

动态场景图像序列中运动目标检测新方法

郑世友^{1,2)} 费树岷²⁾ 刘怀³⁾ 龙飞²⁾

¹⁾(中航雷达与电子设备研究院第六研究室, 无锡 214063) ²⁾(东南大学自动化研究所, 南京 210096)

³⁾(南京师范大学电气与自动化工程学院, 南京 210042)

摘要 在动态场景图像序列中检测运动目标时,如何消除因摄影机运动带来的图像帧间全局运动的影响,以便分割图像中的静止背景和运动物体,是一个必须解决的难题。针对复杂背景下动态场景图像序列的特性,给出了一种新的基于场景图像参考点 3D 位置恢复的图像背景判别方法和运动目标检测方法。首先,介绍了图像序列的层次化运动模型以及基于它的运动分割方法;然后,利用估计出的投影矩阵计算序列图像中各运动层的参考点 3D 位置,根据同一景物在不同帧中参考点 3D 位置恢复值的变化特性,来判别静止背景对应的运动层和运动目标对应的运动层,从而分割出图像中的静止背景和运动目标;最后,给出了动态场景图像序列中运动目标检测的详细算法。实验结果表明,新算法较好地解决了在具有多组帧间全局运动参数的动态场景序列图像中检测运动目标的问题,较大地提高了运动目标跟踪算法的有效性和鲁棒性。

关键词 运动分割 运动目标检测 混合模型 仿射运动模型 动态场景

中图分类号: TN911.73 **文献标识码**: A **文章编号**: 1006-8961(2007)09-1590-08

A Novel Approach to Moving Object Detection in Image Sequence Acquired by a Mobile Camera

ZHENG Shi-you^{1,2)}, FEI Shu-min²⁾, LIU Huai³⁾, LONG Fei²⁾

¹⁾(The Sixth Laboratory, Radar and Avionics Institute of AVIC, Wuxi 214063)

²⁾(Research Institute of Automation, Southeast University, Nanjing 210096)

³⁾(School of Electronic & Automatic Engineering, Nanjing Normal University, Nanjing 210042)

Abstract Moving object detection is a very significant and difficult problem in processing image sequence acquired by a mobile camera. The objective of this research is to present a novel method to resolve the moving target detection difficulties, which bring by several layers with different motion parameters of the sequence images. Considering a moving camera, it was hard to distinguish different motion layers led by 3D depth difference of immobile objects from those caused by target moving in scene. Firstly, a mixture model of image was proposed, and the method of motion segmentation based this model were suggested. Secondly, the reconstruction method of scene depth, which employed the motion parameters by EM, was presented. Finally, the detail theory and algorithm of new moving target detection approach, which were based on the scene 3D depth reconstruction, were described. The results indicate that the moving targets detected by the new technique are more accurate than those detected by traditional methods. In addition, the detection speed becomes more-much. The results seem to suggest that the new method be able to provide efficiency and accuracy advantages in computer version.

Keywords motion segmentation, moving object detection, mixture model, affine motion model, dynamic scene

1 引言

图像序列中运动目标检测与跟踪是当前计算机

视觉研究的一个基础性课题^[1-3],在现代工业、国防和空间技术等领域有着广阔的应用前景。通常将实际的图像序列分为两类:静态场景图像序列和动态场景图像序列。没有全局运动(指因摄像机运动引

收稿日期:2006-03-13; 改回日期:2006-07-13

第一作者简介:郑世友(1972~),男,工程师,2006年于南京东南大学获控制理论与控制工程专业工学博士学位。现研究方向为小波分析、图像分析和理解、计算机视觉等。已在国内外核心期刊和会议上发表论文 10 多篇。E-mail: seebank1972@yahoo.com.cn

起的背景图像运动)的图像序列称为静态场景图像序列,而具有全局运动的图像序列称为动态场景图像序列。在静态场景图像序列中检测和跟踪运动目标相对比较简单,经过多年的研究渐趋成熟。而动态场景图像序列中运动目标检测和跟踪至今仍然存在很多关键问题尚未解决,特别是当背景较复杂时,问题变得更加困难。因此针对动态场景图像序列的研究正成为当前应用视觉领域的热点^[4,5]。

现有动态场景图像序列中运动目标检测方法主要包括全局运动补偿和运动目标分割两个步骤,其中全局运动补偿处于基础性地位。各国学者对此展开研究并提出了大量的方法^[6-9],获得了一定的成功。这类检测方法的基本思想是:首先估计出相邻帧间的全局运动参数;然后运用这些参数补偿全局运动;最后利用差分法完成运动目标检测。但是现有的基于运动补偿的方法^[6-9]大多基于这样一个假设:背景景物的深度差异对图像序列帧间全局运动的影响可以忽略。也就是说,所有的背景景物在图像序列中都假设共享同一组全局运动参数。当背景较简单时,多数图像序列可以满足这个假设;但是当背景较复杂时,位于不同3维深度的背景物体,事实上会在由运动摄影机获得的图像序列中产生多组不同的,但是组内部一致的“运动场”^[10],此时上述假设显然不再满足。虽然一些学者提出一些基于层次化运动模型^[11-15]的运动分割方法,可以把具有多组全局运动参数的图像序列的每帧分割成不同的运动层(每个运动层具有相同的运动参数);Irani等人^[16]进一步针对场景中有运动物体的动态场景图像序列展开研究,提出了一种逐次估计每组显著的运动参数后分别补偿的方法,并给出了判断运动物体层和静态背景层的定性规则。然而,这些方法对于如何判断运动分割后的图像中哪些运动层对应静止背景,哪些运动层对应运动物体还仅限于定性分析,而没有给出定量方法,从而造成运动目标检测实现时的困难。总的说来,对在场景中具有多个运动物体的动态场景图像序列中如何比较精确的分割运动目标,迄今为止在现有文献中仍未见到有效的定量方法。

