

基于自适应分块外观模型的视觉跟踪

赵凌, 冯宾, 邱锦波

(华中科技大学 电子与信息工程系宽带中心, 湖北 武汉 430074)

摘要: 提出将目标和遮挡物视为关联对象, 利用遮挡物的特征来更新模板, 引入高斯混合模型 (GMM) 进行自适应的模板学习。模板中的每个像素利用包含 3 个分量的 GMM 来表示。在当前帧中, 一旦获得模板匹配, 模板中每个像素及其对应的 GMM 都会进行相应的更新。实现结果表明, 对于部分遮挡和外观变化, 算法能够实现顽健的跟踪。

关键词: 部分遮挡; 高斯混合模型; 基于块的模板

中图分类号: TP301.6

文献标识码: A

文章编号: 1000-436X(2011)10-0166-08

Fragment-based visual tracking with adaptive appearance model

ZHAO Ling, FENG Bin, QIU Jin-bo

(Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: A joint object was proposed which combine the target and the occluder in order to exploit the characteristics of the occluder to evolve the target template. To this end, the Gaussians mixture model (GMM) was introduced to enable an adaptive learning mechanism. Each pixel of the template is modeled using a GMM with three components. Once the template-matching is achieved in the current frame, the pixel value is used to update a specific component in its GMM. After that, the pixel will also be updated. Experimental results show that the proposed method is robust with respect to the partial occlusions and natural appearance changes.

Key words: partial occlusions; GMM; fragment-based template

1 引言

在计算机视觉领域中, 对象跟踪是一个富有挑战性的任务。它要求在连续的图像序列中, 稳定地定位目标的位置和大小等信息, 它包括人机交互、视频内容分析和视频监控在内的众多应用的基础和核心问题。研究者们自 20 世纪 80 年代以来对其进行了深入的研究, 但复杂的物体运动、非刚性对

象、混乱的背景以及部分遮挡或全遮挡等^[1]可能出现的困难, 使得实现一个顽健而且有效的对象跟踪算法仍然是一个巨大的挑战。

对象跟踪算法重点要解决如下几个关键问题: 1) 要选择合适的特征对目标进行表达, 如色彩^[2]、梯度^[3]和形状^[4]等; 2) 建立目标模型, 最常用的包括基于关键点^[5,6]、基于模板^[7]和基于轮廓^[8]的建模等; 3) 在两帧之间如何搜索到对象的最佳匹配、

收稿日期: 2010-05-06; 修回日期: 2010-10-18

基金项目: 国家自然科学基金资助项目 (60903172, 60972016); 国家高技术研究发展计划 (“863” 计划) 基金资助项目 (2009AA01Z205); 中央高校基本科研业务费专项基金资助项目 (2010MS092); 湖北省自然科学基金资助项目 (2008CDB329)

Foundation Items: The National Natural Science Foundation of China (60903172, 60972016); The National High Technology Research and Development Program of China (863 Program) (2009AA01Z205); Chinese Universities Scientific Fund (2010MS092); The Natural Science Foundation of Hubei Province (2008CDB329)

粒子滤波器(PF, particle filter)^[8]和均值漂移 (MF, mean shift)算法^[9]就是被广泛使用的搜索策略; 4) 学习机制在根据外观的变化来更新目标模型的过程中也起到非常重要的作用。这些外观变化一般由运动、遮挡和光照条件的变化等因素造成。更新机制通常很难选择, 一方面固定的模板在长时间的观察中难以适应对象外观的变化, 另一方面快速更新的模板又容易偏离目标, 当目标和背景的外观特征相似的时候, 这种情况尤为严重。

本文选取每个像素的灰度值作为特征来进行跟踪, 主要解决上文提到的第 2)~第 4) 个关键问题, 即目标建模、目标匹配和更新机制。重点研究基于分块的模板表示方法, 以实现顽健的遮挡处理。

文献[10]提出了一种新的基于分块的跟踪算法“FragTrack”。和以前的方法不同, 文献[10]使用的分块可以是任意形状和数量的, 而不是将人体分成四肢和躯干^[11]来建模。由于使用了多个分块, 当出现部分遮挡和姿势改变时, 可以采用未受影响的分块来进行匹配, 易于处理。同时, 积分直方图结构^[12]的采用使得算法计算更加高效。但是, 这种方法主要的缺点在于它使用的外观模板是固定的, 无法跟踪有显著外观变化的物体。

