

运动目标自动特征提取与跟踪

李鹏杰 杨树元

(中国科学院声学研究所 北京100080)

摘要 运动目标分割与跟踪越来越成为国内外研究的热点问题,因为它有着很广阔的应用前景和发展趋势,特别是在 MPEG4标准推出以后,如何有效地进行目标与背景的分割,如何不丢失目标的跟踪,成为使用 MPEG4标准达到好的压缩效果所必须解决的问题。不仅是在多媒体视频处理领域,在道路检测、车辆控制、军事侦察和机器人控制等领域目标检测和跟踪有着更重要的意义和应用。而诸如自动控制、道路监测等领域,一般环境比较恶劣,图像的信噪比不高,背景比较复杂,运动目标速度比较快。

本文主要是针对后一种情况,在复杂背景和噪声比较大的情况下,提供一种有效的快速的目标检测、跟踪的算法。我们首先对拍摄的一帧图像进行形态学滤波,再通过两帧之间的比较和运动补偿算法得到粗略的目标图像,通过进一步的图像分割可以得到更精确的目标信息也就完成了目标的检测。目标跟踪是通过在每个目标的参数计算和模板匹配来完成的。

关键词 目标提取,跟踪,特征提取,数学形态学

1. 引言

随着多媒体技术的发展,像 MPEG1、MPEG2那样只是单纯为了提高压缩比而定义的压缩方案已经不能满足人们的需要,因此 MPEG4压缩标准^[1]中通过“视频目标平面”(Video Object Plane 或 VOP)而引进了一些基于内容的功能(content-based functionalities)。VOP 代表视频序列中的一个目标(Object)或具有“语义意义的区域”。每一个目标的形状、纹理和运动信息被编码在不同的比特流中的,这样便允许对每一个目标进行更改和操纵,因而提供“基于内容的交互”(content-based interactivity)功能和“基于内容的可分级性”(content-based scalability)功能。VOP 的引入意味着要想进行视频序列的压缩首先要完成图像中对象的提取和跟踪,由于视频图像一般纹理非常复杂、色彩非常丰富,人物作为视频的核心要素动作形式多样、速度一般比较缓慢而且主要是非刚体运动。也就是要在应用于多媒体的视频分割领域达到比较好的效果,更多地要注意人物的柔性动作、细节动作、复杂颜色、纹理的处理。在这些方面已经有一些比较好的相关算法,比如 Meier 和 Ngan^[2,3]通过提取2值模型并用 Hausdorff 距离来进行跟踪;Wang^[4]通过 watershed 算法和时域跟踪进行视频分割;Deng 和 Manjunath^[5]使用颜色量化和空间分割的方法进行颜色纹理分割等等。

而在其他应用领域如道路、车辆的监控,背景一般是大地、天空等大尺度静止参照物,而其中的目标相对于参照物有明显的运动轨迹和速度,而摄像机用于监控时一般是静止的,跟踪时一般变为动态的。而在这些领域的运动目标一般来讲都是刚体,比如

汽车、轮船、飞机等等,这些目标一般运动速度比较快,图像中有明显的大小变化。以上这些因素都使目标容易形成光流和刚体模型。而另一方面,由于这些领域的应用一般都以自然界做为背景,使得不可预知的情况增多,比如风的因素、车的振动、天上的云等等都会带来各种干扰和噪声,使得图像信噪比下降;而且由于一般摄像机都是大尺度的检测使小目标的检测、物体成像的明显变化和多目标的变化都成为目标检测和跟踪中要解决的重要问题。同样,这个领域的工作也已经进行了一段时间,有一些算法被提出,如 Smith 和 Brady^[6]通过光流进行图像分割然后根据形成的簇进行跟踪,还有使用小波分析进行目标提取的方法^[7],通过图像描述和最优路径搜索进行跟踪的方法^[8]等。

