

指导教师： 杨涛

提交时间： 2015/3/27

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名： 张欢

学号： 2012302407

班号： 10011201



L0 正则静止时间估计人群组分析

易帅 汪小刚 陆策毋 佳家崖

香港大学中国

{syi,xgwang}@ee.cuhk.edu.hk, {cwlu,leojia}@cse.cuhk.edu.hk

摘要

这篇文章我们着手做静止人群分析，是因为它有模拟移动人群类似重要场景，在监视中我们发现了许多应用。我们的关键贡献是提出估计多久前景像素变为静止的稳健算法。这是远比只减去背景更具挑战性，因为由于局部运动目的，照明的变化，并且闭塞造成的单个帧失败可能导致在静止的时间估计上较大的误差。为了达到相当好的结果，混合质点在时间和空间的稀少限制共同加入来塑造一个 3D 静止时间图。它被配制成 L0 优化问题。

除了去除背景，它通过使用在局部共享前景码本。区别不同的在时空空间接近或重叠前景物体。我们所提出的技术用于检测出四种类型的静止群体活动和分析现场围观的结构。我们提出首次公开基准为 dataSet1 的静止时间估计和静止群分析。

1.简介

在视频监控 [25, 1, 4, 26, 15, 18, 32, 30, 6, 31, 20] 中人群分析我们发现了许多重要的应用。人群管理及交通管制是人口密度高的在公共领域常见的问题。现有的工作主要集中在检测运动模式人群 [25, 1, 26, 18, 32, 13, 30, 31, 20] 和分析的运动过程中行人之间的相互作用 [15, 17, 19, 28, 20] 另一方面，静止的人群组分析从未被充分研究 尽管这些团体能够提供令人惊讶的丰富的信息。

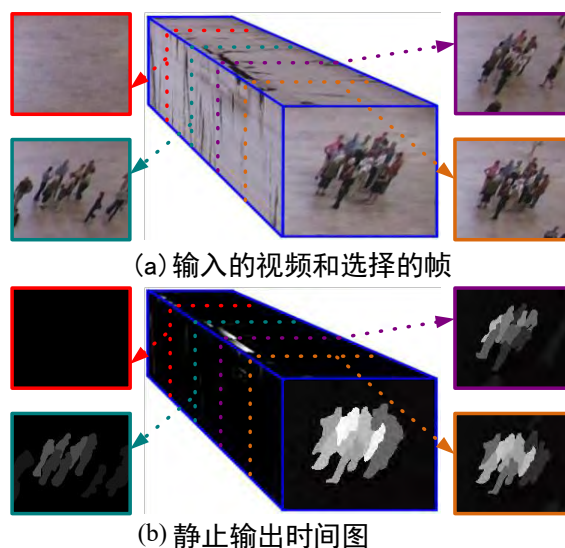


图 1. 估计从一个视频序列中的 3D 静止时间映射。结果显示在一些帧里。多长时间像素静止为一帧是依据强度水平编码。亮的像素对应于较长的时间。

例如，[16] 研究表明，静止人群在某些情况下对不断变化的交通模式比移动人群更大的影响。当行人走动，它们调节的速度而不调整方向来避免冲突。这样自组织行为保持交通流畅。如果行人形成静止的群体，他们强迫别人改变方向会很大程度上影响运输效率。静止群体的突发事件和分散导致人群移动模式动态变化。因此，这是极大的吸引力把静止的群体纳入动态建模人群系统中。研究静止

集团有可能出现位置，和停留的时间是值得的。显示在图三中平均停留时间的地图为人群管理，以及设施和支持提供支持。

静止一段时间团体往往是值得关注。他们的出现，分散，静止持续时间和状态可能招致极大的安全隐患，从检测到的活动中，我们可以发现有价值信息，如人们之间的关系与可能的异常。图 2 示出在本文中待检测的四种类型的静止组活动。

我们估计静止的时间时期 1 的方法允许局部区域中前景像素存在局部运动。如图 1 中，给定的一个视频序列，我们的方法生成在时空空间 3D 静止时间地图。这区别于仅仅的背景去除。我们尝试简单检测前景个别帧，并计算像素已经在前景中停留的时间。结果通常较差。因此，我们把它作为一个新的研究课题。它是进一步分析静止的人群的重要步骤。

