

一种基于 MPEG 压缩域的运动对象分割算法

刘 龙, 刘贵忠, 刘洁瑜, 王占辉

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 针对目前视频压缩域中运动对象分割准确率不高的问题, 提出一种有效的压缩域运动对象的分割算法. 该算法先对运动矢量进行运动累加和中值滤波处理, 然后利用离散余弦变换 (DCT) 系数块之间的相关性对非零运动矢量区域进行校正. 由于在对运动矢量处理的同时加入了空间 DCT 块的相关性校正, 因此分割准确率有了较大的提高. 实验统计结果表明, 对几种不同视频场景中的运动对象进行分割, 其分割率比现有的算法提高了约 15%.

关键词: 分割; 压缩域; 运动对象

中图分类号: TN911.173 **文献标识码:** A **文章编号:** 0253 - 987X(2004)12 - 1264 - 04

New Moving Object Segmentation Algorithm in MPEG Compressed Domain

Liu Long, Liu Guizhong, Liu Jieyu, Wang Zhanhui

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

Abstract: Focusing on the problem that there is low segmentation veracity for moving objects in the video compressed domain, an effective segmentation algorithm was proposed, in which the motion vectors were processed by motive accumulation and median filter, and then the areas of non-zero motion vectors were corrected by using the relativity among discrete cosine transform (DCT) coefficient blocks. Because of adding the relativity correction of spatial DCT blocks while motion vectors were processing, the segmentation veracity is improved greatly. For various video scenes, the experiments results show that the segmentation veracity of the proposed algorithm is improved about 15% compared to that of the other algorithms.

Key words: segmentation; compressed domain; motion vector

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. 基于像素域的分割方法在分割的精确性上要高于在压缩域中的分割, 但是其计算复杂度高. 基于压缩域的分割方法在减少计算复杂度方面有较强的优势, 但在分割的精确性上要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. 文献[1 - 6]均为压缩域中的分割算法, 大致分为 2

类: 一是利用视频压缩域提供的运动矢量进行分割^[1 - 2], 二是利用离散余弦变换 (DCT) 系数或颜色进行分割^[3 - 6].

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, 然后利用 DCT 系数块之间的相关性对非零运动矢量区域进行校正. 实验结果证明, 该算法分割的准确程度有较大的提高.

收稿日期: 2004 - 04 - 14. 作者简介: 刘 龙 (1976 ~), 男, 博士生; 刘贵忠 (联系人), 男, 教授, 博士生导师. 基金项目: 国家自然科学基金资助项目 (60272072); 教育部跨世纪优秀人才基金资助项目; 教育部“十五”211 工程西安交通大学重点科研项目.

1 运动矢量处理

本文对从 MPEG 码流中提取的运动矢量进行处理. 运动矢量的处理包含 2 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累加,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,其中某个宏块的运动矢量表示为 $[m_x(n), m_y(n)]$,它与前后 I 帧的运动矢量累加为

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

式中: i 表示帧序; c 表示当前帧的前后帧数; $[m_x(i), m_y(i)]$ 表示第 i 帧的运动矢量.

本文考虑到后续的处理,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 1 所示.

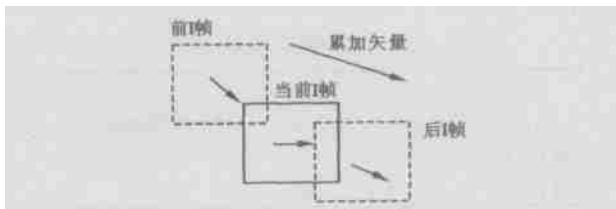


图 1 运动矢量累加

MPEG 标准^[7-8]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. 本文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 帧视频分割的稠密运动矢量场.

2 非零运动矢量区域的校正

211 DCT 系数块的相关性

在像素域中,帧内的像素间存在相关性,同样,

在压缩域中,DCT 系数块之间也存在相关性. 我们把 DCT 系数中具有相关性块称为相似块. 为了判断块间的相关性,定义 DCT 系数块的空间特征向量

$$f = (f_0, f_1, f_2, f_3) = (v(0, 0), \left(\sum_{j=1}^{N-1} v^2(0, j) \right)^{\frac{1}{2}}, \left(\sum_{i=1}^{N-1} v^2(i, 0) \right)^{\frac{1}{2}}, \left(\sum_{i=1}^{N-1} \sum_{j=1}^{N-1} v^2(i, j) \right)^{\frac{1}{2}}) \quad (2)$$

式中: $v(i, j)$ ($0 \leq i, j \leq N-1$) 是大小为 $N \times N$ 块的 DCT 系数; f_0 表示 DCT 系数块的直流系数值; f_1, f_2 和 f_3 分别表示 DCT 系数块的水平、垂直和对角特征值.

