**基于可选线索聚合的自由运动相机中的背景减法**

**摘要：**早前，针对在自由运动相机中的背景减法的方法主要集中在提高运动估计的精确性，在本文中，我们提出通过对前景和背景的可选线索聚合来实现精确的背景减法，我们也提出一个新奇的背景减法框架，称为前景和背景的线索聚合。在这里，前景线索由经过图像对齐补偿的高斯混合模型提取，背景线索由单应性变换后的时空特征提取。随后，该聚合被设计为基于多个层次下的超像素的层次竞争过程，其潜在动机是利用这些线索之间的排他性来补偿它们相应的缺陷。将特定超像素中前景和背景之间的竞争结果作为邻近度，在多个层次上结合超像素和邻近度对前景进行分割。使用标准基准进行的综合评估表明，我们的工作与最新水平相比具有优越性。

**索引词：**背景减法，自由运动相机，前景和背景线索，排他性，超像素

# 1 引言

背景减法是计算机视觉的一个基本任务，它的应用范围广泛，例如视频监控[1]，光学运动捕捉[2]和多媒体应用[3]。该任务旨在将任意像素点分成两类，其中，前景类代表运动物体，背景类代表背景场面。传统上假设通过一个静止的相机来捕捉视频序列，在这种假设下，运动物体成为唯一的运动源，并导致像素强度的变化。在以前的方法中，通过使用统计模型（如高斯混合模型）分析变化来减去背景。然而，在自由运动相机捕获的画面中，背景中的像素不再保持它们各自的位置。此时，背景表现为另一个运动源，导致给定像素出现不规则的观测值，这些不规则的变化使得所提出的静止摄像机算法失效，从而对背景减法问题提出了新的挑战。

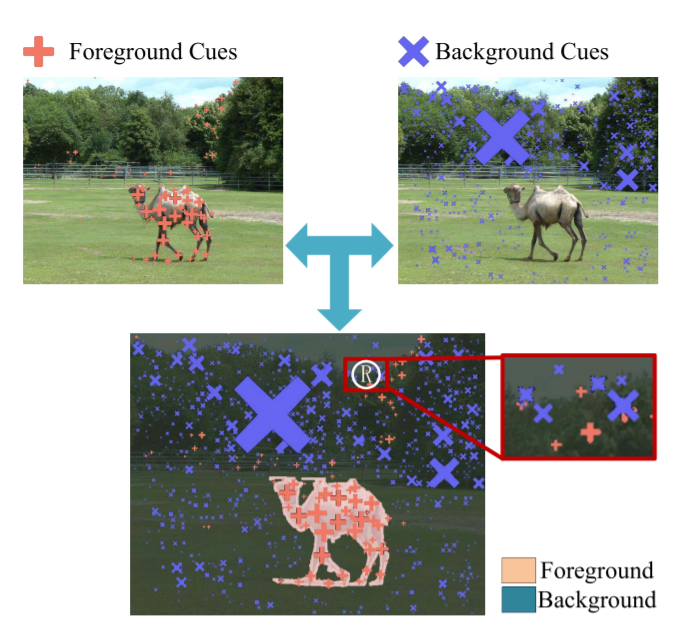


图1 通过整合前景和背景线索的背景减法演示。虽然在属于背景的R区域中捕捉到几种前景线索，但是由于前景线索与背景线索的融合，使得R被正确的标记为背景。

对于自由运动相机，前景的运动和背景的运动重叠，因此很难做出准确的估计。此外，自然场景的复杂性也增添了挑战。在所提出的方法中，不需要精确地估计运动，因为使用可选线索的聚合作为解决方案。如图1所示，由于自然场景的复杂性，前景线索不仅显示在运动目标上，还出现在像图1中R区域这样确定的背景部分。通常，这些摆动的树被错误地检测为运动目标，幸运的是，得益于背景线索的贡献，这些线索的聚合抵消了这些错误，从而产生了一个有效的运动物体的二进制掩膜。受这些见解的启发，我们将注意力集中在聚合可选线索而不是提高运动估计的准确性上，并提出了一种基于前景和背景线索聚合（IFB）的新框架，用于自由运动相机中的背景减除。

在IFB框架中，通过图像对齐延伸的GMM[4]来获取前景线索，用于相机运动的补偿。同时，背景线索由时空特征表示，经过单应性变换[5]滤波后，背景线索在特征之间被认为是相对静态的。随后，考虑到前景和背景线索之间的排斥性，在基于多层次的超像素的竞争过程中，建议将这些可选的线索聚合起来，在那里应用多层次的战略来改进提议方法的准确性。对于每个超像素，前景线索的统计与一个背景线索的统计项相竞争，其结果用作邻近度。在此基础上，利用超像素与邻近度相结合的方法，在提高超像素密度的多个层次上，获取多个邻近图像。最后，将这些邻近图像的平均值与最终前景分割的阈值进行比较。本文的贡献总结如下：

1) 提出了一种新的框架IFB来实现自由运动相机的背景减影。与之前的例子不同，在这里，IFB模型关注的是融合不同的前景和背景线索，而不是提高运动估计的准确性。由于它们的互斥性，这两种线索可以同时使用来弥补各自的缺陷。

2) 将这些线索在多个层次的超像素中进行整合，提高了IFB模型的效率和精度。

3) 通过一个基于背景图像和当前帧的几何约束的图像对齐，我们延伸了GMM来捕获前景线索。从技术上讲，背景图像是没有移动对象的图像，它是从GMM模型捕获的，也表示GMM模型。

4) 提出了一种基于时空特征的背景线索提取方法。利用图像序列的单应性变换估计时空特征的运动，并利用具有相同的相机运动的特征作为背景线索。

# 2 相关工作

近年来，人们提出了许多令人印象深刻的算法来检测使用自由移动相机时的移动目标。我们在接下来的三部分简要讨论了这些算法。首先，第Ⅱ-A部分对基于时空特征的算法进行了综述。然后在第Ⅱ- b节讨论了基于几何约束的算法。最后，由于超像素用于聚合，我们在II-C部分讨论了基于超像素的算法。

## A 基于时空特征的算法

传统背景减法算法假定相机是静止[9]–[11]的，像GMM[4], [12]这样的统计模型已经广泛用于解决这类问题，然而，如果没有固定相机的假设[13]，这会成为一个相当大的挑战，其困难的根本原因在于相机与运动物体同时运动。为了解决这一问题，相当多的早期工作已经使用运动信息对特征进行分类。基于轨迹(如[13]-[20])和光流(如[21]-[28])的方法是自由运动相机中常用的背景减影方法。

Sundaram等人用点轨迹跟踪运动，为了处理无纹理区域，Ochs et al.[17]、Brox和Malik[19]利用颜色信息对长期轨迹进行分析。同样，轨迹的形状也用来减去背景[14]。此外，Elqursh和Elgammal[16]提出了轨迹分类的贝叶斯滤波框架，而且Cui等人通过统计模型对前景和背景的轨迹进行了分类。因此，Berger和Seversky[21]利用轨迹的线性子空间来描述背景。特别地，Wu等人[13]假设运动物体的运动强于摄像机的运动，并通过两种运动之间的差异对运动物体进行分割。

