基于模板匹配的视频对象分割

宋立锋, 韦 岗, 王群生
(华南理工大学电子与信息学院, 广东广州 510641)

摘 要: 视频对象分割是 MPEG-4 标准关键技术. 本文结合模板匹配和基于运动估值和补偿的对象跟踪方法, 提出了一种可以从复杂场景中分割出 MPEG-4 视频对象的新方法. 在使用运动估值和补偿得到分割掩膜后, 以初始帧对象颜色为模板, 在当前帧的轮廓边界区域通过模板匹配检测对象, 使轮廓精确化. 本文方法在一定范围内有效解决了遮挡问题, 并能够以初始帧跟踪任意长序列中的对象.

关键词: 视频对象分割; 半自动视频对象分割; 对象跟踪; 模板匹配
中图分类号: TP391 文献标识码: A 文章编号: 0372-2112 (2002) 07-1075-04

Video Object Segmentation Based on Template Matching

SONG Li-feng, WEI Gang, WANG Qun-sheng
(*Department of Electronic and Communication Engineering, South China University of Technology, Guangzhou, Guangdong 510641, China*)

Abstract: Video object segmentation is the key issue in the MPEG-4 International Standard. This paper presents a new method for MPEG-4 video object segmentation from complicated scenes. In this method, the method of template matching and the object tracking method based on the motion estimation and compensation are combined. After the segmentation mask of the current frame has been generated by the motion estimation and compensation, the object colors in the initial frame are the same as the template and the object is detected in the object boundary area by template matching. So the object contour is refined. This object tracking method is able to solve the problems of occlusion and complicated scenes efficiently to certain extent and also tracks objects from an initial frame over any long time.

Key words: video object segmentation; semiautomatic video object segmentation; object tracking, template; matching

1 引言

多媒体数据压缩国际标准 MPEG-4 采用基于对象的编码方法, 要求在编码前把视频序列的场景分解为多个视频对象平面(VOP: Video Object Plane). 此过程即视频对象分割, 是 MPEG-4 标准关键技术. 然而因为成像过程中的信息丢失和噪声, 使输入图像数据不满足唯一正确解的充分条件(即病态问题), 同时人工智能技术的现状决定计算机不具有人的观察、识别、理解图像的能力, 至今还没有通用的有效方法去解决这个问题. 韩国 Electronics and Telecommunications Research Institute (ETRI) 的"用户辅助视频对象分割工具"[1] 是目前效果最好的分割工具. 它由用户通过图形用户界面生成初始帧分割结果, 再通过对象跟踪自动生成后续帧分割结果. 其处理非刚性物体的对象跟踪方法以轮廓线为对象特征, 以运动估值和补偿结合边缘强度寻优完成对象特征定位. ETRI 方法在实际效果上明显优于仅能检测运动区域且检测结果并不可靠的全自动分割[2~4], 也优于其它半自动分割[5~10]: 文[5] 中方法以区域为对象特征, 以运动估值和补偿与迭代类聚算法进行对象特征定位; 文[6] 中方法以 2D 网格为对象特征, 以 2D 网格配对来完成对象特征定位; 文[7] 中方法以轮廓线为对象特征, 以运动估值和补偿加上形态学水线算法执行对象特征定位; 文[8~10] 中方法建立主动轮廓模型, 以闭合曲线[8,9] 甚至是参数表示的闭合曲线[10] 作对象轮廓线, 用能量最小实现对象特征定位.

本文以甚低码率应用中最常见的头肩像为研究对象. 用 ETRI 最后版本 1.2 演示程序[11] 处理头肩像视频测试序列. 发现用其处理较简单序列 Claire、Alexis、Mother & Daughter、Miss America、Akiyo 等, 结果令人满意, 达到实用要求; 而处理复杂序列 Carphone 和 Foreman, 结果误差很大以致于失去对象含义. 这是因为 ETRI 方法没有利用特定对象知识, 即"没有把图像信息与对象紧密地结合起来"[9], 无法从背景边缘中辨别出对象轮廓, 显得"智力"不足. 它是一个开环过程, 存在误差向后逐帧传递、扩大的问题. 此外, 为了分割复杂序列, 必须处理遮挡问题.

