第 33 卷 第 5 期 计 算 机 工 程 2007 年 3 月
Vol.33 No.5 Computer Engineering March 2007
·博士论文· 文章编号：1000—3428(2007)05—0026—03 文献标识码：A 中图分类号：TP391.43

基于小波模极大值的视频文本区域的提取

李雪妍，郭树旭，郜峰利
(吉林大学电子科学与工程学院，长春 130012)

摘 要：视频图像中包含着许多重要的文字信息。图像和视频文本信息的提取包括文本检测、定位、跟踪、提取、增强和识别等几个部分。将文本的检测、定位与提取，作为文本区域提取的整体来讨论。以文本的检测算法为重点研究对象，提出了应用小波模极大值算法来解决视频图像中文本区域的检测。实验表明，小波模极大值算法所得到的文本区域与其它算法相比具有更好的评价指标。

关键词：文本提取；小波模极大值；滑动窗口

Text Extraction in Video Based on Wavelet Modulus Maximum

LI Xueyan, GUO Shuxu, GAO Fengli
(College of Electronic Science and Engineering, Jilin University, Changchun 130012)

【Abstract】 The text in picture and video includes many useful information. Extraction of text information in picture and video includes text detection, location, tracing extraction, image enhancement and character recognition. Regarding the detection, location and extraction as a whole, the algorithm of the text detection is the important target of the research, it puts forward of the wavelet modulus maximum (WMM) as the function to solve the detection of the text in the videos. The experiment expresses that the result of the WMM is better than other algorithms.

【Key words】 Text extraction; Wavelet modulus maximum(WMM); Slip window

视频图像中包含着许多重要的文字信息，如街道名称、交通标识、字幕等，这些信息对视频资料的索引、压缩等方面有着重要参考价值。能够自动地进行自然环境下的文本理解是一个既令人兴奋又极具挑战性的任务。视频文字识别的关键步骤是如何从视频信号中检测到文本区域。

根据文本对象的存在形式可以将文本分为人工文本和场景文本。目前的研究主要集中于人工文本的研究，而场景文本的研究刚刚起步。根据文本信息提取研究的出发点不同，要解决的问题针对性各异，现在文本提取算法还没有一个通用的评价准则和标准数据库。Sato和Kanade[1]运用规则集检测文字区域，单纯依靠先验的规则进行检测，但是除去文本区域，还存在大量的噪声；张引[2]等研究了面向彩色图像和视频的文本提取算法，提出了一个全面作用在RGB颜色空间3 个分量上的彩色图像边缘检测新算法，这种算法在彩色空间里有较好的效果。小波在文本区域提取中的应用也有一些文献报道，文献[3]中将小波变换和色彩聚类结合使用，以此来得到文本区域。

本文对文本区域提取算法进行了对比，并提出了应用小波模极大值综合滑动窗口的视频文本区域检测算法来提取视频图像中文本的区域。

1 基于小波模极大值算法提取视频图像中文字轮廓

1.1 Lipschitz 指数和连续小波变换

函数的正则性一般用来描述其光滑程度。正则性越高，函数越光滑。通过 Lipschitz 指数 α 来度量函数的正则性。利用 Lipschitz 指数来刻画信号的奇异点。

定理

令 $\alpha>0$，且 n 是不大于 α 的最大整数，$\psi\in L^2(\mathbb{R})$ 满足 $(1+|\omega|)^{\alpha}\hat{\psi}(\omega)\in L^1(\mathbb{R})$ (即小波 $\psi\in C^{\alpha}$)，并且 ψ 具有 n 阶消失矩，即 $\int_{\mathbb{R}} x^k\psi(x)dx=0$，$k=0,1,...,n$，则存在常数 $C>0$，使得对 $\forall f\in L^2(\mathbb{R})\cap C^{\alpha}(\mathbb{R})$，式(1)成立：

$$\left|(W_{\psi}f(a,b))\right|\le C|a|^{\alpha+\frac{1}{2}},\quad (\forall a,b\in\mathbb{R}) \tag{1}$$