本文针对此问题展开研究,在层次化运动模型的基础上,给出了一种基于场景图像参考点3维位置恢复的图像背景判断方法,直接分割出图像静态背景和运动目标,从而较好地解决了在具有多组帧间全局运动参数的图像序列中检测运动目标的

问题。

2 层次化运动模型和运动分割

动态场景图像序列通常具有如下特点:(1)图像中的每一点,存在一个联系该像素和下一图像中像素的运动向量;(2)存在参数不同的运动场集合;(3)总的运动由多个运动场共同生成。显然这些运动场主要有两种成因:(1)仅由摄影机运动造成;(2)由运动物体自身运动和摄影机运动共同造成。如果能够确定出序列图像帧中各个运动场的成因,则就可以比较容易地从其中检测出运动目标。也就是说,由运动目标自身运动和摄影机运动共同造成的运动场所在的区域即是运动目标所在的区域;其他区域则对应背景。因此按像素点运动参数不同把序列图像帧分割成一个个运动层(每层具有相同运动参数)就成为必须解决的首要问题。本文使用文献[10]给出的基于层次化运动模型的运动分割方法来分割图像运动场,以及估计各个场相应的运动参数。下面给出与层次化运动模型相关的一些基本概念:

定义1 图像混合模型^[10,16-18]是指将图像看成一种生成模型。对于图像中的每个像素,计算一个 d 维的特征向量 W (W 可以包含位置、速度、灰度以及纹理信息等),并假设图像由 g 个子模型共同生成,其中第 l 个子模型的概率密度为 $p_l(W|\theta_l)$,其参数为 θ_l ;且选择第 l 个子模型的概率为 π_l 。则生成一个像素向量的概率可表示为

$$p(W|\Theta) = \sum_{l=1}^g p_l(W|\theta_l)\pi_l; \sum_{l=1}^g \pi_l = 1 \quad (1)$$

将这种形式的模型称为混合模型(由于它是一个概率模型的加权和或混合,故也称 π_l 为混合权重)^[10,17-19]。其中参数向量为 $\Theta = (\pi_1, \pi_2, \dots, \pi_g, \theta_1, \theta_2, \dots, \theta_g)$ 。

定义2 层次化运动模型^[10]是利用图像混合模型描述的图像序列的运动场模型。其中子模型是图像的各个运动场。它概括了多组不同的运动场。把其中每一个运动场称为一个运动层,且共享同一组运动参数。

为简单起见本文选用如下六参数仿射运动模型^[10]来描述每个运动层的运动:

假设第 k 帧图像 I_k 中的某像素点 $X = [x, y]^T$ 属于第 l 个运动层,且该像素点在图像 I_{k+1} 中的对应位置为 $X' = [x', y']^T$,则它们之间满足下面的

关系:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} m_1(x, y; \theta_{i,k}) \\ m_2(x, y; \theta_{i,k}) \end{bmatrix} = \begin{bmatrix} a_{11,i,k} & a_{12,i,k} \\ a_{21,i,k} & a_{22,i,k} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_{13,i,k} \\ a_{23,i,k} \end{bmatrix} \quad (2)$$

这个模型称为六参数仿射运动模型,其参数向量为 $\theta_{i,k} = (a_{11,i,k}, a_{12,i,k}, \dots, a_{23,i,k})$ 。

层次化运动模型比较适合研究针对不同运动场分割图像的问题。给定一对由层次化运动模型描述的图像,希望确定以下几点:(1)像素属于哪个运动层;(2)每个运动层的运动参数。这是一个丢失数据问题:丢失的数据是像素所归属的运动场,而参数是表示每个场的参数和混合权重。可以用指示向量 $V_{xy,i,k}$ 来表示这个含义,

$$V_{xy,i,k} = \begin{cases} 1 & \text{如果像素}(x, y)\text{属于第}i\text{个运动场} \\ 0 & \text{其他} \end{cases} \quad (3)$$

假设在获取的灰度图像中,噪声为均值 0,方差为 σ^2 的高斯噪声,且 $I_k(x, y)$ 在 I_{k+1} 中的对应点为 $I_{k+1}(x', y')$,则 $I_k(x, y) - I_{k+1}(m_1(x, y; \theta_{i,k}), m_2(x, y; \theta_{i,k}))$ 即为图像噪声的表现,即服从 $N(0, \sigma^2)$,其中从前帧图像中属于第 l 个运动层中像素点 (x, y) 运动到后帧图像像素点 (x', y') 的参数向量为 $\theta_{l,k}$ 。因此完备的数据对数似然函数为

$$L(V, \theta) = - \sum_{xy,i} V_{xy,i,k} \frac{(I_k(x, y) - I_{k+1}(m_1(x, y; \theta_{i,k}), m_2(x, y; \theta_{i,k})))^2}{2\sigma^2} + C \quad (4)$$

