基于时空信息动态融合的视频显著性检测算法

摘要：

针对目前视频显著性检测算法处理存在背景扰动的视频鲁棒性较差的问题，本文提出了一种基于时空信息自适应融合的视频显著性检测算法。综合视频序列所固有的底层空间特征和时间维度上的运动信息，在空间上, 采用基于区域颜色对比度的方法计算得到空间显著图，保证每帧图像显著区域的空间完整性。在时间上，依据运动的连续性和方向一致性，利用连续多帧的图像的运动信息计算运动能量图和方向一致性特征图，整合为运动（时间）显著图。最后，根据人眼对不同运动特性视频的主观感知，~~设置自适应权重，生成动态融合的时空显著图，得到最终能兼顾动静两种情况的视频显著图。~~通过自适应加权的方式将空间显著图与运动显著图融合，生成能够兼顾不同背景情况的最终显著图。

实验结果表明,该方法能够均匀地凸显视频中的显著区域，具有较好的视频显著性检测效果,在检测精确度和鲁棒性等方面优于现有视频显著性检测算法。

关键字：视频显著性；改进光流法；运动矢量场；显著图

1 引言

人眼每时每刻都在接收海量的信息，视觉系统能够自觉将注意力集中于场景中的显著区域，这些区域相比其它位置较为突出，能够为视觉带来强烈的刺激，在计算机视觉领域这种刺激被称为显著性，图像显著性检测模仿人眼自动搜索显著区域的行为，使用图像处理的方法分析生成显著图，图中每个像素的灰度值表示该像素能够引起注意的程度。显著性检测可以计算和衡量图像中各个位置吸引注意的可能性，对于图像的分析处理以及计算资源的分配具有重要的意义，在运动目标识别，图像检索，自适应压缩，视频索引等计算机视觉领域拥有较广阔的应用前景。

目前，针对静态图像的显著性检测方法已经较为成熟，然而由于计算的复杂性，针对视频图像序列的显著性检测研究较少，本文主要讨论视频显著性。视频由多帧图像组成，与单帧图像相比，帧与帧之间的强相关性使得视频图像在时间维度上包含更丰富的运动信息。神经生物学研究发现, 相比纹理、颜色等图像底层特征,人类视觉系统对场景中的运动信息更为敏感。因此视频的显著性既要考虑单帧图像底层特征所带来的空间显著性，也应考虑视频序列在时间维度上的运动信息对于目标显著性的影响。

2003年，Itti等最早将运动信息引入显著性检测的研究中，在原来的静态Itti模型中加人了运动特征，将连续两帧图像在不同尺度的差异作为运动产生的显著性。2004年，Zivkovic等将混合高斯模型GMM(Gaussian Mixed Model)应用于视频显著性检测中，用多个高斯模型来表征图像中各个像素点，在每一帧图像更新高斯模型，通过像素点与模型匹配得到视频显著图。2006年，Harel等提出了一种基于图的不相似性度量的显著性计算方法(Graph-based Visual Saliency,GBVS)，其主要思想是运用马尔可夫随机场的特点构建二维图像的马尔可夫链，通过求其平衡分布可得到视频显著图。GBVS算法的鲁棒性较好，但在显著区域边缘容易造成大面积的模糊，不能确定显著目标的明显边界。2008年，Guo等将像素点的灰度、 颜色和运动特征组成的四元数组进行四元傅里叶变换，用于提取图像序列在时空域上的相位谱 , 通过相位谱分析得到图像序列的显著区域。该方法仅采用连续两帧之间的差异估计运动信息，当背景的运动较为剧烈时，无法准确定位显著目标的位置。2011年，Barnich等在混合高斯模型的基础上提出了视觉背景提取(Visual background extractor,ViBe)算法，采用邻域像素来创建背景模型，通过比对背景模型和当前输入像素值来检测视频中的显著目标。Vibe算法采用视频序列第一帧初始化背景模型，使得模型的初始化非常迅速，但是当待检测的前景目标在第一帧中出现时,鬼影将伴随着模型的初始化而产生，严重影响检测效果。

2015年，Wang等提出显著性感知测量视频目标分割(Saliency-aware geodesic,SAG)方法，利用时域上具有运动的轮廓边界预测前景目标的位置，通过计算超像素与边缘像素的测地距离生成时空显著图，同时构建前景和背景的全局外观模型和动态位置模型，最终实现时空一致的视频显著目标分割。SAG算法主要依赖于图像的边缘特征和运动梯度特征，在复杂场景下容易产生噪声区域。

现有的视频显著性检测算法通常仅在相邻两帧之间估计运动信息，在处理具有扰动的视频时，易将背景中的干扰误判为运动信息，鲁棒性较差。并且，目前的视频显著性检测算法对于时间和空间的显著性融合多采取简单的线性相加或相乘的方式，通常将运动信息与亮度、对比度等图像底层信息同等对待，没有考虑到不同背景情况下运动信息对显著性影响程度不一的事实，对不同情形视频的处理效果差距较大，不具有通用性。