Prakash 等人提出一种参数型的分块表示方法^[13]。该方法首先利用判别分析将每个对象分为多个一致性的均匀分块; 然后每个分块用一个高斯分布来建模, 进而在后续帧中分别匹配。这种方法在跟踪部分遮挡和外观形变的复杂对象时, 有一定的性能改善。但是, 由于只用一个高斯分布来代表一个分块, 而高斯分布本质上是单峰, 所以跟踪的性能就在很大程度上取决于均匀分块的一致性程度。当对象具有复杂外观时, 对均匀分块的精度要求难以满足。但其对分块进行高斯建模, 并根据部分遮挡和外观形变来更新模型的思想值得借鉴。

本文受到上面两篇文章的启发, 提出了一种基于文献[10]的改进算法, 引入混合高斯模型(GMM)对模板中的像素进行建模, 并在此基础上进行模板的匹配和更新。在某一帧中获得了模板匹配区域后, 通过一个两步的更新机制, 来分别更新模板中每个像素对应的 GMM 模型参数(均值、方差和权重等)以及像素值本身。这种逐帧的更新使得模板可以根据目标外观的变化进行自适应的调整, 提高跟踪的顽健性和准确性。

本文算法的改进主要有以下 2 点。

首先, 本文采用的 GMM 对每个像素进行建模。文献[10]中的模板是固定的, 文献[13]采用一个单一的高斯分布来建模每个一致性分块。

其次, 即使检测到遮挡, 也继续更新模板。文献[10]中利用一个顽健的统计量来处理遮挡情况, 而文献[13]则在发现遮挡时停止更新。本文的处理思路在于, 如果遮挡物在一段时间内随着目标一起运动, 那么将目标和遮挡物视为关联对象, 并利用遮挡物的特征来更新模板, 会比起仅仅利用目标剩余未被遮挡的部分, 更容易实现稳定的跟踪。

2 整体框架

本文将目标和遮挡物视为关联对象的思想基于如下观察。如图 1 所示, 图中有 2 个行人在相对运动, 当他们靠近的时候, 左边的行人被右边的遮挡住。随后, 2 个人一起向右行走, 在左侧的行人持续被部分遮挡住的情况下, 相信通过联合使用右侧行人的特征信息来更新目标模板, 可以获得更好的匹配效果。

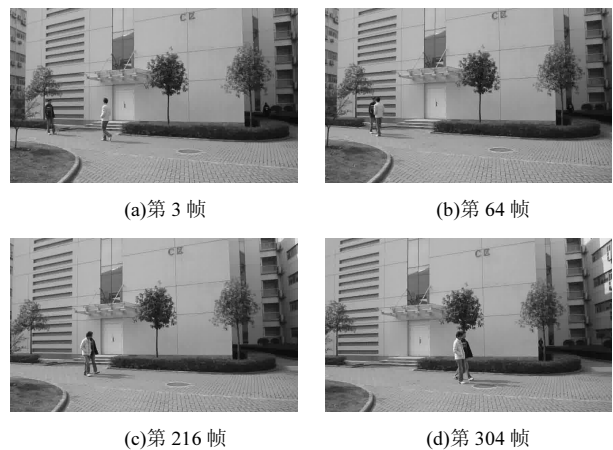


图 1 从手持摄像机拍摄的 450 帧序列中选出的样例图片

本文所提出的算法的流程如图 2 所示。跟踪目标的初始化可以通过人工选择或目标检测算法自动确定, 在本文中采用人工选择的方式, 即在第一帧中通过鼠标选择一个矩形框作为感兴趣区域(ROI), 在后续帧中进行跟踪。根据选定的 ROI, 初始化跟踪模板, 并将其划分成垂直和水平分块的形式, 以积分直方图的结构进行存储, 有利于高效的相似度度量计算。同时, 对于模板中的每一个像素, 初始化包含 3 个分量的 GMM 模型参数。接下来, 在后续的每一帧中进行基于分块的跟踪, 最佳匹配点是通过在前一帧 ROI 位置周围的矩形邻域内遍