本文提出一种可以从复杂场景中分割出视频对象的对象跟踪方法, 把 ETRI 处理非刚性物体的对象跟踪方法和模板匹配结合起来, 以初始帧对象颜色为模板, 首先通过 ETRI 方法得到当前帧的分割掩膜(Segmentation Mask), 然后在当前帧的对象边界区域通过逐像素模板匹配检测对象, 使对象轮廓精确化. 为了提高计算机工作效率和对复杂情况的适应力, 在模板匹配中引入与初始帧的对象颜色匹配和背景颜色匹配, 以及与前一帧的局部背景颜色匹配. 根据这些颜色匹配的排序结果检测对象. 在对象颜色特征与初始帧模板相似的前提下, 本文方法能够有效地解决遮挡和复杂场景等问题, 以初始帧完成任意长序列的对象跟踪. 处理包括 Carphone 和 Foreman 序列在内的多个头肩像视频测试序列得到满意结果.

2 采用模板匹配的轮廓精确化

2.1 模板匹配和 ETRI 方法结合

模板匹配是常用的图形检测方法. 如果以初始帧的对象

收稿日期: 2000-11-27; 修回日期: 2001-11-05
基金项目: 国家自然科学基金; 教育部回国人员科研基金

图像为模板,直接在当前帧做模板匹配,可以检测出对象,但是达不到较高精度,也处理不了复杂图像. 本文把模板匹配和 *ETRI* 方法结合起来, 构成图 1 的方框图. 图中虚线方框是 *ETRI* 处理非刚性物体的对象跟踪方法. 模板匹配在这里的作用是: 利用已知的特定对象知识检测图像中的对象, 从而使分割结果精确化; 形成一个闭环过程, 使后续帧的分割结果受到约束; 能够有条件地解决遮挡问题, 遇到遮挡时可以根据对象在图像特征上的相似性推导出图像内容. 遮挡是“图像中某个对象表面的覆盖/显露问题, 是由于仅占有部分观察场的对象的三维旋转和平移所引起的”[12], 是图像分析难以解决的问题. 目前类似于视频编码中的双向运动补偿方法可以部分解决遮挡问题,即从前面和后面的参考帧求当前帧, 如文献[6]. 而本文应用模板匹配来处理遮挡问题, 这种方法与人认识事物的一般方法类似.

初始帧的对象分割对本文方法至关重要. 作为人关于任意对象的主观定义, 初始帧的分割直接影响整个序列的分割精度. 本文以初始帧对象颜色为主要模板, 进行逐像素颜色矢量值匹配.

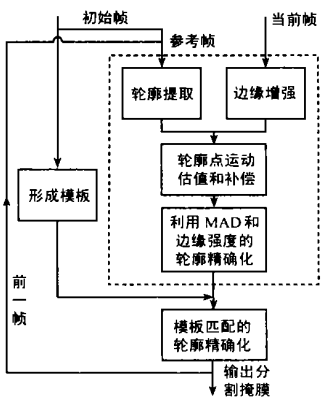

图 1 应用模板匹配的对象跟踪

2.2 模板匹配

在用 ETRI 方法得到初步分割结果后, 以模板匹配作进一步确认. 对于某个需要确认的像素, 计算该像素颜色与初始帧对象颜色以及背景颜色的差距, 由此比较其与对象以及背景的相似度. 以最小颜色距离确定该像素的属性. 这样操作的运算量巨大. 为了减少运算量, 本文采取以下措施:

(1) 减少处理像素数目. 只处理以 ETRI 方法分割出的对象轮廓线为轴线的带状区域内的像素, 即轮廓边界像素, 如图 2 所示. 本文实验结果显示对于 QCIF 格式*(*176× 144*)* 的视频序列, 这样的带状边界区域的宽度设为 5 像素时效果最好.

