

Research on Video Saliency Detection Via Contrast and Self-Adaptive Transfer

Wang Yongguang, Hao Aimin, Li Shuai

State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China

Email address:

478209929@qq.com (Wang Yongguang), ham_buaa@163.com (Hao Aimin), lishuaiouc@126.com (Li Shuai)

To cite this article:

Wang Yongguang, Hao Aimin, Li Shuai. Research on Video Saliency Detection Via Contrast and Self-Adaptive Transfer. *Science Discovery*. Vol. 5, No. 2, 2017, pp. 100-107. doi: 10.11648/j.sd.20170502.14

Received: March 19, 2017; **Accepted:** April 7, 2017; **Published:** April 20, 2017

Abstract: Although a lot of studies in salient motion detection have achieved great success in recent years, many challenges still exist toward the video saliency detection over the non-stationary videos and videos with slowly-moving objects, which supposes to exhibit significant influence on its corresponding subsequent applications. Thus, it urgently needs a more robust, stable, and precise method to solve the above mentioned limitations. In fact, inspired from the basic visualization rule of the human vision system, the human's attention can be easily attracted by two independent factors: the motion saliency clue and the color saliency clue. Hence, this paper develops a novel salient motion detection method by fusing the motion saliency with the color saliency, which refines the preliminary saliency map by self-adaptive transfer via the newly designed intra-frame correlation. Also, comprehensive experimental results of our method toward the state-of-the-art methods over 4 public available benchmarks demonstrate the superiority of our method both in its robustness and high detection precision.

Keywords: Saliency Detection, Contrast, Self-Adaptive, Saliency-Transfer

基于对比与自适应传递的视频显著性检测的研究

王永光, 郝爱民, 李帅

虚拟现实技术与系统国家重点实验室, 北京航空航天大学, 北京, 中国

邮箱

478209929@qq.com (王永光), ham_buaa@163.com (郝爱民), lishuaiouc@126.com (李帅)

摘要: 近年相关领域关于视频显著性检测已经进行了大量的研究, 尽管已经取得一定的成果, 但目前对于其实际应用仍然存在一些传统方法无法克服的局限性, 如对非固定相机视频的运动物体检测、对运动缓慢的物体检测等, 这些局限性将会对后续应用造成重大不利影响, 因此当前急需一种更为鲁棒、更稳定、更准确的显著运动检测方法。根据对人类的视觉成像系统工作原理的机器视觉理解, 即显著运动的物体主要从两方面吸引人类视觉系统: 运动显著性以及颜色显著性, 提出一种结合光流速度场和颜色场的显著运动物体检测方法。首先根据运动场和颜色场结合得到初步的显著值分布图, 并在此基础上根据帧间相关性采用自适应传递对初步的结果进行修正, 从而得到最终的结果。并且, 结合在4个公共数据集上的定量分析结果表明, 本文提出的方法较传统方法能够获得更稳定、精确的视频显著性检测结果。

关键字: 显著性检测, 对比, 自适应, 传递

1. 引言

给定一个视频序列, 视频显著运动检测是将视频序列中的运动物体标记出来, 这些信息将会输入到后续应用(如视频分割[1]、视频物体追踪识别[2]、视频理解[3]等)中, 以形成一个完整的视频应用链。近些年来计算机视觉发展迅速, 作为计算机视觉中最活跃的研究方向之一, 视频显著物体检测也取得了很大的成果, 可以看到现在很多的方法在一些数据集中能够取得一个不错的效果, 对物体的检测也具有相当的精度, 如基于局部建模的方法[4, 5], 以及基于低秩分析的方法[6, 7], 但这些方法都有其各自的局限性, 比如对动态背景、运动缓慢以及物体较大等情况的检测效果不理想。

近些年, 在单帧静态图像上显著性检测算法的研究有很多: Itti[20]通过对多尺度图像显著性特征进行融合, 并将最终结果作为最终的显著性检测结果; Hou[21]提出基于傅里叶谱残差的显著性检测方法; Achanta[22]通过对颜色与亮度等信息的融合得到显著值分布图。经过多年的实际研究, 人们发现衡量单帧图像显著性最重要的方式是进行局部对比, 例如颜色对比、形状对比[8], 在一张图像中, 最显著的前景物体通常较其周围背景部分有着较高的对比度, 而非显著性背景区域则与其周围差异较小。近年, 有很多基于对比的显著性检测方法被相继提出, 如cheng[23]提出的基于区域对比度的显著性检测方法, 类似的, Perazzi[24]将图像进行超像素分割, 并通过对比超像素的颜色独特性以及其在颜色特征空间的分布差异性来得到显著性图。尽管这些基于局部对比度的方法对于一般的自然场景的显著性物体检测有着较为理想的检测结果, 然而, 传统的局限性仍然无法克服, 即由于局部对比度容易对大体积物体的中心区域造成显著性漏检, 并最终导致空洞现象。因此, 如何进行对比度的计算则是显著性检测的一个重要的问题。

根据人类的视觉成像系统, 人眼对于运动物体的敏感度较单一图像的颜色显著性更高。因此, 对单帧图像的显著性检测与视频显著性检测会存在一定的不同, 即视频显著性检测算法需要更加关注运动上的信息。相对于单帧图像颜色的对比, 运动的差异则是衡量视频显著性的一个重要的因素, 并且运动的对比以及颜色对比得到的显著性往往能够互相弥补, 使得结果更加准确。但是, 简单的将这两个显著图进行融合很难得到一个鲁棒的结果, 并且由于对比计算方法的问题, 当颜色显著图以及运动显著图都不准确时, 最终的结果也会出现问题。

