

参赛密码 _____

(由组委会填写)



“华为杯”第十四届中国研究生 数学建模竞赛

题 目 基于监控视频的前景目标提取

摘 要：

通常，在计算机视觉领域中，研究者将每一帧图像中的显著运动部分称为前景，而把当中静止的部分称为背景。本文针对基于监控视频的前景目标提取问题，使用了前景背景分离模型、模糊更新模型、特征点轨迹消抖模型、目标存在概率场模型、瞬时能量模型等并结合随机样本聚类，Mean-Shift 等方法，依托 OpenCV 开源库，利用 matlab 和 VC 编写算法，分别实现在静态背景下，动态背景下，摄像镜头晃动情况下，多摄像头视角下的前景目标提取，以及视频中的异常事件检测。

针对问题一——静态背景下的前景提取问题，我们采用基于样本随机聚类的方法构建背景模型，并基于该模型实现前景背景分离。由于问题一中视频较短，无法提供足够长度的历史数据初始化模型，因此采用基于单帧数据的模型初始化，利用少量数据快速构建背景模型。此外为了能够持续的将背景和前景进行分离，我们还对背景模型进行持续的更新。

针对问题二——动态背景下的前景提取问题,我们针对动态场景中部分物体存在扰动的特点,在问题一方法的基础上结合模糊更新准则,自适应的改变背景模型的检测阈值和更新频率,进而提高动态场景中前景提取的准确性和鲁棒性。

针对问题三——抖动背景下的前景提取问题,我们结合视频中前景、背景目标的运动量具有显著区别的特点,利用历史数据构建特征点轨迹模型,有效对视频的前背景进行区分,并根据捕获轨迹的抖动量,进行抖动补偿,构建视频消抖模型,降低视频抖动对前景提取带来的不良影响。

针对问题四,我们利用所构造的模型对八组视频进行前景提取,取得了很好的效果,模型的平均前景提取精确度达到了 90% 以上。

针对问题五——多视角的运动目标检测问题,我们基于目标存在概率场模型进行多视角的运动目标检测,根据成像原理构建目标窗结构,进而结合目标窗结构的子窗在各视角的图像中包含的图像区域,计算出目标在各个位置存在的概率,获得概率图,然后利用 **Mean-Shift** 算法搜索出局部概率最大的点,该点即为多视角场景下运动目标最可能出现的位置。

针对问题六——视频中的异常事件检测问题,我们首先对视频图像进行网格化处理,并在网格单元内提取特征点,进而基于特征点利用光流法构建运动模型。最后根据动能公式,将运动模型转换成能量模型,并根据视频中计算出的能量值检测出视频中是否出现异常事件。

目 录

1 问题重述	9
1.1 问题背景	9
1.2 需要解决的问题	9
2 问题一的建模与求解	10
2.1 问题分析	10
2.2 模型建立	10
2.3 基于样本随机聚类的背景模型构建方法	10
2.4 基于单帧数据的模型初始化	11
2.5 背景模型更新	12
2.6 模型工作流程	13
2.7 问题求解及分析	13
3 问题二的建模与求解	15
3.1 问题分析	15
3.2 模型初始化	16
3.3 基于样本随机聚类的自适应匹配模型	16
3.4 背景模型更新	17
3.5 模糊更新准则	18
3.6 模型工作过程	18
3.7 问题求解及分析	18
4 问题三的建模与求解	19
4.1 问题分析	19
4.2 特征点轨迹模型	19
4.3 前景特征匹配模型	20
4.3.1 单应矩阵重映射	20
4.3.2 基于单应矩阵重映射的前景轨迹识别	21

4.3.3 基于历史行为的单应矩阵前景轨迹识别模型	22
4.4 视频消抖模型	22
4.5 模型工作流程	24
4.6 问题求解及分析	24
5 问题四（8 组视频分组标注）	25
5.1 CAMPUS 视频标注及分析	25
5.2 CURTAIN 视频标注及分析	27
5.3 ESCALATOR 视频标注及分析	28
5.4 FOUNTAIN 视频标注及分析	29
5.5 HALL 视频标注及分析	30
5.6 LOBBY 视频标注及分析	31
5.7 OFFICE 视频标注及分析	32
5.8 OVERPASS 视频标注及分析	33
5.9 综合分析	34
6 问题五的建模与求解	35
6.1 问题分析	35
6.2 基于目标存在概率场模型的多视角前景目标提取算法	35
6.2.1 目标存在概率场模型	35
6.2.2 目标窗结构模型	36
6.2.3 计算目标存在概率	36
6.2.4 多视角前景目标提取算法流程	37
6.3 模型工作流程	38
6.4 实验分析	38
7 问题六的建模与求解	39
7.1 问题分析	39
7.2 基于光流能量模型的异常事件检测方法	39
7.2.1 基于像素统计分布的运动特征点提取	39
7.2.2 基于光流法的运动模型构建	40
7.2.3 基于瞬时能量模型的异常事件检测方法	42
7.3 模型工作流程	44

7.4	实验分析	44
8	模型评价	45
9	参考文献	46
10	附录	47

图 录

图 1 像素 $v(x)$ 的分类过程.....	11
图 2 像素 $v(x)$ 的 8 个相邻域.....	12
图 3 静态背景下前景提取方法流程图.....	13
图 4 视频 Hall 测试效果.....	14
图 5 视频 Office 测试效果	14
图 6 视频 Pedestrian 测试效果.....	14
图 7 视频 Smoke 测试效果	15
图 8 视频 Airport 测试效果.....	15
图 9 动态背景下前景提取流程图.....	18
图 10 视频 WaterSurface 测试效果.....	19
图 11 基于历史行为的单应矩阵前景轨迹识别模型的工作流程.....	22
图 12 背景平滑示意图.....	23
图 13 抖动背景下的前景提取方法流程图.....	24
图 14 视频 Car6 测试效果.....	24
图 15 视频 Car7 测试效果.....	24
图 16 视频 People1 测试效果.....	25
图 17 视频 People2 测试效果.....	25
图 18 Campus 视频中插入帧的内容	27
图 19 Curtain 视频中插入帧的内容.....	28
图 20 Escalator 视频中插入帧的内容.....	29
图 21 Fountain 视频中插入帧的内容	30
图 22 Hall 视频中插入帧的内容.....	31
图 23 Lobby 视频中插入帧的内容	32
图 24 Office 视频中插入帧的内容	33
图 25 Overpass 视频中插入帧的内容.....	34
图 26 目标存在概率场示意图.....	36
图 27 多视角前景目标提取方法流程图.....	38
图 28 多视角前景提取与匹配.....	39
图 29 特征点分布图.....	40
图 30 异常事件检测方法流程图.....	44
图 31 异常事件检测.....	45

表 录

表 1 Campus 视频中实际出现前景的帧数	26
表 2 Campus 视频中检测出前景的帧数	26
表 3 Campus 视频中出现插入帧的帧数	26
表 4 Curtain 视频中实际出现前景的帧数.....	27
表 5 Curtain 视频中检测出前景的帧数.....	27
表 6 Curtain 视频中出现插入帧的帧数.....	27
表 7 Escalator 视频中实际出现前景的帧数.....	28
表 8 Escalator 视频中检测出前景的帧数.....	28
表 9 Escalator 视频中出现插入帧的帧数.....	28
表 10 Fountain 视频中实际出现前景的帧数	29
表 11 Fountain 视频中检测出前景的帧数	29
表 12 Fountain 视频中出现插入帧的帧数	29
表 13 Hall 视频中实际出现前景的帧数.....	30
表 14 Hall 视频中检测出前景的帧数.....	30
表 15 Hall 视频中出现插入帧的帧数.....	31
表 16 Lobby 视频中实际出现前景的帧数	31
表 17 Lobby 视频中检测出前景的帧数	31
表 18 Lobby 视频中出现插入帧的帧数	32
表 19 Office 视频中实际出现前景的帧数	32
表 20 Office 视频中检测出前景的帧数	32
表 21 Office 视频中出现插入帧的帧数	32
表 22 Overpass 视频中实际出现前景的帧数.....	33
表 23 Overpass 视频中检测出前景的帧数.....	33
表 24 Overpass 视频中出现插入帧的帧数.....	33
表 25 8 组视频前景匹配精确度.....	34

1 问题重述

1.1 问题背景

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。

1.2 需要解决的问题

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。

问题 2：对包含动态背景信息的监控视频，设计有效的前景目标提取方案。

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？

问题 4：对于提供的 8 组视频，利用构造的建模方法，从每组视频中选出包含显著前景目标的视频帧标号，并将其独立成段表示，并注明前景目标是出现于哪一个视频。

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。

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。

2 问题一的建模与求解

2.1 问题分析

问题一要求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，提取出前景目标。解决本问题的思路为利用视频帧间，前景目标的位置差异作为依据进行前景目标的运动建模，并以运动模型为基础将前景目标和背景进行分离。背景差分法是一种常用的前景提取方法，但是通常情况背景差分法都是以概率密度函数为基础。针对特定的颜色空间，如何选择与之相对应的概率密度函数并没有一个准确地标准。在本方法中，我们将视频中的前景提取问题转换成了像素分类问题，通过对像素集合进行对比的方法来代替像素统计模型的分析。此外背景模型的构建通常需要一个过程，但是对于时长较短的视频，如何利用少量样本帧快速构建背景模型也是我们需要解决的问题。

2.2 模型建立

针对静态背景下的前景目标提取问题，我们采用基于样本随机聚类的方法构建背景模型，然后在此模型的基础上设计了基于像素级别的背景差分方法来进行静态背景条件下的运动目标检测算法，将前景目标与背景进行分离。该方法的基本思想总结为如下三个步骤：

1. 构建背景模型以及前后景的分离。在该算法中，在每一个像素位置处都需要建立一个包含 N 个背景样本的集合，然后在欧式灰度空间中根据当前像素与背景样本集合的欧式距离来对像素点进行分类，即该像素点是前景还是背景。
2. 利用单幅帧图像中的数据对背景模型进行初始化。
3. 更新背景模型。在本章基于像素级别的背景差分方法中，利用图像中相邻像素在在时间和空间上分布一致性原则对邻域像素进行实时更新，不断完善背景模型

2.3 基于样本随机聚类的背景模型构建方法

在基于样本随机聚类的背景模型构建方法^[10]中，背景模型中的每一个像素都是由 N 个背景样本组成，这些背景样本可以从历史帧中获得。用 $v(x)$ 表示在欧式空间中处于位置 x 处的像素， v_i 表示 $v(x)$ 的背景样本集合中索引号为 i 的样本，则 $v(x)$ 的背景模型 $M(x)$ 可由式（2-1）所示，

$$M(x) = \{v_1, v_2, \dots, v_N\} \quad (2-1)$$

然后对像素 $v(x)$ 的分类过程如图 1 所示，在构建的背景模型的样本灰度空间中，定义一个以 $v(x)$ 为球心，半径为 R 的球 $S_R(v(x))$ 。当定义的球体空间中包含的背景样本数量小于给定阈值 $\#min$ ，则 $v(x)$ 将被判决为前景，否则， $v(x)$ 被判决为背景，判别表达式如式 (2-2) 所示

$$v(x) = \begin{cases} background & \text{if } \# \{S_R(v(x)) \cap \{v_1, v_2, \dots, v_N\}\} \geq \#min \\ foreground & \text{otherwise} \end{cases} \quad (2-2)$$

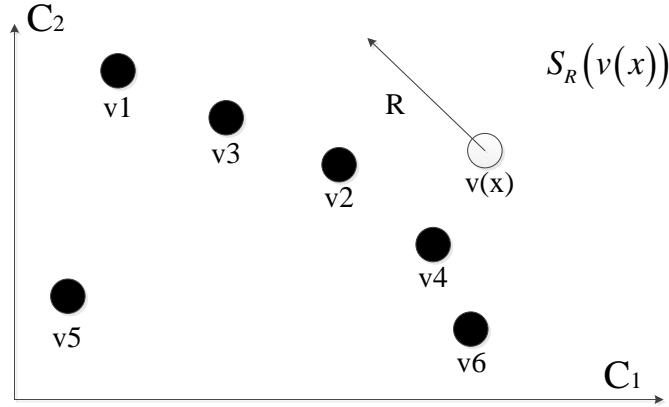


图 1 像素 $v(x)$ 的分类过程

根据式 (2-2) 可看出，阈值 $\#min$ 和球体半径 R 是影响判决的两个重要参数。在本章针对静态背景下的前景目标提取问题中，综合考虑算法效率和提取准确度，将 $\#min$ 设为 2，将 R 设为 20。

2.4 基于单帧数据的模型初始化

现阶段的运动目标检测算法往往需要几十帧图像数据来初始化背景模型，但是在本节问题中，由于视频较短，能够利用的视频数据有限，因此需要利用少量的帧来快速构建背景模型。

利用单帧数据来初始化背景模型存在一个问题，就是单帧数据不包含时间信息。因此我们假设相邻像素在时间和空间上具有一致性特征，这样利用空间上的一致性来弥补时间信息的缺失。以此假设为基础，我们选择像素的邻域内的像素值作为该像素的背景模型样本。如图 2 所示，我们在像素点的周围的八个连通域中随机选择样本。在第一帧中，从 $v(x)$ 的相邻域 $N_G(x)$ 中多次抽取 N 个像素值，然后放入对应的背景模型的 N 个样本中，如式 (2-3) 所示，我们定义单帧背景

模型为 $M^0(x)$,

$$M^0(x) = \{v^0(y | y \in N_G(x))\} \quad (2-3)$$

$V_G(1)$	$V_G(2)$	$V_G(3)$
$V_G(4)$	$V(x)$	$V_G(5)$
$V_G(6)$	$V_G(7)$	$V_G(8)$

图 2 像素 $v(x)$ 的 8 个相邻域

2.5 背景模型更新

为了能持续的将运动前景和背景分离，需要对背景模型进行在线更新。背景模型更新其实就是利用当前处理过的图像中的背景样本来替换历史背景样本。

模型更新策略通常可以分为两大类，保守型更新和激进型更新。保守型更新方式是在图像帧处理过程，只有那些被判决为背景的像素点才用于模型更新。虽然从原理上来讲，保守更新方式较为合理，但是在实际应用过程中，会出现误检的情况，即把那些本应该是前景的像素错误的判决为背景像素，这个时候如果采用的是保守型更新，就会导致模型污染，从而对以后的像素分类产生不良影响。因此这种保守型的更新方式会对前景遮挡较为敏感。