(2) 无位置信息的颜色匹配. 为此按像素矢量值建立对象颜色模板和背景颜色模板. 这样不必逐点匹配.

(3) 在第 2 点基础上, 通过初始帧图像平面的空间重取样和像素值重量化进一步减少模板尺寸. 对于具有部分平坦区域的图像, 第 2、3 点措施可以显著降低运算量, 也减少了内存占用量.

设以初始帧数据形成含有 n_{obj} 个颜色矢量 $\boldsymbol{T}_{obj}$*(* 本文处理的彩色图像以 YUV 表示, 像素颜色值为 YUV 空间的矢量值*)* 的对象颜色模板和含有 n_{back} 个颜色矢量 $\boldsymbol{T}_{back}$ 的背景颜色模板, 并存贮到内存中. 在当前帧带状边界区域内的某个位置 (x, y) 上, 像素颜色矢量值为 $\boldsymbol{I}(x, y)$. 采用模板匹配检测对象如下:

$$D_{obj}(x, y) = \min_{n} \| \boldsymbol{I}(x, y) - \boldsymbol{T}_{obj}(n) \|^2 \quad 0 \leqslant n \leqslant n_{obj} - 1 \qquad (1)$$

$$D_{back}(x, y) = \min_{n} \| \boldsymbol{I}(x, y) - \boldsymbol{T}_{back}(n) \|^2 \quad 0 \leqslant n \leqslant n_{back} - 1 \qquad (2)$$

式中颜色匹配结果 $D_{obj}(x, y)$ 和 $D_{back}(x, y)$ 分别是像素的颜色矢量与对象模板和背景模板之间的最小距离*(*本文采用 YUV 空间的欧氏距离*)*. 如果 $D_{obj} < D_{back}$, 像素(x, y) 与对象最相似, 确定为与对象匹配的对象像素; 否则, 如果 $D_{obj} < D_{back}$, 像素(x, y) 与背景最相似, 确定为与背景匹配的非对象像素*(*背景像素*)*. 相对于后面内容, 这样的匹配称为初始帧全局颜色匹配. 其突出优点是遮挡时可以准确分割.

但是上面的颜色匹配还不足以适应复杂情况. 首先, 在多数场景中, 部分对象颜色与背景颜色相同. 如果这个情况发生在轮廓边界区, 使轮廓不分明, 本文的对象模板匹配和轮廓精确化过程不能可靠地给出正确结果, 只能依靠对象跟踪的其它步骤. 真正必须考虑而且出现较多的情况是: 虽然在初始帧的整个图像平面上部分对象颜色与背景颜色相同, 但是在当前帧轮廓却是分明, 如果这样的轮廓像素与对象颜色模板的距离和背景颜色模板的距离都很小, 就会因为噪声得出错误结果, 如 Carphone 序列人物左肩部分.

为此定义帧间的局部颜色匹配为当前帧轮廓点$(x1, y1)$的邻域与参考帧对应的轮廓点 $x1', y1'$ 的邻域之间的颜色匹配. 相对同一个坐标系, $(x1, y1)$ 与$(x1', y1')$ 的对应关系已经由 ETRI 方法的对象轮廓点运动估值和补偿确定. 也可以在参考帧以$(x1', y1')$为中心, 在逐像素扩大的矩形区域内寻找对象轮廓点, 并把找到的离$(x1', y1')$ 点距离较近的点作为$(x1', y1')$. 对于 QCIF 格式的视频序列, 本文实验设定$(x1', y1')$的邻域是以$(x1', y1')$ 为中心的 33 × 33 的矩形区域, 而$(x1, y1)$的邻域是以$(x1, y1)$为圆心, 半径为 2 的圆区域. 注意所有轮廓像素的这样的邻域形成了图 3 中的带状边界区域. 对于$(x1, y1)$的圆形邻域内一点(x, y), 局部颜色的模板匹配为:

$$LD_{back}(x, y) = \min_{i,j} \| \boldsymbol{I}(x, y) - \boldsymbol{I}'_{back}(x1' + i, y1' + j) \|^2 \quad -16 \leqslant i, j \leqslant 16 \qquad (3)$$

