

运动视频对象分割的一种快速算法

贾振堂^{1),2)} 贺贵明¹⁾ 韩艳芳²⁾

¹⁾(武汉大学软件工程国家重点实验室, 武汉 430072)

²⁾(河南城建高等专科学校计算机工程系, 平顶山 467001)

摘要 为了能够实时地对运动视频对象进行分割,提出了一种对视频序列图象中的运动对象进行快速分割提取的算法.该算法首先对图象进行滤波,并求出连续两帧图象之间的差分,然后应用“同化填充”技术和基于对象的“整体运动估计”来对差分图象进行修正,进而得到对象模板,同时利用模板缓冲区的帧间迭代来维持模板的完整性.该算法不仅不依赖于固定背景,而且能够消除差分图中的显露背景,还能得到运动目标较为精确的形状,并且算法简单、快速,鲁棒性好.

关键词 视频分割 差分图象 显露背景 同化填充 整体运动估计

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2002)11-1123-05

A Fast Algorithm for Moving Video Object Segmentation

JIA Zhen-tang^{1),2)}, HE Gui-Ming¹⁾, HAN Yan-Fang²⁾

¹⁾(The National Key Lab of Software Engineering, Wuhan University, Wuhan 430072)

²⁾(The Computer Department of The College of Henan Urban Construction, Pindinshan 467001)

Abstract A fast moving object segment algorithm on moving video sequences is proposed. Firstly, the difference image is produced between two continuous frames which are filtered in advance by Gauss template. Then the author proposed “assimilation” filling technique and a template modification method employing dominant motion vector are applied to remove the uncovered background and get the final object template. The “assimilation” filling technique only fills the abruptly changed gap on the edge or the inner hole, while leaves the smoothing changed edge untouched. What need to do in the filling process is just to increase or decrease the coordinate, not involving multiplication and division operation, so it is very fast. A memory buffer of objects is maintained through the whole process to remedy the disfigurement of regions gotten from a single frame. The algorithm doesn't rely on a fixed background, and can remove uncovered background to get the accurate size of objects. The algorithm only uses two continuous frames to segment the moving object from background, not like the symmetrical-difference based methods which make use of three frames. It is a simple, but fast and robust algorithm.

Keywords Video segmentation, Difference image, Uncovered background, Assimilation filling, Entire motion estimation

0 引言

MPEG-4 的框架提出以后,人们加紧了基于对象的视频技术研究,其中视频对象的提取是最为重要的一步,因为它是基于对象的视频检索、面向对象的视频压缩和编辑等应用的基础.

视频对象的分割一般从帧内和帧间两个方面来考虑,其中,采用差分图象来检测运动目标是利用帧间信息的最直接方法,如王栓等采用当前图象与固定背景之间的差分来检测运动目标^[1],但它只适用于背景已知的情况. Dong 采用的也是求图象差分的方法^[2],但由于需要对差分图象求梯度,因此产生双线,这无疑会使模板膨胀产生失真,并且该方法没有考虑

基金项目:武汉大学软件工程国家重点实验室建议基金项目(190118)

收稿日期:2001-10-10;改回日期:2002-03-25

“显露背景”的消除,这是一个不容忽视的问题.文献[3]同样利用连续3帧图象,并采用所谓“对称差分”方法,即把相邻的两个差分图相与,但由于该方法需要连续利用3帧图象,因此有一帧时间的滞后.另外,当某一帧图象不运动时(或运动无法检测时),“对称差分”则会使相邻的两帧图象都没有输出.

本文虽同样采用图象差分的方法,但由于只利用相邻两帧图象之间的差分,因此不依赖于固定背景,并且能够消除差分图中的显露背景,从而不仅能得到精确的物体大小,而且算法简单,鲁棒性好.

1 分割算法

1.1 图象差分

从光流约束方程可以看到,当图象存在空间梯度时,由于物体的运动可以从图象序列的时间梯度(帧间差)上体现出来,因此,对具有一定时间间隔的两帧图象进行差分,可以根据差分结果快速确定目标的运动范围.

