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

武汉纺织大学

学　　校	
参赛队号	10495013
队员姓名	1.柳云鹏
	2.毛霞
	3.侯鹏

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：

本文运用混合高斯模型和光流法在静态、动态背景下前景提取方法，在镜头发生抖动情况下用仿射变换，背景补偿帧差的方法提取前景，在多视角视频拍摄情况下，构建跨视角的空间特征，对多视角拍摄视频聚类分析提取前景，最后结合光流追踪和角点平均移动速度识别人群奔跑的异常行为。

针对问题一：静态背景下，摄像机与监控场景保持相对静止，运用混合高斯模型的背景减除法获取运动目标的方法。

针对问题二：动态背景下，运用自适应场境的运动目标跟踪方法，由混合高斯模型实现目标检测，由目标区域 LK 光流预测实现目标跟踪，并融合了“运动”背景局部更新来实现跟踪系统对环境的自适应性。

针对问题三：在相机抖动的情况下，由于目标和背景都是各自独立运动的，在提取前景运动目标时需要考虑由相机自身运动引起的背景变化。采用仿射变换来估计图像间背景变换的方法，从背景变换补偿帧差中得到运动目标的区域。

针对问题四：利用本文所构造的建模方法，从每组视频中选出包含显著前景目标的视频帧标号。

针对问题五：对于多视角视频拍摄，通过基于场景变换检测的视频分割方法，对多视角视频分割为不同场景的视频段，然后对个视角场景视频段特征做主成分分析获取各个场景视频段投影矩阵，构建各视角场景视频段的基空间，然后对构建的跨视角的特征超空间聚类分析，生成多视角的视频摘要。对于多视角图像显著性目标检测，分析场景在不同视角下前景与背景之间的空间关系，通过空间投影计算出不同视角下目标周围被遮挡的背景区域，并将其应用到显著性目标检测中。

针对问题六：对于公共场所人群奔跑异常行为，运用一种基于平均运动速度的检测方法。根检测图像上的角点，采用光流法对这些角点进行跟踪并提取出产生运动的角点，进而获得运动角点在视频序列中连续两帧间的运动向量，计算出人群整体的运动速度，从而检测人群的奔跑行为。

关键词：前景提取；混合高斯模型；光流法；仿射变换；空间聚类；平均运动速度

1 问题重述

1.1 研究背景

视频前景检测是智能视频监控、行为分析等计算机视觉系统的关键步骤，伴随着近几年来人们对于社会安全问题的日益关注而得到广泛的研究和应用。

在各种检测算法中，背景差分被认为是提取前景目标的最有效方法。关于北京建模的方法大致可以归为 2 类。一类方法为基于像素级的背景模型，包括混合高斯模型、贝叶斯模型、马尔科夫随机场模型、背景字典模型、像素采样模型。另一类方法是基于图像级的背景模型，最新的研究热点是基于压缩感知理论的前景检测方法。

1.2 研究问题

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？

问题 4：在附件 3 中提供了 8 组视频。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。

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无人群短时聚集、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 问题分析

2.1 问题一的分析

静态背景下，摄像机与监控场景保持相对静止，在此类摄像机静止，目标运动的情形下，运动目标的检测较为容易，常用的方法包括帧间差分法、背景建模法等。本题中，我们采用的是混合高斯模型，从包含静态背景的视频中提取运动目标，并将其识别为前景。

2.2 问题二的分析

动态背景下的运动目标检测方法主要可以分为基于帧间背景补偿的方法、基于初始背景模型构造的方法和基于多帧运动轨迹的方法三类。这里用高斯混合模型与光流残差相结合的前景目标分割方法。首先利用高斯混合模型建模，计算初略的前景区域，然后利用光流残差法滤除其中动态纹理背景干扰，再采用形态学处理获得最终前景目标区域，通过实验对比，本文方法可获取更为精确的视频前景区域，且对于不同类型动态干扰均有较好的抵抗能力。

2.3 问题三的分析

相机抖动拍摄视频时存在着两个相对独立的运动：相机抖动引起的背景运动

和运动目标在环境中的运动。为了检测出场景中除相机抖动之外的运动目标，需要对图像的背景进行补偿，使连续几帧图像的背景稳定在同一帧图像相同位置上，从而使运动目标凸显出来。

2.4 问题四的分析

本题通过图片显示从每组视频中提取的前景目标，并对选出包含显著前景目标的视频帧标号。

2.5 问题五的分析

多视角视频的结构化分析就是在时间轴上对视频数据完成不同层次的分割，实现多视角视频数据由原来的非结构化的数据格式转化成结构化的数据格式。视频数据的结构化就是把视频帧划分为若干个不同的集合，使其成为不同层次上的结构化的实体。划分一般包含对于视频数据镜头边界的划分和场景边界的划分。在镜头边界划分中通常采用纹理、边缘信息、颜色直方图、运动信息图等方法来实现，场景的划分一般采用相邻镜头之间的相关性和视频内容来实现。

在多视角视频中，一个视角中场景与场景之间在时间轴上具有一定的相关性，不同视角之间空间上相邻的场景视频之间也有一定的联系和相关性。在不同视角中，空间上相邻的视频数据段在内容上具有一定的相似性。

2.6 问题六的分析

在公共场所，场景中的人数变化不定，人与人之间可能存在一定的行为关系，针对人群奔跑异常行为检测，分为两个阶段：①行为描述阶段，即在原视频中提取角点运动方向对奔跑行为进行描述；②异常行为检测阶段，这里运用光流法对这些角点进行跟踪并提取出产生运动的角点，进而获得运动角点在视频序列中连续两帧间的运动向量，计算出人群整体的运动速度，从而检测人群的奔跑行为。

3 问题求解

3.1 问题一的建模

问题一要求，对一个不包含动态背景、摄像头稳定拍摄的监控视频，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，将视频中的前景提取出来，并将背景变换为黑色，以突出前景目标。

GMM 能稳定快速地检测出疑似运动前景，具体定义和更新方法如下：令 $M(x,y,t)$ 表示第 t 帧视频图像 (x,y) 点的像素值。单高斯背景建模法认为背景点的取值是一个服从高斯分布的随机变量，即取值的概率分布函数为：

$$p(M(x,y,t))=\frac{1}{\sqrt{2\pi\delta_t^2(x,y)}}\exp\left[-\frac{\left(M(x,y,t)-\mu_t(x,y)\right)^2}{2\delta_t^2(x,y)}\right] \qquad (1)$$

其中 $\mu_t(x,y)$ 和 $\delta_t^2(x,y)$ 分别为 $M(x,y,t)$ 的均值和方差。

单高斯背景建模算法分为三个基本步骤：

（1）模型初始化

首先对每个像素的均值和方差进行初始化。初始化过程一般基于前 N 帧进行，即

$$\mu_N(x,y)=\frac{1}{N}\sum_{i=1}^{N}M(x,y,i) \tag{2}$$

$$\delta_N^2(x,y)=\frac{1}{N}\sum_{i=1}^{N}\left[M(x,y,i)-\mu_N(x,y)\right]^2 \tag{3}$$

（2）目标检测

初始化完成后就可基于（4）式对后续视频帧进行目标检测。根据输出结果 $O(x,y,t)$ 的取值判断当前像素的属性，其中“1”表示前景像素，“0”表示背景像素。其中 λ 是一个阈值常数，其值可根据实验中的先验知识确定。

$$O(x,y,t)=\begin{cases}1, \left|M(x,y,t)-\mu_{t-1}(x,y)\right|>\lambda*\delta_{t-1}(x,y)\\ 0, \text{其他}\end{cases} \tag{4}$$

（3）模型更新

利用新的视频帧对模型参数进行更新，更新公式如下：

$$\mu_t(x,y)=(1-\alpha)\mu_{t-1}(x,y)+\alpha M(x,y,t) \tag{5}$$

$$\delta_t^2(x,y)=(1-\alpha)\delta_{t-1}^2(x,y)+\alpha\left[M(x,y,t)-\mu_t(x,y)\right]^2 \tag{6}$$

其中 α 是背景更新速率，α 越大，表示更新时当前帧所占的比例越大。

另外，在更新时需要考虑被判断为前景的像素点是否参与更新。若参与更新，运动目标经过的地方会留下残影，出现拖尾现象。图 1 对比了前景像素点参与更新与不参与更新的实验结果。本文实验中采用前景像素点不参与背景更新的方法，即式（5）（6）扩展为式（7）（8）。

$$\mu_t(x,y)=\begin{cases}(1-\alpha)\mu_{t-1}(x,y)+\alpha M(x,y,t), \text{当} O(x,y,t)=0\\ \mu_{t-1}(x,y), \text{当} O(x,y,t)=1\end{cases} \tag{7}$$

$$\delta_t^2(x,y)=\begin{cases}(1-\alpha)\delta_{t-1}^2(x,y)+\alpha\left[M(x,y,t)-\mu_t(x,y)\right]^2, \text{当} O(x,y,t)=0\\ \delta_{t-1}^2(x,y), \text{当} O(x,y,t)=1\end{cases} \tag{8}$$

