

## 局部运动场景中运动对象的压缩域分割算法

刘 龙 韩崇昭 刘洁瑜 王占辉  
(西安交通大学电子与信息工程学院 西安 710049)

**摘 要:** 压缩域视频运动对象的分割技术近年来逐渐被人们所重视, 并且成为较实用的技术之一。该文针对目前压缩域中运动对象分割质量不高的问题, 提出一种压缩域分割的改进算法。该算法首先从 MPEG 压缩码流中提取运动矢量, 并对矢量进行运动累加和中值滤波处理, 最后利用贝叶斯风险决策对非零运动矢量区域进行校正, 最终得到较好的分割结果。该文还定义了压缩域分割质量的评价参数分割率和错分率, 为评价压缩域的分割结果提供了客观评价标准。

**关键词:** MPEG, 分割, 压缩域

**中图分类号:** TN919.8

**文献标识码:** A

**文章编号:** 1009-5896(2005)08-1249-05

## Compressed Domain Moving Object Segmentation in Local Motion Scene

Liu Long Han Chong-zhao Liu Jie-yu Wang Zhan-hui

(Department of Information & Communication Engineering, School of Electronics & Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract** More and more attention has been paid to the study of moving object segmentation in MPEG compressed domain these years, which has already developed into a practical technology. In this paper, an algorithm by using motion vectors and DCT coefficients is put forward for segmenting moving object in MPEG compressed domain. As regards local motion scene, the algorithm produces a better result of segmentation than currently adopted ones. Firstly, the motion vectors are processed by accumulation and median filter; then the areas of non-zero motion vectors are rectified by the Bayesian decision to get high-quality segmentation.

**Key words** MPEG, Segmentation, Compressed domain

### 1 引言

近年来, 视频对象的存储和操作技术日益为人们所重视。视频对象的分割技术是指把视频帧图像按一定的标准分割成区域, 目的是为了从视频序列中分离出有一定意义的实体。对于局部运动场景, 人们往往希望通过视频分割技术提取前景运动对象, 并对其进行操作。如何从数字视频信号中分割和提取视频运动对象是进一步操作和处理的关键环节, 因此视频运动对象的分割技术已经成为当前视频研究领域的热点问题之一。

目前视频分割技术多数是基于像素域的算法, 在压缩域中分割视频运动对象的研究尚不多。基于压缩域的分割方法在减少计算复杂度方面有较大的优势, 但在分割的精确性上逊于像素域中的分割算法。尽管在压缩域中的分割精度不够高, 但目前大量的多媒体数据是以压缩形式存储和传输的,

所以研究压缩域中的视频运动对象分割算法有很大的实用价值。一项重要的应用框架如图 1 所示, 其目的是直接从压缩码流中提取视频运动对象, 并对其进行操作。

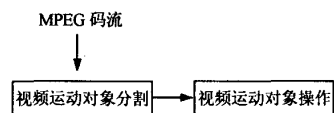


图 1 视频对象的操作

文献[1-6]均为压缩域的分割算法, 大致分为两类, 一类是利用视频压缩域提供的运动矢量进行分割, 如文献[1,2], 另一类是利用 DCT 系数或颜色进行分割, 如文献[3-6]。Venkatesh 和 Ramakrishnan<sup>[1,2]</sup>通过对压缩域中的运动矢量进行 K-means 聚类分析等处理, 并且采用 EM 聚类算法来获得运动对象的数目。Wang 和 Chang<sup>[3]</sup>提出一种在 MPEG 压缩码流中快速确定人脸的算法, 他们使用了皮肤色度的统计数

