一种基于时空联合的视频分割算法

黄 波, 杨 勇, 王 桥, 吴乐南
(东南大学无线电工程系, 江苏南京 210096)

摘 要: 本文提出了一种基于时空联合的视频分割算法. 在空间分割中, 引入了一种新的分水岭分割算法, 具有较快的执行速度; 在运动分割中提出了一种分层块匹配的快速运动估计算法; 最后提出了利用六个仿射参数组成的向量的范数来进行基于运动的区域合并. 实验表明: 这是一种稳健的而且执行速度相当快的视频分割算法.

关键词: 视频分割; 分水岭变换; 运动估计; 区域合并
中图分类号: TN919.8 文献标识码: A 文章编号: 0372-2112 (2001) 11-1491-04

Video Segmentation Based on Spatio-Temporal Information

HUANG Bo, YANG Yong, WANG Qiao, WU Le-nan
(Radio engineering dept., Southeast University, Nanjing, Jiangsu 210096, China)

Abstract: This paper presents a new video segmentation algorithm based on spatio-temporal information. In spatial segmentation, we introduce a new watershed algorithm which has a high implementation speed. In motion segmentation, we put forward a new hierarchical block matching motion estimation algorithm. Finally we use the norm of the six affine parameters to merge the regions. The experimental results show that it is a robust and fast algorithm.

Key words: video segmentation; watersheds transform; motion estimation; region merge

1 引言

在 MPEG-4 等新一代编码压缩标准中, 基于对象的编码和可扩展编码是非常重要的内容. 基于对象的编码就是要把视频图像分割成不同的对象或者把运动对象从背景中分离出来, 然后针对不同的对象采用相应的编码方法, 这样可以达到较高的压缩比, 而 MPEG-4 中可扩展编码也是针对视频对象的, 因此视频对象的自动分割是新一代编码技术中的研究热点和难点.

根据文献[1], 视频分割的一般方法是: 首先, 对原始的视频图像数据进行简化以便于分割, 这种简化可以通过低通滤波、中值滤波、形态滤波等来完成; 然后对视频图像数据进行特征提取, 这些特征可以是颜色、纹理、运动、帧差、位移帧差乃至语义等; 最后是进行分割的决策, 根据提取的特征决定把哪些视频数据归为一类, 这种决策往往都是基于某种均匀性标准的. 从所用的数学工具来看, 视频分割方法主要有三种: 基于贝叶斯估计理论的视频分割方法[2]; 基于聚类理论的视频分割方法[3]; 基于数学形态理论的视频分割方法[4], 其中比较实用的是后两种方法, 特别是基于数学形态理论的视频分割方法得到了广泛的应用. 从分割所在的域来看, 视频分割可以分为空间分割和时间分割, 空间分割一般是一种静态分割, 时间分割一般是运动分割. 在本文中, 我们提出了一种时空联合的视频分割方法, 在空间分割中采用了基于数学形态理论的分割方法, 在时域的运动估计中提出了分层块匹配运动估计方法. 从实验结果来看, 这种分割方法有良好的稳健性, 而且分割的速度相当快. 下文的安排是: 第二节空间分割; 第三节运动估计; 第四节基于运动的区域合并; 第五节实验及实验结果分析.

2 空间分割

大多数的视频空间分割方法都是基于数学形态方法的, 在数学形态方法中分水岭变换以其优良的性能更是引人注目[4]. 自从 Vincent 和 Soille[5] 提出一种快速实现分水岭变换的算法后, 该分水岭算法就广泛地应用于视频分割中. 然而在本文中, 并没有采用 Vincent 和 Soille 提出的算法, 而是引入了一种新的分水岭算法[6], 这种算法执行速度要比 Vincent 和 Soille 提出的算法快, 而且分割的效果良好.

首先简单的阐述一下这种算法. 这种分水岭算法中很多概念与 Vincent 和 Soille 算法是一致的, 但是非常重要的一点是提出了"浮点活动图像"的概念, 并把它作为分水岭变换的输入. 所谓"浮点"是指图像的数据类型是浮点型. 浮点活动图像在物体边界点附近具有较高的亮度值, 而在物体的内部具有较低而且均匀的亮度值, 因此浮点活动图像本身就比较粗糙的指示了物体的边界. 算法分为两步: 第一步先通过对图像进行"淹没"来消除由于噪声等原因引起的不稳定的边界. 所谓"淹没"就是判断当前象素的浮点值是否小于一给定的阈值