由于是静态背景下的前景提取问题，因此我们采用保守型的更新方式，依据邻域像素在时间空间上分布一致性假设，通过扩展空间信息来弥补保守型模型更新方式的缺陷。我们采用的更新方式主要包括三个部分：无记忆更新，扩大更新窗口以及邻域内随机样本更新。

1) 无记忆更新。更新方式需要满足两个条件，一个是背景样本的生存时间应满足单调递减的条件，并且还应该保证较早出现的样本像素不会被无缘由代替。我们采用的随机样本更新策略，假设从 t 时刻开始进行背景模型更新，那么此时背景样本将以 $\frac{(N-1)}{N}$ 的概率被保留下来。这样随着时间的推移，当时刻为 $t+dt$ 时，可以近似算出该样本大致的概率分布，其概率函数为：

$$P(t, t+dt) = \left(\frac{N-1}{N} \right)^{(t+dt)-t} \quad (2-4)$$

公式 (2-4) 也可以表示为：

$$P(t, t+dt) = e^{-\ln\left(\frac{N-1}{N}\right)dt} \quad (2-5)$$

式(2-5)中的背景模型的样本保留时间呈现单调递减的趋势,并且样本的概率密度与时间点 t 无关,因此历史样本对于未来的前景目标提取是没有影响的,正是由于这个原因,这种更新方式被称作无记忆更新模式。

2) 扩大更新窗口。视频流具有两个特点,连续性和周期性。正是由于这两个特点,如果在每处理一个新的图像帧的时候都进行背景模型的更新操作,那么将会带来频繁的重复运算。为了降低运算复杂度,我们适当的提高了背景样本的生命周期,这样就可以降低背景模型的更新频率。我们这里定义了时间重采样参数 ϕ ($\phi=16$),这个参数表示在更新背景模型的过程中,每个背景像素有 $1/16$ 的概率被放入背景模型中。

3) 邻域随机样本更新。由于采用的保守背景模型更新策略对前景遮挡问题较为敏感,为了弥补该缺陷,我们在该像素的8个相邻域内随机选择一个像素进行更新,这种更新方式可以一定程度上缓解前景遮挡问题,让前景遮挡的背景像素持续渐进的更新进背景模型中。

2.6 模型工作流程

静态背景下前景提取模型工作流程如图3所示:

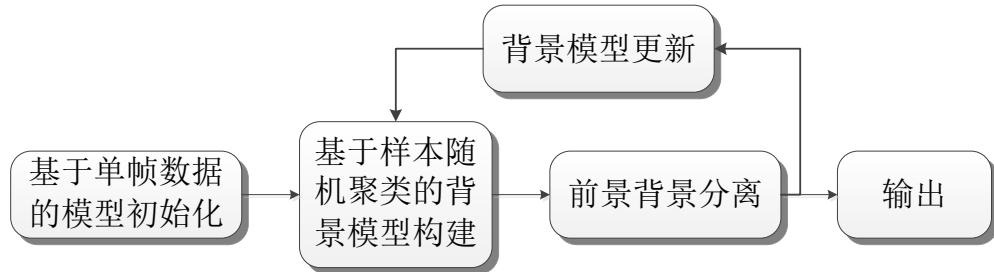


图3 静态背景下前景提取方法流程图

2.7 问题求解及分析

我们利用附件中的视频(Airport, Hall, Office, Pedestrian, Smoke)对本文算法进行了验证,其效果分别下图所示。

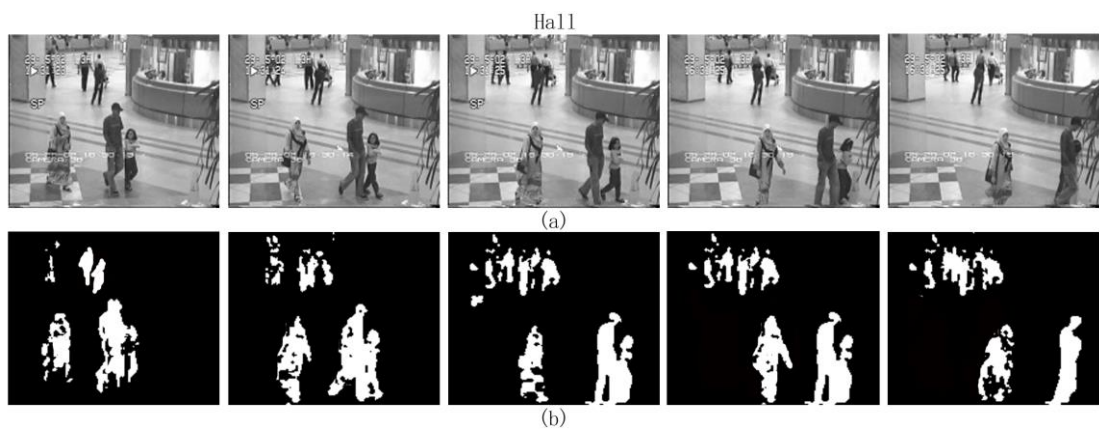


图 4 视频 Hall 测试效果

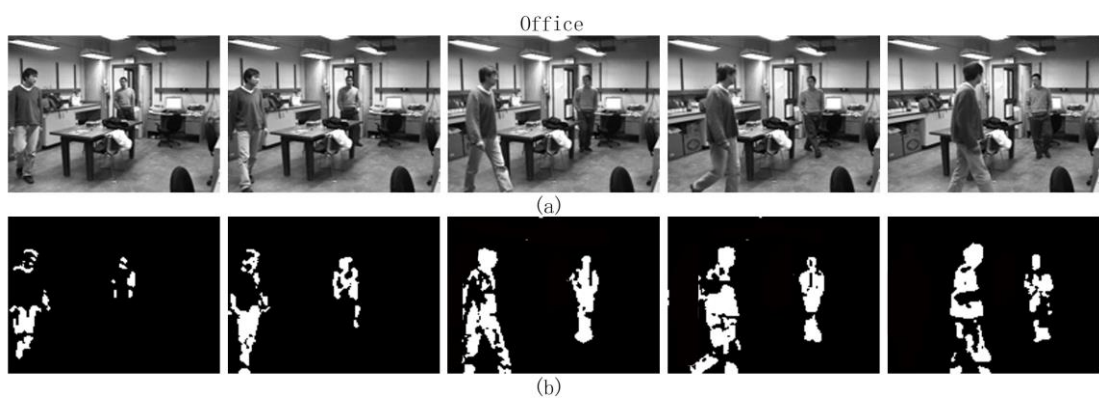


图 5 视频 Office 测试效果

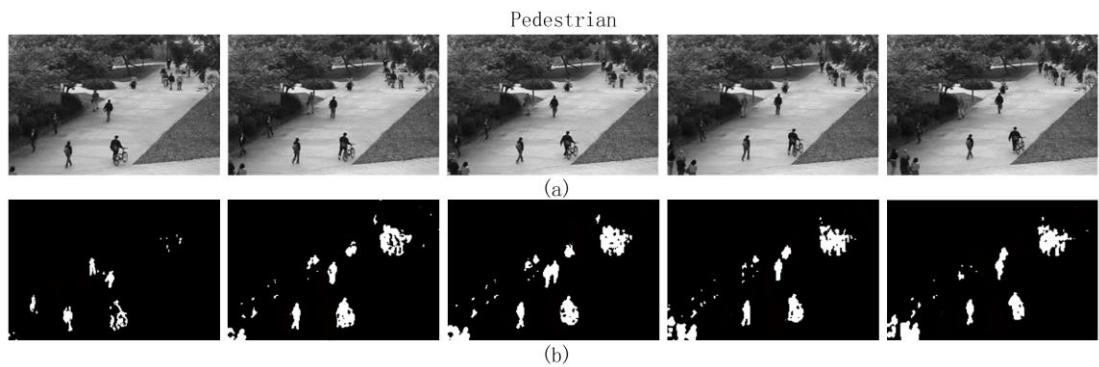


图 6 视频 Pedestrian 测试效果

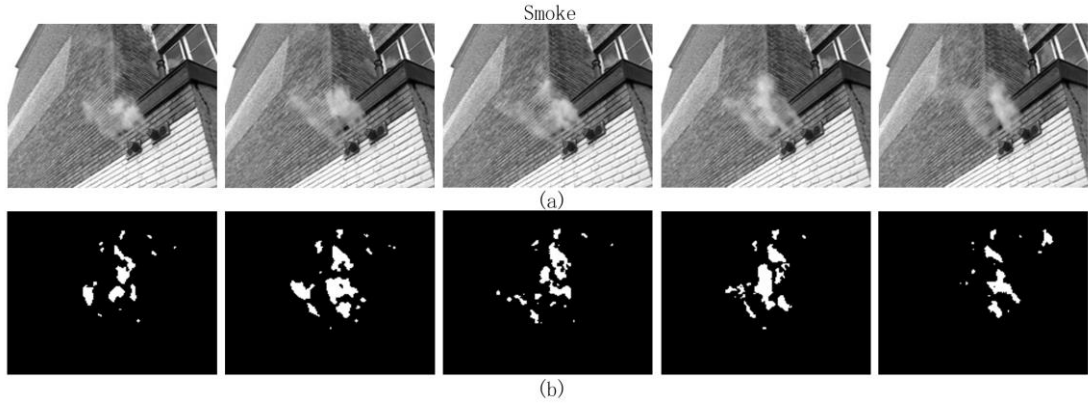


图 7 视频 Smoke 测试效果

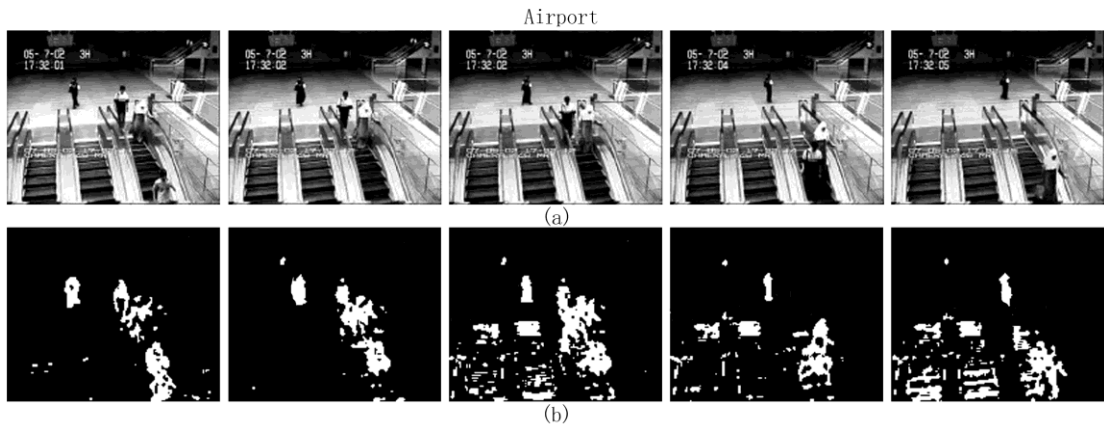


图 8 视频 Airport 测试效果

从这几组图片中可以看出，针对静态背景的视频，本章方法可以有效的将前景进行分离。利用样本随机聚类的方法，构建背景模型。由于采用了单帧数据的模型初始化，所以在视频的开始几帧就能取得较好的前景提取效果。并且随着新图像帧的处理，背景模型也会持续的进行更新，所以前景分离效果越来越好。

但是针对视频 Airport 的前景提取效果并不是非常理想，这是由于本章方法主要针对的是静态背景下的前景提取问题，但是在视频 Airport 中，电梯的移动导致背景出现变化，从而本章方法误将运动的电梯提取出来。下一章我们将对动态背景下的前景提取问题进行建模解决。

3 问题二的建模与求解

3.1 问题分析

通常，在一个拥有静态背景的视频中，相邻的像素点具有相似的时空分布。基于此，基于静态背景的样本随机聚类的建模算法可以通过选择每个像素点相邻的 N 个像素值作为该像素点的背景样本，进而分析出视频的前景。但该方法具

有帧差法中常见的“鬼影”现象，即当前景移走后，上一时刻的前景很可能被误检为前景目标。在分析正常的人类行走或汽车移动时，该方法产生的“鬼影”可以迅速消失；但是当视频中存在持续小波动的移动目标，诸如：水面的波纹、晃动的树叶时，“鬼影”会持续存在，问题一“airport”视频中由于运行的电梯而产生的白色区域即是这样的一种现象。所以在本节中，我们将提出一种可以在动态背景中有效提取前景的样本随机聚类方法，该方法可以利用模糊理论，自适应的改变模型的检测阈值和更新频率，进而有效的减少场景中由于部分物体小规模扰动产生的“鬼影”现象。

3.2 模型初始化

考虑到动态场景中，水波、树叶等持续运动的物体在全部时刻的扰动范围都很小。所以在我们的初始化样本集时，我们抽取任一像素点 x 的前 N 帧图片的像素值初始化我们的背景模型 $M(x)$ ：

$$M(x) = \{v_1, \dots, v_i, \dots, v_N\} = \{I_{t-(N-1)}(x), \dots, I_t(x)\} \quad (3-1)$$

式中， v_i 表示我们取的前 i 个背景的像素值， $i=1, 2, 3, \dots, N$ ， $I_t(x)$ 为 t 时刻的像素值。

3.3 基于样本随机聚类的自适应匹配模型

模型初始化后，则前景的二值图 $F(x)$ 可以表示为：

$$F(x) = \begin{cases} 1 & \text{if } \# \{M(x) \cap S_R(I_t(x))\} < \#_{min} \\ 0 & \text{else} \end{cases} \quad (3-2)$$

