一种基于变异灰度直方图的视频字幕检测定位方法

张佑生,彭青松,汪荣贵
(合肥工业大学计算机与信息学院,安徽合肥 230009)

摘 要: 为实现视频图像中字幕的快速检测与定位,本文提出一种变异灰度直方图 VGH. 文章给出 VGH 的定义,对其中的特征及其与图像中字幕的映射关系作了分析,并在此基础上给出视频字幕的检测与定位方法. 该方法将垂直子图像转换为基于行的 VGH (VGH^R),将水平子图像转换为基于列的 VGH (VGH^C),通过对其中的凹谷特征和梳状凸台特征的识别,实现对字幕的检测与定位. 文章最后给出某些实验结果,表明了算法的有效性.

关键词: 灰度直方图; 变异灰度直方图; 视频图像; 字幕检测与定位
中图分类号: TP391 **文献标识码**: A **文章编号**: 0372-2112 (2004) 02-0314-04

A Method of Caption Detection and Location in Video Images Based on Variant Gray-Scale Histogram

ZHANG You-sheng, PENG Qing-song, WANG Rong-gui
(School of Computer & Information, Hefei University of Technology, Hefei, Anhui 230009, China)

Abstract: The paper presents a variant gray-scale histogram (VGH) for the purpose of detecting and locating captions in video images. The definition of VGH is given and the mapping relationships are analyzed between features in VGH and captions in an image. Then, we proposed an algorithm for detecting and locating captions in an image by recognizing the features, including both valleys in row-based VGH of a vertical sub-image and convexes in column-based VGH of a horizontal sub-image. Some experimental results are given in the end of the paper showing validity of the algorithm.

Key words: gray-scale histogram; variant gray-scale histogram; video image; caption detection and location

1 引言

数字视频信息是人类极为宝贵的信息资源,在新闻、教育、影视和各种多媒体应用方面起着越来越大的作用,其数量巨大、内容丰富,非同寻常. 它一般包含图像、文本和声音等信息. 为了充分使用这一信息源,人们研究出了许多基于内容的检索、分析方法,用于从巨量的存档视频资料中检索和浏览所需视频信息.

在视频数据中,文本/字幕(Text/Caption)往往给出关于视频内容的重要信息. 因此,视频图像中文本/字幕(下文简称字幕)检测对视频内容的分析理解有重要的作用. 但是,可靠检测和定位视频图像中的字幕不是一件容易的事情. 其中存在多方面复杂因素,例如:文本中字符尺寸宽幅变化,字体多种多样;同一文本行的字符颜色可能不同;在一个视频序列中文本可能静止不动,也可能朝某个方向移动,也可能改变大小;文本可能在很复杂的背景中出现;等等[1].

为了从图像和视频信号中提取文本和字幕,人们已提出了多种方法,如文本区检测(Text Area Detection)、文本跟踪(Text Tracking)等. 文献[2]将文本视为服从一定大小约束和水平对齐约束的相连接的单色成分进行抽取. 文献[3]假设文本在视频序列中静止,使用链码从视频图像中分割文本成分,并使用时间信息求精. 文献[4]将视频图像分割成多个不同颜色的子图像,然后考察是否每个子图像包含满足某些启发式知识的文本成分. 文献[5]采用基于边(edge - based)的方法,分析边亮度图,实现对文本区的分割.

本文提出一种基于变异灰度直方图(Variant Gray-scale Histogram, VGH)的检测定位字幕的快速有效方法. 文章给出 VGH 的定义,分析了它的性质,并指出这种定义也适用于子图像;然后,对 VGH 在字幕检测定位中的应用进行了详细讨论. 实验结果表明,该方法完全可行,有关算法十分有效.

2 图像 VGH 及性质

现在广泛应用的灰度直方图(Gray-scale Histogram, GH)是灰度级的函数,描述的是图像中具有某灰度级的像素的出现的次数,其横坐标表示灰度级,纵坐标表示某灰度出现的频率[6]. 虽然,GH 在多个方面有实用价值. 但是,它不能反映图像中的局部特征和对象属性. 而对于视频信号中对象的识别与理解,如基于字幕的检索等来说,局部特征和对象属性的识

收稿日期:2002-07-02;修回日期:2003-08-30
基金项目:安徽省自然科学基金(No. 04-03042207)

别与分析十分重要.出于字幕检测定位的需要,下面给出一种VGH的定义:

定义1　给定一视频图像 $f(x,y)$,设其像素的行数为 L,则其VGH函数可定义为:

$$P(k,Sa)=1-Nr(k,Sa)/Rn,\quad k=0,1,\cdots,L-1 \qquad (1)$$

其中,k 为图像行号,Rn 是一行中的像素总数,Sa 是用灰度级表示的阈值参数,$Nr(k,Sa)$ 是第 k 行中灰度值 $\geq Sa$ 的像素数.

