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

学　　校	同济大学
参赛队号	**10247039**
队员姓名	**1. 张默涵**
	2. 霍文君
	3. 单光旭

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：

本文给出了在不同场景下的监控视频中前景目标提取问题的解决方案，提供了模型在附件视频中前景目标的提取结果，以及每个视频中情景目标出现的视频帧数，本文的最后对不同模型的优劣及使用场景进行了分析。

针对问题一和问题二，本文建立了基于二帧差法的前景目标检测算法和基于高斯混合模型的背景减除法两个模型，并利用所建立模型对附件中的八组视频进行了前景目标的识别提取。

针对问题三，本文建立了抖动场景下的前景目标检测算法模型，该模型利用非参数方法建立背景模型，并通过比较图像与背景模型的差异来判定前景目标。

针对问题五，本文建立了基于全景图像拼接的多角度视频前景目标检测模型，该模型首先将不同角度同一时刻拍摄的视频帧利用基于 SIFT 特征的全景图像拼接算法将其拼接成一个全景图片，之后再进行前景目标的识别和提取。

针对问题六，本文建立了基于长短期记忆神经网络（LSTM）的视频异常事件检测模型，该模型利用 LSTM 将异常事件检测转换为一个序列标注的问题，从而可以有效地检测出视频中异常事件地发生。

关键词：前景目标检测 帧差法 高斯混合模型 背景减除法 全景图

1.问题重述

随着智能化技术的飞速发展以及平安城市等概念的提出和落实，智能化的安防系统成为建设平安城市的趋势和关键。在当今时代，经济增长和治安需要等因素导致城市的各类场所的视频监控数量呈现指数型的增长，大数量的摄像头也为城市的安防人员提供了海量的视频信息，而当对这些视频信息进行分析处理时，如果采用传统的人工方法，则需要大量的人力物力，因此智能化的处理监控视频方法成为最为迫切的需要，其中在监控视频对前景目标的定位和提取成为智能化视频处理的基础。

监控视频中前景目标的提取算法具有广泛的应用场景，比方说，在银行大厅中对客户的面部提取或行为分析可以对不良行为进行警报；另外，警察在破案侦查过程中，可以只对存在犯罪嫌疑人的监控视频中识别分析。

因此，能够针对不同的应用场景给出恰当的目标提取算法十分必要，本文的目标就是针对现实需求，利用附录视频信息，设计算法来解决以下问题：

⑴ 在摄像头稳定拍摄且为静态背景下的监控视频中提取显著前景目标。

⑵ 在摄像头稳定拍摄但包含动态背景信息（如树叶摇动、水波动、喷泉、窗帘晃动）的监控视频中提取显著前景目标。

⑶ 在摄像头具有抖动或偏移情况得到的监控视频中提取前景目标。

⑷ 从不同角度同时拍摄的同一地点的多个监控视频中提取前景目标。

⑸ 利用获取的前景目标信息，提供自动判断监控视频中是否具有异常信息的算法方案，并针对更多的异常事件信息给出相应的事件检测方案。

2.问题分析

前景目标提取作为智能化监控视频分析的基础，它的主要目的是将监控视频中运动的、人们所关心的物体从背景中分离出来，并进行接下来的分析。前景目标提取算法设计的主要难点有：⑴如何确定监控视频中的背景信息，尤其是当视频场景中存在因天气、光照、阴影或其他杂乱干扰时有效分离背景与前景；⑵当视频存在抖动或偏移情况时如何准确确定前景目标。

在目标提取中常用的算法主要有帧间差分法和背景减除法，其中帧间差分法的主要思想是在得到的视频帧序列相邻两帧或三帧间采用基于灰度值的时间差分，并设定判断阈值，当差分值大于阈值时判断为前景 ，差分值小于阈值时判断为背景。帧间差分法具有实时性好、算法复杂度低、敏感性强等优点，但是由于运动的目标灰度值较为相近，因此可能无法提取完整的目标信息，同时该算法受环境噪声影响较大，因此更加试用于静态背景下的目标提取。背景减除法的主要思想是从得到的视频帧中确定背景图像，将当前帧与背景图像进行基于灰度值的差分，并与设定阈值进行比较，其中区别较大的区域被认定为前景目标区域，区别较小的区域被认定为背景区域。该算法的关键是准确地确定背景图像，现有的算法有变带宽核密度估计法、高斯混合模型等。背景减除法的主要优点是能够提取出比较完整的背景信息，并且在具有动态背景的视频中也具有良好的识别效果，但是由于该算法需要首先对完整的视频信息进行分析，因此复杂度较高，且

实时性不如帧间差分法。

针对上述问题，本文给出了在不同场景下的前景识别模型：

⑴问题一是对不包含动态背景且摄像头稳定的视屏中提取前景目标，此问题由于背景稳定，因此帧差法和背景减除法均可以较好得实现，难点在于当背景灯光改变或目标物体运动较慢时，模型是否可以较好地去除噪声。

⑵问题二是在包含动态背景的视频中提取前景目标，此类视频的背景一般包括树叶晃动、水波动、喷泉变化、窗帘摇晃等，此问题难点在于，背景的变化幅度和变化频率等因素都会影响模型的识别效果，因此去除一些背景点的噪声影响十分重要。

⑶问题三是在摄像头发生晃动或偏移的场景下提取前景目标，在抖动场景下，一些位于纹理边缘部分的背景点会因为抖动而被识别为前景点，因此难点在于如何能够正确分离抖动的背景点和真正的前景点。

⑷问题五是通过在不同角度同时拍摄的近似同一地点中的多个监控视频中提取前景目标，由于此问题涉及到的视频彼此之间具有相关性，因此如何能提取这种相关性并且在不同视频之间识别到同一个前景目标十分重要。