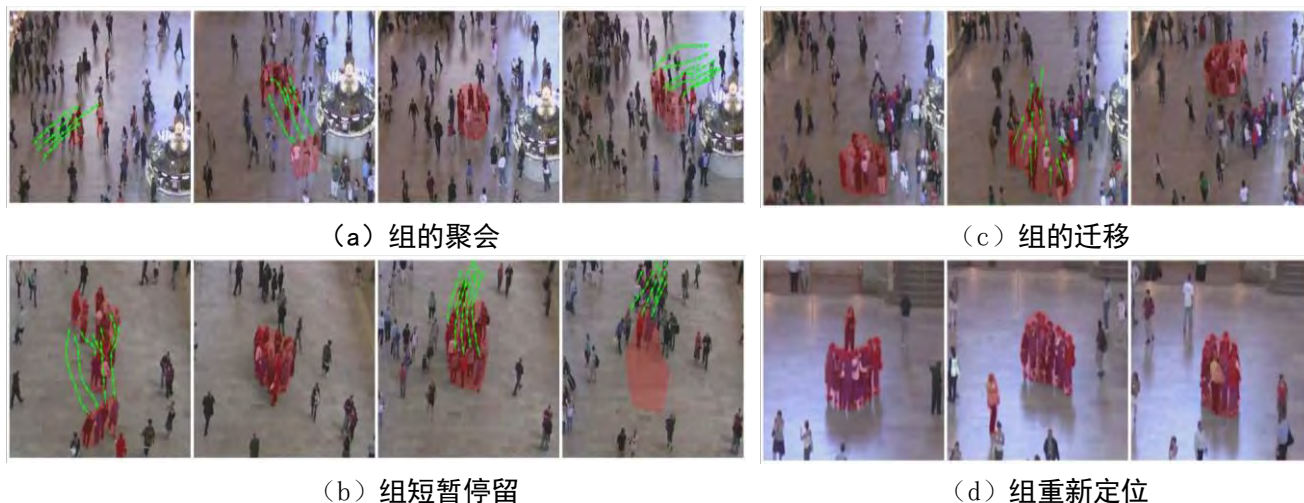


图 2. 我们在工作中检测四个主要的静止团体活动类型，典型场景出现在 CVPR 发布会现场两种口语会话之间空闲时间。（a）人们加入一组从不同的时间不同的方向。当所有的人到达，全团沿着同一目的地移动。（b）有一群人进入视图在一起，呆了一段时间，并一起离开。（c）之后住在一会儿的局部，人移动到另一个位置，并变得静止了。（d）在一组的人都有自己的活动，例如拍照。请查看彩色图像。

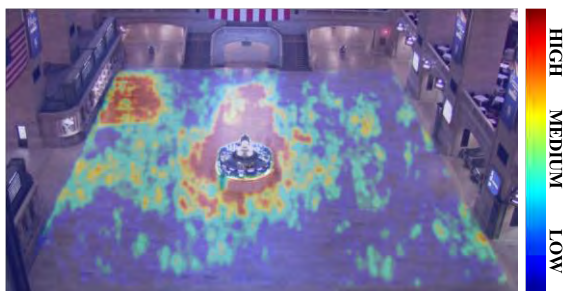


图 3 超过 4 小时. 场均停留时间分布。静止的群体往往出现并围绕信息展位，和售票窗口前停留（用彩图看更好）。

图 4 示出固有的挑战。首先，背景去除并不能区分不同的前景物体。如果两个对象重叠，则估计的停留时间应该比它在重叠区域更长，如图 4（a）和 5。这在拥挤的场面经常发生。

其次，人的局部运动在静止时间过程中很常见如图 4（b）。我们应该继续累积这些局部静止时间即使他们有局部运动，而不是频繁复位时间为 0。匹配局部运动的前景物体尤其是在拥挤的场面是不容易的。第三，大多数的背景去除方法不考虑时间一致性。如果在一帧中的前景像素被错误分类，静止时间可以被错误地重置为 0，如图 4（c）所示。鉴于所有这些挑战加之，没有一个现有的方法中能完备地解决我们的问题。

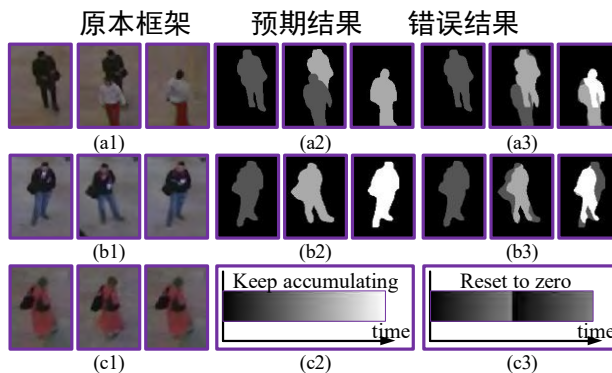


图 4. 挑战静止的时间估计。从结果背景去除是错误的。(a) 两前景对象与时空重叠。对象 (b) 的局部运动也导致估计误差。(c) 假如前景像素被错误分类作为背景在一帧, 静止时间复位为 0, 这是错了。在 (C3), 误分类发生在中间, 使时间复位。

我们的贡献如下。(1) 一个强大的静止时间估计算法。通过使用局部共享的前景码本, 它分离前景物体即使它们在时空空间是接近或重叠的。它还允许匹配在局部区域转移的前景物体。(2) 空间和时间维度的稀疏限制共同加入到构造一个 3D 静止时间映射。这被配制成 L0 优化问题, 较常用的分别对图像和时间空间的增加的局部平整度 (例如 MRF) 有更强大的正规化。一组帧, 而不是个别的被执行。这个过程是健壮对于偶尔局部静止物体, 闭塞, 和误分类的移动。(3) 提出了一套新的描述符描述图 2 所示静止集体活动和用来估计的平稳时间如图 3 (4)。带注释地面实况的一个数据集提供给公众搭建静止时间估计和静止组活动检测, 这是它的第一个应用。

2. 相关工作

背景去除已被很好地研究。斯托弗等人 [23] 提出自适应高斯混合模型。一个流行的做法, 奇夫科维茨等 [33] 提出了改进版。它采用高斯混合适应背景的变化。Kim 等人 [11] 用码本模拟复合背景变化。这些方法面临的挑战已在第一部分 1. 讨论, 鲁棒的 PCA [3] 分离前景和背景作为一个稀疏矩阵和低秩矩阵。因为前景像素长静止时间很可能被归类为背景它是不适合于这个估计任务。

更多和我们的背景去除工作有关是 [21] 中提到的贝叶斯模型。它采用色彩和位置以及执行的非参数, 密度估计的联合特征处理的背景局部运动 (例如波)。为了保持时间的持久性, 一个像素成为前景的可能性以一个常数增加, 如果它为前景中检测的前一帧。则在 MRF 单帧内执行空间的平滑。