其中 $\# \{M(x) \cap S_R(I_t(x))\}$ 可以表示为：

$$\# \{M(x) \cap S_R(I_t(x))\} = \begin{cases} 1 & \text{if } \text{dist}(v_i(x), I_t(x)) < R \\ 0 & \text{else} \end{cases} \quad (3-3)$$

式中， $\text{dist}(\cdot)$ 表示欧式距离。

不同于基于静态背景的样本随机聚类的建模算法中的静态阈值 R ，考虑到动态场景中，不同物体的晃动距离差距较大（树叶晃动的范围和电梯移动的距离不同），所以在基于动态背景的样本随机聚类的建模方法中，我们根据背景的动态程度，自适应的设置阈值 R ，该值和背景的动态移动范围有关，故我们增加一个参数：最小距离集合 $D(x)$ 及平均最小距离以 $d_{min}(x)$ 限制 R ，其中 $D(x)$ 可以表示为：

$$D(x) = \{D_1(x), \dots, D_k(x), \dots, D_N(x)\} \quad (3-4)$$

式中, $D(x) = \min\{\text{dist}(I_t(x), v_i(x))\}$, 表示当前像素 $I_t(x)$ 与其对应背景样本 $v_i(x)$ 的最小欧式距离; \min 表示最小距离对应的样本编号 i 。这里我们采用 N 个 $D_k(x)$ 的平均值 $d_{\min}(x)$ 以表征背景的动态运动情况:

$$d_{\min}(x) = \frac{1}{N} \sum_k D_k(x) \quad (3-5)$$

每次 $I_t(x)$ 与 $v_i(x)$ 匹配成功以后, 我们都记录一次 $d_{\min}(x)$, $d_{\min}(x)$ 和阈值 R 的关系可以表示为:

$$R(x) = \begin{cases} R(x) \cdot (1 - \alpha_{dec}) & \text{if } R(x) > d_{\min}(x) \cdot \xi \\ R(x) \cdot (1 + \alpha_{inc}) & \text{else} \end{cases} \quad (3-6)$$

式中, α_{dec} 、 α_{inc} 和 ξ 都是固定参数。对于背景中不变的物体, $R(x)$ 最终将趋于稳定值 $d_{\min}(x) \cdot \xi$ 。而对于背景中动态扰动的物体, $R(x)$ 会慢慢增加, 以适应扰动。

3.4 背景模型更新

考虑到晃动的叶子、波动的水波极易干扰前景的提取, 有时甚至会让前景被误检为背景, 产生误检率变大的问题。所以为了提高算法的精确度及鲁棒性, 我们利用空间一致性, 即邻域像素的匹配情况判断中心像素模型的准确性。其中对于视频帧 I 中的一个像素点 $x(x_i, x_j)$, 定义其 $k \times k$ 邻域为:

$$N_x = \{y = (y_i, y_j) \in I : |x_i - y_i| \leq k, |x_j - y_j| \leq k\} \quad (3-7)$$

由上式 N_x 可以得到集合 Ω_x , 该参数表示背景模型匹配上的像素:

$$\Omega_x = \{y \in N_x : \#\{M(y) \cap S_R(I(y))\} < \#_{\min}\} \quad (3-8)$$

结合 N_x 和 Ω_x , 我们可以得到邻域一致性因子:

$$NCF(x) = \frac{|\Omega_x|}{|N_x|} \quad (3-9)$$

式中, $||$ 表示集合基数, 一致性因子 $NCF(x)$ 即为背景的一致性匹配度, 一致性因子越大, 则像素点 x 及其邻域越能被背景模型匹配, 背景模型越应该及时更新。

3.5 模糊更新准则

静态背景中，由于大部分物体都是静止的，基于静态背景的样本随机聚类的建模算法可以通过设置固定的模型更新频率以完成背景的匹配。但在动态场景中，由于大部分物体是随机扰动的，固定不变的更新频率无法有效衡量背景检测时机，所以我们引入了这两年较为流行的模糊理论以有效衡量背景更新过程。模糊理论是由美国加州大学 LA. Zadeh 教授于 1965 年创立的模糊集合理论的数学基础上发展起来的一门理论，该技术可以有效利用隶属度函数刻画客观事物中的模糊界线。

为将模糊理论应用于我们的模型中，我们在基于动态背景的样本随机聚类的建模算法中加入了模糊函数 $F_1(x)$ ，该函数可以和随机更新时间因子 $\varphi = F_1(x) \times \varphi$ 组合，进而确定模板随机更新的概率。

$$F_1(x) = \begin{cases} 1 / (2 \times NCF(x)) & NCF(x) \geq 0.5 \\ 0 & NCF(x) < 0.5 \end{cases} \quad (3-10)$$

据 $NCF(x)$ 的计算公式可知， $NCF(x)$ 越大， x 邻域中能被背景模型描述的像素点越多，则用 x 来更新背景模型的概率就越大。上述过程可以反向约束模糊函数 $F_1(x)$ 及随机更新时间因子 φ ，进而自适应的调整背景更新概率，有效提高模型的精确度和鲁棒性。

3.6 模型工作过程

动态背景下前景提取模型的工作流程如图所示：

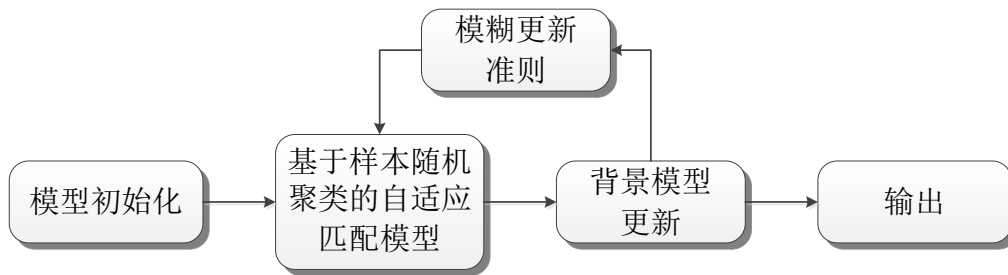


图 9 动态背景下前景提取流程图

3.7 问题求解及分析

我们利用附件中的视频 WaterSurface 对本章针对动态背景的前景提取方法进行了测试。在 WaterSurface 视频中，湖面发生持续的变化，但是从图 10 (b) 分离的前景图中可以看出，本章的方法可以有效屏蔽动态背景的干扰，取得较好的前景提取效果。这是由于本章方法在进行背景模型更新时，结合了模糊更新准

则，反向约束样本聚类方法，从而实现背景模型的自适应构建，这种自适应匹配模型能够更加准确地表征处于持续变化的动态背景。

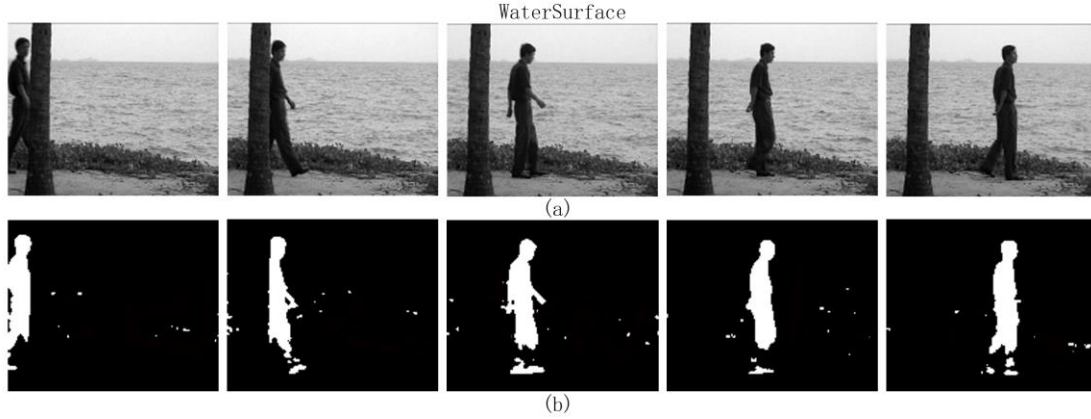


图 10 视频 WaterSurface 测试效果

4 问题三的建模与求解

4.1 问题分析

正如问题描述所说，在现实生活中确实存在的大量的相机抖动现象，比如由于拍摄技术的缺乏或专业设备的缺乏都会导致视频出现抖动频率过高/抖动幅度过大等影响观感的行为。问题中描述的监控视频摄像头确实也很可能由于自然因素（大风、大雨）或人为因素（无意碰撞）而发生晃动或偏移，进而降低监控视频录制的质量，导致视频无法有效提取前景的问题。值得注意的是，这种问题不同于问题二中提出的动态场景，在动态场景中，大部分物体只是在一定程度上的小规模轻微扰动，场景中仍存在大量静止不变的物体；而在镜头出现抖动的场景，所有的物体都处于扰动的状态，此时再利用问题二中针对动态场景的前景检测机制，会出现大量的前景、背景误判问题。针对这一问题，我们针对抖动场景的特点，进一步优化我们的前景检测模型，该模型可以通过对视频的前景背景进行区分，分别分析前景与背景运动轨迹的特点，有效避免了前景物体的特征轨迹对抖动估计和抖动补偿的干扰，进而有效提高了模型在抖动场景下的精确性和鲁棒性。

4.2 特征点轨迹模型

视频消抖的核心在于正确估计出相机本身的运动趋势，在一个视频中，前景由于是不断运动的，我们无法利用前景信息有效估计相机的运动轨迹，但我们注意到视频的背景绝大部分时间都是静止的，它的运动轨迹可以有效反映出镜头的运动轨迹。为实现这一目标，首先我们借用经典的二维图像坐标系的特征点轨迹模型，利用该模型我们即可以利用映射函数 T 得到视频的稳定帧：

$$\begin{aligned} f(X, t) &= f_m(X, t) + f_n(X, t) \\ X &= [x, y]^T \end{aligned} \quad (4-1)$$

其中， f 即表示前景的总运动轨迹， f_m 表示前景的正常轨迹， f_n 表示前景的抖动轨迹。通过观察上式，我们发现由于相机抖动而产生的抖动分量 f_n 相对于正常的前景运动轨迹 f_m 是高频分量，可以利用低通滤波等方法对视频进行平滑处理。但通过简单的仿真我们发现这种方法在消抖方面效果一般，我们进行了简单的分析，发现前景的第一帧信息 f_0 同样是高频分量，它也很可能被低通滤波器过滤掉，这时如果利用映射函数 $T_{f \rightarrow f_m}$ 进行映射，得到的消抖视频效果很差。所以我们结合这几年较为流行的单应矩阵重映射机制优化我们的模型。

4.3 前景特征匹配模型

针对直接利用高斯滤波处理特征点轨迹可能过滤掉前景信息的问题，我们发现必须首先对特征点轨迹进行分类，区分出前景轨迹和背景轨迹，才可以由背景轨迹的运动量计算出相机的扰动量。为实现这一目标，我们设计了一种基于单应矩阵重映射的方法以解决前景特征匹配问题。该方法可以通过计算两帧图片之间的重映射误差，以确定特征点轨迹的分类情况。

4.3.1 单应矩阵重映射

单应矩阵描述的是两幅图像所处平面之间的二维投影关系，为了计算两幅图片之间的单应矩阵可以提取两幅图像中的特征点并通过特征匹配得到两组一一对应的特征点坐标集合 X_1 和 X_2 ，进而利用经典的 RANSAC (Random Sample Consensus) 算法计算，RANSAC 算法可以从 X_1 和 X_2 中随机抽出 4 个样本并保证这 4 个样本之间不共线，计算出单应性矩阵，然后利用这个模型测试所有数据，并计算满足这个模型数据点的个数与投影误差(即代价函数)，若此模型为最优模型，则对应的代价函数最小，单应矩阵 H 的迭代次数可由下式计算：

$$k = \frac{\log(1-p)}{\log(1-w^m)} \quad (4-2)$$

在不考虑任何误差的情况下， $H * X_1$ 应与 X_2 完全相同。

$$S \begin{bmatrix} X_1^2 \\ X_2^2 \\ \dots \\ X_n^2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1n} \\ H_{21} & H_{22} & \dots & H_{2n} \\ \dots & \dots & \dots & \dots \\ H_{n1} & H_{n2} & \dots & H_{nn} \end{bmatrix} \begin{bmatrix} X_1^1 \\ X_2^1 \\ \dots \\ X_n^1 \end{bmatrix} \quad (4-3)$$

但是事实上会有各种干扰可能影响单应矩阵的映射过程，根据 RANSAC 原理，利用 $H * X_1$ 与 X_2 的差值，我们可以检测特征集 X_1 和 X_2 的前景区域。

基于以上原理，当我们需要将单应矩阵重映射机制运用到抖动场景下的前景检测模型时，我们可以认为 X_t 表示视频中第 t 帧的特征点坐标集合， X_s 表示视频中第 s 帧特征点坐标集合，而 H_{st} 表示特征点坐标 X_t 和 X_s 间的单应矩阵，该矩阵表示将 X_s 映射到 X_t 的过程。

4.3.2 基于单应矩阵重映射的前景轨迹识别

考虑到，特征点坐标集合 X_t 和 X_s 的轨迹都由背景轨迹和前景轨迹组成，而背景轨迹所体现的运动趋势（特征点连接起来的长度、形状等）与前景轨迹的运动趋势不同，基于此，我们可以根据单应矩阵的重映射误差对每一条轨迹进行区分。

首先，我们假设 X_t 和 X_s 中属于背景的特征点数量占多数，这样我们才可以的得到一个相对正确的单应矩阵 H_{st} 。对于属于背景轨迹的特征点坐标 p^j 而言，其重投射误差 $e_{s,t}^j$ 要远远小于属于前景轨迹的特征点坐标的重投射误差。为此我们设计了一个阈值 $\tau > 0$ ，当 $e_{s,t}^j > \tau$ 时，我们认为其对应的特征点 j 属于前景轨迹，反之则属于背景轨迹，下式为我们的区分函数：

$$F_{s,t}^j = \begin{cases} 0 & e_{s,t}^j < \tau \\ 1 & e_{s,t}^j \geq \tau \end{cases} \quad (4-4)$$