针对上述问题，本文提出了一种基于时空信息动态融合的视频显著性检测算法(Adaptive fusion of spatial temporal information, AFSTI)。本文采用目前空间显著性检测方法中性能优越的基于区域颜色对比度的(regional contrast based saliency extraction algorithm, RC)方法计算每帧图像的空间显著图，用以均匀完整地突出显著目标所在的区域；依据运动的连续性，本文利用连续多帧图像的运动信息对当前帧的运动显著度进行估计，采用光流法提取像素在每连续两帧之间的位移场，采用改进的中值滤波排除背景干扰造成的奇异值，计算每个像素点在连续多帧中的“运动能量”，从而强调在连续多帧中具有持续稳定运动的区域的显著度，削弱背景中干扰所带来的影响；由于同一运动对象区域的运动矢量应具备相同或近似相同的方向，依据运动的方向一致性，采用夹角余弦法计算每个像素点与其周围像素点进行方向相似值，构成方向一致性特征图，进一步区分显著运动目标和背景；将运动能量图与方向一致性特征图组合构成运动显著图。时空信息融合阶段，根据人眼对不同视频的主观感知，设置自适应融合权重，生成能兼顾动静两种背景情况的视频显著图。

空间显著性：

视频图像序列包含丰富的运动信息和固有的图像底层信息。在众多底层信息中，人类视觉系统对颜色对比度最为敏感，能够快速捕获图像中与周围相比对比度较大的区域。因此，空间显著性检测方面，本文采用目前方法中性能优越的基于区域对比度的空间显著性检测算法(RC)。该算法同时考虑图像底层信息中的颜色对比度和空间相干性，以获得均匀突出空间显著区域的显著图。

为了均匀突出空间显著目标，首先采用基于图的图像分割方法将图像分割为不同的区域，再进行区域级的对比度计算，区域和区域的颜色对比度可表示为：

为颜色在区域的所有种颜色中出现的概率，。

为区域中的颜色和区域中的颜色在颜色空间的距离。

对于每个区域，通过计算它与其它区域的空间距离加权颜色对比度得到该区域的空间显著性值：

其中表示，用以强调较大区域颜色对比度对显著值的贡献。

表示区域和为权值控制系数，空间距离对区域的显著性的影响随着值增大而减小。计算图像中每个区域的显著值，进而得到与该帧图像相对应的空间显著图。

相比于较远区域的高对比度，近邻区域的高对比度可以为一个区域带来更强烈的视觉刺激，并且人类视觉倾向于捕捉视线中具有空间完整性的目标。因此，利用距离加权的颜色对比度计算区域的空间显著值更符合人类的感知过程。

时间显著性

视频序列由时域上具有较强相关性的图像排列而成，其与单帧图像显著特征最大的区别便是增加了时域的运动显著性特征。通常，连续帧中的运动能够迅速引起人类视觉的注意,因此运动信息对于视频显著性的贡献不容忽视。目前，常用的图像帧之间的运动信息提取方法包括帧间差分法、光流法等。帧间差分法虽然速度较快，但是易产生空洞效应，无法完整地提取运动目标。光流法能够在无需知道场景先颜信息的情况下，较好地检测出场景中的运动。本文采用光流法提取视频中的运动信息。从运动能量和方向一致性两方面对时间显著性进行度量，以区分显著目标和扰动背景。

通过光流法在每相邻2帧之间进行计算,得到各帧的运动矢量场，为视频的总帧数，点在第tt+1帧之间的运动矢量为。

相应地，运动幅值和方向分别为

1)运动能量图的计算

光流场的计算对环境中的扰动较为敏感，为了消除光线变化等干扰所带来的奇异值影响，，本文使用改进的中值滤波器对运动幅值进行平滑。

表示经过平滑后的幅值

表示点t~t+τ每相邻两帧之间的第三大运动幅值

一般情况下，采用相邻两帧之间运动幅值的平方表示运动能量的大小。在时间域上较为显著的区域通常具有持续的运动，而背景中包含微小扰动的区域则不具备这样的性质。保持持续运动的点在连续帧段中的运动总能量值较高，而环境中存在突发扰动的点由于幅值的平滑处理以及运动的间歇性，在该帧段中运动能量值较低。~~本文引入运动能量的思想，~~综合考虑t~t+τ之间每相邻两帧的运动幅值，得到每个像素点在该连续帧段中的运动能量,作为该点在第t帧的能量值。这样比仅考虑t与t+1两帧之间的信息更能反映实际运动情况。为了进一步提升算法性能，在计算的过程中引入指数衰落因子以弱化较远帧对当前帧的影响。~~本文采用运动能量E表示像素的运动显著度，E通过运动幅值的平方加权相加获得：~~