关键的问题是确定

$$P\{V_{xy,i,k} = 1 \mid I_k, I_{k+1}, \theta\} \quad (5)$$

也就是确定在已知观测值 I_k, I_{k+1} 和参数 $\theta = (\theta_{1,k}, \dots, \theta_{g,k})$ 的条件下 $V_{xy,i,k} = 1$ 的条件概率,其中 θ 是一个不能直接观测的参数。对于如何求解这个丢失数据问题以及参数 θ ,文献[10]给出了一种基于期望最大化 EM(expectation-maximization)法的运动分割方法。该方法将遵循同一仿射运动模型的像素点聚类到一起形成一个运动层,并且估计出相应层的运动参数(详细的分割方法参见文献[10])。本文正是使用此方法进行运动分割,并在分割图像中对每个像素赋予一个灰度级,灰度级的数值表示了该像素点最可能属于的运动层。

3 场景图像参考点 3 维位置恢复和目标检测

由于运动物体对序列图像的影响主要表现在运

动物体对应的运动层由摄影机运动和它自身运动共同造成,而这点反映在图像序列中时,除了会对它对应的运动层参数带来较大影响外,其他和静止物体并无不同。因此在运动分割时先将运动物体当成一种特殊的静止物体,然后运用第 2 部分的分割方法将图像序列的每帧分割成一系列的运动层,并且估计出每个运动层的运动参数。

显然,上面分割出的运动层有的对应背景,有的对应运动目标。如何判断背景层和运动物体层,从而检测出目标,现有文献中尚未见到有效的定量方法。下面介绍一种新的基于场景图像参考点 3 维位置恢复的判断方法。该方法是基于这样的性质:所有静止背景物体在摄影机运动的同一时刻均遵循相同的投影矩阵,而运动物体则不受此投影矩阵的约束。

所谓场景图像参考点 3 维位置恢复是指从二帧图像中的 N 对有代表性的对应点恢复摄影机的投影矩阵,以及恢复这些参考点在相应 3 维坐标系中的 3 维位置。在摄像机运动中的两个不同时刻观察场景,会得到场景的两幅不同视角的视图,利用它们可以恢复出场景图像参考点的 3 维位置。

3.1 图像运动层配对

依据现有的投影结构恢复理论,在恢复场景图像参考点 3 维位置时,需要先进行对应点配对。由于已经估计出了图像运动层的参数,因此不用逐点配对,而是以图像运动层为匹配单位进行配对,即将场景中同一物体在两帧图像中所处的运动层配对起来。

假设动态场景图像序列中第 k 帧图像 I_k 被分割为 g_k 个运动层;并将分割后的图像记为 I'_k ,将 I'_k 中第 $l(l=1, 2, \dots, g_k)$ 个运动层记为 $L_{l,k}$,参数向量为 $\theta_{l,k}$ 。下面以在 I'_k 中寻找 $L_{l,k-1}$ 的配对运动层为例,介绍运动层配对方法:

(1) 利用式(2)对运动层 $L_{l,k-1}$ 进行变换和插值补偿后(补偿时使用估计好的参数向量 $\theta_{l,k-1}$),获得其在 I_k 中的预测区域 $\hat{L}_{l,k}$ 。

(2) 在 I_k 中除将 $\hat{L}_{l,k}$ 内部对应的像素灰度设置为“1”外(将 $\hat{L}_{l,k}$ 中超出图像 I_k 的部分直接舍弃),其他像素灰度均设置为“0”;这样处理后得到新的二值图像 $\hat{I}_{l,k}$ 。

(3) 求取 $\tilde{I}_{l,k} = \hat{I}_{l,k} \times I'_k$,其中“ \times ”运算如下定义:图像 $\tilde{I}_{l,k}$ 中像素点的灰度值为 $\hat{I}_{l,k}$ 和 I'_k 中对应像素灰度值代数相乘。显然, $\tilde{I}_{l,k}$ 中灰度非零的像素

形成的区域可能包含 $m(m \leq g_k)$ 个运动层的像素。

(4) 记 $\tilde{I}_{i,k}$ 中灰度非零像素点位于第 $n(n=1, 2, \dots, g_k)$ 个运动层 $L_{n,k}$ 中的像素个数为 s_n , 而 $L_{n,k}$ 的像素总数为 t_n , 计算下式

$$j = \arg\left(\max_{n=1}^{g_k} \left\{ \frac{s_n}{t_n} \right\}\right) \quad (6)$$

如果

$$s_j/t_j \geq T_1 \quad (7)$$

则这 I_i 中第 j 个运动层 $L_{j,k}(j \in \{1, 2, \dots, g_k\})$ 即是 $L_{i,k-1}$ 的配对运动层, 记这个配对的 $\Gamma_{(i,j)}^{k-1,k} = \{L_{i,k-1}, L_{j,k}\}$ 。否则认为 $L_{i,k-1}$ 在 I_i 中没有配对运动层。其中 T_1 为预先确定的阈值。

将 I_{i-1} 中的所有运动层进行配对处理将形成多组运动层对。在此基础上就可以恢复出场景的3维位置。为了简化问题, 而专注于新算法的思想, 本文针对运动层对给出下面几个限定假设。

假设1 相邻图像帧的一个匹配运动层对对应3维场景中相同的景物部分。即相邻图像帧中最相似的运动层对应相同的景物, 这显然是合理的。

假设2 相邻图像帧的一个匹配运动层对如果对应静止的3维景物, 则二者具有最相近的景物位置。在图像序列中静止景物的位置变化是由摄影机运动引起的。如果摄影机没有剧烈运动, 则对于相邻两帧来说这个变化可以忽略。因此要满足这个假设要求摄影机没有剧烈运动。