⑸问题六是在识别前景目标的基础上，通过对前景目标信息的分析来判定视频中是否出现人群短时聚集、群体规律性变化、物体爆炸等异常事件，由于不同的异常事件会在视频中有不同的体现，比方说群体规律性变化会产生周期性特征，人群短时聚集会产生线性变化特征等等，因此该问题难点在于如何能通过对前景目标信息的分析来判定是否出现异常事件。

3.模型假设及符号说明

3.1 模型假设

⑴题目中的视频只考虑黑白两种色彩，模型中对视频帧的基本操作的是在附件 1 的程序下进行的；

⑵同一视频下不同帧的长宽比是相同；

⑶问题 1 中前景目标提取的视频中不包含动态背景，摄像头稳定拍摄，即背景结构非常稳定，因此不需考虑背景的动态变化对前景提取的影响；

⑷问题 2 中的背景结构存在着小幅度的波动，背景结构不稳定；

⑸问题 3 摄像头发生晃动和偏移，该类视频近似视为一种线性仿射变换；

⑹问题 4 中的视频帧数是通过题目附件 1 的程序下截取并编号的；

⑺问题 5 中所研究视频的拍摄地点和时间一致，但是拍摄视频的角度不同；

3.2 符号说明

编号	符号定义	符号说明
1	$X_{t,w,h}$	表示某一视频帧（t 为时间，w 为视频宽度，h 为视频高度）

2	$X_{t.g}$	表示t时刻的灰度值（根据假设中同一视频中的w、h相同这里略去w、h，记为X_t）
3	$target_{t,w,h}$	表示含有目标的视频帧（t为时间，w为视频宽度，h为视频高度）
4	$X.length$	表示视频x的总帧数
5	Δ_t	表示当前时刻帧的灰度值与前一时刻的灰度值之差（$0 < t \le X.length$）
6	θ	表示灰度值差值绝对值的阈值
7	a_k	表示每个高斯分布在混合模型中所占的权重 $\left(a_k \ge 0, \sum_{k=1}^{K} a_k = 1\right)$
8	θ_k	表示高斯密度函数的参数
9	$\phi(y \mid \theta_k)$	表示高斯分布的概率密度函数
10	l_i	表示每一像素点i连续出现1和0的次数（$l_i \in [-N, N]$）
11	$H(l)$	表示每一像素点l_i的概率密度函数
12	h_t	表示t时刻LSTM神经元的输出

13	C_t	表示 t 时刻 LSTM 神经元的细胞状态
14	f_t	表示 t 时刻 LSTM 神经元的遗忘门的取值
15	i_t	表示 t 时刻 LSTM 神经元的 *sigmod* 层的取值
16	$\tilde{C}_t$	表示 t 时刻 LSTM 神经元的 *tanh* 层计算出的新候选值向量
17	o_t	表示 t 时刻 LSTM 神经元的输出层的输出比率
18	h_θ	表示 *softmax* 神经元的输出

4.模型建立与求解

4.1 模型建立

4.1.1 模型一：基于二帧差法的前景目标检测算法

一般情况下，视频图像中的运动前景区域像素值相比于上一帧差别较大，利用这一性质，可以利用本帧图像与上一帧图像的像素差值来提取目标前景区域。较为简单的做法是事先认为设定一个阈值，对某一像素点，计算其与上一帧该像素点的灰度差，若大于阈值则认为是前景，否则认为是背景。具体的实现步骤如下：