这种方法我们的完全不同。他们的空间和时间持续性上的约束分开的, 而我们的在 3D 空间中联合模式模仿了他们在稀疏二阶梯度。此外, 贝叶斯模型内帧之间或两帧进行优化, 而我们的方法共同优化了一批视频帧。稀疏 L0 是使用在许多应用中非常强大的正规化的形式。最后, 它没有从不同的前景对象进行区分。由于人群中频繁闭塞, 和发现好的特征点的难度, 密集追踪 [24] 对前景像素的静

止时间估计可能是有问题的。由于沉重的闭塞, 在拥挤的场面行人探测和跟踪执行得也不好。我们的实验表明, 这些方法不能生产和我们的类似良好的结果。y 检测社会团体和分析他们的活动是另一个方向流。Cristani 等 [7] 站在社会学的观点研究了人的相互作用。其他沿着此行的工作主要考虑移动组。行人根据其相对距离和相似性的移动模式 [10, 9, 5] 被分组。各种特征和模型被提出来识别不同的移动组行为 [29, 14, 2, 22, 8, 12]。如在图 2 中所讨论的, 静止的集团有自己的属性和特殊功能来描述它们。

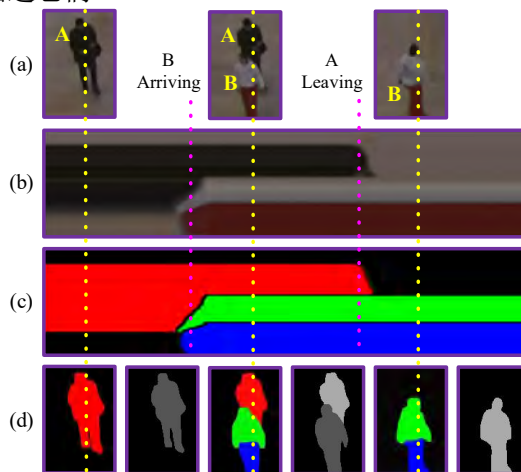


图 5. 使用局部共享前景码本分离在时空空间接近或重叠的前景物体。(a) 来自同一地区的 3 个框架。当人 B 到达, A 离开该集团。(b) 沿黄线时间范围时间片图像。A 和 B 重叠。(c) 前景像素分配有三种不同的代码。他们被很好的分离。(四) 前台代码 (左) 和估计的静止时间 (右)。

3. 静止时间估计模型

像素级静止时间是从一个用重叠技术被均匀地分割成短的剪辑彩色视频序列估计出来的, 以至于前景的代码和停留时间的信息通过修剪变成连续的。

3.1 引导前台模式编码

而相比之下, 背景去除仅指示一个像素是不是前景像素, 我们用多个前景模式标记像素, 使属于人们的不同像素也可以被鉴别出来。

前景像素聚集成 M 模式。每个像素 p 是一个具有 5D 特征量, 即 $I_p = [R_p, G_p, B_p, X_p, Y_p]^T$, 其中 $[R_p, G_p, B_p]$ 和 $[X_p, Y_p]$ 是 RGB 值和的空间坐标。空间位置使得有可能在一个局部区域共享一个模式, 这对小运动有鲁棒性。拥有不同的 RGB 值的行人被

集群化成模式。聚类中心表示作为 $\{\mathbf{d}_1, \dots, \mathbf{d}_M\}$ ($\mathbf{d}_i \in \mathbb{R}^{5 \times 1}, \forall i = 1, \dots, M$), 形成了一个 $5 \times M$ 矩阵 $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_M]$. \mathbf{D} 用移动的平均值来初始化, 并且 M 值因不同的视频剪辑变化。

我们的编码过程中从背景去除结果[11]开始, 其中 $u_p = 0$ 表明 p 是上背景和 $u_p = 1$ 表示前景。这个结果是有争议的。在一个单一的帧错误分类可能导致在静止的时间估计较大的误差。我们的目标是找到一个 M 维编码向量 α_p ($\alpha_p = \{0, 1\}^M$) 对于每个像素 p . α_p 的项只能是 1 或 0, 并且最多一个元件可以是 1, 如果所有条目都为 0, p 在背景上和 $\|\alpha_p\|_1 = 0$. 若第 m 元素是 1, p 属于前台模式 \mathbf{d}_m 并且 $\|\alpha_p\|_1 = 1$.

我们使用 $Q(\mathbf{D}, \alpha)$ 来平衡前景像素和在两种关系下最初背景去除结果偏差的编码错误。它被表示为

$$Q(\mathbf{D}, \alpha) = \sum_{\{p | \|\alpha_p\|_1 = 1\}} \|\mathbf{D}\alpha_p - I_p\|_2^2 + \eta \sum_p (\|\alpha_p\|_1 - u_p)^2. \quad (1)$$

η 是一个参数。再简短的详细说明的话, $Q(\mathbf{D}, \alpha)$ 与其他稀少关系下的最小化 \mathbf{D} 和 α 都将会估计出来。

3.2 稀疏梯度先验

如果一个像素 p 保持与同样的前景标记 α_p 一样, 它的静止时间会增加。由于照明变化, 局部移动, 和闭塞, α 的估算可能有争议。我们的发现是 α_p 在理想的静止的物体而没有上述问题时的变化是很稀疏。在我们据此在标记 α 的估计中加强稀疏先验 $c(\alpha)$ 来抵制争议。它被写为

$$c(\alpha) = \#\{p | \|\partial_{x,t}\alpha_p\|_2 + \|\partial_{y,t}\alpha_p\|_2 \neq 0\}, \quad (2)$$

其中 $\partial_{x,t}$ 和 $\partial_{y,t}$ 为二阶梯度方面到 $x - t$ 和 $y - t$ 的空间生物。# 计算在混合泛音变化数量。

