

基于监控视频的前景目标提取

视频监控是中国安防产业中最为重要的信息获取手段。随着“平安城市”建设的顺利开展，各地普遍安装监控摄像头，利用大范围监控视频的信息，应对安防等领域存在的问题。近年来，中国各省市县乡的摄像头数目呈现井喷式增长，大量企业、部门甚至实现了监控视频的全方位覆盖。如北京、上海、杭州监控摄像头分布密度约分别为 71、158、130 个/平方公里，摄像头数量分别达到 115 万、100 万、40 万，为我们提供了丰富、海量的监控视频信息。

目前，监控视频信息的自动处理与预测在信息科学、计算机视觉、机器学习、模式识别等多个领域中受到极大的关注。而如何有效、快速抽取出监控视频中的前景目标信息，是其中非常重要而基础的问题[1-6]。这一问题的难度在于，需要有效分离出移动前景目标的视频往往具有复杂、多变、动态的背景[7, 8]。这一技术往往能够对一般的视频处理任务提供有效的辅助。以筛选与跟踪夜晚时罪犯这一应用为例：若能够预先提取视频前景目标，判断出哪些视频并未包含移动前景目标，并事先从公安人员的辨识范围中排除；而对于剩下包含了移动目标的视频，只需辨识排除了背景干扰的纯粹前景，对比度显著，肉眼更易辨识。因此，这一技术已被广泛应用于视频目标追踪，城市交通检测，长时场景监测，视频动作捕捉，视频压缩等应用中。

下面简单介绍一下视频的存储格式与基本操作方法。一个视频由很多帧的图片构成，当逐帧播放这些图片时，类似放电影形成连续动态的视频效果。从数学表达上来看，存储于计算机中的视频，可理解为一个 3 维数据 $X \in \mathbb{R}^{w \times h \times t}$ ，其中 w, h 代表视频帧的长，宽， t 代表视频帧的帧数。视频也可等价理解为逐帧图片的集合，即 $X = \{x_1, x_2, \dots, x_t\}$ ，其中 $x_i \in \mathbb{R}^{w \times h} (i = 1, 2, \dots, t)$ 为一张长宽分别为 w, h 的图片。3 维矩阵的每个元素（代表各帧灰度图上每个像素的明暗程度）为 0 到 255 之间的某一个值，越接近 0，像素越黑暗；越接近 255，像素越明亮。通常对灰度值预先进行归一化处理（即将矩阵所有元素除以 255），可将其近似认为 $[0, 1]$ 区间的某一实数取值，从而方便数据处理。一张彩色图片由 R（红），G（绿），B（蓝）三个通道信息构成，每个通道均为同样长宽的一张灰度图。由彩色图片

构成的视频即为彩色视频。本问题中，可仅考虑黑白图片构成的视频。在 **Matlab** 环境下，视频的读取、播放及相应基本操作程序见附件 1。如采用其他编程环境，也可查阅相关资料获得相应操作程序。

题目的监控视频主要由固定位置监控摄像头拍摄，要解决的问题为提取视频前景目标。请研究生通过设计有效的模型与方法，自动从视频中分离前景目标。注意此类视频的特点是相对于前景目标，背景结构较稳定，变化幅度较小，可充分利用该信息实现模型与算法设计。

请你们查阅相关资料和数据，结合视频数据特点，回答下列问题：

问题 1： 对一个不包含动态背景、摄像头稳定拍摄时间大约 5 秒的监控视频，构造提取前景目标（如人、车、动物等）的数学模型，并对该模型设计有效的求解方法，从而实现类似图 1 的应用效果。（附件 2 提供了一些符合此类特征的监控视频）

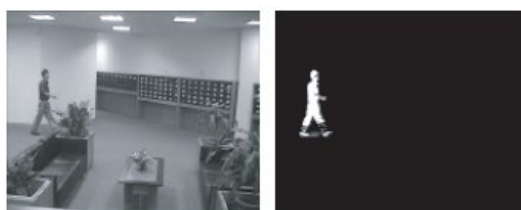


图 1 左图：原视频帧；右图：分离出的前景目标

问题 2： 对包含动态背景信息的监控视频（如图 2 所示），设计有效的前景目标提取方案。（附件 2 中提供了一些符合此类特征的典型监控视频）



图 2 几种典型的动态视频背景：，树叶摇动，水波动，喷泉变化，窗帘晃动

问题 3： 在监控视频中，当监控摄像头发生晃动或偏移时，视频也会发生短暂的抖动现象（该类视频变换在短时间内可近似视为一种线性仿射变换，如旋转、平移、尺度变化等）。对这种类型的视频，如何有效地提取前景目标？（附件 2 中提供了一些符合此类特征的典型监控视频，其它一些典型视频可从

<http://wordpress-jodoin.dmi.usherb.ca/dataset2014/>下载)

问题 4: 在附件 3 中提供了 8 组视频 (avi 文件与 mat 文件内容相同)。请利用你们所构造的建模方法, 从每组视频中选出包含显著前景目标的视频帧标号, 并将其在建模论文正文中独立成段表示。务须注明前景目标是出现于哪一个视频 (如 Campus 视频) 的哪些帧 (如 241-250, 421-432 帧)。

问题 5: 如何通过从不同角度同时拍摄的近似同一地点的多个监控视频中 (如图 3 所示) 有效检测和提取视频前景目标? 请充分考虑并利用多个角度视频的前景之间 (或背景之间) 相关性信息 (一些典型视频可从

<http://cvlab.epfl.ch/research/surv/multi-people-tracking> 下载)



图 3 在室内同一时间从不同角度拍摄同一地点获得的视频帧

问题 6: 利用所获取前景目标信息, 能否自动判断监控视频中有人群短时聚集、人群惊慌逃散、群体规律性变化 (如跳舞、列队排练等)、物体爆炸、建筑物倒塌等异常事件? 可考虑的特征信息包括前景目标奔跑的线性变化形态特征、前景规律性变化的周期性特征等。尝试对更多的异常事件类型, 设计相应的事件检测方案。(请从网络下载包含各种事件的监控视频进行算法验证)

注: 强烈建议深刻考虑问题内涵, 建造合理、高效的数学模型和求解方法, 鼓励进行具有开放思路与创新思维的探索性尝试。

参考文献:

[1] Andrews Sobral & Antoine Vacavant, A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos, Computer Vision and Image Understanding, Volume 122, May 2014, Pages 4-21

[2] B. Lee and M. Hedley, "Background estimation for video surveillance,"

IVCNZ02, pp. 315–320, 2002.

[3] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in *Computer Vision and Pattern Recognition*, 1999. IEEE Computer Society Conference on., vol. 2. IEEE, 1999.

[4] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[5] D. Meng and F. De la Torre, “Robust matrix factorization with unknown noise,” in *IEEE International Conference on Computer Vision*, 2013, pp. 1337–1344.

[6] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, “Robust principal component analysis with complex noise,” in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, 2014, pp. 55–63.

[7] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, “RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2233–2246, 2012.

[8] M. Babaei, D. T. Dinh, and G. Rigoll, “A deep convolutional neural network for background subtraction,” *arXiv preprint arXiv: 1702.01731*, 2017.