3.2 问题二的建模

3.2.1 光流残差法

本节将利用光流残差法滤除动态纹理背景带来的干扰。光流法被广泛用于运动目标分割，但是传统方法计算量大，难以达到实时，为此特征光流法、区域光

流法、金字塔光流法等一些快速算法被提出。光流法基于两个假设：任何物体点所观察到的亮度随时间恒定不变；图像平面内的邻近点以类似的方式进行移动。文献[20]发现动态纹理区域难以满足以上两个假设，其光流残差(optical flow residual，OFR)通常较大，提出了基于光流残差的动态纹理区域检测方法，并取了较好的效果。该方法的基本思想为：

$$R_{of} = \lhd I(x+u, y+v, t+1) - I(x, y, t) \rhd_{\omega}$$

式中，R_{of}为$t+1$帧与t帧之间的光流残差；$I(x,y,t)$表示视频t时刻(x,y)处图像的灰度值；u、v为位移矢量；$<\bullet>_{\omega}$表示与高斯核ω的卷积；$I(x+u,y+v,t+1)$通过插值计算亚像素精度。由于非动态纹理区域满足亮度守恒的假设：

$$I(x+u, y+v, t+1) = I(x, y, t)$$

其对应的光流残差值明显低于动态纹理区域，这一特性可较好地区分动态纹理和非动态纹理产生的前景运动区域。实际应用中，在提取多帧连续的光流残差图像后，需进行三维中值滤波，以降低噪声的影响，提高算法稳定性。

3.2.2 算法及实验结果分析

为了有效提取运动前景区域并克服视频中动态纹理的干扰，本文构造算法如下：

1）获取当前帧视频，利用 GMM 提取运动区域Ω_G。

2）计算当前帧对应光流残差，采用大津法获取动态纹理区域Ω_D。

3）计算滤除动态纹理后的前景运动目标区域$\Omega_O = \Omega_G - \Omega_D$，利用形态学方法对$\Omega_O$中空洞和噪声进行适当修复。

为了验证算法的有效性和鲁棒性，本文采用公开的动态纹理数据集进行对比试验。动态纹理数据集包含自然场景、人物、动物、交通场景等大量测试视频。用高斯混合模型与光流残差相结合的前景目标分割方法，实验结果显示本文方法可提取运动前景目标的同时较好地抵抗动态纹理背景的干扰，可以在有效提取视频运动前景目标的基础上，较好客服不同强度的动态纹理背景干扰，解决了在动态背景下视频前景的分割提取。

3.3 问题三的建模

3.3.1 全方位运动背景补偿

基于运动补偿的思想是将视频中的运动分为由摄像机运动引起的全局运动和由运动目标引起的局部运动，由于全局运动和局部运动是两个相对独立的运动，可以通过对摄像机的自运动补偿来消除全局运动，从而检测出发生局部运动的运动目标。

为了建立两帧图像之间的仿射变换关系，需要构造仿射变换矩阵，得到仿射变换参数。仿射变换模型可以下式表示：

$$\begin{bmatrix} x_i' \\ y_i' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} \quad (1)$$

其中，$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$表示旋转矩阵，$\begin{bmatrix} c \\ d \end{bmatrix}$表示平移矩阵，$\begin{bmatrix} x_i \\ y_i \end{bmatrix}$表示第$j-1$帧图像中第$i$个特征点坐标，$\begin{bmatrix} x_i' \\ y_i' \end{bmatrix}$表示第$j$帧图像中与$\begin{bmatrix} x_i \\ y_i \end{bmatrix}$拟匹配的特征点坐标，$i=1,2,...,k$，$j=2,3,...$。

整理后为：

$$\begin{bmatrix} x_i' \\ y_i' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & c \\ a_{21} & a_{22} & d \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \quad (2)$$

其中，矩阵$\begin{bmatrix} a_{11} & a_{12} & c \\ a_{21} & a_{22} & d \end{bmatrix}$称为仿射变换矩阵。仿射变换矩阵可以通过图像上的特征点来构造。首先在前一帧图像上选取一些特征点，这些特征点一般数量很大，分布于整个图片，然后在后一帧图像上跟踪特征点，通过最小二乘法拟合出整幅图像的仿射变换参数。由式(2)可得，基于k对相邻帧间匹配的特征点坐标的参数估计方程为：

$$\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ c & d \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \cdots & \cdots & \cdots \\ x_k & y_k & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_1' & y_1' \\ x_2' & y_2' \\ \cdots & \cdots \\ x_k' & y_k' \end{bmatrix} \quad (3)$$

式(3)可通过最小二乘法求解。由于机器人视觉图像复杂，为保证算法鲁棒性，k往往大于3。

3.3.2 处理过程

我们以网站 http://wordpress-jodoin.dmi.usherb.ca/dataset2014/ 上traffic 视频为例，对晃动摄像头拍摄的视频进行处理。我们首先从视频中第

1128 帧和第 1130 帧，并进行灰度处理。下面我们将两个帧并排显示，并生成一个红色的青色复合物来说明它们之间的像素差异。由图中可以看出两帧之间显然存在较大的垂直和水平偏移。

第一步：从视频中读取帧

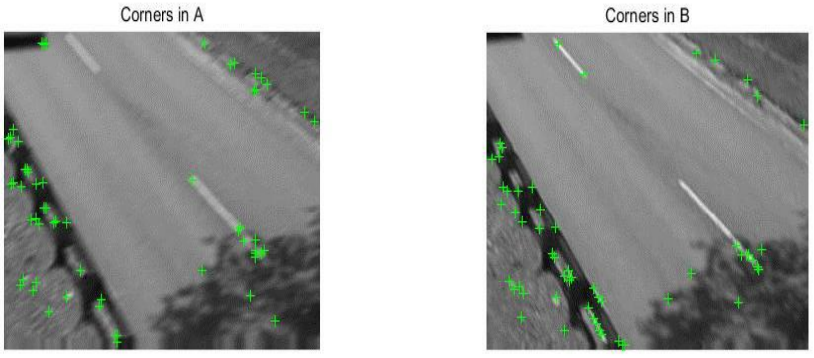
这里我们读取视频序列的前两帧。我们将它们作为强度图像读取，因为稳定算法不需要颜色，因为使用灰度图像可以提高速度。下面我们同时显示两个框架，并且我们生成一个红-青色复合材料来说明它们之间的像素差异。两帧之间显然有一个大的垂直和水平偏移。

第二步：从每个框架收集要点

我们的目标是确定一个能够纠正两帧之间失真的转换。使用 Estimate Geometric Transform 函数，这将返回一个仿射变换。作为输入，我们必须在两个帧之间提供一组点对应关系。为了产生这些对应关系，首先从两个框架中收集候选点，然后选择它们之间的可能对应关系。

在这个步骤中，我们为每个帧生成这些候选点。为了使这些点在其他框架中具有对应点的最佳机会，我们需要围绕突出图像特征的点。为此，我们使用 detectFASTFeatures 函数，它实现了最快的角点检测算法之一。

两帧的检测点如下图所示。观察它们中有多少覆盖了相同的图像特征，例如道路两边的点和中间白线的点。

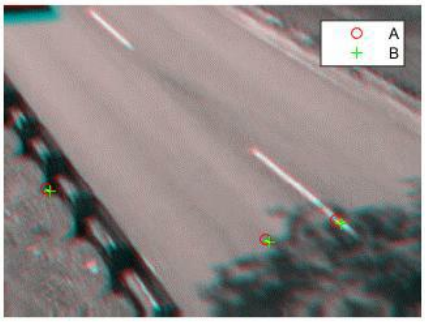


第三步：选择点之间的对应关系

接下来，我们选择上面得出的点之间的对应关系。对于每个点，我们提取以它为中心的快速视网膜关键点（FREAK）描述符。我们在点之间使用的匹配成本是汉明距离，因为 FREAK 描述符是二进制的。帧 A 和帧 B 中的点匹配。注意，没有唯一性约束，因此来自帧 B 的点可以对应于帧 A 中的多个点。

匹配在当前和以前的帧中找到的特征。由于 FREAK 描述符是二进制的，所以 matchFeatures 函数使用汉明距离来查找相应的点。

下面的图像显示了上面给出的相同的颜色复合体，但是添加了来自框架 A 的红色点，以及来自框架 B 的点为绿色的点。点之间绘制黄线，显示上述步骤选择的对应关系。这些通信中的许多都是正确的，但也有大量的异常值。



