用于视频对象平面生成的运动对象自动分割 [1]

俞毅刚 黄 艺

(南京邮电学院信息工程系 南京 210003)

摘 要 新的视频编码标准 MPEG-4 具有基于内容的功能. 它把图像序列分解成视频对象平面 (VOP), 每个 VOP 代表一个运动对象. 文中提出了一种提取运动对象的新的视频序列分割算法, 算法的核心是一个对象跟踪器, 它利用 Hausdorff 距离将对象的二维二值模型与后续帧进行匹配, 然后采用一种新的基于运动相连成分的模型刷新方法对模型的每一帧进行刷新. 初始的模型自动产生, 再利用滤波技术滤除静止背景, 最后, 利用二值模型从序列中提取出 VOP .

关键词 MPEG-4, 视频对象平面, 对象跟踪, 视频序列分割

中图号 TN919.8

1 引 言

传统的视频编码标准如 MPEG-1 , MPEG-2 , H.261 或 H.263 不需要场景的分割和分析. 尽管它们能获取较高的压缩率并且适用广泛, 但随着多媒体应用和基于内容的交互性的普及, 需要有一种新的编码方案.

MPEG-4 标准介绍了视频对象平面 (VOP) 的概念, 它具有基于内容的功能. 输入图像序列的每一帧被分割成任意形状的图像区域, 这样每一个 VOP 描述一种语意的对象或视频内容, 每个 VOP 都包含有形状、运动和纹理的视频对象层.

在 VOP 生成中一个问题是对象没有与颜色、亮度和光流等相似的特性. 传统算法 [1,2] 在生成 VOP 时采用了一种变化检测方法. 尽管它比采用运动场的方法有效得多且利于计算, 但这种方法具有两个缺点: 如果运动对象不含有丰富的纹理, 那么只有边界遮挡区域被标记为变化区域, 而对象内部将不会发生变化; 停止运动一定时间的对象或对象的一部分将被丢弃, 这在基于内容的应用中是不能接受的. 为了解决上面的问题, 变化检测方法使用了一个存储器. 但由于存储器长度的限制, 会使已提取出对象的那部分背景仍被看作为对象, 从而使得最后获取的 VOP 比实际对象要大, VOP 与实际对象的吻合程度取决于运动的速度和存储器的长度. 因此, 传统的分割算法不再适用于 VOP 生成.

光流和运动场在理论上是可用的, 但它们对噪声极其敏感, 并且它们的精确性受到孔径效应和遮挡问题的限制. 精确的图像边界定位比较困难, 而且还必须使用附加的方法来填充对象内的小孔.

本文提出了 VOP 生成的一种新的算法, 它能从视频序列中自动提取出运动对象. 由于这些对象与背景的运动不同, 所以有一部分运动信息必须用于分割算法. 本文提出的算法基于模式识别和对象跟踪原理, 避免了许多与运动估值相关的问题. 算法的核心是一个对象跟踪器, 它在整个视频序列中建立起对象的时间相关性, 这对基于内容的应用是很重要的. 并且当对象以任意长时间停止运动时, 它也能很好地保持跟踪. 跟踪器输出对象的二值模型序列, 从中提取出图像序列的 VOP . 在运动对象自动检测中, 介绍了运动相连成分 (MCC) 这个概念, 为了允许形状方面的较大变化, 本文使用了一种新的模型刷新方法, 它对图像每帧进行刷新. 然后使用滤波器滤除静止背景, 这不但提高了分割的稳定性而且降低了运算的复杂性. 用这种算法获取的 VOP 比其它一些技术所获取的 VOP 要精确得多.