该轨迹可以看作是时间序列[29]中光流的积分，它是在两个连续帧之间捕获的，在运动估计[30]-[33]、目标分割[34]、[35]和运动分割[36]-[38]中有着广泛的应用。kwak等人使用光流初始化运动场，并用贝叶斯滤波器保持背景模型，这与光流估计[39]的贝叶斯推理以及Lim和Han[27]提出的背景减法有关。此外，Zamalieva等人[25]利用Markov Random Fields[40]对传播模型进行了优化。同样，Zhu和Elgammal[29]将背景减法看作是一个多标签问题而不是一个二元标签问题，并使用马尔科夫随机场[40]进行像素级标注。此外，Narayana等人的[23]利用光流的方向来提高算法的鲁棒性，使其与真实世界运动的场景更加一致。出于同样的动机，Sun等人提出了一个完全连接的分层模型来捕捉自然场景中的长程相关性。Bideau和LearnedMiller[28]结合了光流的角度和大小来最大限度地区别物体的运动。最近，Tokmakov等人利用端到端深度学习网络学习光流场特征的粗表示，以减少光流中的误差。

然而，由于光流的亮度一致性假设，光流对光照变化非常敏感。为了解决这个问题，Mohamed等人提出了基于纹理而不是亮度的光照鲁棒光流，而且Kumar等人用随机亮度将亮度分解为反射率和光照分量。遗憾的是，由于自然场景的多样性，这些方法在如此复杂的自然场景中表现得并不令人满意。此外，计算成本是光流[43]的另一个缺陷。即使借助近年来与深度学习相关的[44]、[45]等先进技术，在各种复杂的自然场景中准确捕捉光流仍然是一项颇具挑战性的任务。因此，由于运动物体的不可预测性，运动物体可能产生类似于相机的运动模式。在这种情况下，无论是轨迹还是光流方法都不能正确地标示出运动相机的前景，因此也不能在复杂的场景或从一个快速移动相机获得地视频中工作。在本研究中，带有运动信息的特征只被用来捕捉背景线索，在自由移动的相机中，这些特征与前景线索结合起来进行背景减法。正是这两种线索之间的排他性保证了该方法的有效性。

## B 基于几何约束的算法

由于相机运动导致传统算法失效，因此很多算法使用几何约束来消除相机的运动。

Lim等人使用一个基本矩阵初始化背景和前景标签，该矩阵通过图割算法来进行前景分割。然而，基本矩阵的退化意味着该方法不适用于检测小型移动目标。相反，Liu和Gleicher假设背景在场景中处于支配地位，当一个区域动作的基本矩阵不同于总体背景动作的基本矩阵时，该区域被标记为前景。不幸的是，在Mittal和Huttenlocher 提出的一种实例中指出，目标在不同层中的视差影响意味着［55]方法的效率在复杂场景中很低。在一次设法解决该问题的尝试中，Chen 等人一次性处理背景的几个部分，然后使用总体约束来优化转换的精度。考虑到不同层的影响，具有相同转换的区域仅表现部分背景。为了解决这个问题，Patwardhan 等人在多像素层检测前景。Bugeau和Perez使用概率值验证光流的估计值，允许使用Markov随机域框架来标记像素为前景或背景。除此之外，Yuan等人提出了“平面＋视差”，它是首次在[52]中被提出的，在[52]框架中应用二维层同伦和识别优势层来防止背景运动的估计。此外，Zamalieva和Yilmaz[46]提出了一个2.5D的背景模型来描述场景的外观和几何形状。Zhang等人利用结构、运动和束调整相结合的方法对[48]进行了全三维恢复。他们在[51]之前工作的基础上，在移动的相机中心和3D场景之间进化出了极平面。因此，Bideau等人[38]提出了一种基于透视投影约束和结合语义分割结果的统计模型，它由深度学习网络捕获用于运动分割。

在几何约束的作用下，运动物体是像素变化的唯一来源。在之前的几种方法中，通过统计模型分析像素变化来分割前景。然而，这些前景包含了丰富的噪音，因为反击是不完美的。幸运的是，为集成生成前景线索就足够了，这正是所提议的方法与以前研究中使用的方法的本质区别。

## C 基于超像素的算法

通常采用邻域信息来提高背景减法的鲁棒性(如[13], [59]–[62])，像素不是独立的[59]，应该与其邻域相关，尤其是那些彼此相似的邻域。50]和[63]中描述的基于区域的算法没有考虑像素强度之间的相似性，而像素强度通常在物体的边缘较弱。在[13]中提出的方法中，利用平均位移对轨迹捕捉的前景进行细化，提高了自由运动相机前景检测的鲁棒性，而轨迹的精度也限制了Wu 等人方法的效率[13]。因此，Zhu等人[60]利用超像素对显著性检测结果进行细化[61]，这类似于背景和前景线索的混合过程，该方法与Zhu等人的工作[60]的主要区别在于使用超像素的过程。Zhu等人将突出对象建模为超像素显著性值的优化，在我们的IFB模型中，超像素用于融合前景和背景线索。将某一特定超像素中这些线索的统计量作为其对前景或背景的置信度，并结合这些标记的超像素对运动目标进行分割。由于前景线索和背景线索相互排斥，补偿了这些线索的缺陷，提高了IFB背景模型的鲁棒性和效率。

# 3 IFB框架

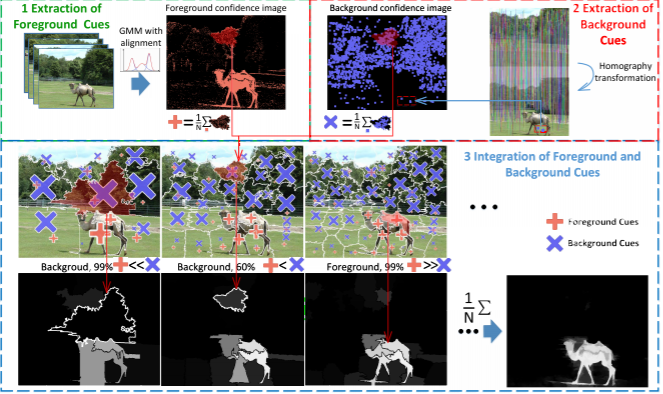
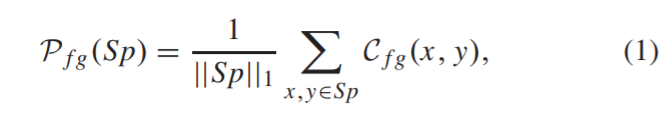
本节解释了IFB框架的细节，FB框架的管道如图2所示，它可以分为3个部分，分别对应于第3- a节，第3- b节和第3- c节。在第3部分a和第3部分b，我们分别介绍了捕捉前景和背景线索的过程。其后，第3- c节解释了超像素在多个层次上对这些线索的整合。

图2 说明了提取前景和背景线索以及这些线索的整合。前景和背景线索分别由保存在前景和背景置信图像中的置信值之和提取。两种线索竞争的结果决定了超像素的接近度，而线索的大小代表了接近度的强弱。线索越多，可能性越大。具有一定邻近值的超像素，构成多层次的邻近图像，并将这些邻近图像的均值作为前景分割的最终结果。

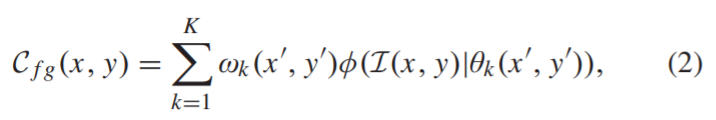
## A 前景线索提取

前景线索被定义为作为前景的超像素的邻近度，它是由一个特定的超像素所包含的置信值的平均值来提取的，如下所示:



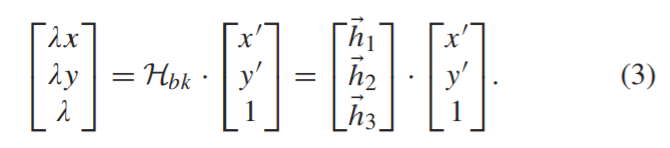
其中Pfg(Sp)为前景线索，Cfg(x, y)为像素在位置(x, y)的置信值，||Sp||1为超像素Sp中包含的像素个数。前景置信度图像Cfg保留当前帧像素的置信度值，如图2第1部分所示。采用高斯混合模型(GMM)[4]对前景置信度图像进行帧对齐步骤补偿。虽然对齐步骤不足以完美地抵消摄像机的运动，但对于生成提取线索的粗糙置信图像是足够的。

GMM是一种经典的基于像素的静态相机背景减法的统计方法，它通过多重高斯函数逼近像素强度的变化，然后将其与权值相结合。我们将位于(x’,y’)上像素的GMM表示为G(x’ , y’) = {ωk (x‘ , y’), θk (x‘ , y’) | k ∈ [1 K] }，ω和θ分别代表每个高斯函数的权重和参数，k是高斯函数的索引，K是高斯函数的个数。在静止相机拍摄的视频中，像素的位置保持不变。因此，位于之前帧(x’，y’) 的GMM直接用于当前帧相同位置的像素上。然而，在自由移动摄像机拍摄的视频中，像素不再保持它们的位置，即使在连续的两帧中也是如此。为了将GMM扩展到可自由移动的摄像机中来捕捉像素强度I(x, y)的置信度，最后一帧中的每个模型G (x’ , y’)通过图像对齐变换到当前帧的对应位置(x, y)，如下所示:

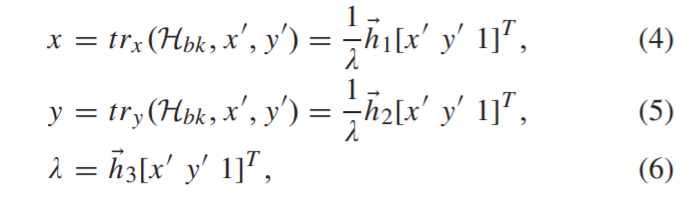


其中，Cfg(x, y)是前景置信图像Cfg中保存的置信值，Φ是高斯函数，θ={μ、σ}代表高斯函数的均值和方差。

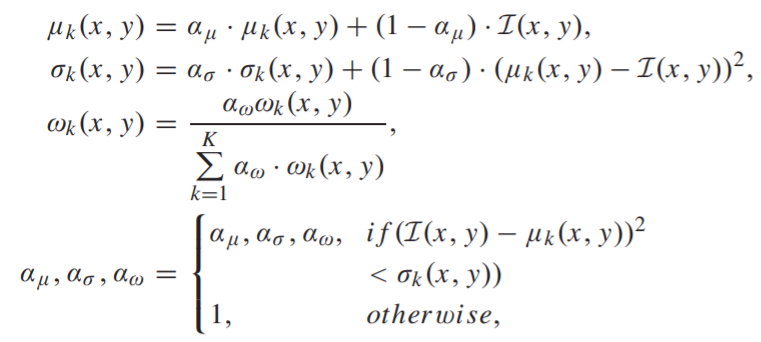
通过一个单应性矩阵[5] Hbk = [h1 h2 h3] T来描述像素(x , y) 与GMM (x’ , y’) 位置之间的转换，如下所示：



通过式3的分解得到(x,y)和(x’,y’)之间的转换步骤，如下所示：



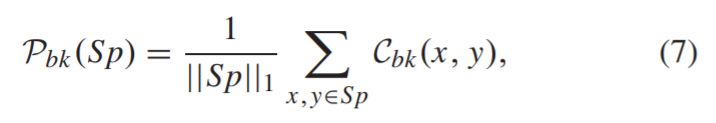
其中，trx() 和 try() 为转换函数，它们将GMM模型的 (x’ , y’) 位置转换到当前帧像素对应的位置 (x, y)，单应性矩阵Hbk由式10得到。关于矩阵Hbk的细节将在后面解释，因为它是在捕获背景线索的过程中得到的。利用式4和式 5将前一帧位置的GMM模型转换为当前帧的对应位置。特别是的，转换到不可能位置的模型被丢弃，并为这些位置创建新的GMM模型，而不需要转换任何GMM模型。为了将这些模型成功地转换到当前帧的新位置，对其进行更新以适应环境的变化，如下所示：



其中，αμ,ασ,αω用于参数更新，K是高斯函数的个数，在实验中，GMM是自实现的，所有的实验都使用默认的参数来强调我们的IFB框架的效率，而不是仅仅用于提取前景线索的GMM模型。采用 {αμ, ασ , αω} = { 0.99, 0.99, 1.05 }，K = 3。此外，视频的第一帧被用于初始化GMM以获取初始线索。

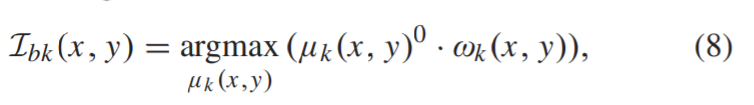
## B 背景线索提取

与前景线索的定义类似，背景线索被定义为作为背景的超像素的邻近度。然而，背景线索的提取是独立于前景线索的提取的，因为整合被设计为前景和背景线索之间的竞争，其中的线索必须相互独立的竞争。通过保存在背景置信度图片（如图2第2部分）中置信值的均值来提取背景线索，如下所示：

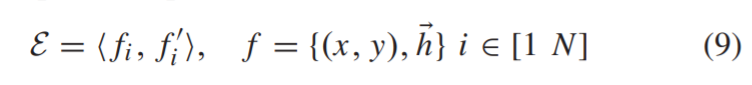


其中，Pbk(Sp)是超像素Sp的背景线索，Cbk (x, y)是位于 (x, y) 像素的背景置信值。

与前景置信图像用于保留扩展高斯混合模型捕捉到的邻近度不同，背景置信图像保留了被认为位于背景中的时空特征关键点。然而，传统的背景假设，即场景的静态部分，在移动摄像机获得的帧中是无效的，因此对象的位置不同使得静态部分不存在。为了解决这个问题，我们假设大部分相对静态的部分是背景。特别是利用时空特征来描述场景的不同部分，将具有相同运动估计的特征识别为相对静态的特征。在此假设下，背景是包含相对静态时空特征数量最多的部分，利用背景置信度图像来保存这些时空特征的位置。根据当前帧和背景图像之间的差异来识别和捕获背景置信度图像，背景图像Ibk是没有运动物体的图像，是GMM模型的表示。通过具有最大权重的高斯函数的均值μ直接获得背景图像中的像素强度，如下所示：

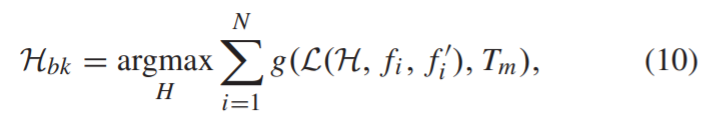


其中，μk (x, y) 和ωk (x, y) 是 (x, y) 处高斯函数的均值和权重。接下来，我们将当前帧和背景图像中得到的时空特征分别表示为 {f1, f2，…fM} 和 { f’1, f’2 ... f’M}，其中M和M’是特性的数目。通常，每个时空特征至少包含两个成分，关键点的描述符h和位置(x, y)。对这些特征的描述符进行交叉匹配，以找到当前帧和背景图像之间的公共部分。特别地，采用欧氏距离的平方来匹配。将 ε 表示为匹配的时空特征对集合，定义如下：

