

基于熵模型的压缩域运动对象检测

徐剑峰¹⁾ 刘志^{1),2)} 张兆杨^{1),2)}

¹⁾(上海大学通信与信息工程学院, 上海 200072) ²⁾(新型显示技术及应用集成教育部重点实验室, 上海 200072)

摘要 随着视频编解码技术的发展, H. 264 已成为最主要的标准之一。为了能从 H. 264 码流中准确有效地分割和提取出运动对象, 提出了一种压缩域下的对象检测算法。该算法主要利用压缩域下对象的运动矢量信息, 先对矢量进行中值滤波预处理, 目的是为了减少运动估计算法和对象实际运动所产生的误差。然后利用基于熵的原理, 建立运动对象在空间和时间上的一致性模型。在该模型基础上, 采用最大熵方法自适应获得阈值, 检测得到运动对象。实验结果证明, 该算法可以获得比较好的检测结果。

关键词 压缩域 对象检测 H. 264

中图分类号: TN941.1 **文献标识码:** A **文章编号:** 1006-8961(2007)10-1815-04

Compressed Domain Moving Object Detection Based Entropy Model

XU Jian-feng¹⁾, LIU Zhi^{1),2)}, ZHANG Zhao-yang^{1),2)}

¹⁾(School of Communication and Information Engineering, Shanghai University, Shanghai 200072)

²⁾(Key Laboratory of Advanced Display and System Application, Ministry of Education, Shanghai 200072)

Abstract With the development of video encoder and decoder technology, H. 264 has become one of the most important standards. In this paper, we proposed novel moving object detection based on spatial-temporal method in compressed domain, which mainly depended on the motion vector field and efficiently segmented the object. Firstly, a spatially median filter was applied to the motion vector to alleviate the error vector due to difference between the motion vector obtained from motion estimation and the real motion information. Secondly, we built the spatial and temporal coherence model with the entropy based method. We obtain the threshold with these models. The effectiveness of our method was demonstrated with its performance on experiment results.

Keywords compressed domain, object detection, H. 264

1 引言

当前, H. 264 标准已经成为人们广泛接受的数字视频压缩标准之一。通过对码流中运动对象信息的分析, 准确有效地定位和检测出对象, 在视频检索、视频监控等领域有广泛的应用。

运动对象检测就是沿着时间轴把图像序列中运动物体的大致轮廓和位置检测出来, 以便为其他的

应用作后处理。检测信息一般包含两种: 视觉属性(如颜色^[1]、形状^[2]和纹理等)和运动信息(如运动矢量等)。目前大多数检测算法都是基于像素域这种算法精确度比较高, 但是其计算复杂度也高^[3]。此外, 现在大多数视频都是以压缩的形式存储的, 在压缩域下进行检测不但可以避免全解码和视频的再编码过程引起的使视频质量退化, 又可提高检测速度。

近年来, 已经开始研究 MPEG 压缩域的运动对

基金项目: 国家自然科学基金项目(60572127, 60602012); 上海市教委发展基金项目(05AZ43); 上海高校选拔培养优秀青年教师科研专项基金项目(2006); 上海市重点学科项目(T0102)

收稿日期: 2007-07-05; **改回日期:** 2007-07-25

第一作者简介: 徐剑峰(1982 -), 男, 上海大学信号与信息处理专业硕士研究生。主要研究方向为数字视频信号处理等。E-mail: xjfe_mail@yahoo.com.cn

象分割。例如 Babu 首先采用时域累积和空域内插的方法估计一个稠密运动矢量场,然后采用期望最大化(EM)算法进行聚类以获得不同的运动对象^[4]。文献[5]对帧间预测宏块采用聂曼-皮尔逊准则进行检测。H.264 域中能直接用来分割的信息只有运动矢量。Zeng 等人^[6]提出了一种基于块的 MRF(markov random field)模型从稀疏运动矢量场中分割运动对象的方法,根据各个块运动矢量的幅值赋予各个块不同类型的标记,通过最大化 MRF 的后验概率标记出属于运动对象的块,但这种方法只局限于静态背景的序列。Liu 等人^[7]提出了一种基于二叉树的运动对象分割方法,通过二叉树的遍历选择结点来表示合适的运动对象。