第四步：估计嘈杂信件的转换

在上一步中获得的许多点对应是不正确的。但是，我们仍然可以使用 M 估计器 SAmple Consensus（MSAC）算法来导出两个图像之间的几何变换的鲁棒估计，该算法是 RANSAC 算法的一个变体。MSAC 算法在估计量测转换函数中实现。该函数在给出一组点对应关系时，将搜索有效的 inlier 对应关系。从这些，它将导出仿射变换，使得来自第一组点的内部值与来自第二组的内在关系最紧密地匹配。该仿射变换将是一个 3 乘 3 矩阵的形式：

$$\begin{bmatrix} a_1 & a_3 & 0 \\ a_2 & a_4 & 0 \\ t_x & t_y & 1 \end{bmatrix}$$

参数 a 定义变换的缩放，旋转和剪切效果，而参数 t 是转换参数。该变换可以用于扭曲图像，使得它们的相应特征将被移动到相同的图像位置。

下面是一个颜色复合材料，显示与重新投影的框架 B 重叠的框架 A，以及重新注射的点对应。 结果是非常好的，内部的通信几乎完全一致。图像的核心都很好地对准，使得红-青色复合材料在该区域变得几乎纯黑白色。

内联对应关系是否全部在图像的背景中，而不是前景，其本身不对齐。这是因为背景特征足够远，它们的行为就像在无限远的飞机上一样。因此，即使仿射变换被限制为仅改变成像平面，这里足以对准两个图像的背景平面。此外，如果我们假设背景平面在帧之间没有移动或显着改变，则该变换实际上是捕获相机运动。 因此，纠正这将使视频稳定。只要帧之间的相机的运动足够小，或者相反，如果视频的采样时间足够高，则该状态将保持。

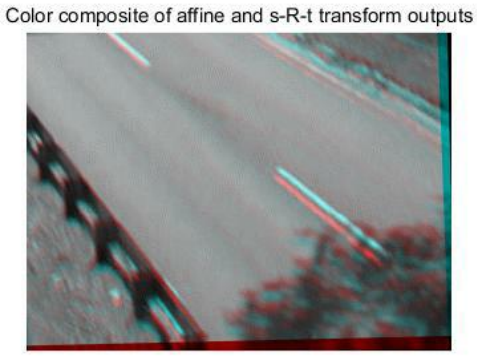
第五步:变换近似和平滑

给定一组视频帧 T_i，$i = 0,1,2,...$，我们现在可以使用上述过程来估计所有帧之间的失真 T_i 和 T_{i+1} 作为仿射变换 H_i。因此，相对于第一帧的帧 i 的累积失真将是所有先前帧间变换的乘积，或

$$H_{cumulative,i} = H_i \prod_{j=0}^{i-1}$$

我们可以使用上面的仿射变换的所有六个参数，但是为了数值简单性和稳定性，我们选择将矩阵重新拟合为更简单的缩放-旋转平移变换。与完全仿射变换的六：一个比例因子，一个角度和两个翻译相比，这只有四个自由参数。这个新的变换矩阵的形式是：

$$\begin{bmatrix} s*\cos(ang) & -s*\sin(ang) & 0 \\ s*\sin(ang) & s*\cos(ang) & 0 \\ t_x & t_y & 1 \end{bmatrix}$$

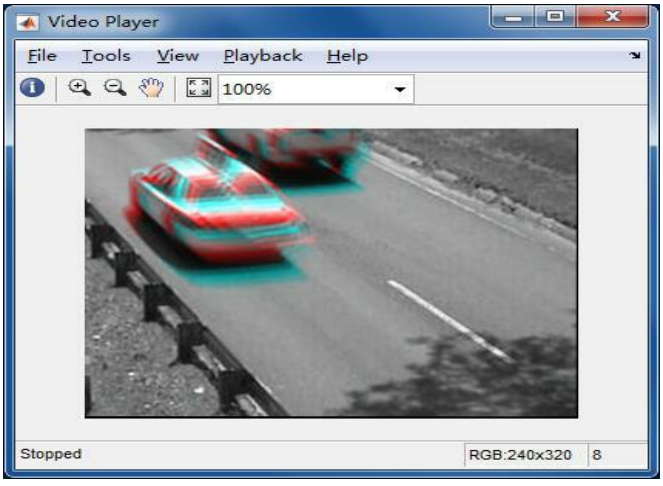


我们通过将上述获得的变换 H 与等值旋转转换等价物 H_{sRt} 拟合，从而显示以下转换过程。为了表明转换转换的错误是最小的，我们用两个变换重新投影框架 B，并将下面的两个图像显示为红-青色复合。 当图像看起来是黑色和白色时，显然，不同重现之间的像素差异可以忽略不计。

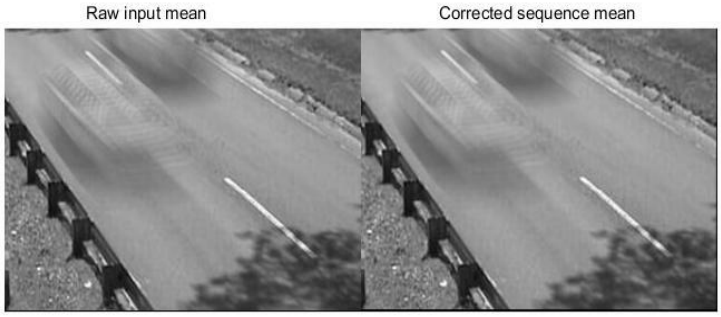
第六步：在完整视频上运行

现在我们应用上述步骤来平滑视频序列。 为了可读性，估计两个图像之间的变换的上述步骤已经被放置在 MATLAB®函数 cvexEstStabilizationTform 中。函数 cvexTformToSRT 还将一般的仿射变换转换为缩放-旋转-转换变换。

在每个步骤中，我们计算当前帧之间的 H 变换。我们把它作为一个 s-R-t 变换，H_{sRt}。 然后我们结合这个累积变换 $H_{\text{cumulative}}$，它描述了自第一帧以来的所有相机运动。 平滑视频的最后两帧在视频播放器中显示为红-青色复合。



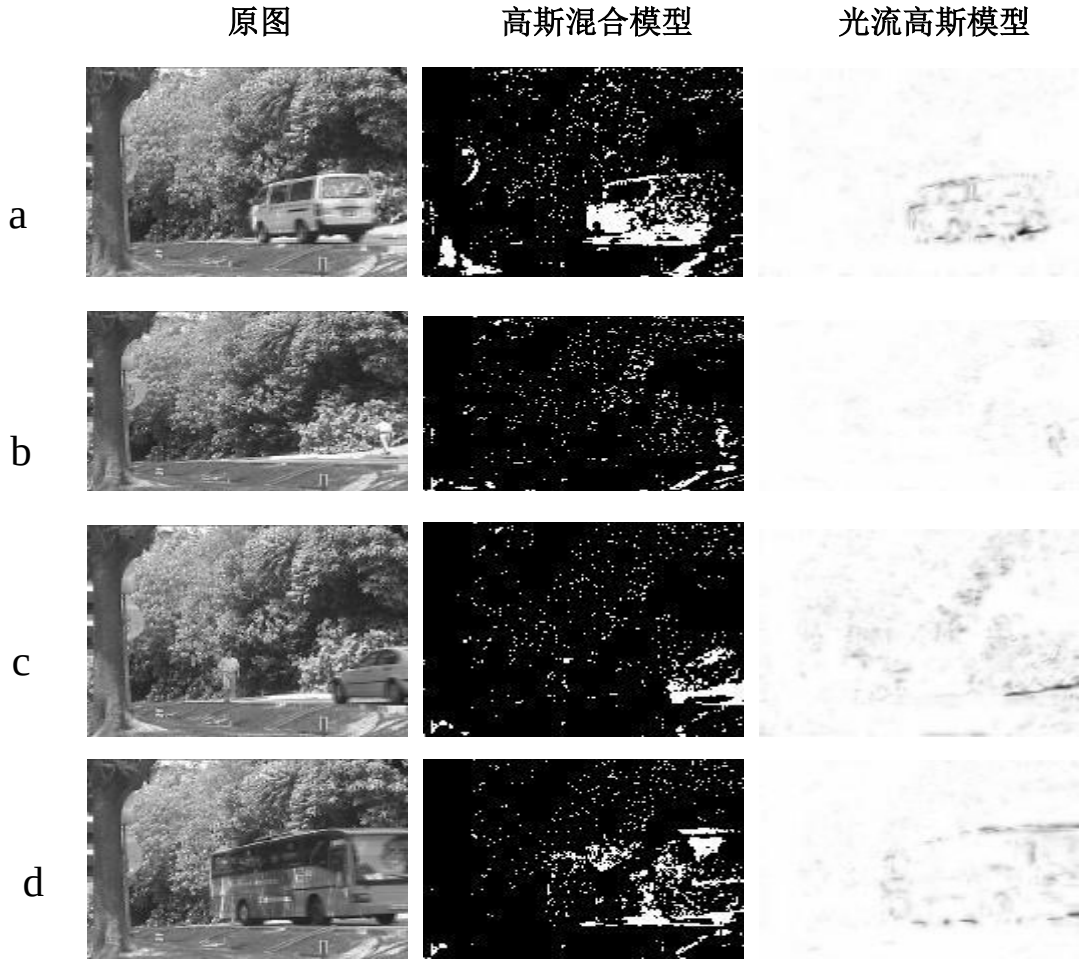
在计算过程中，我们计算了原始视频帧和校正帧的平均值。这些平均值在下面并排显示。左图显示原始输入帧的平均值，证明原始视频中存在大量失真。然而，右侧的校正帧的平均值显示了几乎没有失真的图像核心。虽然前景细节已经模糊。，这显示了稳定算法的功效。