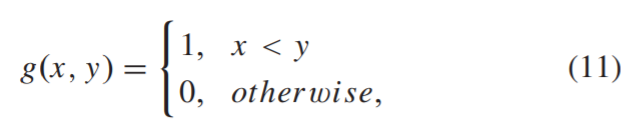


其中，i是共有部分特征的索引，N是共有部分的数目。在实验中，SIFT(尺度不变特征变换)[6]、SURF(加速鲁棒特征)[7]、KAZE[8]特征被用做时空特性的例子，以此展现不同特征输入下IFB框架的有效性。

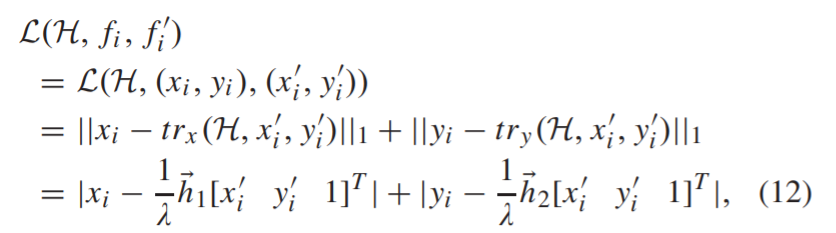
如前所述，背景被定义为包含了最多数量的时空特征，这些特征遵循用于描述背景运动的相同变换矩阵。我们将背景运动的变换矩阵表示为Hbk，它的数学表达式如下：



其中，用 g(L(H, fi, fi’), Tm) 判断fi与fi’ 之间的变换是否在阈值Tm所描述的误差范围内跟随矩阵H，g(x, y) 是一个分段函数：

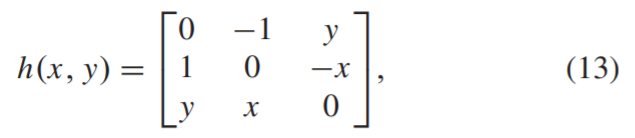


L()是通过矩阵H = [h1 h2 h3]T将背景图像中特征fi’ 的位置变换为当前帧中特征fi的对应位置时，用于误差评估的损失函数。L()的定义如下：

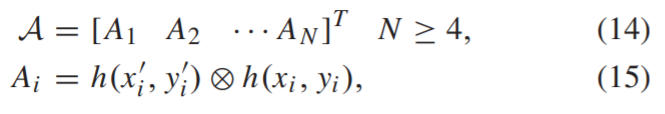


其中，trx和try分别为式 4和式 5所示的变换函数，(xi, yi) 和 (xi’, yi’)分别是特征fi和fi的关键点的位置，λ的定义在式6。

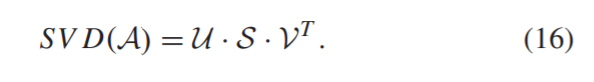
为了获取Hbk，采用单应性变换和RANSAC。为了本文的完整性，我们简要介绍了通过单应变换和RANSAC捕获Hbk的方法，详情请参阅[5]。由于单应性变换的性质，ε 中的四对特征(如式9所示)主要需要获取一个单应性矩阵H，我们用 {(xi, yi) , (xi’ , yi’) | i ∈ [1 N], N ≥ 4} 来表示这些特征对中关键点的位置。首先，获取每个位置的斜对称矩阵如下:



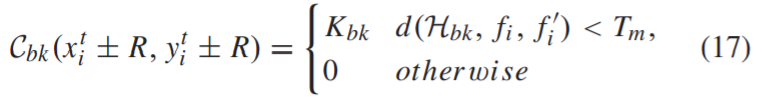
其中，h(x, y)是返回位置 (x, y) 的斜对称矩阵，然后，这些斜对称矩阵组合如下:



其中，⊗表示克罗内克积，(xi, yi) 和 (xi’ , yi’) 分别是ε中各条目的位置，也是当前帧和背景图像中各关键点的位置。最后，矩阵A的奇异值分解(SVD)结果为这四个位置间的单应矩阵的解，如下所示：



对应最小奇异值的“右奇异向量”(V中的一列)是单应矩阵H的解，然后，RANSAC算法随机选择四对 ε 若干次，得到若干个单应矩阵Hi，以式10得到最大值的Hi作为背景运动估计的Hbk。在RANSAC算法中，选择的数目固定为400，这是RANSAC的默认参数。跟随背景运动的 ε中的特征被认为是背景的一部分，其位置被保留在背景置信度图像中，如下所示：

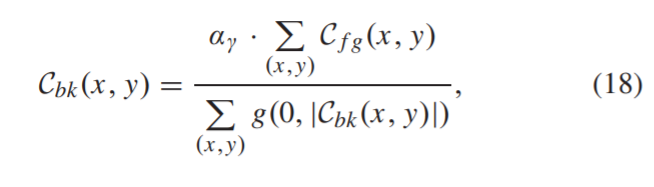


其中，Cbk是背景置信图像，R是关键点的半径，Kbk是表示点贡献的常量，由于关键点是不可见的，为了使点的视觉清晰度，我们简单地将点更改为一个矩形，R是边的长度，在这些实验中，R固定为5。

## C 前景和背景线索的聚合

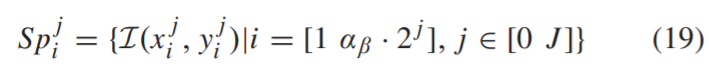
我们的IFB模型效率的主要因素是前景线索和背景线索之间的排他性，因为一种线索中存在的任何缺陷都会被另一种线索抵消。基于这种动机，整合被设计成一个前景线索和背景线索竞争的过程，如图2的第2部分所示。竞争结果被用来作为属于前景或背景的超像素的接近度。然后，一个接近图像，由这些具有邻近度的超像素组成，产生在一个特定的层次上。随后，在多个层次上对多个邻近度图像进行分层捕获。这些邻近度图像的均值被用作最终的接近图像，然后将其与阈值进行比较来分割移动对象。

然而，在自由移动摄像机获取的视频中，用于提取背景线索的Cbk中保存的关键点数量并没有保持不变，因为时空特征的数量是随着背景场景的变化而变化的。相反，由于前景置信度图像是从每个像素建立的GMM模型中获取的，因此受场景变化的影响较小。为了处理Cfg和Cbk之间的不平衡，根据采集到的前景置信图像Cfg调整Cbk中保留的置信值，如下所示：



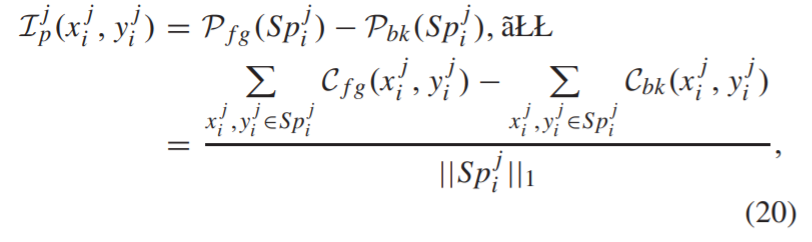
其中，αγ控制背景线索贡献度的参数，Cfg和Cbk分别为前景置信度图像和背景置信度图像，g() 是一个分段函数，已在式11中定义。