由于 τ 的设置对判定结果直接影响，直接设置一个固定且合理的 τ 十分困难，所以我们设计一个自适应机制以满足 τ 的需求：

$$\tau = \frac{1}{2\omega} \sum_{s=t-\omega, \dots, t-1, t+1, \dots, t+\omega} \frac{\lambda}{B} \sum_{j=1}^{B_s} e_{s,t}^j \quad (4-5)$$

其中， B_s 表示 X_s 中轨迹的数量， $e_{s,t}^j$ 表示对应轨迹的特征点重投影误差， λ 表示微调参数，可以根据场景需要设置为 $[0,1]$ 间的任何实数。

考虑到单帧图像可能具有偶然性，我们对 $2\omega+1$ 帧图像都进行重映射过程并计算了相应的 F_t^j ：

$$F_t^j = \sum_{s=t-\omega, \dots, t-1, t+1, \dots, t+\omega} F_{s,t}^j \quad (4-6)$$

对于每一条轨迹 p^j 中， F_t^j 表示在 2ω 次重映射中对应轨迹被判定为前景轨迹的次数，若 $F_t^j > \omega$ ，即大于一半的帧都满足条件时，则认为该条轨迹是前景轨迹，否则为背景轨迹。

上述机制利用单应矩阵重映射的特点，有效区分了视频中前景物体与背景物体的运动轨迹。但该机制仍然存在问题，即景深变化对运动速度的影响。当前景物体里相机很远时，相机本身的移动对前景物体的运动轨迹影响很小；但当前景物体离相机很近时，即使相机本身的移动范围很少，其反映在运动轨迹的变化却很大。此时若仍然利用单应矩阵重映射机制分析视频，其产生的重映射误差将会严重影响视频的前景提取。

4.3.3 基于历史行为的单应矩阵前景轨迹识别模型

为解决上述问题，我们在基于单应矩阵重映射的前景轨迹识别模型的基础上，引入了自适应机制，以解决景深对于前景物体的提取误差问题。

我们发现，景深对于前景目标的判定上产生误差的原因主要是由于前景目标的不断变大（离镜头越来越近），所以我们的基于历史行为的单应矩阵前景轨迹识别模型在前景分析上加入了一步前景判定机制，即一旦某条轨迹在 t 时刻被认定为前景轨迹，则在 t 之后的时间内，它都将被认定为前景轨迹，我们设计的前景判决模型将不会对它继续分析。该机制的好处在于它有效避免了继续计算导致的重映射误差变大的问题，而且只计算尚未被判定为前景轨迹的轨迹可以有效降低模型的运行成本。基于历史行为的单应矩阵前景轨迹识别模型的工作流程如下：

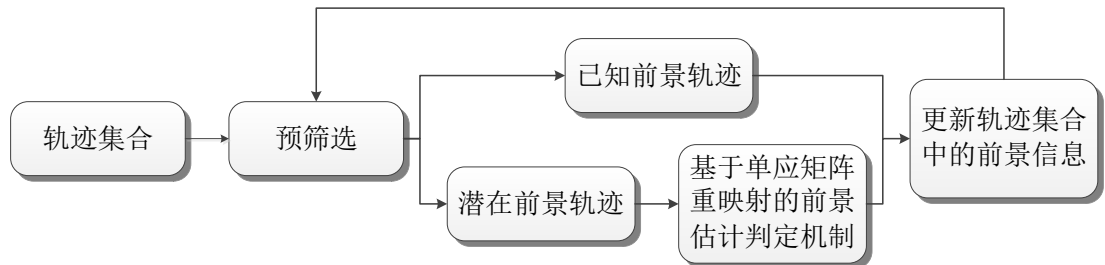


图 11 基于历史行为的单应矩阵前景轨迹识别模型的工作流程

4.4 视频消抖模型

前景路径判定后，剩下所有的轨迹都是背景轨迹。基于背景轨迹和前景轨迹

我们可以分别得到相机在最近若干帧中的抖动量 $f_n(X, t)$ 及前景在最近若干帧中的运动量 $f_m(X, t)$ 。为有效获取背景的运动轨迹，我们采用高斯低通滤波对背景轨迹进行平滑处理：

$$\begin{aligned} \tilde{f}_{Bg} &= (g * f_{Bg}) \\ g_t &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}}, t \in [-\sigma, \sigma] \end{aligned} \quad (4-7)$$

其中， $*$ 代表卷积符号， g_t 表示高斯低通滤波器， f_{Bg} 表示处理前的背景轨迹集合， \tilde{f}_{Bg} 表示处理后的背景轨迹集合。高斯低通滤波可以有效的滤除背景轨迹中的高频分量，从而消除背景轨迹中的抖动量。背景平滑过程如图所示：

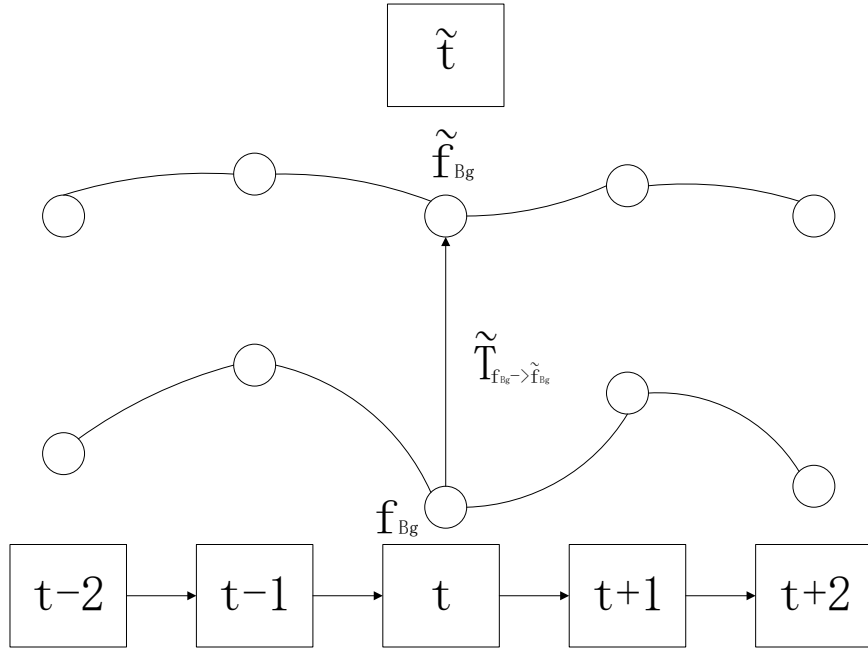


图 12 背景平滑示意图

根据平滑前后的背景轨迹信息，我们可以计算出一个单应矩阵，该单应矩阵如下式所示：

$$\min \sum_{i=1}^B \|f_{Bg}^i - \tilde{f}_{Bg}^i\| \quad (4-8)$$

其中， B 为背景轨迹数量。通过上式，我们可以得到第 t 帧从原始视角到稳定视角的投影关系，利用变换函数 $\tilde{T}_{f \rightarrow f_m}$ 对当前帧进行透视变换，即可得到当前帧在稳定视角下的稳定帧。

完成背景轨迹的平滑及单应矩阵的计算后，我们可以对背景轨迹进行计算，得到抖动量的正确估计，进而可以对视频进行消抖补偿，算出视频中正确的前景

轨迹。

4.5 模型工作流程

视频消抖模型工作流程如图 13 所示：

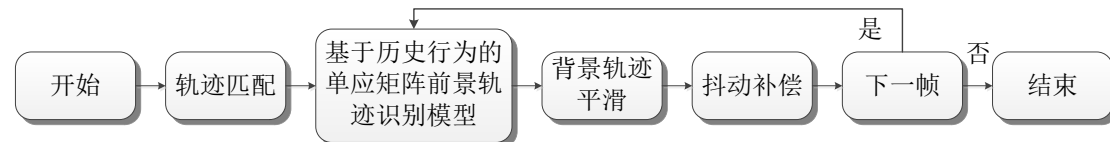


图 13 抖动背景下的前景提取方法流程图

4.6 问题求解及分析

本章利用了附件中四组包含抖动背景的视频（Car6, Car7, People1, People2）进行测试。背景抖动会导致视频中所有的前景背景均发生运动，只是运动的幅度有所区别。针对此问题，我们构建了视频消抖模型，根据视频历史数据捕获前景和背景的运动轨迹，并提炼出运动量来区分出前景和背景。从图 14, 15, 16, 17 中可以看出，随着处理帧数的增加，前景提取效果逐渐变好，这是由于随着历史数据的增多，前景和背景的运动轨迹捕获的越真实，有利于建立更加准确地视频消抖模型。

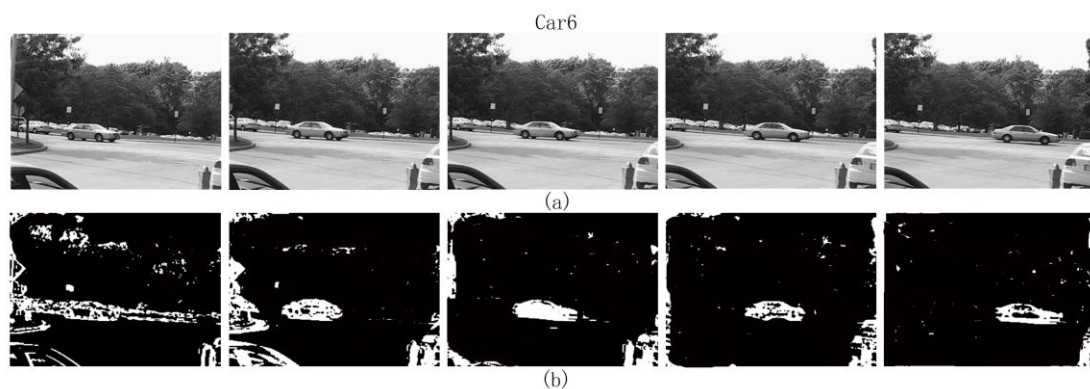


图 14 视频 Car6 测试效果

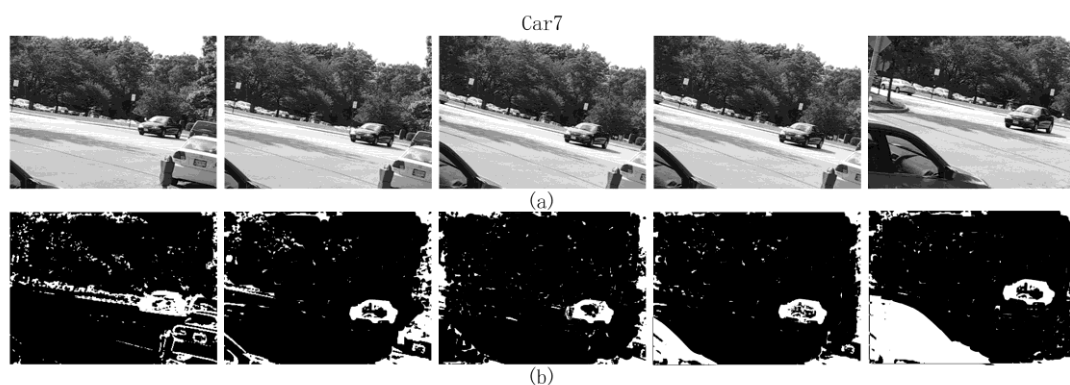


图 15 视频 Car7 测试效果

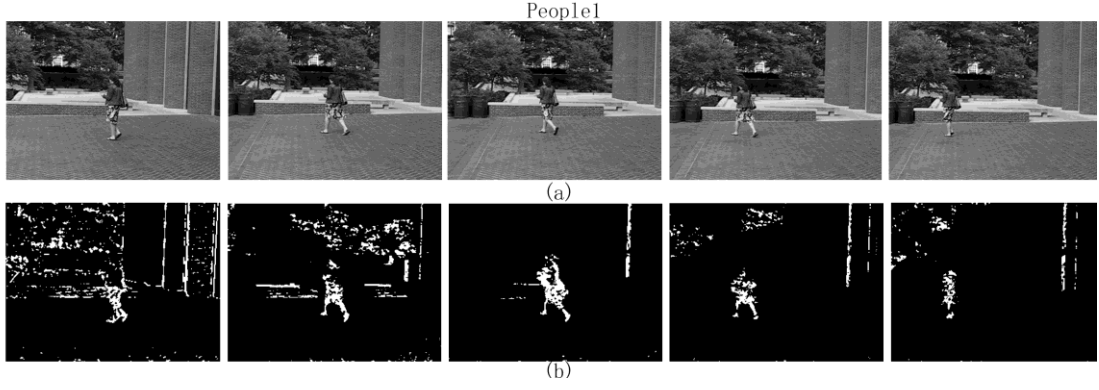


图 16 视频 People1 测试效果

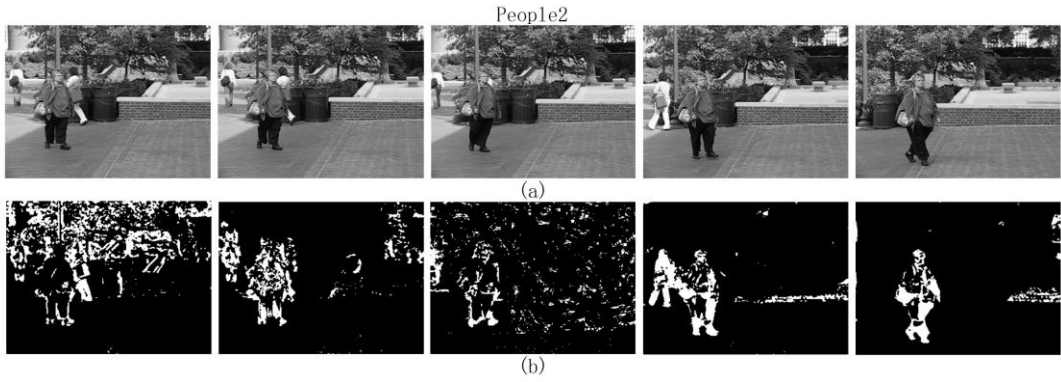


图 17 视频 People2 测试效果

5 问题四（8 组视频分组标注）

本节中，我们利用前文提出的模型对八组视频进行了分组标注，八组视频分别为：Campus、Curtain、Escalator、Fountain、Hall、Lobby、Office 和 Overpass。我们采用精确度作为我们的测评指标：

$$\text{Acc} = \frac{\sum_{i=1}^n \tilde{f}_i \cap f_i}{n} \quad (5-1)$$

其中 Acc 就是我们使用的精确度， n 为总帧数， \tilde{f}_i 表示前景实际出现的帧数， f 表示我们模型测出的前景出现帧数。下面我们分视频对各个视频进行分析。