⑴计算当前帧与前一帧的图像像素点灰度值差并取绝对值；

⑵如果差值大于阈值，可以认为该像素点为前景区域并标记；

⑶保存并记录含有前景目标的帧。

算法流程图如图 4.1.1

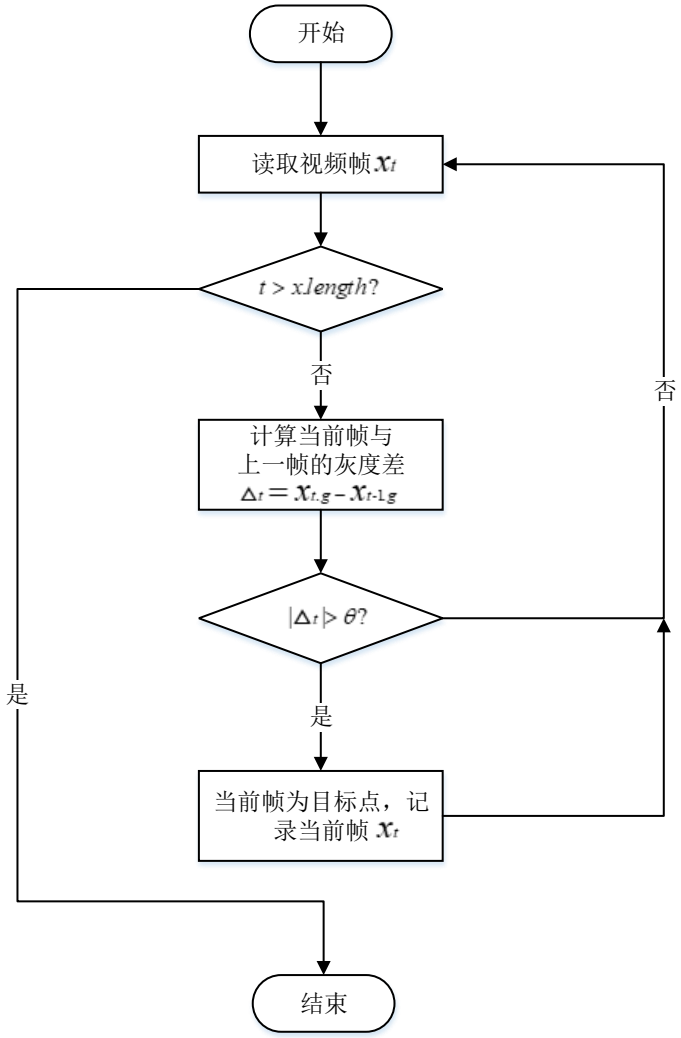

图 4-1-1 基于二帧差法的前景目标检测算法流程图

基于二帧差法的前景目标检测算法描述如表 4-1-1 所示。

表 4-1-1 基于二帧差法的前景目标检测算法描述

基于二帧差法的视频目标检测算法	
输入	视频 *x*
输出	目标编号及视频帧 *target*
01	**for**(int *t*=1; *t*<=*x.length*;*t*++)
02	$\Delta_t = X_{t.g} - X_{t-1.g}$
03	**if**($|\Delta_t| > \theta$)
04	$target_t = x_t$
05	**return** *target*

4.1.2 模型二：基于高斯混合模型的背景减除法

背景减除法的主要思想是首先通过视频确定出背景模型，再将视频中的每一帧与背景逐像素比较，差异较大的像素点为前景区域，差异较小的像素点为背景区域。

在背景减除法中，背景模型的确定是至关重要的一环，此算法适合于在摄像头固定的情况下进行背景建模，当背景具有复杂的动态效果，比如树叶晃动、水

面波动、光线反射等时，基于像素的高斯混合模型由于具有多峰分布的特点，也依然可以适应动态背景的变化，准确提取前景目标。

高斯混合模型是用多个单一高斯模型的混合来拟合某一事件的概率分布，它的概率密度函数如公式 4.1 所示：

$$P(y)=\sum_{k=1}^{K}a_k\phi(y\mid\theta_k) \tag{4.1}$$

其中 a_k 是权重系数，代表每个高斯分布在混合模型中所占的权重，并且满足 $a_k\geq 0,\sum_{k=1}^{K}a_k=1$ ； $\phi(y\mid\theta_k)$ 表示高斯分布的概率密度函数，其中 $\theta_k=\left(\mu_k,\sigma_k^2\right)$ 表示高斯密度函数的参数，μ_k 为均值，σ_k^2 为方差，密度函数表示如公式 4.2 所示：

$$\phi\left(y\mid\theta_k\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(y-\mu_k)^2}{2\sigma_k^2}} \tag{4.2}$$

上述模型中的主要参数的拟合常用 EM(Expectation Maximization)算法

基于混合高斯模型的背景减除法有训练背景模型和分离前景目标两个步骤，在训练背景模型中，对图像中的每一个像素点的灰度值分别用 K 个高斯分布构成的混合高斯模型来建模，得到每一个像素点灰度值的混合高斯模型；在前景目标分离中，对图像中的每一个像素点与上一步得到的每一个高斯模型进行匹配，如果差别不大的话则认为是背景，否则认为是前景。整体的算法框架如下：

⑴设定训练的图像帧并进行背景的高斯混合模型的建立；

⑵对视频中的每一帧与背景模型进行逐像素点的比较并确定每一帧的前景目标区域；

⑶保存并记录含有前景目标的帧。

具体的算法流程如图 4-1-2：

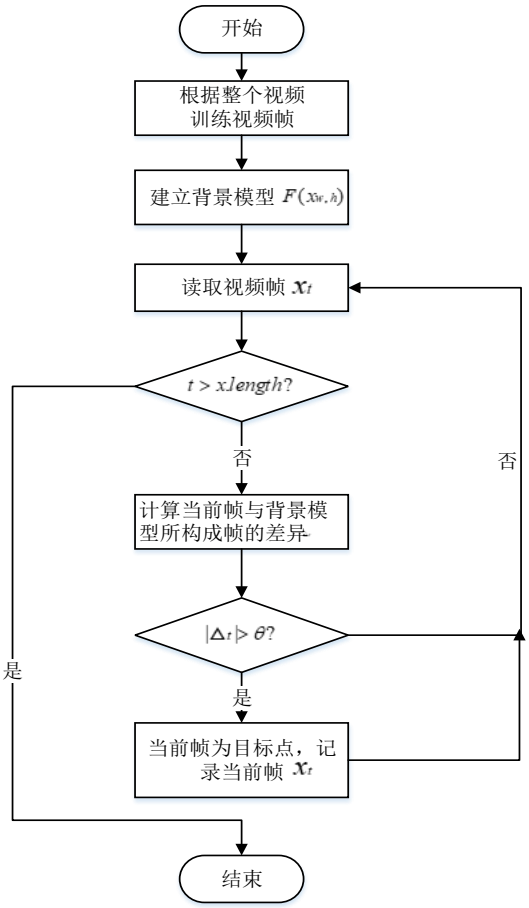

图 4-1-2 基于高斯混合模型的背景减除法流程图

基于高斯混合模型的背景减除法算法描述如表 4-1-2：

表 4-1-2 基于高斯混合模型的背景减除法

基于高斯混合模型的背景减除法的视频目标检测算法	
输入	视频 *x*
输出	目标编号及视频帧 *target*
01	int n = *x.length*
02	**while**(n--!=0)
03	$F(x_{w,h})$=*practice(x.n)*
04	**for**(int *t*=1; *t*<=*x.length*;*t*++)
05	$\Delta_{i,j} = (\mu - \sigma < \lvert x_{t,i,j} - F(x)_{i,j} \rvert < u + \sigma)?1:0$
06	$\Delta_t = \sum_{i,j}^{w,h} \Delta_{i,j}$
07	**if**($\lvert\Delta_t\rvert > \theta$)
08	$target_t = x_t$
09	**return** *target*