[1] 1999-06-24 收到, 1999-11-20 定稿
邮电部中青年基金资助课题

图 1 Hausdorff 距离的计算， $h(O,I)$ 和 $h(I,O)$ 中的最大值即为 $H(O,I)$

2 算法原理

本文算法首先提出广义 Hausdorff 距离，用于自动获取对象的模型并能与后续帧很好地进行匹配，然后利用一种新的刷新技术，对每帧进行刷新以适应图像形状的旋转和变化，并提出了一种新的滤波技术滤除静止背景，最后提取出 VOP 作为 MPEG-4 的输入。

2.1 Hausdorff 对象跟踪

Hausdorff 对象跟踪是本文算法的核心，它在整个视频序列中建立起对象的时间相关性。这对基于内容的功能相当重要，即使对象以任意长时间停止运动，算法也始终保持跟踪对象。

为了利用 Hausdorff 距离实现对象模型与后续帧的匹配，首先要获取对象的边界图像模型。由于灰度图像对亮度的变化过于敏感，它们一般不适用于模块或对象的匹配 [3]。本文利用二值边界图像，模型的边界点不受对象边界的限制，并采用 Canny 操作 [4] 来获取边界图像。

在获取边界图像的二值模型以后，要将它与后续帧进行匹配。一个可靠的匹配方法必须能检测出形状在进行变换、旋转和变化的对象。这使得那些通过模型和后续帧的相互关联获取新的位置的基本的模块匹配方式不再适用。这是因为这种匹配方法在处理形状发生变换、旋转和变化的图像时计算量极大。为此，本文通过最小化 Hausdorff 距离来获取视频对象模型和后续帧之间的最佳匹配。

令 $O=\{o_1,\cdots,o_m\}$ 为构成需跟踪的对象模型的边界像素的一组特征点， $I=\{i_1,\cdots,i_n\}$ 为边界图像的边界像素的另一组特征点。其中 m,n 分别是对象模型和边界图像的像素点。

定义 Hausdorff 距离为

$$H(O,I)=\max\{h(O,I),h(I,O)\}=\max\{\max_{o\in O}\min_{i\in I}\|o-i\|,\quad \max_{i\in I}\min_{o\in O}\|i-o\|\} \tag{1}$$

即对每个模型点 O，计算出到最近的图像点 i 的距离。 $h(O,I)$ 就是其中的最大值。同样，对每个图像点 i，计算出到最近的模型点 O 的距离， $h(I,O)$ 就是其中的最大值。 $h(O,I)$ 和 $h(I,O)$ 较大的值就是 Hausdorff 距离 $H(O,I)$(见图 1)。

(1) 式的定义会引起一些问题，如果有一个模型或图像点远离中心，即使其它所有点匹配得很好， Hausdorff 距离也会很大，因此，本文提出广义 Hausdorff 距离，将 (1) 式中定义的距离以增序排序，然后抽样第 k 个值，即

$$h_k(O,I)=k\text{-th}_{o\in O}\min_{i\in I}\|o-i\| \tag{2}$$

当 $k=m$ 时，(2) 式与 (1) 式等效；当 $k<m$ 时，$m-k$ 个点远离中心但并不增大 Hausdorff 距离，这个特性在处理部分遮挡或形状快速变化的对象时是很有用的。类似的 $h_l(I,O)$ 定义为增序距离中的第 l 个值。利用参数 k 和 l，可以选择靠近图像点的模型点点数，反之亦然。

由于 Hausdorff 距离自动地选择 k(或 l) 个最佳匹配点，所以模型点和图像点之间不需要点相关性。这在对象改变形状时很有帮助。当所有与图像相关的模型变换的模型和图像之间的 Hausdorff 距离最小时即为最佳匹配。

Hausdorff 距离可以利用距离变换来计算。距离变换 $t=(t_x,t_y)$ 通过对边界像素置 0，而对非边界像素赋无穷或足够大的值进行初始化，然后依次计算每个像素的距离并依次刷新，最终得到每个像素到最近的边缘像素的距离。首先对边界图像进行距离变换，求出每个像素到最近像素的距离，然后，对所有的变换 $t=(t_x,t_y)$，计算出 $h_{k,t}(O,I)$，$h_{k,t}(O,I)$ 中的 t 表示 $h_k(O,I)$ 依赖于变换 t。因此，利用矢量 t 对对象 O 进行变换，模型点 O 位置的距离变换直接给出了点 O 与最近的边缘像素之间的距离。这些距离然后以增序排列，选择第 k 个值作为 $h_{k,t}(O,I)$。注意 $h_{k,t}(O,I)\leq T$，T 为一阈值，凡是 Hausdorff 距离大于 T 的匹配都将被滤去，这可以通过一个前端扫描终端实现。同样以类似的方法求得这些变换的 $h_{l,t}(I,o)$，然后得出 $H_t(O,I)$。根据最小 Hausdorff 距离 $H_t(O,I)$ 得出新的位置，也就是模型所进行的变换 t。