3.4 问题四的建模

在附件 3 中提供了 8 组视频。使用本文第二部分所构造的建模方法，从每组视频中选出包含显著前景目标的视频帧标号。

① 在 Campus 视频中，我们提取的前景来自 200-249（a），306-522（b），600（c），644-683（d），691-712（e），742-876（f），1005-1036（g），1264（h），1331-1370（i），1377-1405（j）帧。

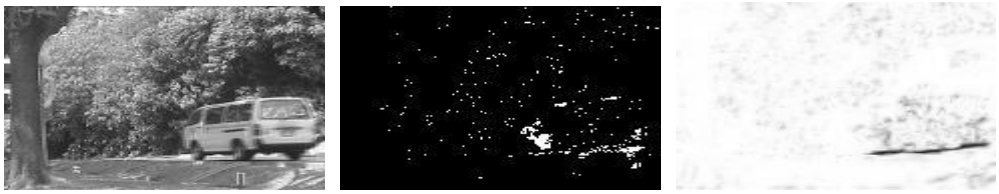


e

f

g

h



i

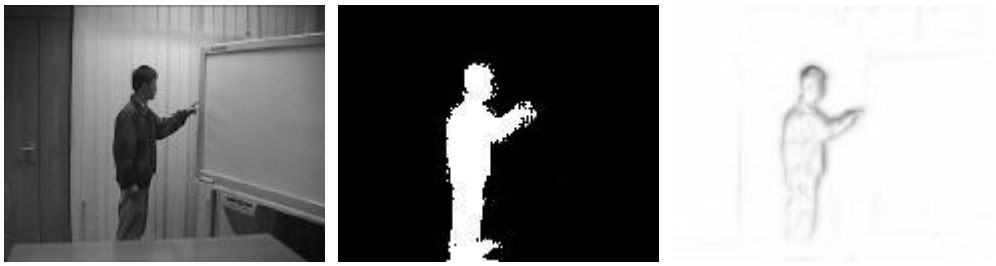
j

② 在 Curtain 视频中，我们提取的前景来自 411（a），967（b），1762-1905（c），2126（d），2175-2314（e），2642（f），2769-2931（g）。

原图　　高斯混合模型　　光流高斯模型

a

b

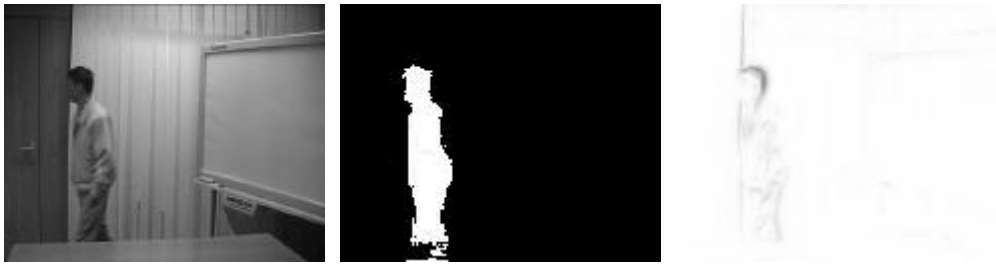


c

d

e

f



g

③ 在 Escalate 视频中，我们提取的前景来自 312（a），464（b），914（c），

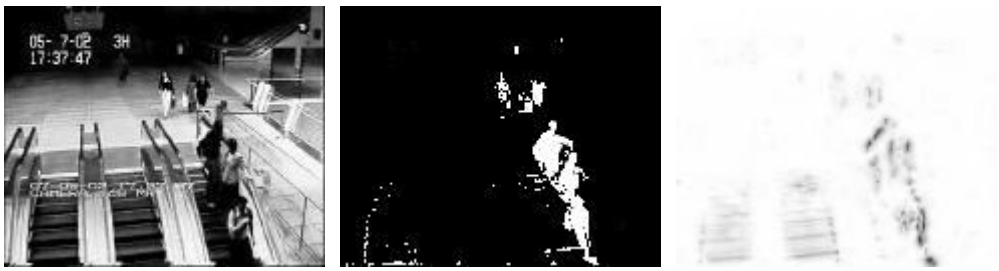
1825（d），2381（e），2754（f），2829（g），3189（h），3301（i），3417（j）。

g

h

i

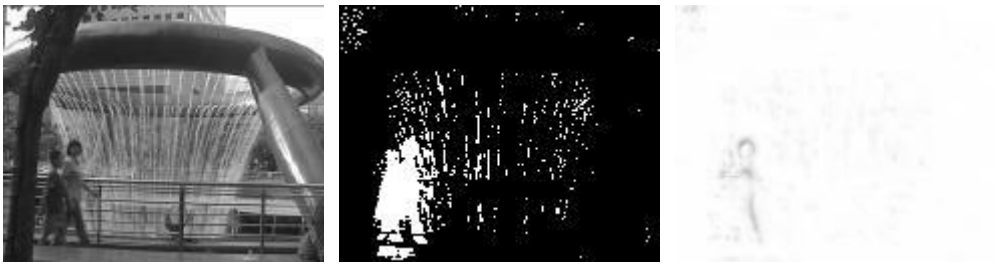
j



④ 在 Fountain 视频中，我们提取的前景来自 141（a），153-213（b），259（c），335（d），403-523（e）。

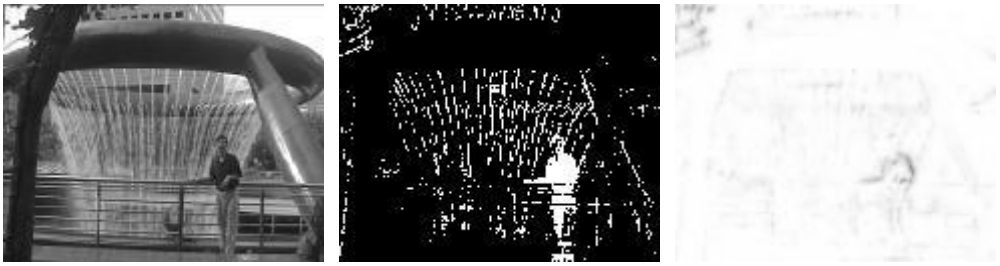
原图　　高斯混合模型　　光流高斯模型

a

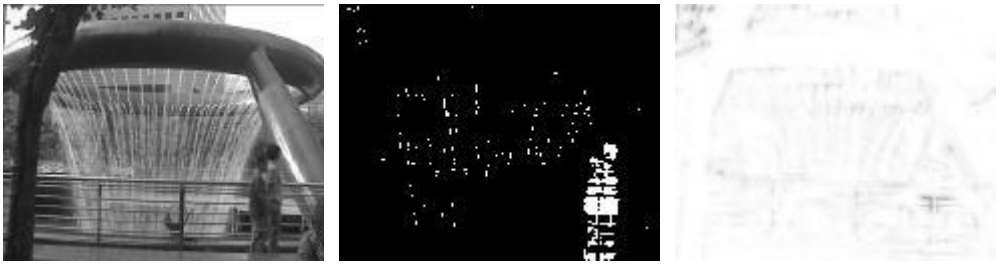


b

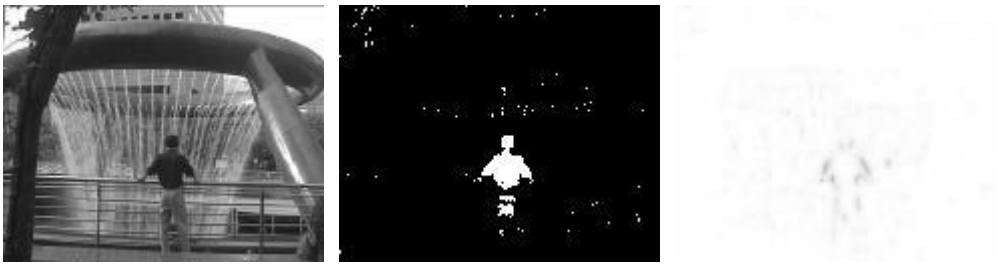
c



d



e



⑤ 在 Hall 视频中，我们提取的前景来自 3（a），122（b），431（c），670（d），3104（e），3195（f），3275（g），3483（h）。

