基于灰度差分和二维最大熵阈值的新闻字幕检测 *

陈树越，张世林
（常州大学 信息科学与工程学院，江苏 常州 213164）

摘　要：针对新闻视频帧中文本区域的定位提取问题，提出了一种有效的字幕定位提取方法。通过灰度差分和变异灰度直方图对新闻视频帧字幕区域定位，再经改进的二维最大熵阈值方法对分割出的文字区域进行二值化，得到可识别的文字图片。最后对文本定位和 OCR 识别情况进行了算法对比。实验表明，与传统的投影法和最大熵方法相比，该方法可有效地提高文本定位的查全率和 OCR 的识别率。

关键词：文本定位；文本提取；灰度差分统计；变异灰度直方图；二维最大熵；印刷体识别
中图分类号：TP391.1　　**文献标志码**：A　　**文章编号**：1001-3695(2011)08-3195-03
doi:10.3969/j.issn.1001-3695.2011.08.110

Detection of news captions based on gray-scale difference statistics and two-dimensional maximum entropy threshold

CHEN Shu-yue, ZHANG Shi-lin
(*School of Information Science & Engineering, Changzhou University, Changzhou Jiangsu* 213164, *China*)

Abstract: As to the problem to locate and extract the text region of the news in video frames, this paper proposed an effective methodology for caption location and extraction. It used the gray-scale difference statistics and local gray-scale histogram to locate the text region of the video frame, then applied improved two-dimensional maximum entropy threshold to extract the segmented text area to obtain a binarized image. Finally, analyzed the abilities of several algorithms by comparison for text location and OCR. Experimental results show that the method is more effective to improve the recall ratio and OCR recognition rate compared with traditional projection technique and maximum entropy approach.

Key words: text location; text extraction; gray-scale difference statistics; variant gray-scale histogram(VGH); two-dimensional maximum entropy; optical character recognition(OCR)

0　引言

新闻视频中的标注文本[1]是对视频内容的高度概括，有效地定位提取这些文字对新闻的理解以及对新闻视频的检测起着非常重要的作用[2]。目前已经有很多方法用于提取新闻视频帧中的字幕，现有文字提取方法包括以下几个步骤：检测、定位、增强、提取和识别[1~3]。各种方法的区别在于字幕区域检测以及文字提取所采用的方法不完全相同。

视频字幕的检测定位主要有基于边缘检测、基于纹理和基于区域的方法等[2~5]。基于边缘检测的方法虽然效率较高，但鲁棒性较弱，在检测背景复杂的文字时错误率较高[2,6]；基于纹理的方法可以有效检测出复杂背景下的文字区域，但由于涉及全图的微分运算，因此算法效率较低[2,7]；基于区域的方法对图像文本字符的大小、文本与背景的对比度以及分辨率等要求比较高[3,6]。对于视频文字提取来说，一般的文字分割方法是利用阈值的方法。基于阈值的方法包括 Otsu 和 Niblack 方法，这种阈值的方法比较简单，对于一些复杂的图像来说效果不理想[2,3]。

本文根据新闻视频字幕的区域特征，以及分辨率低、文字难以提取的问题，给出一种基于灰度差分行定位与变异直方图列定位的新闻字幕区域定位方法，以及改进的二维最大熵阈值法对新闻字幕区域的分割。

1　新闻字幕检测算法

1.1　新闻视频字幕定位

本文采用基于灰度差分和变异灰度直方图的新闻字幕定位方法，其主要思路是：通过灰度差分统计对视频帧中文字区域进行行定位，然后再利用变异灰度直方图对文字区域进行列定位。具体步骤如图 1 所示。

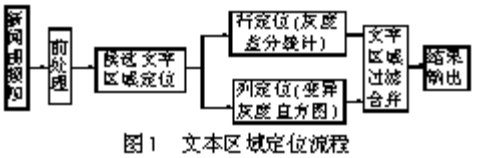


图 1　文本区域定位流程

1.1.1　文字区域行定位

通过灰度差分统计对视频帧中的文本区域进行行定位。下面给出灰度差分统计的定义。

定义 1　给定一视频帧图像 $f(x,y)$，设其像素的行和列分别为 m 和 n，则其灰度差分统计定义为[2,8]

收稿日期：2010-11-26；**修回日期**：2011-01-14　　**基金项目**：江苏省高校自然科学基金资助项目(08KJB520002)

作者简介：陈树越(1963-)，男，河北定州人，教授，博士，主要研究方向为信号与信息处理、测试技术等(csyue2000@yahoo.com.cn)；张世林(1986-)，女，江苏南通人，硕士研究生，主要研究方向为数字图像处理.