5.1 Campus 视频标注及分析

我们首先利用“会声会影”软件，分析了 Campus 视频中前景物体出现的帧数，我们认为 Campus 视频中的前景为运动的人和车。下表为具体前景物体及其出现的帧数：

表 1 Campus 视频中实际出现前景的帧数

前景物体的实际出现帧数											
帧数	200	306	352	644	691	739	810	1006	1330	1348	1377
	-	-	-	-	-	-	-	-	-	-	-
	224	522	392	682	711	905	834	1035	1354	1374	1405
前景	存在	存在	存在	存在	存在	存在	存在	存在	存在	存在	存在

进而我们利用我们在上文中提出的模型，对视频进行了分析，下表是我们模型分析出的帧数范围：

表 2 Campus 视频中检测出前景的帧数

模型分析的前景物体帧数											
帧数	200	306	354	647	695	741	811	1007	1331	1349	1377
	-	-	-	-	-	-	-	-	-	-	-
	224	519	392	682	711	903	833	1035	1354	1374	1405
前景	存在	存在	存在	存在	存在	存在	存在	存在	存在	存在	存在

我们进行了对比分析。实验证明，在 Campus 视频中，我们的模型的检测精度可以达到人眼（经过了大量的样本训练）的检测精度的 96.69%。除此之外，我们的算法还在视频中发现了一些被额外插入的帧。

表 3 Campus 视频中出现插入帧的帧数

帧数	85	600	1193	1264
前景	存在	存在	存在	存在

这些帧存在长度很短，由于视觉残留现象，人的肉眼通常很难在视频中区分出这些帧的前景信息，但我们的方法可以有效突破人类视觉的极限，清晰的展示出这些插入帧的内容。具体的分辨效果如下图所示：

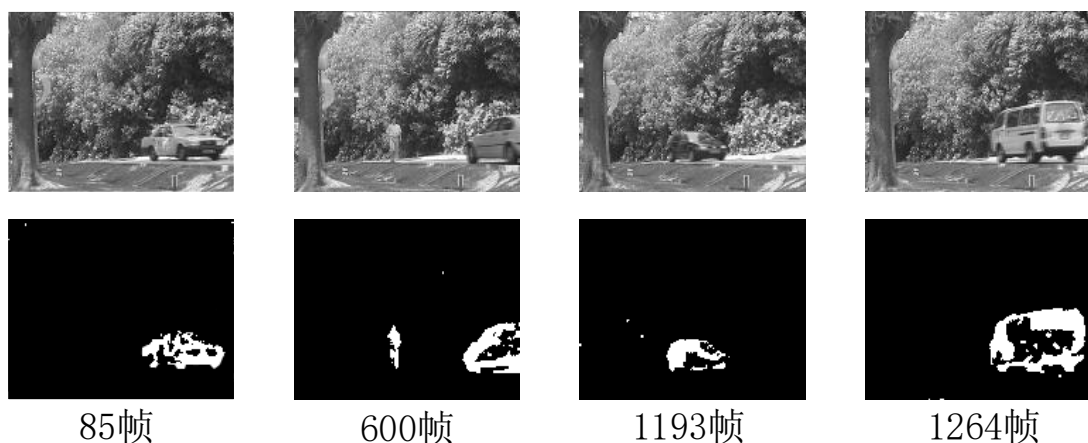


图 18 Campus 视频中插入帧的内容

5.2 Curtain 视频标注及分析

类似于 Campus 视频的分析思路，我们认为 Curtain 视频中的前景为运动的人，基于此情景我们分析了 Curtain 视频的实际视频帧数：

表 4 Curtain 视频中实际出现前景的帧数

前景物体的实际出现帧数			
帧数	1757-1903	2170-2313	2765-2930
前景	存在	存在	存在

进而我们利用我们在上文中提出的模型，对视频进行了分析，下表是我们模型分析出的帧数范围：

表 5 Curtain 视频中检测出前景的帧数

模型分析的前景物体帧数			
帧数	1758-1893	2174-2313	2768-2929
前景	存在	存在	存在

我们也采用精确度衡量模型的前景分析能力，实验证明，在 Curtain 视频中，我们的模型的检测精度可以达到人眼（经过了大量的样本训练）的检测精度的 95.81%。在这个视频中我们也发现了一些被额外插入的帧。

表 6 Curtain 视频中出现插入帧的帧数

帧数	411	967	1561	2126	2642
前景	存在	存在	存在	存在	存在

这些单独帧的具体的分辨效果如下图所示：

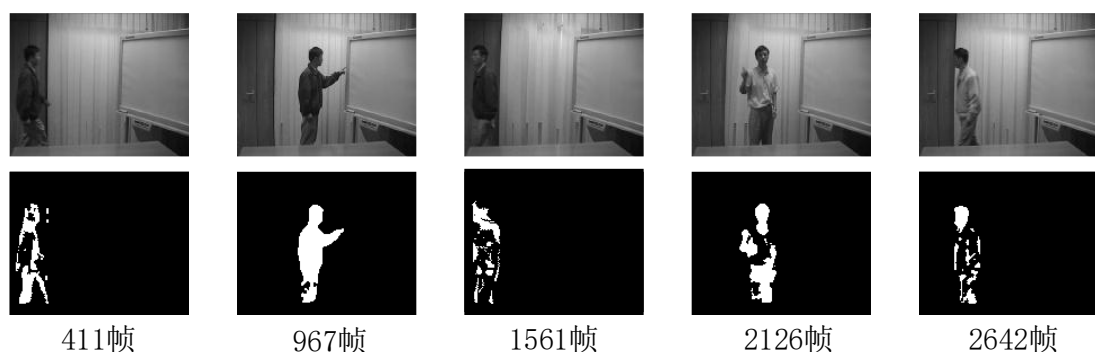


图 19 Curtain 视频中插入帧的内容

5.3 Escalator 视频标注及分析

类似的，我们认为 Escalator 视频中的前景为运动的人群，基于此前景我们分析了 Escalator 视频的实际视频帧数：

表 7 Escalator 视频中实际出现前景的帧数

前景物体的实际出现帧数			
帧数	1-168	206-2391	2769-3417
前景	存在	存在	存在

进而我们利用我们在上文中提出的模型，对视频进行了分析，下表是我们模型分析出的帧数范围：

表 8 Escalator 视频中检测出前景的帧数

模型分析的前景物体帧数			
帧数	1-143	283-2390	2773-3417
前景	存在	存在	存在

我们也采用精确度衡量模型的前景分析能力，实验证明，在 Escalator 视频中，我们的模型的检测精度可以达到人眼（经过了大量的样本训练）的检测精度的 96.43%。在这个视频中我们也发现了一些被额外插入的帧。

表 9 Escalator 视频中出现插入帧的帧数

帧数	2415	2539	2754
前景	存在	存在	存在

这些单独帧的具体的分辨效果如下图所示：

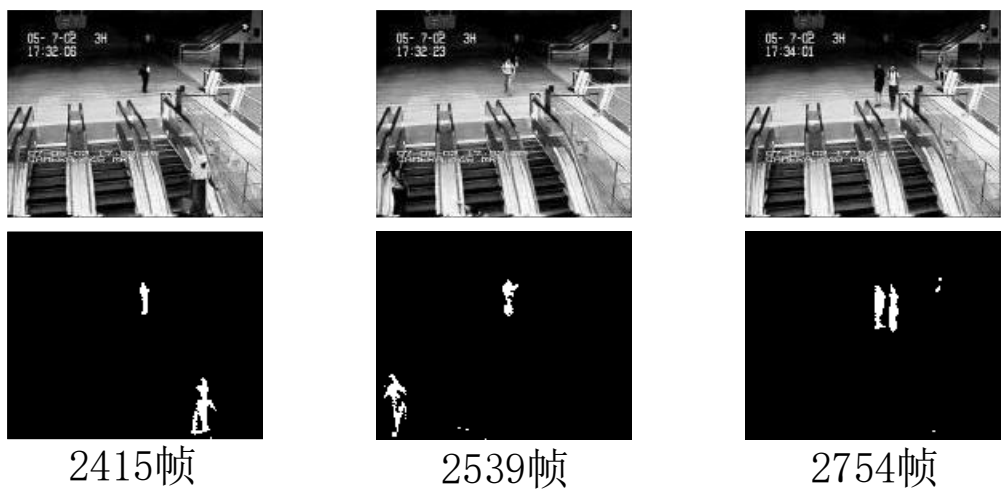


图 20 Escalator 视频中插入帧的内容

5.4 Fountain 视频标注及分析

类似的，我们认为 Fountain 视频中的前景为运动的人，基于此前景我们分析了 Fountain 视频的实际视频帧数：

表 10 Fountain 视频中实际出现前景的帧数

前景物体的实际出现帧数		
帧数	154-212	403-523
前景	存在	存在

进而我们利用我们在上文中提出的模型，对视频进行了分析，下表是我们模型分析出的帧数范围：

表 11 Fountain 视频中检测出前景的帧数

模型分析的前景物体帧数		
帧数	157-209	408-523
前景	存在	存在

我们也采用精确度衡量模型的前景分析能力，实验证明，在 Fountain 视频中，我们的模型的检测精度可以达到人眼（经过了大量的样本训练）的检测精度的 93.82%。在这个视频中我们也发现了一些被额外插入的帧。

表 12 Fountain 视频中出现插入帧的帧数

帧数	141	259	335
前景	存在	存在	存在

这些单独帧的具体的分辨效果如下图所示：

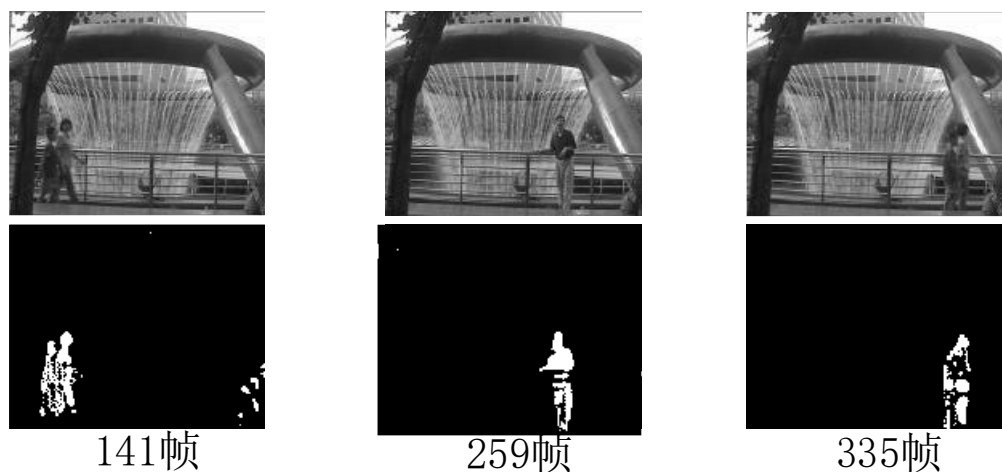


图 21 Fountain 视频中插入帧的内容

5.5 Hall 视频标注及分析

类似的，我们认为 Hall 视频中的前景为运动的人群，基于此前景我们分析了 Hall 视频的实际视频帧数：

表 13 Hall 视频中实际出现前景的帧数

前景物体的实际出现帧数									
帧数	1	601	818	1154	1279	2722	2817	3025	3404
	-	-	-	-	-	-	-	-	-
	509	746	1053	1212	2583	2776	2869	3349	3534
前景	存在	存在	存在	存在	存在	存在	存在	存在	存在

进而我们利用我们在上文中提出的模型，对视频进行了分析，下表是我们模型分析出的帧数范围：

表 14 Hall 视频中检测出前景的帧数

模型分析的前景物体帧数									
帧数	2	653	820	1154	1280	2734	2817	3011	3405
	-	-	-	-	-	-	-	-	-
	462	736	1049	1211	2582	2775	2865	3348	3534
前景	存在	存在	存在	存在	存在	存在	存在	存在	存在

我们也采用精确度衡量模型的前景分析能力，实验证明，在 Hall 视频中，我们的模型的检测精度可以达到人眼（经过了大量的样本训练）的检测精度的 95.58%。在这个视频中我们也发现了一些被额外插入的帧。

表 15 Hall 视频中出现插入帧的帧数

帧数	578	795	1246
前景	存在	存在	存在

这些单独帧的具体的分辨效果如下图所示：

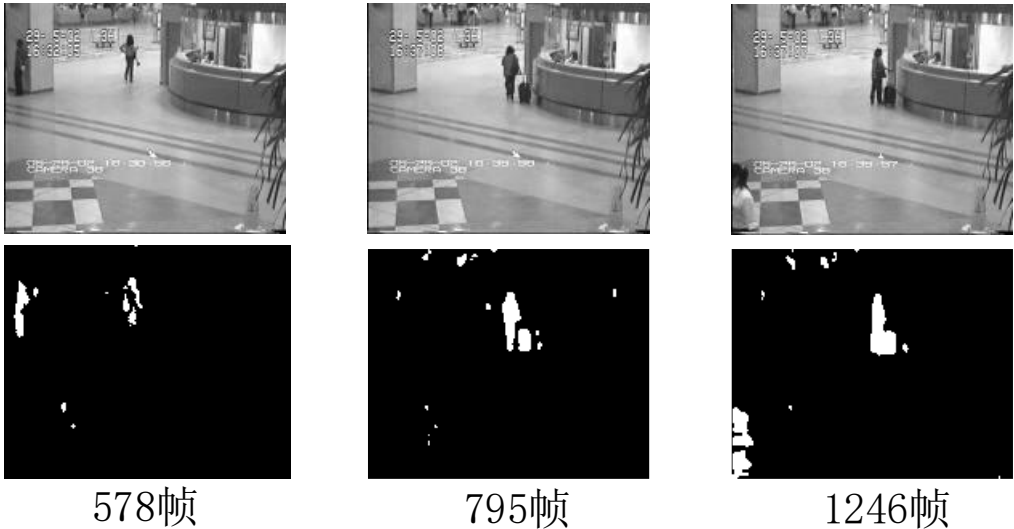


图 22 Hall 视频中插入帧的内容

5.6 Lobby 视频标注及分析

类似的，我们认为 Lobby 视频中的前景为运动的人，基于此前景我们分析了 Lobby 视频的实际视频帧数：

表 16 Lobby 视频中实际出现前景的帧数

前景物体的实际出现帧数							
帧数	154-195	345-392	623-666	963-1036	1239-1282	1336-1376	1377-1538
前景	存在	存在	存在	存在	存在	存在	存在