式中局部颜色匹配结果 $LD_{back}(x, y)$ 是像素颜色矢量 $\boldsymbol{I}(x, y)$ 和$(x1', y1')$ 的矩形邻域内背景像素颜色矢量 $\boldsymbol{I}'_{back}(x1' + i, y1' + j)$ 之间的最小距离. 这样定义的局部颜色匹配是与像素位置有关的逐点匹配. 如果参考帧分割结果正确, 通过比较像素颜色与对象颜色模板以及参考帧背景颜色的相似程度就能得到正确的轮廓精确化结果. 特别是在类似 Carphone 序列中的人物左肩部分的情况下, 帧间局部颜色匹配的结果比初始帧全局颜色匹配的结果更准确. 但是帧间局部颜色匹配结果并不可靠. 在参考帧分割结果不正确或对象显露的情况下, 局部匹配不能

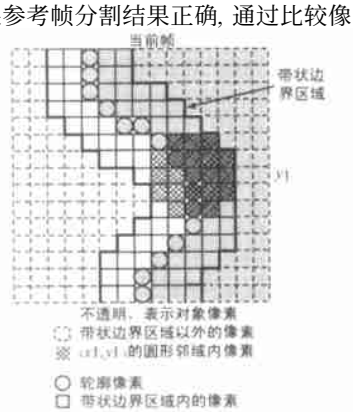

图 2 带状边界区域内的轮廓精确化

得到正确分割结果.

2.3 对象轮廓精确化

用 ETRI 方法得到初步的分割掩膜, 从中提取轮廓点. 以每个轮廓点(如图 2 中的$(x1, y1)$)为中心形成一个半径为 2 的圆形邻域. 区域内像素为轮廓边界像素. 所有这样的像素形成一个以轮廓线为轴线的带状区域. 对每个这样的轮廓边界像素执行上述颜色匹配. 在得到每个轮廓边界像素的全局对象颜色匹配结果 D_{obj}、全局背景颜色匹配结果 D_{back}和帧间局部背景匹配结果 LD_{back}后, 根据排序情况确定像素属性是对象还是背景, 如表 1 所示.

表 **1**　像素标记判决表

情况	升序 1	升序 2	升序 3	匹配情况	像素标记
1	D_{obj}	LD_{back}	D_{back}	确定的对象	对象
2	D_{obj}	D_{back}	LD_{back}	确定的对象	对象
3	D_{back}	LD_{back}	D_{obj}	确定的背景	背景
4	LD_{back}	D_{back}	D_{obj}	确定的背景	背景
5	LD_{back}	D_{obj}	D_{back}	不确定的对象	对象
6	D_{back}	D_{obj}	LD_{back}	不确定的背景	如果$\lvert LD_{back}-D_{obj}\rvert > D_{obj}$同时 $D_{obj}<4$, 认作对象; 否则, 认作背景.

由于局部匹配结果是不可靠的, D_{obj}和 D_{back}起主要作用, LD_{back}起次要作用. 对于表 1 中的第 1、2、3、4 这四种情况, LD_{back}使比较 D_{obj}和 D_{back}的结果更加可靠; 而第 5 和 6 这两种情况是不确定情况. 对于情况 6, 考虑到上述 Carphone 序列人物左肩部分区域的情况, 在本文程序中设定如果$\lvert LD_{back}-D_{obj}\rvert > D_{obj}$同时 $D_{obj}<4$, 认作对象; 否则, 认作背景. 对于情况 5, 因为没有实例去拒绝较可靠的全局匹配结果, 本文暂且将像素归为对象.