图象差分不仅可以作用于当前图象与固定背景,也可作用于连续的两帧图象.由于考虑到运算的实时性要求,不能允许过多的帧缓冲延迟,因此这里采用当前图象 f_i 和前一帧图象 f_{i-1} 来求差分图 d_i .

$$d_i(p) = W * f_i(p) - W * f_{i-1}(p)$$

并对 d_i 取阈值,得到二值图象 D_i

$$D_i(p) = \begin{cases} 255 & |d_i(p)| \geq T \\ 0 & \text{其他} \end{cases}$$

其中, $p = (x, y)$ 是图象中的任一点, W 是一个用来滤波的窗口函数, $*$ 表示卷积运算, T 是一个阈值.

由于原始图象中往往存在一定的噪声,而直接求差分则受噪声影响比较大,因此,必须在差分运算之前,先对图象进行滤波处理.从测试的结果来看,高斯滤波的效果优于中值滤波.

阈值 T 用于滤除摄像机所带来的系统噪声(包括信号数字化过程所产生的噪声、光照扰动等)的影响,以保证静止的背景在差分图象中没有输出.该阈值主要由摄像机性能和所采用的预滤波方法所决定.滤波效果好的, T 可以取小一点,否则就取大一点.对于固定的滤波方法, T 的值主要取决于摄像机的性能,而且是相对固定的值,而与图象内容关系不大(仅有个别摄像机的信噪比会随图象亮度的变化而略有波动),可以预先测出.

对于所使用的摄像机(用于可视电话的简易摄

像头),在使用之前,需对由它所摄入的完全静止的视频画面求差分,通过直方图可以明显看出,由于其差分绝对值完全分布在 $0 \sim 6$ 之间,因此本文给出的测试结果中, T 取 7.如果考虑到人物移动时所造成的局部光照扰动以及阴影的影响, T 可以取得再大一些,比如 10,但也不能过大,否则就会滤掉真正的运动部分.在计算差分的过程中,同时可以方便地得到包含差分图的最小矩形框,这个矩形框用“ box ”表示(下面要用到).

对于多目标的情况,可通过模糊聚类技术来对多目标进行划分^[4,5],然后分别处理,而这里仅讨论单目标的情况.

1.2 模板的“同化”填充

以上步骤得到的二值图象可以作为对象的原始模板.由于物体内部通常是平坦的区域,其不仅空间梯度不明显,而且造成时间梯度(即帧间差)也不明显,因此模板中往往存在一定的空隙.这里利用本文提出的“同化”算法来进行填充,即在模板所处的矩形框 box 内,扫描所有的 0 值点.对于任一 0 值点 p ,考虑通过该点的任一条直线 L ,如果 L 上存在两个点 p_1 和 p_2 ,且满足

- (1) p_1, p_2 点的值都是 255
- (2) p_1, p_2 分布在 p 的两侧
- (3) $\|p - p_1\| < R, \|p - p_2\| < R$

则把 p 点置为 255,否则仍保持为 0.如果 p 点被置为 255,则认为“点 p 依据直线 L ,在 R 内被同化”.其中, R 是一个距离阈值,这里不妨称为“同化半径”.实际编程时,为了减小计算量,可不必计算所有的方向,只需考虑如图 1 所示的 4 条直线,即

$$L_0: (x-1, y) - (x+1, y),$$

$$L_1: (x-1, y+1) - (x+1, y-1),$$

$$L_2: (x, y-1) - (x, y+1),$$

$$L_3: (x-1, y-1) - (x+1, y+1).$$

由于搜索过程只是在直线上按照一定的步长增加或减小坐标,而不涉及乘除运算,因此速度比较快.

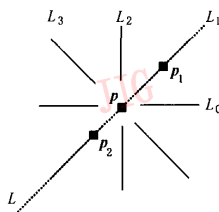


图1 “同化”示意图

这种同化填充算法的优点是:只填充变化陡峭的内凹空隙(缺口)以及内部空隙,而对变化缓慢的边缘部分则基本不做修改(比如脖子部位).这符合一般的实际情况.