综上所述, 本文提出了一种基于运动-颜色梯度图的对比度的计算方法, 该方法能够很好的对前景物体以及其周围的超像素点赋予较高的显著值, 并且, 本文还新提出一种时空平滑方法, 使得最终输出的视频显著性检测结果具有较好的时空连续性, 从而进一步的提高了视频显著性检测的精度。与此同时, 本文采用自适应传递的方式对前面得到的初步时空平滑结果进行进一步的修正, 大幅度的提升了视频显著性检测的精度。

2. 方法简述

如图1所示, 本文采用的方法主要有以下几步: 1) 通过运动-颜色梯度图来指导对比度的计算, 对光流场以及颜色场分别进行局部差异度对比得到颜色显著值分布图和运动显著值分布图, 并通过融合的方式作为初始视频显著性信息; 2) 通过对前景区域的颜色特征描述进行低秩分析, 并以此从视频段中选取一部分的视频帧作为可靠帧; 3) 以可靠帧为基点, 以自适应传递的方式对初步显著性结果(步骤1)进行修正, 并进行时空平滑得到最终结果。

本文提出的方法不同于传统的视频显著性检测方法, 即传统的方法通常采用逐帧或者整体的方式来计算显著值, 但这样一来, 由于缺乏对错误检测结果的抑制机制, 某些由外界干扰(如相机抖动, 光照变化等)造成的瞬时检测误差融合对剩余的检测结果造成错误的累积, 并最终造成错误的检测结果。因此, 为了避免错误累积这一问题, 本文方法将整个视频序列等分划分为多个短的帧段, 然后分别对其进行视频显著性检测。

另一方面, 为了减少视频显著性检测的计算负担, 本方法对所有待检测的视频帧进行边界敏感的平滑方法[9]来消除一些不必要的信息, 并使用简单线性迭代聚类(SLIC Simple Linear Iterative Clustering)[10]方法对其进行超像素分割, 即采用超像素点作为基本的显著性对比单元, 以此来代替原本的像素级的显著性对比计算, 大幅度的降低计算代价。与此同时, 在计算对比度时提出运动-颜色梯度图(MCG Motion-Color Gradient)来控制显著对比的计算范围, 使高的显著值能够分配到前景物体上。然后采用时空平滑策略对得到的两张显著图分别进行平滑, 通过图2可以看到平滑前后两种显著图的区别, 效果的改进非常明显。在之前的基础上进一步提出基于低秩分析的初步结果可靠度计算方法, 并基于可靠度高的帧对视频前后进行传递修正。与传统方法比较, 本文在精确度和连续性上都能取得一个好的结果。

3. 方法详述

3.1. 基于对比的初步显著信息

本文采用基于对比的显著性计算, 不同于传统的采用逐帧或整个视频序列方式的视频显著性检测方法[11,12], 本文采用逐帧段的方式来对视频显著性进行检测, 将视频序列等分为若干个小序列 $\mathbf{B}_i = \{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_m\}$, 其中 \mathbf{I}_m 代表在第 i 个帧段中第 m 帧图像。

给定一个视频序列 $\mathbf{S} = \{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_n\}$, 显著运动检测是找到在每一帧中的显著物体, 即前景物体。在本文第一步需要得到的是每一帧初步的显著值, 在这里使用两种方法计算初步的显著值信息, 基于光流方法使用相邻两帧图像得到的运动场, 以及根据单帧图像中颜色分布的颜色场。

本文采用2011年C. Liu[13]所提供的光流代码, 根据这个方法能得到每一帧图像的运动场, 即在每一帧图像像素点的运动速度和方向, 采用RGB颜色得到颜色场。

