

使用修改的豪氏道夫距离自动提取运动对象

史立 张兆扬

(上海大学通信与信息工程学院, 上海 200072)

摘要 新的视音频编码标准 MPEG-4 增加了支持基于内容的功能, 它把视频序列分割成语义意义上的视频对象 (VO) 视频对象在某一瞬时的“快照”称为视频对象平面 (VOP), 且一系列 VOP 表示一个运动对象, VOP 分割相当困难, 这主要是因为物理对象通常不以亮度、彩色或光流等低级特征来表达, 所以经典的分割方法无法获得有意义的分割结果。为了对这种视频运动图象进行有效的提取, 提出了一种基于修改的豪氏道夫对象跟踪器的自动 VOP 分割方法, 首先提取出初始模型, 然后用跟踪器在序列后继帧中跟踪此对象, 再对模型逐帧修改, 以适应对象在后继帧中形状的旋转和变化, 最后根据一系列二值模型来提取出视频对象, 此外, 为了提高分割效果和减少复杂性, 还使用了静态背景滤除技术来滤除静态背景, 实验结果表明, 该算法是有效的。

关键词 MPEG-4 对象跟踪 视频对象平面 豪氏道夫距离

中图法分类号: TN911.21 **文献标识码:** A **文章编号:** 1006-8961(2001)07-0689-05

Extraction of Video Object Plane Using Modified Hausdorff Object Tracker

SHI Li, ZHANG Zhao-yang

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

Abstract The new video&audio coding standard MPEG-4 is enabling content-based functionalities. It takes advantage of a decomposition of sequences into semantically meaningful video object (VO). A snapshot of a VO is named video object plane (VOP) at a given time and a series of VOPs represent one moving object. This is a very challenging task, because physical objects are normally not homogeneous with respect to low-level features such as color, luminance, or optical flow. Hence, conventional segmentation algorithms will fail to obtain meaningful partitions. In this paper, a new automatic VO segmentation algorithm based on modified Hausdorff object tracker is presented. A binary model for moving object is automatically derived and tracked in subsequent frames using the modified Hausdorff distance. First initial model is extracted and tracked based on the proposed object tracker against subsequent frames in the sequence. Then the model is updated every frame to accommodate for rotation and changes in shape. The video object is extracted by a series of binary models in the end. Furthermore, to improve the quality of segmentation and to reduce the computational complexity, stationary background is filtered by a novel technique. Experimental results demonstrate the performance of our proposed algorithm.

Keywords MPEG-4, Object tracking, Video object planes, Hausdorff distance

0 引言

MPEG-4 视频编码标准支持基于对象的内容

表达, 通过分别编码和解码物理对象, 提供一套使用户可交互式地操作视频对象的工具, 基于内容的可分层性允许编解码器在不同层次、以不同分辨率分配可用的比特流。

基金项目: 上海市自然科学基金(992D14035)和上海市教委重点项目(s990123)

收稿日期: 2000-05-15; 改回日期: 2001-04-09

图1是体现上述思想的编解码结构图^[1].先将视频序列的场景分割成语义意义上的各个视频对象(VO).VO在视频序列中某个时刻的帧图象称为视频对象平面(VOP),每个VOP所包含的形状、纹理信息和运动矢量可单独编码,用户可根据需要,在编

码端通过控制单元对各个VO分配不同的码率或通过复用单元来改变组合参数;也可在解码端,通过组合单元来改变VO的位置及组合顺序或通过解复用单元来请求解码所需的码流,从而充分实现了与用户的交互性.

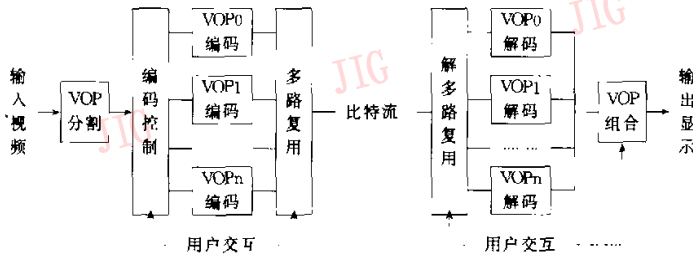


图1 MPEG-4 VM编解码结构

由上可见,要实现 MPEG-4 的交互功能,从输入视频序列中分割出 VO 是必须的,虽然在称为第 2 代图象编码技术中曾出现一系列基于诸如光流场、彩色、强度等低层次特征区域聚类的分割方法,但这对分割具有语义意义的对象是无效的.