进而我们利用我们在上文中提出的模型，对视频进行了分析，下表是我们模型分析出的帧数范围：

表 17 Lobby 视频中检测出前景的帧数

模型分析的前景物体帧数						
帧数	155-196	346-392	623-647	1240-1258	1280-1281	1336-1538
前景	存在	存在	存在	存在	存在	存在

我们也采用精确度衡量模型的前景分析能力，实验证明，在 Lobby 视频中，我们的模型的检测精度可以达到人眼（经过了大量的样本训练）的检测精度的 66.46%。在这个视频中我们也发现了一些被额外插入的帧。

表 18 Lobby 视频中出现插入帧的帧数

帧数	79	259	521	870	1161
前景	存在	存在	存在	存在	存在

这些单独帧的具体的分辨效果如下图所示：

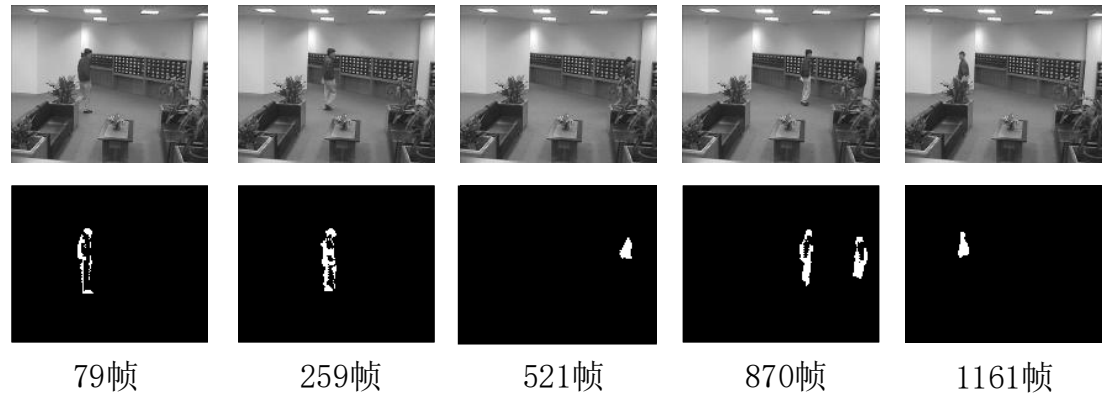


图 23 Lobby 视频中插入帧的内容

5.7 Office 视频标注及分析

类似的，我们认为 Office 视频中的前景为运动的人，基于此前景我们分析了 Office 视频的实际视频帧数：

表 19 Office 视频中实际出现前景的帧数

前景物体的实际出现帧数	
帧数	586-2040
前景	存在

进而我们利用我们在上文中提出的模型，对视频进行了分析，下表是我们模型分析出的帧数范围：

表 20 Office 视频中检测出前景的帧数

模型分析的前景物体帧数	
帧数	587-2035
前景	存在

我们也采用精确度衡量模型的前景分析能力，实验证明，在 Office 视频中，我们的模型的检测精度可以达到人眼（经过了大量的样本训练）的检测精度的 99.58%。在这个视频中我们也发现了一些被额外插入的帧。

表 21 Office 视频中出现插入帧的帧数

帧数	197	372	501	2080
前景	存在	存在	存在	存在

这些单独帧的具体的分辨效果如下图所示：

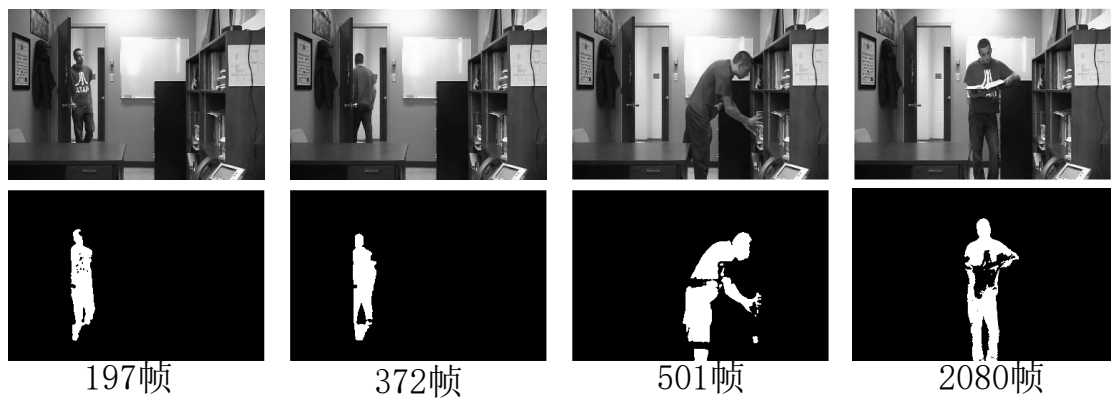


图 24 Office 视频中插入帧的内容

5.8 Overpass 视频标注及分析

类似的，我们认为 Overpass 视频中的前景为运动的人和船，基于此前景我们分析了 Overpass 视频的实际视频帧数：

表 22 Overpass 视频中实际出现前景的帧数

前景物体的实际出现帧数		
帧数	548-634	2337-2958
前景	存在	存在

进而我们利用我们在上文中提出的模型，对视频进行了分析，下表是我们模型分析出的帧数范围：

表 23 Overpass 视频中检测出前景的帧数

前景物体的实际出现帧数		
帧数	556-634	2336-2956
前景	存在	存在

我们也采用精确度衡量模型的前景分析能力，实验证明，在 Office 视频中，我们的模型的检测精度可以达到人眼（经过了大量的样本训练）的检测精度的 98.73%。在这个视频中我们也发现了一些被额外插入的帧。

表 24 Overpass 视频中出现插入帧的帧数

帧数	374	968	1551	1881	2098
前景	存在	存在	存在	存在	存在

这些单独帧的具体的分辨效果如下图所示：

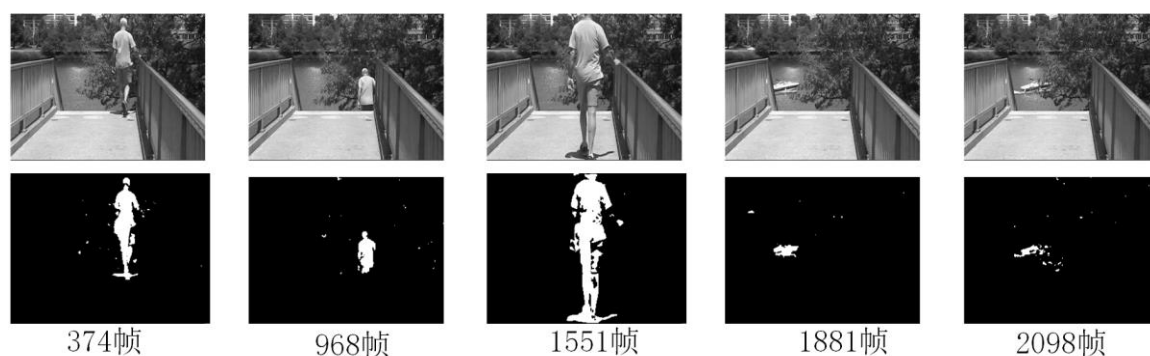


图 25 Overpass 视频中插入帧的内容

5.9 综合分析

通过对上述八个视频进行分析，我们发现我们的模型可以较好的对这八个视频的前景进行有效提取，每个视频的前景匹配精确度如下表所示。结合视频及我们的分析结果，我们发现除了 Lobby 视频外，我们的模型拥有 93% 以上的前景匹配率。进一步我们对 Lobby 视频进行了分析，发现 Lobby 视频中的光照变化严重的影响了我们模型的检测结果，在 lobby 视频中的第 427 帧停电后，由于背景颜色的突然改变，所有的背景都会被误检为前景。虽然我们模型的自学习能力可以不断的消除光照变化对视频检测的影响，但这仍然大大弱化了我们的前景检测能力。

表 25 8 组视频前景匹配精确度

视频	前景匹配精确度
Campus	96.69%
Curtain	95.81%
Escalator	96.43%
Fountain	93.82%
Hall	95.58%
Lobby	66.46%
Office	99.58%
Overpass	98.73%

针对以上问题，我们进行了进一步深入研究。我们认为人眼之所以优于我们前景提取机制的原因，在于人类的眼睛经过了大量前景样本的训练，单纯的光照影响难以对人眼的前景提取产生影响，而我们的机制没有经过大量的样本训练，仅根据像素的颜色差别对前景进行区分，这种机制必定会受到光照变化的剧烈影响。所以针对这个问题，我们未来期望利用深度学习机制优化我们的模型，减少

光照因素对前景提取的影响。

6 问题五的建模与求解

6.1 问题分析

在多摄像头监控视频中进行前景提取需要完成两个任务，一是要在各个摄像头视角中将前景目标提取出来，二是将在不同视角中出现的同一目标对应起来。运动目标检测算法和匹配技术是多摄像系统的核心技术，是后续进行视频智能分析的基础。多摄像头监控的难点在于要将各独立摄像头提取到的前景进行关联，通常的做法是基于图像空间的方法。该方法首先在各自视角下进行前景目标的提取，然后基于提取结果，结合几何、纹理、颜色特征等外观信息特征匹配实现目标对应，但是这种方法难以解决遮挡问题，对此本章采用融合空间的检测算法来提取前景以及实现对应匹配。

6.2 基于目标存在概率场模型的多视角前景目标提取算法

6.2.1 目标存在概率场模型

图 26 表示多个摄像头在对同一块区域进行监控，用符号 Π 表示该区域。然后将此区域等分成 m 个格点 $i \in \{1, 2, \dots, m\}$ ，每个格点代表该区域中的一个位置坐标， m 为划分的格点总数。假设区域 Π 为平面区域，在给定不同摄像头拍摄的图像的条件下，对任意位置 i 定义目标存在的概率为

$$p(e_i = 1|V) \quad (6-1)$$

其中， $V = \{I, \Theta\}$ ， I 表示观测图像的集合， Θ 表示所有摄像机的参数集合， e_i 为一个随机变量， $e_i = 1$ 表示区域 Π 的位置 i 处存在目标， $e_i = 0$ 则表示不存在目标。

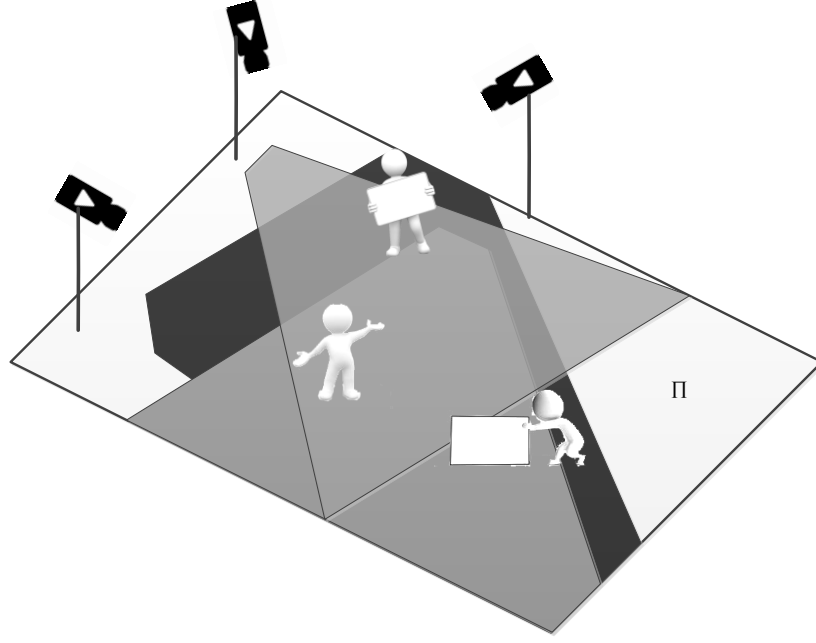


图 26 目标存在概率场示意图

6.2.2 目标窗结构模型

为了有效表示目标存在的概率，我们依据成像的原理构建了目标窗结构^[9]。定义符号 S 表示目标窗结构， $S = \{M, W\}$ ，其中 M 表示空间目标模型，利用圆柱体表示。 M 包含的参数有位置 i ，高度 h ，半径 r 。其中参数 h 和 r 的选择要能够涵盖目标的绝大部分区域。 $W = \{w_1, w_2, \dots, w_n\}$ ， w_k 表示目标模型在第 k 个摄像头的投影区域。

对于存在场中的位置 i 而言，如果存在目标，则无论该目标是否被背景遮挡，目标的投影总会落入对应的摄像机视角的子窗口中。这表明目标在位置 i 是否存在的概率只与目标窗结构的子窗在各视角的图像中所包含的图像区域的大小有关。并且在实际情况下，真实目标并非一个点，因此其占据的位置也不会仅仅是一个点，而应该是一块区域。处于目标占据位置中心在所有摄像机投影的子窗口中得到的前景区域最完整，随着目标远离中心，各个子窗口包含的前景区域将会逐渐减小。

