

基于小波变换和 SVM 的文本区域定位

孙红星¹, 赵楠楠², 徐心和¹

(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110004; 2. 辽宁科技大学 电信学院, 辽宁 鞍山 114044)

摘 要: 提出了一种基于小波变换和支持向量机(SVM)在数字图像中定位文本的方法. 对图像进行小波变换,并在低频概貌和高频能量空间应用 SVM 提取文本的纹理特征,由 SVM 来决定当前的像素是文本类还是非文本类. 因为 SVM 的分类结果可能存在噪声或错误,用形态学去噪和计算纹理能量的方法对 SVM 的分类结果进行后处理. 小波变换和 SVM 的结合,不仅降低了输入空间样本的数量,而且利用了 SVM 适合于高维空间工作的特点,提高了文本提取的效率. 实验结果表明,提出的方法可以快速有效地定位数字图像中的文本区域.

关 键 词: 文本检测;纹理分析;小波变换;支持向量机(SVM);形态学

中图分类号: TP 181 **文献标识码:** A **文章编号:** 1005-3026(2007)02-0165-04

Text Region Localization Using Wavelet Transform in Combination with Support Vector Machine

SUN Hong-xing¹, ZHAO Nan-nan², XU Xin-he¹

(1. School of Information Science & Engineering, Northeastern University, Shenyang 110004, China; 2. School of Electronics and Information Engineering, Liaoning University of Science and Technology, Anshan 114044, China. Correspondent: SUN Hong-xing, E-mail: neushx @sohu.com.)

Abstract: A method based on wavelet transform in combination with support vector machine (SVM) is proposed for localizing text region. After the image was decomposed by wavelet transform, SVM is used to extract the texture characteristics of text from the low-frequency vague image sub-space and high-frequency energy sub-space and determine what the current pixel belongs to, i. e., the text class or non-text class. Then, because there are possible noise and false text after SVM classification, the methods of morphological denoising and textural energy calculating are used to reprocess the results of SVM classification. The proposed method utilizes the characteristic that SVM suits high-dimension space so as to improve the efficiency of extracting text in addition to reducing the number of spatial sample inputs. Experimental results showed that the method can rapidly and effectively localize the text region in digital image.

Key words: text detection; texture analysis; wavelet transform; support vector machine (SVM); morphology

数字图像和视频中的文本通常能给人们提供简短而重要的信息,比如图像中出现的各种标牌,商标名称,人物介绍,视频中的字幕等等. 这些信息通常对图像的自动理解起着非常重要的作用. 只有准确的定位文本区域才能保证进一步的文本识别正确性,因此数字图像中文本的定位已成为一个广受关注的研究课题.

人们已经提出了许多文本定位方法,它们主

要分为基于连通元的方法和基于纹理的方法^[1-5]. 基于连通元的方法实现简单,但对于复杂背景情况和在文本处有噪声的图像情况下效果不佳. 基于纹理的方法虽然适合于复杂背景下的文本提取,但是通常都很耗时.

本文提出的文本检测方法也是基于纹理的方法,它把小波变换^[6]和支持向量机(SVM)相结合,不仅取得了很好的定位效果而且计算量也不

收稿日期: 2006-03-14

基金项目: 国家自然科学基金资助项目(60475036).

作者简介: 孙红星(1963-),男,辽宁铁岭人,东北大学博士研究生; 徐心和(1940-),男,河北临榆人,东北大学教授,博士生导师.

大,采用 SVM 进行分类,无需事先分析纹理特性,而是直接在训练中建立纹理模型,获得文本和非文本的纹理特征,构成区别它们的 SVM。图像经小波变换后,把高频和低频系数分开再合并,再送入 SVM 进行分类。因为文本区域都包含丰富的高频信息,这样相当于对图像的高频部分做了增强。最后的分类结果中可能含有噪声和虚假的文本,并不能作为最后的定位结果输出。这里利用图像形态学的去噪算法和计算平均垂直纹理能量的方法对由 SVM 分类得到的图像进行后处理。

另外由于下列原因,计算速度也有了很大程度的改善:由于在小波的高频和低频域上分别提取特征,使得需要计算的像素个数变成了原来的 1/4;提取特征的采样窗口是“米”字型的,它不仅比“口”字型窗口更符合文本的笔划特征,而且 SVM 输入向量的维数少了,训练和分类的计算速度必然会被加快;SVM 的训练采用“解靴算法”这样做不仅解决了 SVM 不适合大样本的训练集合的问题,也提高了训练速度。

