

基于局部能量特征的视频字幕分割

周 军 徐 奕 周源华

(上海交通大学图象通信与信息处理研究所, 上海 200030)

摘 要 视频字幕能够给人们提供高度浓缩的与视频内容相关的信息,但字幕通常均以图象方式存在于视频、图象中,为了提取视频字幕,提出了一种基于局部能量的视频字幕分割方法,该方法利用局部能量与图象中边缘、轮廓特征之间的对应关系进行字幕的自动分割.为了快速地进行局部能量的计算,在分析局部能量模型基础上,将局部能量计算方法进行了推广,即先通过选用具有对称性的双正交小波基及 Hilbert 变换,构造了适于局部能量计算的 90° 相移滤波器,然后通过多分辨率小波变换实现了信号的多分辨率频带分割,并在此基础上计算局部能量.实验结果表明,该方法可获得较好的分割效果.

关键词 视频字幕 局部能量 小波变换 Hilbert 变换

中图法分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2002)11-1134-05

Video Captions Segmentation Based on Local Energy Features

ZHOU Jun, XU Yi, ZHOU Yuan-hua

(Inst. Image Comm. & Info. Proc., Shanghai Jiao Tong Univ., Shanghai 200030)

Abstract Caption text presented in the video plays an important role in video retrieval and browser as it provides highly condensed information about the contents of the video. The caption text only can be used after it is extracted from video. According to Morrone's phase congruency theory, image features such as edges, shadows and bars always occur at points of maximum phase congruency, and the maxima of local energy occur at points of maximum phase congruency. Based on this theory, a video caption text segmentation approach is presented in this paper. Instead of Morrone's approach in calculating of local energy, we extended Morrone's approach through constructing a quadrature pair filter from a biorthogonal wavelet by Hilbert transform. The local energy is then calculated from the multiresolution decomposed image in octave bands. According the relationship between local energy and caption text's edge features, we then segment the image using local energy projection. This is the first step of caption segmentation. After that, colour segmentation can be applied to the first segmentation results. The experiments in this paper show that this approach can achieve good caption region segmentation results.

Keywords Video captions, Local energy, Wavelet transform, Hilbert transform

0 引 言

目前,图象和视频数据已广泛应用于教育、娱乐以及多媒体等领域,因此数据海量增加,而对于不断增长的视频数据,则要求有一种高效的存储、传输、检索和游览的管理方法.尽管在图象和视频数据的存储、传输时,可以不考虑视频数据的内容,但对大量图象、视频数据进行检索、浏览时,则必需考虑

与视频数据相关的内容,即需利用所提取到的、反映图象、视频数据内容的信息进行视频检索^[1,2].

大家知道,很多图象及视频数据中均含有人工添加的字幕信息,如海报、电影、新闻节目等均含有人工添加的字幕,由于这些人工添加的字幕信息与视频内容密切相关,因此如果能够分割提取,并识别出这些字幕信息,那么它们就可以为视频检索提供重要检索依据.如目前已有的 OCR 系统就基本解决字符识别问题,但由于其在进行 OCR 识别

基金项目:国家自然科学基金项目(69905003)

收稿日期:2001-11-01; 改回日期:2002-05-13

前,必须将图象中的文本与背景相分离,因此必需对视频字幕进行分割^[3]。

图象分割是图象处理领域中的一个基本问题,它是图象模式识别中的一个关键技术,也是目标特征提取、识别的基础。在进行图象分割时,根据所处理对象的特点不同,可以采用不同的方法进行,如文献[4]采用神经网络进行灰度图象的自适应分割。对于图象和视频中的字幕来说,由于它与背景中的自然景物有着明显的差别,且一般字幕中,字符排列有序,字体颜色基本相同,最重要的是字符本身与背景之间存在着明显的边缘轮廓,因此在图象分割时,如能有效利用视频字幕的上述特点,对于最终字幕的分割效果是十分重要的。目前许多学者研究了视频数据中字幕的提取方法,它们基本上可分为基于纹理的和基于结构的两类分割方法^[5]。其中,基于纹理的分割方法主要利用文字区域所具有的特殊纹理,而基于结构的分割方法,则根据文字的轮廓几何排列、字符的同色等特征来进行分割。