众所周知，一个像素不是独立的，而是与其邻域相关的，特别是与其相似的邻域。相似的像素很有可能在前景或背景上共享相同的标签。因此，使用由相似像素组成的超像素来整合前景和背景线索。特别是选择SLIC[61](简单线性迭代聚类)算法来捕获超像素，因为它基于k均值，这使得超像素的尺寸更方便，并且由超像素的数量来控制。利用几何级数来控制超像素的数量，让我们表示超像素在多个层次的数量为 :{ αβ·2j | j∈[0 J] }，这些数字控制的超像素如下所示：

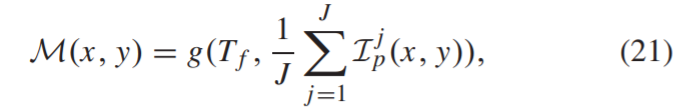


其中，j是层次的索引，i是j层下的超像素的索引，J是控制层次数量的用户参数，点(xij , yij) 表示构成超像素Spij的像素的位置。

考虑到前景和背景线索之间的排他性，线索的聚合被设计成一个竞争的过程。对于每个超像素，将前景线索与背景线索进行竞争，获取超像素是作为前景还是作为背景的邻近度，分别从式 2和式 17的置信度均值中提取前景线索和背景线索。数学上，竞争过程如下：

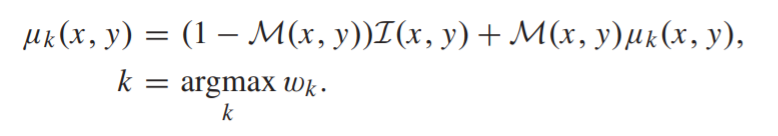


其中，Pfg(Spij) 和Pbk(Spij) 是位于超像素Spij的前景和背景线索，(xij , yij ) 和 ||Spij|| 是超像素之间的位置和像素个数，Ijp(xij , yij) 是特定层次上的邻近图像，如前所述，多个层次捕获的超像素构成不同的邻近图像，将邻近图像的均值与前景分割的阈值Tf进行比较，如下所示：



其中，M(x, y)是运动物体的二进制掩码，J是层数，g()是式11中定义的分段函数。

不幸的是，在用单应性矩阵转换GMM模型的过程中，会积累对齐误差。为了解决这一问题，提出了一种强更新策略来处理误差的累积，它直接用当前帧的像素强度来代替GMM模型的均值，强更新是在前景分割后完成的，在式 8中，我们从具有最高权值的高斯函数的平均值uk中获取背景图像的强度，因此，在强更新过程中，权值最高的uk被对应像素的强度所代替。为了避免运动目标的影响，只对背景像素进行强更新，如下所示：



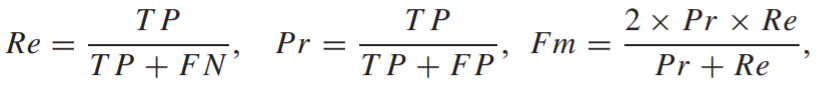
# 4 实验

在这一部分，我们提出了综合实验来评估我们的IFB模型，在IV-A节介绍数据集和评价指标，前景和背景线索之间的排他性的好处在IV-B节中得到了证明，我们还设计了另一个背景线索被消除的IFB模型，并将其与完整的IFB模型进行了比较，最后，在IV-C节中，我们对IFB进行了定量评价，然后将IFB模型与FBMS[16] 基准中的GBSSP[26]、calMoSeg[27] 和MLayer[28] 相对比。

## A 评估度量和数据集

据我们所知，目前还没有公认的标准基准来评估自由移动相机的背景减法方法，已经提出了许多数据集进行评估(如[14]、[17]、[24]、[27]、[28])，其中FBMS[16]数据集最大。因此，我们选择FMBS[16]作为评价IFB模型的基准，FBMS数据集[16]是一个非常有挑战性的数据集，包含59个视频序列，摄像机的运动包括平移、旋转和缩放变换，以及增加其复杂性的场景多样性。因此，任何使用FMBS数据集的比较都是公平和全面的。

在比较过程中，Re(Recall)， Pr(Precision)和Fm(F-measure)指标用于评估，Re和Pr分别是完整性和准确性的度量，Fm是Re和Pr的结合，这些度量的定义如下：



其中，TP和FP是真正类和假正类，这里的正表示前景，而负表示背景。真表示这个检测的结果是正确的，而假表示相反。因此，TP意味着检测的结果是前景同时也是真相。

## B 前景和背景线索之间的排他性

在本节中，我们将讨论前景和背景线索之间的排他性，IFB的效率来自于前景和背景线索之间的排他性，没有这种排他性，IFB就不能准确地分割移动的物体。在IFB模型中，前景线索被用来产生一个粗略的前景，同时，显示场景背景部分的背景线索可以用来补偿前景线索的错误。为了证明这一点，我们设计了另一个背景线索被消除的IFB模型(IFBnobk)，并将IFBnobk与IFB模型进行比较，以证明背景线索的必要性。特别地，已经从FBMS数据集中选择了6个视频进行适当的比较定量和定性比较分别见图4和表1。此外，各方法的ROC曲线如图3所示，以减少比较中主观因素的影响。

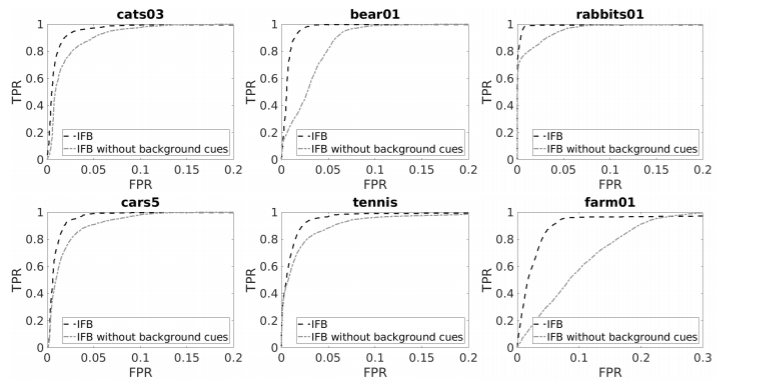
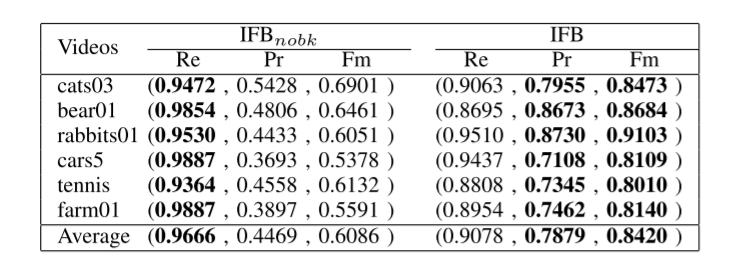