假设 2 个 DCT 系数块的空间特征矢量为 f_1 和 f_2 ,若 $|f_1 - f_2| < T$, T 一般取量化参数 (QP) 的 4/3 倍,则认为这 2 个 DCT 块相关. 相关块合并后的区域称为相似区域, n 个相关块合并成相似区域,其空间特征向量为

$$= \frac{1}{n} \sum_{i=1}^n f_i \quad (3)$$

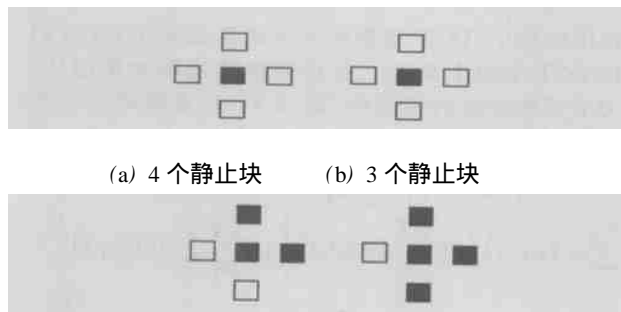
式中: f_i ($1 \leq i \leq n$) 表示第 i 个块的空间特征向量.

212 校正

本文将属于非零运动矢量区域的 DCT 系数块称为运动块,反之称为静止块. 将不属于运动对象区域但属于非零运动矢量区域的 DCT 系数块称为伪运动块. 将属于运动对象区域但不属于非零运动矢量区域的 DCT 系数块称为伪静止块. 校正就是把非零运动矢量区域中的伪运动块去除,并把伪静止块划归非零运动矢量区域,从而使校正后的非零运动矢量区域与运动对象区域尽可能一致. 校正的关键就是要辨别伪运动块和伪静止块. 本文认为伪运动块和伪静止块基本都集中在运动块与静止块的混合区域,图 2 显示了以运动块为中心、运动块与静止块混合分布的 4 种可能,黑色块代表运动块,白色块代表静止块.

图 2 所示的 4 种情况要根据 DCT 系数块的相关性判别伪运动块和伪静止块,并进行校正. 将以运动块为中心的 5 个相邻块记为集合 S ,其中运动块子集合为 S_m ,静止块子集合为 S_s ,那么 $S = S_m \cup S_s$,校正的过程如下:

- (1) 在分割帧找到如图 2 所示的一个中心运动块,记为 a ,其特征向量记为 f_a ;
- (2) 在集合 S 中合并相似区域,记为 S' ,若 $|f_a - f_i| < T$,则执行步骤(3),若 $|f_a - f_i| \geq T$,则认为中心运动块为伪运动块,在非零运动矢量区域中去除该块,寻找



(a) 4 个静止块 (b) 3 个静止块

(c) 2 个静止块 (d) 1 个静止块

图 2 边界区域运动块与静止块混合分布的 4 种情况

另外的中心运动块,重新执行步骤(2);

(3) 按公式(3)计算空间特征向量,记为 f_a ;

(4) 在集合 S 中任取一静止块,记为 $b_i, 1 \leq i \leq n$. 按公式(2)计算 b_i 空间特征向量,记为 f_{b_i} ;

(5) 如果 $|f_a - f_{b_i}| > |f_a - f_{a'}|$,则认为 b_i 是伪静止块,将该块划归中心运动块 a 所属的非零运动矢量区域;

(6) 返回执行步骤(3),直到集合 S 中所有静止块取完;

(7) 返回执行步骤(1)直到分割帧内所有符合图 2 中 4 种情况的运动块都处理完毕,结束执行.

以上循环执行 3~6 次就能得到比较满意的分割结果.

3 实验结果分析

3.1 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对分割质量进行评价. 为了从统计意义上分析压缩域分割质量的优劣,本文引入了分割率和错分率 2 个评价参数.

在压缩域视频运动对象的分割中,分割结果常常是不完善的,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,这些块被称为错分块. 另一方面,有些属于运动对象区域的块被划分为背景区域,这些块被称为未分块. 图 3 显示了 Tennis 的第 15 帧分割后的错分块和未分块,其中黑色块表示错分块,白色块表示未分块.

本文采用分割率 p_{ocs} 和错分率 p_{oes} 来评价压缩域的分割质量,计算公式如下

$$p_{ocs} = \frac{\text{对象块数} - \text{未分块数}}{\text{对象块数}} \quad (4)$$

$$p_{oes} = \frac{\text{错分块数}}{\text{分割块数}} \quad (5)$$

分割率 p_{ocs} 反映了分割区域包含运动对象的程

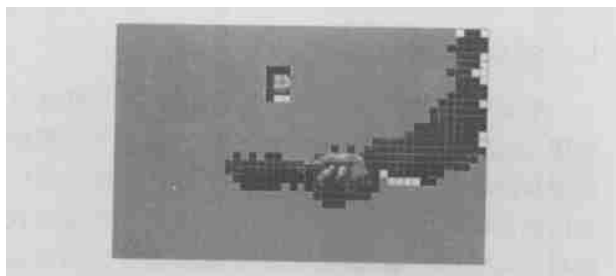


图 3 分割图像的错分块和未分块示意

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 p_{oes} 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.