1 基于小波变换的局部能量计算

由视频中字幕的特征及字符与自然景物之间的区别可知,若在图象中有人工添加字幕的区域,则该区域内的局部能量较其他区域高且集中。由此可见,局部能量可以为视频字幕的分割提供依据。

1.1 局部能量模型

Morrone 等从 Fourier 级数定理出发,分析了各类信号傅氏级数在各点处的相位特点,提出了相位叠合模型(PC:Phase Congruency)^[6],即将信号 $f(x)$ 展开为如下傅氏级数

$$f(x) = \sum_n A_n \cos(n\omega x + \varphi_n - \theta) \quad (1)$$

并定义相位叠合函数 $F_{PC}(x)$ 为

$$F_{PC}(x) = \frac{\max_{\pi \in [0, 2\pi]} \sum_n A_n \cos(n\omega x + \varphi_n - \theta)}{\sum_n A_n} \quad (2)$$

在该模型中,由于直接利用式(2)计算各点处的相位叠合很困难,因此 Morrone 利用局部能量特征进行相位叠合计算,即

$$E(x) = F_{PC}(x) \sum_n A_n \quad (3)$$

其中,局部能量为

$$E(x) = \sqrt{\sum_n (f(x) * M_n^e)^2 + \sum_n (f(x) * M_n^o)^2} \quad (4)$$

其中, M_n^e 和 M_n^o 为频带 n 下的偶对称和奇对称带通滤波器, $*$ 为卷积运算。

在该模型中,相位叠合可通过局部能量进行计算,而局部能量的计算则需先对信号进行带通滤波,以得到信号的局部频率特征。一般采用 Gabor 滤波器进行带通滤波,该滤波器实质上是一个加 Gaussian 窗的 Fourier 变换子,其实部和虚部之间互为 Hilbert 变换,且实部为偶对称滤波器,虚部为奇对称滤波器。经该滤波器滤波后,其所得结果的实部和虚部之间也互为 Hilbert 变换。

根据 Hilbert 变换的特点及其与实信号解析构造之间的关系,可将局部能量计算方法进一步推广,即先利用具有线性相位及频带划分分解特性的滤波器来对实信号进行滤波分解,以便得到各通带下的局部频率特征,然后通过 Hilbert 变换来构造该分解信号的解析信号,并由该解析信号计算局部能量。

1.2 基于小波变换的局部能量计算

小波变换不仅定义了一个函数 $\psi(x) \in L^2(\mathbf{R})$, 而且该函数的一个伸缩、平移系 $\{\psi_{a,b}(x)\}$ 是 $L_2(\mathbf{R})$ 的一个规范正交基^[7], 该规范正交小波基不仅可以信号分解为一系列具有不同空间分辨率、不同频率特性和方向特性的子带信号,而且这些子带信号具有良好的局部性时域和频域特性,同时小波变换还克服了 Fourier 分析在处理非平稳信号时不能分析信号局部特征的局限性。在定义了小波变换的基础上, Mallat 提出了信号的多分辨率分析方法^[8], 该方法发展了子带编码理论,它以级联方式来使用双子带编码,并自底向上建立了小波变换。通过塔式分解就可以得到信号的一个多分辨率描述。由此可见,小波变换具有多分辨率分析的特点,它可通过级联的塔式分解来得到信号的一个倍频频带划分。

由于除 Harr 基外,其余满足紧支正交规范化小波条件的函数均缺乏对称性,即正交小波除 Harr 基外均无线性相位特性,为此需引入具有紧支集对称(奇或偶)的双正交小波。由于这种双正交小波具有对称性,因此它天生具有线性相位特点。由局部能量计算的推广方法及双正交小波基的特点可知,双正交小波基可用于与能量计算相关的多分辨率带通滤波,但在推广方法中,还需进一步对滤波结果进行 Hilbert 变换,以构造相应的解析信号。由于 Hilbert 变换是一个不定积分,且直接对滤波结果进行 Hilbert 变换来构造解析信号较困难,因此可根据解析信号的构造流程(图 1(a))来进行变换处理,即通