图3 在cats03、bear01、rabbits01、cars5、tennis和farm01视频序列中，没有背景线索的IFB和IFB的ROC曲线。

表1 使用RE,PR, FM指标的IFB和IFBnobk 的定量比较



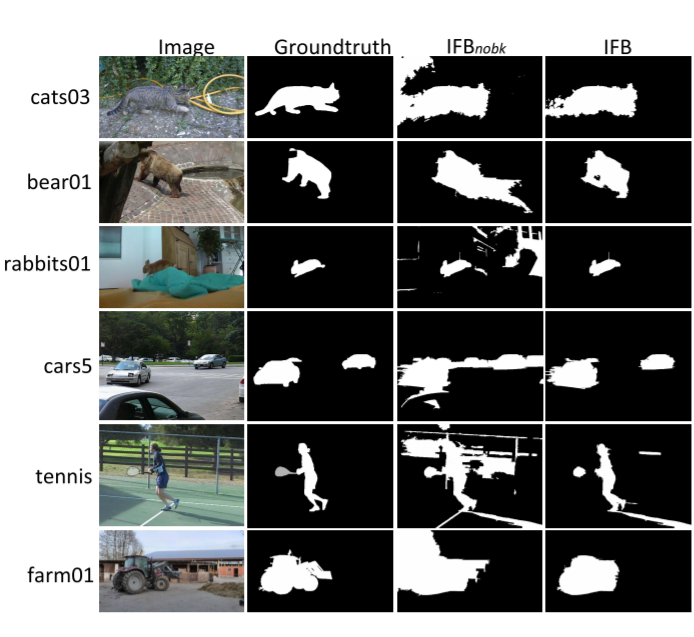
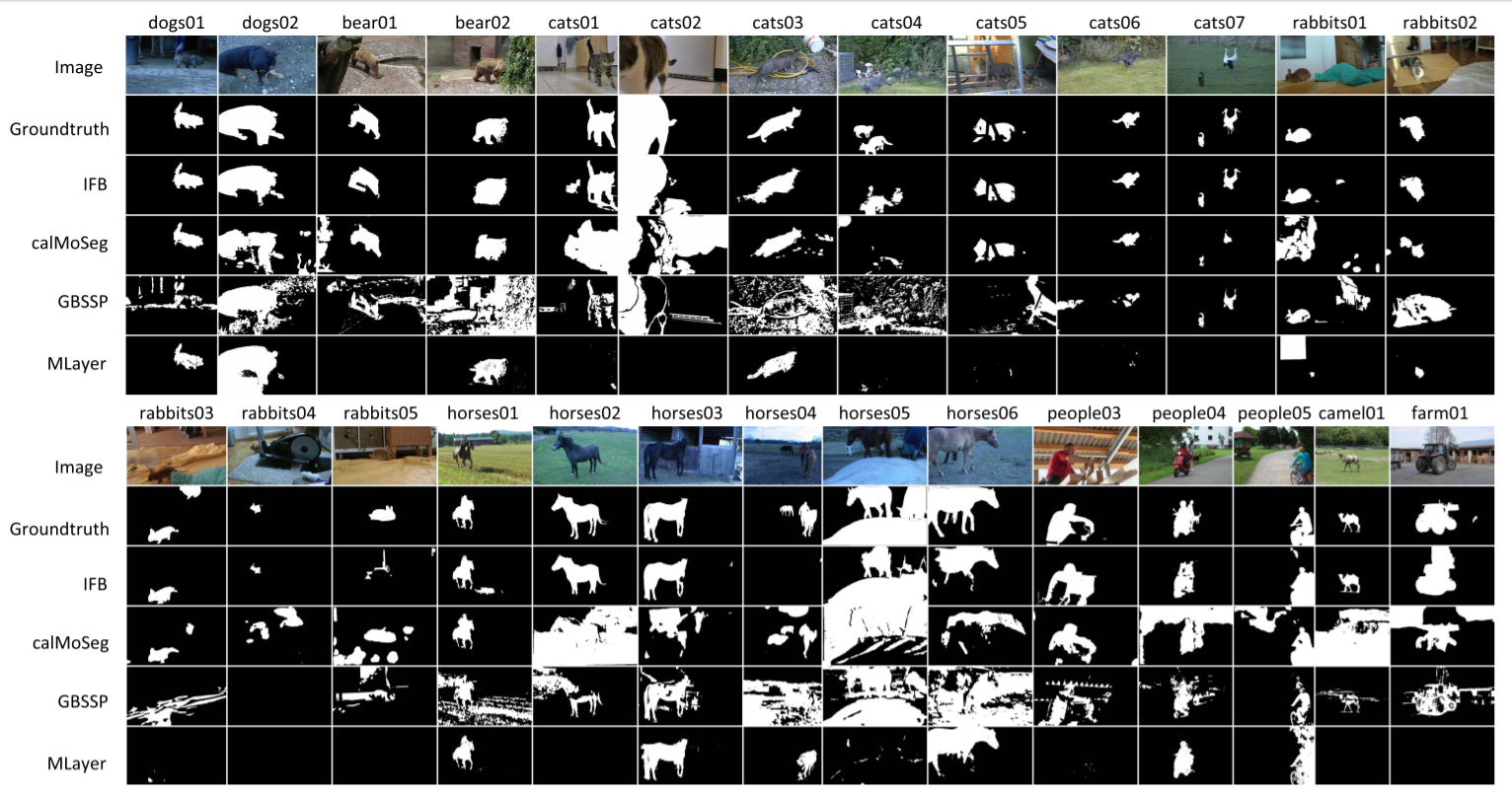


图4 IFB与IFBnobk的定性比较

如表1所示，IFB模型的Fm评分高于没有背景提示的IFB模型，因为Fm度量是最重要的，所以有理由宣称IFB产生的前景比IFBnobk分割的前景更好。此外，由于缺少背景线索，所有的前景线索，无论是对的还是错的，都被用于前景分割。因此，IFBnobk分割的前景更加完整，Re具有更高的分数。但是，由于前景线索是粗糙线索，错误地检测到几个作为前景的背景部分，所以IFBnobk的Pr分数低于IFB。在IFB模型中，由于背景线索被用来弥补前景线索的缺陷，所以IFB在Pr指标上有了明显的改善，但在Re分数上有很小的牺牲。最终，IFB模型的Fm评分高于IFBnobk，背景线索的重要性也比较突出。

如图3所示，IFB的ROC曲线更靠近左上角，而不是IFBnobk，这一现象表明，IFB模型在前景分割方面的收益大于成本，图4所示的前景也反映了这一现象，因为IFB的前景具有较高的准确性，在完整性方面的牺牲可以忽略不计。



