并且使得处理的数据量保持最小, 降低了计算复杂度.

## 2 系统概述

整个系统 (如图 1) 主要包括 3 部分: (1) 运动向量处理; (2) 粗糙的视频对象分割; (3) 像素域对象细节提取. 视频码流包括 3 种类型的帧: 帧内编码帧 (I 帧), 预测帧 (P 帧) 和双向预测帧 (B 帧). I 帧用 DCT 独立编码, 视频序列开始于 I 帧. P 帧和 B 帧利用参考帧 (I 帧或 P 帧)、运动向量和残差 DCT 系数进行编码.

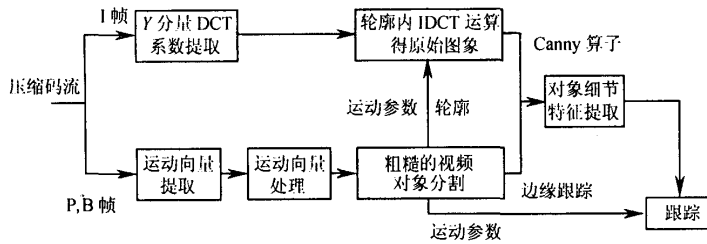


图 1 系统框图

## 3 运动向量处理

压缩域中运动向量只是用于编码目的, 它们是有噪声、不可靠的, 直接用于运动分割会产生很大的错误. 为了消除噪声, 增强运动向量的可靠性, 需进行以下处理.

### 3.1 噪声运动向量滤除

由于图像噪声等原因, 用块匹配算法估计运动向量时, 静止的区域往往产生噪声运动向量, 且幅值相对比较小. 于是我们就可以设定一个阈值  $T_{|MV|}$  将噪声运动向量滤除, 即

$$MV_{i,j} = \begin{cases} 0, & |MV_{i,j}| \leq T_{|MV|} \\ MV_{i,j}, & \text{其他} \end{cases} \quad (1)$$

其中  $MV_{i,j}$  是索引为  $(i, j)$  的块的运动向量.

### 3.2 空间检测和校正

运动向量是有噪声的, 部分运动向量并不能表示对象的运动<sup>[7]</sup>. 我们对 wang<sup>[7]</sup> 的运动向量空间检测准则进行修改, 提出了一种运动向量空间检测和校正准则, 以得到可靠的运动向量场. 强度和方向的检测是看当前运动向量背离相邻运动向量的程度, 如果背离程度大 (即强度或方向突变), 则认为当前运动向量是不可靠的. 将不可靠的运动向量检测出来, 利用相邻运动向量中值进行填充.

$$Dev1_{i,j} = \left| \left( M_{i,j} - \sum_{\Theta} \frac{M}{N} \right) / \sum_{\Theta} \frac{M}{N} \right| \quad (2)$$

其中  $Dev1_{i,j}$  为索引  $(i, j)$  处运动向量强度  $M_{i,j}$  背离相邻的非零运动向量的程度.  $\Theta$  表示 8 邻域结构.  $N$  是邻域内非零运动向量的个数.  $M$  是邻域内运动向量的强度.

$$Dev2_{i,j} = \sum_{\Theta} |D(MV_{i,j}) - D(MV)| / N \quad (3)$$

其中  $\text{Dev}2_{i,j}$  为索引  $(i,j)$  处运动向量与邻域运动向量夹角绝对值之和的平均值。  $\Theta$  和  $N$  同上。  $D(MV)$  是运动向量  $MV$  的方向。

强度背离程度值  $\text{Dev}1_{i,j}$ ，方向背离程度值  $\text{Dev}2_{i,j}$  越大，说明当前运动向量越不可靠。于是我们设定阈值  $T_M$  和  $T_D$ ，将不可靠的运动向量检测出来进行处理。

最后对运动向量进行二维中值滤波处理。中值滤波的作用可以平滑运动向量。经过上述处理过程，运动向量能够较真实地反映运动对象的运动。

#### 4 粗糙的视频对象分割

运动分割是将运动一致的区域（这里我们处理的是运动向量）进行合并。我们用一种概率的方法（即 Estimation Maximum, EM 算法<sup>[8]</sup>）进行聚类，将运动一致的区域看成一个运动对象提取出来。EM 算法需要预先知道分割对象数目，而运动对象个数的确定牵扯到模式识别等技术，不在本文讨论的范围之内，本文假设运动对象的个数已知。