这正是 L_0 梯度最小化问题。Xu 等 [27] 中讨论了类似的问题, 并表明 L_0 规范, 不像 L_1 和 L_2 稀疏, 在解决大量的各种图像的问题具有许多优异的性能。他们为了限制所产生的非零梯度稀疏结构通过在全球范围控制数量取得像样的成绩。我们采用了一套解决程序似于在项目网站一个公开可用的

程序[27]。我们利用二阶梯度在处理时空信息进行修改, 以下详述。

为什么选择二阶梯度? 为了在分割和前景区域的标记中保持平滑, 一个简单的掺入前一阶的梯度沿每个维度也可以使用, 表示为

$$c'(\alpha) = \#\{p | \|\partial_x\alpha_p\|_2 + \|\partial_y\alpha_p\|_2 + \|\partial_t\alpha_p\|_2 \neq 0\}. \quad (3)$$

我们比较这个形式在 (2) 中的公式 以显示在静止时间估计上二阶梯度更有效。

在方程 (3), 在 x , y , 或 t 梯度任何非零值会导致非零 c' , 当计算图 6(c), 静止人生成图 6(c) 所示的结果, 其中所有身体的界限不可避免地产生许多非零值。

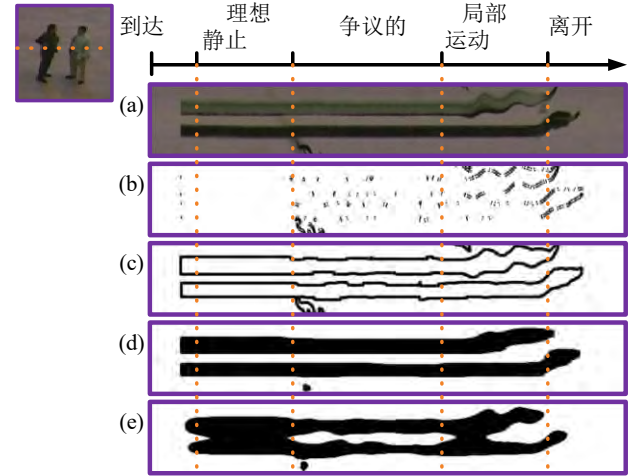


图 6. 通过添加稀疏之前, 我们最好从有争议和/或局部地移动的物体估计 α (a) 两个行人到达, 停留, 局部移动, 并离开阶段 (x 轴: 时间; y 轴: 扫描线的像素橙色高亮显示). (b) 非零 ∂_x 值显示为黑色. (c) 非零梯度 $\partial_x + \partial_t$ 黑色显示. (d) 我们用 $\partial_{x,t}$ 标记的结果. (e) 错误的标记导致 $\partial_x + \partial_t$.

当使用 c' 作为正规化的先验值, 所有这些边界像素将被转正, 这不是我们的本意。

在方程 (2), 中没有这样的问题, 如果对象是静止的为了这些像素; 计算 $\partial_{x,t}$ 时作为空间边界产生的非零梯度将会被淘汰。如 6(d) 和 (e). 所示的两个先验比较我们系统的最终结果。该二阶梯度有效地针对不同的前景区域产生合理的标记。

3.3 联合优化

方程 (1) 和 (2) 代入并统一优化得到

$$\begin{aligned} & \min_{\mathbf{D}, \boldsymbol{\alpha}} \{Q(\mathbf{D}, \boldsymbol{\alpha}) + \lambda c(\boldsymbol{\alpha})\}, \\ & \text{s.t. } \boldsymbol{\alpha}_p = \{0, 1\}^M, \|\boldsymbol{\alpha}_p\|_1 \leq 1. \end{aligned} \quad (4)$$

从数百强度水平 M 个中层语义中产生数据词 $Q(\mathbf{D}, \boldsymbol{\alpha})$ ，它显著简化了局部配套前景像素。该先验值 $c(\boldsymbol{\alpha})$ 对于在时空空间一个静止的物体的每种模式捕捉稀疏结构性。

3.4 Pixel-Wise 静止时间估计

它是容易估计基于前景模式（或标记）的改变的静止时间。如果一个像素的前台模式是从帧 t_1 出发到 \mathbf{d}_i ，将其改变为在框架 t_2 不同的前台模式 \mathbf{d}_j ，或 t_2 框架下的背景，这个像素的静止时间为 $t_2 - t_1$ 。如果像素从背景改变到前景模式，它在局部搜索对于与在其前一帧的像素具有相同模式。如果这样的匹配的象素被发现，它的静止时间，将被当前像素继承，而不是从零开始计数。这避免局部运动引起的低估。

如果帧接近一个视频剪辑，估计边界可能不是同样可靠。我们的系统使用共享缓冲框架的重叠视频剪辑。只有在帧外面的缓冲中估计出的停留时间因为可靠性的缘故而被保持。如果对象比剪辑的持续时间停留地更长，他们的前景模式可以通过剪辑匹配使停留时间可以在堆积中继续。

4.求解

\mathbf{D} 和 $\boldsymbol{\alpha}$ 在方程(4) 联接，问题是高度非凸。我们引进了一组腋向量 $\boldsymbol{\alpha}_p^0 \in \mathbb{R}^M$ 来放宽原来的问题，表现为

$$\begin{aligned} & \min_{\mathbf{D}, \boldsymbol{\alpha}, \boldsymbol{\alpha}^0} \left\{ Q(\mathbf{D}, \boldsymbol{\alpha}) + \beta_1 \sum_p \|\boldsymbol{\alpha}_p - \boldsymbol{\alpha}_p^0\|_2^2 + \lambda c(\boldsymbol{\alpha}) \right\}, \\ & \text{s.t. } \boldsymbol{\alpha}_p = \{0, 1\}^M, \|\boldsymbol{\alpha}_p\|_1 \leq 1, \boldsymbol{\alpha}_p^0 = \{0, 1\}^M, \|\boldsymbol{\alpha}_p^0\|_1 \leq 1. \end{aligned} \quad (5)$$