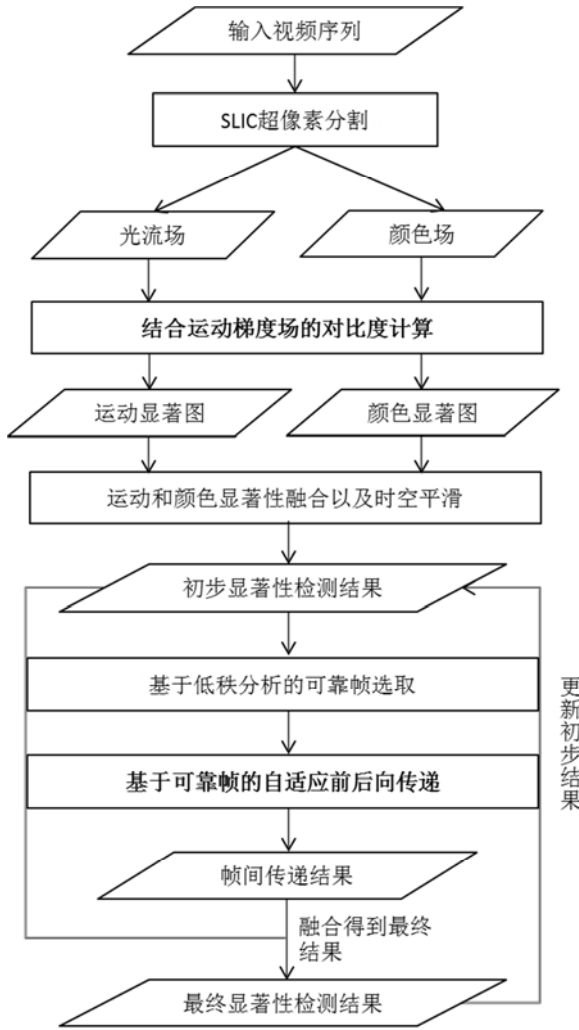


图1 方法流程图。

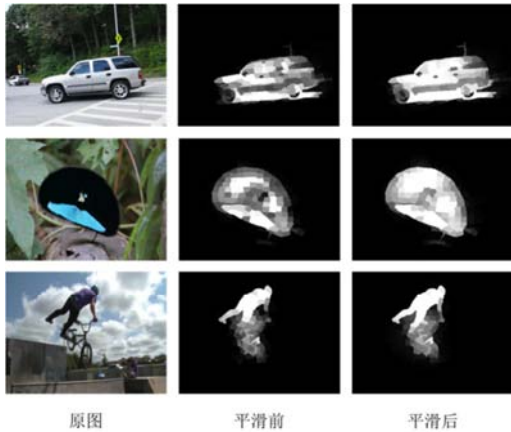


图2 平滑效果对比。

不同于传统的方法直接计算对比得到显著信息,根据[1, 14]中提到的运动梯度图比单纯的运动对比计算更具有鲁棒性,本文引入运动-颜色梯度图(MCG)来对计算对比的范围进行约束,使得显著值能够更多地分配到显著物体的周围。第*i*个超像素的运动显著计算如公式(1):

$$\mathbf{MS}_i = \sum_{\mathbf{p}_j \in R_i} \frac{\|\mathbf{v}_i \mathbf{v}_j\|_1}{\|\mathbf{p}_i \mathbf{p}_j\|_1} \quad s. t. R_i = \min_{0 < j < n} \|\mathbf{p}_i, \mathbf{MG}_j\|_2 \quad (1)$$

$\|\cdot\|_1$ 在这里代表的是L1范式, $\|\cdot\|_2$ 代表L2范式, \mathbf{p}_i 表示第*i*个超像素的位置中心点, \mathbf{V} 表示两个方向上的光流速度, n 代表在这一帧中所有超像素的数量, R_i 表示计算对比的范围,它由第*i*个超像素于运动-颜色梯度图MCG的最短欧式距离计算得到,而运动-颜色梯度图的计算方式如公式(2):

$$\mathbf{MCG} = \|\nabla \mathbf{I}\|_2 \odot \|\mathbf{v}_x, \mathbf{v}_y\|_2 \quad (2)$$

\odot 表示逐元素相乘的阿达玛积(Hadamard Product), $\nabla \mathbf{I}$ 代表颜色梯度图, $\|\mathbf{v}_x, \mathbf{v}_y\|_2$ 表示运动梯度图。同样,颜色显著值分布图的计算如公式(3)

$$\mathbf{CS}_i = \sum_{\mathbf{p}_j \in R_i} \frac{\|\mathbf{c}_i \mathbf{c}_j\|_2}{\|\mathbf{p}_i \mathbf{p}_j\|_1} \quad (3)$$

上式 \mathbf{C} 表示RGB颜色, $\|\cdot\|_2$ 代表L2范式。

3.2. 显著信息平滑与融合

尽管本文使用了梯度信息对对比的计算范围进行了约束,但是由于有的物体过大,处于物体中心的区域由于其周围的区域相似度较高,使其被误认为背景,导致这一部分超像素的显著值较低,出现明显错误。并且,由于本文采用的方法是基于传递的,因此,一旦有一帧有较大的错误,那么即使周围的帧都是对的,该错误的结果也会对其余视频帧造成误差的累积,使得最终的检测结果出现错误。因此,需要一个对错误检测结果进行一定修正的机制来弥补这一缺陷,即在得到运动显著值分布图(MS)和颜色显著值分布图(CS)之后,本文提出一种对这两个显著图进行一个修正平滑的方法,使得它们的结果能够更加的鲁棒,如公式(4)(5)(6)所示:

$$\mathbf{S}_{t,i} = \mathbf{S}_s / \mathbf{W}_s \quad (4)$$

$$\mathbf{S}_s = \sum_{k=t}^{t+2} \sum_{j \in \varphi_i} \mathbf{S}_{k,j} \times \exp(-\|\mathbf{c}_{t,i}, \mathbf{c}_{k-1,j}\|_2 \times \alpha) \quad s. t. \quad (5)$$

$$\varphi_i = \{R_i < \|\mathbf{p}_{t,i}, \mathbf{p}_{k,j}\|_1 < \beta\}$$

$$\mathbf{W}_s = \sum_{k=t}^{t+2} \sum_{j \in \varphi_i} \exp(-\|\mathbf{c}_{t,i}, \mathbf{c}_{k-1,j}\|_2 \times \alpha) \quad s. t. \quad (6)$$

$$\varphi_i = \{R_i < \|\mathbf{p}_{t,i}, \mathbf{p}_{k,j}\|_1 < \beta\}$$

在这里 t, i 表示要求的超像素的下标,即第*t*帧第*i*个超像素。 \mathbf{S} 为显著值, \mathbf{C} 表示超像素RGB颜色, β 为平滑范围上限(其具体值通过试验确定,即 $\beta = 100$), R_i 为公式(1)中的范围, $\alpha = 10$ 作为平滑的强度, $\mathbf{p}_{t,i}$ 是第*t*帧第*i*个超像素的空间位置,公式中超像素*j*的空间位置距离待平滑的超像素*i*大于 R_i 小于 β 。