定理 1 说明信号 f 在 x_0 点的局部Lipschitz正则性依赖于信号 f 的小波系数的模 $|(W_{\psi}f(a,b))|$ 在 x_0 点的衰减性，这个衰减性可以由 $|(W_{\psi}f(a,b))|$ 的模极大值点来控制。如果 x_0 点是使小波变换的模 $|(W_{\psi}f(a,b))|$ 在尺度 a 上取极大值，则 $\left.\frac{\partial(W_{\psi}f(a,b))(x)}{\partial x}\right|_{x=x_0}=0$。所以只要在所有 a 尺度上找到严格的模极大值点，就能找到所有的边界点(奇异点)。

1.2 二维小波变换模极大值算法[4]

输入：离散的二维图像 $\{f(k,l)|k=0,1,...,K;l=0,1,...,L\}$

输出：图像的边界点(k, l)

(1)计算二维图像小波变换的模：$\{M_sf(k,l)|k=0,1,...,K;l=0,1,...,L\}$和梯度方向编码：$\{CodeA_sf(k,l)|k=0,1,...,K;l=0,1,...,L\}$

(2)确定阈值 T>0 对 k=0,1,…,K 和 l=0,1,…,L，如果

1) $|M_sf(k,l)|\ge T$

2) $|M_sf(k,l)|$ 沿梯度方向 $CodeA_sf(k,l)$ 达到模极大值

计算二维函数的导数时，要考虑方向问题，此时的边界点是指沿着梯度方向达到模极大值的点，这里梯度表示为

$$grad(f*\theta_s)(x,y)=i\frac{\partial}{\partial x}(f*\theta_s)(x,y)+j\frac{\partial}{\partial y}(f*\theta_s)(x,y) \tag{2}$$

首先要找寻那些像素点，这些点在梯度方向上达到模极

作者简介：李雪妍(1980－)，女，博士生，主研方向：图像处理；郭树旭，教授、博导；郜峰利，博士生

收稿日期：2006-04-21 **E-mail：**lxy001225@126.com

大值，如式(3)。

$$\left|grad(f*\theta_s)(x,y)=\sqrt{\left|\frac{\partial}{\partial x}(f*\theta_s)(x,y)\right|^2+\left|\frac{\partial}{\partial y}(f*\theta_s)(x,y)\right|^2}\right| \quad (3)$$

然后这些像素点组成了边界。

定义

$$\psi^1(x,y)=\frac{\partial\theta}{\partial x}(x,y) \qquad \psi^2(x,y)=\frac{\partial\theta}{\partial y}(x,y)$$

当 $\theta(x,y)$ 具有好的局部化特征时，即它必须满足下面的条件：

$$\int_{\square}\int_{\square}\psi^1(x,y)dxdy=\int_{\square}[\theta(+\infty,y)-\theta(-\infty,y)]dy=0 \quad (4)$$

$$\int_{\square}\int_{\square}\psi^2(x,y)dxdy=\int_{\square}[\theta(x,+\infty)-\theta(x,-\infty)]dx=0 \quad (5)$$

$\psi^1(x,y)$ 和 $\psi^2(x,y)$ 称为二维小波。

$$\left|grad(f*\theta_s)(x,y)=\frac{1}{s}\sqrt{\left|(f*\psi_s^1)(x,y)\right|^2+\left|(f*\psi_s^2)(x,y)\right|^2}\right| \quad (6)$$

