基于 DCT 压缩域的图象字符定位

黄祥林 沈兰荪
(北京工业大学信号与信息处理研究室, 北京 100022)

摘 要 为了能够利用图象中所含的文字信息来进行图象的快速高效浏览检查,其中,快速字符定位是很重要的工作,为此设计了一种直接在图象压缩域中进行字符定位的方法. 该方法主要是利用图象中字符纹理所具有的方向性特点,首先直接在 DCT 域中提取字符的横向、竖向、斜向纹理的方向信息,然后根据各自的阈值把字符区域从图象背景中分割出来. 在处理过程中,用形态滤波的方法可有效地消除噪音点. 该算法可直接处理 JPEG、MPEG 等以 DCT 为编码基础的压缩数据,仅需少量的解码过程(Huffman 解码)即可完成字符定位,因此要处理的数据量较少,用该算法既提高了处理速度,又减少了对计算机资源的需求. 试验结果表明,此方法具有较高的准确率.

关键词 字符定位 DCT 变换 压缩域处理 形态滤波
中图法分类号: TP391.4 TN919.81 **文献标识码**: A **文章编号**: 1006-8961(2002)01-0022-05

Character-Localization in DCT-Compressed Domain

HUANG Xiang-lin, SHEN Lan-sun
(*Signal and Information Processing Lab*, *Beijing Polytechnic University*, *Beijing* 100022)

Abstract Segmenting character-regions in an image is very important because these characters contain clear clues of retrieving and browsing images from video/image-databases efficiently and effectively. In this paper, We propose a method to locate character-regions of video/image in DCT-compressed domain directly. With the distinguishing characteristics of character's texture (such as horizontal lines, vertical lines, or slant lines in a character) that can be extracted directly in DCT-compressed domain, the character-regions are segmented from their backgrounds quickly, and the image-noises rising during the processing period can be removed by morphological filter. With this method, the compressed bit-streams, which are encoded by DCT-based encoding algorithm such as JPEG, MPEG-1/2, etc., can be processed directly to locate the character-regions in image, just a very small amount of decoding is required (Huffman-decoding only). So, the amount of data which want to process is smaller, the processing speed is faster and the demand of computer memory is less. The experimental results show that the correct-localization rate of this algorithm is higher.

Keywords Character-localization, DCT, Compressed-domain processing, Morphological filtering

0 前 言

如何对大量多媒体数据进行快速高效地浏览、检索,是当前多媒体领域中的一个重要研究方向,而基于多媒体内容的操作是其中的关键技术之一. 对图象(包括视频)来说,由于其所含的文字信息在一定程度上反映了本幅图象(或一小段视频)的部分重要内容,对帮助人们理解图象的内容、检索相关图象有着重要的作用,因此,自动定位图象中的字符区域,并抽取这些文字信息,对于图象理解、检索查询是很重要的.

目前已有许多对图象中的字符进行自动定位的方法[1~5]. 这些方法主要分如下两大类:一类是基于字符纹理特征的定位方法,由于字符区的纹理特征与背景的纹理特征不同,据此可分离字符区与背景

基金项目:863-317 计划资助(863-317-01-10-99)
收稿日期:2000-07-27;改回日期:2001-01-20

区;另一类是根据字符的结构特点来定位的方法,其主要是根据字符的结构特点(字符边缘的几何特点、字符颜色的一致性等)来提取字符信息,以便从图象背景中分割出字符区域,但由于这些方法基本上都是在原象素域操作的,当处理 JPEG、MPEG 等压缩数据时,首先需要对压缩数据进行全部解码,在得到原始象素域数据后,才能再实施字符的定位处理,这样它们基本上没有利用压缩数据的特性,操作时间也比较长,不利于实时处理.