4.1.3 模型三：抖动场景下的前景目标检测算法

在抖动场景下，背景的某些边缘像素部分会因为抖动而被认为是前景目标像素，为了能在抖动场景下依然能够提取前景目标，需要恰当地建立背景模型。对

一些抖动场景下的视频进行分析发现：背景点会因为相机抖动而产生运动信息，而且背景像素点会因为其所在位置和相机的抖动频率而产生不同的运动信息，但是同一背景像素点因抖动而产生的运动信息是确定的，这种运动信息可以使用无参数方法中的核密度估计进行量化。对于前景目标，它的运动信息相对于背景像素点的运动信息是不确定的，利用这种差异性可以有效分离前景像素点和背景像素点。

本文利用背景与前景运动信息的差异性建立了一种在抖动场景下的前景目标检测算法，算法的整体框架如图 4-1-3 所示，具体的算法步骤如下：

⑴利用模型二中的背景减除法提取候选前景，此步提取的候选前景包含因相机抖动而产生运动信息的背景点，其中每一帧的候选前景像素标记为 1，背景像素标记为 0，生成二值图像；

⑵连续采集 N 个二值图像，统计每一像素点连续出现 1 和 0 的次数，记为 l_i，其中 $l_i \in [-N, N]$，符号为负表示连续出现 0，符号为正表示连续出现 1，利用非参数方法中的核密度估计法拟合每一像素点 l_i 的概率密度函数 $H(l)$ 作为该像素点的运动信息模型；

⑶对于待判定前景目标的二值图，假定该帧只与它的前 W 帧有关，其中 $W \ll N$，用同样的方法计算这 W 帧中每一像素点 l_i' 的概率密度函数 $H(l)'$，比较 $H(l)'$ 与 $H(l)$ 的相似性，该相似性可以用 $H(l)'$ 隶属于 $H(l)$ 的概率 P 表示，并将 P 与阈值 θ 进行比较，若 $P < \theta$，说明差异性较大，则该点为前景点，否则为背景点。

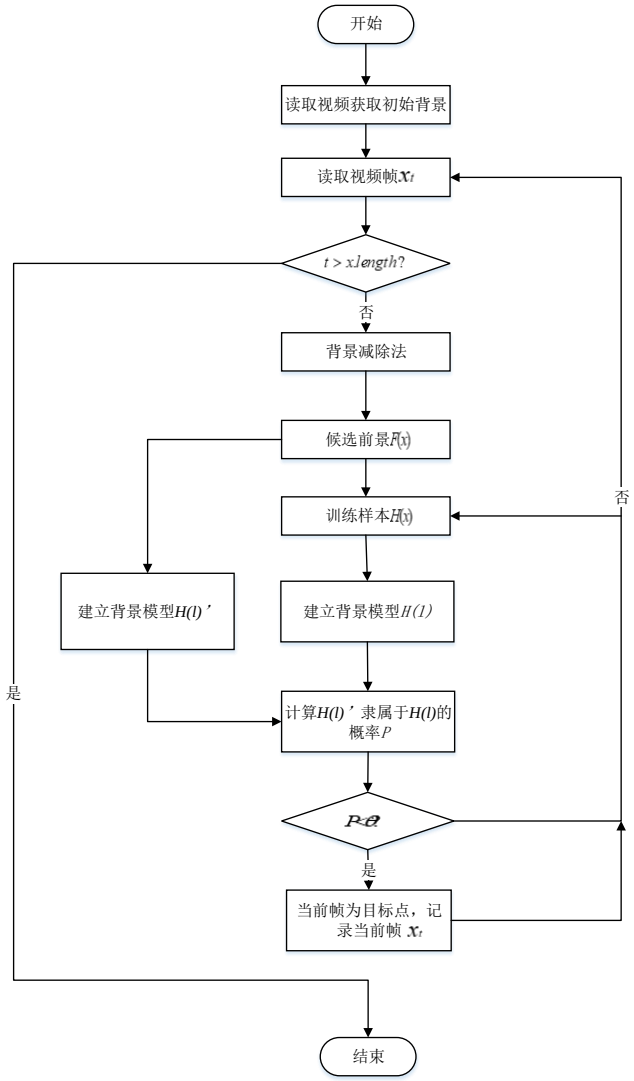

图 4-1-3 抖动场景下的前景目标检测算法流程图

4.1.4 模型四：基于全景图像拼接的多角度视频前景目标检测模型

针对问题 5 提出的如何更好的利用从不同角度同时拍摄的近似同一地点的多个监控视频更加有效检测和提取视频前景目标的问题，解决思路是将多个视频的同一时刻的视频帧，拼接成一个全景图像，之后再利用模型二，进行前景目标的检测和提取，算法框架如图 4-1-4 所示。