$$E(x)=\sum_{y=1}^{n-1}|f(x,y)-f(x,y+1)| \quad (1\leqslant x\leqslant m,1\leqslant y\leqslant n-1) \quad (1)$$

其中：$E(x)$表示为视频 x 行的相邻两个像素灰度差的绝对值累加之和，$f(x,y)$、$f(x,y+1)$分别为相应像素点的灰度值。

1.1.2　文本区域列定位

通过以上处理就可以得到候选文字区域的上下边界，其结果中还是包含一些非文本区域。于是，必须在行定位的基础上进行列定位，滤除非文本区域。这里可通过变异灰度直方图来实现。

传统的数字图像灰度直方图是灰度级的函数，它反映出图像中对应于各个不同灰度级的像素数，其横坐标表示灰度级，纵坐标表示某灰度出现的频率。但是其像素的统计在整个图像中进行，不能反映图像中的局部特征，而局部特征信息对于文字区域定位而言十分重要。

下面给出一种变异灰度直方图的定义（这里以按行生成变异灰度直方图为例）。

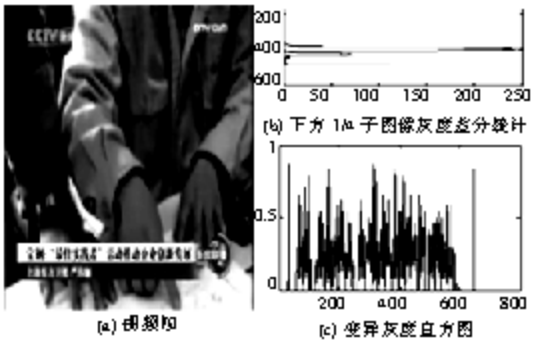
定义 2　给定一视频帧图像$f(x,y)$，取其由任意连续的行（数值为 L）所构成的子图像进行处理，生成变异灰度直方图（VGH）。

设子图像的像素的行和列分别为 L 和 n，则其 VGH 定义为[13]

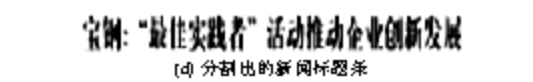
$$\mathrm{VGH}(y)=\frac{\sum_{x=1}^{L}f(x,y)}{L} \quad (1\leqslant x\leqslant L,1\leqslant y\leqslant n-1) \quad (2)$$

其中：$f(x,y)$为像素点(x,y)处的灰度值。

图 2 给出用基于灰度差分和变异灰度直方图的定位过程和结果。



(a) 视频帧　(c) 变异灰度直方图



(d) 分割出的新闻标题条

图 2　视频帧及定位结果

1.2　基于二维最大熵阈值的文字提取

经上述新闻标题文字定位，需进一步对文本图像分割处理，得到二值化的图像，作为接下来文字识别的输入。由于新闻视频字幕分辨率比较低，先将切割出的新闻标题条用双三次插值方法放大 1.5 倍再进行处理。采用二维最大熵阈值的方法对文字区域进行提取，并对传统二维最大熵阈值作了一些改进。具体算法流程如图 3 所示，其中 $H_O(s,t)$和 $H_B(s,t)$分别表示目标和背景的熵。

二维最大熵利用了图像的灰度信息和邻域的空间相关信息，能有效地抑制图像噪声，得到较高精度的分割结果。传统方法直接采用中心邻域灰度值与 $k\times k$ 邻域的灰度均值构成二维直方图。为了有效地突显文字区域的特性，本文采用中心邻域灰度值与 $k\times k$ 邻域灰度差分构成二维直方图得到二维最大熵。

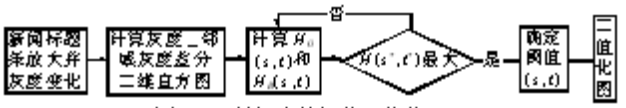


图 3　二维最大熵阈值二值化流程

对视频帧图像$f(x,y)$，设其行数和列数分别为 M 和 N，图像具有 L 级灰度级，相应的像素邻域平均灰度也有 L 级灰度级。图像灰度差分为

$$e(x,y)=\sum_{m=-[k/2]}^{[k/2]}\sum_{n=[k/2]}^{[k/2]}|f(x,y)-f(x+m,y+n)| \quad (3)$$

其中：$0<x+m<M$，$0<y+n<N$；$f(x,y)$表示(x,y)处的灰度值（记为 i）；$e(x,y)$表示(x,y)处的 $k\times k$ 邻域的灰度差分值（记为 j）；$[k/2]$表示对 $k/2$ 取整。