收稿日期: 2000-11-27; 修回日期: 2001-06-16
基金项目: 国家自然科学基金(No. 60072013 与 No. 60002008)

(该阈值称为“地下水平线”,如图1),如果小于,就把当前象素和其满足此条件的邻域象素合并.在“淹没”这一步中会产生一定数量的“湖”(或称为“峡谷”),它由低于给定阈值的所有象素点构成.通过这一步可以减弱噪声的影响并且可以减少“过分割”;第二步,若把当前象素视为“雨滴”,那么算法就要决定“雨滴”该沿地形表面象哪一个方向流,其实质也就是当前象素该分到哪一类中去,或者说当前象素该和周围的哪一个已经标记好的象素合并.一般把当前象素和邻域(可以是四邻域或八邻域)中速度下降最快的那个象素合并.下降最快是指当前象素和邻域象素差值最大.通过对图像从左至右,从上到下的扫描,不断的进行象素合并,最后形成不同分割块.

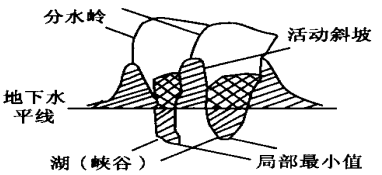

图1 分水岭算法示意图

文献[6]中是先用(1)计算出梯度图像,然后对梯度图像作一定的处理作为浮点活动图像,式(1)中f代表原始图像.

$$grad(f)=(\partial f/\partial x)^2+(\partial f/\partial y)^2 \quad (1)$$

然而通过实验发现,在对MPEG-4测试序列进行分割时,对于大部分测试帧,分割结果良好,但也有一部分测试帧分割结果不是很好.因此采用了一种新的浮点活动图像,先计算出图像的形态梯度,然后对形态梯度图像做一定处理,作为浮点活动图像.因为浮点活动图像要达到的目的就是要在物体的边界点附近具有较高的亮度值,而在物体的内部具有较低且均匀的亮度值,形态梯度恰好具有此特点.定义形态梯度如下[7]:

$$grad(f)=(f\oplus g)-(f\ominus g) \quad (2)$$

其中f代表原始图像,g代表结构元素,$\oplus$代表膨胀运算,$\ominus$代表腐蚀运算,浮点活动图像由式(3)得到.

$$fimg(f)=grad(f)*grad(f)/255.0 \quad (3)$$

形态梯度的作用可以用图2来表示(结构元素采用3×3的矩形结构).用这种浮点活动图像进行分割得到的结果更加稳健.分水岭分割能够比较准确的得到运动物体的边缘,但是分水岭算法容易造成图像的过分割,如在我们的实验中,分水岭变换后,图像一般都被分割成几百块至上千块(参见图4),因此分水岭分割往往得不到具有语义的物体,要得到具有语义物体的必须联合时间分割,对分割后的小块进行合并.因此,下面先进行运动估计,然后再进行基于运动的块合并.

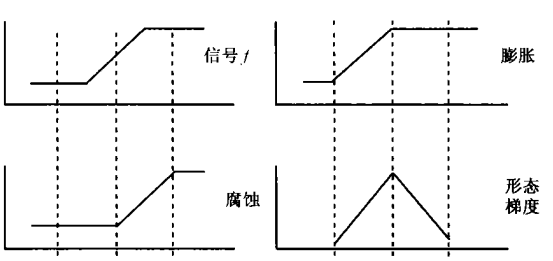

图2 形态梯度作用示意图

3 运动估计

运动估计是基于运动的块区域合并的需要,对于同一物体,或者物体的某一部分,通常具有相似的运动,可以通过这种相似性来进行区域合并.运动估计可以采用光流估计,但是光流估计的运算量往往过大,如果采用光流估计,那么视频分割的决大部分运算都消耗在运动估计上了.因此采用块匹配运动估计方法.考虑到视频分割的特点,提出了一种新的分层块匹配运动估计方法[8].

在MPEG-4测试序列中,有一类图像序列如“特雷弗(Trevor)”、“母亲和女儿(Mother and daughter)”、“美国小姐(Miss Amercia)”、“苏珊(Suzie)”、“卡耐瑞(Claire)”等,它们的背景几乎是静止的,前景的运动量也不大.在实际应用中,也大量的存在这种图像序列,如视频电话、视频会议的图像序列.在利用块匹配进行运动估计的时候,没有必要对每一个子块都在一个较大的搜索区域进行匹配,对不同的子块应该有不同大小的搜索区域且大多数子块应该限制在一个小的搜索区域内进行匹配,这样会较大的提高运动估计的速度,我们提出的分层块匹配算法就是基于这种考虑,它分三层进行,如图3.