因此,近两年出现了一些从背景中半自动或自动分割具有语义性的前景对象的探索方法^[2~4].尽管这些方法基本上可归结为基于变化检测模板(CDM)的 VOP 提取,但用这些方法来解决运动对象位置变化的显露问题仍需要复杂的运动估计.

由于本文所提出的算法是基于对象的跟踪技术,从而有效避免了许多与运动估计有关的问题,其算法的核心是以修改的基于豪氏道夫距离的对象跟踪器去匹配后继帧中对象的二维二值模型,其所求的最佳匹配不仅表达了对象发生的平移,而且通过二值模型的逐帧修改,还可以容纳对象的形变.最后该算法根据一系列二值模型自动地从序列中导出 VOPs.

1 基于修改的豪氏距离(MHD)的对象跟踪器

1.1 HD 和 MHD 对象跟踪匹配准则

豪氏道夫距离(HD)测度是计算从对象模型和测试图象中提取的两套边缘点集间的距离值.经典的 HD 算法^[5]基于衡量两套点集之间的相似度,对于大小分别为 N_o 和 N_l 的两个特征点集: $O = \{o_1, o_2, \dots, o_{N_o}\}$ 和 $I = \{i_1, i_2, \dots, i_{N_l}\}$, 定义豪氏道夫距离(HD)为

$$H(O, I) = \max\{h(O, I), h(I, O)\} \quad (1)$$

其中, $h(O, I)$ 和 $h(I, O)$ 是集合 O 和 I 之间的直接距离,分别定义为

$$h(O, I) = \max_{o \in O} \min_{i \in I} \|o - i\| \quad (2)$$

和

$$h(I, O) = \max_{i \in I} \min_{o \in O} \|i - o\| \quad (3)$$

上述两者之间的较大值即为 $H(O, I)$. 但在实际应用中,式(1)~式(3)的定义是有缺陷的.例如若模型 O 或图象 I 中有一特征点远离中心,即使二者匹配良好,其所得的 $H(O, I)$ 却很大,所以 Huttenlocher 等.提出了局部 HD 算法^[5,6]用以比较包含严重遮挡区域和退化的图象.局部 HD 定义如下

$$h_K(O, I) = K_{oc}^n d_l(o) \quad (4)$$

其中, $d_l(o) = \min_{i \in I} \|o - i\|$, 表示点 o 到点集 I 的距离极小值, K_{oc}^n 表示第 K 个排序的 $d_l(o)$ 值.另外,此种 HD 度量需要一个参数 $f = K/N_o$, 它的范围为 $0.0 \sim 1.0$. 根据 f 所表示的小数值,它的性能变化伸缩性很大,实验证明, f 通常选择 0.6 为宜.

通过修改排序阶统计特性, Azencott 等.提出了 CHD 度量准则用来比较二值图象^[7]. CHD 的直接距离定义为

$$h_{K,q}(O, I) = P_{oc}^n Q_{il}^q \|o - i\| \quad (5)$$

其中, P_{oc}^n 为 $Q_{il}^q \|o - i\|$ 的第 P 个排序值, 而 Q_{il}^q 表示欧氏距离集的第 Q 个排序值.另外, CHD 算法需要如下两个参数: $p = P/N_o$ 和 $q = Q/N_l$, 其中 N_s 为欧氏距离集的大小.该两个参数的范围从 $0.0 \sim 1.0$. 实验结果表明, p 取 0.8 和 q 取 0.9 时,可获得良好的匹配效果.

Dubuisson and Jain 提出了基于平均距离值的 MHD 算法^[8],而 MHD 直接距离定义为

$$h_{MHD}(O, I) = \frac{1}{N_o} \sum_{o \in O} d_I(o) \quad (6)$$

其中, $N_o = p$ 是集合 O 中的特征点数. 实验证明, 由于此区域匹配准则与局部 HD 和 CHD 算法相比, 对噪声不敏感, 且简单易实现, 因此在实际应用中最有效; 另外, 由于不需要手工提供参数, 从而提高了算法的自适应性. 本文的算法中, 就以这个匹配准则来跟踪后继帧中的运动对象.