原图 **高斯混合模型** **光流高斯模型**

a


b
c
d
e
f
g

h



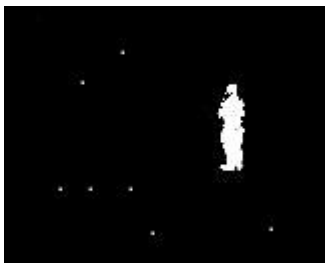
⑥ 在 Lobby 视频中，我们提取的前景来自 79（a），169（b），349（c），521（d），641（e），870（f），1161（g），1247（h），1338（i）。

原图　　高斯混合模型　　光流高斯模型

a



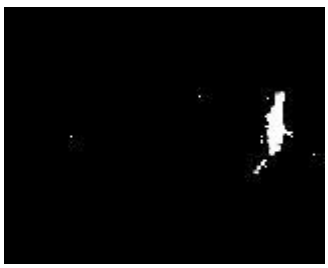
b


c

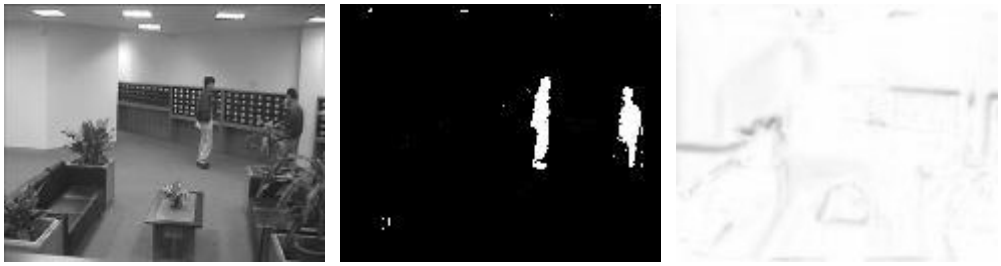


d

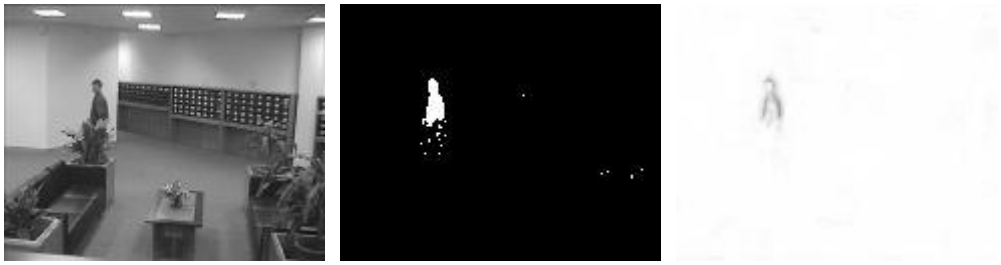




e

f



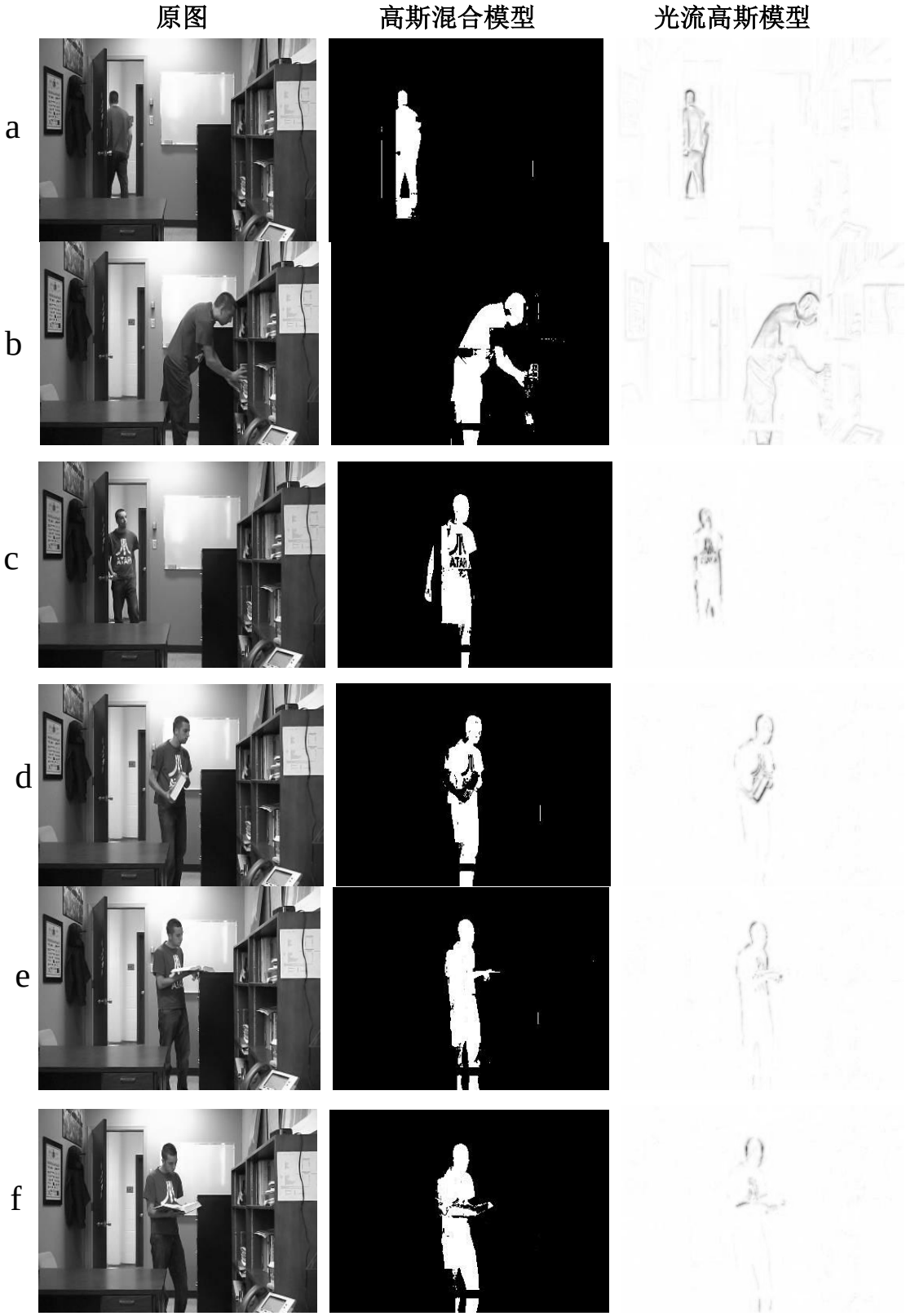
g



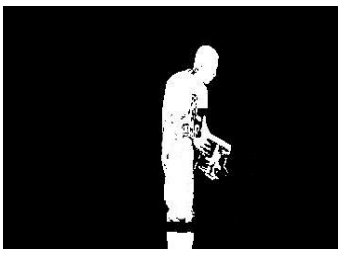
h

i

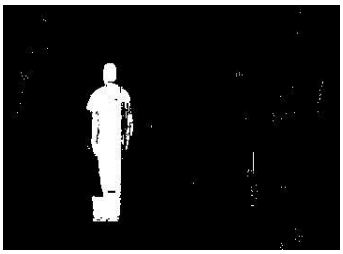
⑦ 在 Office 视频中，我们提取的前景来自 372（a），501（b），650（c），728（d），1049（e），1376（f），1879（g），2006（h）。


原图
高斯混合模型
光流高斯模型
a
b
c
d
e
f

g

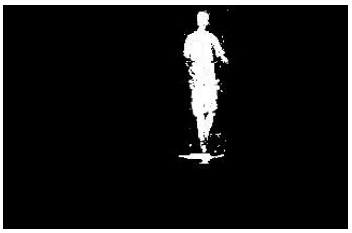

h

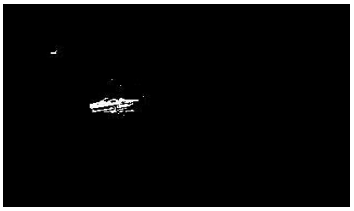
⑧ 在 overpass 视频中，我们提取的前景来自 374（a），591（b），968（c），1551（d），1881（e），2368（f），2554（g）。