过 Hilbert 变换来构造一个解析的复小波 $\tilde{\psi}(x)$ (图 1(b))

$$\tilde{\psi}(x) = \psi(x) + j\psi_H(x) \quad (5)$$

其中, $\psi_H(x)$ 为 $\psi(x)$ 的 Hilbert 变换对, 在此基础上, 通过该复小波可直接进行信号的带通滤波(图 1(c)), 所得结果就是解析信号。

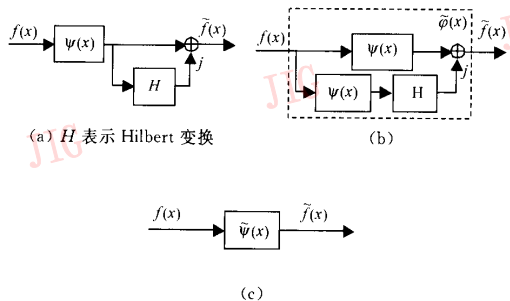


图 1 信号滤波及解析构造的等价变换

利用式(5)所定义的复小波 $\tilde{\psi}(x)$, 并采用 Mallat 分解算法, 经过变换后就可以得到一个多分辨率的解析信号。由于在多分辨率分解过程中, Mallat 算法只取用一个滤波器组, 并且是通过亚抽样频域变换技术^[9]来进行级联变换, 以保证变换后数据的非冗余性, 因此分解后, 各频带信号的采样频率不同, 在计算局部能量 E 时, 需进行频带逆变换, 以保证不同频带具有相同的采样率, 为此可通过插零来进行频带压缩, 并由尺度滤波器滤波来得到与下一级频带采样频率相同的解析信号。这样, 就可以由式(4)来计算信号的局部能量。

根据视频中字幕的特征及字符与自然景物之间的区别可知, 在图象中, 由于人工添加字幕区域内的局部能量较其他区域高, 且集中, 因此局部能量可以为视频字幕的分割提供依据。

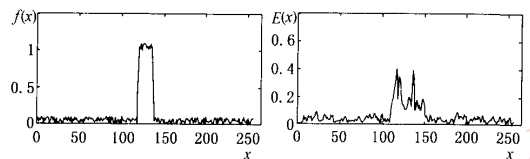
2 视频字幕特点及分割

在视频数据中, 文字字幕具有一些特点, 如其可能存在于视频中的任何位置, 又如有些字幕是描述视频的重要信息, 而有些则不是。根据文字字幕的特征, 视频中的字幕基本上可以分为场景中的字幕和人工添加的字幕两类。对于场景中的字幕, 它们是作为视频内容的一部分而存在的, 在视频中是随机出现的; 对于人工添加的字幕, 这些字幕则是在视频数据的后处理过程中, 根据视频内容添加进去的, 通常是通过视频字幕机添加的, 由于这些字幕所描述的

与视频内容密切相关, 因此对视频数据中人工添加的字幕信息进行提取, 将有助于视频数据内容的分析、检索。

一般而言, 人工添加的字幕信息具有以下一些特征^[3]: ①字幕中的文字均处于前景位置, 不存在遮挡问题; ②文字基本上是同色的; ③文字与背景间存在着较强的对比度; ④文字的大小有限, 一个字符不可能占据整个屏幕; ⑤同一帧中的文字均按水平或垂直对齐等; ⑥同一行或列中字符大小基本相同。提取视频字幕时, 利用图象、视频字幕中存在的这些特征将有利于进行人工添加字幕的分割。