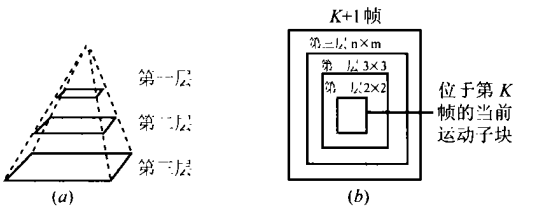

图3 分层块匹配算法

算法的第一步:把当前帧中要进行运动估计的子块(块大小为2×2)直接向下一帧投影,找到下一帧中相对应的子块(参看图3(b)),计算两个块之间的均值绝对差分MAD,如果MAD小于一个预先给定的阈值$T0$,则停止下一层的搜索,认为当前子块的运动向量$\boldsymbol{MV}$为零.这一步的实质也就是把搜寻区域限制在和子块大小一样的区域中.

算法的第二步:如果第一步中的MAD大于阈值$T0$,则算法继续在下一帧中找到与当前子块相对应的3×3的搜寻区域进行块匹配(参看图3(b)),计算各自对应的MAD,并找到这些MAD中的最小值minMAD1,如果minMAD1小于预先给定的阈值$T1$,则停止下一层的搜索,并认为当前子块的运动向量$\boldsymbol{MV}$是与minMAD1相对应的运动向量$\boldsymbol{MV}1$,在这一步中块匹配采用完全搜索法.

算法的第三步:如果第二步中的minMAD1大于阈值$T1$,则算法继续在下一帧中找到与当前子块相对应的$n\times n(n>3)$的搜寻区域进行块匹配(参看图3(b)),计算各自对应的MAD,并找到这些MAD中的最小值minMAD2,把与minMAD2相对应的运动向量$\boldsymbol{MV}2$作为当前子块的运动向量$\boldsymbol{MV}$,在这一步中块匹配可以采用完全搜索法,也可以采用三步搜索法,十字搜索法,视n的大小而定.实验证明这是一种性能优良的运动估计方法,估计速度很快.

4　基于运动的块区域合并

4.1　参数模型

在进行块运动合并时, 一般不采用非参数的稠密运动场, 而采用参数模型. 非参数模型运动估计由于不适定问题和遮挡显露问题的存在因而不精确, 不利于分割. 常见的参数模型有八参数模型、仿射参数模型等[9]. 本文采用仿射参数模型, 它由式(4)定义.

$$\begin{cases} d_x = a_1 + a_2 x + a_3 y \\ d_y = a_4 + a_5 x + a_6 y \end{cases} \tag{4}$$

其中 d_x, d_y 是象素(x, y)的运动向量.

4.2　仿射参数的计算

对于每一个分割块, 利用最小二乘法对误差函数求最小值来估计仿射运动参数. 误差函数由式(5)定义.

$$E(a_1, \cdots, a_6; S) = \sum_S [(d_x - a_1 - a_2 x - a_3 y)^2 + (d_y - a_4 - a_5 x - a_6 y)^2] \tag{5}$$

其中 d_x, d_y 是分割块 S 中象素的运动向量, 把式(5)分别对 $a_1, \cdots, a_6$ 求偏导数, 并令其等于零, 可以得到两个三元一次方程组(6)和(7):

$$\begin{cases} \sum_S (d_x - a_1 - a_2 x - a_3 y) = 0 \\ \sum_S (d_x - a_1 - a_2 x - a_3 y) x = 0 \\ \sum_S (d_x - a_1 - a_2 x - a_3 y) y = 0 \end{cases} \tag{6}$$

$$\begin{cases} \sum_S (d_y - a_4 - a_5 x - a_6 y) = 0 \\ \sum_S (d_y - a_4 - a_5 x - a_6 y) x = 0 \\ \sum_S (d_y - a_4 - a_5 x - a_6 y) y = 0 \end{cases} \tag{7}$$

不妨把每个块的六个仿射运动参数看成一个六维的向量, 那么每个块的仿射参数就是这六个方程的解向量.