6.2.3 计算目标存在概率

在 6.2.2 节中，我们已经分析了，目标存在于某个位置的概率仅仅与目标窗结构的子窗口包含的局部区域大小有联系，因此在计算目标存在概率时，仅需要考虑子窗内的局部图像。

在推导概率计算公式前首先需要作出两个假设：

假设 1：各摄像头监控的视频之间条件独立。

假设 2：各摄像头监控的视频自身的噪声相互独立，则

$$\begin{aligned}
p(e_i = 1|V) &= p(e_i = 1|I, \Theta) \\
&= \frac{p(I|e_i = 1, \Theta) p(e_i = 1|\Theta)}{p(I|\Theta)} \\
&= \prod_{k=1}^n p(I_k|e_i = 1, \Theta) \frac{p(e_i = 1|\Theta)}{\prod_{k=1}^n p(I_k|\Theta)} \\
&= \frac{p(e_i = 1|\Theta)}{\prod_{k=1}^n p(I_k|\Theta)} \prod_{k=1}^n \frac{p(e_i = 1|I_k, \Theta) p(I_k|\Theta)}{p(e_i = 1|\Theta)} \\
&= \frac{1}{(p(e_i = 1|\Theta))^{n-1}} \prod_{k=1}^n p(e_i = 1|I_k, \Theta)
\end{aligned} \tag{6-2}$$

其中 $(p(e_i = 1|\Theta))^{n-1}$ 是由投影参数决定，对于各个位置点都一样，因此得到

$$p(e_i = 1|V) \propto \prod_{k=1}^n p(e_i = 1|I_k, \Theta) \tag{6-3}$$

由式子（6-3）可以发现，只要得到各个图像的 $p(e_i = 1|I_k, \Theta)$ ，就可以计算出目标存在概率。前面我们已经分析过，目标在多摄像头监控位置存在的概率只与目标窗结构的子窗口有联系，所以在计算 $p(e_i = 1|I_k, \Theta)$ 的时候，仅需考虑目标窗的对应子窗。

然后依据公式（6-3）计算 $p(e_i = 1|I_k, \Theta)$

$$p(e_i = 1|I_k, \Theta) = \frac{fore(w_k^i)}{area(w_k^i)} \tag{6-4}$$

其中， $fore(w_k^i)$ 表示在存在场位置 i 处的目标窗结构在图像 k 的子窗内前景的面积和， $area(w_k^i)$ 表示在存在场位置 i 的目标窗结构在图像 k 的子窗面积。

将（6-4）代入到（6-1），可以得到存在概率的计算公式，

$$p(e_i = 1|V) \propto \prod_{k=1}^n \frac{fore(w_k^i)}{area(w_k^i)} \tag{6-5}$$

图 2 表示依据（6-5）求解出的一个存在概率场的计算结果。

6.2.4 多视角前景目标提取算法流程

（1）前景提取：利用各个摄像头捕获的图像构建前景和背景模型

(2) 存在概率场计算：利用步骤(1)构建的前景和背景模型来计算目标存在概率场，概率值越大表示该位置存在目标的可能性越大

(3) 优化搜索：得到概率场了以后，我们采用 Mean-Shift 算法搜索空间极值，确定目标位置。

(4) 视频间的目标检测和匹配：根据步骤(3)确定出的目标位置，根据各个位置的目标窗结构，将结果重新投影到各个图像上，对应的目标窗结构的子窗将其对应视角的运动目标检测出来，并根据映射关系确定各目标的对应关系。

6.3 模型工作流程

多视角前景目标提取模型的工作流程如下图所示：

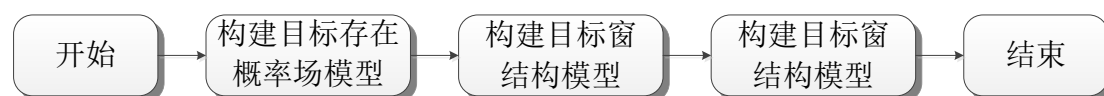


图 27 多视角前景目标提取方法流程图

6.4 实验分析

我们利用一组多视角拍摄的视频进行方法验证。在本组视频中，由四个摄像头拍摄而成的多视角视频，其中视频中有 4 个移动的目标。我们利用本章的基于目标存在概率的多视角前景提取方法不仅可以有效提取出 4 个前景目标，并且在确定出目标位置以后，根据各个位置的目标窗结构，将结果重新投影到各个图像上，根据映射关系确定出个目标的对应关系，测试效果如图 28 所示，可以看出本章方法可以有效提取出多视角拍摄视频中的前景，并确定对应关系。

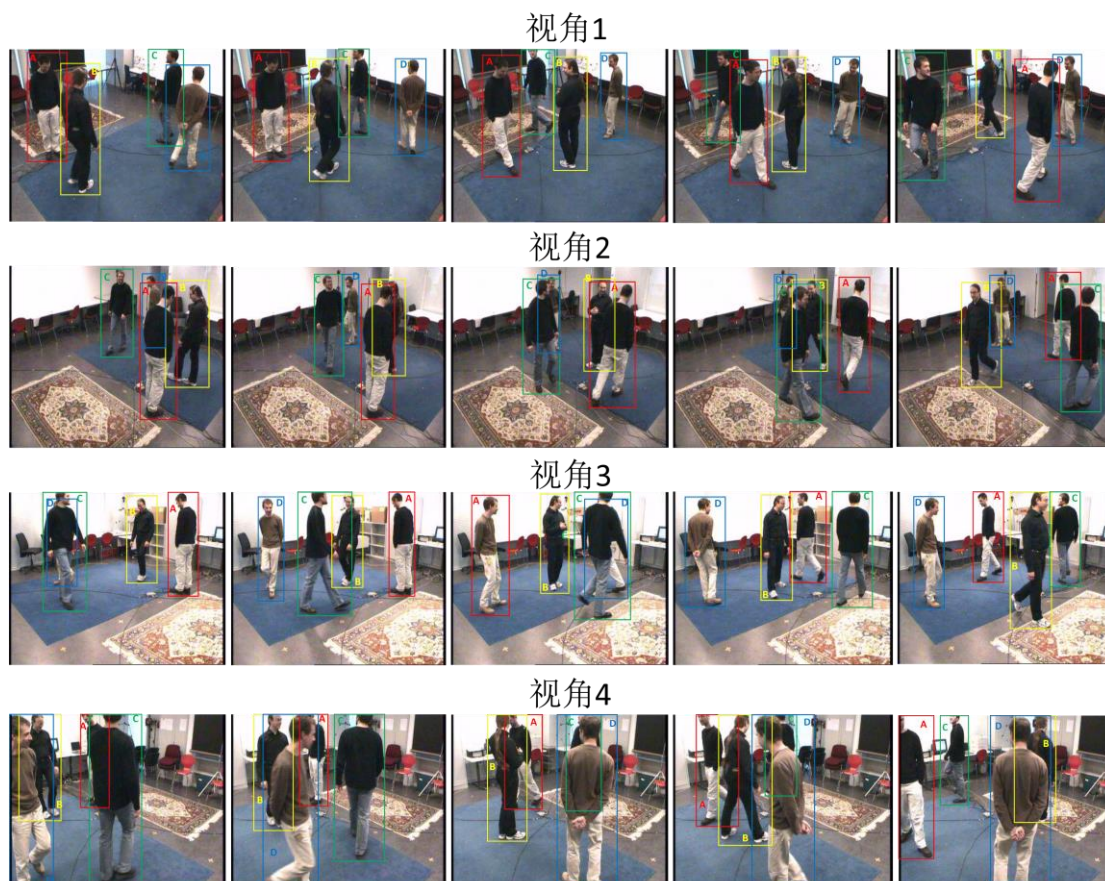


图 28 多视角前景提取与匹配

7 问题六的建模与求解

7.1 问题分析

近年来恐怖袭击事件频发，为了有效保证公共安全，智能视频监控需要具备异常事件检测的能力。异常事件检测的难点在于定义视频的异常特征，以及异常特征的提取。这里，我们以群体的运动特征作为视频的异常特征。但是如何自动捕获群体的有效特征信息并基于特征信息做出准确的判断则是一个难点。针对此问题我们利用光流法对运动群体进行运动建模，并将运动模型转换成瞬时能量模型来检测是否有异常事件的发生。本章采用的方法主要分为三个部分，基于像素统计分布的运动特征点提取、基于光流法的运动模型构建以及基于瞬时能量模型的异常事件检测方法。

7.2 基于光流能量模型的异常事件检测方法

7.2.1 基于像素统计分布的运动特征点提取

在进行目标运动特征点提取前，需要对运动目标进行检测，我们采用问题三中的方法对运动目标进行检测，获取前景运动区域，然后在前景运动区域中提取

运动特征点。

首先对前景运动区域进行网格化处理。为了有效反映运动群体在前景中的密度分布，我们采用基于前景像素面积占比的方法来提取运动区域的特征点。假设每个网格单元中包含的像素点数为 N_{mes} ，在网格内前景像素所占的面积比为 r ，前景运动区域在每个网格单元内的像素数量为 N_{for} ，因此前景像素面积占比 r 可由式（7-1）表示

$$r = \frac{N_{for}}{N_{mes}} \quad (7-1)$$

然后根据面积占比 r 来确定单元网格内的特征点个数 n 以及特征的位置分布，我们根据 r 的不同取值范围设置了四组特征点分布图

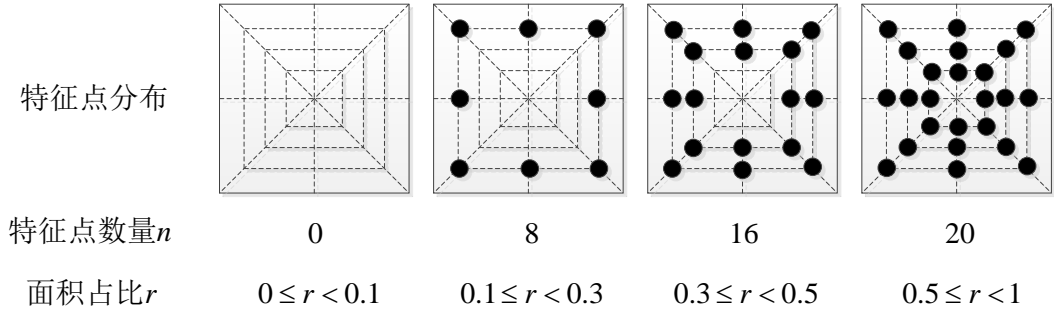


图 29 特征点分布图

当 $0 \leq r < 0.1$ 时，可以认为该网络所在区域内没有运动群体的活动，因此不进行特征点的提取，通过这种方法可以有效节约特征提取带来的时间开销，提高算法效率。随着面积占比的增加，可以认为运动群体逐步占据网格，因此设置更多的特征点来捕获群体运动行为。

这种基于像素统计分析的运动特征点提取方法，可以有效避免一些无用特征点的提取，避免其干扰运动建模以及提高运算代价，并且可以有效反映运动群体的密度分布情况，有利于后续的运动建模。

7.2.2 基于光流法的运动模型构建

光流的概念是由美国学者 Gibson 首次提出，它将物体在空间运动状态的变化转换成二维成像平面上所对应的像素的瞬时变化。光流法依据每一帧视频中像素的灰度值变化表征相邻帧之间物体的速度信息和位置变化，换言之，光流可以直接表示运动目标在相邻帧之间的瞬时移动速度。本节我们依据 L-K 光流法对视频中的运动目标进行运动建模。

L-K 光流法假设在待处理的像素周围的局部区域内，光流场是恒定不变的。基于该假设，在待处理像素点周围的局部区域内的所有像素点满足基本光流方程，如式（7-2）所示

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y = -\frac{\partial I}{\partial t} \quad (7-2)$$

其中， V_x 和 V_y 分别为光流在 x 轴和 y 轴上的分量。 $\frac{\partial I}{\partial x}$ ， $\frac{\partial I}{\partial y}$ 和 $\frac{\partial I}{\partial t}$ 分别表示图像的灰度值在 x 轴， y 轴以及时间上的偏导数。然后利用最小二乘法求解式(7-2)。

假设以 p_c 作为中心像素点，在其局部邻域内的所有像素点 $\{p_1, p_2, \dots, p_n\}$ 需要满足：(1) 运动目标在连续的两帧中移动距离较小；(2) 所有像素点 $\{p_1, p_2, \dots, p_n\}$ 具有相同的光流。基于假设(1)和(2)，我们可以对邻域内的所有的像素点 $\{p_1, p_2, \dots, p_n\}$ 列出如下基本光流方程

$$\begin{cases} \frac{\partial I}{\partial x}(p_1)V_x + \frac{\partial I}{\partial y}(p_1)V_y = -\frac{\partial I}{\partial t}(p_1) \\ \frac{\partial I}{\partial x}(p_2)V_x + \frac{\partial I}{\partial y}(p_2)V_y = -\frac{\partial I}{\partial t}(p_2) \\ \dots \\ \frac{\partial I}{\partial x}(p_n)V_x + \frac{\partial I}{\partial y}(p_n)V_y = -\frac{\partial I}{\partial t}(p_n) \end{cases} \quad (7-3)$$

其中 $\frac{\partial I}{\partial x}(p_i)$ ， $\frac{\partial I}{\partial y}(p_i)$ 和 $\frac{\partial I}{\partial t}(p_i)$ 分别表示当前时刻下，图像 I 在像素点 p_i 位置处对变量 x ， y ， t 的偏导数。式子(7-3)可简写为

$$Av = b \quad (7-4)$$

其中，

$$A = \begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \dots & \dots \\ I_x(p_n) & I_y(p_n) \end{bmatrix}, \quad v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}, \quad b = \begin{bmatrix} I_t(p_1) \\ I_t(p_2) \\ \dots \\ I_t(p_n) \end{bmatrix} \quad (7-5)$$

其中， $I_x = \frac{\partial I}{\partial x}$ ， $I_y = \frac{\partial I}{\partial y}$ ， $I_t = \frac{\partial I}{\partial t}$ 。显然，方程数大于未知数，因此公式(7-4)