历计算每一个候选匹配区域和模板之间的相似性（相异性）程度，选择相异性度量最小的匹配点得到的。为了能够顽健地应对遮挡情况，对于每一个候选匹配区域，分别计算区域内每个分块与模板内对应分块间的相异性度量，并按从小到大进行排序，采用第 Q 个最小的度量值作为该候选匹配点的相异性度量值来与其他的匹配点进行比较。这样处理的出发点是基于如下的假设：即使跟踪对象发生了部分遮挡，仍然有至少 Q （本文采用 25%）个分块的区域是未被遮挡的，利用这些分块可以获得较可信的匹配，进而实现顽健的跟踪。

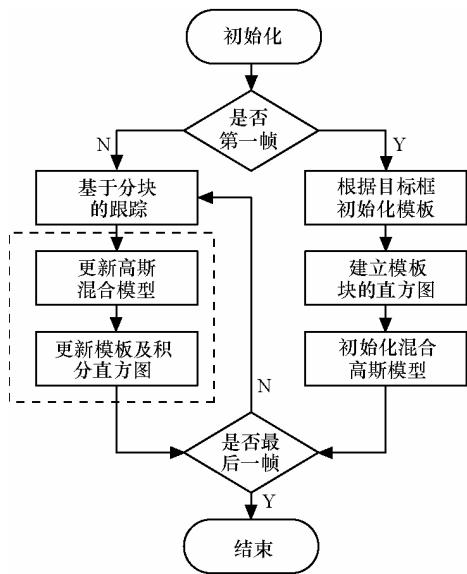


图 2 基于自适应分块外观模型的视觉跟踪流程

在每一帧中获得了匹配区域后，需要更新对象的外观模型，以适应由于外观变化或光照以及其他因素造成的影响。对于模板中的每个像素，本文采用一个两步的更新机制，即首先根据匹配区域中的各个像素值，更新模板中与其对应像素的 GMM 模型，再利用更新后的 GMM 模型来更新模板中的该像素值。更新后的模板仍然以积分直方图的结构存储下来，用来在下一帧中进行匹配跟踪。

3 基于高斯混合模型的自适应模板建模

GMM 在背景建模中应用的非常广泛，它可以适应光照变化、复杂背景以及长期的场景变化，获得顽健的背景模型^[14]。在本文中引入 GMM 对模板中的每个像素的演变过程进行建模，以适应跟踪对象外观的变化。由于单独的高斯函数是单峰的，而具有不同表面特征分布的对象可能在同一个像素

点位置重叠，因此采用多个高斯函数分布是很有必要的。

3.1 混合建模

在第一帧中通过鼠标选择一个矩形框作为感兴趣区域(ROI)之后，就可以对感兴趣区域中的像素进行建模，得到模板的高斯混合模型。对于 t 时刻位于帧内坐标 (x, y) 的像素，采用 GMM 对其灰度值 $I(x, y, t)$ 来建模。GMM 包含 3 个高斯分量， $A(x, y, t) = \{\eta_i; i = S, O_L, O_R\}$ ，其中 η_S 代表稳定分量 S ，它对应模板外观变化缓慢的特征，因此表征的是在很长一段时间内保持稳定的对象属性。 η_{O_L} 和 η_{O_R} 分别代表 2 个离群分量 O_L 和 O_R ，对应由于噪声、遮挡和姿势改变等因素造成的明显偏离 S 分量的值分布，图 3 给出的是 3 个分量间的关系。

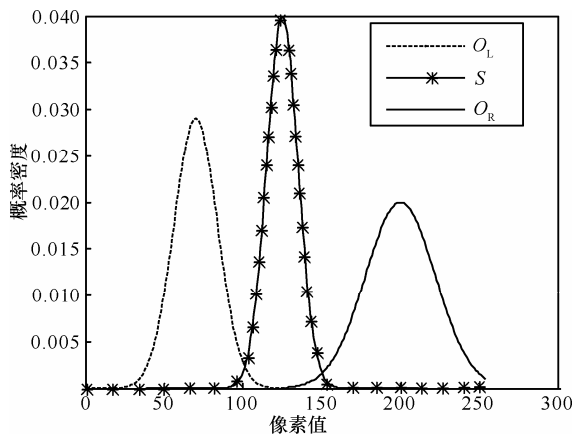


图 3 GMM 中 3 个分量之间关系

为了与文献[10]方法进行公平的比较，本文算法仅使用灰度值信息，但它能够很方便地扩展为包含色彩信息。这里 3 个参数 $\{\mu_i(x, y, t), \sigma_i(x, y, t), w_i(x, y, t); i = S, O_L, O_R\}$ ，分别表示每个分量的均值、标准差和权重，其中 3 个权重值的和为 1，权重则表征了一个像素属于某一分量的置信度。