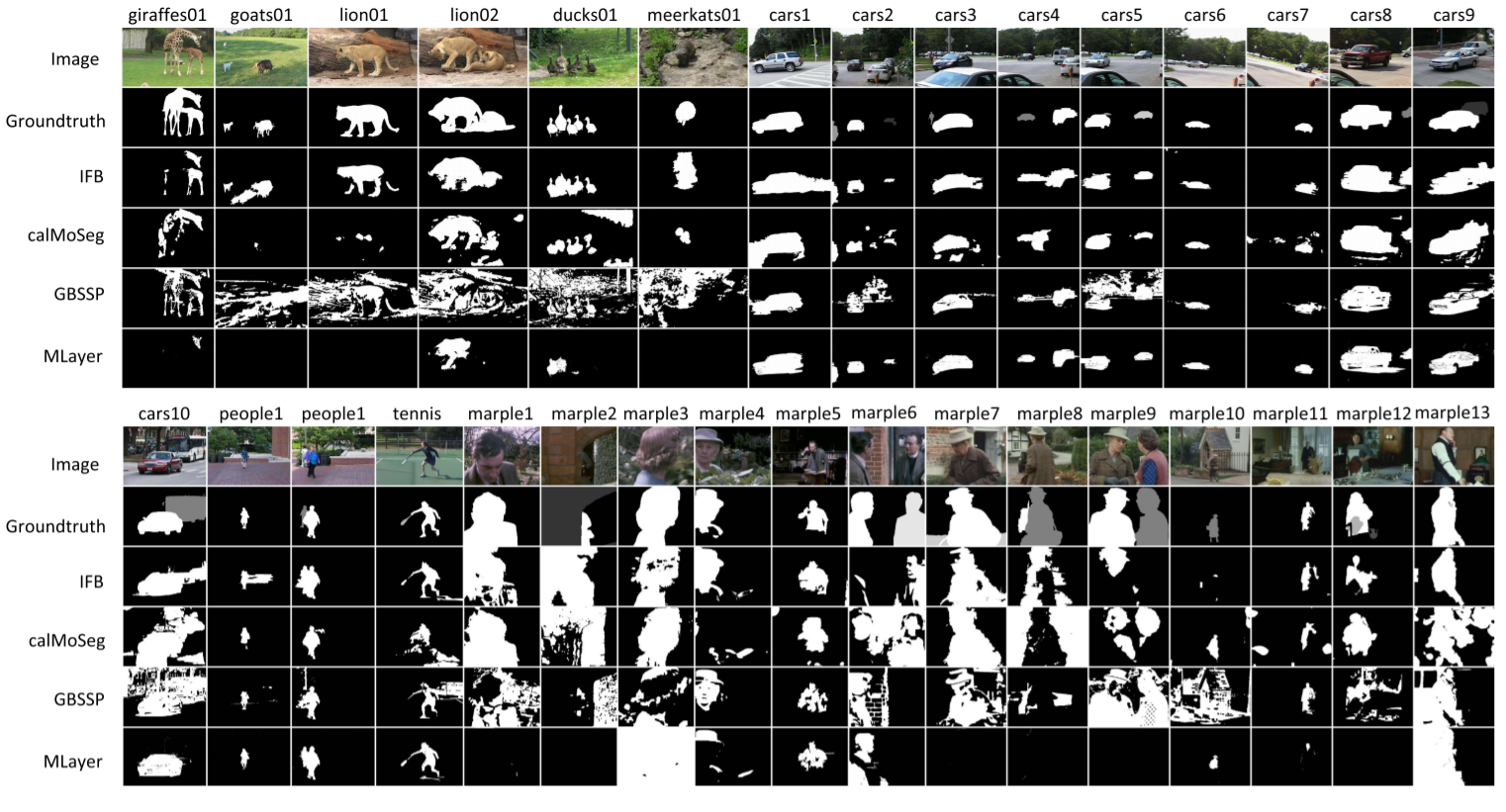


图5 IFBSI的定性比较，它的输入是FBMS数据集中的SIFT特征[6]、GBSSP[26]、calMoSeg[27] 和 MLayer[28]

## C IFB的定量评估

将IFB模型与FBMS[16]基准中的GBSSP[26]、calMoSeg[27]、MLayer[28]进行比较，这两类比较算法都是基于光流的，特别是MLayer是基于大位移光流(LDOF)[30]，这是一种著名的光流提取算法，定量比较见表2，定性比较见图5。通过作者提供的实现，获得了比较算法的结果。对于我们的IFB模型，所有结果都是在与表3相同的参数下执行的。此外，IFBSU、IFBKA、IFBSI的结果分别由SURF[7]、KAZE[8]和SIFT[6]特征输入的IFB模型捕获，考虑到本文篇幅的限制，仅将IFBSI的结果如图5所示，并将同类别的视频序列放在一起讨论。

表2 IFBSU、IFBKA、IFBSI的定量比较，它们是分别具有SURF[7]和KAZE[8]和SIFT[6]特征输入的IFB模型，并且三种最先进的算法对FBMS数据集使用FM度量