从上面的公式可以看到, 本文不仅使用了空间范围的平滑, 还使用时间范围的平滑。对一个超像素 SP_i , 使用它周围的超像素对它进行平滑, 并且根据前后两帧在同一个位置的范围内其他超像素对它进行平滑, 这样的做法能使平滑之后的显著图不仅消除物体中间空洞的问题, 还能使前后帧能保持良好的连续性。

另一方面, 单纯的运动显著分布图或者颜色显著图很难得到一个良好的检测结果。同样, 由于视频序列的多样性, 对于一些类型的视频序列, 它们表现出良好的检测结果, 但是对于其它的视频序列, 它们的检测结果会出现一些比较严重的错误, 而运动显著图与颜色显著图能够在效果上互补, 为了得到更加鲁棒的结果, 在最后会将颜色显著值分布图与运动显著值分布图进行融合, 得到显著值的初步结果 PS , 如公式 (7):

$$PS = MS \odot CS \quad (7)$$

\odot 表示逐元素相乘, 从图3可以看出, 经过融合后的效果比单独的颜色或者运动显著值分布图效果会好很多。

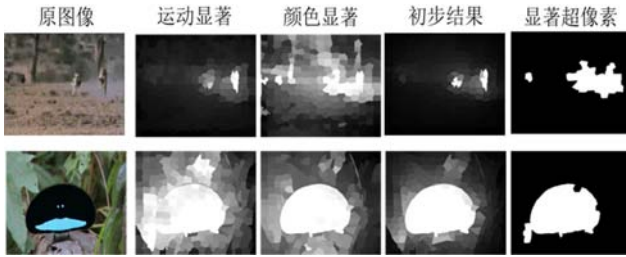


图3 初步结果以及相应显著超像素的计算及结果。

3.3. 显著超像素的选取

正如前面所提到的, 本方法所有的计算单位都是以超像素为单位, 在得到初步的显著值分布之后, 接下来需要找出表现为前景的超像素。考虑到由于有些视频比较长, 前景物体有可能会在大小上有比较大的变化, 根据之前进行的分段处理方式, 在每个帧段计算一个阈值 γ , 把满足公式 (8) 的超像素作为显著超像素 SSP (salient super-pixel):

$$PS_{t,i} > \text{mean}(PS_t) \times \gamma \text{ s.t.}$$

$$\gamma = (0.25 \times SpNum) / \text{MeanSPLength} \quad (8)$$

$SpNum$ 为第 t 帧超像素个数, MeanSPLength 为显著超像素的平均值, $PS_{t,i}$ 为第 t 帧第 i 个超像素的显著值, 初始 $\gamma = 2$ 。在这里首先计算第一帧的显著超像素, 以此来估算前景物体的大小, 并进一步更新 γ 使得更准确的将前景物体包含在显著超像素中。

大多数情况下, 光流的计算结果都是相对正确的, 但是在某些时候 (如相机剧烈的抖动或物体快速的发生移动) 会存在一些计算误差, 如图4所示。而颜色对比的计算同样如此, 特别在前景物体与背景颜色相差很小的时候, 颜色显著值分布图就会出现比较大的错误, 导致会有融合后的显著值分布图出现较大的错误, 所以接下来就是对这些有错误的分布图进行修正。



原始图像

光流结果

图4 光流结果示意, 右上与右下分别为错误与正确的结果。

3.4. 可靠帧的选取

在之前的步骤中, 通过对显著超像素的选取, 将前景物体包含在显著超像素中, 接下来对这些显著超像素进行处理。对于一个视频帧序列, 其前景物体在整个视频序列中一般是不变的, 所以它们在颜色空间的变化相对来说较小, 可以认为它们的前景部分颜色分布在整个视频序列具有低秩的特性, 对其进行低秩分析, 低秩部分的颜色分布就是前景物体的颜色分布。对于在一段视频帧序列中稀疏度较低的视频帧, 可以认为这一帧的初步结果相对于它周围的帧具有更高的准确度, 所以在这一步需要找出在这些帧当中可靠的帧作为后面处理的基础。

首先对前面计算得出的显著超像素进行颜色直方图统计, 得到每一帧显著超像素的颜色分布情况。由于本文使用 RGB 3通道对颜色进行表示, 所以需要将其进行投影转换到1维空间中以便后续的处理。其具体做法为: 将 RGB 颜色0-255分为10个区间, 将每一帧中的每一个超像素的颜色值投影到一个 $10 \times 10 \times 10$ 的三维颜色空间中, 统计每一个点的颜色分布情况。由于 $10 \times 10 \times 10$ 相对于使用的超像素个数来说较大, 为了防止颜色分布很稀疏的情况, 使用类似高斯球模型对这样一个三维空间进行扩散, 以避免在低秩分析时出现所有的颜色都是稀疏的情况, 如公式 (9) 所示:

$$R_{i,j,k} = \sum_{\text{pos}=px}^{ps} R_{\text{pos}} \times \text{pow}(2, -d) \text{ s.t.}$$

$$d = \|\text{pos}, (i, j, k)\|_1,$$

$$ps = (i + 2, j + 2, k + 2),$$

$$px = (i - 2, j - 2, k - 2) \quad (9)$$

在这里 i, j, k 为三维空间下标, pos 为三维空间位置, 表示坐标为 i, j, k 的值 $R_{i,j,k}$ 的取值为以它为中心的一个小正方体上各点的加权和, 这个小正方体边长为4, 权重为 $\text{pow}(2, -d)$, 即2的 d 次方, d 为各点与坐标为 i, j, k 的L1距离。