当 β_1 足够大时， $\boldsymbol{\alpha}_p^0$ 完美地接近 $\boldsymbol{\alpha}_p$ 。它使得原本具有挑战性的问题归结于两大子问题。满意的结果通过每次迭代后反复地增加 β_1 求解这两个优化问题来实现。这种策略被用于[27]。这是相当有效地解决 L_0 梯度最小化的问题。

4.1 求解 \mathbf{D} 和 $\boldsymbol{\alpha}_p^0$

当 $\boldsymbol{\alpha}_p$ 静止，稀疏现有术语是常数并可以因此可省略。我们重写公式(5) 为

$$\begin{aligned} & \min_{\mathbf{D}, \boldsymbol{\alpha}^0} \left\{ Q(\mathbf{D}, \boldsymbol{\alpha}^0) + \beta_1 \sum_p \|\boldsymbol{\alpha}_p - \boldsymbol{\alpha}_p^0\|_2^2 \right\}, \\ & \text{s.t. } \boldsymbol{\alpha}_p^0 = \{0, 1\}^M, \|\boldsymbol{\alpha}_p^0\|_1 \leq 1. \end{aligned} \quad (6)$$

\mathbf{D} 和 $\boldsymbol{\alpha}_p^0$ 的迭代估计。鉴于 $\boldsymbol{\alpha}_p^0$ ，通过求解最小二乘问题得到 \mathbf{D} 。鉴于 \mathbf{D} ， $\boldsymbol{\alpha}_p^0$ 通过搜索前景和背景 ($M + 1$) 的可能性模式得到。

4.2 求解 $\boldsymbol{\alpha}_p$

鉴于 \mathbf{D} 和 $\boldsymbol{\alpha}_p^0$ ，第二优化问题是

$$\min_{\boldsymbol{\alpha}} \left\{ \beta_1 \sum_p \|\boldsymbol{\alpha}_p - \boldsymbol{\alpha}_p^0\|_2^2 + \lambda c(\boldsymbol{\alpha}) \right\}. \quad (7)$$

约束 $\boldsymbol{\alpha}_p = \{0, 1\}^M$ 首先省略，然后当 $\boldsymbol{\alpha}_p$ 收敛时再用临界值添加回来。公式(7) 是非凸。我们进一步采用腋窝向量 \mathbf{h} 和 \mathbf{v} 来接近 $\partial_{x,t} \boldsymbol{\alpha}$ 和 $\partial_{y,t} \boldsymbol{\alpha}$ ，类似公式(5) 的做法。它产生

$$\begin{aligned} & \min_{\boldsymbol{\alpha}, \mathbf{h}, \mathbf{v}} \left\{ \beta_1 \sum_p \|\boldsymbol{\alpha}_p - \boldsymbol{\alpha}_p^0\|_2^2 + \lambda c(\mathbf{h}, \mathbf{v}) \right. \\ & \left. + \beta_2 \sum_p (\|\partial_{x,t} \boldsymbol{\alpha}_p - \mathbf{h}_p\|_2^2 + \|\partial_{y,t} \boldsymbol{\alpha}_p - \mathbf{v}_p\|_2^2) \right\}. \end{aligned} \quad (8)$$

我们再次用两个子优化问题迭代，类似解决方程(5) 来解决方程(8)

估计 (\mathbf{h}, \mathbf{v}) 方程 (8) 简化为

$$\begin{aligned} (\hat{\mathbf{h}}, \hat{\mathbf{v}}) = \arg \min_{\mathbf{h}, \mathbf{v}} & \left\{ \lambda c(\mathbf{h}, \mathbf{v}) + \beta_2 \sum_p \|\partial_{x,t} \boldsymbol{\alpha}_p - \mathbf{h}_p\|_2^2 \right. \\ & \left. + \beta_2 \sum_p \|\partial_{y,t} \boldsymbol{\alpha}_p - \mathbf{v}_p\|_2^2 \right\}. \end{aligned} \quad (9)$$

作为该问题变成 p 的独立，像素的解决方案被产生出来

$$(\hat{\mathbf{h}}_p, \hat{\mathbf{v}}_p) = \begin{cases} (\mathbf{0}, \mathbf{0}) & \text{if } \lambda/\beta_2 \geq \|\partial_{x,t}\alpha_p\|_2^2 + \|\partial_{y,t}\alpha_p\|_2^2 \\ (\partial_{x,t}\alpha_p, \partial_{y,t}\alpha_p) & \text{elsewhere} \end{cases}$$

更大的 λ 意味着结构术语是更重要的，并且更多的非零梯度被设置为零。 β_2 迭代增加。

估计 α 公式 (8) 更新为(10)

$$\hat{\alpha} = \arg \min_{\alpha} \left\{ \beta_1 \sum_p \|\alpha_p - \alpha_p^0\|_2^2 + \beta_2 \sum_p (\|\partial_{x,t}\alpha_p - \mathbf{h}_p\|_2^2 + \|\partial_{y,t}\alpha_p - \mathbf{v}_p\|_2^2) \right\}, \quad (10)$$

这是一个封闭式的二次优化问题的解决方案。求解器的解决方案和相关的证明是被包含在可从该项目的网站下载辅助材料里。

β_1 和 β_2 初始化为 1。他们在迭代中自动增加，如在 [27] 中描述。我们的优化经过 3-5 次迭代迅速收敛。一个例子示在图 7 中的实验 [27] 及此方法对于 β_1 和 β_2 初始化值收敛不敏感