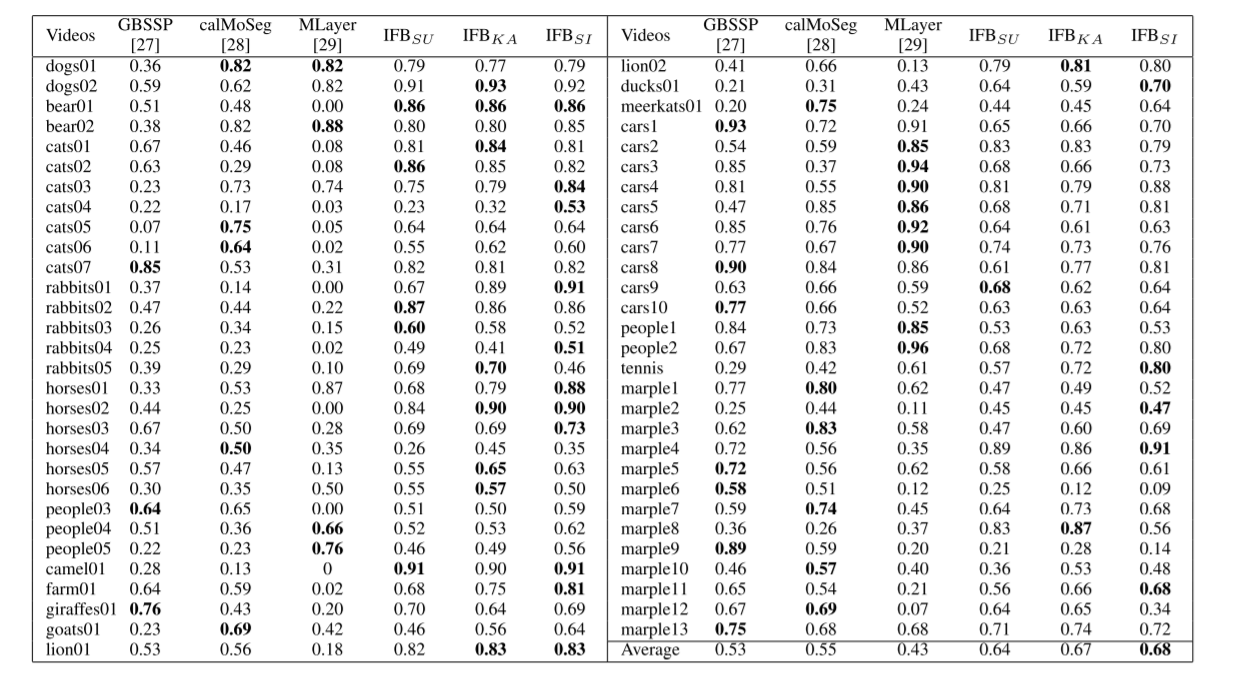


表3 IFB框架和GMM模型的参数值

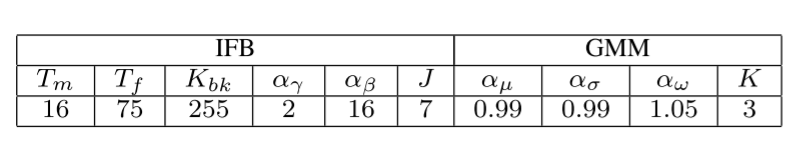


表2给出了不同算法在FBMS数据集[16]视频中的Fm评分，IFBSU、IFBKA和IFBSI得益于前景和背景线索的互补性，在平均Fm度量方面取得了不可否认的进步。此外，在不同的输入条件下，IFB的计算结果比较接近，这进一步证明了我们的IFB的效率与输入无关。在dog01-02的视频中，可以看到狗在镜头前走来走去，这些视频序列的挑战来自于摄像机运动的复杂性，包括旋转和缩放变换。例如，在dog01的视频序列中，一只狗走过场景，摄像机的运动对应于平移变换，在这种情况下，光流有足够的能力对前景和背景进行分类。但是在dog02视频序列中，摄像机的运动变得更加复杂，光流无法处理这种情况。因此，GBSSP[26]和calMoSeg[27]都是基于光流的，性能相对较差。与上述情况相比，由于大位移光流(LDOF)[30]的优良性能，基于LDOF的MLayer[28]获得了令人满意的结果。类似地，所提出的方法中使用的单应性变换对这两种变换都是强健的。因此，我们的IFB模型可以很好地实现对两个视频序列的高性能。在bear01-bear02的视频序列中，熊移动缓慢，摄像机的运动比运动物体的运动要强烈得多。在这种情况下，光流不够敏感，无法区分相机的运动和运动物体的运动。因此，在bear01视频序列中，MLayer[28]呈现无效。在我们的IFB模型中，使用GMM模型来获取前景线索，它是一个统计模型，用于处理缓慢移动的对象。因此，IFB模型的背景线索与前景线索相平衡，所以IFB模型具有从强烈的摄像机运动中分割慢速运动物体的能力。

IFB模型在cats01-07和rabbits01-05视频序列中表现良好，在cats01-07的视频序列中，猫产生了一个突然的动作，这是这些序列的主要挑战，在短时间内，摄像机快速移动。在这种情况下，很难用光流来精确地描述突然的运动，这就是为什么GBSSP[26]、calMoSeg[27]和MLayer[28]在这类视频中都不能很好地工作，因此，图5所示的前景也证明了比较算法之间的准确性问题。相反，我们的IFB模型的背景图像在每一帧都会更新。因此，IFB的单应性变换具有良好的处理相机快速运动的能力。因此，我们的IFB模型在这些视频序列中表现良好。然而，不幸的是，我们的IFB模型不能解决猫外套的伪装，因为前景线索是从基于颜色信息的GMM中捕获的。这就是为什么我们的IFB模型在cats04视频序列中只能获得限制的Fm分数的原因。

在horses01-06视频序列中，IFB模型得到了显著的改善。在这些视频中，马在自然场景中奔跑得很快，除了强大的摄像机运动，复杂的自然场景是这些视频序列的另一个挑战。在复杂的自然场景中进行准确的运动估计是非常具有挑战性的，这也是为什么相比较的算法在这些视频中表现不佳的原因。幸运的是，粗略的线索足以使我们的IFB模型产生准确的前景。由于前景和背景线索的融合提高了IFB模型的效率，我们的IFB模型在这些视频序列中表现出了良好的性能。在people03-05和camel01的视频片段中，我们观察到场景的环境更加复杂。特别是在camel01视频序列中，GBSSP[26]和calMoSeg[27]的Fm评分分别为28%和13%，而MLayer[28]则完全失效。然而，在我们的IFB模型中，扩展的GMM捕获前景线索进行整合，前景线索和背景线索的交替使用提高了前景分割的准确性。最后，IFB模型的Fm分数达到91%，远远高于其他算法。

在farm01、giraffes01、goats01、lion01-02、ducks01、meerkats01和tennis视频片段中，所有的场景都是自然的，这些动物的动作是不可预测的，场景的环境是复杂的。在这类场景中，很难准确地估计摄像机的运动，然而，我们的IFB模型的重点是整合可选的线索，而运动的准确估计是不必要的。采用超像素进行积分，提高了精度和效率。因此，我们的IFB模型在这些视频序列中也显示出良好的性能。不幸的是，在这些FBMS数据集[16]的视频中，其中包含了cars1-10、people1-2和marple1-13，几个视频的帧数是有限的。例如，cars1视频仅包含19帧。在这种情况下，像GMM这样的统计模型没有足够的帧来产生足够用于聚合的前景线索，运动物体的大部分被识别为背景。因此，在没有足够的学习帧的情况下，GMM模型获取的背景图像中包含了运动的物体，影响了背景线索的提取。前景和背景线索的失败导致了IFB模型在这些视频序列中的有限性能。实验证明，当视频至少包含40帧时，我们的IFB模型工作良好。此外，在marple6和marple9的视频序列中，移动的对象被包含在初始帧中，并且在其他时间序列中可以看到移动的对象，与视频帧数有限的情况类似，IFB无法捕获足够的线索进行整合，这是IFB在这些视频序列中表现不佳的主要原因。

## D 实现细节和时间复杂性

我们的IFB框架的完整源代码可以在https://github.com/zhaochenqiu/IntegrationFgBk找到。IFB框架已经通过Matlab和c++的混合实现，利用c++实现了IFB框架的几个关键部分(如GMM模型的转换、线索的集成)，利用Matlab将这些部分组合成一个框架。SURF[7]、KAZE[8]和SIFT[6]特征是通过默认参数的MexOpencv[64]获取的。在聚合部分中，采用多个线程用于加速。

表4 IFB框架不同部分的运行时间，运行在分辨率为340×270的帧内(秒/帧)

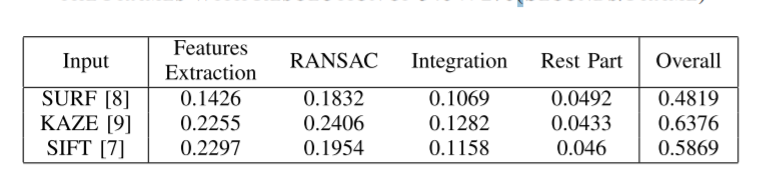
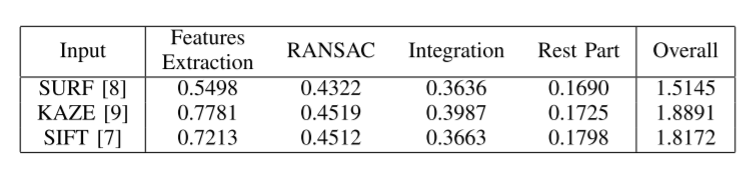


表5 IFB框架不同部分的运行时间，运行在分辨率为680×540的帧内(秒/帧)



该方法的运行时间如表4和表5所示，其中，IFB框架的运行时间以帧为单位进行评估，分辨率分别为680×560和340×270，除了IFB的整体运行时间外，还展示了IFB关键部分的运行时间，包括SURF[7]、KAZE[8]和SIFT[6]特征的提取、RANSAC、前景和背景线索的整合，以及IFB框架的其余部分。实验是在ubuntu16.04的Matlab 2017b和g++环境下，在i7-6900k八核处理器的PC机上进行的。在评价实验中，所有的图像都没有归一化。

# 5 结论

在本文中，我们提出了自由运动相机背景减法的IFB框架。不像以前的工作，试图提高运动估计的精度，我们的IFB集中于集成替代前景和背景线索的前景分割。特别地，通过背景运动估计的GMM模型来检测前景线索，而通过单应性变换从时空特征中捕获背景线索，其中SURF[7]、KAZE[8]和SIFT[6]特征为例。然后，利用多层面的超像素来整合这些线索。如之前所示，IFB的效率来自于前景和背景线索之间的排他性，以及超像素在多个层次上的利用。通过综合实验验证了该方法在不同场景下的有效性。通过与其他先进方法的比较，证明了该框架的高效性，具有实际应用的潜力。