鉴于视频序列图象的特殊性,文献[4]针对 MPEG 数据提出了一种基于压缩域的字符定位方法,其基本思想为:若某帧含有字符区,则其相邻帧的对应区域与此必有较大偏差,且这种偏差是通过重建低分辨率的图象来检测的(使用 DC 系数或再加上几个 AC 系数);而文献[1]提出的方法是,首先直接检测 MPEG 数据中的 P 帧或 B 帧,然后根据字符区出现与消失时的大偏差来定位,但由于这两种方法都是根据 MPEG 数据中相邻两帧的大偏差来检测字符区的,并未充分利用字符本身的特征信息,若字符是通过渐变的方式出现和消失,则检测失效,并且这两种方法均不能应用于 JPEG 图象数据上,为此,本文提出了一种直接在 DCT 域进行字符定位的方法,其对于 JPEG、MPEG 压缩数据只需进行简单的 Huffman 解码,就可直接在 DCT 压缩域提取字符区的纹理特性,这就充分利用了 DCT 域数据的特点,因而处理时间大大缩短.

1　DCT 域的字符定位方法

图象经 DCT 变换后所得到的系数位置及其幅值所反映的是该变换图象的空间频率及其能量[6].对于 8×8 的图象子块 $y[i,j](0\leq i,j<8)$,经 DCT 变换后,所得到的 64 个系数 $Y[u,v](0\leq u,v<8)$就反映了这个图象子块的频率能量及其分布情况

$$Y[u,v]=\frac{1}{4}\sum_i\sum_j C(i,u)C(j,v)y[i,j] \quad (1)$$

其中,$C(i,u)=A(u)\cos\dfrac{(2i+1)u\pi}{16}$

$$A(u)=\begin{cases}\dfrac{1}{\sqrt{2}} & u=0\\ 1 & u\neq0\end{cases}$$

式(1)中,u、v 分别表示图象子块的水平、垂直频率,在实际的编码过程中,必须要对系数 $Y[u,v]$ 的值进行量化.这里,若将 8×8 的 DCT 系数划分为 0 区、1 区、2 区、3 区等 4 个区域,则每个区域代表不同纹理的方向:0 区表示的是直流分量(即 8×8 子块的平均值),1 区表示的是竖向纹理(即水平方向的频率变化),2 区表示的是斜向纹理(即斜向的频率变化),3 区表示的是横向纹理(即竖直方向的频率变化),如图 1 所示.

0	1	1	1	1	1	1	1
3	2	1	1	1	1	1	1
3	3	2	2	1	1	1	1
3	3	2	2	2	1	1	1
3	3	3	2	2	2	2	1
3	3	3	3	2	2	2	2
3	3	3	3	2	2	2	2
3	3	3	3	3	2	2	2

图 1　DCT 块 4 个区域的具体位置

由于字符具有特殊的线条结构,它基本上可归为横向、竖向、斜向的线条组合(统计特性)(如汉字、英文字符等).在图象中,字符的这些线条主要显示出特殊的纹理特点,其灰度(颜色)与背景相差较大,即边缘变化较剧烈,且表现出明显的竖向、斜向、横向纹理特征;而在 DCT 域中则主要表现为在图 1 的 1、2、3 区的中高频部分的系数值较大,即竖向线条的变换系数主要集中于 1 区,斜向线条的变换系数主要集中于 2 区(1 区与 3 区也有一定的分布),横向线条的变换系数主要集中于 3 区,这是图象所含字符在 DCT 域中所表现出的纹理特征,只要对这些纹理特征进行适当组织,就能分割出这些字符在图象中的具体位置.

对于每个 DCT 块中的 3 个区域(图 1 中的 1、2、3 区),实验时,只考虑它的中高频部分,并且把其频率下限视为可调参数.如果对每个区域求其能量值(这里定义为系数绝对值之和),那么这 3 个能量值分别表示横向、斜向、竖向的频率变化程度,并将其作为字符的纹理特征来检测字符区.3 个区域的能量表达式分别为

$$E_h=\sum_{i,j\in\Omega_1}|Y(i,j)|$$

$$E_x=\sum_{i,j\in\Omega_2}|Y(i,j)|$$

$$E_v=\sum_{i,j\in\Omega_3}|Y(i,j)|$$

其中,$Y(i,j)$表示 DCT 的系数值,E_h、E_x、E_v 分别表示水平、斜向、竖直的频率变化能量值,Ω_1、Ω_2、Ω_3 分别

表示各个方向的频率系数所属区域,其在本算法中是可调的动态参数,实验时,采用了如图 2 所示的区域.