本文是在研究了以上的算法之后提出的一种快速的对噪声不敏感跟踪效果比较稳定的方法。整个算法分为分割和跟踪两个部分。目标的提取有图像分割来完成,主要经过了数学形态学滤波,运动预测,边界提取等几个步骤。形态学滤波是对图像的一个预处理步骤,通过形态学的处理我们可以对噪声进行比较好的抑制而对图像内容没有什么损害;通过快速的运动预测我们可以得到两帧图像间的运动信息,正如我们以前所分析的,我们的目标就存在于这些运动像素之中;运动预测可以让我们得到一个粗略的目标图像,进一步的目标提取则通过边缘检测和跟踪来完成。在进行了目标提取之后,我们就可以在每一帧图像中找到要跟踪的目标,跟踪步骤是使每一帧的目标一一对应。本文使用的是基于模板的参数方法,通过对目标各种参数的计算,在一定的规则下进行目标的匹配。算法的主要流程见图1。

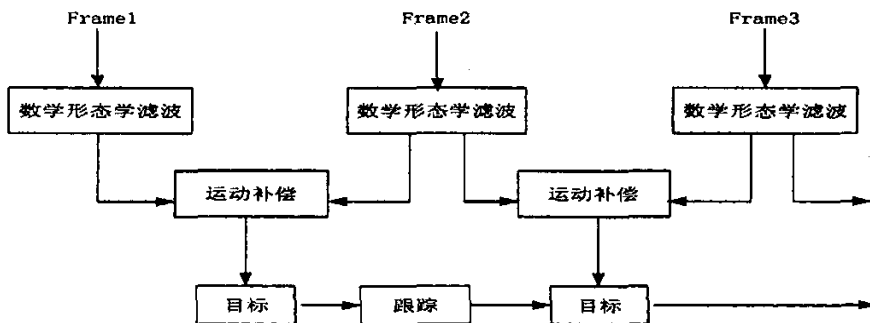


图1 目标检测和跟踪流程

本文以后的章节将做如下安排:第2部分我们将介绍一下需要的数学形态学和摄像机运动模型的背景知识;第3部分和第4部分分别介绍本算法的目标提取和目标跟踪方法;在文章的第5部分我们将给出一些实验结果并进行相关的分析;最后我们将做一些总结和展望。

2. 背景

2.1 数学形态学

数学形态学的主要内容是设计一系列的变换、算法、概念用以描述图像的基本特征。这些工具不同于常用的频域或空域的方法,而是建立在积分几何和随机集论的基础之上的,这是因为积分几何可以得到各种几何参数的间接测量,并且描述图像的体式性质,而随机集论则适用于描述图像的随机性质。

数学形态学的变换都是在结构元素的作用下完成的,这种结构元素类似于卷积核,它在图像上逐像素的移动以完成对整幅图像的处理。数学形态学的基本操作是腐蚀和膨胀。简单的腐蚀是消除物体的所有边界点的一种过程,其结果是剩下的物体沿其周边收缩。一般的腐蚀概念定义为:

$$E = I \otimes S = \{x, y | S_{xy} \subseteq I\} \quad (1)$$

其中 E 表示腐蚀以后的结果, I 表示被处理的图像, S 表示一个结构元素, S_{xy} 表示原点在 (x, y) 时的结构元素。

简单的膨胀操作是将与物体接触的背景点合并到该物体中的过程,其结果是使像素看起来向外扩展了。膨胀的定义为:

$$D = I \oplus S = \{x, y | S_{xy} \cap I \neq \emptyset\} \quad (2)$$

也就是说膨胀产生的图像 D 是当结构元素 S 的原点移到 (x, y) , 结构元素与图像的交集不为空时得到的。

数学形态学对于图像的处理都是腐蚀、膨胀这两个操作的组合,在这两个操作的基础上最常用的两个形态学运算是开运算和闭运算。开运算是先腐蚀后膨胀的过程。定义为:

$$Open = I \cdot S = (I \otimes S) \oplus S \quad (3)$$

开运算可以消除细小物体、在纤细点处分离物体,而且在平滑较大物体的边界时又不明显地改变其面积。由此可见开运算不仅可以消除图像中的分散的噪声而且适当地选择结构元素可以消除一些较大的噪声点簇,平滑图像边缘。