据, 形状和亮度 DCT 系数的能量分布来确定人脸区域。Sukmarg 和 Rao<sup>[4]</sup>采用文献[3]的算法确定视频对象的区域, 然后采用自适应 K-means 算法对颜色信息进行空间聚类, 最后部分解码进行对象细化。Eng 和 Ma<sup>[5]</sup>用压缩域中的空间信息(颜色信息和 DCT 系数)来对视频对象进行分割。Ji 和 Park<sup>[6]</sup>利用 DCT 块的空间相关性进行分割。文献[3-6]中的算法均不是针对局部运动场景中的运动对象分割, 只有 Venkatesh 和 Ramakrishnan 提出的运动矢量分析算法<sup>[1,2]</sup>是针对局部场景中的运动对象进行分割的, Venkatesh 和 Ramakrishnan 在文献[2]中的算法是文献[1]提出算法的改进, 所以文献[2]算法分割效果比文献[1]质量更好, 但是分割结果仍存在较大的误差。本文算法主要与文献[2]的算法进行了对比。

本文针对局部运动场景中运动对象的分割问题, 提出一种新的压缩域分割算法。该算法首先从 MPEG 压缩码流中提取运动矢量场, 对运动矢量进行积累和中值滤波处理, 最后利用贝叶斯决策对非零运动矢量区域进行校正, 最终得到较好的分割结果。Venkatesh 和 Ramakrishnan 提出的分割算法流程和本文提出的分割算法流程如图 2 所示。

本文的结构安排如下: 第 2 节介绍运动矢量处理, 第 3 节为校正, 第 4 节为试验结果, 最后第 5 节给出了本文的结论。

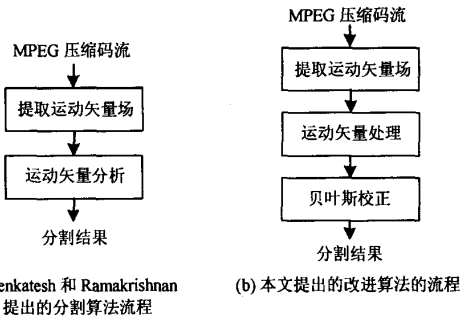


图 2 基于 MPEG 压缩域的二次分割流程结构

## 2 运动矢量处理

在像素域中, 分割算法利用空间信息与运动信息进行分割。空间信息包括灰度, 颜色等; 运动信息主要包含运动矢量等。经过视频压缩处理后, 像素域中的颜色和灰度转换为压缩码流中的矢量和 DCT 系数, 矢量和 DCT 系数经过熵编码最终形成压缩码流, 所以从压缩码流中可以获得运动矢量场和 DCT 系数。

MPEG 视频压缩码流是根据 MPEG1/2 视频压缩编码标准<sup>[7,8]</sup>而生成的视频压缩码流。在编码过程中, 视频帧被分为

3 种: I 帧, P 帧和 B 帧。I 帧为基帧, P 帧为预测帧, 在解码端需要根据运动矢量从 I 帧中恢复; B 帧属于双向预测帧。I 帧的编码模式为帧内编码, P 帧和 B 帧的编码模式为帧间编码。I 帧, P 帧和 B 帧的预测关系可见图 3。

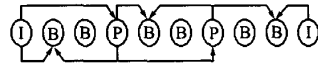


图 3 帧间预测示意图

帧内编码模式是将一帧图像划分为多个 8×8 大小的块, 然后对每一个块进行 DCT 变换, 量化和熵编码的处理, 最后按照扫描顺序排列形成二进制码流。帧间编码主要处理运动矢量场, 运动矢量场估计是 P 帧基于 I 帧或 P 帧的宏块运动估计, 其中的每一个运动矢量分别代表每一个宏块的运动轨迹, 运动矢量通过熵编码最终也排列到二进制码流中。在解码端, MPEG 码流经过反熵编码可以直接获得 P 帧的运动矢量场, 以及 I 帧的二维 DCT 系数。

本文中所采用的矢量是指 I 帧后的第一个 P 帧与其参考帧 I 帧形成的运动矢量, 其反映的是 I 帧内的宏块相对于该 P 帧的运动轨迹, 每个宏块都对应着一个运动矢量。称帧内具有非零运动矢量的宏块区域为非零运动矢量区域, 具有非零运动矢量的宏块很有可能属于运动对象的一部分, 所以非零运动矢量区域可以大致确定为运动对象区域, 因此文献[1,2]直接从 MPEG 码流中提取的运动矢量, 并经过一定的分析运动对象的分割。