利用图像像素点的灰度值和其 $k\times k$（文中 $k=3$）邻域的灰度差分值组成的灰度组对(i,j)来表示图像。设灰度对出现的次数记为 $m_{i,j}$，则相应的联合概率密度 $P_{i,j}$为

$$P_{i,j}=\frac{m_{i,j}}{M\times N} \quad (i,j=0,1,\cdots,L-1) \quad (4)$$

以二维矢量(s,t)作为阈值分割图像，则二维直方图就被分成目标、背景、边缘、噪声四个区域。二维熵法中一般假设边缘和噪声区域的概率近似为 0，目标和背景出现的概率和则近似为 1。目标和背景的熵分别由式(5)和(6)计算得到：

$$H_o(s,t)=\log_2(P(s,t))+\frac{h(s,t)}{P(s,t)} \quad (5)$$

$$H_B(s,t)=\log_2(1-P(s,t))+\frac{H(L,L)-h(s,t)}{1-P(s,t)} \quad (6)$$

其中：

$$P(s,t)=\sum_{i=1}^{s}\sum_{j=1}^{t}P_{i,j},h(s,t)=-\sum_{i=1}^{s}\sum_{j=1}^{t}P_{i,j}\log_2 P_{i,j}$$

$$H(L,L)=-\sum_{i=1}^{L}\sum_{j=1}^{L}P_{i,j}\log_2 P_{i,j}$$

则最优阈值(s,t)应当满足

$$H(s^*,t^*)=\max_{1<s,t<L}\{H_o(s,t)+H_B(s,t)\} \quad (7)$$

通过获得的阈值(s,t)对图像进行二值化处理，标题文字提取结果如图 4 所示。

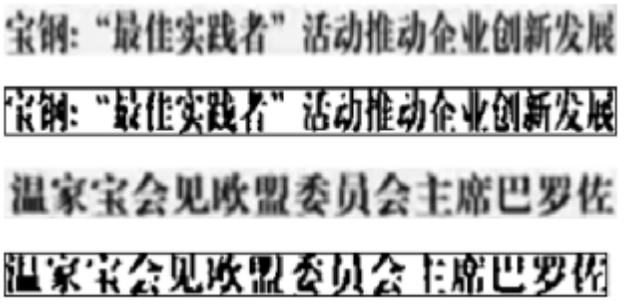


图 4　新闻标题文字提取结果

2　实验结果与分析

为了验证算法的有效性，将所提出的方法在一组视频上进行测试。测试环境为：Intel Core2 T5600 CPU；内存 512 MB；操作系统为 Windows XP；开发工具为 MATLAB7.1；算法流程如图 1、3 所示。测试视频帧是随机抽取 CCTV-1 的新闻联播

2009 和2010 年节目中具有字幕的视频帧。帧画面的分辨率多是为 352×240 和 486×648。采用文献[3]给出的评定文字定位算法的标准来检测新闻标题定位的有效性,即

查全率 = 正确定位的文字块数目/所有真实的文字块

错检率 = 错误定位的文字块数目/所有真实的文字块

本文方法与投影法[13]的文字定位查全率和错检率的结果如表1 所示,其中查全率提高 5.94%,错检率降低了 8.42%。

对于文字提取,实验结果用 OCR 的识别率和查准率来评测算法的有效性,即

识别率 = 正确识别字符的数目/所有真实字符的数目

查准率 = 正确识别字符的数目/所有识别字符的数目

与传统最大熵法相比,如表 2 所示,本文方法在识别率上提高了 3.68%,查准率提高了 3.67%。

表1　文字定位结果

文字定位	查全率/%	错检率/%
投影方法	87.85	15.37
本文方法	93.79	6.95

表2　文字提取结果

文字提取	识别率/%	查准率/%
传统最大熵法	76.69	79.25
本文方法	80.37	82.92

实验表明本文算法在查全率上与投影法相比得到提高,在识别率、查准率上与传统灰度均值二维最大熵方法相比有所提高。

3　结束语

本文采用灰度差分和变异灰度直方图对新闻视频帧文字区域进行定位,有效消除噪声和背景干扰的同时,可以准确定位新闻视频的标题文字行数及其区域。文字提取采用改进二维最大熵法,有效地提取出了新闻字幕文字,相比传统最大熵法在识别率和查准率上均得到提高。实验结果表明算法效果良好,不受文字大小影响,且易于实现。该算法对字幕文字提取后所得到的识别率还不够高,进一步提高文字提取的准确性以及建立一个自动新闻视频检索系统,是笔者下一步的工作。

参考文献:

[1] 章东平. 视频文本的提取[D]. 杭州:浙江大学,2006.

