

基于显著区域的台标检测与识别*

汪文英^{1,2,3}, 李锦涛^{1,2}, 张勇东^{1,2}, 张冬明^{1,2}, 徐杰^{1,2,3}

(1. 中国科学院计算技术研究所 虚拟现实技术实验室, 北京 100190)

(2. 中国科学院计算技术研究所 智能信息处理重点实验室, 北京 100190)

(3. 中国科学院研究生院 北京 100190)

摘要: 台标是电视台的标志, 它已成为基于内容的多媒体检索中重要的语义信息之一。台标检测中帧差法是常用的方法, 但在一些实时系统中无法取得多帧信息以致帧差法不适用。为此本文设计了一种在单帧图像上提取显著区域并进行台标检测与识别的算法, 首先改进了 Scale Saliency 方法, 利用相对熵度量显著性, 提出参考上下文的显著区域检测算法, 它在图像中提取的显著区域能够较准确地包含台标区域。然后, 通过计算显著区域与模板之间的 Hausdorff 距离的方法实现台标的检测与识别, 这种算法具有不需要进行训练和能够较好的排除背景干扰的优点。

关键词: 台标检测; 视频分析; Kullback-Leibler; Hausdorff 距离

Detection and Recognition of TV Logo Based on Salient Region

Wenyang Wang^{1,2,3}, Jintao Li^{1,2}, Yongdong Zhang^{1,2}, Dong ming Zhang^{1,2}, Jie Xu^{1,2,3}

(1. Virtual Reality Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190)

(2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190)

(3. Graduate School of the Chinese Academy of Sciences, Beijing 100049)

Abstract: TV Logo as the sign of TV station has become important information for Content Based Information Retrieval. The conventional method of detection of TV Logo is using the difference between multi frames, however this method can't work when the background is still, and in some real time systems only single frame can be available. We designed a method for detection and recognition TV Logo based on single frame of video. First, we redefined the saliency of a region, and proposed a Context-dependent Scale Saliency detection method, which use relative entropy between local region and context around as the measurement of saliency. This method can detect the region of Logo more precisely. Second, we implement a method of recognition of TV Logo by using Hausdorff distance. This method need not train and can exclude the impact of background.

key words: TV Logo Recognition; Video Analysis; Kullback-Leibler; Hausdorff Distance

1 引言

台标作为电视台的标志是基于内容的多媒体检索中重要的语义信息之一, 同时台标检测也是学术界和商业界近年来的研究热点之一。[1,2,3,4,5,6]利用了连续帧帧差的办法来提取台标的区域进行识别, 算法中都假设视频中除了台标区域静止不动外都会不断变化。这些算法都利用到了多帧信息, 而后采用帧差的办法求得台标区域。

基金资助: 973 项目资助, 项目编号 2007CB311100

863 项目资助, 项目编号 2007AA01Z416

作者简介: 汪文英(1982-)男, 安徽桐城人, 博士研究生, Email: wangwenying@ict.ac.cn

然而帧差法在遇到静止背景、动画台标以及在一些实时系统中很难得到多帧信息等情况下就不适用了。这就提出了如何在单帧视频图像上进行台标识别的问题。本文设计了在单帧视频图像中提取显著区域，并计算显著区域中的边缘与台标模板中的边缘之间 Hausdorff 距离的方法进行台标检测与识别。

显著区域算法是用来在图像中寻找特征区域的算法，它能够克服图像分割的弱点提取特征信息进行对象识别。效果比较好的显著区域的算法有 Kadir and Brady[7]提出的 Scale Saliency 算法，这种方法是针对 Gilles [8]算法的改进。然而 Scale Saliency 算法在本文台标识别的应用中仍显不足，比如没有考虑到上下文对显著性的影响，实验中会发现使用 Scale Saliency 圈定台标区域往往将背景区域也包含进显著区域，或缠绕在显著区域的边缘上。针对这一点本文利用相对熵提出了新的显著区域提取算法，能够较好圈定台标区域。

台标识别的算法比较多，比如不变矩，支持向量机等算法。然而由于台标区域一般都包含比较多的背景噪音，且一般很难将台标与背景噪音完全分割开来。所以本文采用了只利用边缘信息进行识别的办法，将显著区域的边缘归一化后计算与模板之间 Hausdorff 距离的方法，能够取得比较好的识别效果，这种方法计算复杂度比较低，与模板的大小(也就是像素点数)成一次线性复杂度。