3 个分量的高斯概率密度由式(1)给出（省略变量 x, y 和 t ）。

$$\eta_i(I, \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(I - \mu_i)^2}{2\sigma_i^2}\right\} \quad (1)$$

3.2 在线更新

在每一帧得到最佳匹配后，都需要对模板进行在线更新。模板的在线更新是在给定 $t-1$ 时刻的估计结果以及 t 时刻的观察量条件下，对当前时刻 t 模板内所有像素的模型参数进行估计。如果用标准的期望值最大化(EM)^[15]算法来解决这个问题，需

要保留之前时刻的数据并进行复杂的迭代，计算复杂度高、需要的内存空间大，因此采用在线 K 均值近似法^[14]。

首先，根据在 $t=0$ 时刻人工选择的跟踪对象，使用式(2)~式(6)来初始化对象的外观模型参数（省略变量 x 和 y ），其中除了 $I(x, y, 0)$ 外，等式中其他参数通过实验预先确定。这里的 β 是一个标量系数，表示 3 个分量之间的分布距离。即假设在不越界的情况下，稳定分量均值与两个离群分量的均值差为 $2\beta\sigma_s(0)$ 。

$$\mu_s(0) = I(0), \sigma_s(0) = \sigma_0, w_s(0) = w_0 \quad (2)$$

$$\mu_{O_L}(0) = \max[I(0) - 2\beta\sigma_s(0), 0] \quad (3)$$

$$\mu_{O_R}(0) = \min[I(0) + 2\beta\sigma_s(0), 255] \quad (4)$$

$$\sigma_{O_L}(0) = \sigma_{O_R}(0) = \sigma_1 \quad (5)$$

$$w_{O_L}(0) = w_{O_R}(0) = w_1 \quad (6)$$

接下来，经过在邻域内的遍历搜索，每个 $t > 0$ 时刻将获得一个新的匹配区域，然后就可以根据该匹配结果来更新对象的外观模型。本文采用一个两步的更新机制，首先将匹配区域中的各个像素值与模板中对应位置像素在 $t-1$ 时刻获得的 GMM 模型中 3 个分量分别进行比较，以确定其属于哪一个分量。如果该像素值和 S 分量均值差的绝对值小于 β 倍的标准差，就表示与 S 分量匹配。否则，就意味着出现了像素值显著偏离 S 分量的均值情况(如由于遮挡物的出现)，接下来根据像素值和 S 分量均值间的大小关系，判断该像素属于 O_L 或者 O_R 分量。如图 3 所示，由于 3 个分量的高斯函数覆盖了像素值的全部范围，即 0~255，所以每个像素都会和一个分量相匹配，这意味着即使发生遮挡也会继续更新跟踪模板，不同的是当对象外观发生显著变化时， O_L 或 O_R 分量更新，否则 S 分量进行更新。匹配分量的均值和方差自适应更新如下：

$$\mu_{mc}(t) = (1 - \rho)\mu_{mc}(t-1) + \rho I(t) \quad (7)$$

$$\sigma_{mc}^2(t) = (1 - \rho)\sigma_{mc}^2(t-1) + \rho(I(t) - \mu_{mc}(t))^2 \quad (8)$$

其中， mc 代表匹配分量，即 $mc=S, O_L, O_R$ ，参数 ρ 代表均值和方差调整的幅度大小，由式(1)和学习率参数 α 计算得到。学习率参数 α 决定了模型以多快的速度“遗忘”原来的数据并更新参数。这表明更近的观察量在更新分布参数的过程中越重要。 α 必须仔细选取，因为一方面为了处理遮挡需要保持长