在得到三维颜色空间统计之后, 把它转换为1000的向量来代表一帧的颜色分布, 在得到所有帧的颜色分布之后, 将其列为一个矩阵, 对这个矩阵进行低秩分析, 如公式(10)所示. 本文采用的是RPCA (Robust Principal Component Analysis) 方法进行求解, 可以得到一个稀疏矩阵以及一个低秩的矩阵. 根据这个得到的稀疏矩阵, 对每一帧求其累计偏差 $\mathbf{E_Sum}_i$, 找到在每一段中最小的一帧, 将其作为一个可靠帧作为下一步的基础。

$$\mathbf{M} = \mathbf{E} + \mathbf{A} \quad (10)$$

\mathbf{E} 是稀疏部分, \mathbf{A} 是低秩部分, $\mathbf{E_Sum}_i$ 代表第 i 帧整体的残差。

3.5. 传递

传递以及传递过程中自适应的更新是本文的重点。在完成上面过程之后, 已有的数据包括每一帧的显著超像素以及它与对应的约束节点, 所有帧的显著值分布图, 相对可靠的一些帧。而在初步结果中, 必定有好的结果以及坏的结果, 那么用结果好的帧对坏的结果的帧进行修正, 不断循环迭代, 就可以得到正确的结果。现在需要做的就是如何利用可靠帧进行扩散修正初步结果错误的帧。

对第 i 帧第 k 个显著超像素 $\mathbf{SP}_{i,k}$, 第 $i+1$ 帧为待传递帧。在第 $i+1$ 帧的 $\mathbf{P}_{i,k}$ (超像素 $\mathbf{SP}_{i,k}$ 的位置) 附近 Dis 范围之内找到所有的候选超像素 $\{\mathbf{SP}_{i+1,j}, \dots, \mathbf{SP}_{i+1,j+n}\}$, 并且根据这些超像素为其寻找附近的可能约束节点, 其中 Dis 根据2.3节显著超像素的大小预设的值。以候选超像素 $\mathbf{SP}_{i+1,j}$ 为例, 在 $\mathbf{SP}_{i+1,j}$ 的空间位置 $\mathbf{P}_{i+1,j}$ 附近寻找所有与它的距离小于 $Dis \times 2$ 的超像素作为其可能的约束节点, 如此便得到一个 $m \times (n+1)$ 的矩阵, m 代表第 $i+1$ 帧中与 $\mathbf{P}_{i,k}$ 距离小于 Dis 的候选超像素的数量, n 代表第 $i+1$ 帧中所有距离候选超像素 $\{\mathbf{SP}_{i+1,j}, \dots, \mathbf{SP}_{i+1,j+n}\}$ 的距离小于 $Dis \times 2$ 的超像素的数量, 根据这些信息就可以计算出每一个候选超像素与传递超像素的一个相似程度。矩阵的每一行代表一个候选超像素以及它可能的约束节点, 接下来需要确定哪些超像素是约束节点, 以 $\mathbf{SP}_{i+1,j}$ 为例, 分别对第2列到第 $n+1$ 列求与 $\mathbf{SP}_{i,k}$ 的三个约束节点 $\mathbf{SPP}_{i,k}$ 的颜色距离, 对每一个约束节点找出颜色距离最小的超像素, 认为此超像素为 $\mathbf{SP}_{i+1,j}$ 的约束节点, 如此可以得到 $\mathbf{SP}_{i+1,j}$ 以及它的三个约束节点, 最终得到一个 $m \times 4$ 的矩阵, 第一列代表 $\mathbf{SP}_{i,k}$ 与 $\{\mathbf{SP}_{i+1,j}, \dots, \mathbf{SP}_{i+1,j+n}\}$ 的颜色距离, 第二、三、四列表示 $\mathbf{SPP}_{i,k}$ 与约束节点的颜色距离。用这四列作为计算传递的权重的一部分, 如公式(11) (12) (13) 所示:

$$\mathbf{W}_{i,j} = \mathbf{W}_m + \mathbf{W}_p \quad (11)$$

$$\mathbf{W}_m = 3 \times \left\| \mathbf{RGB}_{\mathbf{SP}_{i,k}} - \mathbf{RGB}_{\mathbf{SP}_{i+1,j}} \right\|_2 \quad (12)$$

$$\mathbf{W}_p = \sum_{p=1}^3 \left\| \mathbf{RGB}_{\mathbf{SPP}_{i,k,p}} - \mathbf{RGB}_{\mathbf{SPP}_{i+1,j,p}} \right\|_2 \quad (13)$$