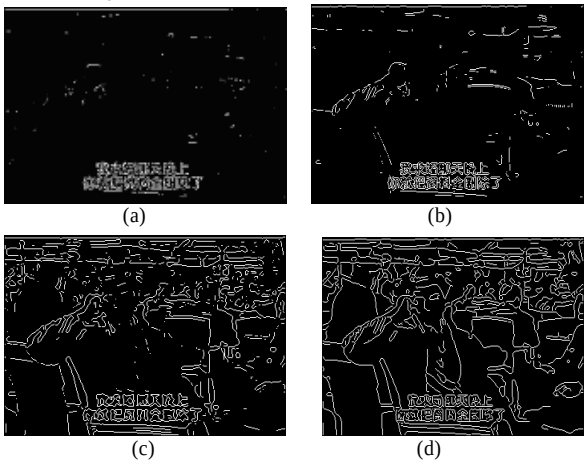
$$=\frac{1}{s}\sqrt{\left|W_s^{\psi^1}f(x,y)\right|^2+\left|W_s^{\psi^2}f(x,y)\right|^2}$$

$$\left|W_s^{\psi^i}f(x,y)\right|=(f*\psi^i)(x,y)=\int_{\square}\int_{\square}f(x-u,y-u)\psi_s^i(u,v)dudv \qquad i=1,2$$

$f(x,y)$ 相应的 ψ^1 和 ψ^2 小波变换的模定义如下：

$$M_sf(x,y)=\sqrt{\left|W_s^{\psi^1}f(x,y)\right|^2+\left|W_s^{\psi^2}f(x,y)\right|^2} \quad (7)$$

显然，$\left|grad(f*\theta_s)(x,y)\right|=\frac{1}{s}M_sf(x,y)$。计算一个光滑函数的导数沿梯度方向的模极大值等价于计算其小波变换的模极大值。图 1(a)为使用小波模极大值算法提取的视频图像中的文字轮廓，图 1(b)为使用 Robert 算子的结果，图 1(c)为使用 Canny 算子的结果，图 1(d)为使用 LOG 算子的结果。



(a) (b) (c) (d)

图 1 使用小波模极大值算法提取的图像文字轮廓检测结果

2 特征提取

特征向量的选取关系到整个系统的效果，是研究的重点。本文应用滑动窗口来扫描经过小波模极大值变换后得到的特征图像。对窗口中像素值对比了两种方法取得的特征，即直接使用原始特征和使用统计量作为特征。

2.1 原始特征向量

直接使用窗口范围内特征图像像素值作为特征，则得到了一个 *m* n* 维特征，窗口水平方向和竖直方向滑动的步长分别为 *steph*、*stepv*，将得到 *(M*N)/(steph*stepv)* 个特征向量，这里 *M*、*N* 为图像的宽度和高度。对于使用小波模极大值算法得到特征图像，实际得到 *m*n* 维特征。

2.2 统计特征

直接得到特征固然简便，但其数据量大。可以考虑使用内部的统计量来得到特征。在研究中使用了如下的 5 个统计量作为特征，这里 G 为对特征图像使用滑动窗口得到的矩阵，$\overline{G}$ 为此矩阵的均值：

(1)密度

$$Density=\sum_{i=1}^{m}\sum_{j=1}^{n}G(i,j) \quad (8)$$

(2)均值

$$Mean=\frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}G(i,j) \quad (9)$$

(3)二阶矩

$$moment2=\frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(G(i,j)-\overline{G}\right)^2 \quad (10)$$

(4)三阶矩

$$moment3=\frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(G(i,j)-\overline{G}\right)^3 \quad (11)$$

(5)标准差

$$std=\left(\frac{1}{m*n-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(G(i,j)-\overline{G}\right)^2\right)^{1/2} \quad (12)$$

对于滑动窗口得到的系数矩阵求取上述 5 个统计量，得到一个 5 维特征量。

2.3 可分类判据

特征提取与选择追求的是求出那些对分类识别最有效的特征，使最小维数特征空间中异类模式点相距较远，而同类模式点相距较近。人们提出了多种可分性判据，如基于几何距离的判据、基于概率密度的判据[5]等。文本提取作为一个两类分类问题，采用基于几何距离的可分性判据是合适的。

下面给出了所要用到的几个判据的定义：设 N 个模式 $\{x_1\}$ 分属 c 类，$\omega_i=\{x_k^{(i)},k=1,2,\ldots,N_i\}$，$i=1,2,\ldots,c$，定义如下的判据。