对象边界是噪声较大的区域, 为使分割结果正确, 本文采取以下措施:

(1) 以轮廓线为轴线、宽为 3 的带状区域为不确定区域(此不确定区域小于上面的轮廓精确化区域). 只有不确定区域以外的像素值可以作为模板颜色或作为帧间局部颜色匹配的背景颜色.

(2) 在确定了当前帧对象边界像素属性后, 考虑邻域同一性约束: 仅当某个轮廓点邻域像素全为对象属性时, 才把该邻域边界点中的原背景像素改为对象像素; 或者当该邻域像素全为背景属性时, 才把该邻域的原对象像素改为背景像素; 否则不改变像素原属性.

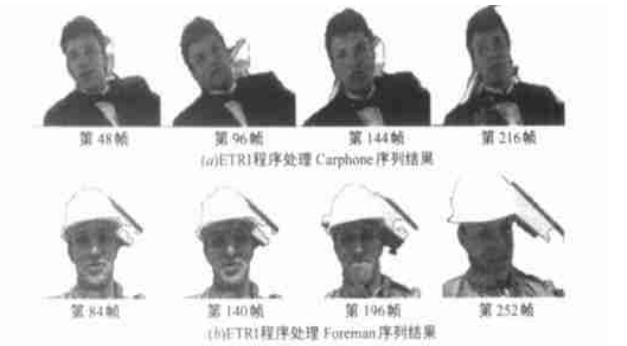

图 3　ETRI 程序分割结果

3　实验结果和结论

实验数据为 QCIF(176× 144, Y: U: V 4: 1: 1)格式的彩色视频测试序列. 用 Adobe Photoshop 5.0 工具对每个序列的第 1 帧图像勾画出对象轮廓线, 并以此生成初始帧分割掩膜. 读入每帧图像后即转化为 Y: U: V 4: 4: 4 格式. 用 Microsoft Visual C++ 6.0 工具把图 1 的方框图编成 C 语言程序. 为了减少噪声, 对每帧输入图像数据进行中值滤波, 对每帧输出分割掩膜进行先开后闭的形态学滤波. 用 Pentium Ⅱ 400MHz CPU、64MB 内存的微机运行此程序处理视频序列, 有关实验数据如表 2 所示. 虽然表 2 的运行时间数据仅供参考, 但是也反映当图像复杂、颜色较多时, 本文的模板匹配过程会消耗较多的机时.

目前对图像分割的评价以主观评价为主. 为方便观察, 把每帧的分割结果与其原始图像结合起来, 或形成轮廓线图像, 或去除背景形成前景对象图像. 在本文的实验中还用 ETRI 的版本为 1.2 的演示程序[11]处理相同数据以便比较, 在图 3、4、5 列出部分图像. ETRI 的演示程序的所有的输出功能都被取消了, 图 3 中的 ETRI 图像是经过截图等一系列处理后得到的结果. 图 4、5 是本文程序处理这两个序列的结果, 包括初始帧、末尾帧以及在序列中以等距离抽取的帧的分割结果, 这样的分布有助于了解整个序列的情况. 在图 4、5 中各有四帧图像与图 3 的图像相同, 以供对照. 这里没有列出其它序列的分割图像, 有 Claire、Alexis、Mother & Daughter、Miss America、Akiyo 等. 因为 ETRI 方法处理这些简单序列的误差不明显.