在得到相似度之后, 利用相似程度来传递第 i 帧的显著值到第 $i+1$ 帧中, 传递方式如公式(14) (15) (16) 所示:

$$\mathbf{NSM}_{i+1,j} = \mathbf{SSUM} / \mathbf{WSUM} \quad (14)$$

$$\mathbf{SSUM} = \sum_{k=0}^n (\mathbf{SM}_{i,k} \times \exp(-\mathbf{W}_{i,k})) \quad (15)$$

$$\mathbf{WSUM} = \sum_{k=1}^n \exp(-\mathbf{W}_{i,k}) \quad (16)$$

其中 j 满足 $\|\mathbf{P}_{i+1,j}, \mathbf{P}_{i,k}\|_1 < Dis$, $\mathbf{SM}_{i,k}$ 是第 i 帧第 k 个超像素的显著值, 并且 k 为第 i 帧的所有满足条件的超像素。公式(15)表示在第 i 帧中的所有距离第 $i+1$ 帧中目标超像素距离在一定范围内的超像素都会向其传递显著值, 但是最后会除以一个最终的权重作为平均。对所有的显著超像素传递之后便得到一个新的第 $i+1$ 帧的显著值分布。

为了得到更好更鲁棒的结果, 本文将传递分为两个过程, 对于每一个可靠图像帧, 我们会使其向左以及向右传递两次, 并且每一个可靠帧的传递范围仅限于它的上一个可靠帧以及下一个可靠帧之间的所有图像帧。对于每一个图像帧, 它都会有两个方向上传递而来的传递结果, 由于这两个结果都具有相同的可靠度, 所以对其进行求平均得到结果。如图5所示, 可以看到传递的结果比传递之前的结果更为准确。



图5 传递效果对比。

3.6. 更新显著值分布图

对于从可靠帧传递而来的新的显著值分布图, 我们可以相信它的正确性, 但是很多时候由于相邻帧物体运动的问题使得它还是会有部分瑕疵。并且对于每一个帧段只取一个可靠帧, 那么就是说同样一个帧段, 不与初步结果进行融合, 在传递时候很容易出现错误累积的情况, 特别是在远离可靠帧的那些视频帧, 会出现比较大的错误。所以在得到新的显著值分布之后, 就需要对两个显著值分布图进行一个融合, 而融合方法的选取会很大程度上影响到最终结果的好坏, 重点是如何合理的分配能使原显著值分布图中错误的被更正, 正确的却不会被传递成错误的。在这里本文使用了区分的思想, 如公式(17)所示:

$$\mathbf{SM}_{i,k} = \begin{cases} \mathbf{SM}_{i,k} \times 0.1, & \mathbf{NSM}_{i,k} = 0 \\ \mathbf{SM}_{i,k} \times 0.5, & \overline{\mathbf{NSM}_{i,k}} < \mathbf{NSM}_i \times \gamma \\ (1 - \mathbf{E_sum}_i) \times \mathbf{SM}_i + \mathbf{NSM}_i, & \text{else.} \end{cases} \quad (17)$$

对于在传递的结果 \mathbf{NSM}_i 中为0, 即传递中认为是背景的部分, 所以有理由认为它是前景的可能性较低, 所以对其显著性进行降低, 对于小于其平均值的 γ 倍的部分 (γ 为2.3节中计算显著超像素时的阈值), 其为背景的可能性也比较大, 对其进行不同幅度的降低, 剩余其他部分则为前景的可能性大。在传递之前的初步结果中对其进行低秩分析会得到一个帧的可靠性衡量, 即 $\mathbf{E_Sum}_i$ 。当 $\mathbf{E_Sum}_i$ 较高的时候, 代表第 i 帧的初步结果可靠度较低,

那么此时我们就要将最终结果更多的由传递结果来决定,而达到一个好的结果。同时在这里,为了防止 E_Sum_i 过高或者过低使得结果过于偏向于传递结果或者初步结果的情况,我们需要将 E_Sum_i 归一化到0.4-0.8之间。结果如图6所示。

4. 实验

为了衡量本文所采取方法的效果,本文在4个公共的benchmark上进行测试,这4个benchmark为SegTrack v1 [15], SegTrack v2 [16], BMS [17], DS [18] 数据集。它们包含了各种现实生活中可能出现的情形。其中SegTrackv1数据集包含了6段视频序列,它们的特点是物体运动速度快,并且伴有较为复杂的背景。而SegTrack v2数据集则有10段视频,有适当的运动速度,背景有静态以及非静态的。BMS共有26段视频,它们的运动各有特色,包含多种情况。DS有10段视频,里面的物体运动速度很慢,并且背景也是动态的,检测的难度也会有所提升。在这4个benchmark上面进行方法的测试能够体现出我们方法的优越性以及先进性。

为了与其他先进的显著性检测方法进行对比,本文选取了国际上这方面5个先进的方法,包括SA15 [12], GF15 [14], ST14 [3], MC15 [19], SU14 [11]。为了