类内均方欧式距离

$$\overline{d}^2(\omega_i)=\frac{1}{N_i}\sum_{k=1}^{N_i}\left(x_k^{(i)}-m^{(i)}\right)^T\left(x_k^{(i)}-m^{(i)}\right) \quad (13)$$

两类间欧式距离

$$\overline{d}(\omega_i,\omega_j)=\frac{1}{N_iN_j}\sum_{k=i}^{N_i}\sum_{l=1}^{N_j}d\left(x_k^{(i)},x_l^{(j)}\right) \quad (14)$$

类内离差矩阵

$$S_{w_i}=\frac{1}{N_i}\sum_{k=1}^{N_i}\left(x_k^{(i)}-m^{(i)}\right)\left(x_k^{(i)}-m^{(i)}\right)^T \quad (15)$$

总的类内离差矩阵

$$S_w=\sum_{i=1}^{c}P_iS_{w_i}=\sum_{i=1}^{c}P_i\frac{1}{N_i}\sum_{k=1}^{N_i}(x_k^{(i)}-m^{(i)})(x_k^{(i)}-m^{(i)})^T \quad (16)$$

总的类间离差矩阵

$$S_B=\sum_{i=1}^{c}P_i(m^{(i)}-m)(m^{(i)}-m)^T \quad (17)$$

总体离差矩阵为

$$S_T=\sum_{i=1}^{c}(x_l-m)(x_l-m)^T \quad (18)$$

这里 P_i 为 ω_i 类的样本频率；$m^{(i)}$ 为 ω_i 类样本均值矢量；m 为总的样本均值矢量。利用特征空间中 S_W、S_B、S_T 式可以构造许多可分性判据，如

$$J_1=Tr[S_W+S_B] \quad (19)$$

$$J_2=Tr[S_W^{-1}S_B] \quad (20)$$

2.4 窗口滑动步长的影响

对于大小为 $m*n$ 的滑动窗口其滑动的步长将会影响到计算的速度和分类的准确性，当步长为 1 时，得到一个与原图像大小相同的特征图像，原图像中的每个点将参加到 $m*n$ 个窗口的分类中去；当步长增加到与窗口的大小一样时意味着将原始图像均匀的分布到特征图像中，这将会导致原始图像中的每个点只由一个窗口的分类决定，此时算法的运算量将是步长为 1 时的 $1/(m*n)$。为了衡量步长对分类结果的影响，本文比较了不同步长下分类情况，表 1 给出了使用 8*8 窗口。从表中可见步长的变化对于分类的结果影响并不十分的明显，为了兼顾效率和正确率，通过实验，认为步长为窗口大小的一半是一个较好选择。

表 1 不同步长的分类结果

步长	文本正确率
2	88.22%
4	91.17%

3 实验结果和结论

分别使用了 Robert、Canny、LOG 边缘算子与小波模极大值算法得到特征图像。为了从本质上比较图像变换对文本提取的影响，计算在特征图像中使用 8*8 的滑动窗口取得 64 维原始特征的可分类判据。因为 8*8 的窗口在图上滑动的过程中，窗口有可能落在文本和背景的交界位置，使得提取的 64 维特征既包含文本部分又包含背景部分，按照窗口中文本和背景的面积来进行归类，文本面积大的即认为是文本，反之亦然。

表 2 原始特征判据

	文本区域类内距离	背景区域类类距离	J1	J2
Robert	9.77	0.76	6.09	1.39
Canny	9.79	4.51	7.48	0.98
LOG	9.91	3.02	7.23	0.96
DB2	3.93e+004	1.19e+003	1.55e+004	1.00

表 2 给出了提到的算法的判据。为了降低特征维数，使用所给出的 5 个统计量作为特征，表 3 是使用统计特征时的分类判据，从中可以看出统计特征也可以将文本和背景在一定程度上区分出来。