点在第t帧运动能量值为

其中为指数衰落因子，τ为连续帧数，本文中取5

则第t帧的运动能量图可表示为：

2）方向一致性特征图的提取

运动的方向一致性特征是指在当前帧中属于同一运动对象区域的运动矢量应具备相同或近似相同的方向。如果某个区域内的运动矢量表现出方向不一致性，则该区域有可能位于运动区域的边缘或处于背景当中。通过光流法计算得到的运动矢量本质属于向量的范畴，本文采用夹角余弦法在一定的空间邻域范围内对运动矢量进行方向一致性度量，提取出运动的方向一致性特征图Dmap。

首先对所有运动向量进行无量纲化处理，使各元素都为正，用以保证夹角余弦的取值范围为[0,1]，且与向量夹角呈现单调递减的关系，取值越大表明两向量夹角越小，即方向一致性越高，当值为1时表示两个向量的方向完全相同。与之间的夹角余弦值可表示为：

计算点在第t帧的运动矢量，并确定其邻域，记邻域的大小为（本文中取值为3）。分别计算与其邻域内对所有其它运动矢量的夹角余弦值，实现运动方向的一致性度量，将在邻域内计算得到的所有值累加求和，并进行归一化处理，最终得到运动方向一致性特征图。

其中表示以像素为中心，窗口长度为的邻域的像素集合，属于集合，为计算值的次数。某点的值越大，说明该点与周围点的运动方向一致性越高，表明该点位于显著的运动区域范围内，由于同一运动区域内各点的运动方向相同，则该区域中的每个点的值较高且相近。

第t帧的运动显著图可表示为：

的计算综合考虑近邻τ帧的运动能量以及邻域范围内的运动方向一致性

特征，可以凸显具有连续运动的区域。

**显著图融合**

通过计算单帧图像的底层空间信息和几帧图像之间的运动信息分别得到空间显著图和运动显著图，根据心理学研究，当背景静止不动时，相对于空间显著性特征，运动显著性特征更能吸引人的注意；当摄像头平动时，视频中的背景随运动物体一起移动，此时局部运动信息不再突出，图像底层特征所带来的对比度，亮度等空间特征更能引起人的注意。因此，针对实际情况，本文采用自适应动态加权的方式对空间显著图和运动显著图进行融合，得到能够适应静、动不同背景情况的整体显著图，定义如下式：，

其中，为自适应动态权值，，为当前运动显著图中的最大值和平均值。当视频中的背景不动，只有小范围的区域存在运动，则运动信息在整个运动显著图上的均值很小，且与的峰值相差很大，此时增大，强调运动信息对显著性的影响；背景随显著目标一起运动时，视频中的运动信息较为平均，则减小，从而弱化运动信息的影响，为空间显著图赋予一个较大的权值。

采用自适应动态加权的融合方式，使得本视频显著性检测算法能够根据视频内容变化，自适应调整空间和运动显著图的比重，既考虑了运动容易引起注意的特点，又合理地度量了空间显著性，符合人类感知过程的相关规律。

得到了静态和动态特征显著图之后，下面的工作就是根据物体静态和动态显著性的关系，对两者进行融合从而生成整体显著图。（1）时空显著性相互关系

根据心理学研究，相对于静态显著性特征，动态显著性特征更能吸引人的注意。例如，在一段视频中，动态目标相对于静止物体具备更显著的特性。另一方面，在摄像机扫视的情况下，也即是说局部运动不显著的情况下，静态的亮度、颜色、方向等特征则更容易引起人的注意。

因此，对于时空显著性可以得出以下关系[14]

如果视频时空信息中的运动信息非常显著，那么动态显著性将比静态显著

性更重要，即组合时应该更加倾向于时域信息。

另一方面，如果视频片段中局部运动并不显著，那么组合时应当更倾向于代表

空域信息的静态特征显著图。

在研究了时空显著性之间的关系后，可以得出：采用固定权重的线性组合来

生成整体显著性是不符合实际情况的，必须采用自适应动态加权的方式对静态和

动态特征图进行融合。如果动态特征显著图中的运动程度比较剧烈，则对动态特

征显著图赋予一个大的权值；类似地，如果动态特征显著图中运动程度相对比较

低，则对动态特征显著图赋予一个较小的权值。这样，可以使得模型能够根据视

频内容变化，自适应调整静态和动态显著图的比重，即空域和时域显著度的比例。

**实验结果与分析**

为验证本文算法的有效性和鲁棒性，将本文提出的算法在四组视频显著性研究领域具有挑战性且被多次引用的视频数据库[17][18]共12段视频上进行定性和定量评价。所选数据库中的视频全面并具有代表性，包含不同的目标运动状况、场景复杂度、背景局部扰动情况，拍摄设备移动条件，涵盖了目标大小不一、背景信息不同的各种情况。以上数据库都提供了手工标记真实显著前景Ground Truth，用于后续定量分析比较各算法的显著性检测效果。选择目前5种主流算法与本文算法进行比较，这些算法包括FT、GBVS、PQFT、Vibe、SAG。实验的软件环境为VS2013，硬件为Inter Celeron CPU G1840，4.00 GB内存。