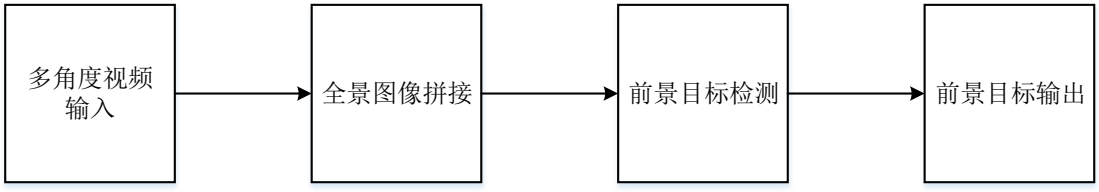

图 4-1-4 基于全景图像拼接的多角度视频前景目标检测模型流程图

因此解决该问题的关键是全景图像的拼接，针对此我们提出了基于 SIFT 特征的全景图像拼接算法，算法的流程如图 4-1-5 所示。

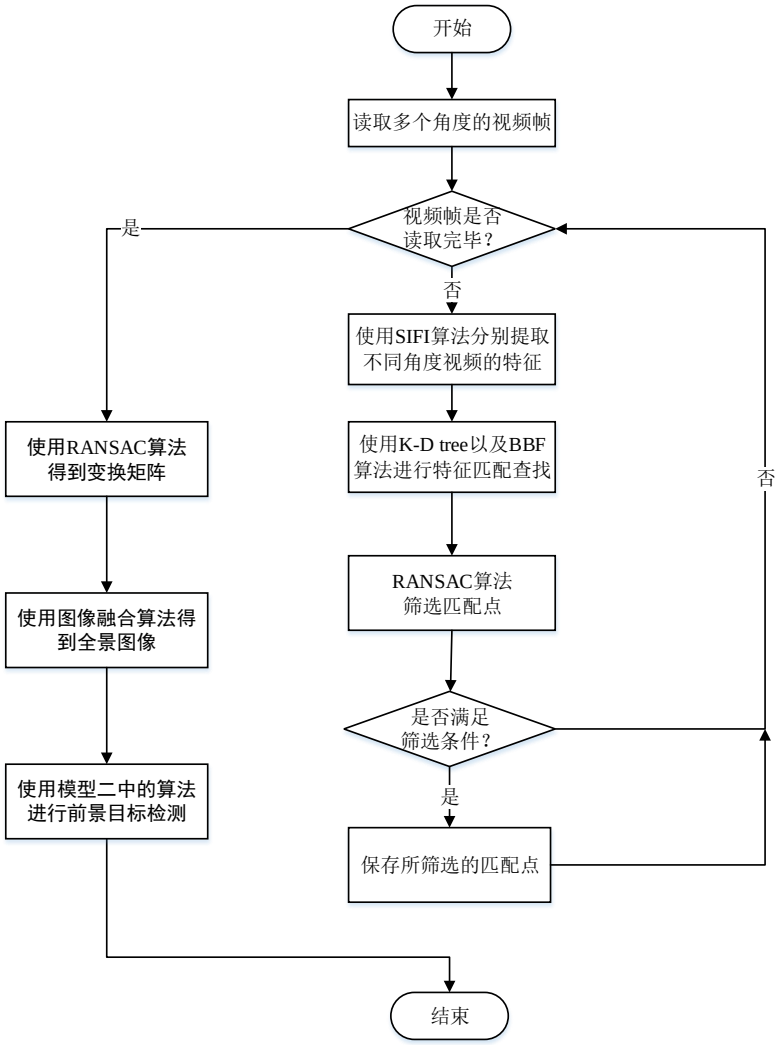

图 4-1-5 基于 SIFT 特征的全景图像拼接算法流程图

算法首先使用 SIFT 算法对多个角度同一时刻的视频帧分别进行特征提取。尺度不变特征转换(Scale-invariant feature transform，SIFT)算法是一种用来侦测与描述影像中的局部性特征，它在空间尺度中寻找极值点，并提取出其位置、尺度、旋转不变量的算法。其主要分为下面 4 步：

⑴尺度空间极值检测：搜索所有尺度上的图像位置。通过高斯微分函数来识别潜在的对于尺度和旋转不变的兴趣点。

⑵关键点定位：在每个候选的位置上，通过一个拟合精细的模型来确定位置和尺度。关键点的选择依据于它们的稳定程度。

⑶方向确定：基于图像局部的梯度方向，分配给每个关键点位置一个或多个方向。所有后面的对图像数据的操作都相对于关键点的方向、尺度和位置进行变换，从而提供对于这些变换的不变性。

⑷关键点描述：在每个关键点周围的邻域内，在选定的尺度上测量图像局部的梯度。这些梯度被变换成一种表示，这种表示允许比较大的局部形状的变形和光照变化。

然后使用 K-D tree 算法和 BBF 算法进行特征匹配查找，并根据最近邻和次近邻距离比值进行初步筛选。K-D tree 算法本质上是一棵搜索二叉树，它的每一

层将特征空间分成两部分，树的顶层节点按一维进行划分，下一层节点按另一维进行划分，以此类推，循环往复。这里我们使用其构建出一棵 SIFT 算法提取出的特征空间搜索二叉树，之后使用 BBF 算法进行特征搜索和查找。

BBF 算法（近似最邻近算法），其基本思想是借助优先队列实现。从树的根开始，在 K-D tree 上寻路时将路过的节点加入到优先队列里直到叶子节点；然后再从队列里取出目前节点值最小的，重复上述过程。

在对 SIFT 特征进行查找匹配之后，我们使用 RANSAC 算法筛选匹配点并计算变换矩阵。RANSAC 算法是采用迭代的方式从一组包含离群的被观测数据中估算出数学模型的参数，最终会得到一个图像的变换矩阵。

基本算法思想如下：

⑴构建一个最小抽样集的势为 n 的模型(n 表示初始化模型参数所需要的最小样本个数)和一个样本集 P，集合 P 的样本数 P.*count* 满足关系 P.*count* $> n$)；

⑵从集合 P 中随机抽取包含 n 个样本的 P 的子集 S 初始化模型 M；

⑶选出差集 SC=P-S 中与模型 M 的误差小于某一设定阈值 t 的样本集 S，SC 与集合 S 取并集之后构成 S^*。（S^* 可以认为是内点集，它们构成 S 的一致集）；

⑷若满足条件 $\#(S^*) \geq N$，那么可以认为得到的参数是正确的模型参数，然后利用集 S^* (内点)采用最小二乘法重新计算新的模型 M^*；重新随机抽取新的 S，重复过程(1)-(4)，如果不满足条件，那么执行(5)；