首先，将静止的物体（通常是背景）看成一层，它的运动向量为零，将这些点排除在外，以减少 EM 方法的计算量。应用 EM 算法将剩余的块分割成不同的层（对象）。

##### 4.1 E 步

假设每个运动模型参数  $u_j$ ，E 步计算了每个运动向量属于特定运动模型的后验概率：

$$P_j(B) = \Pr(\text{运动向量 } v(B) \text{ 属于第 } j \text{ 个运动模型 } |u_j) \quad (4)$$

这里，  $B = [p \ q]^T$  是运动向量对应的象素块在图像中的位置坐标，  $u_j^T = [v_x^j \ v_y^j]$  是第  $j$  个模型运动参数向量。设

$$r_j(B) = \|u_j - v(B)\|^2 \quad (5)$$

是第  $j$  个模型运动参数向量  $u_j$  和  $B$  处运动向量之差的平方。所以运动向量  $v(B)$  属于第  $j$  个运动模型的后验概率为

$$P_j(B) = e^{-r_j(B)/(2\sigma^2)} / \sum_{i=1}^N e^{-r_i(B)/(2\sigma^2)} \quad (6)$$

这里  $N$  是已知运动模型的个数，  $\sigma^2$  控制了运动模型适合运动向量场的保真度（典型值为 0.01~0.09）

##### 4.2 M 步

M 步重新估计每个模型（对象）的运动参数。使下列误差函数最小

$$u_j = \arg \min \sum_{vB} P_j(B) r_j(B) \quad (7)$$

可以得到运动模型更新的运动参数：

$$u_j = \left[ \sum_{vB} P_j(B) \right]^{-1} \sum_{vB} P_j(B) v(B) \quad (8)$$

具体地计算步骤如下：

- (1) 任意选取  $N$  个运动模型参数  $u_1^{(1)}, u_2^{(1)}, \dots, u_N^{(1)}$ 。
- (2) 在第  $i$  次迭代中，根据式 (5) 和式 (6) 计算每个运动向量属于第  $j$  个运动模型的概率  $P_j^{(i)}(B)$ 。
- (3) 根据式 (8) 更新运动模型参数。

(4) 如果  $u_j^{(i+1)} = u_j^{(i)}$ ,  $j = 1, 2, \dots, N$ , 那么迭代结束. 否则, 返回 (2) 继续执行.

经过 E 步和 M 步的几次迭代, 每个运动向量被分配给相应的运动模型. 这里我们采用了一个阈值  $T_P$  (典型值为 0.7~0.9), 将每个运动向量分给特定的运动模型, 根据:

$$Z_j(B) = \begin{cases} 1, & P_j(B) > T_P \\ 0, & \text{其他} \end{cases} \quad (9)$$

如果  $Z_j(B) = 1$ , 则表示运动向量  $v(B)$  属于第  $j$  个运动模型. 最后可以得到每个运动模型包含的运动向量, 而且得到了运动模型的平移运动参数  $u_j$ . 从 EM 算法的机理分析, 它不适合于新对象出现或是对象消失的情况.

运动向量代表像素块帧间的位移, 所以聚类了运动向量也就是聚类了像素块. 但是得到的是运动对象的初始定位和粗糙的轮廓, 精度是像素块的大小. 如果只需要粗略的轮廓, 这一步就够了, 但是我们往往需要对象的细节信息, 而单纯地利用运动向量进行分割, 无法得到精确的对象轮廓. 这就需要在像素域对运动对象进行处理.

### 5 对象的细节特征提取

在许多应用中, 往往需要知道对象的精确形状和像素域特征, 比如视频检索和对对象识别等. 这时压缩域的特征不足以满足要求. 但是我们利用压缩域的运动向量进行了对象的初始定位和粗略的对象轮廓提取. 于是我们只需对对象区域进行解码, 得到该区域的原始图像. 利用 Canny 边缘检测算子<sup>[9]</sup> 就可以得到运动对象的精确轮廓.