然而,有些差分图象中会出现较大的空洞,若 R 值取得过小,则无法完全填充,如图 2(a)所示;若 R 值取得过大,则会使对象边缘的平缓转弯处也被填

充(这是不应该的).为此,本文进一步采用区域反填充技术来提取差分模板,这时仍然只考虑 box 内部区域.提取模板时,首先用灰色(0~255 之间的某一值,比如 200)从 box 内壁填充所有 0 值区域(这是简单的区域填充问题),即得到如图 2(b)所示的结果,然后取灰色区域的补集,即得到对象的差分模板,如图 2(c)所示.

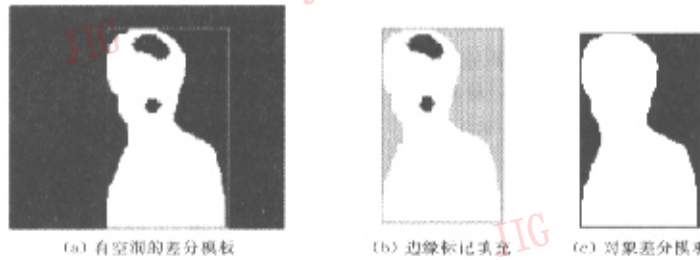


图 2 用反填充技术解决内部空洞

由于采用区域反填充技术,同化填充只需保证对象模板具有闭合的边缘即可.实际上,因为对象边缘通常是空间梯度较明显的部分,所以在差分图象中,边缘部分一般不会有较大的缺口(只要物体存在运动),另外,由于大的空洞通常出现在对象内部的平坦区域,因此同化半径取 3~5 即可.但有时,即使是边缘,缺口也可能比较大(物体与背景的对比不太明显),为此可以连续实施两次同化填充,这样一般都能得较好的结果.

1.3 消除显露的背景

在差分模板中,如果目标区域和显露背景同时存在,则必须去除显露背景部分,才能得到实际的目标模板.假如对差分图象做如下处理,则可以直观地看到模板中的显露背景部分(图 3):

$$D_i(x, y) = \begin{cases} 255 & |d_i(x, y)| > T \\ 0 & |d_i(x, y)| < -T \\ 127 & \text{其他} \end{cases}$$



图 3 包含显露的背景

图 3 右边的白色部分就对应显露的背景.据此容易想到,虽然去除白色边缘部分就可得到运动目

标,但是,实际处理时,左边的黑色部分也可能对应显露背景,这要看目标和背景的相对亮度以及目标的运动方向来确定.由此可见,去除背景并不是一件容易的事情.本文采用运动补偿的方法来处理这个问题,即先求出运动矢量,再根据运动矢量的方向和大小来消除显露的背景部分.

(1) 求对象运动矢量

一般采用划分宏块的方法来估算运动矢量.由于宏块的运动矢量一般比较杂乱,且每一个矢量并不一定真正代表实际的物体运动,所以单个地使用运动矢量得不到理想的结果,此时需要求得平均运动矢量(或做其他较为复杂的处理),来近似对象的整体运动矢量 V_0 .

本文是采用另一种方法来直接求目标的整体运动矢量.该方法求取目标的整体运动矢量是以整个运动物体为单位来进行匹配搜索的,即

$$V_0 = \arg \{ \min_v \text{Sad}(f_i, f_{i-1}, \text{box}, v) \}, v \in \Omega$$

$$\text{Sad}(f_i, f_{i-1}, \text{box}, v) = \frac{1}{S} \sum_{p \in \text{box}'} |f_i(p) - f_{i-1}(p-v)|$$

其中, $\text{box}' = \text{box} \cap (\text{box} + v)$, S 为 box' 的面积, Ω 表示矢量的可能范围,比如:

$$\Omega = \{v | v = (d_x, d_y), d_x \in [-15, 15], d_y \in [-15, 15]\}$$

图 4 给出了搜索过程的直观说明.