原图　　高斯混合模型　　光流高斯模型

a

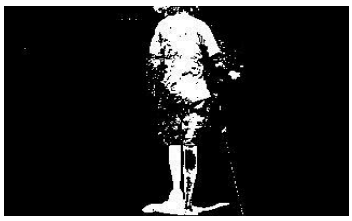

b

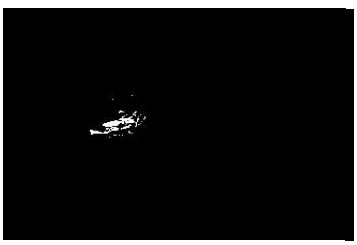
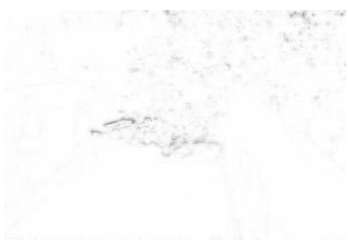
c



d

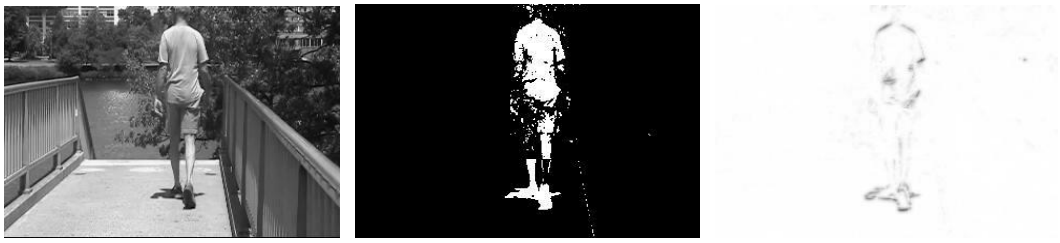




e

e

g



3.5 问题五的建模

多视角联合显著性目标检测包含四个步骤：

第一步：通过视角投影技术得到不同视角间的空间关系。

多视角图像序列中，不同视角的图像都是对同一场景的描述，对多视角图像的处理首先需要将不同视角的图像放置到同一个视角下，然后在相同的观察角度上分析各幅图像之间的异同。此处利用场景的深度图实现视角的投影转换。

第二步：计算不同视角的显著图。

通常不同的视角有不同的显著性图，获得各个视角下的显著性图是进行多视角图像显著性分析的基础，本文方法可以灵活地选取现有显著性检测算法计算单视角显著性图，并在此基础上完成多视角图像融合。

第三步：不同视角显著性图的投影。

通过前两步获得不同视角之间的空间关系和各视角的显著性图后，将其中两侧视角的显著性图根据视角空间关系投影到中间目标视角中，并完成视角投影后的裂纹修复。

第四步：多视角显著性图的融合。

将投影过来得显著性图与当前视角的原始显著性图融合在一起。首先利用投影空洞消除物体周围的干扰背景，然后分析两种背景区域：一是物体两侧空洞以外的区域，二是边缘背景概率模型得到的高概率背景区域，对这两种背景区域进行抑制，最终得到图像中清晰的物体，完成多视角图像的显著性目标检测。

3.5.1 多视角空间关系

数字图像可以抽象为$M \times N$矩阵，矩阵中每个元素代表图像的一个像素。依像素矩阵，以O_0为原点，定义图像坐标系，u、v，各个矩阵元素的坐标代表其在像素矩阵中的列数和行数，进一步用物理单位描述图像坐标系，令图像中心O_1

为坐标原点，X 轴和Y轴分别与u 轴和v轴平行，建立图像物理坐标系。设O_1在图像坐标系中对应点为(u_0,v_0)，单个像素对应的 X 轴物理长度为dX ，对应的Y轴物理长度为dY ，其图像坐标系和物理坐标系中点的转换关系为：

$$\begin{cases} u=\dfrac{x}{dx}+u_0 \\ v=\dfrac{y}{dy}+v_0 \end{cases} \tag{1}$$

为计算简便，以上坐标对应关系改为齐次式：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}=\begin{bmatrix} \dfrac{1}{dX} & 0 & u_0 \\ 0 & \dfrac{1}{dX} & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2}$$

实际应用中需要给定外界环境的坐标系来确定摄像机的位置，称此坐标系为世界坐标系(X_w,Y_w,Z_w)。世界坐标系与摄像机坐标系之间的对应关系如式：

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}=\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}=M_2\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{3}$$

其中，R为3×3正交单位矩阵，t为平移向量，M_2为摄像机的外部参数。

设f为摄像机焦距，(X,Y)为P点图像坐标，(x,y,z)为P点在摄像机坐标系中的坐标，摄像机的透视投影可表示为：

$$s\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}=\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{4}$$

其中，s为比例因子。将式（2）和式（3）代入式（4），可得P点在世界坐标系和图像坐标系中的对应关系：

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}\frac{1}{dX}&0&u_0\\0&\frac{1}{dX}&v_0\\0&0&1\end{bmatrix}\begin{bmatrix}f&0&0&0\\0&f&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}R&t\\0^T&1\end{bmatrix}\begin{bmatrix}X_w\\Y_w\\Z_w\\1\end{bmatrix}=\begin{bmatrix}\alpha_x&0&u_0&0\\0&\alpha_y&v_0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}R&t\\0^T&1\end{bmatrix}\begin{bmatrix}X_w\\Y_w\\Z_w\\1\end{bmatrix}=M_1M_2X_w=MX_w \tag{5}$$

其中，α_x和α_y分别为u轴和v轴上的归一化焦距，与u_0、v_0都是相机内部参数。由式（5）可知，即使有多个视角，只要确定了摄像机的内、外部参数，就可以计算出不同视角之间像素的对应关系。

3.5.2 深度图与空间坐标转换

式（5）中 M 是不可逆的，给定图像坐标(u,v)，由式（5）可以得到三个关于(x,y,z)的三元一次方程，因此，要确定该点的世界坐标至少需要预先知道其 z 坐标。一般摄像机的感光器件获取的图像是二维平面的，无法感知场景中物体的远近距离。本文使用深度图像采集设备获取的深度图像（Depth Image）来确定空间坐标转换时需要的 坐标。深度图由测距传感器得到，通常与原始图像分辨率是一样的，其各个像素描述场景中各个位置与传感器的空间距离。深度图补充了图像的深度这一维度信息，与普通图像共同描述完整场景。下面以两个视角（Cam1 和 Cam2）为例详细介绍坐标转换方法。

设 Cam1 中点P_1的图像坐标为(u_1,v_1)，该点在世界坐标系中对应的点为$P_w(x,y,z)$，P_w的z坐标由深度图计算得到，由式（5）可得P_1点在两个坐标系下的对应关系：

$$s_1(u_1,v_1,1)^T=M_1(x,y,z,1)^T=\begin{bmatrix}m_{11}&m_{12}&m_{13}&m_{14}\\m_{21}&m_{22}&m_{23}&m_{24}\\m_{31}&m_{32}&m_{33}&m_{34}\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix} \tag{6}$$

其中，M_1为 Cam1 的投影矩阵。通过式（6）可计算出P_1点在世界坐标系中的准确位置(x,y,z)：

$$\begin{cases} x = \dfrac{m_{12}m_{24} - m_{14}m_{22} + m_{22}u - m_{12}y + m_{12}m_{23}z - m_{13}m_{22}z}{m_{11}m_{23} - m_{12}m_{21}} \\ x = \dfrac{-(m_{12}m_{24} - m_{14}m_{21} + m_{21}u - m_{11}y + m_{11}m_{23}z - m_{13}m_{21}z)}{m_{11}m_{23} - m_{12}m_{21}} \\ z = z \end{cases} \tag{7}$$

设 M_2 为 Cam2 的投影矩阵，(u', v') 为点 P_w 在 Cam2 中的投影点，有 $s_2(u_2, v_2, 1)^T = M_2(x, y, z, 1)^T$，解方程得：

$$\begin{cases} s_2 = M_{23}[x, \quad y, \quad z, \quad 1]^T \\ u_2 = M_{21}[x, \quad y, \quad z, \quad 1]^T / s_2 \\ v_2 = M_{22}[x, \quad y, \quad z, \quad 1]^T / s_2 \end{cases} \tag{8}$$

其中，M_{21}、M_{22}、M_{23}为组成投影矩阵M_2的行向量。

至此，通过世界坐标系已将两个视角图像的像素点(u_1, v_1)和(u_2, v_2)对应起来，可以通过式（7）和式（8）将某一视角中所有像素映射至新的视角。

3.5.3 多视角图像显著性目标检测

通过以上的视觉投影，获得了不同视角之间的空间映射关系，本节利用不同视角中前、背景空间关系进行多视角图像的显著性目标检测。本文独立计算各视角的显著性图，然后投影显著性图，对多视角的显著性图加权融合，并利用多视角中物体的空间遮挡关系消除背景干扰。这里需要借助带限图像非均匀采样恢复原理填补图像中的裂纹区域。带限图像可定义为如下函数：

$$f(x, y) = \int_{-\Omega}^{\Omega} \int_{-\Omega}^{\Omega} F(\zeta, \eta) e^{2\pi j(x\xi + y\eta)} d\xi d\eta \tag{9}$$

其中，Ω为图像信号范围，F 为傅里叶变换

给定采样集 $X = \{(x_i, y_i), i \in I\}$，如果采样点是均匀的，根据香农采样原理，$f(x, y)$可重建为：

$$f(x, y) = \sum_{k \in Z} \sum_{l \in Z} f\left(\frac{k}{2\Omega}, \frac{l}{2\Omega}\right) \frac{\sin 2\pi\Omega\left(x - \frac{k}{2\Omega}\right) \sin 2\pi\Omega\left(y - \frac{l}{2\Omega}\right)}{\pi^2 \left(x - \frac{k}{2\Omega}\right)\left(y - \frac{l}{2\Omega}\right)} \tag{10}$$