5. 实验及结果

两个数据集被用于评价：一种是大中央火车站的数据集 [32]，另一种是我们收集的数据集。对于每个前景像素，其停留时间到达当前帧是手动注释的。17 帧（带超过 8 万像素）一律从两次采样

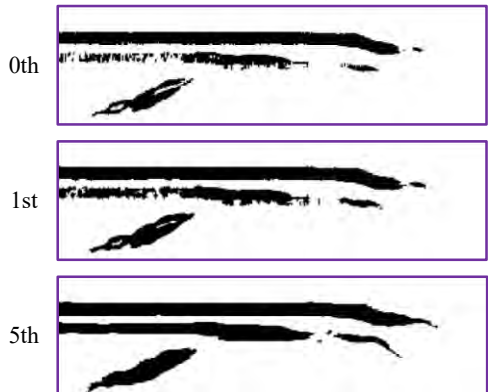


图 7. 在 $x-t$ 平面收敛的说明。在不同的迭代初始估计和更新被示出。争议逐步消除。

数据集注释在像素级别。注释静止时间图的例子示于图 8。这些数据集的详细内容包括在表 1。

每个视频被分割成一分钟的片段。在邻近剪辑之间有 25% 的重叠。我们目前基于英特尔 CPU @ 3.3GHz 的未优化 MATLAB 需要 12 分钟来处理一分钟的视频剪辑。

在方程 (1) 的 η 和方程(4)的 λ 是将被设置的参数。他们的选择取决于初始背景去除结果的质量和稀疏先验的信心。我们经验设定 $\eta = 1.5$ 和 $\lambda = 20$ 。

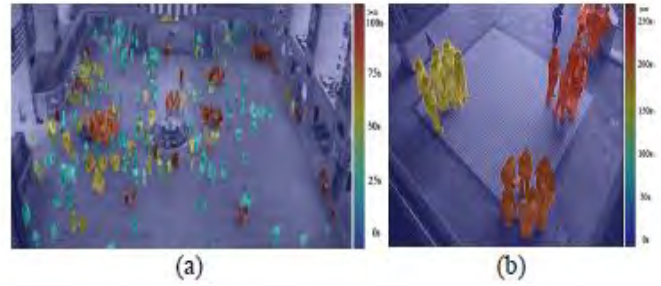


图 8. 注释静止时间地图上的大中央数据集 [32] (a) 和由我们所收集的数据集 (b) 中，看彩图更好。

评估的不同类型在进行中。对所有前台像素平均的静止的时间 (ET) 估计误差被获得。此外，我们计算每个前景像素估计误差和地面实况之间的比率。然后把前景像素所有的比率平均。这项措施表示为在静止的时间 (ERT)，平均估计的错误率。如果像素变得静止的到当前帧时间超过 10 秒，它被认为是静止的。使用的几种检测措施包括 (1) 误报率 (FAR)，(2) 漏检率 (MDR)，和 (3) 的总误差率 (TER)。

我们与几个背景去除的方法的结果进行比较，包括改进的自适应高斯混合模型 [33]，码书模型 [11]，和自适应贝叶斯模型 [21]。如果一个静止的像素被检测为前景静止时间积累。我们在检测到的前景像素 [11] 中测试密集追踪的 [24]。因为一个像素变为前景，静止时间被估计为所述轨迹的长度。我们指出在我们的方法中我们用一阶那些结果 (公式 3)) 代替二阶梯度 (式 (2)) 的结果。这种简化的版本被表示为“我们的 (FOrder)”。

6.应用

我们提出静止组的一些应用和它们的形成时间的估计。

6.1 静止小组活动检测

我们应用我们的方法来检测静止的集体活动。地面中央火车站图像 [32] 包含各个静止集体活动。我们发现四种类型在图 2 所示的活动中，因为它们的人群监视是常见的。地面实况被手动注释在这个数据集里。我们只考虑静止时间长于 30 秒，尺寸大于 2500 像素点的那些群体，因为长时间静止的大集团在监控应用中很容易引起注意。探测器单独训练每种类型的活动，检索整个视频序列。如果在时空空间检测组和地面实况之间的重叠大于 50% 一个真正类就被计数。训练的号码和测试样本总结在表 4 中。

一个好的探测器有三个主要特点。(1) 应该准确估算在静止期间，确定突发事件和静止组的分散活性。(2) 应该成组群集检测静止像素。我们的方法采用集群均值漂移给予我们相当合理像素级的时间估计。(3) 运动的描述表明导致的突发事件，扩散和组变形的特征。我们提出 12 个由 KLT 跟踪器引出基于关键点轨迹静止组描述符 ($\{D_1, \dots, D_{12}\}$)，下面详述。

D_1 - D_4 表征的突发事件的过程中，即成员是否从同一方向在短期内期间加入一组或从多个方向在延长的时期加入一组。传入轨迹随时间 (E_T) 和方向 (E_A) 的直方图被计算出来。两个直方图的占主导地位的模式 (M_T 和 M_A) 是由均值移动获得的。我们设定 $D_1 = |M_T|/|E_T|$ 和 $D_3 = |M_A|/|E_A|$ ，其中 $|\cdot|$ 表示的一个模式或直方图的大小。 D_2 和 D_4 的方差是 E_T 和 E_A 。

同样， D_5 - D_8 表征疏散过程中，即无论是成员在大约同一时间朝同一个方向离开一个组，或在不同时间的多个方向。它们都是基于传出轨迹。 D_9 是一组中心的空间变化和和 D_9 可用于检测动迁组 (移动到另一个局部)。 D_{10} - D_{12} 表示静止组是否保持其内部结构稳定。他们计算静止组中的特征点的拓扑变化。对于组中一个点 i ，其 K -近邻 $N_K(i)$ 和和邻居的拓扑结构趋于稳定。