闭运算是先膨胀后腐蚀的过程,其定义为:

$$Close = I \cdot S = (I \oplus S) \otimes S \quad (4)$$

闭运算可以填充物体内细小的空洞、连接临近物体,也可以在不明显改变物体面积的情况下平滑其边界。连续的开闭运算可以填充物体区域的一些错判的孔,消除背景区域上散布的一些小的噪声物体。

数学形态学操作的对象一般都是2值图像,通过腐蚀、膨胀操作的各种组合来完成图像的细化、抽骨架、分割等运算^[9]。不过这些形态学运算推广到灰度图像一样有着很好的效果,本文推荐的算法就是以数学形态学运算为基础,通过对图像的开闭运算组合达到滤波的效果。

2.2 摄像机运动估计

一个视频序列中物体的运动是由其自身的运动和摄像机的运动叠加而成的。虽然在我们所讨论的领域里有很多时候摄像机是静止的,比如十字路口、银行里摄像机的监测,这时图像序列背景是固定的,物体的运动很容易被分离出来。但是在自动控制、跟踪、制导等领域摄像机是随检测到的情况运动的,这时为了得到纯粹的目标的运动情况我们需要消除摄像机的影响。

对于摄像机运动的研究已经进行了比较长的时间,并且有不少论文对此进行了讨论。我们使用的是投影和双线性模型来计算摄像机的运动,用这两种模型差不多可以计算诸如旋转、平移、放缩等所有摄像机的运动形式。投影模型需要8个参数 $(m_i, i=1, 2, \dots, 8)$ 来描述,方程为:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (5)$$

其中 (x, y) 是原始坐标下的值,经过变换后的坐标

用 (x', y') 表示 x' 和 y' 由下式算出:

$$\begin{aligned} x' &= \frac{m_1x + m_2y + m_3}{m_7x + m_8y + 1} = \frac{u}{w} \\ y' &= \frac{m_4x + m_5y + m_6}{m_7x + m_8y + 1} = \frac{v}{w} \end{aligned} \quad (6)$$

双线性模型与投影模型类似,也是需要8个参数 $(q_i, i=1, 2, \dots, 8)$ 来描述摄像机的运动。方程为:

$$\begin{aligned} x' &= q_1xy + q_2x + q_3y + q_4 \\ y' &= q_5xy + q_6x + q_7y + q_8 \end{aligned} \quad (7)$$

由以上两个模型可以得知通过4个对应点的计算,能够得到相应的8个参数,继而利用这些参数可以计算出运动物体由摄像机附加的运动向量,从计算的运动向量中减去这部分附加分量就可以得到运动目标实际的运动。

3. 目标检测

为了能近似实时地处理图像序列,从中提取目标并进行进一步的跟踪,要求我们的算法相比其他算法更加快速有效。我们提出的算法主要分三个步骤来进行,首先是对图像的先期处理,主要是图像的去噪和滤波,这里就使用了前部分我们介绍的数学形态学方法;其次是对先期处理过的图像进行目标提取;最后是对每一帧的目标进行参数计算实现目标的跟踪。在这一部分我们将介绍我们的目标提取算法包括图像的先期处理,在下一部分我们会介绍目标的跟踪算法。

图像序列中的目标提取基本上都要利用目标的运动信息,由于我们所关注的环境对图像质量的影响比较大,图像的信噪比不高,直接运用光流或者运动预测的方法会造成较复杂的误差分布不利于进一步的图像分割,所以先期的滤波处理是有必要的。本算法采用的是前面提到数学形态学的方法,利用开闭运算能够填充物体区域的一些错判的小孔,能够消除背景区域上散布的一些小噪声物体的性质,我们对每一帧图像进行先期的滤波。这种处理会使噪声为正常图像像素所覆盖,以利于进一步的运动预测和分割。要消除摄像机运动的影响,可以通过对应点的参数计算来完成。由于一般目标物体都在图像的中间部位,所以对应点的选取可以从四个角来得到,通过4组对应点的选取由式(6)或者(7),可以得到摄像机运动的各种参数,继而可以消除摄像机运动给图像序列带来的影响,完成图像的配准(这里假设我们的摄像机运动速度不快,相邻两帧图像的背景变化不大)。进行完配准的图像背景是静止的,目标物体是运动的,这样我们就完成了图像的预处理,得到了有利于进行目标提取的图像序列。