对于字幕中的文字来说, 它们都是由一些基本笔画构成的, 且每一笔画都有一定的粗细, 而每一笔画的粗细与所采用的字体有关。当它们被添加到视频数据中后, 字体中每一笔画的截面可由一梯形函数进行描述, 其梯形顶部表示文字的前景色亮度, 底部表示背景色亮度, 两边则反映了文字与背景图象间存在的边缘轮廓强度, 而且一般视频字幕中的文字大小有限, 故梯形函数的顶部宽度有限, 根据局部能量模型可知, 在梯形的顶部端点处具有最大的局部能量(图 2)。对于任一字符来说, 它都由一些基本的笔画构成, 且由于画面中含有字符区域能量较高, 因此利用这一特点, 就可以通过局部能量来进行字符区域分割。



(a) 加噪梯形波形 (b) 基于小波变换得到的局部能量

图 2 加噪梯形函数及其局部能量

此外, 由于字幕中的文字呈规则排列, 一般为水平或垂直方向排列, 因此通过对局部能量进行水平或垂直方向投影就可以初步提取出含有字符的区域, 在此基础上, 还可以进一步利用颜色特征分割出视频字幕中的字符。

3 实验结果

前面主要介绍了基于小波多分辨率分解下的局部能量计算方法, 并在此基础上, 结合视频字幕的特点及视频中人工添加字幕区域局部能量之间的对应关系, 论述了基于局部能量的字幕分割方法, 为验证本算法

的实用性,在此,对电视剧、电视新闻以及各类 VCD 等视频中含有字幕的图象,进行了字幕提取分割实验.

图 3(a)为从新闻节目中截取的含有人工添加字幕的原始视频图象,在该图象中有两处字幕信息;图 3(b)为通过小波变换计算得到的局部能量结果,从该结果可以看出,视频中的字幕区域,其局部能量相应较高,同时图象中含有边缘轮廓处的局部能量

也相应较高;图 3(c)为采用本文的局部能量投影方法得到的初步分割区域.在该结果中,两处字幕均能分割出来,但下方一处的短字符未提取到,这主要是由于图象本身存在的边缘轮廓特征会对投影分割方法产生干扰,同时也由于字符本身特征不是特别明显所致.图 3(d)为采用阈值化分割后得到的二值化字幕信息,该结果可直接用于 OCR 进行字符识别.

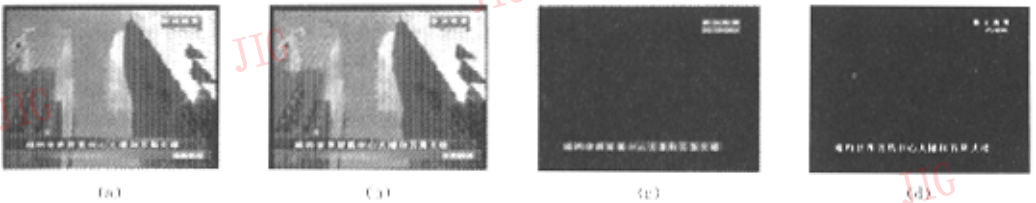


图 3 新闻视频图象中字幕的分割

对于 VCD 这一类视频,其上面添加的字幕一般较明显,对于宽银幕格式视频,字幕通常在视频下方的黑色空白区域,这种情况采用局部能量可以得到很好的粗分割效果(图 4(a),(b)).对于一般格式的视频,采用能量分割方法也能得到很好的分割效果(图 4(c),(d)).此外,根据字幕信息长度进行字幕识别时,视频中的字幕可能是一行,也可能是连续多行,一般以两行居多.对于多行字幕情况,根据局部能量基本上也可以将多行同时分割出来(图 4

(e),(f)),但在利用局部能量进行投影判别时,可能会将较长一行的两头截掉,这主要还是由于投影分割造成的,故还需寻找一种适合于局部能量的分割方法(图 4(g),(h)).

在单行字幕分割中,对于像“一”这一类单笔字符,由于该类字符的图象特征较少,相应的局部能量也就较少,特别是在字符颜色与背景颜色对比度较小的情况下,因此往往不能进行成功分割(图 4(i),(j)).