为了进一步加速匹配过程，应对搜索区域进行限制。变换将被限制在与前一帧对象位置相关的各个方向上的特定的几个像素上。

2.2 运动对象的初始化

利用 Hausdorff 距离自动获取对象模型并与后续帧进行精确匹配后，就应对运动进行初始化。由于对象的初始位置是未知的，为了进行运动对象分割，必须定出运动对象的初始位置。对于非静止背景或用运动摄像机获取的场景，经常要进行全局的运动补偿。

首先，假设背景是静止的，在整个场景上只有一个运动对象。这也可以扩展为多个对象，只要它们不相互重叠。利用两帧之间颜色和亮度的差异是检测变化区域的最有效的方法之一，较大差值表示对象正在运动或改变形状。如果对象的纹理不是很密，就只能看到运动对象的边界而看不到对象本身。这也正是本文为跟踪器获取模型所需要的。

差值图像所需的阈值取决于序列的运动速度、亮度和噪声变化等特性。在本文的算法中，通过统计测试和背景活动测量，以一个输入参数的形式给定。在确定阈值时，如果阈值选得太小，随后的细化步骤将确保差值图像的所有部分都是一个像素宽；如果阈值太大，刚开始丢失的对象成份在每一帧对模型进行刷新时可重新获取。这两点对于运动对象的初始化是很重要的。

遮挡区域一般不止一个像素宽，可以通过对差值图像进行腐蚀来细化。在细化过程中相连成份不被分开是很重要的，在这个过程中，孤立的噪声像素被消除了，原因是它们不像是一个对象。

经过细化后，属于对象的像素相互连接，而噪声像素孤立成群。在二值图像中找出相连成分的一个简单算法是相连成分标记。大于特定阈值的成分被记为属于运动对象，称它为运动相连成份 (MCC)。因为噪声成份比那些属于对象的成份要明显小得多，所以能自动获取这个阈值。这样，通过发现 MCC 就可以检测到运动对象。当然，仍然需要通过抽样二值边界图像中离 MCC 较近距离 (1–2 像素) 的所有像素来获取对象跟踪器的初始模型。这可以利用距离变换很容易地实现。

在非静止背景的情况下必须进行全局运动补偿。因此，相关矢量场利用等级匹配法进行计算。全局运动估值比较容易实现，这是因为：全局运动相对简单，且包括图像变换，镜头移动，可能还有放缩。在很多应用中，与独立的运动对象相比，背景区域要大得多，因此，遮挡区域的影响或相关域的错误极小。可靠的参数估值技术将获得很好的结果。

本文使用仿射模型和最小均方准则来估计参数。获取初始模型的一个直接方法是计算排列的帧间差值。但由于运动模型和估计的相关场的不精确性，这种差值图像中噪声太多。所以，本文提出了一种新的方法。块运动估值算法把帧分成多个方块，通过将估值相关矢量与从仿射全局运动模型综合得到的矢量相比较可以确定与全局运动不同运动的块。与全局运动不同的一致运动的相连块就是运动对象。初始模型在这些块里包含有边缘像素。

2.3 模型刷新

当一个跟踪对象经过视频序列时，它的形状可能会发生旋转或变化，因此，模型就必须每帧进行刷新。这个步骤在具有杂乱背景或运动摄像机的场景的情况下的实现是相当困难的。

在有些情况下，对象的一部分比对象的其余部分变化或运动快得多 (如行走的人的手、腿比身体运动快得多)。因此本文提出了一种新的刷新技术，它包括两个成分。一个成分负责缓慢运动部分，另一个成分负责比整体对象运动快的部分。这两个成分的结合就构成了一个可靠的刷新算法。