1 SVM 提取文本的纹理特征

1.1 SVM 的基本理论

SVM 是统计机器学习理论(SLT)的核心内容,它基于 VC 维理论和结构风险最小化原理^[7],在很大程度上克服了传统机器学习方法中维数灾难、易陷入局部极小点以及过学习等难以克服的困难,具有良好的泛化能力,所以是一个比传统的神经网络(NN)更好地解决分类问题的数学工具。本文中,输入向量的维数是由分析窗口的大小决定的,为了提取文本的纹理属性,分析窗口至少要覆盖半个字符高度。如果选择神经网络作为分类机,将需要大量的神经元和权值,所以本文选择 SVM 作为分类工具。

给定一个训练样本集合 $\{x_i, y_i\}$,其中, $i = 1, 2, \dots, l$, $x_i \in \mathbf{R}^n$, $y_i \in \{-1, +1\}$ 。这里 x_i 是 n 维特征空间的第 i 个向量,即第 i 个样本点,它对应的类别是 y_i 。SVM 学习的目的是找到最优的分类超平面使得两类训练样本之间的分类间隔最大。因为输入的样本通常不是线性可分的,这个最优的分类超平面很难找到。SVM 的解决办法是,把输入向量映射到一个更高维的特征空间,然后在这个高维空间中寻找使两类样本分类间隔最大的超平面。

$$f(x) = \text{sign} \left(\sum_{i=1}^{N_{sv}} y_i K(x_i, x_j) + b \right), \quad (1)$$

其中, $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 是核函数, N_{sv}

是支持向量的个数。对于非线性支持向量机,常用的核函数有多项式核,高斯径向基函数,多层感知器等。本文采用的是高斯径向基函数,即

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2}. \quad (2)$$

利用支持向量机把输入图像分为文本类和非文本类,解出的支持向量也有两类,文本类支持向量和非文本类支持向量。因其具有适合于小样本的特点,用有限的样本,就可以得到性能优良的分

1.2 输入向量的组成

SVM 工作在变换后的小波域上。首先,对图像进行一级小波变换,得到 4 个频率子带,分别是 LL, HL, LH, HH。小波变换有很好的时间和频率定位性,各个子带上的对应位置分别代表了它在当前子带上的频率特性。把 LH, HL, HH 三个子带组成一个高频能量子带 HE。

$$HE_{i,j} = \sqrt{HL_{i,j}^2 + LH_{i,j}^2 + HH_{i,j}^2}, \quad (3)$$

其中, $\sqrt{\quad}$ 是大于 1 的系数,这相当于对图像的高频部分做了适当的增强^[8]。在 LL 和 HE 上用“米”字型 $M \times M$ 的分析窗口提取 SVM 的输入向量,如图 2b 所示。这样组成输入向量的好处是:样本点减少了,图像经一级小波变换后,样本点变为原来的 1/4。因为采用“米”字型窗口,这不仅更符合文本的笔划特征,而且输入向量的维数也减少了,由原来的 $4 \times M \times M$ 变成现在的 $8M - 6$ 。本文中选取 $M = 7$ 。对图像高频部分做了增强。

1.3 SVM 的训练

SVM 适合少数样本的分类问题,因为当样本点的数量很大时,运行时需要大量的内存空间和运行时间。但希望的情况是训练样本的选择应能够涵盖整个的输入空间,也就是说,对于本文的分类任务,样本点选得越多, SVM 提取的特征就越完备。如何建立一个既比较完备又有代表性而且易于执行的训练数据集,是训练 SVM 一个重要问题。本文中借鉴了 Sun 和 Poggio 提出的“解靴算法”^[9]的思想,提出了一个寻找 SVM 训练所用的小样本的方法。

图像中的纹理模式被分为文本模式和非文本模式。SVM 输出的符号表示了分类的类别,其中 +1 代表文本类, -1 代表非文本类。训练 SVM 的过程如图 1 所示。首先,让所有的文本模式和部分的非文本模式组成一个初始训练集合。用这个初始训练集合先训练出一个初始 SVM。接着,把这个初始 SVM 应用到不含有文本的图像上,结果发现,有很多正的文本模式输出了,显然这是一

个错误的分类,应给予纠正.把这些错误的分类和本次训练得到的非文本支持向量取代初始训练集中非文本样本,再训练出一个 SVM.即训练这个新的 SVM 的样本集合包含 3 个部分,一是原来的完全的文本模式,二是解出的非文本类的支持向量,三是把上一次得到的 SVM 应用于不含文本图像的错分样本.考虑部分非文本样本的纹理特性有相似性,本文只随机选择了这些错分类的 15 % 加入到训练样本集中.用这个集合的样本训练出一个新的 SVM,再把它应用于非文本图像.这个过程反复的迭代,直到在不含文本的图像中没有错分类产生为止.最后将得到一个最具有代表性的训练集合,虽然它是部分的,但是它所代表的信息都是重要的.可以认为,用它来训练出的 SVM 和完全样本训练出来的 SVM 相比,效果只有很小的差异.