时间的一致性，但另一方面模型也需要根据对象的外观变化及时进行自适应改变。匹配分量的权重使用式(10)更新。之后，3 个分量需要进行归一化，使其和为 1。

$$\rho = \alpha \eta_{mc}(I(t), \mu_{mc}(t-1), \sigma_{mc}(t-1)) \quad (9)$$

$$w_{mc}(t) = (1 - \alpha)w_{mc}(t-1) + \alpha \quad (10)$$

完成了对模板中每个像素的 GMM 模型的更新以及归一化以后，接下来需要进行第二步的跟踪模板更新。首先，根据 3 个分量各自的权重和方差的比值大小进行排序，这个比值越大，当前像素属于该分量的可能性越高；然后，选择比值最高的分量，用其高斯分布的均值来更新模板中的对应像素值。该模板更新过程如式(11)和式(12)所示。式(11)中分母应该与一个小的常数相加来避免被零除。注意 $\tilde{w}_i(t)$ 是 $w_i(t)$ 归一化的结果，同时 $\hat{I}(t)$ 是第 t 帧的更新结果并且在下一时刻用作 $I(t+1)$ 。

$$i^* = \arg \max_{i \in \{S, O_L, O_R\}} (\tilde{w}_i(t) / \sigma_i^2(t)) \quad (11)$$

$$I(t+1) = \hat{I}(t) = \mu_{i^*}(t) \quad (12)$$

需要注意的是，模板中 GMM 分量的个数可以进行灵活的调节，来适应对象分布的变化。但在本文实验中发现，采用 3 个分量可以在计算复杂度和跟踪性能两方面获得较好的折衷。

4 基于分块的跟踪

在上一节，讨论了怎样定义自适应外观跟踪模型，以及一旦找到新的匹配观察量后，模板如何更新的问题。接下来的问题是：根据前一帧给出的模板最佳估计结果，如何在当前帧中找到和模板最佳匹配的结果。一般有 2 种解决方法可以解决这个问题：确定性方法和随机性方法。均值漂移(MS)^[9]和粒子滤波器(PF)^[8]分别是 2 种方法的代表。不过，MS 本质上是梯度下降算法，容易陷入局部最小；虽然由于候选假设大部分是随机选取的，粒子滤波器能够避免局部最小值，但计算复杂度的问题，使其在实际应用中受到限制。

事实上，连续帧之间有丰富的时间冗余，这意味着当前帧的最佳匹配的位置通常在上一帧目标位置的相邻区域，这也是视频编码框架中运动估计算法的基础^[16]。根据这一思路来搜索当前帧中模板的最佳匹配。如图 4 所示，模板被水平和垂直划分为分块集合。对于每个块来说，在当前

位置的一个 $(2r+1)\times(2r+1)$ 邻域内, 使用穷尽搜索来定位最佳匹配位置。参数 r 指定了搜索区域的大小, r 越大, 需要估计的候选假设区域就越多。虽然块的形状原则上是任意的, 但由于矩形在相似性计算上的高效性, 这里采用矩形结构(详见第 5 节)。

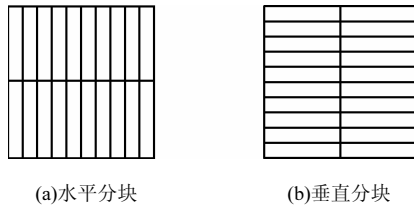


图 4 匹配模板中的分块模式

对每个候选假设位置, 需要估计相邻两帧对应块之间的相似性。采用直方图数据结构来进行估计。虽然它也许不是最好的非参数密度估计, 但能够满足实时应用的低计算复杂度需求。常用的测量 2 个直方图之间的相似性的方法包括 Bhattacharya 系数^[17]等, 文献[18]提出了一种更有效的方法——EMD(the earth mover's distance)。它度量的是将一个直方图分布改造成另一种直方图分布(包括将当前直方图的某个部分移到新的位置)所需要的开销。EMD 方法适用于任何维数, 并且可以提供更加直观的匹配结果, 所以这里用 EMD 来度量匹配结果之间的相似性。在计算了搜索区域中的所有候选假设之后, 可以得到一个“强度”图像。“强度”图像中的每一点的值代表的是目标块和对应的候选块之间的相异程度, 然后需要结合每个块的“强度”图像来为整个模板估计出最佳匹配位置。文献[10]提出了利用最小均方值估计(LMedS)来处理由遮挡块或显著的外观变化所引起的离群点, 其基本思路是即使有一部份分块发生了遮挡, 对于剩下的可见块来说, 它们的“强度”图像也是可信的。因此, 对于某一候选假设, 将所有块的“强度”图像中对应坐标点的“强度”值(即这一候选假设所有分块与模板的对应分块的相似性值)进行排序, 然后选择第 Q 个最小的值(因为 EMD 实际上是不相似性测量, 值越大, 表明越有可能是离群点)来代表对当前候选假设的综合评估值。 Q 值的选取反映了对于至少有多少比例的分块是始终保持可见的一个先验估计。在获得了候选区域中所有候选点的综合评估值之后, 值最小的点即为最佳匹配结果。