通过上文显著性计算和投影，得到了中间视角的显著性图以及两侧视角投影到中间视角的显著性图。下面将完成两种显著性图的融合以及投影空洞区域的背

景消除，得到边缘清晰的显著目标。

在投影显著性图融合时，这些不同视角下被遮挡的区域可以视为是图像的背景，即投影中产生的大的不闭合区域，即上文所述的空洞，两侧的投影显著性图都会产生空洞，将所有的空洞区域标记得到空洞掩膜图 $Mask_h$，将空洞在中间视角显著性图 S_m 和左、右两侧视角投影显著性图 S_l、S_r 中的对应区域都置为零，如式（11），完成对目标周围空洞区域背景的消除。

$$\begin{cases} S_m = S_m * Mask_h \\ S_l = S_l * Mask_h \\ S_r = S_r * Mask_h \end{cases} \quad (11)$$

本文根据视角间最优权重分配算法将中间视角显著性图和左、右两侧视角投影显著性图相加进行融合，最优权重分配表述为：

$$\begin{cases} r_i = \dfrac{1}{\sqrt{2\pi}\sigma} e^{\frac{(a\Delta_i - u)^2}{2\sigma^2}} + g_0 \\ \omega_i = \dfrac{r_1}{\sum r_1} \end{cases} \quad (12)$$

其中 $i \in \{l, m, r\}$，表示不同的视角，r_i 为视角相关系数，ω_i 为视角权重，Δ_i 表示该视角与中间视角的距离，左、中、右视角与中间视角的距离分别为-1、0、1，u 和 σ 分别为高斯拟合函数的均值和方差，α 和 g_0 分别为尺度因子和平移参数，最终多视角融合的显著性图定义为：

$$S = \omega_L S_l + \omega_M S_m + \omega_R S_r \quad (13)$$

在 Ballet 多视角序列中，可以计算得视角的权重，从而实现多视角图像显著性检测。

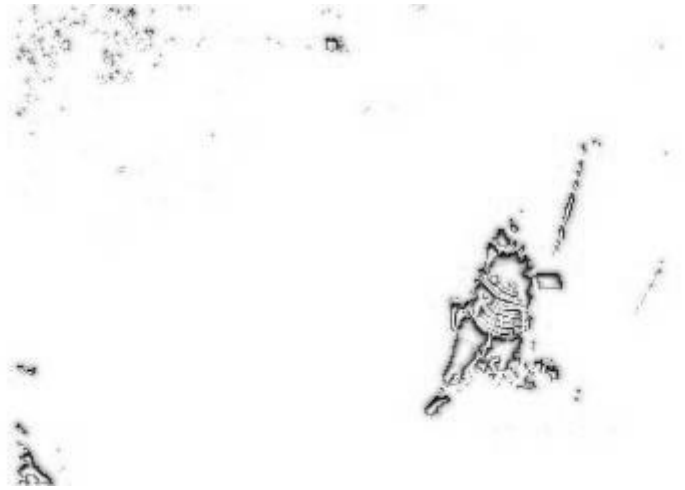
3.5.4 综合检测结果

本题所选视频来自网址 https://m.v.qq.com/play.html?&vid=z03232819j7&ptag=v.qq.com%23v.play.adaptor%232&mreferrer=https%3A%2F%2Fv.qq.com%2Fx%2Fpage%2Fz03232819j7.html。该视频为多角度拍摄的女孩拿取快递的视频。本视频共从三个角度拍摄，第一段视频是从远角俯视拍摄，第二段视频是从门框外俯视拍摄，第三段视频是门框内俯视拍摄。从处理结果可以看出，三段视频经过投影显著性图融合处理后，视频中的女孩被显著地凸显出来。

第一段视频（原视频第 438 帧）

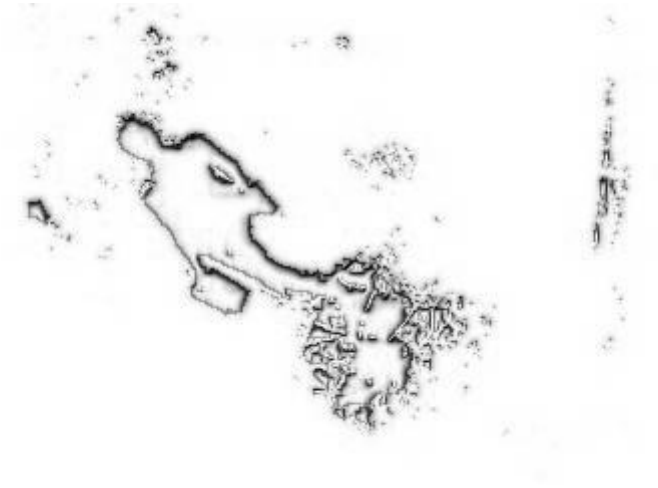
第一段视频（投影显著性图融合后视频第 438 帧）

第三段视频（原视频第 138 帧）



第二段视频（投影显著性图融合后视频第 138 帧）

第三段视频（原视频第 193 帧）



第三段视频（投影显著性图融合后视频第 193 帧）

3.6 问题六的建模

3.6.1 特征选择

角点作为图像上的特征点，在保留图像图形重要特征的同时，可以有效地减

少信息的数据量，有效地提高了计算的速度，有利于图像的可靠匹配，使得实时处理成为可能。其在三维场景重建、运动估计、目标跟踪、目标识别、图像配准与匹配等计算机视觉领域起着非常重要的作用。在本文中，选择 Harris 角点作为图像的特征点进行计算，用特征点的运动状态估计人群整体的运动状态。

3.6.2 光流跟踪

由于需要根据检测出的角点的运动估计行人的运动状态，而背景中固定的角点或者由于环境的干扰（风、光照等的影响）造成检测出的角点产生微小的移动，并不能够代表场景中行人的运动状态，因而需要从检测到的角点集合中删除背景中的固定角点和移动微小的角点，提取出运动的角点作为人群整体运动的特征。

删除非运动角点的方法是，首先获得含有待检测人群的连续 2 帧图像，即第 n 帧和第 $(n+1)$ 帧，对第 n 帧图像进行 Harris 角点检测，将检测到的角点位置集合作为光流跟踪的初始点集合，记为 P_1；然后利用金字塔 $L-K$ 光流算法，对运动进行跟踪，在第 $(n+1)$ 帧得到跟踪成功的角点位置集合，记为 P_2。对于 P_1 中的每一个角点，判断其在 P_2 中是否同样存在，若存在则在 P_2 中删除该角点，反之则将该角点保留在集合 P_2 中，并对运动的角点进行标记。

经过上述运动角点判断处理后，就删除了大部分无用的角点信息，保留了运动的角点。

3.6.3 角点运动速度估计

对于给定的两帧连续图像 $I_n(x,y)$、$I_{n+1}(x,y)$，第 n 帧图像 $I_n(x,y)$ 上某一角点 $u=(u_x,u_y)$，在第 $n+1$ 帧图像 $I_{n+1}(x,y)$ 中找到具有相似图像强度的一点 $v=u+d=[u+d_x,u+d_y]^T$，其中向量 $d=[d_x,d_y]^T$ 为角点 u 的速度向量，即 u 点的光流。

角点运动速度 v 为

$$v=\sqrt{d_x^2+d_y^2}\times fps$$

式中，fps 代表视频帧率。

3.6.4 人群平均运动速度

在人群运动的连续过程中，行人的运动速度会围绕速度的均值发生一定程度的波动，正常情况下，这种波动在一个较小的范围内，当发生了人群的奔跑后，人群的运动速度加快，超过了速度均值正常的波动范围，由此可以检测出人群的奔跑行为。

在进行人群奔跑行为检测时，用特征点的平均运动速度来估计人群整体的运动速度，当人群整体平均运动速度超过预先设置的阈值时，认为发生了人群的奔跑行为。

每一帧图像人群运动的平均速度V_{avg}定义为

$$V_{avg} = \frac{1}{k} \times \sum_{i=1}^{k} v_i$$

式中，k表示检测到运动角点的数量。

3.6.5 奔跑阈值的设置

在设计中，通过对一定连续时间内场景中行人运动平均速度的分析估计奔跑的阈值T_2。，在同一固定场景下，人群中行人的正常平均运动速度曲线存在较小范围的波动，几乎成一条直线。在实验中，将人群的奔跑阈值T_2设置为 40，当人群运动的平均速度V_{avg}持续大于 40 时，则可以认为发生了人群的奔跑行为。