由于去除了摄像机运动的影响,图像序列中的运动特征变得更为明显,我们可以根据目标的运动信息将其从背景中分离出来。典型的运动目标的提

取方法是使用光流方程(式(8)),

$$v_x E_x + v_y E_y + E_t = 0 \quad (8)$$

式中 v_x, v_y 分别为沿着 x, y 方向上的运动速度, E_x, E_y, E_t 分别为亮度值在 x, y, t 上的梯度分量。通过计算光流场虽然可以得到运动物体的位置,不过这样的计算需要对图像中的每个点进行,而且得到的运动边界也并不很清晰。我们采用运动预测+边缘检测的方法希望达到好的分割效果并节约运算时间。

由于经过配准的图像序列相邻两帧间有一个大致相同的静止的背景,而目标物体是运动的,所以进行时域的差分能很容易地抑制背景,并得到两帧间的运动区域,这样就大大缩小了我们关注领域。在两帧间的运动区域里,我们进行基于块的运动预测,我们使用 4×4 的块,并把搜索区域基于运动区域进行适当的扩充,根据最小均方误差准则(MSE)进行搜索。预测误差的计算公式如式(9)所示。

$$MSE(d_1, d_2) = \frac{1}{N^2} \sum_{n_1, n_2 \in B} [I_0(n_1, n_2, t) - I_1(n_1 + d_1, n_2 + d_2, t + 1)]^2 \quad (9)$$

其中 $N=4, B$ 代表 4×4 的块 n_1, n_2 为块中的元素, I_0, I_1 为相邻两帧。完成运动预测之后不仅可以进一步地确定运动目标的位置,而且也能在一定程度上确定两帧间目标区域的对应关系。

经过运动预测所得到的运动目标还是不能令人满意的,因为它有明显的块效应,这也是基于块的运动预测在提供快速检索的同时带来的负效应。更精确的目标提取是通过边缘检测得到的,由于目标图像已经被缩到比较小的范围而且噪声已经经过了过滤,这时候采用canny检测的方法不会有太敏感的噪声的影响,而且能够得到精确的目标图像的边缘。进一步进行边缘图像外边界的提取就能够完成对目标的检测和分割。

4. 目标跟踪

我们采用基于参数和多特征融合的方法进行目标的跟踪。对于每一个提取出来的物体,我们通过计算目标的运动轨迹、大小、纹理和不变距来描述目标物体,通过计算各目标物体之间的距离实现对目标的跟踪。

首先要计算的是目标对象的质心坐标,这个坐标可以确定物体的位置,继而通过它计算物体的运动轨迹和不变距。质心坐标 (c_x, c_y) 的计算方法如式(10)所示。

$$\begin{aligned} c_x &= \left(\sum_{(i,j) \in O} p_{i,j} \cdot i \right) / \left(\sum_{(i,j) \in O} p_{i,j} \right) \\ c_y &= \left(\sum_{(i,j) \in O} p_{i,j} \cdot j \right) / \left(\sum_{(i,j) \in O} p_{i,j} \right) \end{aligned} \quad (10)$$

式中 O 表示目标图像区域, $p_{i,j}$ 为坐标 (i, j) 上的灰

度值。质心可以确定目标的位置,通过前面得到的运动向量可以对下一帧的物体位置做预测,预测方法可以基于刚体的运动准则 $S = v_0 t + \frac{1}{2} a t^2$,其中 S 为移动距离, v_0 为物体的初始速度, a 为目标的加速度。因为两帧之间的变化是个差分关系,原式变为 $\Delta S = v_0 + a t$ 。

目标区域面积的变化可以由式(11)计算,这里我们是使用了一个散布变量来描述,通过归一化可以消除亮度变化的影响。

$$disp = \frac{(\sum_{(i,j) \in O} \sqrt{(i-c_x)^2 + (j-c_y)^2} \cdot p_{i,j})}{(\sum_{(i,j) \in O} p_{i,j})} \quad (11)$$