3.1.2 统计分析

记文献[2]中的算法为算法 1,本文提出的改进算法为算法 2. 从 Tennis、Silence 和 Hall Monitor 这 3 种局部运动的视频场景序列中分别抽取 5 帧作分割实验,实验参数 p_{ocs} 和 p_{oes} 的均值列在表 1 和表 2 中.

表 1 分割率实验 %

	p_{ocs}		
	Tennis	Silence	Hall Monitor
算法 1	81	62	89
算法 2	96	92	97

表 2 错分率实验 %

	p_{oes}		
	Tennis	Silence	Hall Monitor
算法 1	15	3	4
算法 2	16	3	5

表 1 中的数据表明,算法 2 在分割区域中包含运动对象的程度明显高于算法 1,说明本文提出的校正法使得原先的一部分伪静止块成为分割区域的一部分,从而提高了分割质量. 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,说明校正没有使大量的静止块纳入最终的分割区域. 从 2 个参数的权衡对比看,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.

3.1.3 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 2 种不同算法进行分割实验,结果见图 4~图 6.

从图 4~图 6 可以看出,本文提出的运动对象压缩域的分割算法在主观上比 Venkatesh 和 Ra2 makrishnan^[2]提出的基于运动矢量分割的算法效果要好.



(a) 算法 1 的分割结果 (b) 算法 2 的分割结果

图 4 Tennis 序列的第 15 帧分割结果



(a) 算法 1 的分割结果 (b) 算法 2 的分割结果

图 5 Silence 序列的第 73 帧分割结果



(a) 算法 1 的分割结果 (b) 算法 2 的分割结果

图 6 Hall Monitor 序列的第 41 帧分割结果

4 结 论

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 块特征矢量的概念.改进该算法利用 MPEG 压缩码流中的运动矢量场与 DCT 系数分割视频运动对象,再对矢量场进行运动累加和中值滤波处理,然后通过空间块的相关性对非零运动矢量区域进行校正,最终得到高质量的分割结果.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.实验结果表明,本文提出的分割算法优于以往的分割算法,取得了较好的分割效果.

参考文献:

- [1] Venkatesh R, Ramakrishnan K R. Content-based video retrieval using motion descriptors extracted from compressed domain [A]. IEEE International Symposium on Circuits and Systems, Arizona, USA, 2002.
- [2] Venkatesh R, Ramakrishnan K R. Compressed domain motion segmentation for video object extraction [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, USA, 2002.
- [3] Wang H, Chang S F. A highly efficient system for automatic face region detection in MPEG video [J]. IEEE Trans on Circuit and System for Video Technology, 1997, 7(4): 615 - 628.
- [4] Orachat S, Rao K R. Fast object detection and segmentation in MPEG compressed domain [A]. TENCON, Kuala, Malaysia, 2000.
- [5] Eng H, Ma K K. Spatiotemporal segmentation of moving video objects over MPEG compressed domain [A]. IEEE International Conference on Multimedia and Expo, New York, 2000.
- [6] Ji S, Park H W. Moving object segmentation in DCT based compressed video [J]. IEE Electronics Letters, 2000, 36(21): 1769 - 1770.
- [7] ISO/IEC. JCT1/SC29 CD1117222, MPEG1, international standard for coding of moving pictures and associated audio for digital storage media at up to about 115Mb/s [S].
- [8] ISO/IEC. JCT1/SC29 CD13818 - 2, MPEG2, coding of moving pictures and associated audio [S].

(编辑 刘 杨)

本刊“学术争鸣与评论”栏目征稿启事

为了活跃学术气氛,开展学术讨论,本刊研究决定自 2005 年开始,开设“学术争鸣与评论”栏目,从现在开始面向国内外读者和审者广泛征集稿件,具体要求如下.

(1) 评论稿件是对本刊已发表学术论文中存在的缺陷或可能的失误提出讨论和评述,并尽量给出评论者认为更好或正确的观点,也可以对本刊已发表的论文中的某些在学术界有争议的观点进行争鸣.

(2) 批评和评论应当本着实事求是的态度,对自己提出的观点和解释提供必要而且充分的佐证资料,避免泛泛而谈,并不得对作者进行人身攻击.

(3) 评论要求有理有据、短小精悍,一般在 1 000 个汉字字符以内,稿酬从优.

(4) 除特殊情况以外,来稿一般不进行同行审稿,但会尽力联系被评论论文的作者,请作者提出答辩,需要时也安排发表.

(5) 评论稿件可以署笔名发表,但来稿时应当提供真实姓名和具体联系方式.

投稿邮箱:710049 西安市咸宁西路 28 号 西安交通大学学报编辑部

本刊网址: <http://journal.xjtu.edu.cn/zrb> E-mail: xuebao@mail.xjtu.edu.cn