⑸在完成上述抽样次数后，还未找到一致集则算法失败，否则选取抽样后得到的最大一致集判断内外点，算法结束。

最后进行图像的融合，融合的方法是选择一个参考帧，然后对其他帧使用 RANSAC 算法得到的变换矩阵进行处理，然后加入到参考帧中，最后得到的新的图像就是全景图像。

在得到所有帧的全景图像之后，再利用模型二进行前景目标检测。

4.1.5 模型五：基于长短期记忆神经网络(LSTM)的视频异常事件检测模型

针对问题 6 提出的判断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的任务，提出了一种基于 LSTM 神经网络的视频异常事件检测模型。

模型主要分为两个部分：视频预处理部分和视频帧标注部分。视频预处理部分主要是利用针对问题 2 提出的模型将视频进行处理，将视频每一帧进行分离出前景目标，作为第二部分的输入。视频帧标注部分，主要是利用 LSTM 神经网络对每一个视频帧根据当前帧的特征以及该帧之前的帧的部分特征综合进行标注，标注的结果是将当前帧分为：正常帧(N-Frame)、异常事件起始帧(B-Frame)、异常事件中间帧(M-Frame)和异常事件结束帧(E-Frame)，四种标签中的一种。

流程图如图 4-1-6 所示：

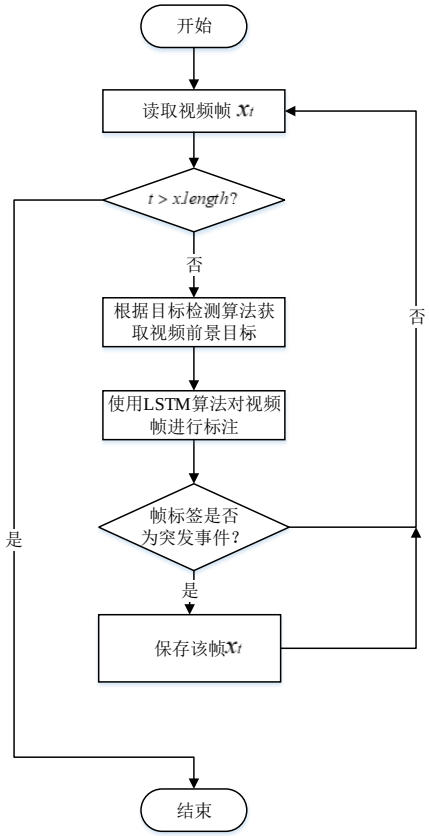

图 4-1-6 基于 LSTM 的视频异常事件检测模型流程图

模型的第一部分在前文已经详细介绍过，这里就不再介绍。下面主要介绍模型的第二部分，基于 LSTM 神经网络的视频帧标注算法，算法的网络结构图如图 4-1-7 所示：