4.3　块区域合并

仔细观察一下式(6)、(7)这两个方程组, 当分割块中的象素运动向量全部为零时, 那么解出来的六个仿射运动参数也全为零, 因此由这六个仿射运动参数组成的解向量的 2—范数(下面简称范数)也为零. 如果分割块中象素的运动向量都很小时, 那么由上述两个方程组解出的六个仿射参数所组成的解向量的范数也很小. 在 MPEG-4 标准测试序列中, 对“特雷弗”、“母亲和女儿”、“美国小姐”、“苏珊”、“卡耐瑞”等序列来说, 它们的背景几乎是静止的, 前景(运动物体)具有一定的运动, 因此, 如果分割块属于前景, 则它们六仿射参数组成的解向量的范数应该较大, 反之, 如果属于背景, 则它们的解向量的范数应该很小. 于是, 可以设定一个阈值, 如果分割块的仿射参数组成的解向量的范数大于阈值则为前景, 否则为背景.

最后总结一下本文提出的视频分割算法:

(1)对原始的视频图像数据进行中值滤波, 减少噪声的影响;

(2)用式(2)、(3)计算浮点活动图像;

(3)进行分水岭变换;

(4)利用分层块匹配进行运动估计;

(5)对于象素个数大于某一常数的分割块计算它的仿射运动参数, 并求出仿射参数组成的解向量的范数;

(6)基于运动的块合并;

(7)进行形态滤波, 对分割后的视频图像进行后续处理.

5　实验及实验结果分析

5.1　实验结果

随机从“美国小姐”、“特雷弗”、“卡耐瑞”三组测试序列中抽出三帧视频图像(QCIF 格式), 用本文前面提出的方法进行视频分割, 结果如下:

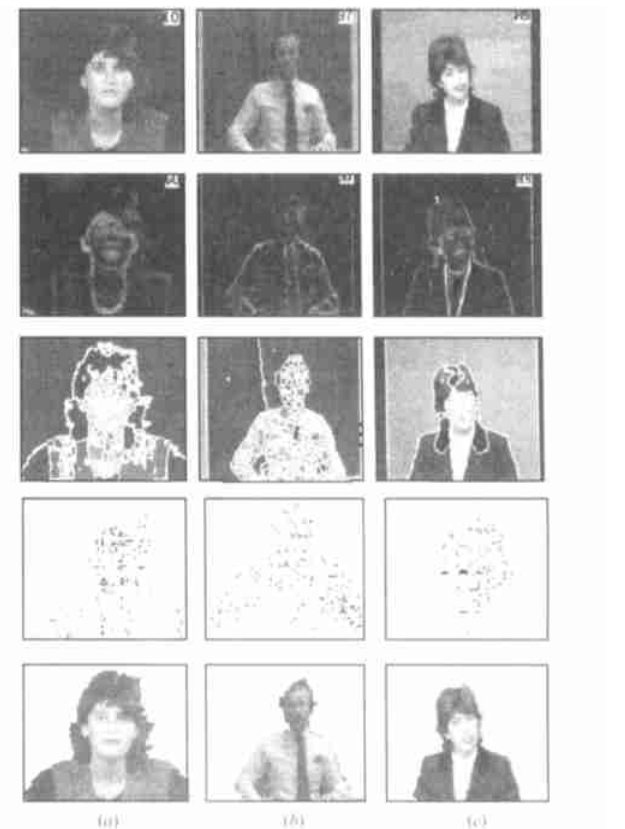

图 4　三帧标准测试序列图像的分割结果

图 4 中的(a), (b), (c)图分别是标准测试序列“美国小姐”、“特雷弗”、“卡耐瑞”的 68 帧、97 帧和 89 帧的处理结果; 从上到下分别是三帧测试图像的原图、形态梯度图、空间分割结果、运动估计结果和最后的分割结果. 把 Vincent 和 Soille 提出的分水岭算法称为 V-S 方法, 把本文中采用的分水岭分割算法称为 PDS 方法, 则表 1 是我们重复文献[6]的实验得到的两种分水岭算法对不同图像大小的标准测试图像辣椒(PEPPERS)进行分水岭分割所用的时间比较. 把不分层的全局搜索块匹配运动估计算法称为 BM 方法, 把本文提出的分层块匹配算法称为 HBM 方法, 表 2 是这两种方法对测试序列“特雷弗”的 60、70、80 帧进行运动估计所用的时间比较, 表 3 是本文分割算法中各个部分所用的时间. 实验中使用的是 PII

233MHZ 计算机.