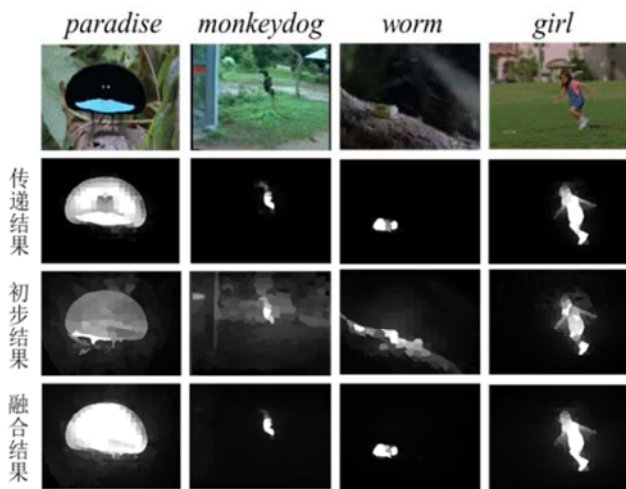


图6 融合效果对比。

更好的体现本文方法的优点,本文采用了precision-recall (PR)作为评估的标准,这是一种公认的评估标准。并且本文将视频显著性的检测结果用阈值 $T[1, 255]$ 分割(更好的观测结果),那些区域大于阈值 T 的作为前景物体,然后与ground truth进行比较,以得到一个precision-recall,这个结果能够很好地表现结果的优劣。本文的方法结果与其他方法进行的一个比较,如图7、8、9、10、11所示。

从图中的结果可以看到,我们的方法对于显著运动物体的检测精度较高,对于四个数据集都能够得到一个好的结果。

5. 结论

本文首先介绍了视频显著检测的一些通用方法以及相应的不足,并据此提出基于光流运动场以及颜色场的初步显著结果计算方法。并以此为基础提出一种基于自适应传递的视频显著性精度提高算法。通过实验仿真结果,可以明显看出本文提出的算法在精确度上相对于已有的算法有着大幅度提升。