为了避免由于人群运动速度正常波动造成的奔跑误报警，需要对V_{avg}的值进行峰值滤波处理。如果当前帧的V_{avg}值为相邻连续 5 帧中的最小值或者最大值时，取这连续 5 帧的均值作为新的V_{avg}的值。这样可以减小波动，同时不丢失原来的峰值特性。当V_{avg}的值在某一帧突然发生较大的变化，认为其是噪声，系统不做出反应。只有当连续 5 帧V_{avg}的值大于其阈值时，认为发生了人群的奔跑行为，系统发出报警信号。

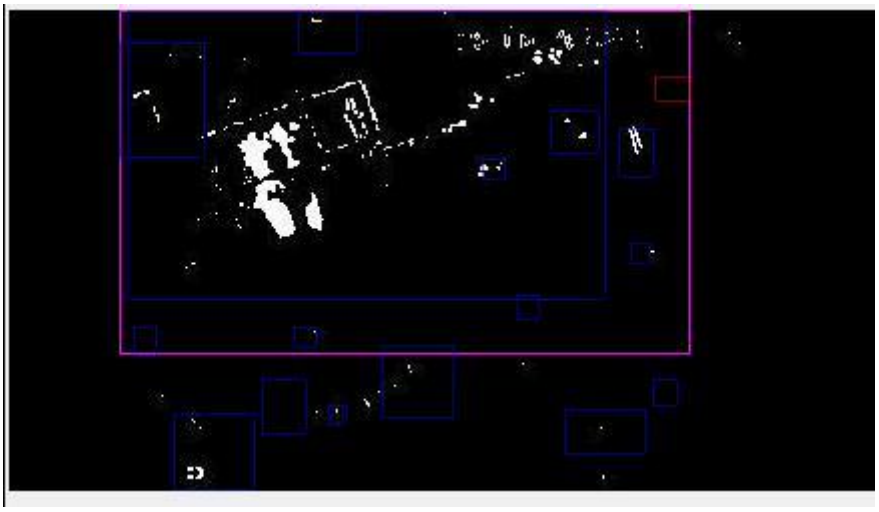
3.6.6 综合检测结果

本段所选视频来自网址 http://my.tv.sohu.com/us/270065766/84303476.shtml。该视频为土耳其伊斯坦布尔机场监控在发生爆炸袭击是拍摄的。因本程序计算量较大，截止到论文提交时仍未运行完成，我们从已运行完毕的部分选取了相近的三张帧图片，并分别

显示了原视频，原视频的处理结果和运用高斯混合模型提取目标后的结果显示出来。待程序运行完毕后，我们会将人群奔跑的识别结果在附件中给出。

原视频（第 36 帧）

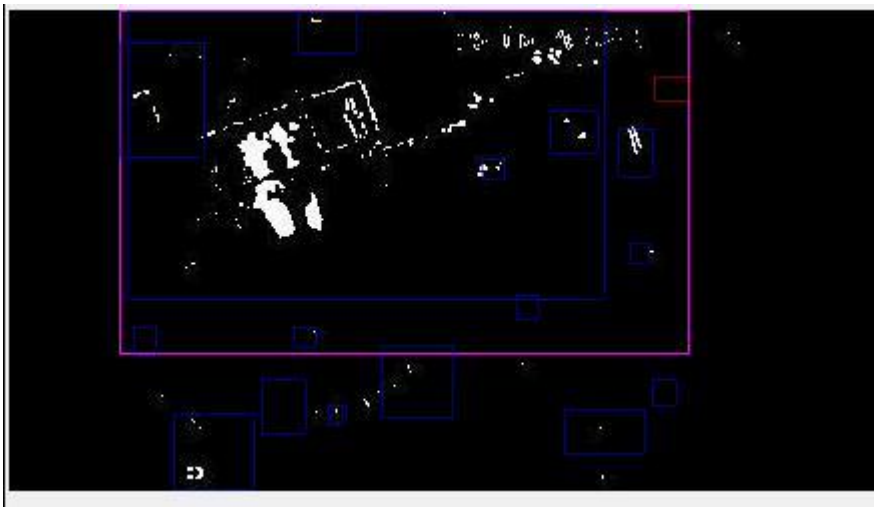
原视频的处理结果（第 36 帧）



运用高斯混合模型提取目标后的结果（第 36 帧）

原视频（第 43 帧）

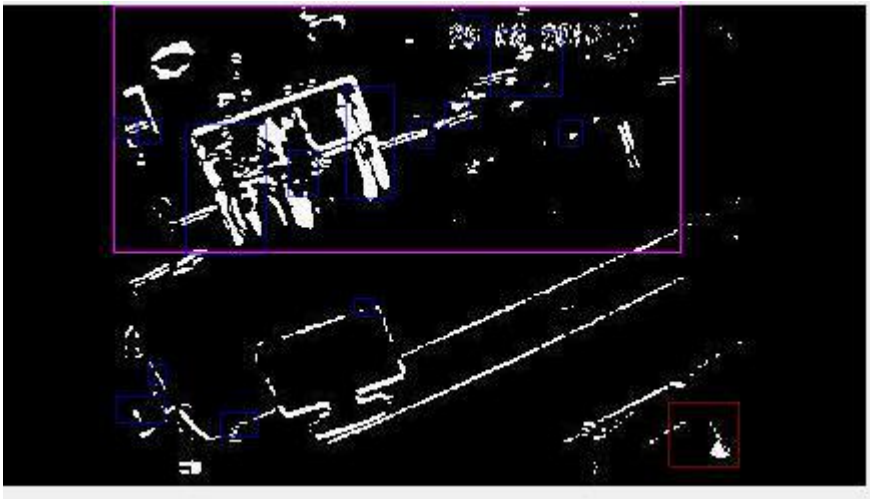
原视频的处理结果（第 43 帧）



运用高斯混合模型提取目标后的结果（第 43 帧）

原视频（第 52 帧）

原视频的处理结果（第 52 帧）



运用高斯混合模型提取目标后的结果（第 52 帧）

从原视频，第 36 帧、第 43 帧和第 52 帧图像的处理结果可以看出，人群处于缓慢移动状态，我们运用蓝色方框中为识别出的缓慢移动的目标，粉色较大方框中为目标较多的区域。从运用高斯混合模型提取目标后视频，第 36 帧、第 43 帧和第 52 帧图像的处理结果可以看出，蓝色方框中为识别出的聚集的人群，但

视频中白色的噪点也会被误认为目标；粉色较大方框中为识别出的人群集中的区域；红色小方框为原视频自带水印造成的，可以不给予关注。以上图像为人群缓慢移动状态下的处理结果，因程序计算量较大，人群奔跑状态下的处理结果将会在附件给予说明。

4 总结与展望

基于场景变换检测的视频段分割是值得进一步进行研究的，对视频数据场景分割的好坏直接影响着视频摘要效果的好与坏。跨视角超空间的构造是多视角视频摘要技术的重点研究内容，后续工作中应对跨视角超空间的构造作出进一步的研究，可以对子空间映射产生的超空间中的元素做进一步的处理，去除一些无用或冗余的元素，防止其干扰聚类结果。

在运动信息提取方面，本文使用了光流特征，但是当整个场景中行人数量多，存在严重遮挡时，光流法得到的光流特征就不足以提供有效的运动信息。之后的算法中，可以加入行人跟踪技术与多摄像头协同技术，对判断异常行为的行人目标记录其个人特征与运动轨迹。

本论文所使用的程序、网站下载的视频，以及视频处理结果将在附件中给出。

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. Cand`es, 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. Babaee, D. T. Dinh, and G. Rigoll, “A deep convolutional neural network for background subtraction,” arXiv preprint arXiv: 1702.01731, 2017.

[9]周同雪, 朱明. 视频图像中的运动目标检测[J]. 液晶与显示, 2017, 32(01):40-47.

[10] 潘志安, 朱三元. 移动摄像视频的多运动目标实时跟踪算法[J]. 控制工程, 2017, 24(04):836-843.

[11] 桑海峰, 陈禹, 何大阔. 基于整体特征的人群聚集和奔跑行为检测[J]. 光电子·激光, 2016, 27(01):52-60.

[12] 刘冬, 赵凯旋, 何东健. 基于混合高斯模型的移动奶牛目标实时检测[J]. 农业机械学报, 2016, 47(05):288-294.

[13] 兰红, 周伟, 齐彦丽. 动态背景下的稀疏光流目标提取与跟踪[J]. 中国图象图形学报, 2016, 21(06):771-780.

[14] 陈成, 庄越挺, 肖俊. 相机运动条件下的视频前景提取[J]. 浙江大学学报(工学版), 2009, 43(06):973-977+982. .

[15]汤一平, 胡大卫, 蔡盈梅, 黄珂, 姜荣剑. 基于全景视觉的移动机器人的运动

目标检测[J]. 计算机科学,2015,42(11):314-319.
[16] 陈杰,邓敏,肖鹏峰,杨敏华,梅小明,刘慧敏. 基于分水岭变换与空间聚类的高分辨率遥感影像面向对象分类[J]. 遥感技术与应
用,2010,25(05):597-603.