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。

人的视觉系统在感知显著目标时通常是捕捉  
一个概念上完整可定义的目标. 本文选取了目前  
在图像显著性处理中性能突出的基于图像 SLIC 超  
像素分割[14]的 DSR 算法[12]计算当前帧的空间显著  
图, 以获得空间完整的显著目标

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。

并且，空间距离在人类注意力方面也起到非常大的作用，比如，与区域r颜色对比度相同的两个区域ra和rb，距离区域r更近的ra相比于rb能够为r的显著度带来更大的贡献。

时间显著性

视频序列是由时域上的图像排列而成，其与单帧图像显著特征最大的区别便是时域上的运动显著性特征。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信息的情况下，较好地检测出独立运动的物体。本文采用光流法提取视频中的运动信息。

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，则第t帧中点的运动矢量为。

在实际视频中，明显运动的像素点通常具有较大的运动位移，本文采用运动幅值对其进行表征，点在第tt+1帧之间的运动幅值为。

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幅值

综合考虑t~t+τ之间每相邻两帧的运动幅度，可以得到图像在该连续帧段中的运动能量,具有连续运动的区域运动能量较高，而环境中存在突发扰动的区域由于幅值的平滑处理以及运动的不连续性，运动能量值较低。这样比仅考虑t与t+1帧之间的运动信息更能反映实际的运动情况。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：

其中为指数衰落因子，τ为参与计算的连续帧数

t帧的时间（运动）显著图可表示为：

，