假设3 图像中所有运动目标所在运动层的总面积远小于背景所在运动层的总面积。本文暂不考虑大运动物体(或多运动物体几乎遍布整个图像)的情况。

3.2 基于场景参考点3维位置恢复的背景判断方法

将相邻两帧图像看成由双目摄像机对同一景物观察获得的两幅图像, 可以很容易地估计出每个运动层中对对应景物点的3维位置。由于所有静止物体在摄影机运动的相同时刻均遵循相同的投影矩阵, 而运动物体则不受此投影矩阵的约束。因此如果用每帧图像静止背景对应的投影矩阵去恢复满足假设1~3的序列图像帧中参考点3维位置值, 则恢复出的每个图像背景点对应的景物3维位置值在相邻帧间将没有显著变化, 而图像中运动物体点对应的3维位置值将表现出明显变化。利用这个特点就可以判断背景区域和运动物体区域。

本节先介绍如何从同一场景的两幅图像中恢复

指定点3维位置的基本理论和方法, 然后再给出新的背景区域判断方法。在3维重构中, 如果摄影机是标定过的, 它们的内参数和外参数都已知, 则由外参数就可以得到图像点对应的景物点在世界坐标系的位置。然而在实际应用中常常面对的却是较复杂的情形: 摄影机的位置甚至是内参数都是事先不知道的, 而且会随着时间变化。特别是在图像序列的研究中, 拥有的只是一帧帧图像, 缺乏摄影机有关的直接信息。显然, 由于标定过程本身的难度以及其内在计算复杂性, 随时标定这些参数以获取投影矩阵参数并不是明智的选择。如果能够从图像序列本身直接恢复出各自的投影矩阵和指定点的3维位置信息, 则可能使问题得到简化。

给定 n 个点 $P_j(j=1, 2, \dots, n)$ 和它们在 m (这里 $m=2$) 个摄影机观察下的图像齐次坐标 $p_{ij} = (x_{ij}, y_{ij}, 1)^T$, 则透视投影方程如下:

$$\begin{cases} x_{ij} = \frac{\varphi_{1i} \cdot P_j}{\varphi_{3i} \cdot P_j} \\ y_{ij} = \frac{\varphi_{2i} \cdot P_j}{\varphi_{3i} \cdot P_j} \end{cases} \quad (8)$$

其中, $\varphi_{1i}, \varphi_{2i}$ 和 φ_{3i} 表示 3×4 投影矩阵 Φ_i 的各行转置后形成的列向量。 Φ_i 是第 i 个摄影机在某个坐标系下的投影矩阵, P_j 是点在同一坐标系下的齐次坐标 $P_j = (X_j, Y_j, Z_j, 1)^T$ 。

如果 Φ_i 和 P_j 是式(8)的解, 对任意非零的 λ_i 和 $\mu_j, \lambda_i \Phi_i$ 和 $\mu_j P_j$ 也是解。因此满足方程的 Φ_i 和 P_j 只定义到比例层次, 其中 Φ_i 有11个独立参数, P_j 有3个独立参数。同样若如果 Φ_i 和 P_j 是式(8)的解, 则 $\Phi_i' = \Phi_i Q$ 和 $P_j' = Q^{-1} P_j$ 也是解, 其中 Q 是一个投影变换矩阵(任意的非奇异 4×4 矩阵)。矩阵 Q 是定义在比例上的, 有15个自由参数。因为式(8)对 $11m$ 个矩阵参数和 $3n$ 个向量 P_j 参数共有 $2mn$ 个约束, 再加上问题本身的不确定性, 所以要使问题有有限数量的解, 需要满足 $2mn \geq 11m + 3n - 15$ 。故 $m=2$ 时, 7对对应点就可以确定两个投影矩阵和其他点的位置。

但是直接恢复方法计算量比较大, 结合外极线约束可以找到更高效的投影矩阵和点3维位置的恢复方法, Faugeras 和 Hartley^[10] 对此进行了原创性研究, 于1992年分别独立提出了用两个未标定摄影机获得的同一场景两幅图像恢复场景投影结构的方法。Ponce^[10] 也于1993年从新的角度提出了一种几何方法。这些方法在假设两幅图像中的 n 个投影

点成功匹配的前提下,能够从这些图像匹配点估计出图像中任一点的 3 维空间位置坐标,以及每幅图像(每个摄影机)对应的投影矩阵。

基于上述基础理论,下面建立详细的基于场景图像参考点 3 维位置恢复的背景区域判断方法:

(1) 在 $I_{i,k-1}$ 和 I_k 中选择匹配运动层对 $\Gamma_{(i,j)}^{k-1,k} = \{L_{i,k-1}, L_{j,k}\}$,然后在 I_k 和 $I_{i,k+1}$ 中选择对应的匹配运动层对 $\Gamma_{(j,d)}^{k,k+1} = \{L_{j,k}, L_{d,k+1}\}$ 。

(2) 在 $\Gamma_{(i,j)}^{k-1,k}$ 和 $\Gamma_{(j,d)}^{k,k+1}$ 中选择参考点对。选取方法如下:对 $L_{i,k-1}$ 中的某个像素点 $q_{e,i,k-1}$,利用估计出的运动参数 $\theta_{i,k-1}$,用式(2)可以很容易计算出它在 I_k 中的对应点 $q_{f,j,k}$ 。如果 $q_{f,j,k} \in L_{j,k}$,则认为 $q_{e,i,k-1}$ 和 $q_{f,j,k}$ 为一对配对点;否则认为 $q_{e,i,k-1}$ 没有合适的配对点,在 $L_{i,k-1}$ 中选择新的点重新组对。对选好的 $q_{f,j,k}$ 用同样的方法求取其在 $L_{d,k+1}$ 中的匹配点。直到在 $\Gamma_{(i,j)}^{k-1,k}$ 和 $\Gamma_{(j,d)}^{k,k+1}$ 中都选择好了 $N(N \geq 7)$ 对配对点为止。注意:选择 $L_{i,k-1}$ 中的参考像素点时需要保证它在匹配层 $L_{j,k}$ 和 $L_{d,k+1}$ 中都有配对点;并且尽量避免参考点共线。