			1	1	1	1	1
			1	1	1	1	1
			2	1	1	1	1
3	3	2	2	2	1	1	1
3	3	3	2	2	2	2	1
3	3	3	3	2	2	2	2
3	3	3	3	2	2	2	2
3	3	3	3	3	2	2	2

图 2 试验用的 DCT 系数区域

为了定位图象中的字符区域(图 3 是一幅原始图),在对图象进行 8×8 的 DCT 变换后,首先求出每个 DCT 子块的 E_h、E_v,并根据各自的阈值 T_h、T_v 进行初选,这时凡是满足阈值的竖向、横向纹理区都被选出(如图 4 所示). 由于图象中的字符一般都是成片的,很少单个存在,因此对于初选区域中的单个噪音点,可采用形态滤波的方法去除. 去除噪音点后,再用闭运算的方法填补候选区的空穴和间隙(见图 5). 此时的候选字符区已消除了单个的噪音点,并且相邻接的区域也已合并,以后进一步的工作就是要消除非字符区.

图 3 含字符区的图象

图 4 满足阈值的横、竖向纹理区

由于字符在统计意义上具有一定的斜向能量,故最后用斜向能量来排除非字符区,但由于并不是每个字符都具有斜向能量(如对于某些汉字),因此这里采用候选字符区的平均斜向能量来量度. 假设图象中有

图 5 形态运算后的结果

M 个候选字符区 $D_k(k=1,\cdots,M)$,且 D_k 中共包含 b_k 个 8×8 的 DCT 子块,每个 DCT 子块的斜向能量为 $E_x^j(j=1,\cdots,b_k)$. 则 D_k 的平均斜向能量为

$$P_k=\frac{1}{b_k}\sum_{j\in D_k}E_x^j$$

对于每个候选字符区 D_k 的平均斜向能量 P_k,用 T_x 作为阈值进行精细筛选,最后把满足条件的所有 $D_t(0\leq t\leq M)$确定为真正的字符区,而对每个字符区,则用能包含这个区域的最小矩形框来表示,其结果如图 6、图 7 所示.

图 6 满足斜向纹理阈值的区域

图 7 最终定位结果

2 压缩域的处理方法

基于 DCT 压缩编码的基本算法[7]是:①对输入图象进行分块,块的大小为 8×8;②对每个 8×8 的子块实施 DCT 变换,使其能量按频率集中,并对 DCT 系数按量化矩阵进行量化;③对已量化的 DCT 系数进行 ZIG-ZAG 排序,由于这样会出现成片的零

值,因而便于下一步处理;④对已排序的系数进行游程长(RLE)、Huffman 编码,以得到最终压缩码流.

在压缩域进行处理时,首先要对 JPEG、MPEG 的压缩数据进行 Huffman 解码(因为经 Huffman 编码后的数据流为非字节对齐,其数据项是非结构化的,并不利于计算机处理,故在 DCT 压缩域分析处理数据时,一般都要进行 Huffman 解码),在得到 RLE 域的数据码流后,就可进行处理. 可见,这里的压缩域处理就是对 RLE 压缩码流的处理,具体步骤如下:

(1)首先将区域模板(图 2)和 DCT 系数的量化矩阵进行 ZIGZAG 排序,使其与 RLE 中的 DCT 系数顺序一致.

(2)对 RLE 码流数据,根据排序后的区域模板求每个子块的 E_h、E_v 值(注意,求取该值时,要乘上 DCT 的量化系数),并根据阈值 T_h、T_v 来确定候选字符区.

(3)用形态学算子滤除单个噪音点,并连通邻近区域.

(4)对于每个候选字符区,计算其平均能量 P_k,并根据 T_x 阈值来确定最终的字符区.

(5)对于每个字符区,用能包含此区域的最小矩形框来表示.

可见,对于 JPEG、MPEG 压缩码流,在定位其图象中的字符区时,并不需要对其码流进行全部解码,而只需要进行部分解码(Huffman 解码),即可直接在 RLE 域,根据 DCT 域的特性来提取字符的纹理特征,由于其充分利用了压缩数据的特性,节省了许多处理时间,故有利于实时处理,实验结果证明,这种方法是比较有效的.