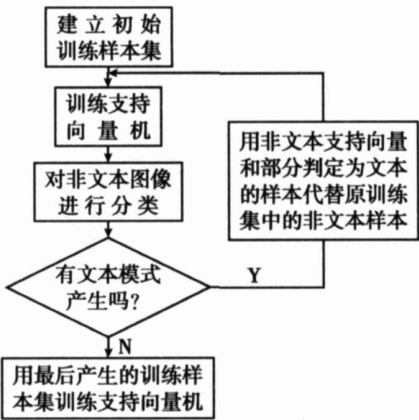


图 1 SVM 训练过程流程图
Fig. 1 SVM training process

2 文本区域的精确定位

用上一部分训练出来的 SVM 对图像上的文本和非文本进行分类,图 2c 是分类结果.从图中可以看出,得到的结果只是文本可能的区域,这里还有噪声和虚假文本的存在,还不能作为最终的文本区域检测结果输出.本文用数学形态学的去噪算法整理了 SVM 分类的结果.滤波的过程是先开启后闭合.

$$(A \circ B_1) \otimes B_2,$$
 (4)

其中, A 是要滤波的二值图像, B₁ 是开启算子,选择 5 × 2 矩形, B₂ 是闭合算子,选择 3 × 3 的“+”字型,处理后的结果如图 2d 所示.

经过以上的处理后,剩余的候选文本区域一般是以较大的连通区域出现,这些连通区域可能包含多行文本,为了便于进行字符分割和识别,需将多行文本区域划分成若干个单行文本.本文采

用的方法是在原始图像中计算文本区域中每行的平均垂直纹理能量^[10],逐行比较这些能量,当某行的平均垂直纹理能量相对于相邻的上下两行都小很多时,就将该行判为分割行.分割的结果如图 2e 所示.

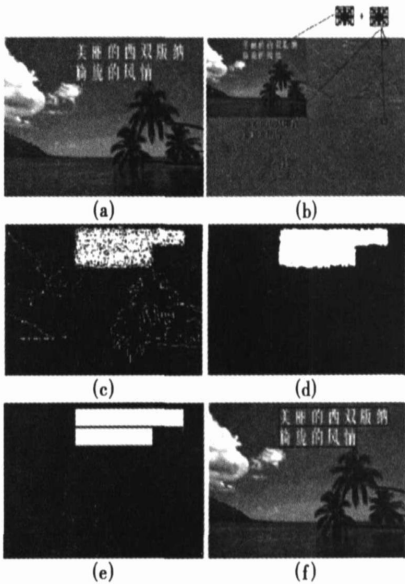


图 2 文本提取过程
Fig. 2 Extracting process of text
(a) —原始图像; (b) —小波变换图像;
(c) —SVM 分类结果; (d) —用闭合算法处理后的图像;
(e) —文本分割图像; (f) —文本定位结果.

实验中发现,文本矩形区域应满足以下条件:
整个文本区域的面积大于 80 个像素,小于整个图像的 1/6,并且矩形的高度在 7 ~ 30 之间.
矩形的宽度与高度之比大于 1.2.

3 算法步骤

为了检测不同大小的文本,可以把原始图像的尺寸逐渐缩小,对每种尺寸的图像都应用同样的算法.文本定位算法的步骤如下:

- 输入原始图像;
- 对输入图像进行小波变换,在小波域上提取 SVM 的输入向量,送 SVM 分类,得到文本类和非文本类的二值分类结果;
- 对得到的分类结果用式 (4) 进行形态学去噪;
- 用文献[10]的方法分割出文本行;
- 是否满足文本矩形区域的约束条件? 满足,则判断为文本输出,并把当前的图像上该位置的像素值全部置 0.
- 是否达到合适的缩小尺寸? 是则对图像进行缩小,并返回步骤 2,否则退出.