2 Scale Saliency

Gilles[8]利用香农熵来定义一个区域的信号复杂性并用之衡量显著性。Kadir and Brady 修改了 Gilles 的算法提出 Scale-Saliency[7]算法：首先，Scale-Saliency 能够在多尺度上监测显著区域，而不像 Gilles 的算法在固定尺度上来监测显著区域。其次，使用概率分布(PDF)的尺度间变化程度来加权局部熵的峰值，这种方法使得尺度间的变化越剧烈越得到更高的显著值，使得监测的区域更加稳定。下面简单介绍了 Scale-Saliency 算法。

I. 区域的局部香农熵 H_D 是尺度 s 及中心坐标位置 x 的函数，并定义为：

$$H_D(s, x) = - \int_{d \in D} p(d, s, x) \log_2 p(d, s, x) \cdot dd \quad (1)$$

其中 $p(d, s, x)$ 是中心位置为 x 尺度(相当于半径)为 s 的区域的概率分布， d 为区域的描述子，且取值空间为 D 。

II. 尺度间的概率总体变化作为加权 $W_D(s_p, x)$ ，被定义为：

$$W_D(s_p, x) = s \int_{d \in D} \left| \frac{\partial}{\partial s} p(d, s, x) \right| \cdot dd \quad (2)$$

III. 假设在尺度 s_p 时局部熵得到了局部极大值

$$s_p = \left\{ s : \frac{\partial H_D(s, x)}{\partial s} = 0, \frac{\partial^2 H_D(s, x)}{\partial s^2} < 0 \right\} \quad (3)$$

则此区域的显著性被定义为：

$$y_D(s_p, x) = H_D(s_p, x) \times W_D(s_p, x) \quad (4)$$

总之，Scale Saliency 综合信息熵和几何约束的方法来计算显著性，首先在特征空间使用香农熵来描述显著性，然后使用 PDF 的变化情况来增加几何约束。

3 参考上下文的显著区域检测算法

如上节所述 Scale Saliency 增加了多尺度和几何约束, 然而 Scale Saliency 算法没有考虑上下文对区域显著性的影响, 这使得 Scale Saliency 在本文的应用中显示了不足, 它圈定的显著区域往往不能准确地包含台标区域, 往往是缠绕在台标或其他物体的边缘上, 或落入局部复杂区域, 如后面实验部分图 1a 所示。

本文算法将局部区域周围的上下文作为区域显著性计算的重要因素。算法的基本思想是不单独从局部特性上判断区域的显著性, 而是将局部区域与其外围上下文区域进行比较, 并且认为显著区域应该为与上下文比较起来显示出差异比较大的区域。首先, 将图像局部区域的显著性定义为一个区域与其外围上下文的差异程度。其次, 使用相对熵也就是 Kullback-Leibler 距离来计算这种差异程度, 从而得到了新的区域显著性计算公式:

$$KL_D(s \| c, x) = - \int_{d \in D} p_s(d, s, x) \frac{\log_2 p_s(d, s, x)}{\log_2 p_c(d, s, x)} dd \quad (5)$$

其中 $P_s(d,s,x)$ 是中心位置在 x , 尺度为 s 的局部区域的信息概率密度, $P_c(d,s,x)$ 该区域外围上下文区域的信息概率密度。 d 为信息的描述子, 其取值空间为 D , 比如灰度信息 d 为每个像素的灰度值, 取值区间为 $[0, 255]$ 。在信息理论中相对熵通常被称为 Kullback-Leibler 距离, 它是一个计算两种概率密度分布之间差异的常用方法。如实验部分图 1.b 是本文算法得到的结果, 可以看出本文算法考虑到了上下文对显著性的影响, 其提取的显著区域能够较好的包含台标区域。具体实现算法如下:

- 01 对于图像 I 中的每个像素位置 (x,y) :
- 02 对于在 S_{min} 和 S_{max} 之间的尺度 s :
- 03 IS =在距离像素位置 (x,y) 小于 s 的范围内的统计值集合
- 04 $P_s(d,s)$ =估计局部统计值集合 IS 的统计概率分布 PDF;
- 05 IC =在距离像素位置 (x,y) 大于 s 小于 $s+\Delta$ 的环形范围内的统计值集合
- 06 $P_c(d,s)$ = 估计外围统计值集合 IC 的统计概率分布 PDF;
- 07 $KLD(s,c,x)$ =计算概率分布 $P_s(d,s)$ 与 $P_c(d,s)$ 之间的 Kullback-Leibler 距离;
- 08 对于相对熵的峰值的 $KLD(s,c,x)$: 显著度 $YD(SP,x,y) = KLD(s,c,x)$