(3) 利用 $\Gamma_{(i,j)}^{k-1,k}$ 中的 N 对参考匹配点,使用传统双目估计方法^[18]估计出针对 $L_{i,k-1}$ 的投影矩阵 $\Phi_{i,k-1}^{k-1,k}$ 和针对 $L_{j,k}$ 的投影矩阵 $\Phi_{j,k}^{k-1,k}$,以及 N 对点对应的 3 维位置坐标。估计时使用如下透视投影模型:

给定运动层 $L_{i,k-1}$ 的 N 点中的任一点 $q_{e,i,k-1}$ ($q_{e,i,k-1} \in L_{i,k-1}, e = 1, 2, \dots, N$),其齐次坐标为 $[x_{q_{e,i,k-1}}, y_{q_{e,i,k-1}}, 1]$ 以及它对应的 3 维景物坐标 $P_{q_{e,i,k-1}} = [X_{q_{e,i,k-1}}, Y_{q_{e,i,k-1}}, Z_{q_{e,i,k-1}}, 1]^T$,则可写出如下透视投影模型:

$$\begin{cases} x_{q_{e,i,k-1}} = \frac{\varphi_{1,i,k-1} \cdot P_{q_{e,i,k-1}}}{\varphi_{3,i,k-1} \cdot P_{q_{e,i,k-1}}} \\ y_{q_{e,i,k-1}} = \frac{\varphi_{2,i,k-1} \cdot P_{q_{e,i,k-1}}}{\varphi_{3,i,k-1} \cdot P_{q_{e,i,k-1}}} \end{cases} \quad (9)$$

其中, $\varphi_{1,i,k-1}, \varphi_{2,i,k-1}$ 和 $\varphi_{3,i,k-1}$ 表示 3×4 投影矩阵 $\Phi_{i,k-1}^{k-1,k}$ 的各行转置后形成的列向量, $\Phi_{i,k-1}^{k-1,k}$ 是第 $k-1$ 帧时第 i 个区域内像素的投影矩阵。

(4) 计算

$$\begin{cases} X_{i,k-1} = \frac{1}{N} \sum_{e=1}^N X_{q_{e,i,k-1}} \\ Y_{i,k-1} = \frac{1}{N} \sum_{e=1}^N Y_{q_{e,i,k-1}} \\ Z_{i,k-1} = \frac{1}{N} \sum_{e=1}^N Z_{q_{e,i,k-1}} \end{cases} \quad (10)$$

并称 $(X_{i,k-1}, Y_{i,k-1}, Z_{i,k-1})$ 为运动层 $L_{i,k-1}$ 的 3 维恢复位置。

(5) 利用投影矩阵对 $\Phi_{i,k-1}^{k-1,k}$ 和 $\Phi_{j,k}^{k-1,k}$ 来估计 $I_{i,k-1}$ 其他运动层中 N 个参考像素点对应的 3 维位置,并利用式(10)计算出这些运动层的 3 维恢复位置。直到估计出 $I_{i,k-1}$ 中所有运动层在此投影阵约束下的 3 维恢复位置为止。同样,选择 $I_{i,k-1}$ 其他运动层中参考像素点时,也需要保证它在 I_k 和 $I_{i,k+1}$ 中对应的匹配层内都有对应点。

(6) 对 $\Gamma_{(j,d)}^{k,k+1} = \{L_{j,k}, L_{d,k+1}\}$ 用同样的方法估计出投影矩阵对 $\Phi_{j,k}^{k,k+1}$ 和 $\Phi_{d,k+1}^{k,k+1}$,并计算出 I_k 中所有运动层在此投影阵约束下的 3 维恢复位置 $(X_{j,k}, Y_{j,k}, Z_{j,k})$ 。估计时选用估计 $\Phi_{i,k-1}^{k-1,k}$ 和 $\Phi_{j,k}^{k-1,k}$ 时使用过的各层参考像素点或其对应点。

(7) 由于 3 维位置估计常常只定义在比例层次,不能直接比较这些值,故利用下面的方法在比例层次上进行 3 维位置比较:

根据假设 2,如果选择的运动层对 $\Gamma_{(i,j)}^{k-1,k}$ 和 $\Gamma_{(j,d)}^{k,k+1}$ 对应的是静止的背景景物,则在理想情况下,用投影矩阵对 $\Phi_{i,k-1}^{k-1,k}$ 和 $\Phi_{j,k}^{k-1,k}$ 估计出的其他静止背景景物对应的运动层中像素点在 $k-1$ 帧时对应的 3 维位置值,和用投影矩阵对 $\Phi_{j,k}^{k,k+1}$ 和 $\Phi_{d,k+1}^{k,k+1}$ 估计出的相应静止背景景物对应的运动层中匹配像素点在 k 帧时对应的 3 维位置值将保持恒定的比例;如果运动层对对应的是运动物体,则恢复出的 3 维位置值一般不满足这个性质。考虑噪声和计算误差的影响,假设对另一个运动层对 $\Gamma_{(a,b)}^{k-1,k} = \{L_{a,k-1}, L_{b,k}\}$ 来说,如果

$$\max \left\{ \left| \frac{X_{a,k-1}}{X_{a,k}} - \frac{X_{i,k-1}}{X_{j,k}} \right|, \left| \frac{Y_{a,k-1}}{Y_{b,k}} - \frac{Y_{i,k-1}}{Y_{j,k}} \right|, \left| \frac{Z_{a,k-1}}{Z_{b,k}} - \frac{Z_{i,k-1}}{Z_{j,k}} \right| \right\} \leq T, \quad (11)$$