在基于目标的整体运动估计过程中,由于 Sad 的计算被限定在 box' 中,故计算量比较小.另外,为了进一步减少计算量, Sad 可以在 box' 中,通过抽点进行计算,比如按 1/2 或 1/4 抽点.还可以采用三步

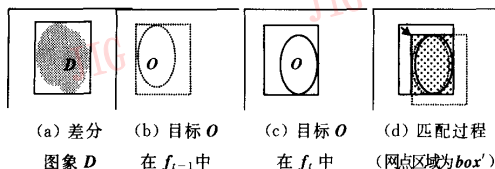


图4 目标的整体运动估计

搜索法或者其他快速算法来进行计算,以提高效率。

(2) 消除显露背景

根据上面计算的运动矢量对图象模板做如下的修正,即可消除显露的背景部分,从而得到对象模板 O 。

$$O(p) = \begin{cases} 255 & D(p) = 255 \text{ 且 } D(p - V_o) = 255 \\ 0 & \text{其他} \end{cases}$$

其中, $p \in \text{box}'$ 。

修正的过程也可以简单地描述为:先把差分模板 D 按照 V_o 平移,再把平移之后的模板同未平移时的差分模板叠在一起,这样两模板中重叠的非零部分即是运动目标 O 。

这种修正方法,是假定物体只存在平移运动,且运动的过程中基本没有形变。当然这样的假设在一定程度上是合理的,因为差分模板来自于连续的两帧,由于帧间的时间间隔很短,故物体的变形不大。由于使用了整体运动矢量及消除了局部运动矢量固有的不可靠性,从而可得到比较准确的结果。

2 帧间投射

帧间投射的目的是要根据物体运动的连续性,用前一帧得到的对象模板来弥补当前帧中出现的缺陷。比如当目标的一部分与其背景灰度相近时,物体的这一部分在当前的差分图象中便没有信号输出,而事实上,它仍是物体的一部分。为此,采用下面的更迭策略。

假设上一帧得到的最终模板为 O_{t-1} ,当前得到的差分二值图象为 D ,估计出的当前帧目标运动矢量为 V_t ,那么,现在的任务是得到当前帧的模板 O_t 。

(1) 帧间投射

根据当前估计得到的运动矢量,再把上一帧的模板投射到当前模板缓冲区(需预先将当前对象缓冲区清空),即形成投射对象 O_{proj} 。依据 V_t ,当前图象中任一坐标点 p 均可以在上一帧图象中确定其对应位置 $p' = p - V_o$,如果 p' 在 O_{t-1} 内,而且两帧对应点的像素灰度差值小于某一阈值 T ,则认为 p 属于新的对象。

$$O_{\text{proj}} = \{p \in f_t \mid (p - V_o) \in O_{t-1} \text{ 且 } |f_t(p) - f_{t-1}(p - V_o)| < T\}$$

(2) 获取当前帧的模板

首先根据对象运动矢量来得到修正的当前差分模板 D_t (如本文第一节图象差分所述),然后与 O_{proj} 叠加(求“并集”),最后再通过填充来得到当前对象模板 O_t 。

由此可见,整个分割过程中,维持着一个对象模板缓冲区,且该缓冲区最初置为空(也可以通过场景切换检测,在适当的时候把缓冲区重新置空)。

3 测试结果

(1) 算法的部分实际测试结果如图5所示。

由图5可见,最后得到的对象模板是很理想的。

(2) 算法时间消耗

测试程序的运行环境是:CPU: P III 866,内存:128MB,操作系统: Windows98,编程语言: C++ Builder 4.0;图象格式:176×144 YUV。

处理连续的30帧,每帧的处理时间(单位为ms)为:70,55,85,60,91,55,85,65,85,70,70,55,68,60,70,68,65,60,65,50,55,56,60,55,75,65,75,55,84,66,平均每帧处理时间为66.6ms,其中包括滤波、差分、同化填充、运动估计、背景消除等步骤。由此可见,在现有测试条件下,处理速度可达15帧/s。