本文提出了一种新的在 H.264 压缩域下运动对象检测的方法,该方法利用运动矢量在空间和时间上的一致性建立熵模型,然后根据阈值检测出对象。

2 检测算法

本文提出的 H.264 压缩域内运动对象检测算法如图 1 所示。首先从 JM8.6 编码器中提取运动矢量场,然后对矢量场作预处理。在得到了可靠的运动矢量场之后,利用熵原理建立空间和时间一致性模型,最后用最大熵的方法自适应选择阈值,检测出运动对象。

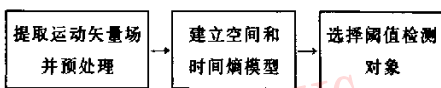


图 1 运动对象检测流程图

Fig. 1 Flowchart of the proposed algorithm

2.1 运动矢量预处理

采用中值滤波算法对运动矢量作预处理。一个运动对象内部的运动矢量应该一致。但是基于块匹配准则的,加之噪声等影响使运动估计得到的运动矢量并非都是如此。因这些噪声和非真实运动矢量的介入对视频对象检测结果有比较大的干扰。所以利用中值滤波在空域或时域上对运动矢量场作平滑预处理是必要的。

中值滤波是指一个窗口中心位置像素的值是这个窗口内所有像素的中间值,这里采用的是 3×3 窗口,定义如下:

$$A_i \in \text{Median}[A_j] \quad i \in (0,8)$$

A_i 表示 3×3 窗口的中间位置的像素值,Median 表示作中值滤波处理。

2.2 建立运动矢量熵模型

因为视频编码中帧内和帧间的运动矢量分别在空间和时间上存在一定的相关性,利用这些相关性,我们可以建立运动矢量熵模型来观察运动矢量是否具有^[8]一致性。

对于空间熵模型,如果 $N \times N$ 窗口范围内的运动矢量幅度或者相位没有太大变化,则说明该空间上的信息量比较少。信息量的大小反映了运动矢量变化是否具有^[8]一致性。

$$EP_{sc}(n) = \frac{HIST_{i,j}^*(n)}{\sum_{l=1}^m HIST_{i,j}^*(l)} \quad (1)$$

$$SC(i, j) = - \sum_{n=1}^m EP_{sc}(n) \log(EP_{sc}(n)) \quad (2)$$

其中, (i, j) 表示当前运动矢量的位置, SC 表示运动矢量的空间信息量, EP_{sc} 是直方图 $HIST$ 的相应概率分配函数,即直方图中每个样本在该直方图空间中的概率。 m 为直方图空间大小。 w 表示窗口尺寸。类似地,在时间域上也可以建立同样的模型。

$$EP_{tc}(n) = \frac{HIST_{i,j}^*(n)}{\sum_{l=1}^m HIST_{i,j}^*(l)} \quad (3)$$

$$TC(i, j) = - \sum_{n=1}^m EP_{tc}(n) \log(EP_{tc}(n)) \quad (4)$$

以上模型分别只从空间或者时间角度来考虑,具有一定局限性。为了使模型能同时反映时空一致性,本文在此基础上对模型进行了改进,将原空间 $N \times N$ 的窗口和时间 $1 \times 1 \times L$ 的窗口调整为 $N \times N \times L$,其中, L 表示连续的帧数,使得运动矢量的一致性表现得更加充分。改进后的效果如图 4 所示。

比较改进前后两种结果,可以看出改进后的图 2(e) 是人放好东西后弯腰起身的样子;图 2(f) 是人提着东西离开场景的动作。结合两种相关性后检测出的结果比较清晰,更加逼近实际运动。