运动矢量的处理包含两个部分: 运动矢量累加处理和滤波处理。视频运动对象在场景推进过程中会出现局部的停滞或者突变。为了提取真正反映视频对象运动的矢量就需要将当前帧前后几帧的运动矢量叠加, 本文把这一过程称为运动矢量累加。

过程采用的是 P 帧基于 I 帧估计的运动矢量, 每个宏块的运动矢量累加如下所述: 设当前 I 帧为第  $n$  帧, 宏块的中心坐标为  $(k, l)$ , 与之对应的运动矢量表示为  $(m_x^{k,l}(n), m_y^{k,l}(n))$ , 与前后 I 帧的运动矢量累加按式(1)计算:

$$(m_x^k, m_y^l) = \sum_{i=n-c}^{i=n+c} (m_x^{k,l}(i), m_y^{k,l}(i)) \quad (1)$$

在本文中考虑到后续的处理, 因此只采用了当前 I 帧的前 I 帧和后 I 帧进行累加( $c=1$ ), 如图 4 所示。

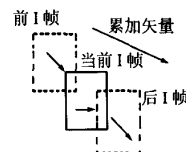


图 4 运动矢量累加

MPEG 标准<sup>[7,8]</sup>利用块匹配法, 并采用最小均方误差准则估计 P 帧中宏块相对于 I 帧中宏块的运动矢量, 这样得到的运动矢量中有些非零运动矢量是由于前后两帧中存在噪声而产生的。在视频场景中, 这些非零运动矢量所对应的宏块实际并未运动, 我们把这些非零运动矢量称为非运动信息矢量。编码解码关心的只是 P 帧的恢复, 并不处理非运动信息矢量, 但用这样的运动矢量来分割运动对象则会造成很大的偏差。为了去除运动矢量累加后带来的误差, 需要对累加后的运动矢量进行中值滤波处理从而得到真正反映视频对象运动的运动矢量。运动矢量中值滤波过程如下: 对于每一个非零运动矢量, 用相邻的运动矢量中值代替其值。

运动矢量通过累加和中值滤波可以得到用于当前 I 帧视频分割的运动矢量场, 如图 5 所示。

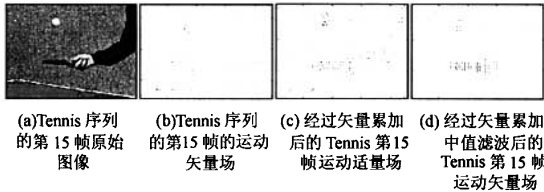


图 5 运动矢量场的累加和中值滤波

### 3 贝叶斯校正

视频分割质量的好坏在于能否较准确地将场景中运动对象区域与背景区域进行分离, 图 6 显示了 Tennis 序列的第 15 帧运动对象区域和背景区域。视频场景的运动变化分为两种: 一种是局部运动, 另一种是全局运动。局部运动的场景背景是不随时间变化移动的, 全局运动的场景背景和运动对象都是随时间变化而移动的。对于局部运动场景, 对象运动有时存在局部静止, 这时整个场景分为零运动矢量区域和非零运动矢量区域, 非零运动矢量区域包含在运动对象区域内, 零运动矢量区域包含了静止区域和对象区域的一部分, 它们的关系如图 6 所示。要取得较好的分割质量需要使得非零运动矢量区域尽量接近对象区域。针对局部运动场景中运动对象的分割, 本文在对运动矢量进行运动积累和中值滤波处理的基础上, 利用贝叶斯风险决策对非零运动矢量区域进行校正, 从而得到较准确的运动对象区域, 较好地提高了分割的质量。