另外需要说明的是,在编程时,主要是注重算法的原理和程序结构的清晰,而对程序的执行速度并



(a) 原帧图象

(b) 估计运动后的帧图象

(c) 运动估计后的差分图

(d) 同化填充后图象

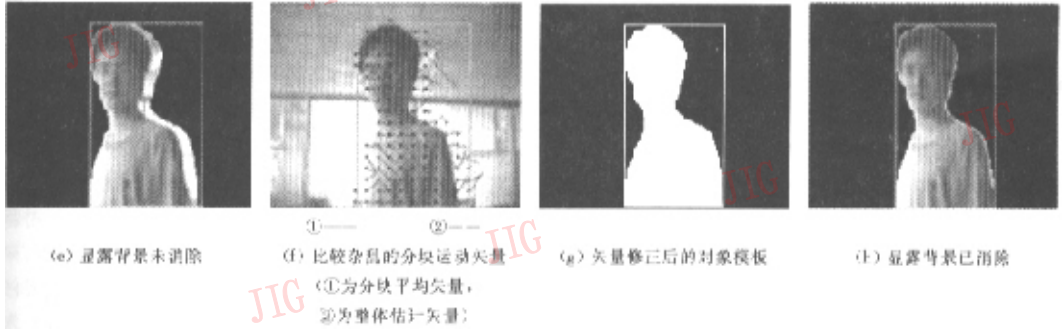


图 5 部分测试结果

没有在编程技巧上进行特别的优化, 如果对程序的结构和语句, 按照速度目标进行细致的优化, 那么相信处理速度还会有所提高。

4 结束语

本文提出了一种利用连续两帧图象之间的帧间信息来对视频序列中的运动对象进行快速分割提取的方法, 该算法不仅不依赖于固定背景, 而且能够消除差分图中的显露背景, 同时可得到运动目标较为精确的形状, 并且算法简单、快速, 鲁棒性好。经测试, 对于 176×144 分辨率的视频图象, 处理一帧图象所用的全部时间基本上在 66ms 左右, 处理速度可以达到 15 帧/s。

如果存在摄像机运动, 则可首先估计出背景的全局运动, 然后再求出与背景有相对运动的物体, 这样就可将视频运动对象分割转化成静止背景分割问题, 由于篇幅有限, 文中没有对这一问题进行讨论。本文算法的基本假设是运动物体在连续两帧之间基本不发生形变, 而对于形变比较大(或者多目标)的情况, 则必须首先根据运动一致性标准来进行分割, 然后再分别跟踪。另外, 本算法讨论的是运动对象的分割问题, 如果物体不运动或运动不明显, 则差分图象会过于零碎, 从而使本算法无法得到正确的语义分割结果。

参 考 文 献

1 王栓, 艾海舟, 何克忠. 基于差分图象的多运动目标检测域跟踪 [J]. 中国图象图形学报, 1999, 4A(6): 470~475.

2 Dong Kwon Park, Ho Seok Yoon, Chee Sun Won. Fast object tracking in digital video [J]. IEEE Transaction on Consumer Electronics, 2000, 46(3): 785~790.
 3 陈朝阳, 张桂林. 基于图象对称差分运算的运动小目标检测方法 [J]. 华中理工大学学报, 1998, 26(9): 34~35.
 4 裴继红, 谢维信. 聚类的密度函数方法. 西安电子科技大学学报, 1997, 24(4): 463~467.
 5 郑南宁. 计算机视觉与模式识别[M]. 北京: 国防工业出版社, 1998.



贾振堂 1969年生, 河南城建高等专科学校计算机工程系讲师, 现为武汉大学软件工程国家重点实验室博士研究生。主要从事视频压缩和基于人脸识别、表情识别的智能人机接口技术研究。



贺贵明 武汉大学软件工程国家重点实验室教授, 博士生导师。主要研究领域为多媒体、网络通信、模式识别等。



韩艳芳 1970年生, 河南城建高等专科学校信息中心讲师。主要从事视频压缩和流媒体方面的研究和应用开发。