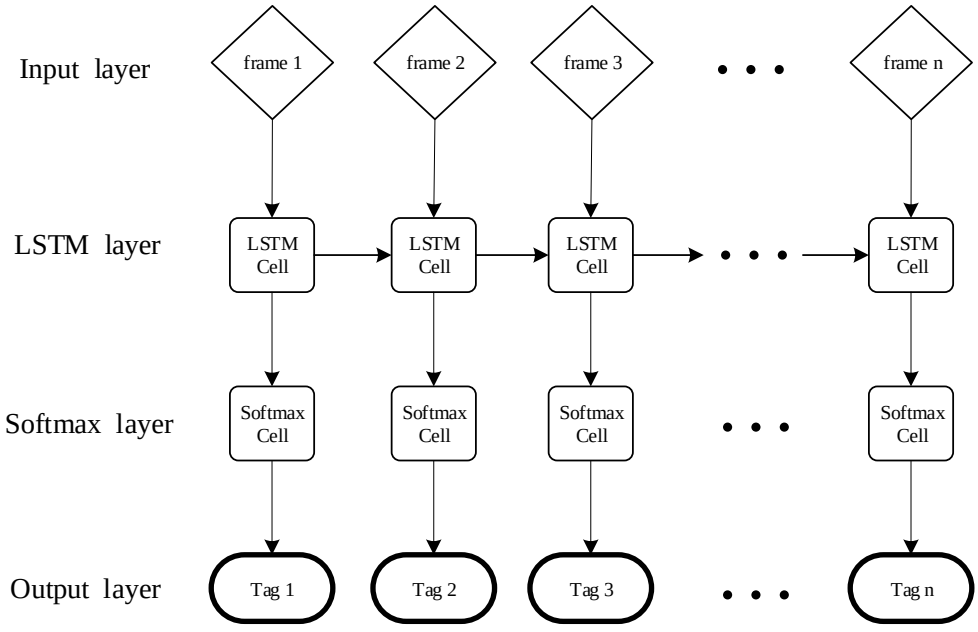

图 4-1-7 基于 LSTM 神经网络的视频帧标注算法

在基于 LSTM 神经网络的视频帧标注算法中的 LSTM 层中的各个神经元之间共享一套参数，在 Softmax 层中的各个神经元之间也共享同一套参数。

其中 LSTM 层的神经元的结构图如图 4-1-8 所示：

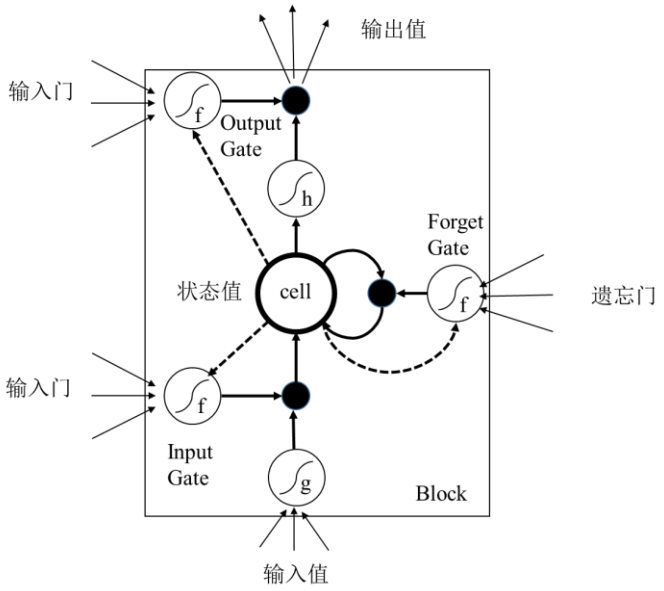

图 4-1-8　LSTM 层的神经元的结构图

LSTM 层神经元的运作方式如下：

⑴决定从细胞状态中丢弃什么信息。这个决定通过一个称为遗忘门层完成。该门会读取 h_{t-1} 和 x_t，输出一个在 0 到 1 之间的数值给每个在细胞状态 C_{t-1} 中的数字。1 表示“完全保留”，0 表示“完全舍弃”。公式如 4.3 所示：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4.3}$$

⑵确定什么样的新信息被存放在细胞状态中。这里包含两个部分。第一，sigmoid 层称 “输入门层” 决定什么值将要更新。然后，一个 tanh 层创建一个新的候选值向量，其会被加入到状态中。下一步，会将这两个信息来产生对状态的更新。公式如 4.4，4.5 所示：

$$i_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_i) \tag{4.4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4.5}$$

⑶更新旧细胞状态，C_{t-1} 更新为 C_t。前面的步骤已经决定了将会做什么，现在就是实际去完成。把旧状态 C_{t-1} 与 f_t 相乘，丢弃掉确定需要丢弃的信息。接着加上 $i_t * \tilde{C}_t$。这就是新的候选值，根据所决定的更新每个状态的程度进行变化。公式如 4.6 所示：

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4.6}$$

⑷确定输出什么值。这个输出将会基于的细胞状态，但是也是一个过滤后的版本。首先运行一个 sigmoid 层来确定细胞状态的哪个部分将输出出去；接着把细胞状态通过 tanh 进行处理（得到一个在 -1 到 1 之间的值）并将它和 sigmoid 门的输出相乘，最终仅仅会输出确定输出的那部分。公式如 4.7，4.8 所示：

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{4.7}$$

$$h_t = o_t * \tanh(C_t) \tag{4.8}$$

softmax 层中的神经元进行的是 softmax 回归，其各个神经元的参数进行共享，其输入为对应时刻的 LSTM 神经元的输出。其分类函数如公式 4.9 下：

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)}=1 \mid x^{(i)};\theta) \\ p(y^{(i)}=2 \mid x^{(i)};\theta) \\ \vdots \\ p(y^{(i)}=k \mid x^{(i)};\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{4.9}$$

最终得到一个 k 维的向量（向量元素的和为 1），选择概率值最大的一维所对应的标签为分类的最终的结果。

这里需要提及的一句是 LSTM 和 softmax 层神经元的参数的求解采用的是反向传播算法（Backpropagation）进行求解。

4.2 模型求解

本文利用模型建立中的模型一和模型二对附录三中的八组视频进行对前景目标的识别，具体如图 4-2-1~4-2-8，其中(a)图为原视频图，(b)图为二帧差法，(c)图为背景减除法。

(a) (b) (c)

图 4-2-1 campus 视频前景目标对比图

(a) (b) (c)

图 4-2-2 curtain 视频前景目标对比图

图 4-2-3 escalator 视频前景目标对比图

图 4-2-4 fountain 视频前景目标对比图

图 4-2-5 hall 视频前景目标对比图

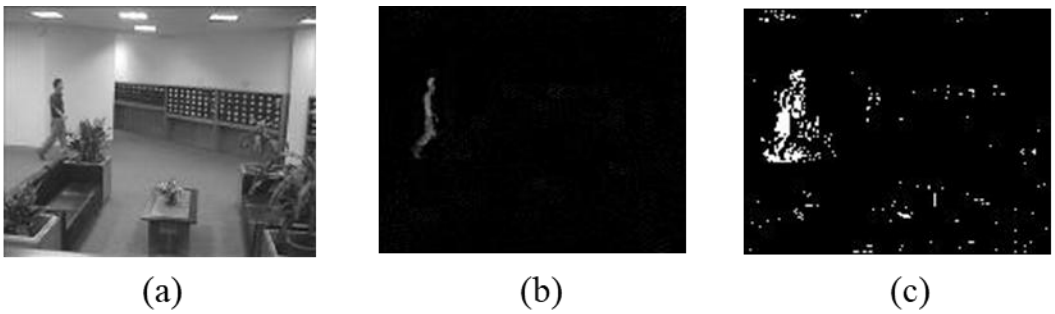

图 4-2-6 lobby 视频前景目标对比图

(a) (b) (c)

图 4-2-7 office 视频前景目标对比图

(a) (b) (c)

图 4-2-8 overpass 视频前景目标对比图

针对每一视频将目标出现时的检测，在本章的结果显示表现较好的一组算法，并将每组视频中出现前景目标的帧数结果展示如表 4-2-1~4-2-8。

表 4-2-1 campus 视频目标识别

问题分类	算法模型	问题求解		
		目标编号	目标出现时的帧编号	目标离开时的帧编号
动态下的目标检测	背景减除法	1	200	225
		2	305	524
		3	645	684
		4	693	713
		5	747	905
		6	1331	1374
		7	1378	1407

表 4-2-2 curtain 视频目标识别

问题分类	算法模型	问题求解		
		目标编号	目标出现时的帧编号	目标离开时的帧编号
动态下的目标检测	背景减除法	1	967	967
		2	1761	1905
		3	2126	2126
		4	2175	2315
		5	2769	2930