2.3 自适应阈值选择

为了得到更准确的分割结果,采用基于最大熵原理自适应阈值选择方法^[9]。

在建立的熵模型基础上,对任意位置 (x, y) 的信息量用 $f(x, y)$ 表示。在一帧图像中,最小信息量为 $\min[f(x, y)]$,用信息等级 0 来表示,最大信息量为 $\max[f(x, y)]$,用信息等级 $l-1$ 表示。 $R = |0,$

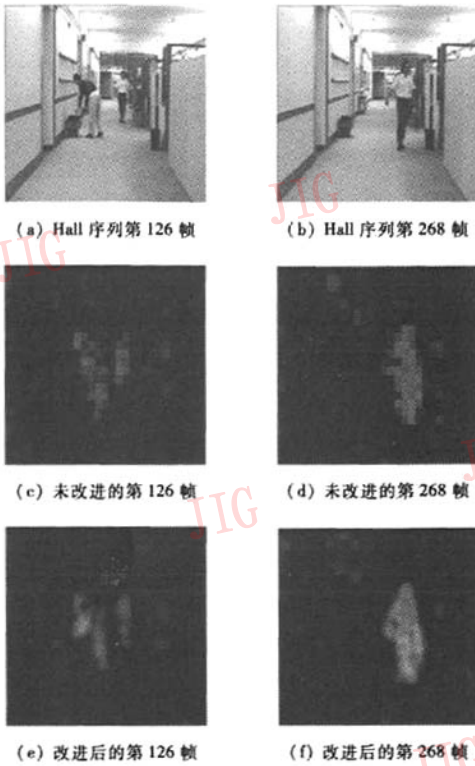


图 2 改进窗口后的一致性模型

Fig. 2 Coherence model of proposed window

$1, \dots, l-1$ 表示信息量等级的集合。定义 $N_i (i \in R)$ 为信息量等级 i 时的像素点数量。对于阈值 $t \in R$, 低于阈值的信息量表示如下:

$$E_B = - \sum_{i=0}^l \frac{N_i}{\sum_{m=0}^l N_m} \log \left(\frac{N_i}{\sum_{m=0}^l N_m} \right) \quad (5)$$

高于阈值的信息量如下:

$$E_A = - \sum_{j=l+1}^{l-1} \frac{N_j}{\sum_{n=l+1}^{l-1} N_n} \log \left(\frac{N_j}{\sum_{n=l+1}^{l-1} N_n} \right) \quad (6)$$

所以, 自适应的最佳阈值为

$$t = \arg \max_{\tau} (E_A + E_B) \quad (7)$$

由式(7)可以得到对象和背景分界点的阈值。一般对于静止背景的视频, 大于阈值的像素点认为是检测对象。

从上述的计算公式来看, 信息量的大小和概率分布函数有关。概率越大, 信息量反而越小。本算法除了考虑到一帧图像的空间信息外, 还考虑到了信息量的分配关系和特定阈值下信息量的大小, 而

一般的阈值选择算法都只考虑到了空间信息, 检测出的对象不够准确。

3 实验结果和分析

本文使用了 JM8.6 版本的 H. 264 编码器提取运动矢量场。采用 Baseline Profile, 每 30 帧插入 1 个 I 帧, 1 个参考帧, 运动估计的搜索范围为 $[-32, 32]$, 量化参数为 30。

图 3 是标准测试序列 Hall (CIF 格式, 352×288) 第 88 帧和第 155 帧的检测结果。两组结果分别表示的是 HALL 序列中单个运动对象进入场景和两个运动对象相遇的情况, 利用空间相位一致性建立的模型。图 3(c) 和图 3(d) 是在熵模型的基础上, 利用最大熵的方法获得最佳阈值后得到的结果。使用阈值最大的好处是去除了一些和运动对象不相关的点, 能比较准确地检测出运动对象。