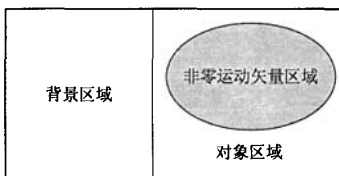


图 6 区域关系

本文所称的校正就是要将对象区域中的 DCT 块纳入非零运动矢量区域, 最终使其接近对象区域。在 DCT 系数块中, 较大系数分布主要集中在直流系数, 一般大小的系数主要分布在最上边一行和最左边一列, 较小的系数分布在其它点, 根据这个特点定义 DCT 系数块的特征值  $X$ , 计算如下:

$$X = \frac{v(0,0) + \sqrt{\sum_{j=1}^{N-1} v(0,j)^2} + \sqrt{\sum_{i=1}^{N-1} v(i,0)^2} + \sqrt{\sum_{i=1}^{N-1} \sum_{j=1}^{N-1} v(i,j)^2}}{4} \quad (2)$$

其中  $v(i,j)$ ,  $0 \leq i, j \leq N-1$ , 是  $N \times N$  大小块的 DCT 系数。

在局部运动场景中, 背景一般与运动对象有明显的区别, 因此可以认为 DCT 块特征值满足两个条件: (1)背景区域内的 DCT 块特征值分布为均匀分布并且分布广泛; (2)对象区域内的 DCT 块特征值分布为正态分布, 并且分布范围较集中。贝叶斯校正过程如下:

$$R_0(X) = C_{00} \cdot \rho(\omega_0 / X) + C_{10} \cdot \rho(\omega_1 / X) \quad (3)$$

$$R_1(X) = C_{01} \cdot \rho(\omega_0 / X) + C_{11} \cdot \rho(\omega_1 / X) \quad (4)$$

$$R_0(X) < R_1(X) \Rightarrow X \in \omega_0 \quad (5)$$

$$R_0(X) > R_1(X) \Rightarrow X \in \omega_1 \quad (6)$$

其中  $\omega_0$  和  $\omega_1$  分别代表对象区域和背景区域。  $\rho(\omega_1 / X)$  代表先验概率, 即在给定样本特征值  $X$  的情况下是  $\omega_1$  的概率。  $C_{00}$  和  $C_{11}$  是正确分类的代价权值,  $C_{01}$  和  $C_{10}$  是错误分类的代价权值。式(7)是贝叶斯公式,  $\rho(X / \omega_i)$  代表条件背景块或对象块的概率密度函数:

$$\rho(\omega_i / X) = [\rho(X / \omega_i) \cdot \rho(\omega_i)] / \rho(X) \quad (7)$$

利用贝叶斯公式, 代入式(3), (4), 则可以得到式(8),(9)

$$\rho(X / \omega_0) > \tau \Rightarrow X \in \omega_0 \quad (8)$$

$$\rho(X / \omega_0) < \tau \Rightarrow X \in \omega_1 \quad (9)$$

其中

$$\tau = \frac{C_{10} - C_{11}}{C_{01} - C_{00}} \cdot \frac{\rho(\omega_1)}{\rho(\omega_0)} \rho(X / \omega_1)$$

$(C_{10} - C_{11}) / (C_{01} - C_{00})$  是个可以调节的值;  $\rho(\omega_1) / \rho(\omega_0)$  可以被认为是个固定值; 由于运动块的 DCT 特征值通常在很小的范围内, 所以静止块的 DCT 系数分布可以认为是均匀分布, 因此可以  $\rho(X / \omega_1)$  可以近似认为是个常数。因此  $\tau$  在实验中是个可以调节的阈值, 在实验中需要所有的 DCT 系数块的特征值归一化后进行贝叶斯判断。因为在不同的视频序列中的视频对象其 DCT 系数块的特征值  $X$  的条件概率分布参数是不同的, 所以阈值  $\tau$  针对不同的视频对象而变化的。阈值  $\tau$  的确定要根据对所分割的视频对象的实验统计得到。

假设非零运动矢量区域内有若干个 DCT 块，这些 DCT 块的特征值则可以作为对象 DCT 块特征值分布的样本。由于对象 DCT 块特征值分布成正态分布，如公式(10):

$$\rho(X/\omega_1) = \frac{\exp\left(-\frac{1}{2}(x-\varepsilon)^T \sigma(x-\varepsilon)\right)}{2\pi\sqrt{\sigma}} \quad (10)$$