由此定义可以看出,在VGH中,横坐标表示图像的行号,纵坐标表示一行中灰度值小于 Sa 的像素所占的比例.这种直方图具有如下的特点和性质:

(1)参数 Sa 的影响　当 Sa 取图像最大灰度级(S_{max})时,P 为全1.0;当 Sa 取图像最小灰度级(S_{min})时,P 为全0;适当选取 Sa 的值,如 $Sa=(0.3\sim0.6)S_{max}$,将使VGH产生特定对象的映射特征.

(2)局部描述能力　图像中局部区域的灰度变化,将在VGH中反映出来.也就是说,图像中的特定对象,如视频字幕、汽车牌照、商标与条码等将在VGH中产生特定的映射特征.这种局部描述能力使VGH具有实用价值.

(3)局限性　VGH能反映图像特定区域的存在,但不能提供关于区域内容的更进一步的信息.

上述VGH是按行生成的,称为基于行的VGH,记为 VGH^R.与此相应,我们可定义基于列的VGH即 VGH^C 如下:

定义2　给定一视频图像 $f(x,y)$,设其像素的列数为 L,则 VGH^C 函数可定义为:

$$P(j,Sa)=Nc(j,Sa)/Cn,\quad j=0,1,\cdots,L-1 \qquad (2)$$

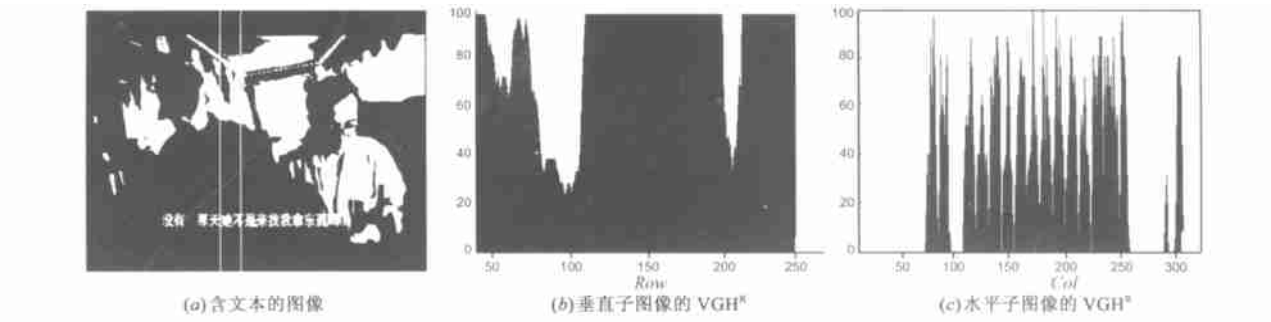
其中,j 为像素的列号,Cn 是一列中的像素总数,$Nc(j,Sa)$ 是第 j 列中灰度值 $\geq Sa$ 的像素数.由此定义可知,其横坐标表示图像列号,而纵坐标表示一列中灰度值大于 Sa 的像素所占的比例.VGH^C 具有与 VGH^R 相似的性质.

VGH的两种定义从不同侧面对图像中对象进行描述,不但可以互为验证,而且可以分别提供特定对象所在行或列的信息,互为补充.因此,实用中可根据需要选择其一为主,而以另一种为辅.此外,上述两种VGH定义均适用于图像中的任意子图像,如若干连续行构成的水平子图像和若干连续列构成的垂直子图像.从识别映射特征考虑,对水平子图像生成 VGH^C,对垂直子图像生成 VGH^R,将两者配合使用.

3　VGH中的映射特征

图像字幕形状比较规则,其中的字符具有明显高于(或低于)其邻近区域的亮度,故可在VGH中生成凹谷或梳状凸台映射特征.图1(a)给出一帧含字幕的图像,其中字幕的位置为:(row:200 - 213,col:70 - 252).在其中取垂直子图像(row:1 - 252,col:121 - 140,如白线所示),生成 VGH^R,其右半部存在一典型的凹谷(起止行号为200,213),如图1(b)所示;参照凹谷的起止行号,取水平子图像(row:200 - 213,col:1 - 320),生成 VGH^C,如图1(c)所示,其中有一梳状凸台(其起止列号为70,252).

可见,凹谷与凸台特征与图像中的字幕对应.若分别检测出图1(b)与(c)中凹谷与凸台,则可以判定图像中存在字幕,其位置可根据凹谷和凸台的位置确定.



(a)含文本的图像　(b)垂直子图像的 VGH^R　(c)水平子图像的 VGH^C

图1　含文本图像及其中子图像VGH

为了清楚而准确地说明子图像VGH中的文本映射特征的存在性,下面给出一个相关命题(证明略).