精确定位对象轮廓问题涉及到边缘的检测和跟踪. 一种典型的解决方法就是先利用边缘检测算子得到对象的边缘, 但低亮度比的边缘可能检测不到, 这时就要利用边缘跟踪来连接对象的边缘. 边缘连接的最基本问题就是确定边缘的跟踪方向. 这里我们利用最简单的边缘跟踪方向, 即跟踪的方向是第 4 节得到的粗略对象的边缘方向. 利用这种简单的边缘连接方法将轮廓连通, 得到比较精确的对象轮廓.

考虑到压缩码流中只有 I 帧有原始图像的 DCT 系数, 所以只对帧内编码的帧进行 IDCT 运算. 当初始分割是紧接 I 帧的 P 帧或 B 帧时, 将初始分割的模板用运动参数映射到 I 帧, 对 I 帧模板区域进行 IDCT 运算得原始图像. 在运动向量处理阶段, 对运动向量进行了检测、校正和中值滤波, 这些处理可能滤出了对象边缘的运动向量, 还有就是边缘的像素块的运动估计可能失败而没有运动向量, 这样模板区域可能小于对象区域 (如图 2), 于是我们对模板进行膨胀, 使得模板包含整个对象区域, 然后进行边缘检测和连接.

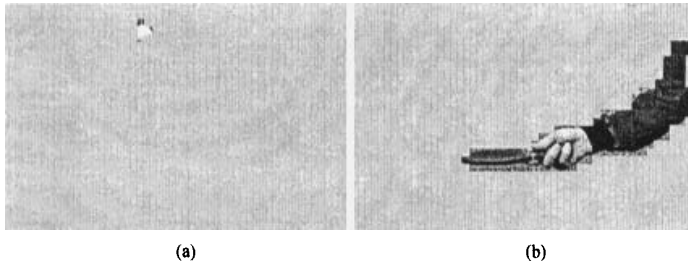


图 2 轮廓内区域进行 IDCT 得原始图像

### 6 试验结果

试验基于 H.263 编解码系统 (CMU Version 3.1), 如图 3 所示. 将 Table Tennis 测试序列经过 H.263 编码, 编码格式为 IPPPP. 部分解码压缩码流, 得到 I 帧的 DCT 系数和相邻 P 帧运动向量. 去除噪声运动向量的阈值  $T_{|MV|} = 1$ , 强度背离程度阈值  $T_M = 0.3$ , 方向背离程度阈值  $T_D = 0.523$  (即  $30^\circ$ ). EM 算法聚类时所取阈值  $\sigma^2 = 0.04$ ,  $T = 0.7$ , 于是得到 P 帧内运

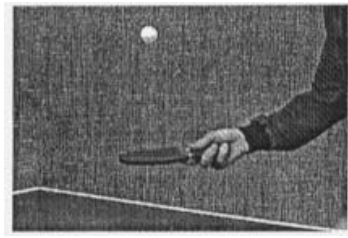
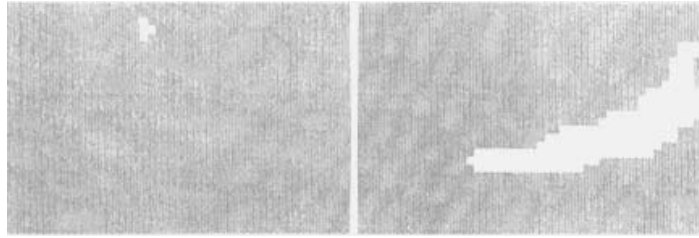


图 3 I 帧原始图像

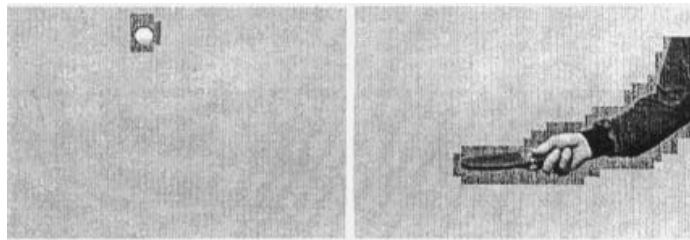
动对象的模板. 将运动对象模板映射到 I 帧 (如图 4), 并将模板向外膨胀  $8 \times 8$  像素块大小, 然后对模板内的块进行 IDCT 得到模板内原始图像 (如图 5), 再对模板内的图像进行 Canny 边缘检测和边缘跟踪得到运动对象精确边界和细节特征 (如图 6, 图 7).