图 4 VCD 视频中字幕的分割结果

由于基于局部能量的字幕分割方法是利用局部能量所反映的字符边缘、轮廓特征来进行字符提取的,因此该方法对于字幕中字符颜色存在变化的情况同样适用,且具有较高的鲁棒性,但在此情况下,根据区域分割结果,利用颜色进行字符的提取就较困难.在所选用的视频图象(图 5(a))中,由于字体

表面颜色具有明显的光照特征,因此利用局部能量最后可得到较好的分割结果(图 5(b)).此外,对于一串字符中含有不同颜色的字体也可以正确识别,如图 5(c),(d)的“经济半小时”中的“半”字与其他字符的颜色不同,为红色.由此可见,基于局部能量的分割方法与字符颜色无关.



图5 视频中字体颜色变化条件下的分割结果

此外,在从50多幅电视、VCD影碟等媒体中截取的视频图象字符进行的分割中,有88%的字幕信息能够成功分割。另外12%的字符没有正确分割,其原因部分是因为字符太小或是简笔画字符,还有一部分是因字符较模糊或与背景对比度较小所致。此外,在实验中,还出现一些误分割现象,主要是因为分割区域中边缘、轮廓较多,而含有较高的局部能量信息所致,但这些部分的局部能量分布特征与字符区域的局部能量分布特征有明显区别,相信通过进一步的判别可以消除这部分误分割区域。

4 结论

综上所述,局部能量与相位叠合模型相同,都能用于图象特征描述。若根据局部能量计算的基本原则,并进一步将之推广,即利用具有线性相位特性的双正交小波基,并通过 Hilbert 变换来构造一个具有 90° 相移特性的复小波基,则该复小波基在局部能量求取过程中同样满足频带分割滤波的要求,而且在频带的分割滤波中,还可以应用小波变换所特有的多分辨率分析方法来进行分割。

根据视频中人工添加字幕及字符图象的特点,由于视频中人工添加字幕区域的局部能量通常较大,因此局部能量很适合于视频中人工添加字幕的提取,并且具有较强的抗干扰性,如字体表面颜色有变化等情况,另外还能从整个图象中得到较好的定位效果,进而可为下一步的字符提取、识别提供基础,实验结果也证明了这是一种相当不错的字幕提取方法。

参考文献

- 1 Bertini M, Bimbo A D, Pala P. Content-based indexing and retrieval of TV news[J]. Pattern Recognition Letters, 2001, 22: 503~516.

- 2 金红. 基于内容检索和视频分析技术的研究[D]. 上海:上海交通大学图象通信与信息处理研究所, 2000, 7.
- 3 Lienhart R, Effelsberg W. Automatic text segmentation and text recognition for video indexing[J]. Multimedia Systems, 2000, 8(1): 69~81.
- 4 叶芾芸, 戚飞虎, 蒋隽. 基于选择性多分辨率 Kohonen 网络的自适应灰度图象分割方法[J]. 红外与毫米波学报, 1998, 17(1): 48~53.
- 5 Zhong Y, Karu K, Jain A K. Locating text in complex color images[J]. Pattern Recognition, 1995, 28(10): 1523~1536.
- 6 Morrone M C, Owens R A. Feature detection from local energy[J]. Pattern Recognition Letters, 1987, 6: 303~313.
- 7 Daubechies I. Orthonormal bases of compactly supported wavelets[J]. Comm. Pure and Appl. Math., 1988, 41(11): 909~996.
- 8 Mallat S G. Multiresolution signal decomposition: the wavelet representation[J]. IEEE Trans. on PAMI, 1989, 11: 674~693.
- 9 Crochiere R E, Rabiner L R. Interpolation and decimation of digital signals—a tutorial review[J]. Proc. of the IEEE, 1981, 69(3): 300~331.



周 军 1997年获上海交通大学博士学位,现为上海交通大学图象通信与信息处理研究所副教授。当前主要的研究领域为图象通信、计算机视觉、小波变换理论及应用、软件中间件技术。



徐 奕 上海交通大学图象通信与信息处理研究所攻读博士研究生,专业为通信与信息系统。目前主要从事图象处理、计算机视觉及小波变换理论及应用方面的研究。

周源华 上海交通大学图象通信与信息处理研究所教授,博士生导师。主要研究领域为图象处理与通信、分形及小波变换理论等。