命题1　如果由若干连续列构成的子图像穿越(或紧包含)某一字幕(假定字符亮度高于背景),且式(1)中的 sa 值适当,则该子图像的 VGH^R 中必有与字幕对应的凹谷特征.如果由若干连续行构成的子图像穿越(或紧包含)某一字幕(假定字符亮度高于背景),且式(2)中的 sa 值适当,则该子图像的 VGH^C 中必有与字幕对应的梳状凸台特征,且凸台两边的邻近区域函数值很小甚至为0.

对于VGH中的文本映射特征,可给出它们的描述参数,作为判别文本存在的依据.

凹谷的描述参数有:1)边界点:指凹谷的下降缘起点和上升缘终点;2)底部均值:指底部函数值的平均值;3)深度:指左右边界点函数平均值与底部均值之差;4)边缘斜度:指下降缘或上升缘的函数变化率.

梳状凸台由多个凸单元(简称单元)构成,其描述参数包括:1)凸台位置与宽度:位置指第一单元的起点和最后一个单元的终点,该两点之间的列数为凸台宽度;2)邻近区域宽度:指其左右两边邻域的宽度;3)单元高度:指各个单元的函数平均值;4)单元宽度:指各个单元所占列数,各个单元宽度的平

均值为单元平均宽度;5) 单元数目:指凸台包含的单元个数;6) 间隙宽度:指单元之间的间隙所占列数,所有间隙宽度的平均值为间隙平均宽度.

4 VGH中的特征检测与定位

凹谷的特点是具有一对陡峭的下降、上升边缘,在两边缘之间函数值较小而边缘顶部函数值较大(即有较大深度),底部波动不很大,等等. 为了有效地检测凹谷,需要充分利用这些特点.

我们知道,曲线向下拐时,如凹谷下降缘起点和上升缘终点,函数二阶导数出现负的极大值;向上拐时,如下降缘终点和上升缘起点,二阶导数出现正的极大值. 因此,可以通过对函数 $P(k)$ 求二阶差分、寻找二阶差分的极大值及其符号模式$\{-,+,+,-\}$,可得到下降、上升边缘对,从而找到候选凹谷.

这种方法对简单图像是可行的. 但是,当图像比较复杂并存在噪声数据影响时,$P(k)$ 函数存在较大的局部波动,需要进行平滑处理. 这不但增加了计算量,检测效果也欠佳. 经进一步研究,我们得到另一种更有效的方法——函数值连续下降(上升)累计法,其基本思想是:对 $P(k)$ 进行逐点考察,当遇到连续点序列$\{k\mid P(k+1)<P(k)\}$或$\{k\mid P(k+1)>P(k)\}$或$\{k\mid P(k+1)=P(k)\}$时(设 $k=j_1,\cdots,j_2$),对数组 VL 作如下赋值:

$$VL(1,k1)=P(j_2)-P(j_1),\ VL(2,k1)=P(j_1),$$
$$VL(3,k1)=j_1,\ VL(4,k1)=j_2.$$

其中,$k1$ 为 VL 数组的列号. 当 $P(k)$ 的全部点都考察完时,就生成了全部下降段($VL(1,k1)<0$)和上升段($VL(1,k1)>0$). 这样,找出一对相邻的满足$|VL(1,*)|>\xi$(预定阈值)的下降、上升段,就找到了一个凹谷,其始止行号、边缘及其变化速率、底部宽度等参数也都可计算出来.

根据这种思想,可设计出快速检测定位凹谷的算法(简称VDLA). 该算法描述如下:

Step1 对当前子图像生成 VGH^R,在其中检测下降、上升边缘对$\{k_{i1},k_{i2}\}$,$i=1,\cdots,nk$;

Step2 若 $nk>0$,则初步认为找到了 nk 个凹谷及其边界,于是逐个计算出凹谷宽度 W_i,其两边缘的函数变化率绝对值 r_{i1} 与 r_{i2},并通过计算凹谷顶部函数均值与底部函数均值,得到它的深度值 d_i,转 Step3;否则,转 Step4;

Step3 若 $\min(r_{i1},r_{i2})>r_0$, and, $d_i>d_0$. and. $W_i>W_0$,则令 $Valley(i)=(k_{i1}$的起始行号, k_{i2}的终止行号). 处理完时返回调用程序;

Step4 令 $Valley=0$,返回调用程序.

该算法中的 r_0、d_0 和 W_0 为门限参数,可根据经验或通过对多幅图像进行检测实验得到.