在公式(5)中相对熵 KLD 可以被看成在给定外围上下文的概率密度 $P_c(d,s,x)$ 作为先验概率, 得到的中心区域的概率分布相当于后验概率 $P_s(d,s,x)$ 后所得到的信息。相对熵越大得到的信息就越多, 就说明局部区域越显著。通过这种方法就可以更好的解释显著性的产生。

4 台标识别算法

台标识别首先需要建立需要识别的台标模版库, 为此首先抽取台标比较的清晰的视频帧, 并提取台标的边缘信息建立掩码, 再统一归一化到一定的尺度; 然后在识别的过程中, 通过上节中提取出的显著区域, 对显著区域提取边缘, 并归一化到跟模版同样的标准大小, 通过计算 Hausdorff 距离的办法来进行识别。这种识别的算法由于只利用到了边缘信息, 能够比较好的减少背景对识别的干扰, 不需要做前景背景的分割, 且能够实现多分辨率的识别。

设模版为上点集为 B , 提取的某个显著区域的边缘的点集为 A 则它们的 Hausdorff 距离为:

$$h(A,B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (6)$$

这样得到的距离是点集 A 中每一点距离点集 B 中点的最近距离的最大值，Hausdorff 距离比较好的特性是它不需要点集 B 与点集 A 完全的匹配。微小地移动 A 中某一点对结果影响不是太大。这是基于形状的认识比较重要的优点。为了准许一部分的偏离或缺失，可以使用部分 Hausdorff 距离：

$$h_K(A,B) = K^{\text{th}} \min_{a \in A} \min_{b \in B} \|a - b\| \quad (7)$$

比如 $K=0.9|A|$ 时部分距离 $h_K(A,B)$ 小于一定阈值 d 时表示 90% 边缘点能够跟模版匹配上。但是这种求距离的算法时间复杂度比较高，首先需要穷举出点集 A 中点与点集 B 中点的最小距离，需要时间为 $O(|A||B|)$ 。其次需要对这些最小距离数组选出第 K 个最小距离，所以需要排序时间复杂度为 $O(|A|^2)$ 。

为了减少时间消耗和更加容易实现，本文将 A, B 使用矩阵 Ma, Mb 的方式来表示，用 1/0 二值来表示是否为集合的点，设 Mb^d 表示将台标模板的边缘使用相态学函数向两侧扩展 d 个像素后的掩码，则有：

$$h_K(A,B) \leq d \Leftrightarrow Ma \bullet Mb^d \geq K \quad (8)$$

$$Ma \bullet Mb^d = \sum_{x=1..width} \sum_{y=1..height} Ma(x,y) \times Mb^d(x,y) \quad (9)$$

其中 \bullet 表示如(9)式表示的两个矩阵的点乘。通过这样点乘就能够计算出与模板匹配上的边缘点数。当它大于阈值 K 的时候则判定上它匹配了该模板。修改后的算法时间复杂度为模板的大小即 $O(\text{width} \times \text{height})$ 。

5 实验与分析

在实验中首先选择了一些电视台的台标提取模板如图 3 所示。然后提取一些包含台标和不包含台标的视频帧各 300 张，采取人工标注的方法来进行测试，将所有含有台标的视频帧被正确检测的比率为查全率，就不包含台标但被错误地识别出台标或包含某电视台台标但被错误地识别为其他电视台的台标的比率为误检率，其中公式(8)中 d 设为 1，K 为 $0.8|A|$ ，实验结果如表 1 所示。

如图 1 所示本文算法相对于 Scale Saliency 算法的优势。图 1.a 是 Scale Saliency 算法得出的结果，由于没有考虑到上下文对显著性的影响，使得检测出的区域落入复杂局部区域，缠绕在边缘，没有能够圈出台标区域。图 1.b 本文算法得到的结果，由于考虑到了上下文对显著性的影响，能够较好地圈出台标区域。需要注意的是，虽然在图 1 标出所有的显著区域，在实际的识别过程中只检测拐角上的显著区域。

如图 2 所示本文算法在遇到静态背景、动画台标登情况时会有明显的优势。当背景是静态的情况下使用帧差法无法检测出台标区域，而本文算法能够从单帧上通过提取显著区域从而圈出可能的台标区域。图 4 给出一些帧的显著区域的提取结果。