定性比较

在视频数据库中，图2为car前景目标低速运动，对应的视频为行人和车辆相向运动穿插并离开，该视频背景静止且场景包含大块匀质区域，真实的显著目标为行人和车辆；图3birdfall为前景目标快速运动，对应的视频中一只鸟在复杂森林背景中向下飞行，该视频背景中存在树叶飘动等微小扰动且纹理较为复杂，真实的显著目标为向下运动的鸟；图4surfing为前景目标低速运动，对应的视频为运动员在海上冲浪，该视频背景存在动态干扰，摄像机随目标的前进平稳移动但背景中的水波浮动的较为剧烈，真实的显著目标为两个运动员；图5girl为前景目标快速运动，对应的视频为女孩在草地上奔跑，摄像头平动且背景包含大块匀质区域，真实显著目标为奔跑的女孩；图6cheetah为前景目标快速移动，对应的视频为一只羚羊在土地上奔跑，该视频中目标与背景颜色相似且背景纹理较为复杂，摄像机剧烈抖动带来较大的干扰，真实的显著目标为奔跑的羚羊。

各个视频显著性检测算法在不同数据库的检测结果表明，本文所提出的视频显著性检测算法在较为精确地检测出视频中的显著性区域的同时，能够很好地克服背景中的微小扰动以及相机运动等干扰的影响。FT算法是基于静态图像的显著性检测,从底层颜色特征出发，只能反映图像具有复杂纹理、强烈空域反差出的信息, 没有考虑视频图像序列间在时间维度上的运动信息，并且过于依赖于图像的边缘特征，因此不能很好地区分显著运动的物体与背景；包含图的抖动特征通道的 GBVS方法只能反映出运动目标的大致位置,但在显著区域内部容易产生大面积的模糊，并且在背景纹理复杂或者摄像机运动时也存在较多噪声;PQFT算法将颜色、亮度、运动向量作为四元组计算显著性，没能将图像底层特征与运动信息对于显著性的贡献进行权衡，在处理小目标运动且背景复杂的视频时，引入了很多不相关的空间信息，错将背景中的干扰判断为显著区域，例如birdfall，car。Vibe采用首帧图像初始化背景模型，当显著目标出现在第一帧时 ，后续显著图中会出现“鬼影”现象，严重影响显著性检测的效果，例如girl，cheetah，car。本文提出的算法能在不同视频场景中取得较优效果是因为本文利用基于运动矢量场的方法计算运动显著性，可以有效地反映视频的运动信息，虽然背景中的动态扰动与前景目标均处于运动状态，但位于背景和前景区域中的像素点在运动强度和方向上明显不同，本文通过计算“运动能量”强调在连续帧具有持续稳定运动的显著区域，运动方向一致性值的计算进一步滤除了背景中扰动噪声的影响；计算运动显著性的同时，本文也从视频图像的底层特征出发，计算视频图像固有的空闲显著性，从而确保显著区域的空间完整；进一步从时空角度采用自适应加权融合的方式对显著图进行优化，以适应不同的背景情况。因此在大多数测试视频上本文提出的算法都有更好的背景区域抑制能力，且检测出的显著性区域也更完整。

**结语**

本文提出了一种基于时空信息自适应融合的视频显著性检测算法。综合视频序列所固有的底层空间特征和时间维度上的运动信息，采用基于区域颜色对比度的方法计算得到空间显著图，以确保显著目标的空间完整性，依据运动的连续性和方向一致性，利用连续多帧的图像的运动信息计算运动能量图和方向一致性特征图，整合为运动（时间）显著图。运动显著图能够强调在连续多帧中具有持续稳定运动的区域的显著度，有效地削弱了背景中干扰所带来的影响。最后，通过自适应加权的方式将空间显著图与运动显著图融合、优化得到视频显著图。

在 个测试视频上，将本算法与 种视频显著性检测模型进行对比，实验结果表明，本文算法在主观分析和客观指标上均取得了较好的效果，能够较好地解决复杂扰动背景情况下的视频显著性检测问题，在召回率，精度以及F值等性能指标上有所提升。本文采用光流法计算运动矢量，计算耗时较大，因此，如何提升算法的计算效率是本文后续的重点研究方向。