则称 $\Gamma_{(a,b)}^{k-1,k}$ 和 $\Gamma_{(i,j)}^{k-1,k}$ 具有相同的 3 维位置比。其中, T 为预先设定的阈值。计算出所有与 $\Gamma_{(i,j)}^{k-1,k}$ 具有相同 3 维位置比的运动层对,并求取这些运动层的并集 $\cup_i L_{i,k-1}$ 和 $\cup_j L_{j,k}$ 。根据假设 3,只有当

$$\eta_{(i,j)} = \min \left\{ \frac{A(\cup_i L_{i,k-1})}{A(I_{k-1})}, \frac{A(\cup_j L_{j,k})}{A(I_k)} \right\} \geq T. \quad (12)$$

$\cup_i L_{i,k-1}$ 和 $\cup_j L_{j,k}$ 才可能对应 3 维中的静止景物。其中 $A(*)$ 为求面积操作。

(8) 选择其他与 $\Gamma_{(i,j)}^{k-1,k}$ 及 $\Gamma_{(j,d)}^{k,k+1}$ 具有不同 3 维位置比的运动层对,重复步骤(2)~(7),直到计算完所有可选的运动层对。选择

$$(u, v) = \arg(\max_{(u, v)} |\eta_{(u, v)}|) \quad (13)$$

则认为运动层对 $\Gamma_{(u, v)}^{k-1, k}$ 对应的就是实际的3维静止景物。 I_{k-1} 中与 $\Gamma_{(u, v)}^{k-1, k}$ 具有相同3维位置比的所有运动层和 $L_{\mu, k-1}$ 的并集即形成图像 I_{k-1} 的背景区域; 而其他运动层的并集形成图像 I_{k-1} 的运动物体区域。用同样的方法可以确定图像 I_k 的背景区域和运动物体区域。如果找不到符合要求的 (μ, v) , 则新选 I_k 和 I_{k+1} 组成相邻图像帧对后返(1)继续判断, 直到找到符合要求的 (u, v) 为止。

这里存在两个问题:

(1) 如果摄影机没有运动, 则对于相邻图像帧而言它只是单目摄影机, 将无法进行上述静止场景3维位置恢复, 也无需进行3维位置恢复(因为此时没有全局运动)。因此3维位置恢复前, 需要首先判断摄影机是否具有运动。摄影机是否运动的判断和处理方法。

显然在理想情况下, 如果一个运动层对 $\Gamma_{(u, v)}^{k-1, k}$ 的参数为 $\theta_{i, k-1} = \theta_0 = (1, 0, 0, 1, 0, 0)$, 则表示 $L_{i, k-1}$ 和 $L_{j, k}$ 间没有全局运动。考虑噪声和计算误差的影响, 如果 $|\theta_{i, k-1} - \theta_0| \leq T_s$, 则认为运动层 $L_{i, k-1}$ 和 $L_{j, k}$ 间没有全局运动。其中, T_s 为预先设定的阈值。根据假设3可知, 如果第 $k-1$ 帧和 k 帧图像中没有全局运动区域的面积远大于具有全局运动区域的面积, 则第 $k-1$ 帧和第 k 帧之间没有全局运动, 也判定此时摄影机没有运动。当摄影机没有运动时是静态场景图像序列, 很容易使用传统相邻帧差法进行目标检测, 限于篇幅本文不再详细介绍。

(2) 如果摄影机发生剧烈运动, 此时场景发生剧烈变化, 则相邻图像帧差异较大, 可能会造成算法失效, 需要单独处理。

判断 I_k 到 I_{k+1} 是否发生场景切换的判断方法:

(1) 对 I_k 中所有已判断为背景的运动层和 I_{k+1} 中运动层进行配对处理, 且记成功配对的背景运动层并集的面积 $A_{k, k+1}^{bg}$, 如果

$$\frac{A_{k, k+1}^{bg}}{A(I_{k+1})} < T_m \quad (14)$$

则认为发生了场景切换, 其中 T_m 为预先设定的阈值; 否则

(2) 对 I_k 中所有成功配对的背景运动层进行补偿后形成新的图像帧 \hat{I}_{k+1} 。

(3) 检查 I_{k+1} 和 \hat{I}_{k+1} 在背景区域内灰度的差是否超过了一个预先设定的阈值 T_c 。用下面的规则来检测场景切换:

$$\begin{cases} \text{出现场景切换} \\ \frac{1}{N_{bg}|P|} \sum_{P \text{ 为 } I_k \text{ 背景区域内内部像素}} (I_{k+1}(p) - \hat{I}_{k+1}(p))^2 > T_c \\ \text{没有场景切换} \\ \frac{1}{N_{bg}|P|} \sum_{P \text{ 为 } I_k \text{ 背景区域内内部像素}} (I_{k+1}(p) - \hat{I}_{k+1}(p))^2 \leq T_c \end{cases} \quad (15)$$