则其期望  $\varepsilon$  和方差  $\sigma^2$  值为非零运动矢量区域内的所有 DCT 系数块特征值的均值和二矩矩。

在校正过程中，对零运动矢量区域内的 DCT 块进行贝叶斯判断，将判断为对象块的 DCT 块纳入非零运动矢量区域，最终得到分割结果。图 7 为通过贝叶斯校正得到分割结果的过程。

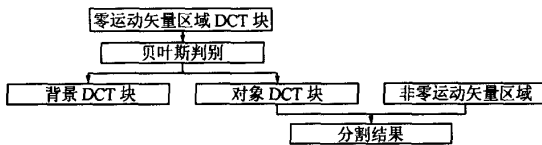


图 7 贝叶斯校正

### 4 试验结果分析

由于本文提出的算法针对局部运动场景中的视频对象分割，因此选取 Tennis, Silence 和 Hall Monitor 3 种局部运动的视频场景序列来验证本文提出的压缩域的分割算法，与 Venkatesh 和 Ramakrishnan 在文献[1,2]中提出的分割的算法进行比较。

#### 4.1 评价参数

目前，所有的视频压缩域的分割算法的文献中都使用主观对比的方法来对分割质量进行评价。为了从统计上分析压缩域分割质量的好坏，本文引入了分割率和错分率两个评价参数。

在压缩域视频运动对象的分割中，分割结果常常不是完善的，在这种情况下总有一些属于背景的块被错误地划分到运动对象区域，我们称这些块为“错分块”；另一方面，有些属于运动对象区域的块未被划分为背景区域，我们称这些块为“未分块”。图 8 显示了 Tennis 的第 15 帧分割后的错分块和未分块。



(a) Tennis 序列的第 15 帧运动对象区域 (b) 分割图像的错分块和未分块示意

图 8 错分块和未分块

本文定义两个参数分割率 Pocs (percentage of correctly segmented)和错分率 Poes (percentage of error segmented) 来评价压缩域的分割质量，计算公式如下:

$$Pocs = \frac{\text{对象块数} - \text{未分块数}}{\text{对象块数}} \quad (11)$$

$$Poes = \frac{\text{错分块数}}{\text{分割块数}} \quad (12)$$

分割率反映分割区域包含运动对象的程度，分割率越高，说明分割区域中包含运动对象的程度越高，反之亦然；错分率反映分割运动对象的准确程度，错分率越高，说明分割运动对象的准确性越差，反之亦然。

#### 4.2 统计分析

记文献[2]中的算法为算法 1，本文提出的算法为算法 2。从 Tennis, Silence 和 Hall Monitor 3 种局部运动的视频场景序列中分别抽取 5 帧作分割试验，试验参数 Pocs 和 Poes 的均值列在表 1 和表 2 中。

表 1 Pocs 参数试验统计表

	Tennis	Silence	Hall monitor
算法 1	15%	3%	4%
算法 2	16%	3%	5%

表 2 Poes 参数试验统计表

	Tennis	Silence	Hall monitor
算法 1	81%	62%	89%
算法 2	96%	93%	96%

表 1 中的数据表明算法 2 在分割区域中的包含运动对象程度明显高于算法 1，说明本文提出的校正使得原先的一部分伪静止块成为分割区域的一部分，从而提高分割质量；表 2 中的数据表明算法 1 与算法 2 的错误分割数基本相当，说明校正的没有使大量的静止块纳入最终的分割区域。从两个参数的权衡对比看，本文提出的算法 2 在对局部运动场景中的运动对象进行分割的质量高于算法 1。

#### 4.3 主观对比

分别对 Tennis 序列的第 15 帧, Silence 序列的第 73 帧和 Hall Monitor 的第 41 帧采用 3 种不同算法进行分割试验，结果如图 9—图 11 所示。