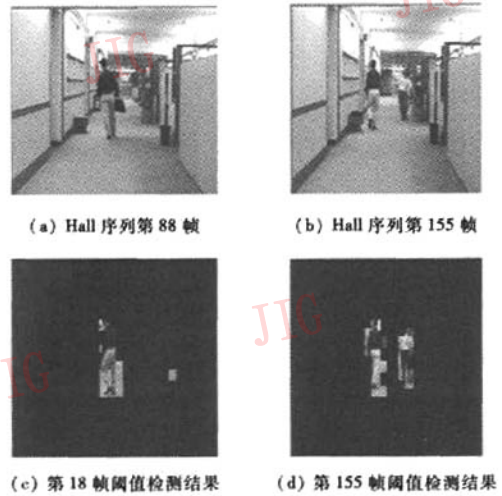


图 3 采用自适应阈值后 Hall 序列检测结果

Fig. 3 Adaptive threshold detective result of sequence Hall

图 4(a) 和 (b) 分别是头肩测试序列 Akiyo (CIF 格式, 352×288) 第 20 帧和第 200 帧, 而图 4(c) 和图 4(d) 分别是采用自适应阈值后对这两帧的检测结果。

图 4(c) 反映的是人脸部的运动, 同时肩部也有一些轻微的运动, 所以阈值检测结果基本上把整个人的轮廓都检测出来。而图 4(d) 主要表现脸部的运动, 肩部基本上不运动, 检测结果得到的只是人的脸部。



图 4 采用自适应阈值后 Akiyo 序列检测结果

Fig.4 Adaptive threshold detective result of sequence Akiyo

4 结 论

基于 H. 264 的运动矢量场,提出了一种在压缩域内的运动对象检测算法。先通过中值滤波算法对矢量场作预处理,然后利用运动矢量空间和时间的相关性建立一致性模型,最后采用基于最大熵原理的自适应阈值选择算法。经过实验证明,整个检测算法是有效的,能较清楚准确地检测出运动对象。

参考文献 (References)

- 1 Fieguth P, Terzopoulos D. Color-based tracking of heads and other mobile objects at video frame rates [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], San Juan, Puerto Rico, 1997; 21 ~ 27.
- 2 Faloutsos C, Equitz W, Flickner M, et al. Efficient and effective querying by image content [J]. Journal of Intelligent Information Systems, 1994, 17(1): 231 ~ 262.
- 3 Liu Long, Liu Guan-zhong, Liu Jie-yu, et al. New moving object segmentation algorithm in MPEG compressed domain [J]. Journal of Xi'an Jiaotong University, 2004, 38(12): 1264 ~ 1267. [刘龙,刘贵忠,刘洁瑜等.一种基于 MPEG 压缩域运动对象分割算法[J].西安交通大学学报, 2004, 38(12): 1264 ~ 1267.]
- 4 Venkatesh Babu R, Ramakrishnan K R, Srinivasan S H. Video object segmentation: A compressed domain approach [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(4): 462 ~ 474.
- 5 Chen Shan, Yu Hong-yang. Moving object detection from MPEG compressed field [J]. Signal Processing, 2004, 20(6): 628 ~ 631. [陈杉,于鸿洋.基于 MPEG 压缩域的运动对象检测方法[J].信号处理, 2004, 20(6): 628 ~ 631.]
- 6 Zeng W, Du J, Gao W, et al. Robust moving object segmentation on H. 264/AVC compressed video using the block-based MRF model [J]. Real-Time Imaging, 2005, 11(4): 290 ~ 299.
- 7 Liu Z, Zhang Z Y, Shen L Q. Moving object segmentation in the H. 264 compressed domain [J]. Optical Engineering, 2007, 46(1): 017003.
- 8 Ma Y F, Hua X S. A generic framework of user attention model and its application in video summarization [J]. IEEE Transactions on Multimedia, 2005, 7(5): 907 ~ 919.
- 9 Wong A K C, Sahoo P K A. A gray-level threshold selection method based on maximum entropy principle [J]. IEEE Transactions on Systems And Cybernetics, 1989, 19(4): 866 ~ 871.