我们可以用一个特征向量来描述图像的纹理,这个特征向量是基于图像的灰度关系,借用图像检索的8像素法^[10]来产生。8像素法用一个字节描述了一个像素同其周围像素的灰度关系,再通过计算图像的直方图生成特征矢量 T 。图2给出了8像素法的像素模型,其中 x 为被处理的像素。

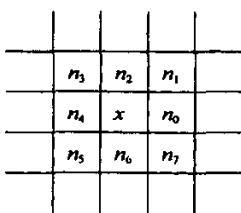


图2 8像素模型

不变距因为与物体放大、平移、主轴的旋转无关,因此在目标的识别和跟踪方面有着很重要的作用。不变距是以中心距的计算为基础的,中心距的计算由式(12)定义,主轴的旋转角由式(13)计算。

$$M_{jk} = \sum_{(i,j) \in O} (i-c_x)^j (j-c_y)^k \cdot p_{i,j} \quad (12)$$

$$\tan 2\theta = \frac{2M_{11}}{M_{20} - M_{02}} \quad (13)$$

在计算中心距前,对主轴旋转相应的角度,得到的中心距具有旋转不变性,再用面积规范后,中心距成为与物体运动无关的不变量。由于三阶或更高阶的距经过规范化后不能保持不变性,因此我们使用规范后的2阶距 M_{11} 来作为目标的不变距。

我们的目标跟踪策略就是基于对象的这几个参数来制定的。由目标的质心计算的对象的运动轨迹充分体现了目标刚体的运动特性;目标的面积参数主要体现的是目标距摄像机的距离和角度,因为目标本身的大小是不会有剧烈变化的,该参数的变化主要是成像引起的;目标的纹理向量能体现目标区域的纹理也就是像素之间的大小关系,这为成像面积差不多的目标之间的选择提供了区分的依据;不变距在目标跟踪方面最有用,因为它不仅反映了物

体的形状还反映了其内部的密度分布,而且在物体放大、平移、旋转时保持不变。这4个参数,质心和面积参数主要体现物体的运动情况,纹理向量和不变距则体现的是目标物体本身的一些属性。我们使用前两个参数考察从下一帧图像中提取的目标是否符合运动估计的结果,使用后两个参数考察两帧间目标图像本身的差别。

由于每帧图像可能被提取出多个目标物体,如果在实现物体的跟踪时每两个目标之间一一比对的话,目标的匹配就需要较长时间而且计算上的花销也比较大。我们在实际中希望目标跟踪能近似实时最好,所以必须减少匹配上的时间,我们使用一些方法和规则来达到这个目的,简化目标间的比较。由于目标参数在计算时大多使用了目标质心的坐标,而我们在进行目标提取时进行了运动估计得到了目标对象的运动向量,所以我们首先根据目标的运动方向减少要比对的目标数量,那些在目标运动方向相反的对象将不被比较。接下来我们根据运动向量计算下一帧中目标物体的质心的预期位置,然后和下一帧图像中要比较的目标的质心位置进行比较,当两质心的距离大于一个阈值时,则进一步舍弃那些目标。进一步的匹配通过其他3个参数的比较一一得出,对于每个参数比较都用一个阈值来限定,在使用一个整体阈值做全局门限,任何一个参数的差值大于这个全局门限都可以舍弃相应的比较目标。如果一一比较完还有多个目标作为待定目标则计算距离和(式14),距离最小的为要找的目标。

$$dif = \omega_c \sqrt{(c_x - c'_x)^2 + (c_y - c'_y)^2} + \omega_{disp} |disp - disp'| + \omega_T \|T_3 - T'_3\| + \omega_M |M - M'| \quad (14)$$

对于运动情况不复杂的情况,比如公路上对汽车的监控,由于目标运动变化不会太大,所以通过适当的缩小阈值可以大大减少比较目标的数量,而对于目标速度变化比较快的情况可以通过加大运动参数比较的阈值,主要通过纹理向量和不变距来进行目标比较。