式中, N_{bg} 为 I_k 中所有背景区域内内部像素个数。如果检测到场景切换, 则分割算法需要丢掉前面帧的计算信息, 全部重新开始计算。

3.3 静止背景去除和运动目标分割方法

检测出背景区域后就可以很方便的将它们去除, 从而只留下运动物体所在的区域。方法是: 将 I_{k-1} 和 I_k 中背景区域对应的像素灰度设置为“0”, 而保留它们运动物体区域对应的像素灰度值。

3.4 动态场景图像序列的目标检测

运动目标检测的详细算法(流程图如图1所示):

- (1) 获取给定图像序列的相邻3帧 I_{k-1} 、 I_k 和 I_{k+1} ;
- (2) 对相邻两帧分别进行运动分割, 并将分割后图像的运动层进行配对处理;

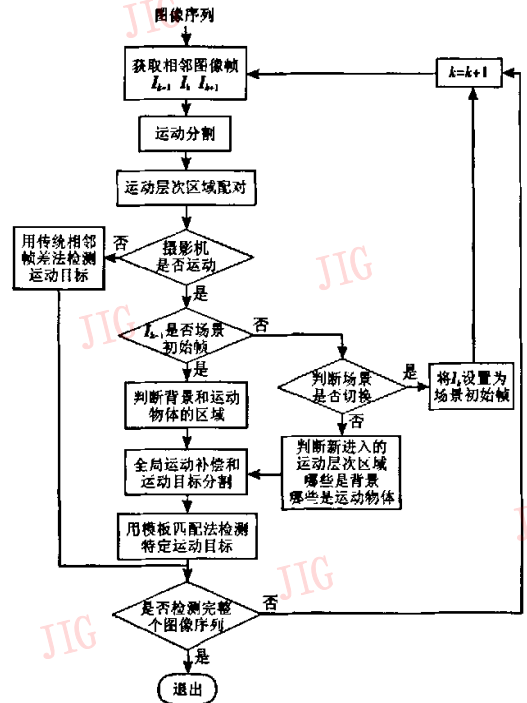


图1 动态场景运动目标检测流程图

Fig. 1 The flow chart of moving object detection in image sequence acquired by a mobile camera

(3) 判断摄影机是否有运动。如果没有运动,则直接用传统的相邻帧差法检测特定目标,并令 $k = k + 1$, 然后转第 2 步, 直到检测完整个图像序列; 否则判断前帧 I_{k-1} 是否为场景初始帧。如果不是场景初始帧, 则转第 6 步; 否则用基于场景 3 维位置恢复的背景区域判断方法区分背景区域和运动物体区域;

(4) 然后背景区域去除和运动目标分割。即直接将 I_{k-1} 和 I_k 中背景区域对应的像素灰度设置为“0”, 而保留它们运动物体区域对应的像素灰度值。

(5) 用模板匹配的方法在保留灰度原值的区域内检测特定的目标。令 $k = k + 1$ 转第 2 步继续算法, 直到检测完整个图像序列。其中模板的选择和更新方法详见文献[1]。

(6) 判断是否发生场景切换。如果没有发生场景切换, 则认为 I_k 中与 I_{k-1} 已判断为背景区域相配对的区域也是背景, 并利用此背景区域的投影矩阵去判断其他运动的运动区域(包含已判断为运动物体的区域和新进入图像的区域)。转第 4 步; 否则令 $k = k + 1$, 并将前帧 I_{k-1} 设置为场景初始帧, 并转第 2 步, 直到检测完整个图像序列。

4 实验及结果

我们对一个实拍的图像序列进行目标跟踪实验验证新运动检测算法的有效性。图像序列是用平移运动的摄影机采集的, 拍摄时将摄影机架设在了一辆平移的小车上, 运动目标为奔驰的摩托车, 摩托车时速为 35km/h, 摩托车在摄影机视场内时离摄像机的距

离范围为 10 ~ 15m。静止景物中树离摄影机较近, 房屋和草离摄影机相对较远。因此, 在图像序列中树看起来平移的相对快些, 而房屋相对慢一些。显然不满足传统全局运动补偿方法对“背景遵循同一组运动参数”的假设, 此时如果使用传统的先补偿全局运动, 然后差分以消去背景确定运动目标的潜在区域, 最后通过精确匹配以定位运动目标的方法来检测本序列中的运动目标, 则虽然可以通过运动补偿消除掉占背景主要部分的房屋和草等部分, 然而没有消除掉离摄影机较近的树等, 而无法进行有效的背景去除, 要检测出运动目标需要搜索很大的图像区域, 给目标检测和跟踪带来很大的困难。而本文给出的方法则是采用先运动分层, 然后进行运动层配对, 最后判断这些运动层是否背景的思路, 从而避免了传统方法不能消除树等影响的困难, 较好地解决这类图像序列中的目标检测和跟踪问题。跟踪结果如图 2 所示(这里只给出了第 1, 15 和 25 帧的跟踪情况)。

实验时使用文献[20]给出的目标遮挡检测和处理方法。总的说来, 传统的补偿差分检测算法原理简单易于实现, 在解决只有一组全局运动参数的图像序列中目标检测问题时可以取得很好的效果, 但是在解决具有多组全局运动参数的图像序列中目标检测时, 则往往不能高效准确地工作; 而新算法在具有多组全局运动的动态场景图像序列能比较准确地检测跟踪目标; 同时使整个运动目标检测方法有了更好的自适应性, 跟踪也更加稳定, 但是新方法比较紧密地依赖于图像运动层次分割的效率和每层运动参数估计的准确度。