(a) Tennis 序列的第 15 帧原始图像 (b) 算法 1 的分割结果 (c) 算法 2 的分割结果

图 9 Tennis 分割结果

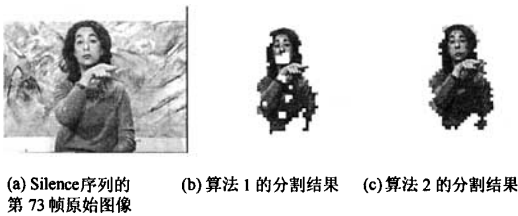


图 10 Silence 分割结果

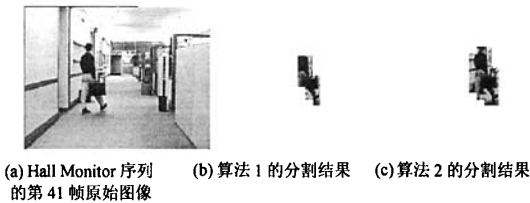


图 11 Hall Monitor 分割结果

如图 9—图 11 所示, 本文提出的运动对象压缩域的分割算法在主观上比 Venkatesh 和 Ramakrishnan 在文献[2]中提出的基于运动矢量分割的算法效果好。

## 5 结束语

本文针对目前压缩域中运动对象分割质量不高的问题, 提出一种压缩域分割的改进算法。引入了 DCT 块特征值的概念。该算法从 MPEG 压缩码流中提取运动矢量场与 DCT 系数分割视频运动对象。首先对矢量场进行运动累加和中值滤波处理, 最后利用贝叶斯风险决策对非零运动矢量区域进行校正。本文还定义了压缩域分割质量的评价参数: 分割率和错分率。实验证明本文提出的压缩域分割算法对局部运动场景中运动对象分割有很好的效果。

## 参 考 文 献

[1] Venkatesh R, Ramakrishnan K R. Content-based video retrieval using motion descriptors extracted from compressed domain. IEEE International Symposium on Circuits and Systems(ISCAS),

Arizona, 2002, vol. 4: 141 – 144.

- [2] Venkatesh R, Ramakrishnan K R. Compressed domain motion segmentation for video object extraction. Int'l Conf. on Acoustics, Speech, and Signal Processing, Orlando, Sept. 2002, vol.4: 3788 – 3791.
- [3] Wang H, Chang S F. A highly efficient system for automatic face region detection in MPEG video. *IEEE Trans. on Circuit and System for Video Technology*, 1997, 7(4): 615 – 628.
- [4] Orachat Sukmarg, Rao K R. Fast object detection and segmentation in mpeg compressed domain. TENCON 2000, Kuala Lumpur, Malaysia, Sept. 2000: 1769 – 1770.
- [5] Eng How-Lung, Ma Kai-Kuang. Spatiotemporal segmentation of moving video objects over mpeg compressed domain. IEEE Int'l Conf. on Multimedia and Expo(ICME), New York, 2000, Vol.3: 1531 – 1534.
- [6] Ji Salkmann, Park Hyun Wook. Moving object segmentation in DCT-based compressed video. *Electronics Letters*, 2000, 36(21): 1769 – 1770.
- [7] ISO/IEC.JCT1/SC29 CD11172-2, MPEG1, international standard for coding of moving pictures and associated audio for digital storage media at up to about 1.5Mb/s. 1991.
- [8] ISO/IEC.JCT1/SC29 CD13818-2, MPEG2, coding of moving pictures and associated audio. 1993.

刘 龙: 男, 1976 年生, 博士生, 研究兴趣包括基于对象的视频编码与检索技术、无线视频传输技术和视频分析等。

韩崇昭: 男, 1942 年生, 教授, 博士生导师, 主要研究兴趣为信息融合、雷达信号处理、图像处理等。

刘洁瑜: 女, 1970 年生, 博士生, 研究兴趣包括多媒体技术、小波理论、时频分析等。

王占辉: 男, 1978 年生, 硕士生, 研究兴趣包括图象处理、视频分析与传输控制技术、小波分析等。