5. 实验结果和分析

根据上面我们提供的算法,对一些道路监控的图像序列进行了试验,图3是序列中的几帧图像。

这个图像序列是一辆汽车迎面开来的图像,图像背景比较复杂,信噪比不高,目标个数不多。从图像的变化可以看出汽车的形体变化比较明显,背景只有细微的扰动,由于汽车的运动方向,物体的运动信息和光流并不明显。

应用我们的算法,首先进行运动预测得到大致的目标范围,再进行进一步的目标检测得到目标图像图4,相邻帧的图像被提取出来以后通过计算参数进行目标跟踪,由于我们的算法中使用了面积参数,

使得这个序列的跟踪变得比较简单,如图4所示。从实验结果可以看出,本检测和跟踪算法能有效地提

取和跟踪自然环境、低信噪比下的目标。

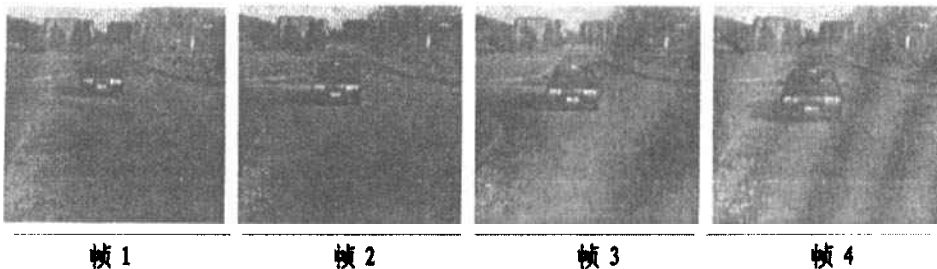


图3 测试图像序列



图4 提取和跟踪出的目标

由于我们的跟踪算法采用多级审核较早地排除了一些不可能的目标,节省了很多时间,使得跟踪更加快速有效。本算法通用适用于新目标的产生,原有目标的消失等情况。对于目标的遮挡和再出现,由于我们在计算时采用的相邻帧计算法,相当于一个不断改变的模板,使得遮挡对于目标图像的变化变得轻微,能部分地解决这种问题。

结论 本文开始对于各种目标检测和跟踪的应用进行了分类,按情况的不同提出了解决不同问题的研究重点,我们认为图像处理的价值就在于处理实际问题,由于自然世界的图像信息极为丰富、情况复杂多变,只有选定相关的环境才能更好地设计、应用相应的图像处理算法。基于此,我们对自然环境下低信噪比的情况的目标检测和跟踪进行了研究,并提出了相应的算法,从实验结果看算法是快速有效的,而且能适用于比较复杂的情况。目标检测和跟踪算法的核心就在于快速、准确。我们也将从这两个方面考察改进我们的算法

参考文献

1 MPEG-4 visual fixed draft international standard, ISO/IEC

14496-2, Oct. 1998

- 2 Meier T, Ngan K N. Automatic Segmentation of Moving Objects for Video Object Plane Generation. *IEEE Trans. On Circuits and Systems for Video Technology*, 1998, 8(5)
- 3 Meier T, Ngan K N. Segmentation and tracking of moving objects for content-based video coding. *IEE Proc. Visual Image Signal Processing*, 1999, 146 (3): 144~150
- 4 Wang Demin. Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking. *IEEE Trans. Circuits and Systems for Video Technology*, 1998, 8(5)
- 5 Deng Yining, Manjunath B S. Unsupervised Segmentation of Color-Texture Regions in Images and Video. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2001, 23(8)
- 6 Smith S M, Brady J M. ASSET-2: Real-Time Motion Segmentation and Shape Tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1995, 17(8)
- 7 Wang Y, Doherty J F, Van Dyck R E. Moving Object Tracking in Video. 0-7695-0978-9/00, 2000 IEEE
- 8 Cohen I, Medioni G. Detecting and Tracking Moving Objects for Video Surveillance. 0-7695-0149-4/99, 1999 IEEE
- 9 Serge BEUCHER, Recent Advance in Mathematical Morphology
- 10 Berman A, Shapiro L. A flexible image database system for content-based retrieval. *Computer Vision and Image Understanding*, 1999, 75(1/2): 175~195