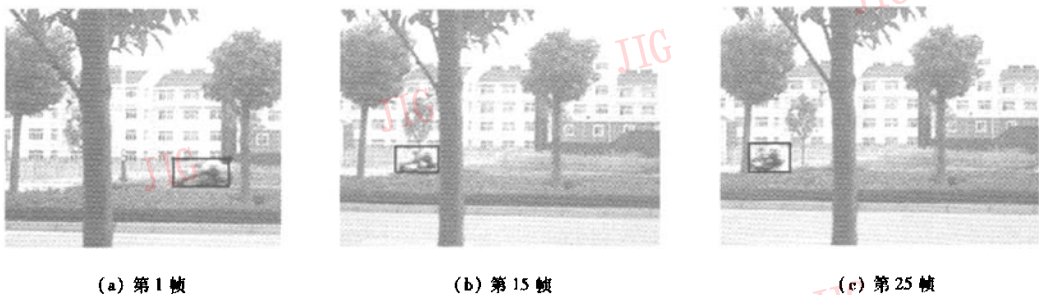


图 2 新方法对具有多组全局运动参数的图像序列中目标跟踪结果

Fig. 2 The result of moving object detection in image sequence by the proposed approach

5 结论

本文简单介绍了传统的动态场景图像序列中运

动目标检测方法, 并且分析了其利弊。针对其存在的局限, 给出了全新的基于场景参考点 3 维位置恢复的运动目标检测方法。该法首先引入了一种层次化运动模型, 并用它来进行图像帧的运动分割; 然

后,给出了运动层配对方法,以及基于场景参考点3维位置恢复的背景区域判断方法。最后,给出了动态场景图像序列中运动目标检测的详细算法。实验结果表明,新算法可以直接判断出图像中的静止背景区域,并且予以去除,以提取出运动目标区域,从而快速、准确地检测出运动目标。本文的贡献之一在于突破了传统动态场景图像序列中运动目标检测方法要求“背景景物的位置差异对图像序列的帧间全局运动的影响可以忽略”的限制。另一个贡献是给出了一种场景切换的检测方法,使算法具有更好的鲁棒性。

参考文献 (References)

- 1 Zheng Shi-you, Fei Shu-min, Long Fei. A new algorithm of moving target detection in image sequence based on the lifting scheme[J]. *Journal of Image and Graphics*, 2005, 10(5): 596~602. [郑世友, 费树岷, 龙飞. 基于小波提升框架的图像序列中运动目标检测算法[J]. *中国图象图形学报*, 2005, 10(5): 596~602.]
- 2 Mech 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 Paul Withagen, Klammer Schutte, Frans Groen. Object detection and tracking using a likelihood based approach[A]. In: *Proceedings of the ASCI 2002 Conference*[C], Lochem, The Netherlands, 2002: 248~253.
- 4 Dufaux F, Konrad J. Efficient, robust and fast global motion estimation for video coding [J]. *IEEE Transactions on Image Processing*, 2000, 9(3): 497~500.
- 5 Sarkar A, Biswas M K, Sharma K M S. A simple unsupervised MRF model based image segmentation approach[J]. *IEEE Transactions on Image Processing*, 2000, 9(3): 801~812.
- 6 Zhang Tong, Carlo Tomasi. Fast, robust, and consistent camera motion estimation [A]. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* [C], Los Alamitos, CA, USA, 1999, 1: 164~170.
- 7 Tsaig Y, Averbuch A. Automatic segmentation of moving objects in video sequences: A region labeling approach[J]. *IEEE Transactions on Circuits and System for Video Technology*, 2002, 12(7): 597~612.
- 8 Neri A, Colonnesse S, Russo G, *et al.* Automatic moving object and background separation[J]. *Signal Processing*, 1998, 66(2): 219~232.
- 9 Chien S Y, Ma S Y, Chen L G. Efficient moving object segmentation algorithm using background registration technique [J]. *IEEE Transactions on Circuits and System for Video Technology*, 2002, 12(7): 577~586.
- 10 David A Forsyth, Jean Ponce. *Computer Vision: A Modern Approach* [M]. New Jersey: Prentice Hall, 2002.
- 11 Weiss Y, Adelson E H. A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models[A]. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* [C], San Francisco, California, USA, 1996: 321~326.
- 12 Etoh M, Shirai Y. Segmentation and 2D motion estimation by region fragments [A]. In: *Proceedings on International Conference on Computer Vision* [C], Seattle, USA, 1994: 192~199.
- 13 Dellaert F, Seitz S, Thorpe C, *et al.* Structure from motion without correspondence [A]. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'00)* [C], South Carolina, USA, 2000: 557~564.
- 14 Wang J Y A, Adelson E H. Representing moving images with layers [J]. *IEEE Transactions on Image Processing*, 1994, 3(5): 625~638.
- 15 Baker S, Szeliski R, Anandan P. A layered approach to stereo reconstruction[A]. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* [C], Washington, DC, USA, 1998: 434~441.
- 16 Irani M, Anandan P. A unified approach to moving object detection in 2D and 3D scenes[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(6): 577~589.
- 17 Weiss Y. Smoothness in layers: Motion segmentation using nonparametric mixture estimation [A]. In: *IEEE Conference on Computer Vision and Pattern Recognition* [C], San Juan, Puerto Rico, 1997: 520~526.
- 18 Figueiredo M A T, Jain A K. Unsupervised learning of finite mixture models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(3): 381~396.
- 19 Zoran Zivkovic, Ferdinand van der Heijden. Recursive unsupervised learning of finite mixture models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(5): 651~656.
- 20 Jia J, Wu T, Tai Y, *et al.* Video repairing: inference of foreground and background under severe occlusion[A]. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)* [C], Washington, DC, USA, 2004, 2(1): 364~371.