表 1 两种分水岭算法计算时间的比较

PEPPERS	128* 128	256* 256	512* 512
V-S	160	570	1780
PDS	40	180	730

表 2 两种运动估计计算时间的比较

特雷弗	第 60 帧	第 70 帧	第 80 帧
BM	339	340	340
HBM	86	77	108

表 3 本算法中各个部分所用的计算时间

时间 / 序列	分水岭变换	运动估计	仿射参数估计及合并	算法执行时间
美国小姐	145	93	55	420
特雷弗	95	110	56	270
卡耐瑞	95	83	60	266

注: 算法执行时间没有包括滤波时间, 表 1、2、3 中的时间单位均为毫秒.

5.2 实验结果分析

(1) 本算法的稳健性良好. 通过测试“美国小姐”、“特雷弗”、“卡耐瑞”、“苏珊”、“母亲和女儿”等标准测试序列发现只要前后两帧图像的内容不发生大的改变, 该算法都能取得较好的分割结果.

(2) 本算法具有相当快的执行速度. 从表 1、2、3 中可以看到本文提出的分割算法中各个部分的执行速度都很快. 从国内外的关于视频分割方面的参考资料来看, 分割一帧视频图像是比较耗时的. 如文献[10]中, 在 Ultrasparc 2/2000 的计算机上, 分割一帧 MPEG-4 的标准的彩色图像(QCIF 格式)需十几秒钟, 而本文提出的算法对一帧图像进行分割所用的时间不超过 0.5 秒, 这个速度是相当快的.

(3) 本文提出的算法中有几个阈值: 一个是分水岭变换中“地下水平线”阈值; 另外两个是运动估计时的阈值 $T0$、$T1$; 还有一个是基于运动的区域合并时仿射参数范数的阈值. 对于运动估计时的两个阈值 $T0$、$T1$, 在算法中预先设定后, 对实验中不同测试序列都适用, 因此真正要调整的是“地下水平线”阈值和仿射参数的范数阈值. 对于同一测试序列来说, 这两个阈值调整一次后, 不需要再调整. 下一步我们将在自动确定分割中所用的阈值方面做一些工作, 提高本算法的性能, 以期达到自动的视频分割.

参考文献:

[1] P Salembier, F Marques. Regiorr-based representations of image and video: segmentation tools for multimedia services [J]. IEEE Trans. CSVT, 1999, 9(8): 1147- 1169.

[2] M M Change, A M Tekalp, M I Sezan. Motiorr field segmentation using an adaptive MAP criterion [A]. IEEE ICASSP93 [C], Minneapolis, MN, 1993, V: 33- 36.

[3] R Castagno, T Ebrahimi, M Kunt. Video segmentation based on multiple features for interactive multimedia applications [J]. IEEE Trans. CSVT, 1998, 8(5): 62- 571.

[4] F Meyer, S Beucher. Morphological segmentation [J]. Journal of Visual Communicaton and Image Representation, 1990, 1(1): 21- 46.

[5] L Vincent, P Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations [J]. IEEE Trans. PAMI, 1991, 13(6): 583- 598.

[6] P D Smet. Performance and scalability of high optimized rainfalling watersheds algorithm [A]. CISST' 98 [C], 1998, Lasvegas, Novada, USA.

[7] 崔讫. 图像处理与分析数学形态学方法及应用 [M]. 科学出版社, 2000 年.

[8] 黄波, 杨勇, 等. 一种用于视频分割的快速运动估计方法 [J]. 电路与系统学报, 2001, 6(1): 69- 71.

[9] A Murat Tekalp. Digital Video Processing [M]. Englewood Cliffs, NJ: Pretice-Hall, 1995.

[10] S Herman, H Mooshofer, H Dietrich, W Stechele. A video segmentation algorithm for Hierarchical Object Representations and Its Implemtation [J]. IEEE Trans. CSVT, 1999, 9(8): 1204- 1215.

作者简介:

黄 波 男. 1975 年出生于湖南. 东南大学无线电系, 博士研究生, 主要研究方向为数字视频技术, 多媒体信息处理技术.

杨 勇 男. 1973 年出生于陕西. 东南大学无线电系, 博士研究生, 主要研究方向为数字视频技术, 多媒体信息处理技术.

王 桥 男. 1966 年出生于安徽. 理学博士, 副教授. 现从事信号分析与表示及量子信息论等相关研究, 目前主持一项国家自然科学基金项目.

吴乐南 男. 1952 年出生于福建. 教授, 博士生导师. 从事多媒体信息处理等研究与开发工作, 目前主持一项国家自然科学基金项目.