可解，L-K 光流法利用最小二乘法对方程组进行求解，可得

$$v = (A^T A)^{-1} A^T b \quad (7-6)$$

然后将式(7-6)进行展开可以得到光流 (V_x, V_y) ，其表达式如下：

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n I_x(p_i)^2 & \sum_{i=1}^n I_x(p_i)I_y(p_i) \\ \sum_{i=1}^n I_y(p_i)I_x(p_i) & \sum_{i=1}^n I_y(p_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_{i=1}^n I_x(p_i)I_t(p_i) \\ -\sum_{i=1}^n I_y(p_i)I_t(p_i) \end{bmatrix} \quad (7-7)$$

根据式 (7-7) 即可获得中心像素点 p_c 的光流。

我们基于 K-L 光流法对网格内的目标进行运动建模, 首先利用 7.2.1 节的方法获得图像中的特征点, 然后利用 K-L 光流法持续的提取特征点的运动信息。在利用光流法进行运动建模后, 第 j 个网格内的第 i 个特征像素点 p_i^j 的运动矢量可由式 (7-8) 表示,

$$\vec{v}_i^j = (V_{l,x}^j, V_{l,y}^j) \quad (7-8)$$

则所有在第 j 个网格中的 n 个特征点的运动矢量可以组合成运动向量集合 M_j ,

$$M_j = \{\vec{v}_1^j, \vec{v}_2^j, \dots, \vec{v}_n^j\} \quad (7-9)$$

网格中获得的运动向量 M_j 可以有效表征视频中运动群体的运动状态, 然后下面我们将视频目标的运动模型转换成能量模型, 并根据能量模型的取值判断视频是否出现异常。

7.2.3 基于瞬时能量模型的异常事件检测方法

在物理世界中, 万事万物都处于运动状态, 我们通常利用动能 E_k 来表示运动物体所蕴含的能量,

$$E_k = \frac{1}{2}mv^2 \quad (7-10)$$

其中 m 表示运动物体的质量, v 表示运动物体的速度。因此我们可以将上一节基于光流法的运动模型转换成能量模型^[11]。

在上节中, 我们将视频中的前景运动区域进行了网格化处理, 并且能够获得单个网格的运动向量 M_j 。在后续的处理中, 我们以单个网格作为基本处理单元, 通过计算出单个网格的动能从而获取整幅图像的能量流动。网格运动状态的整体速度矢量 \vec{v}^j 可以近似由式 (7-11) 表示,

$$\vec{v}^j = \frac{1}{m} \sum_{l=1}^m \vec{v}_l^j \quad (7-11)$$

整体速度矢量 \vec{v}^j 在 x 轴和 y 轴上的分量 V_x^j 与 V_y^j 分别为:

$$V_x^j = \sqrt{\frac{1}{m} \sum_{l=1}^m V_{l,x}^2} \quad (7-12)$$

$$V_y^j = \sqrt{\frac{1}{m} \sum_{l=1}^m V_{l,y}^2} \quad (7-13)$$

但是如果想要最终获得网格的动能 E_j ，我们还缺少能够有效表征网格质量的物理量。这里我们可以利用 7.2.1 节求得的运动区域像素面积占比 r 来近似的表示网格的质量。因此最终，第 j 个网格的动能 E_j 可由式 (7-14) 表示

$$E_j = r_j \cdot \left[(V_x^j)^2 + (V_y^j)^2 \right] \quad (7-14)$$

在监控视频中，正常情况下，监控的运动目标会以较为平稳的速度进行移动，因此运动群体呈现出的动能也应该是平稳变化的。此时对于视频图像中的第 j 个网格来说，从第 k 帧开始连续两帧的动能差 $\Delta E_j[k]$ 可以表示为

$$\Delta E_j[k] = E_j[k+1] - E_j[k] \quad (7-15)$$

并且满足:

$$\Delta E_j[k] \leq \varepsilon \quad (7-16)$$

其中， ε 表示动能差的阈值，该值可以通过大量的实验统计正常视频中运动目标的动能差值的最大值来获得。

通过对包含异常事件的视频进行分析可以发现，当视频中出现异常事件，如人群短时聚集、人群惊慌逃散、物体爆炸、建筑物倒塌等，往往会出现局部特征速度的突变，使得视频图像中运动目标的动能从一个量级跳变为另外一个量级。为了表示这种能量的突变，我们提出瞬时能量模型来表征这种动能的突变，并通过捕获动能突变这种情况来检测视频中的异常事件。

定义从第 k 帧开始，连续两帧视频中总动能的差值为 $\psi[k]$ ，

$$\psi[k] = \sum_{j=1}^{l[k]} \Delta E_j[k] \quad (7-17)$$

其中 $l[k]$ 表示在第 k 帧图像中包含的前景运动区域网格的个数，且 $\psi[k]$ 满足:

$$\psi[k] = \begin{cases} \psi[k] & \text{when } \psi[k] > l[k] \cdot \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (7-18)$$

定义人群瞬时能量 $\phi[k]$ 如下：

$$\phi[k] = \kappa \sum_{k=k_0}^k \psi[k] \quad (7-19)$$

其中 κ 为总动能差 $\psi[k]$ 的放大因子，通过放大因子 κ 可更加准确的表征运动群体出现异常时的变化特性，这里我们取 $\kappa=10$ 。此外， k_0 满足：

$$\begin{cases} \phi[k_0 - 1] = 0 \\ \psi[k_0] \neq 0 \end{cases} \quad (7-20)$$

当 $k - k_0 = \tau$ 时，令 $\phi[k] = 0$ ，也就是瞬时能量 $\phi[k]$ 只需要计算 τ 帧的视频图像，然后就自动的被置为 0，这样做的目的是为了降低对异常事件的误报。

当运动群体处于平稳的运动状态时，瞬时能量 $\phi[k] = 0$ 。如果视频中出现了异常情况时，满足 (7-18) 阈值条件的总动能差 $\psi[k]$ 被放大因子 κ 放大以后会逐渐累加，这样可以使 $\phi[k]$ 增加，当超过异常阈值 α 时，则认为视频中出现异常事件，如式 (7-21) 所示。

$$R(\phi[k]) = \begin{cases} normal & \phi[k] < \alpha \\ abnormal & \phi[k] \geq \alpha \end{cases} \quad (7-21)$$

7.3 模型工作流程

异常事件检测模型的工作流程如下图所示：

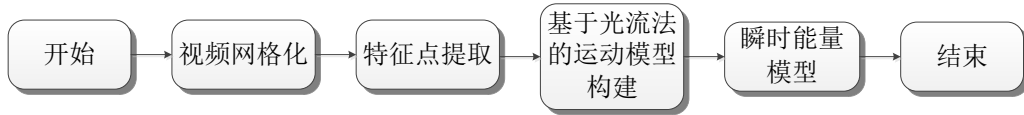


图 30 异常事件检测方法流程图

7.4 实验分析

我们从网上下载了四组视频，其中一组为包含了正常情景的视频，其余三组为包含了**爆炸**，**人群逃散**等异常事件的视频。我们首先对这些视频进行了网格化处理，然后在网格中提取运动特征点，并利用光流法构建运动模型。如图 31 所示，我们从视频中选取了兩帧图像，这两帧图像相隔 5 帧，根据这两组图像中特征点的移动构建了光流模型，如图 31 中最右侧子图。当视频中有异常事件时，

这两组图像中会出现大幅度大范围的特征点运动，而光流模型可以有效的捕获特征点的运动，如图 31 (b) (c) (d) 所示。为了对特征点的运动幅度进行量化，我们将特征点的运动转换成动能，然后根据动能的大小来检测出视频中是否出现异常事件。

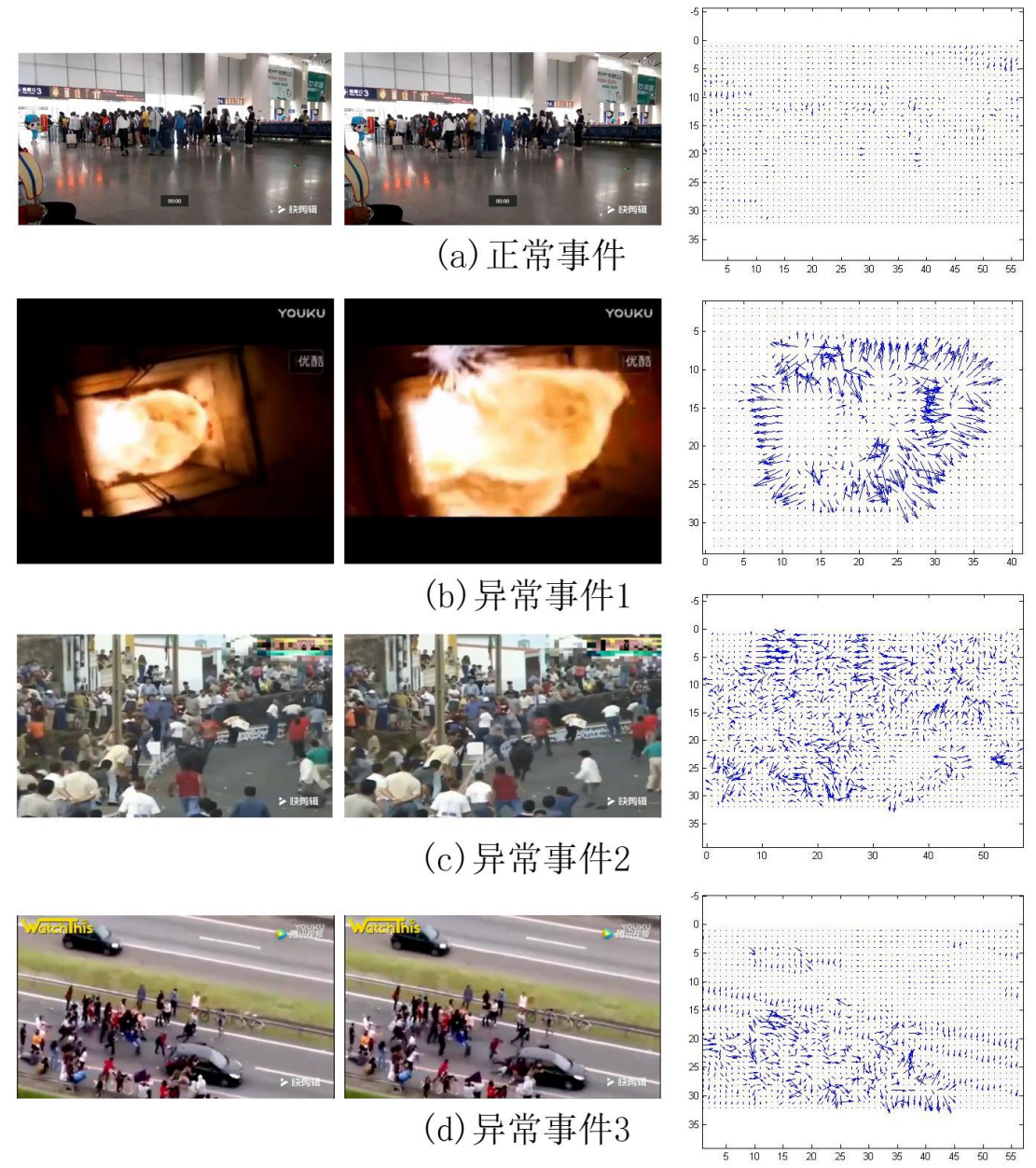


图 31 异常事件检测

8 模型评价

针对问题一——静态背景下的前景提取问题，我们采用了基于样本随机聚类的方法构建背景模型，并基于单帧数据快速的对背景模型进行初始化。该模型能

够在静态环境下有效的分离运动前景和背景，并且在进行模型初始化的时候，仅仅依靠单帧数据，因此在视频初期就能取得较好的效果。但是，该模型对动态背景较为敏感。

针对问题二——动态背景下的前景提取问题，我们在问题一方法的基础上结合了模糊更新准则，自适应的改变背景模型的检测阈值和更新频率，提高针对动态变化的背景模型构建的准确性。但是，针对视频抖动问题，该方法无法有效解决。

针对问题三——抖动背景下的前景提取问题，我们采用特征点轨迹模型，有效捕获前景运动目标的运动轨迹，完成视频前、背景运动轨迹的有效区分，然后根据若干帧中运动轨迹的抖动量，构建抖动补偿模型，用来降低背景抖动产生的不良影响。

针对问题四——八组视频的前景匹配问题，实验证明我们的模型在大部分场景中（七组视频中）都可以拥有 93% 以上的前景检测精确度；但是由于光照的变化，我们模型在 Lobby 模型中的前景检测精度仅有 66%，这主要是由于我们模型是基于像素值的变化进行设计的，该模型对光照会比较敏感。未来，我们期望结合深度学习模型，以提高我们模型在光照改变场景中的鲁棒性。

针对问题五——多视角前景提取问题，我们采用目标存在概率模型，依据成像的原理构建了目标窗结构来表示目标存在的概率。在得到目标概率存在场以后，利用 Mean-Shift 算法搜索局部最优点。根据各个位置的目标窗结构，将结果重新投影到各个图像上，对应的目标窗结构的子窗将其对应视角的运动目标检测出来，并根据映射关系确定各目标的对应关系。

针对问题六——异常事件检测问题。我们首先对视频图像进行网格化，然后从单元网格内提取运动特征点，然后利用光流法构建运动模型捕获图像帧间的运动。然后利用动能定理将运动模型转换成瞬时能量模型，并根据动能大小来检测是否出现异常。

9 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.
- [9]. 徐剑, 丁晓青, 王生进. 基于目标存在概率场的多视角运动目标检测与对应算法[J]. *自动化学报*, 2008, 34(5): 609-612.
- [10]. Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences[J]. *IEEE Transactions on Image processing*, 2011, 20(6): 1709-1724.
- [11]. 张兵. 智能视频监控中人群密度分析及突发异常行为检测[D]. 北京交通大学, 2016.

10 附录

在本文中，主要利用 Matlab, VC 并结合 OpenCV 开源库进行建模与算法实现，模型的选取涉及到 Vibe 模型，POM 模型，光流模型等。由于计算机程序和代码数量多，所使用的命令和编写的计算机源程序不在论文中一一录入，将整理成附件上传。