5 基于 VGH的字幕检测定位方法

根据上述分析可知,VGH^R 可确定字幕的起止行,VGH^C 可确定其起止列. 因此,我们将这两种变异直方图结合起来,提出一种基于 VGH 的字幕检测定位方法. 该方法的基本步骤为:

(1) 将一幅图像划分为一系列垂直子图像,逐个对其进行以下各步的处理;

(2) 对当前子图像,选取适合的阈值,按式(1)生成其 VGH^R,并采用凹谷快速检测定位算法(VDLA),以查找凹谷特征;

(3) 若(2)中找到了 nk 个($nk>0$)凹谷特征,则逐个用其起、止行号构造水平子图像,生成其 VGH^C,再用凸台快速检测定位算法(CDLA),查找梳状凸台;

(4) 若(3)中找到了凸台特征,则对上述的凹谷与凸台特征逐对计算出描述参数,进行综合判别,以确认文本区,并得到它在图像中的准确位置.

算法 VDLA 已在第 4 节中给出,算法 CDLA 的思想与它基本相同. 由于梳状凸台的描述参数比较多,故其检测定位的算法 CDLA 复杂一些.

6 字幕检测实验与结果分析

上述字幕检测与定位方法相关算法已用 Java 语言编程实现. 经过对多幅图像的检测定位实验,得到了令人满意的结果. 对于图 1(a)的图像,将它划分成一系列的垂直子图像(宽度 30). 当处理图中两白线所界定的子图像时,得到数组 $VL(1:4,1:nv)$,(nv 为数组的列数). 下面列出该数组对应于图 1(b)右边凹谷附近的若干列数据,表中函数差值为-42.86 和 42.89 的两列对应于凹谷的下降与上升边缘,它们之间的各列为凹谷的底部有关数据. 从表中第三行可见,凹谷底部函数值都比较小.

表 1

函数差值	0.0	-42.86	0.0	-14.29	4.76	-19.05	14.29	0.0	28.57	-14.29	42.89	0.0
边缘起点	112	200	201	202	203	204	206	208	209	211	212	213
起点值	100.0	100.0	57.14	57.14	42.86	47.62	28.57	42.85	42.86	71.43	57.14	100.0
边缘终点	200	201	202	203	204	206	208	209	211	212	213	255

图 2 给出另外两例处理结果:(a)为一电影图片,(b)为一广告图片. 程序在图像中标出的方框,表明两幅图像中的字幕和文本都被正确检测出来,且被准确定位. 程序运行条件是:从第 1 列开始扫描,所用子图像的宽度为 20、30 或 40,扫描间隔为 12、15 或 18,(当图像总列数小于 320 时取其中的小值,大于 640 时取大值). 由图可见,尽管两图像都比较复杂,

且图(b) 出现了多行文本(列方向不对齐)的情况,但都获得了好的结果.

图 2　图像中文本检测定位结果 (a)电影图片中的字幕;
(b)广告图片中的文本

上述方法由于将二维图像转换为一维曲线,而以曲线作为主要处理对象,且所有二维至一维的转换操作都限于行(列)数很少的子图像,因此计算复杂度明显降低,处理速度加快.

7　结束语

本文提出一种变异的灰度直方图,通过对其中映射特征的分析,得到它们与图像中字幕的映射关系,并在此基础上提出一种检测定位视频字幕的方法.计算实验表明,该方法快速有效.

由于视频图像的复杂性,字幕的检测定位受多种因素的影响,这些因素都有某种程度的不确定性,因此,为了更有效地在视频序列中检测定位字幕,需要综合考虑更多影响因素,进行概率分析,建立适当的概率模型.此外,VGH 的特征描述参数具有模糊性,因此,采用模糊技术对凹谷和梳状凸台特征进行判别,可望取得好的结果.我们现正进行这两方面研究.

参考文献:

[1] Y Zhong, H J Zhang, A K Jain. Automatic caption localization in compressed video [J]. IEEE Trans. on PAMI. 22, 2000: 385 - 392.

[2] Y Zhong, K Karu, et al. Locating text in complex color images [J]. Pattern Recognition, 1995, 28(10): 1523 - 1536.

[3] J C Shim, C Dorai, et al. Automatic text extraction from video for content-based annotation and retrieval [A]. Proc. 14th Int'l Conf. Pattern Recognition [C]. UK: World Scientific, 1998. 618 - 620.

[4] A K Jain, B Yu. Aotumatic text location in images and video frames [J]. Pattern Recognition, 1998, 31(12): 2055 - 2076.

[5] W Qi, et al. Integrating visual audio and text analysis for news video [A]. 7th IEEE Int. Conf. on Image Processing (ICIP 2000) [C]. Vancouver, British Columbia, Canada, 2000. 10 - 13.

[6] Kenneth R Castleman. 数字图像处理[M]. 朱志刚, 等, 译. 北京: 电子工业出版社, 1998. 9.

作者简介:

张佑生　男,1941 年生于湖南省浏阳市,教授,博士生导师,感兴趣的领域为图像处理、计算机图形学和智能 CAD 等.

彭青松　男,1976 年生于江苏省连云港市,博士生,感兴趣的领域为图像处理.