表 4-2-3 escalator 视频目标识别

问题	算法	问题求解

分类	模型	目标编号	目标出现时的帧编号	目标离开时的帧编号
动态下的目标检测	背景减除法	1	1	171
		2	220	2399
		3	2539	2540
		4	2574	2574
		5	2777	3417

表 4-2-4 fountain 视频目标识别

问题分类	算法模型	问题求解		
		目标编号	目标出现时的帧编号	目标离开时的帧编号
动态下的目标检测	背景减除法	1	141	141
		2	157	212
		3	259	259
		4	335	335
		5	408	523

表 4-2-5 hall 视频目标识别

问题分类	算法模型	问题求解					
		目标编号	目标出现时的帧编号	目标离开时的帧编号	目标编号	目标出现时的帧编号	目标离开时的帧编号
静态背景下的目标检测	帧差法	1	2	23	14	819	848
		2	28	36	15	894	1048
		3	38	45	16	1155	1214
		4	46	69	17	1280	1550
		5	79	88	18	1558	1760
		6	89	106	19	1990	2010
		7	111	141	20	2092	2159
		8	166	219	21	2383	2445
		9	220	226	22	2519	2565
		10	227	280	23	2812	2865
		11	325	480	24	3031	3306
		12	578	578	25	3460	3534
		13	601	698			

表 4-2-6 lobby 视频目标识别

问题分类	算法模型	问题求解		
		目标编号	目标出现时的帧编号	目标离开时的帧编号
静态背景下的目标检测	帧差法	1	79	79
		2	154	199
		3	259	259
		4	345	394
		5	521	521

		6	622	669
		7	870	870
		8	963	1038
		9	1238	1283
		10	1333	1538

表 4-2-7 office 视频目标识别

问题分类	算法模型	问题求解		
		目标编号	目标出现时的帧编号	目标离开时的帧编号
静态背景下的目标检测	帧差法	1	197	197
		2	372	372
		3	501	501
		4	579	2043
		5	2080	2080

表 4-2-8 overpass 视频目标识别

问题分类	算法模型	问题求解		
		目标编号	目标出现时的帧编号	目标离开时的帧编号
动态背景下的目标检测	背景减除法	1	374	374
		2	471	712
		3	968	968
		4	1551	1551
		5	1881	1881
		6	2098	2098
		7	2333	2968

5. 模型评价及改进

本文针对不同场景下监控视频的前景目标识别问题给出了不同的模型方法，每个模型方法都有利有弊且适用于不同场景的前景目标识别。

⑴二帧差法通过相邻帧逐像素灰度值求差来提取前景目标，该方法实时性好、运算速度快、算法简单，并且不需要对背景更新，因此复杂度也较低，根据算法原理，该方法对快速出现或移动的前景目标具有很好的识别效果，但是该算法只能检测出目标周围的运动信息，因此只能检测出目标的轮廓，出现空洞现象这在模型求解中图 4-2-7 中有所体现。

⑵基于高斯混合模型的背景减除法首先通过高斯混合模型建立背景模型，并通过将视频帧与背景模型进行比对来确定前景目标，该方法通过对背景信息进行建模，能非常有效地识别出前景目标，虽然该算法在动态背景下可能会受到噪声的影响，但是通过对高斯模型个数的改变以及背景信息的实时更新可以使该算法在动态背景下也能很好地适应。

⑶抖动场景下的前景目标识别在模型二的基础上进行了改进和创新，由于在

抖动场景下，某些背景点会产生运动信息而被认为是前景目标，但是背景点的运动信息是确定的，而前景目标的运动信息是不确定的，利用这一特点可以有效地提取前景目标。该方法在进行背景信息建模时使用了非参数方法的核密度估计，这样可以有效提取背景点的运动信息。

⑷基于全景图像拼接的多角度视频前景目标检测模型针对一些比较复杂的多角度视频前景检测效果并不好，主要原因是在生成全景图像时，图像融合采用的方式较为简单，导致最终的全景图像效果并不是很好，还存在一定的改进的空间。

⑸基于长短期记忆神经网络（LSTM）的视频异常事件检测模型的核心是将视频异常事件检测转化为一个序列标注问题，使用单向的 LSTM 神经网络对视频帧进行特征的提取，提取到的特征既包含了当前帧的信息又包含了该帧之前帧的信息，之后利用 softmax 回归进行标注分类，或者说对提取到的特征进行解码。但是使用 softmax 回归进行标注分类，每一帧的标注结果并不直接受到之前时刻帧的标注结果的影响。最好希望之前时刻帧的标注结果可以直接影响当前时刻帧的标注结果，这样做更加符合真实的情况。所以本模型可以使用条件随机场算法（conditional random field ,CRF）算法来替换掉原来的 softmax 回归算法对视频帧进行标注分类。

6.参考文献

[1] Brahme Y B, Kulkarni P S. An Implementation of Moving Object Detection, Tracking and Counting Objects for Traffic Surveillance System[C]// International Conference on Computational Intelligence and Communication Networks. IEEE Computer Society, 2011:143-148.

[2] Liao J, Dong R, Li B, et al. A Non-Parametric Motion Model for Foreground Detection in Camera Jitter Scenes[J]. IEEE Signal Processing Letters, 2014, 21(6):677-681.

[3]Zivkovic Z, Ferdinand V D H. Efficient adaptive density estimation per image pixel for the task of background subtraction[J]. Pattern Recognition Letters, 2006, 27(7):773-780.

[4] Barnich O, Droogenbroeck M V. ViBe: A Universal Background Subtraction Algorithm for Video Sequences[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2011, 20(6):1709-1724.

[5] Greff K, Srivastava R K, Koutnik J, et al. LSTM: A Search Space Odyssey[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, PP(99):1-11.