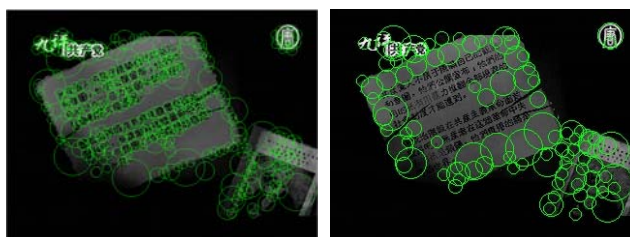


图1.a

图1.b

图1.a为Scale Saliency算法得出的结果，可看出由于没有考虑上下文对显著性的影响，使得检测出的区域落入复杂局部区域，缠绕在边缘，没有能够圈出台标区域。图1.b为本文算法得到的结果，由于考虑到了上下文对显著性的影响，能够比较准确的圈出台标区域。

图1 Scale Saliency与本文算法的比较

Fig.1 Compare between Scale Saliency and our method



图 2.a

图 2.b

图 2.a 由于背景相对静止，所以通过帧差法并没有检测出台标区域(其中颜色加黑区域为多帧中变化较小的区域)。图 2.b 通过本文的算法提取的显著区域能够较好从图像包含台标区域。

图 2.帧差法与本文算法的比较

Fig.2 Compare between difference of Multi Frames and our method



图 3.台标模板

Fig.3 TV Logo templates

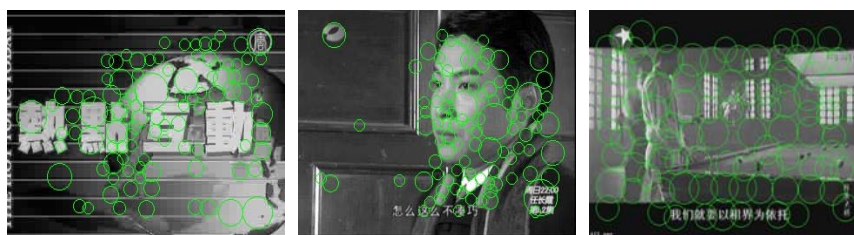


图 4.本文算法提取显著区域的部分结果

Fig.4 Saliency detection result of several pictures

表 1 台标检测与识别结果

Tab.1 Result of TV Logo detection

台标名称	查全率%	误检率%
湖南卫视	83.2	3.5
新唐人	94.1	2.6

凤凰卫视	92.7	2.0
东方卫视	92.9	2.4

6 结束语

本文实现了利用单帧信息进行台标检测与识别的算法,在文献中利用单帧视频图像进行台标检测的算法还很少有报道。本文首先重新定义了图像局部区域显著性的度量方式,提出参考上下文的显著区域检测算法,利用中心区域与外围上下文区域灰度值分布概率的相对熵度量中心区域的显著性,能够更加准确地在图像中圈定包含台标的显著区域。其次,计算显著区域与台标模板之间的 Hausdorff 距离实现了台标识别,这种算法有不需要进行训练且能够较好的排除背景干扰的优点。实验显示该算法有比较好的结果。当然这种算法还有不少需要改进的地方,比如这种算法比较倾向于圈出 Blob-Like 区域,对于现状比较扁长的台标比如 cctv 等台标就不适用了。在以后的工作中我们就继续研究如何快速的自适应地提取各种形状的显著区域。

参考文献

- [1] Katrin Meisinger, Tobias Troeger, Marcus Zeller, and Andr'e Kaup, "Automatic tv logo removal using statistical based logo detection and frequency selective inpainting," Proc. European Signal Processing Conference'05, September 2005.
- [2] Wei-Qi Yan, JunWang, and Mohan S. Kankanhalli, "Automatic video logo detection and removal," ACM Trans. on Multimedia System, July 2005.
- [3] A. Albial, M. J. C. Full'a, A. Albial, and L. Torres, "Detection of tv commercials," Proc. ICASSP'04, May 2004.
- [4] Wang, J., Duan, L., Li, Z., Liu, J., Lu, H., Jin, J.S.: A robust method for tv logo tracking in video streams. Proc. ICME'06 (2006)
- [5] Jinqiao Wang, Qingshan Liu, Lingyu Duan, Hanqing Lu, and Changsheng Xu :Automatic TV Logo Detection, Tracking and Removal in Broadcast Video. LNCS 4352, Part II, pp. 63-72, 2007.
- [6] Alberto Albiol , Mar'ia Jos'e Ch. Full'a, Antonio Albiol. Detection of TV Commercials. ICASSP 2004
- [7] T. Kadir and M. Brady, "Saliency, scale and image description," International Journal ofComputer Vision, pp. 83-105, 2001.
- [8] S. Gilles, "Robust description and matching of images," PhD thesis, University of Oxford, 1998.