1.2 豪氏道夫距离的实现

在确定了对象跟踪的匹配准则后, 余下的就是如何计算 HD. 为了能较早排除错误的匹配, 需要设置阈值 T , 若豪氏距离小于阈值 T , 就能得到匹配.

计算豪氏距离的步骤如下: 首先, 对边缘图象通过距离变换来得到最近边缘象素的距离; 然后对所有平移 $t = (t_x, t_y)$, 计算 $h_{M..}(O, I)$. $h_{M..}(O, I)$ 根据平移 t 进行计算, 也就是说, 对象根据矢量 t 平移, 通过某一模型点 o 位置上的距离变换来直接给出点 o 到最近边缘象素间的距离, 将所有这些距离累加再除以模型点数 N_o , 其值赋给 $h_{M..}(O, I)$, 以类似的方法计算出 $h_{M..}(I, O)$, 最终得到 $H_{M..}(O, I)$. 这一最小的豪氏道夫距离 $H_{M..}(O, I)$ 就表明模型平移了 t 的新位置.

2 视频对象平面分割算法

图 2 给出了一个从静态背景中自动提取运动对象的分割算法流程. 流程图主要由对象检测、二值模型初始化、模型修改和 VOP 提取等组成. 用 Canny 算子^[9]检测出边缘图象二值模型(见图 3)后, 再使用修改的豪氏道夫距离跟踪器在后继帧中自动匹配这个二值模型, 模型通过逐帧修改以容纳对象形状的旋转和变化, 再根据生成的二值模型序列, 来提取模型 VOP 序列所构成的视频对象(VO). 此外, 为了有效地跟踪匹配, 流程图中的滤波器还用来滤除凌乱静态的背景.

2.1 运动对象初始化

因为运动对象位置未知, 所以, 初始化时, 需由前后两帧之间的强度差来检测对象的变化区域. 将该差值阈值化, 其大于阈值的部分即为运动的对象或改变了形状的对象(见图 4(a)). 对其中超过一个象素宽度的遮挡区域, 可通过文献[10]给出的腐蚀

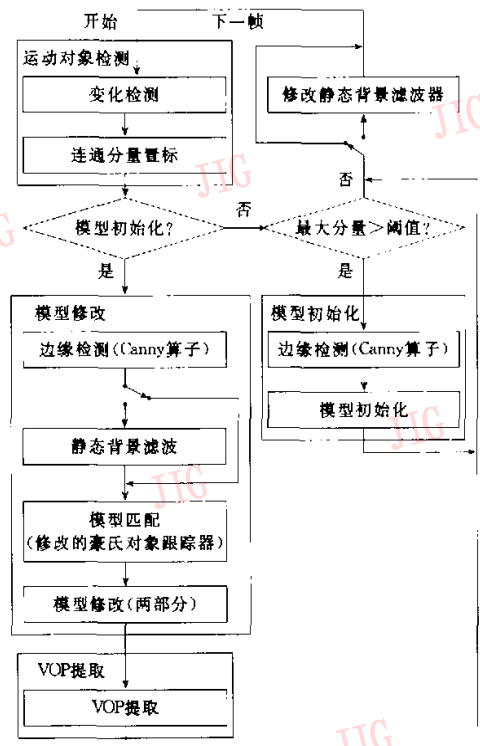


图2 VOP分割算法流程图

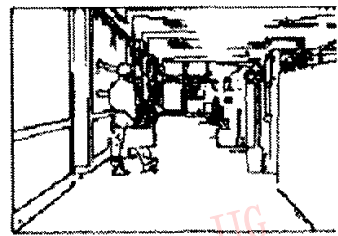


图 3 Hall monitor 序列第 40 帧使用 Canny 算子所获得的二值边缘图象

算法来细化(见图 4(b)). 根据连通分量置标^[10]在腐蚀的图象中寻找连通的分量(见图 4(c)), 其中, 大于一指定阈值的那些分量是属于运动对象, 这里把它称作运动连通分量(MCC), 通常它们远大于噪声分量.