5 积分直方图

如前文所述, 利用了直方图来建模矩形分块的密度分布。由于将匹配模板划分成了多个分块, 而且每个分块需要与多个候选假设进行比较, 所以直方图的有效提取和计算对于本文所提算法的实用性至关重要。

文献[12]提出的积分直方图结构可以近似满足这个需求。积分直方图是对积分图像概念的扩展^[19]。积分图像是从输入图像中计算得到的, 其中每个点的值是输入图像中该点位置和图像左上角原点之间构成的矩形区域内所有像素值之和。得到积分图像后, 计算图像中任意大小矩形区域内像素值的和的任务固定为 4 个点的加减运算操作, 而与图像的尺寸无关, 非常高效。在积分图的思想上进行延伸, 计算积分直方图则是为直方图的每个区间计算一个积分图像, 该积分图像中每个像素点的值表示的是, 原始图像中在该点位置和左上角原点之间的区域中, 落入到当前颜色量化区间内的像素点个数。有了积分图像, 直方图的提取就非常容易了。

6 实验结果

本文采用多个自然图像序列进行测试, 其中大多数序列来自公共数据库。通过与文献[10]提出的使用固定模板的分块跟踪(“FragTrack”)比较, 验证本文算法(“AdapFrag”)的优势, 该算法的原始代码由 Adam 提供^[20]。其中, 目标模板通过在起始帧人工标记一个矩形区域来确定, 搜索区域参数 r 设置为 7 个像素, 目标模板用 RGB 空间表示, 每个颜色通道量化成 16 个区间, 分位数 Q 值为 25%, 这些设置均与文献[10]相同。而基于 GMM 的外观模板的参数设定如下: α 为 0.2, β 为 0.25, σ_0 为 3, σ_1 为 5, w_0 为 0.8, w_1 为 0.1, 实验中所有算法参数都是固定的。

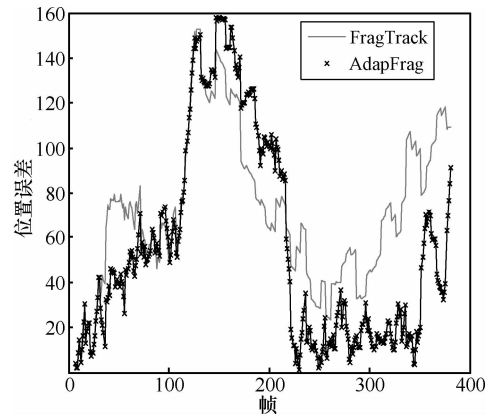
图 5 为 2 个公共测试序列“David”和“Faceocc2”^[20]的测试结果。“David”序列中有光照和外观变化, “Faceocc2”序列中有遮挡和姿势改变。跟踪结果用矩形框来表示, 其中黑色矩形框对应“AdapFrag”的结果, 灰色矩形框对应“FragTrack”的结果。从图中可以看出, “AdapFrag”算法效果更好, 因为它可以顽健地处理由遮挡引起的外观改变和离群值, 而“FragTrack”却明显地偏离了目标。2 个序列的定量比较结果如图 6 所示, 图中曲线表示的是

2 种算法的跟踪结果相对于正确标定位置^[21]之间差的均方根值，结果表明在多数情况下，本文的算法比“FragTrack”效果更好。图 7 在本文算法 2 个序列的跟踪结果的右上角，还分别给出了当前帧所使用的匹配模板，从中可以看出，模板中主要保持的是对象的长期特征，同时也包含有部分短期特征，如“David”序列中面部的阴影，以及“Faceocc2”中人面部有部分书本的纹理等。