Carphone 序列背景复杂, 轮廓不分明, 有大量的背景颜色和对象颜色重叠, 被头部覆盖的车窗会显露出来; Foreman 序列包含复杂的头部局部运动和镜头变动所引起的全局运动, 人物肩部和面部的遮挡情况严重. 结果显示本文采用模板匹配后明显提高了分割精度, 在序列各帧图像的人物颜色与初始帧相似的情况下能够适应遮挡、复杂背景和复杂运动. 另

表 **2**　实验数据

序列	Claire	Alexis	Miss America	Mother & Daughter	Akiyo	Carphone	Foreman
序列长度(帧)	158	110	150	100	50	250	前 290 帧
初始帧对象大小(像素)	7819	6319	12628	11537	9513	10025	8995
对象颜色模板大小	1343	951	2296	2202	1346	1615	1531
背景颜色模板大小	730	1416	782	1140	1687	2274	2226
两个模板的相同颜色数	0	473	327	7	18	489	41
本文程序运行时间(秒)	221	173	253	205	86	519	573
每帧平均时间(秒)	1.41	1.59	1.70	1.88	1.76	2.08	1.98

图 4 本文方法处理 Carphone 序列结果

图 5 本文方法处理 Foreman 序列结果

外, 本文方法的结果没有出现分割误差向后逐帧传递的现象, 因为本文的模板匹配用存贮在内存的初始帧数据约束后续帧的分割, 在对象跟踪过程中构成一个闭环.

需要指出的是, 视频对象分割和其它图像分析问题一样是病态问题. 本文提出的方法仍然没有使计算机达到高"智力", 因为仅仅从初始帧获得特定对象知识是显然不够的. 另外, ETRI 的"用户辅助视频对象分割工具"包括用于非刚性物体的对象跟踪方法和用于刚性物体的对象跟踪方法这两部分. 本文的用于分割视频对象的对象跟踪方法是以 ETRI 用于非刚性物体的对象跟踪方法为基础设计的, 不能用于处理刚性物体, 如车、船等. 本文的目标是生成足够精度的分割结果进行 MPEG-4 标准基于对象编码, 而非解决视频对象分割问题. 本文方法实现了这个目标.

参考文献:

[1] ISO/IEC 14496-2. Information Technology-Coding of Audio-Visual Objects-Part 2 [S].

[2] Mesh R, Wollborn M. A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera [J]. Signal Processing, 1998, 66(2): 203- 217.

[3] Neri A, et al. Automatic moving object and background separation [J]. Signal Processing, 1998, 66(2): 219- 232.

[4] Kim M, et al. A VOP generation tool: Automatic segmentation of moving objects in image sequences based on spatio-temporal information [J]. IEEE Trans on CSVT, 1998, 9(8): 1216- 1226

[5] Zhong D, Chang S F. An integrated approach for content-based video object segmentation and retrieval [J]. IEEE Trans on CSVT, 1999, 9(8): 1259- 1268.

[6] C Toklu, et al. Semiautomatic video object segmentation in the presence of occlusion [J]. IEEE Trans on CSVT, 2000, 10(4): 624- 629.

[7] Gu C, Lee M C. Semiautomatic segmentation and tracking of semantic video objects [J]. IEEE Trans on CSVT, 1998, 8(5): 572- 584.

[8] Leymarie F, Levine M D. Tracking deformable objects in the plane using an active contour model [J]. IEEE Trans on PAMI, 1993, 15(6): 617- 634.

[9] Jang D S, Choi H I. Active models for tracking moving objects [J]. Pattern Recognition, 2000, 33(7): 1135- 1146.

[10] 赵雪春, 戚飞虎. 用可变形模板进行基于内容的图像分割算法 [J]. 电子学报, 2000, 28(4): 69- 72.

[11] ISO/IEC JTC 1/SC 29/WG11, M4479. Electronics and Telecommunications Research Institute (ETRI) [S].

[12] Tekalp A M. Digital Video Processing [M]. 北京: 清华大学出版社, New York: Prentice Hall, 1998.

作者简介:

宋立锋 男, 1967 年出生于陕西西安, 1989 年和 1992 年分别获得华南理工大学工学学士、工学硕士学位. 现为华南理工大学通信与信息系统专业博士生, 研究方向为图像和视频信号处理、多媒体数据压缩编码以及通信等.