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显著性信息动态融合的视频显著性检测算法。该算法从两个方面对视频的显著性进行度量。首先，在空间上, 利用基于区域对比度的显著性检测算法计算区域颜色差异，提取图像中与周围相比颜色对比度较大的区域，得到每帧的空间显著图；在时间上，本文提出改进的光流法，利用运动的连续性在连续多帧图像中提取稳定鲁棒的运动矢量场，计算得到能够凸显真实运动区域的运动显著图，实现鲁棒的运动显著性度量。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。

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

人类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，图像中每一个像素的显著性程度用显著值来表示，可计算的显著性检测模型能够模仿人眼自动搜索显著区域的行为，对图像进行分析生成显著图，用以计算和衡量图像中各个位置能够吸引注意的可能性，实现对可能吸引观察者注意力的区域的位置进行预测。

现实中人眼对环境的观测是连续的过程，神经生物学研究发现, 相比图像的底层特征, 例如纹理、颜色等，人类视觉系统对场景中的运动信息更为敏感。视频是由多帧图像组成的，与单帧图像相比，帧与帧之间的强相关性使得视频图像包含更丰富的运动信息。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，也应考虑连续视频序列中运动信息对于目标显著性的影响。视频显著性检测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。因此，本文主要着重讨论视频显著性。

现有的视频显著性检测算法通常仅在相邻2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，没有突出运动信息对于目标显著性影响的重要性。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(dynamic fusion of spatial temporal information)，采用基于区域对比度的（RC）方法计算空间显著图，确保显著区域的空间完整性；依据运动的连续性，本文提出改进的光流法，在连续多帧图像中提取鲁棒的运动矢量场，利用运动矢量的幅值计算得到能够排除背景扰动的运动显著图；为了进一步提升算法性能，在运动矢量的提取过程中引入指数衰落因子，弱化较远帧对当前帧的影响，强调较近帧对当前帧的影响；时空信息融合阶段，根据人眼对不同视频的主观感知，设置融合权重，生成能兼顾动静两种情况的视频显著图。

针对目前视频显著性检测算法对环境中的干扰鲁棒性较差的问题，本文提出了一种基于改进光流的视频显著性检测方法(A video saliency detection method based on improved optical flow)IOFVSD，用于视频序列中具有明显运动和对颜色比度较强的显著区域检测。IOFVSD由运动显著性检测，空间显著性检测，显著图融合三部分组成。流程如图1。

运动显著性检测：依据运动的动态连续性，本文提出改进的光流法用于运动显著性的检测，不同于其他视频显著性检测方法在相邻帧对间进行计算，本文提出的改进光流法在连续视频帧段中提取运动向量。在提取运动向量的过程中引入指数衰落因子，弱化较远帧对当前帧的影响，强调邻域帧对当前帧的影响。改进光流法充分利用连续视频帧中的运动信息，实现视频中运动显著性的度量，进而获得能够凸显视频中运动区域运动显著图。

空间显著性检测部分，我们采用MingMingCheng提出的基于区域对比度的显著性检测方法，该方法提取视频序列中每帧图像的颜色、方向、对比度等初级视觉特征，相比于其他空间显著性检测方法，该方法能够得到均匀突出显著区域的空间显著图；

在显著性融合阶段，利用”skew-max”融合方法将空间显著图和运动显著图融合到一起，通过融合传统单帧图像的空间显著性以弥补运动显著性中色彩信号的损失，构建得到最终的视频显著性图。

通过以上分析可以看出, 现有的算法在提取运动显著性上存在一些局限. 虽然大部分算法都采用了空间和时间的显著性融合的方法[11], 但是融合的方式和权重的确定对结果的影响并没仔细分析.例如, Itti 的算法[5]和 PQFT 算法[3]都是把运动信息作为 4 个信息源通道中的一个通道来进行等权重的融合, 没有考虑运动特征和空间特征的差异; Kim 的算法[8]将时间和空间作为一个整体来考虑, 虽然不需要人为调节权重, 事实上把时间和空间的信息  
同等考虑减弱了运动分量所起的作用, 最终结果也不够理想. 针对上述问题, 根据生物学研究, 人的视觉系统对运动信息更敏感这一特点[1], 本文提出根据时间显著性的效果, 动态设置权来自适应调节空间和时间的显著性融合, 以达到合理地均  
衡两者的信息. 此外, 目前大多数视频显著性算  
法中仅使用相邻 2 帧[3]或 3 帧进行运动估计, 这对  
带有扰动背景的处理效果很差. 本文提出利用多  
帧的运动估计来代替仅使用相邻帧的估计, 利用  
显著目标在运动的方向和速度大小上具有一致性  
的特点, 能够克服在扰动背景下的时间显著性的提取  
提取