表 1 数据集的详细信息数据集

	Dataset I [32]	Dataset II
Scene type	Indoor	Outdoor
Video length	3,500 seconds	800 seconds
Frame rate	24 fps	24 fps
Resolution	960×540	768×576
Number of annotated frames	8	9
Number of stationary pixels on the annotated frames	147,930	553,500
Total number of pixels on the annotated frames	4,147,200	3,981,300

表 2 的结果是以秒为单位。

Methods	FAR	MDR	TER	ET	ERT
Ours	0.29%	3.49%	0.39%	10.04	12.21
Ours (FOrder)	0.51%	5.90%	0.69%	16.12	26.77
GMM [33]	0.27%	24.51%	1.11%	29.46	43.98
Codebook [11]	0.26%	21.03%	0.93%	29.51	40.14
Bayesian [21]	0.33%	20.18%	1.01%	26.70	39.16
Tracking [24]	0.30%	24.26%	1.09%	40.78	56.49

表 3 对数据集 II 静止时间估计的结果。

E 是以秒为单位。

Methods	FAR	MDR	TER	ET	ERT
Ours	0.91%	0.54%	0.86%	15.88	8.67%
Ours (FOrder)	1.37%	0.98%	1.32%	16.90	10.68%
GMM [33]	0.92%	16.24%	3.06%	57.41	39.76%
Codebook [11]	1.03%	13.37%	2.75%	58.28	40.67%
Bayesian [21]	1.05%	12.26%	2.60%	45.20	32.19%
Tracking [24]	0.92%	5.75%	1.60%	54.14	38.86%

5.1 结果分析

列于表 2 和 3 两个数据集的结果，总体而言，我们的方法优于所有其他同时在室内和室外的数据集的替代品。由于高稀疏先验产生的平滑效果的原因，其假的报警率比几个方法稍高。其误检率 6-15 倍更低。静止时间估计误差也至少低于其它时间方法 2.5 倍。如果一些阴影不能通过初始背景去除的方法完全去除，可能引起误报。

随着大误检率和估计错误，背景去除是不适用于静止时间估计的。在一般情况下，自适应贝叶斯模型 [6][21] 的作品比其他方法更好，因为它增加了在空间域中和两个连续帧之间的平滑约束。但它仍然没我们的方法好，因为在第 1 和第 2 部分讨论的各种原因，平滑先验会导致比我们更多的误报，这体现了必要采用二阶梯度稀疏先验。

表 4. 活动检测结果 (虚警/误检测)

Activities	Gather	Stop by	Relocate	Deform
Training samples	30	30	30	30
Test samples	45	58	27	50
Ours	3/6	5/6	4/1	6/4
GMM [33]	4/23	6/25	4/9	7/19
Codebook [11]	3/22	4/23	4/8	7/18
Bayesian [21]	2/23	4/24	3/8	6/17
Tracking [24]	4/25	5/28	5/12	6/20

我们使用 $\mu_t(i) = 1 - |N_t(i) \cap N_{t-\Delta}(i)|/K$ 去测量从时间 $t - \Delta$ 到 t . 临近变化的部分。该 K 的临近不变量根据其到 i . 的距离排名。

$R_t(i) = [\sigma_t^1(i), \dots, \sigma_t^{K_0}(i)]$ 和 $R_{t-\Delta}(i) = [\sigma_{t-\Delta}^1(i), \dots, \sigma_{t-\Delta}^{K'}(i)]$ 是近邻中的排名在时间 t 和 $t - \Delta$. $\zeta_t(i)$ 是保留时间 $R_t(i)$ 和 $R_{t-\Delta}(i)$. 之间的距离。同样 $\kappa_t(i)$ 是基于对角的排名中计算出来的。最后, 对基于 $\mu_t(i)$, $\zeta_t(i)$, 和 $\kappa_t(i)$ 的整个静止期间中所有在的特征点的平均值, D_{10} - D_{12} 分别地被计算出来了

给出 D_{10} - D_{12} 了解到, 线性 SVM 被用作分类器。更多细节在我们的辅助材料里。表 4 通过不同的方发报告误报和漏检的数量。所有的方法都使用相同的群集方法和组描述符。它们由前面所述的静止时间估计方法的不同来区分。由于静止时间估计是至关重要的在这个应用程序中, 我们的方法由于它抑制争议的稳健性达到应有的最佳效果。

6.2 现场了解

静止时间估计可以帮助现场了解并提供了计算时间的宝贵数据。例如, 一个平均静止时间图计算出的在四小时大中央火车站视频各组的数据显示在图 3, 它表示其中静止组倾向于出现的地方, 他们一般停留的时间。这些资料对人群管理, 设施设计, 事件监控和交通管制是重要的。一个简单的情况是, 如果静止群体往往出现在大楼的入口, 警报器可以为采取进一步的行动而触发, 以改善那里的交通。

7. 结论和未来工作

我们已经探索出了静止众人分析一个新的研究课题, 其中还有许多重要的应用。获得有用信息的一个基本步骤是估计前景像素的静止时间, 那是一个不能被现有的背景去除技术解决的问题。我们

提出了一个强大的算法, 采用在局部共享前景码本, 并使用二阶梯度塑造 3D 静止时间图。它被表述为 L_0 最小化问题, 并且通过切实有效的方案而解决。我们所提出的方法, 以及作为研究的课题, 可以应用到检测静止小组活动和群众场面的理解。

我们相信它会发现许多有趣和有价值应用前景。例如, 它可以被掺入到现有系统来模仿不断变化的移动交通静止组的影响和预测行人之间的社会关系。研究这种问题和部署我们的解决方案的潜力是无穷的。