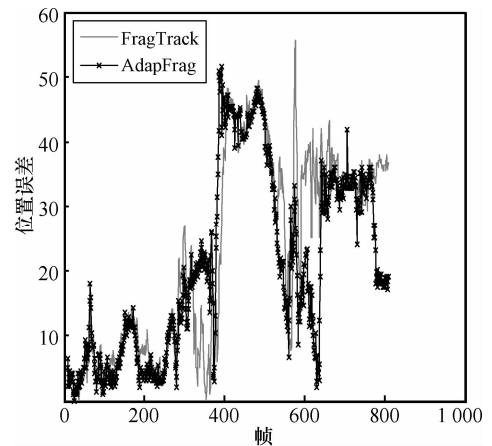


图 5 “David”和“Faceocc2”序列的结果

来自 CAVIAR 数据库^[22]的一个序列的测试结果如图 8 所示，这个序列具有明显的尺度变化。而在 2 个跟踪算法中，并没有引入尺度自适应能力，因为更多地去关注自适应外观模型与固定外观模型相比性能的提升，这也是本文的主要贡献。对于背着书包的目标人物，“AdapFrag”能可靠地跟踪，而“FragTrack”却完全丢失目标，并且无法恢复，这突出表明了使用自适应的外观模型的优点。



(a) David 序列



(b) Faceocc2 序列

图 6 “David”和“Faceocc2”序列定量比较结果



(a) David 序列 345 帧



(b) Faceocc2 序列 709 帧

图 7 “David”和“Faceocc2”序列选取的图像帧的结果



(a)初始帧



(b)第 362 帧



(c)第 383 帧



(d)第 422 帧

图 8 CAVIAR 序列结果

图 9 中给出使用手持摄像机拍摄的序列的部分比较结果。向右运动的行人为跟踪对象，在这个过程中，他被穿白色外套的行人遮挡。在共同行走了一段距离后，他们又分开向反方向运动。在这种情况下，“FragTrack”明显地偏离了跟踪对象。因为当目标被其他对象遮挡的时候，它仍然利用目标的特征进行跟踪。但是，“AdapFrag”仍然可以很好地匹配，因为当遮挡发生的时候，算法会利用遮挡物的特征来适当地更新匹配模板。这体现了将目标和遮挡物视为关联对象来更新外观模型的优势。



(a)初始帧



(b)第 238 帧



(c)第 340 帧



(d)第 390 帧

图 9 自己拍摄序列的部分结果

在跟踪过程中，由于需要不断地更新跟踪模板，相对于固定模板的算法而言，计算量会有一定的增加。表 1 给出的是本文算法相对于“FragTrack”算法，处理时间所增加的比例。从表中可以看出，复杂度的增加最大不超过 4%，平均在 2%左右，处于可以接受的范围内，并且该比较结果是在没有进行特殊的代码优化条件下测试得到的，另外，由于只使用像素级的操作和更新判决，因此可以充分利用并行处理架构的优势来降低复杂度。

表 1 “AdapFrag”与“FragTrack”的复杂度比较结果

序列	增加的时间比/%
David	2.51
Girl	1.58
Faceocc2	3.98
CAVIAR	1.92
自拍序列	1.80
平均	2.36

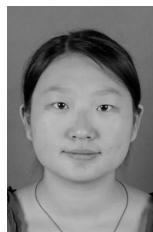
7 结束语

本文提出了一种基于自适应分块外观模型的视觉跟踪算法, 该算法可以根据运动对象的外观变化、姿势变化和遮挡等现象自适应调整模板, 实现顽健的跟踪。它引入了 GMM 对跟踪模板中的各个像素进行建模, 其中每个 GMM 包括 3 个分量: 稳定分量 S 和 2 个离群分量 O_L 和 O_R 。当前帧中, 一旦模板匹配, 每一个像素值被用来更新 GMM 中生成该像素值概率最高的分量的模型参数。然后, 利用对应权重和方差的比值对 GMM 中的 3 个分量进行置信度排序, 用置信度最高的分量来更新模板中的像素值。通过全面的定性和定量实验结果证明, 本文算法的跟踪效果要优于现有的同类算法。

参考文献:

- [1] YILMAZ A, JAVED O, SHAH M. Object tracking: a survey[J]. ACM Journal of Computing Surveys, 2006, 38(4):13.
- [2] MCKENNA S, RAJA Y, GONG S. Tracking color objects using adaptive mixture models[J]. Image and Vision Computing, 1999,17: 225-231.
- [3] BIRCHFIELD S. Elliptical head tracking using intensity gradients and color histograms[A]. Proc IEEE Conf on Computer Vision and Pattern Recognition[C]. Santa Barbara, California, 1998. 232-237.
- [4] KALE A, JAYNES C. A joint illumination and shape model for visual tracking[A]. Proc IEEE Conf on Computer Vision and Pattern Recognition[C]. New York City, NY, 2006.602-609.
- [5] ZHOU H, YUAN Y, SHI C. Object tracking using SIFT features and mean shift[J]. Computer Vision and Image Understanding, 2009, 113(3): 345-352.
- [6] HE W, YAMASHITA T, LU H, *et al.* SURF tracking[A]. International Conference on Computer Vision[C]. Kyoto, Japan, 2009.
- [7] MEYER F, BOUTHEMY P. Region-based tracking in an image sequence[A]. European Conf on Computer Vision[C]. Santa Margherita Ligure, Italy, 1992. 476-484.
- [8] ISARD M, BLAKE A. Condensation: conditional density propagation for visual tracking[J]. Int Journal of Computer Vision (IJCV), 1998, 29(1):5-28.
- [9] COMANICIU D, RAMESH V, MEER P. Real-time tracking of non-rigid objects using mean shift[A]. Proc IEEE Conf. on Computer Vision and Pattern Recognition[C]. Hilton Head, South Carolina, United States, 2000. 142-149.
- [10] ADAM A, RIVLIN E, SHIMSHONI I. Robust fragments-based tracking using the integral histogram[A]. Proc IEEE Conf on Computer Vision and Pattern Recognition[C]. New York City, NY, 2006. 98-805.
- [11] SIGAL L, BHATIA S, ROTH S, *et al.* Tracking loose-limbed people[A]. Proc IEEE Conf on Computer Vision and Pattern Recognition[C]. Washington, DC, USA, 2004.421-428.
- [12] PORKILI F. Integral histogram: a fast way to extract histograms in cartesian spaces[A]. Proc IEEE Conf on Computer Vision and Pattern Recognition[C]. San Diego, CA, USA, 2005. 829-836.
- [13] PRAKASH C, PALURI B, PRADEEP N, *et al.* Fragments based parametric tracking[A]. Asian Conference on Computer Vision[C]. Tokyo, Japan, 2007.522-531.
- [14] STAUFFER C, GRIMSON W. Adaptive background mixture models for real-time tracking[A]. Proc IEEE Conf Computer Vision and Pattern Recognition[C]. Ft Collins, CO, USA, 1999.246-252.
- [15] DEMPSTER A, LAIRD N, RUBIN D. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society, 1977, 39 (Series B):1-38.
- [16] Information Technology—Coding of Audio-Visual Objects— Part 10: Advanced Video Coding[S]. Final Draft International Standard, ISO/IEC FDIS 14 496-10.
- [17] BHATTACHARYYA A. On a measure of divergence between two statistical populations defined by their probability distributions[J]. Bulletin of the Calcutta Mathematical Society, 1943,35: 99-109.
- [18] RUBNER Y, TOMASI C, GUIBAS L. The earth mover's distance as a metric for image retrieval[J]. Int Journal of Computer Vision (IJCV), 2000,40(2):91-121.
- [19] VIOLA P, JONES M. Robust real time object detection[A]. IEEE ICCV Workshop on Statistical and Computational Theories of Vision[C]. Vancouver, Canada, 2001.
- [20] http://vision.ucsd.edu/~bbabenko/project_miltrack.shtml[EB/OL].
- [21] <http://www.cs.technion.ac.il/~amita/fragtrack/fragtrack.htm>[EB/OL].
- [22] <http://groups.inf.ed.ac.uk/vision/caviar/caviardata1/>[EB/OL].

作者简介:



赵凌 (1985-), 女, 湖北宣城人, 华中科技大学硕士生, 主要研究方向为目标跟踪。

冯镔 (1978-), 男, 湖北武汉人, 博士, 华中科技大学副教授, 主要研究方向为视频编码与多媒体通信。

邱锦波 (1977-), 男, 湖北武汉人, 博士, 华中科技大学讲师, 主要研究方向为宽带无线与多媒体通信。