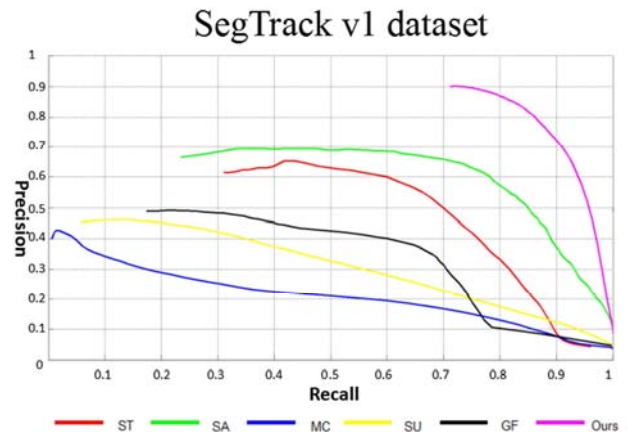


图7 SegTrack v1 PR曲线显示。

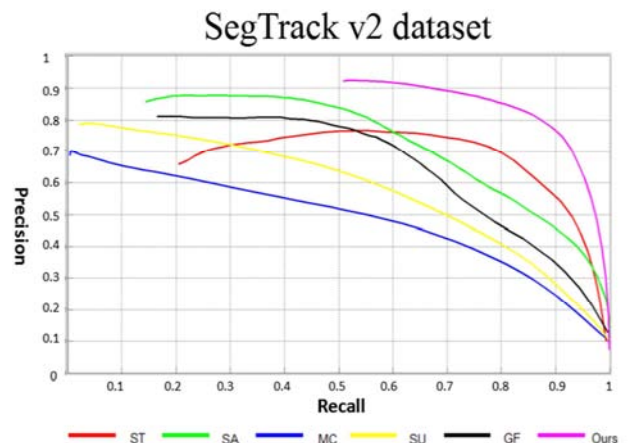


图8 SegTrack v2 PR曲线显示。

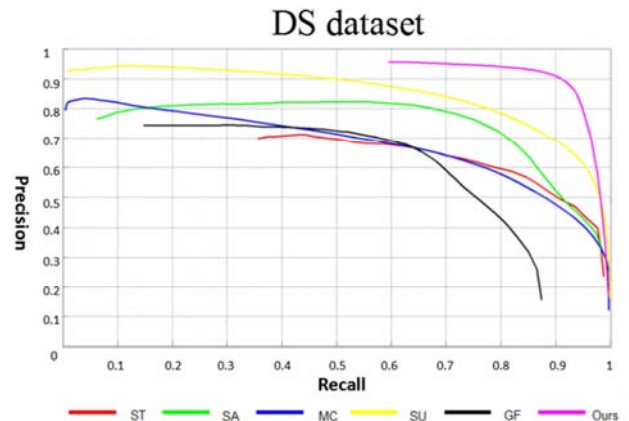


图9 DS PR曲线显示。

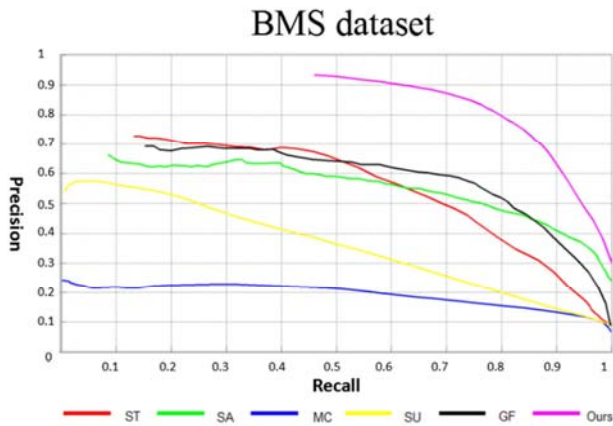


图10 BMS PR曲线显示。

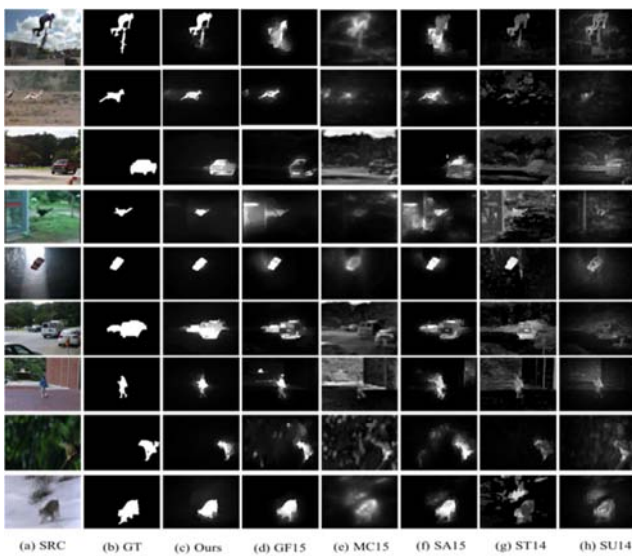


图11 结果对比。

致谢

本文研究工作是在教育部博士点基金新教师项目(项目编号: 20131102120016)的资助下完成的。

参考文献

- [1] Zhang D, Javed O, Shah M. Video object segmentation through spatially accurate and temporally dense extraction of primary object regions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 628-635.
- [2] Chen C, Li S, Qin H, et al. Real-time and robust object tracking in video via low-rank coherency analysis in feature space [J]. Pattern Recognition, 2015, 48 (9): 2885-2905.
- [3] Zhou F, Bing Kang S, Cohen M. F. Time-mapping using space-time saliency [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3358-3365.
- [4] Liang D, Kaneko S, Hashimoto M, et al. Robust object detection in severe imaging conditions using co-occurrence background model [J]. International Journal of Optomechatronics, 2014, 8 (1): 14-29.
- [5] St-Charles P. L, Bilodeau G. A, Bergevin R. Subsense: A universal change detection method with local adaptive sensitivity [J]. IEEE Transactions on Image Processing, 2015, 24 (1): 359-373.
- [6] Chen C, Li S, Qin H, et al. Robust salient motion detection in non-stationary videos via novel integrated strategies of spatio-temporal coherency clues and low-rank analysis [J]. Pattern Recognition, 2016, 52: 410-432.
- [7] Gao Z, Cheong L. F, Wang Y. X. Block-sparse RPCA for salient motion detection [J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36 (10): 1975-1987.
- [8] Cheng M. M, Mitra N. J, Huang X, et al. Global contrast based salient region detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (3): 569-582.
- [9] Gastal E. S. L, Oliveira M. M. Domain transform for edge-aware image and video processing [C]// ACM Transactions on Graphics (TOG). ACM, 2011, 30 (4): 69.
- [10] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 34 (11): 2274-2282.
- [11] Fang Y, Wang Z, Lin W, et al. Video saliency incorporating spatiotemporal cues and uncertainty weighting [J]. IEEE Transactions on Image Processing, 2014, 23 (9): 3910-3921.
- [12] Wang W, Shen J, Porikli F. Saliency-aware geodesic video object segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3395-3402.
- [13] Liu C. Beyond pixels: exploring new representations and applications for motion analysis [D]. Massachusetts Institute of Technology, 2009.
- [14] Wang W, Shen J, Shao L. Consistent video saliency using local gradient flow optimization and global refinement [J]. IEEE Transactions on Image Processing, 2015, 24 (11): 4185-4196.
- [15] Tsai D, Flagg M, Nakazawa A, et al. Motion coherent tracking using multi-label MRF optimization [J]. International journal of computer vision, 2012, 100 (2): 190-202.

- [16] Li F, Kim T, Humayun A, et al. Video segmentation by tracking many figure-ground segments [C]// Proceedings of the IEEE International Conference on Computer Vision. 2013: 2192-2199.
- [17] Brox T, Malik J. Object segmentation by long term analysis of point trajectories [C]// European conference on computer vision. Springer Berlin Heidelberg, 2010: 282-295.
- [18] Fukuchi K, Miyazato K, Kimura A, et al. Saliency-based video segmentation with graph cuts and sequentially updated priors [C]// IEEE International Conference on Multimedia and Expo. IEEE, 2009: 638-641.
- [19] Kim H, Kim Y, Sim J. Y, et al. Spatiotemporal saliency detection for video sequences based on random walk with restart [J]. IEEE Transactions on Image Processing, 2015, 24 (8): 2552-2564.
- [20] Itti L, Koch C, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20 (11): 1254-1259.
- [21] Hou X, Zhang L. Saliency detection: A spectral residual approach [C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.
- [22] Achanta R, Estrada F, Wils P, et al. Salient region detection and segmentation [C]// International conference on computer vision systems. Springer Berlin Heidelberg, 2008: 66-75.
- [23] Cheng M. M, Mitra N. J, Huang X, et al. Global contrast based salient region detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (3): 569-582.
- [24] Perazzi F, Krähenbühl P, Pritch Y, et al. Saliency filters: Contrast based filtering for salient region detection [C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 733-740.