致谢

支持这项工作的香港研究资助局主办的综合研究基金

(项目编号 417110, 417011, 429412, 413113 和)。

参考

- [1] S. 阿里和 M. 沙阿。拉格朗日粒子动态方法针对流动人群的细分和稳定性分析。在 CVPR, 2007 年。
- [2] M. R. 阿米尔和 S. 托多罗维奇。用于定位链路模式视频中的小组活动的参与者。在 ICCV, 2011。
- [3] E. J. Candès, X. 李, Y. 马, 和 J. 赖特。强大的主成分分析? 期刊 ACM, 58 (3), 2011。
- [4] A. B. 陈 Z. S. 梁和 N. 洛斯。隐私保护人群监测: 计数人没有人模型或跟踪。在 CVPR 2008。
- [5] M. 张, N. Krahnstoeber, ANDW. 戈。概率组新手上路运动分析和情景识别。在 ICCV, 2011。
- [6] K. 陈, 龚 S., T. 翔, 和 C. C. 洛伊。累积属性空间年龄和人群密度的估计。在 CVPR, 2013。
- [7] M. Cristani, L. 巴扎尼, G. Paggetti, A. 福萨蒂, D. Tosato, A. 德尔武卫, G. Menegaz 和 V. 穆里诺。社交发现由 F-编队统计分析。在 监视中心, 2011。
- [8] Y. 富, M. Hopedales, T. 翔和 S. 宫。属性学习理解非结构化的社交活动。在 ECCV, 2012。
- [9] W. 戈, R. T. 科林斯和 R. B. Ruback。基于视觉的分析在步行的人群小团体。IEEE 跨。上 PAMI, 34: 1003 年至 1016 年, 2012。
- [10] 一 Haritaoglu 和 M. Flickner。检测和跟踪购物商店组。在 CVPR, 2001 年。
- [11] K. 金, TH Chalidabhongse, D. 哈伍德和 L. 戴维斯。使用码本的实时前景, 背景分割模型。实时成像, 11: 172-185, 2005 年。
- [12] T. 兰, 王勇, 王 W., SN Robinovitch 和 G 森。判别潜在的车型识别上下文

- [13] 集体活动。IEEE 跨。在 PAMI, 34: 1549 年至 1562 年, 2012。
- [14] C. C. 洛伊, T. 翔和 S. 宫。增量建模活动在多个不相交的摄像头。IEEE 跨。在 PAMI, 34: 1799 年至 1813 年, 2012。五马哈德, X. 李五 Bhalodia 和 N. 洛斯。异常检测在拥挤的场面。在 CVPR, 2010。
- [15] R. 迈赫兰, A. 小山, 和 M·沙阿。异常人群行为检测利用社会力模型。在 CVPR, 2009 年。
- [16] M. Moussaid, N. Perozo, S. 卡尼尔, 黑尔宾 D. 和 G. Theraulaz。行人社会的行走行为团体及其对人群的动态影响。公共科学图书馆 ONE, 5, 2010。
- [17] S. 佩莱格里尼, A. Ess 来, K. 辛德勒和 L. 面包车 GOOL。你会永远不会独行建模多目标的社会行为跟踪。在 ICCV, 2009 年。
- [18] M·罗德里格斯, J. Sivic, 一拉普捷夫海和 J. AUDIBERT。数据驱动人群分析视频。在 CVPR, 2011。
- [19] P. Scovanner 和 M. F. 塔彭。学习行人动态从现实世界。在 ICCV, 2009 年。
- [20] J. 邵 C. C. 洛伊和十王。场景组独立分析人群。在 CVPR, 2014 年。
- [21] Y. 谢赫和 M·沙阿。动态贝叶斯模型场景检测对象。模式分析与机器情报, 在 IEEE 交易, 27 (11): 1778-1792 年, 2005 年。
- [22] B. 索尔马兹, B. Moore 和 M·沙阿。识别行为使用稳定性分析动力系统的人群的场面。IEEE 跨。在 PAMI, 34: 2064 至 70 年, 2012。
- [23] C. 斯托弗 ANDW. E. L. Grimson。自适应背景混合模型的实时跟踪。在 CVPR 1999。
- [24] H. Wang, A. Klaser, C·施密德和 C.-L. 刘。动作识别通过密集的轨迹。在 CVPR, 2011。
- [25] 王十, 十马, 和 E. Grimson。无监督活动通过感知分层贝叶斯模型。在 CVPR, 2007 年。
- [26] 王十, 十马, 和 E. Grimson。无监督活动在感知采用分层拥挤, 复杂的场景贝叶斯模型。IEEE 跨。在 PAMI, 31: 539-555, 2009 年。
- [27] L. 徐, 陆 C., Y. 徐和 J. 佳。通过 10 图像平滑梯度最小化。ACM 跨。显卡, 30, 2011。
- [28] K. 山口, AC 伯格, T. 伯格和 L. 奥尔蒂斯。谁是你和你要去哪里。在 ICCV, 2011。
- [29] Z. 钟, W. 焯, 王南, 杨米和 Y. 许。人群能源和特征分析。在 ICIT, 2007 年。
- [30] B. 周, 唐 X. 和 X. Wang。相干过滤: 检测连贯的动作, 从人群的杂波。在 ECCV, 页 857-871. 2012。
- [31] B. 周, 唐 X. 和十王。测量的人群集体主义。在 CVPR, 2013。
- [32] B. 周, 王 X., 和 X. 汤。了解集体人群行为: 学习动力的混合模式行人代理商。在 CVPR, 2012。
- [33] [Z. 日夫科维奇。改进的自适应混合高斯模型背景扣除。在 ICPR, 2004 年