最后通过选择 MCC 小距离(1~2 个象素)内的边缘图象所有象素, 即获得初始化二值模型.

2.2 对象跟踪

所谓对象跟踪就是用的修改的豪氏对象跟踪器在序列中跟踪二值模型.

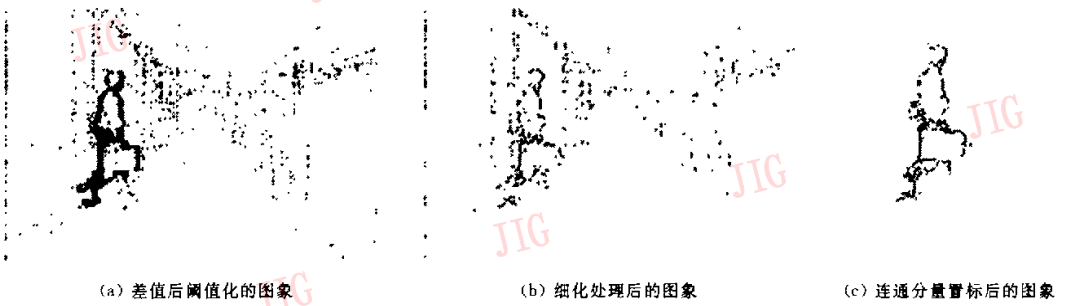


图4

图5是Hall Monitor序列中第40帧在使用修改的豪氏距离跟踪器跟踪后,所找到的匹配位置并叠加在第41帧时的结果.从图中可以看到,除了右脚外(右脚的运动和身体其他部位运动不同,属于非刚性运动),匹配是相当精确的.

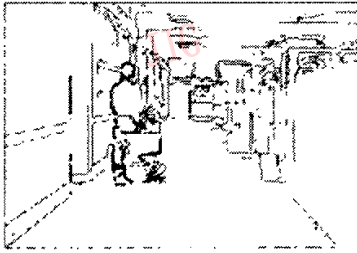


图5 跟踪匹配结果

2.3 模型修改

因为新位置上的模型形状发生了变化,所以必须逐帧修改模型以容纳对象的平移和形状变化.模型修改包括两个分量:一是缓慢变化部分,二是相对于运动对象整体运动的快速变化部分.

缓慢变化部分实际上是准刚性分量.对缓慢变化部分的模型修改就是将先前帧的模型根据最佳匹配平移到当前帧的新位置上,且假定靠近这个平移的旧模型的象素是属于对象的.这样,在平移的旧模型的指定距离 T_s (典型的为 $1\sim 3$ 个象素) 内所有的边缘象素均被分配给新模型.这种修改如图6(a)所示,除了右脚外,新模型包含旧模型平移后的边缘象素.

快速变化部分实际上是非刚性运动部分.其模型修改是通过帧间差或变化检测模板来实现.这样通过平移后的模型和当前帧的帧间差,就可单独检测到快速变化的部分,再通过连通分量置标,把这一部分叠加到平移后的旧模型上,就实现了整个模型的修改,如图6(c)所示.

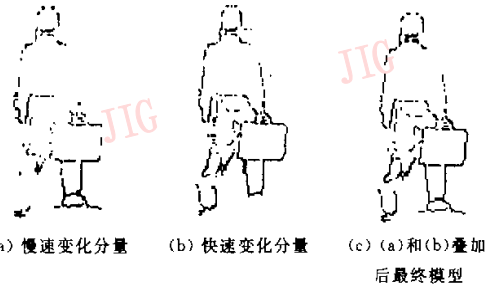


图6 模型修改

2.4 滤除静态凌乱的背景

在模型修改阶段,发现阈值越大,越有可能把整个对象包括进新模型中,但也增加了把背景包括进去的可能.为了避免这种情况,希望在进行模型修改前就滤除掉静态的背景,更重要的是,为了减少图象点数目和降低匹配时间,人们更希望使用过滤后的边缘图象来进行豪氏匹配.