4 实验结果与分析

本文选用 300 幅 JPEG 图像作为实验数据,其中 100 幅用于训练 SVM,另外 200 幅用于做检验。这些图像中大部分都是包含文本的图像,还有一些不包含文本的图像,其中的文本包含中文、英文和数字。文本提取过程如图 2 所示。文本区域提取方法的性能由以下两个指标来评价,分别是误判率和检测率。

误判率 = 总的误判次数 / 文本块总数,

检测率 = 总的检测到的文本数 / 文本块总数。

其中,文本块总数 = 总的误判次数 + 总的检测到的文本数 + 总的漏检的文本数。

用本文提出的方法对 200 幅 JPEG 图像进行文本定位实验,从统计的结果来看,平均检测率达 97%,平均误判率达 1.7%,结果表明,本文的算法取得了较高的检测率,对于文本区域的边界定位比较准确,在背景比较复杂的情况下也能取得较好的提取效果。运用本文的方法可以较好地提取出图像中的水平文本,用类似的方法也可以提取出垂直方向的文本或任意角度的文本。图 3 是一些文本定位结果的实例,从图中可以看出,对图像中太小和太大的字符以及灰度值接近背景的字符,出现了漏检。

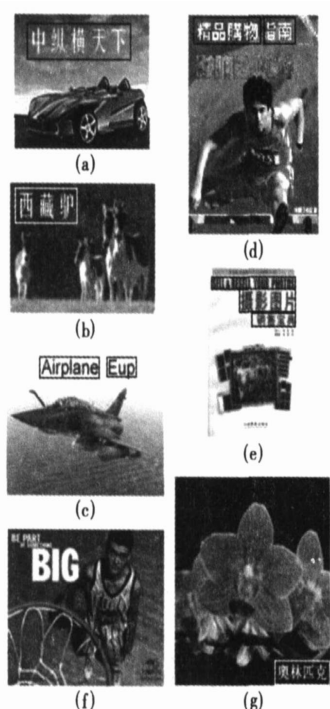


图 3 文本提取的几个实例

Fig. 3 Examples of text extraction

5 结 论

本文提出了一种基于纹理的检测图像上文本区域的方法。首先,用小波变换对输入图像进行处理,然后用 SVM 在小波变换域上分析文本的纹理特征。通过小波变换,不但减少了检测样本的数量而且降低了样本空间的维数,支持向量机的训练采用解靴算法,在大的样本集合中选择最具代表性的样本训练 SVM。最后利用形态学的闭合算子和计算纹理能量的方法对 SVM 的分类结果进行处理,准确地找到了文本的位置。

参考文献:

- [1] Jain A K, Yu B. Automatic text location in images and video frames[J]. *Pattern Recognition*, 1998, 31(12): 2055 - 2076.
- [2] Wu V, Manmatha R, Riseman E M. Textfinder: an automatic system to detect and recognize text in image[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 20(11): 1224 - 1229.
- [3] Jain A K, Zhong Y. Page segmentation using texture analysis[J]. *Pattern Recognition*, 1996, 29(5): 743 - 770.
- [4] Li H, Doermann D, Kia O. Automatic text detection and tracking in digital video[J]. *IEEE Trans Image Processing*, 2000, 9(1): 147 - 156.
- [5] Zhong Y, Karu K, Jain A K. Locating text in complex color image[J]. *Pattern Recognition*, 1995, 28(10): 1523 - 1536.
- [6] Mallat S. A theory for multiresolution signal decomposition: the wavelet representation[J]. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 1989, 11(7): 674 - 693.
- [7] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995: 50 - 75.
- [8] 曾鹏鑫, 么建石, 陈鹏, 等. 基于小波变换的图像增强算法[J]. 东北大学学报: 自然科学版, 2005, 26(6): 527 - 530. (Zeng Peng-xin, Yao Jian-shi, Chen Peng, et al. An approach to wavelet-based image enhancement algorithm[J]. *Journal of Northeastern University: Natural Science*, 2005, 26(6): 527 - 530.)
- [9] Sung K K, Poggio T. Example-based learning for view-based human face detection[J]. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 1998, 20(1): 39 - 51.
- [10] 王建, 周源华. 一种基于纹理能量的 JPEG 图像文本定位算法[J]. 上海交通大学学报, 2004, 38(9): 1492 - 1495. (Wang Jian, Zhou Yuan-hua. A text localization algorithm based on texture energy for JPEG images[J]. *Journal of Shanghai Jiaotong University*, 2004, 38(9): 1492 - 1495.)