第一个成分刷新缓慢运动部分。如果靠近被移动的旧模型的的像素是对象的一部分，前一帧的模型将移动到最佳匹配的新的位置。这样，与移动的旧模型特定距离 T_s 内的所有边界像素 (一般是 1–3 个像素) 被安排给新模型。这可以通过计算旧模型的距离变换和寻找到边界图像中距离变换值小于或等于 T_s 的所有点来实现。T_s 值越大，整个对象被包括进模型的可能性就越大。但是，这也增大了背景成为模型一部分的可能性。为了避免选出杂乱的背景，事先滤除静止背景更为可取。

第二个成分选出较快运动部分。像初始化过程一样，它基于运动相连成分这个概念。通过增加在这些 MCC 特定距离内的所有边界像素，与跟踪对象相连的成分被用来刷新相应的模型。结合这两个成分得出的一个刷新模型把缓慢和快速运动这两部分很好地选取出来。

2.4 滤除静止的杂乱背景

如果边界图像的所有像素都属于对象，那对象跟踪的获取就相对比较容易，但实际的很多序列中包含有杂乱的背景，因此有必要在模型匹配和刷新之前先滤除这些杂乱的背景。否则，如果背景边界点与模型相当近的话，模型刷新时可能会选出这些背景边界点。注意： Hausdorff 匹配可以很好地处理杂乱背景，但在滤波以后使用边界图像更为可取，这样可以减少图像点的数目和减少运算时间。

为了滤除静止背景，可以用滤除前一帧已是边界像素的所有边界像素来完成。但是这种简单的二进制差值法对噪声很敏感，它同时会滤除停止运动的对象。

本文提出了一种新的滤波技术，它对每个像素作为边界的次数进行计数。如果计数值超过一个阈值，这个像素就被认为是背景部分而被滤除。计数器只对没有被对象遮挡的像素进行刷新。当所有对象的位置已知时，计数值在处理完一帧图像刷新计数器时获取。因此，本文只能收集到真正被归类成背景的像素的信息。在刚开始的几帧，计数器不能给出可靠的结果，因此在收集到足够的数据之前，可以使用连续帧间的简单的二进制差值。

本文提出的滤波器保存属于对象的边界像素要比简单差值法好很多，它对噪声不敏感；更重要的是，即使在对象停止运动任意长时间的情况下它照样起作用。因为计数器值不会由于对象的位置而增加，所以它绝不会超出滤波的阈值。

静止背景的假设对许多应用都很有效。但是，由于全局运动估值和补偿不能完善地调整边界图像，滤除运动背景将比较困难。

2.5 VOP 的提取

在前面论述的基础上， VOP 的提取步骤为： (1) 利用 Hausdorff 距离自动获取对象模型并与后续帧进行匹配； (2) 运动对象初始化； (3) 每帧对图像进行刷新； (4) 滤除静止背景。

经过这些过程以后，跟踪器输出了对象模型的二值边界图像序列。后面的步骤是从视频序列中提取出相应的对象，也就是提取出对象的 VOP 。这可以通过在每一行找到第一个和最后一个模型点来完成。他们中间的像素被安排给 VOP ，然后对每一列重复进行类似的处理过程。

3 实验结果与讨论

本文利用两个图像序列对算法进行测试，以检测本文算法的有效性。在第一个大厅监视器的图像序列中，背景没有移动，但很杂乱，并有很高的噪声。原图像序列中第 32 帧中人的二值

模型的 VOP 见图 2 。模型对于较大的形状变化也能很好地适应，得到的 VOP 比其它一些算法更清晰、更准确。 VOP 中的微小的齿状边缘是由于提取算法简单造成的。

在第二个海岸警卫队的图像序列中摄像机跟踪船运动，这样背景是运动的。图 3 所示的结果表明船能被很好地分离出来。最大的问题是船下的波浪，因为它们是时变的，并且与跟踪对象很近，这使得跟踪器将波浪也包括进模型中去。

本文提出的广义 Hausdorff 距离解决了传统定义中个别点远离中心而引起 Hausdorff 距离增加的问题。由于只抽样其中的第 k(或 l) 个值，其它点远离中心并不会增大 Hausdorff 距离。这个特性在处理部分遮挡或形状变化的对象时是很有用的，利用 Hausdorff 距离获取的匹配非常精确。新的基于运动相连成分的刷新方法通过结合分别负责缓慢和快速运动的两个成分每帧进行刷新，能够很好地适应图像形状的旋转和变化。对边界像素进行计数的滤波技术对噪声不敏感，更重要的是即使对象停止运动任意长时间它也起作用，这种方法比简单差值法要好得多。