采用象素“计数器”法来滤除静态背景.其思想是通过计算每个象素归类为边缘图象的频度来进行滤除.若计数器大于某个阈值,就认为它是属于背景部分,然后即将其从边缘图象中删除.由于计数器仅对没有被对象遮挡的象素进行修改,因此在所有的对象位置都知道时,可每处理一帧后,就进行计数器修改.这样,即可仅收集那些真正被标识为背景的象素的信息.通常,仅仅利用初始的几帧将无法获得可靠的数据.

2.5 VOP的提取

跟踪器的输出是一系列拟提取对象的二值边缘模型.将这些二值边缘模型“置于”对应的帧中,进行“切割”而得VOP.先扫描每行中开始和结束的模型点,再扩展到每列,最后便得到VOP和对应的视频对象(VO).

3 实验结果

使用 CIF 格式标准序列 Hall Monitor 来测试本文的算法。实现平台是操作系统 Windows 9x/NT 的 Visual C++ 6.0, 在奔腾 II 400 上进行。实验过程



(a) 原始第 45 帧



(b) 二值模型



(c) 提取的 VOP

图 7

4 结 语

本文提出了一种新的基于修改过的豪氏道夫距离的对象跟踪器提取 VOP 的方法。实验结果表明, 本文的算法可有效地提取处于静态背景的运动视频对象。

提取运动背景下的 VOP 其困难之处在于如何标识运动背景下的运动对象。这需要结合形态运动检测和图象理解技术来实现。此外, 本算法计算较复杂, 而更好的边缘检测和对象跟踪技术尚需进一步研究。

参 考 文 献

- 1 Ebrahimi T, Home C. MPEG-4 natural coding—an overview. *Signal Processing Image Communication*, 2000, 15(13): 365~385.
- 2 Mech R, Wollborn M. A noise robust method for segmentation of moving objects in video sequences. In: *IEEE Int. Conf. On Acoustics, Speech, and Signal Processing (ICASSP 7)*, Munich, Germany, April 1997, 4: 2657~2660.
- 3 Neir A, Colonnese S, Russo G *et al.* Automatic moving object and background separation. *Signal processing*, 1998, 65(2): 219~232.
- 4 Meier T, Ngan K N. Automatic segmentation of moving objects for video object plane generation (Invited paper). *IEEE Trans. on Circuits and Systems for Video Technology*, 1998, 8(5): 525~538.
- 5 Huttenlocher D P, Klanderman G A, Rucklidge. Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal.*

说明, 提取一个 VOP 的整个过程大概需要 5~6s, 其中边缘提取和匹配算法占去大部分时间。部分实验结果如图 3~图 6 所示, 其最终结果如图 7 所示。

图 7(a) 为 Hall Monitor 序列第 45 帧原始图象, 图 7(b) 为跟踪匹配和模型修改后所生成的二值模型, 图 7(c) 为最终提取的视频对象平面 VOP。

Machine intell., Sept. 1993, 15: 850~863.

- 6 Huttenlocher D P, Noh J J, Rucklidge W J. Tracking nonrigid objects in complex scenes. In: *Proceedings of the Fourth International Conference on Computer Vision*, 1993: 93~101.
- 7 Azencott R, Durbin F, Faumard J. Multiscale identification of building in compressed large aerial scenes. In: *Proc. 13th Int. Conf.*, Aug. 1996, 2: 974~978.
- 8 Dubuisson M P, Jain A K. A modified hausdorff distance for object matching. In: *Proc. 12th Int. Conf. Pattern Recognition*, Jerusalem, Israel, Oct. 1994: 568~568.
- 9 Canny J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 1986, PAMI-8: 679~698.
- 10 Rosenfeld A, Pfaltz J L. Sequential operations in digital picture processing. *J. ACM.*, Oct. 1966, 13: 471~494.



史立 1971 年生, 1998 年获安徽大学硕士学位, 现为上海大学通信学院通信与信息系统专业博士研究生。主要研究方向为图象通信与音频处理。已发表文章数十篇。



张兆扬 1938 年生, 教授, 博士生导师, 1962 年毕业于西安交通大学无线电工程系, 现为上海大学通信与信息学科带头人。长期从事图象处理与多媒体通信的研究。已完成多项国家自然科学基金及重大项目研究, 发表学术论文 80 余篇。