[2] 叶静. 基于内容的新闻视频检索[D]. 上海:上海大学,2001.

[3] 宋砚,刘安安,张勇东,等. 基于聚类的视频字幕提取方法[J]. 通信学报,2009,30(2):136-140.

[4] 朱成军,李超,熊璋. 视频文本检测和识别技术研究[J]. 计算机工程,2007,33(10):219.

[5] LYU M R,SONG J G,CAI Min. A comprehensive method for multilingual video text detection,localization,and extraction[J]. IEEE Trans on Circuits and Systems for Video Technology,2005,15(2):243-255.

[6] 赵亚琴. 一种有效的新闻视频主题字幕检测方法[J]. 计算机工程与应用,2009,45(33):175-178.

[7] YE Qi-xiang,GAO Wen,HUANG Qing-ming. Automatic text segmentation from complex background[C]//Proc of IEEE International Conference on Image Processing. [S. l.]:IEEE Press,2004:2905-2908.

[8] 许家佗,孙炀,张志枫,等. 基于差分统计方法的舌象纹理特征的分析与识别[J]. 上海中医药大学学报,2003,17(3):55-58.

[9] 张佑生,彭青松,汪荣贵,等. 一种基于变异灰度直方图的视频字幕检测定位方法[J]. 电子学报,2004,32(2):314-317.

[10] WONG A K C,SAHOO P K. A gray level threshold selection method based on maximum entropy principle[J]. IEEE Trans on Systems Man and Cybernetics,1989,19(4):866-871.

[11] 龙建武,申铉京,魏巍,等. 基于均值—中值—梯度共生矩阵模型的最大熵分割算法[J]. 计算机应用研究,2010,27(9):3575-3578.

[12] 彭静,章宝歌,刘小明. 一种基于边界特征的二维最大熵分割算法[J]. 计算机与数字工程,2008,36(7):18-20.

[13] 徐峰,梁学战. 新闻视频帧中的标题字幕探测[J]. 中国科技信息,2009(23):117-120.

(上接第 3194 页)该算法不需要考虑阈值的选取问题。在本算法中通过改变 K-SVD 字典学习算法中的约束条件,并在实验的基础上分析了不同的噪声标准差与非零元个数在图像去噪中峰值信噪比的规律,从中发现对于含较大标准差的噪声图像,去噪后峰值信噪比随非零元个数的增加而下降;而对于含较小标准差的噪声图像,去噪后峰值信噪比随非零元个数的增加而增加,当达到一定非零元个数时,峰值信噪比趋于不变后变化很小。因此通过上述规律,采用两个固定数目的非零元来去噪,得到了一个简便实用且效果更好的去噪算法。实验结果表明,与小波类去噪方法相比,本算法能够更好地去除图像中的高斯白噪声,对于高噪声的图像尤其有效,在保留图像的边缘和细节信息,提高去噪图像的 PSNR 值,且具有较好的视觉效果。

图5　含噪图像 cameraman 的去噪效果

参考文献:

[1] 张文革,刘芳,焦李成,等. 基于 Bandelets 域逐子块阈值的图像去噪[J]. 电子学报,2010,38(2):290-294.

[2] DONOHO D L. Denoising by soft thresholding[J]. IEEE Trans on Information Theory,1995,41(3):613-627.

[3] 杨晓慧,焦李成,李伟,等. 基于第二代 bandelets 的图像去噪[J]. 电子学报,2006,34(11):2063-2067.

[4] 杨帆,赵瑞珍,胡绍海. 基于 Contourlet 系数相关特性的自适应图像去噪算法[J]. 光学学报,2009,29(2):357-361.

[5] ELAD M,AHARON M. Image denoising via sparse and redundant representation over learned dictionaries[J]. IEEE Trans on Image Processing,2006,15(12):3736-3745.

[6] AHARON M,ELAD M,BRUCKSTEIN A. K-SVD:an algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Trans on Signal Process,2006,54(11):4311-4322.

[7] CHEN S,BILLINGS S A,LUO W. Orthogonal least squares methods and their application to non-linear system identification[J]. International Journal of Control,1989,50(5):1873-1896.

[8] DAVIS G,MALLAT S,ZHANG Z. Adaptive time-frequency decompositions[J]. Optical Engineering,1994,33(7):2183-2191.