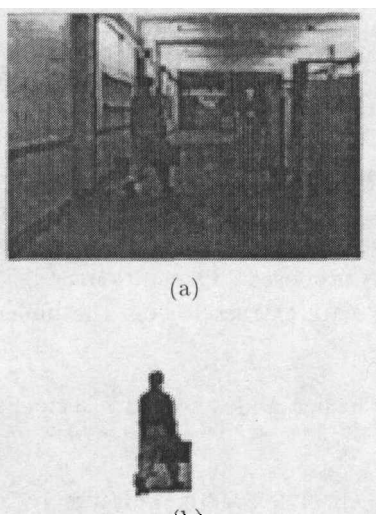

图 2　(a) 原始图像　(b) 提取出的 VOP

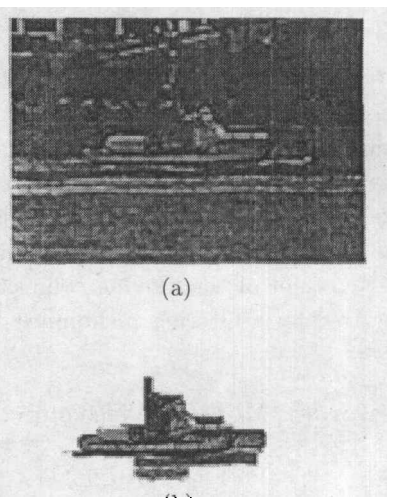

图 3　(a) 原始图像　(b) 提取出的 VOP

4 结　论

本文提出了一种新的基于对象跟踪的视频序列分割算法。利用 Hausdorff 距离自动获取了对象的模型并能与后续帧很好地进行匹配。为了允许形状的旋转和变化，该模型通过利用一种新的刷新技术每帧进行更新。这种新的技术包含的两个成分分别对应缓慢和快速变化和运动的部分。本文还更进一步提出了一种新的滤波方法来改进在静止背景下的性能。实验结果表明该算法可以从静止和运动背景中较好地提取出 VOP 。

参　考　文　献

[1] R. Mech, M. Wollborn, A nice roust method for segmentation of moving objects in video sequences, in IEEE Int. Conf. Acoust., Speech, Signal Processing, ICASSP'97. Munich, Germany, 1997, 4(4), 2657-2660.

[2] R. Mech, P. Gerken, Automatic segmentation of moving objects (partial results of core experiment n2) in ISO/IEC JTCI/SC29/WG11MPEG97/ m1949, Bristol, U.K., Apr., 1997.

[3] G. Adiv, Determining three-dimensional motion and structure from optical flow generated by several moving objects, IEEE Trans. on Pattern Anal. Machine Intell., 1985, 7(7), 384–401.

[4] J. Canny, A computational approach to edge detection, IEEE Trans. on Pattern Anal. Machine Intell., 1986, 8(11), 679–698.

GENERATING VIDEO OBJECT PLANE BY AUTOMATIC SEGMENTATION OF MOVING OBJECTS

Yu Yigang Huang Yi

(*Dept. of Info. Eng., Nanjing Institute of Posts and Telecomm., Nanjing* 210003, China)

Abstract The new video coding standard MPEG4 is enabling content-based functionality. It decomposes sequences into video object planes (VOPs) so that each VOP represents one moving object. This paper presents a new automatic video sequence segmentation algorithm that extracts moving objects. The core of this algorithm is an object tracker that matches a two-dimensional binary model of the object against subsequent frames using the Hausdorff distance. The initial model is derived automatically and a new model update method based on the concept of the moving connected components is proposed. The stationary background is removed by a filtering technique. Finally, the VOPs are extracted from the binary model sequence.

Key words MPEG-4, Video object plane, Object tracking, Video sequence segmentation

俞毅刚： 男， 1975 年生，硕士生，研究方向为多媒体通信.

黄 艺： 男， 1968 年生，博士生，研究兴趣为视频信号处理、小波和分形编码以及多媒体通信等.