(a)

(b)

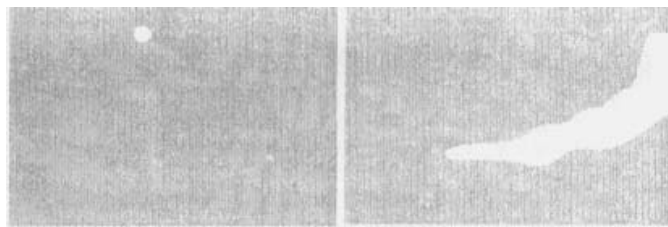
图 4 P 帧分割结果映射到 I 帧



(a)

(b)

图 5 膨胀后轮廓区域进行 IDCT 得原始图像



(a)

(b)

图 6 精确的对象轮廓

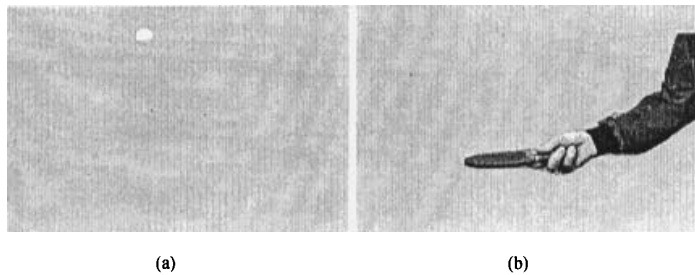


图 7 像素域分割结果

## 7 结论

本文提出的利用压缩域信息进行预分割和空域精细分割的方法简单、有效。不需要解码整个压缩数据,只需要提取运动向量和解码部分区域,所以整个过程需要的时间将比单纯的利用像素域分割的时间大大减少,提高了处理速度。虽然从压缩域提取的对象特征是粗糙的,但是它包含了足够的信息用于快速的视频浏览和对象跟踪。然而,对被提取对象的细节信息需要高级的检索和识别时,我们进一步解码相应的区域进行处理,得到了精确的对象轮廓。试验证实本文提出的分割系统是有效的。

## 参 考 文 献

- [1] Moscheni F, Bhattacharjee S, Kunt M. Spatiotemporal segmentation based on region merging. *IEEE Trans. on PAMI*, 1998, PAMI-20(9): 897-915.
- [2] Dufaux F, Moscheni F, Lippman A. Spatio-temporal segmentation based on motion and static segmentation. In Proc. of ICIP, Washington, USA, 1995: 306-309.
- [3] Torres L, Garcia D, Mates A. A robust motion estimation and segmentation approach to represent moving images with layers. In Proc. of ICASSP, Munich, Germany, 1997: 2661-2664.
- [4] Jarnrozik M L, Hayes M H. A compressed domain video object segmentation system. In Proc. of ICIP, New York: USA, 2002: 113-116.
- [5] Wang H, Chang S F. A highly efficient system for automatic face region detection in MPEG video. *IEEE Trans. on Circuits and Systems for Video Technology*, 1998, 7(4): 615-628.
- [6] Babu R V, Ramakrishnan K R. Compressed domain motion segmentation for video object extraction. In Proc. of ICASSP, Florida, USA, 2002: 3788-3791.
- [7] Wang R, Zhang H J, Zhang Y Q. A confidence measure based moving object extraction system built for compressed domain. In Proc. of ISCAS, Geneva, Switzerland, 2000: 21-24.
- [8] Dempster A, Laird N, Rubin D. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 1977, 39(1): 1-38.
- [9] Canny J F. A computational approach to edge detection. *IEEE Trans. on PAMI*, 1986, PAMI-8(6): 679-698.

王占辉: 男, 1978 年生, 硕士生, 主要研究兴趣包括模式识别、图像处理、视频分析等。  
 刘贵忠: 男, 1962 年生, 教授, 博士生导师, 1989 年 6 月获荷兰 Eindhoven 大学博士学位, 主要从事非平稳信号的分析与处理、音视频数据压缩、模式识别、反演等方面的理论、方法与应用研究。  
 刘 龙: 男, 1976 年生, 博士生, 主要研究兴趣包括基于对象的视频编码技术、视频分析等。  
 刘洁瑜: 女, 1970 年生, 博士生, 主要研究兴趣包括基于对象的视频编码、无线视频传输技术等。