表 3 5 维统计特征判据

	文本区域类内距离	背景区域类间距离	类间距离	*J1*	*J2*
Robert	53.42	8.75	67.23	Inf	1.05
Canny	27.35	27.58	48.90	Inf	0.92
Log	21.38	21.75	43.07	Inf	0.87
DB2	8.66e+007	2.18e+005	5.53e+008	1.90	0.17

为了验证各种特征的分类效果，把样本图像提取的特征送入分类器中，比较分类的准确率来评价特征的优劣。对 700 张图片进行测试，其中 130 张图片作为训练样本。表 4 给出了使用两层的前馈神经网络分类的结果。根据各步骤的分析结果，使用小波模极大值算法作为提取文本区域的整体算法的效果最佳。

表 4 使用统计特征的神经网络分类结果

Robert	Canny	LOG	DB2
82.97%	83.54%	86.65%	91.03%

参考文献

1 Sato T, Kanade T, Hughes E, et al. Video OCR: Indexing Digital News Library by Recognition of Superimposed Caption[J]. Multimedia System, 1999, 7(5): 385-395.

2 Zhang Yin, Pan Yuehe. A New Approach for Text Extraction from Color Image and Video[J]. Journal of Computer-aided Design & Computer Praphics, 2002, 14(1): 36-39.

3 黄晓东，周源华. 用小波变换及颜色聚类提取的视频图像内中文字幕[J]. 计算机工程, 2003, 29(1): 43-44.

4 唐远炎，王 玲. 小波分析与文本文字识别[M]. 北京: 科学出版社, 2005.

5 孙即祥. 模式识别中的特征提取与计算机视觉不变量[M]. 北京: 国防工业出版社, 2001.

~~~~~~

（上接第 22 页）

由图 1 可以看出，当固定 σ^2,ε 时，预测样本的平均相对误差随 C 的增加而减小，当增加到最优参数附近后逐渐变化平稳；由图 2 可见，当 ε 增加时，预测样本的平均相对误差在小于 0.19 时变化相对平稳，大于 0.19 后，开始迅速增加；图 3 中，随 σ^2 的增加，预测样本的平均相对误差变化相对平稳。从预测样本的精度变化曲线可以看出，3 个参数在最优值附近，皆有较大的容许变化范围，而精度的变动相对较小，这与文献[5]中的研究结果相类似。

5 结束语

本文综合考虑了影响路段交通量的 3 种因素，首次应用了时空联合预测模型和支持向量回归方法对交通量进行了预测研究，通过对 LOO 误差上界求最小化值的变尺度算法选取支持向量回归中的最优参数。研究结果表明，通过对平均相对误差的比较，该方法对路段交通量的预测精度较高，显示了该方法的良好的预测性能。

参考文献

1 贺国光, 李 宇, 马寿峰. 基于数学模型的短时交通流预测方法探讨[J]. 系统工程理论与实践, 2000, 20(12): 51-56.

2 陈淑燕, 王 炜. 交通量的灰色神经网络预测方法[J]. 东南大学学报(自然科学版), 2004, 34(7): 541-544.

3 Friedrichs F, Igel C. Evolutionary Tuning of Multiple SVM Parameters[J]. Neurocomputing, 2005, 64(1): 107-117.

4 Chapelle O, Vapnik V, Bousquet O, et al. Choosing Multiple Parameters for Support Vector Machines[J]. Machine Learning, 2002, 46(1): 131-159.

5 Chang Mingwei, Lin Chihjen. Leave-one-out Bounds for Support Vector Regression Model Selection[J]. Neural Computation, 2005, 17(5): 1188-1222.

6 Kim K J. Financial Time Series Forecasting Using Support Vector Machines[J]. Neurocomputing, 2003, 55(3): 307-319.

7 Hobeika A G, Kim C K. Traffic-flow-prediction Systems Based on Up Stream Traffic[C]//Proc. of Vehical Navigation & Information System Conference. 1994.

8 Chang C C, Lin Chihjen. LIBSVM: a Library For Support Vector Machines[Z]. 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm.
~~~~~~