3　实验结果

为了验证本文所提的字符区域定位算法的效果,这里用该方法对 JPEG 图象和 MPEG 的 I 帧图象进行了处理试验(对于彩色图象,只对 Y 分量进行处理). 这些图象主要来自 VCD 视频、电视节目、网上下载的 JPEG 图片和用数码相机拍摄的 JPEG 照片共3 800幅,所含字符区域共4 320个,其中包含的字符既有汉字,也有英文. 实验时,用统一阈值:$T_h=T_v=420$,$T_x=330$. 试验的部分结果见图 8～图 10,统计结果见表 1.

图 8　字符区定位部分结果(白线框为定位出的字符区)

图 9　对大尺寸、实心、低对比度字符引起漏判

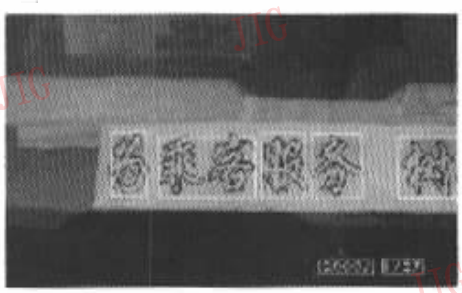

图 10　对大尺寸、空心、强对比度字符的正确定位(白线框为定位出的字符区)

表 1 检测结果(共4 320个字符区域)

实际检测到的字符区域数	正确检测率(%)	未检测到的字符区域数	漏检率(%)	检测到的错误字符区域数	错误检测率(%)
4 212	97.5	108	2.5	38	0.88

由表 1 可见,有一定的漏判、误判率,据分析,引起误判的主要原因是某些具有像字符一样的纹理图案构成了一定的似字符区域(如图 8 右下图中的中央电视台台标),引起漏判的主要原因是:(1)字符尺寸较大,且笔划为实心,在 8×8 的子块中显示不出其频率的变化特性(图 9),但如果字符笔划为空心,则能正确定位(如图 10 所示);(2)如果字符的亮度比背景亮度低很多,且不满足阈值 T_h、T_v、T_x,则字符区域定位失败,但如果降低阈值,又使其算法失去一般性,这样在处理别的图象时,容易产生误判,其解决的办法是设计自适应阈值,以达到动态调整的目的,这将是下一步的研究方向.

4 讨 论

本文提出了一种直接在 DCT 压缩域定位字符区域的算法,对于压缩图象,用这种算法只需进行少量解码(Huffman 解码),即可在 DCT 域中,根据字符具有横向、竖向、斜向频率变化的统计特性,用阈值进行字符区域定位. 从试验结果来看,这种算法具有较高的可靠度,因而具有一定的实用性. 进一步的工作主要研究以下几个方面的内容:(1)对于彩色图象,目前本算法只用到了 Y 分量,如果能结合 C_r、C_b 分量,将会进一步提高字符区域定位的准确性;(2)采用自适应的阈值方法,使字符区域定位不受背景与字符区亮度的变化影响;(3)引入多分辨率技术,使本算法对各种尺寸的字符都能正确定位;(4)进一步研究 MPEG 码流中,对 P、B 帧所含字符区的定位方法.

参 考 文 献

1 Gargi U, Antani A, Kasturi R. Indexing text events in digital video databases[A]. In:Proc. 14th Int'l conf. on Pattern Recognition(ICPR)[C]. Brisbane, 1998:916～918.

2 Shim J C, Dorai C, Bolle R. Automatic text extraction from video for content-based annotation and retrieval[A]. In:Proc. 14th Int'l Conf on Pattern Recognition(ICPR)[C]. Brisbane, 1998:618～620.

3 Ohya J, Shio A, Akamastsu S. Recognizing characters in scene images[J]. IEEE Trans. Pattern Analysis and Machine Intelligence. 1994,16:214～220.

4 Yeo B L, Liu B. Visual content highlighting via automatic extraction of embedded captions on MPEG compressed video[A]. In:SPIE digital video compression: algorithms and technologies[C]. San Jose, CA USA,1996,2668:38～47.

5 Zhong Y, Karu K, Jain A K. Locating text in complex color images[J]. Pattern Recognition, 1995,28(10):1523～1536.

6 Chiptrasert B, Rao K R. Discrete cosine transform filtering[J]. Signal Processing, 1990,19(3):233～245.

7 Wallace G K. The JPEG still picture compression standard[J]. Comm ACM,1991,34(4):31～44.

黄祥林 1967 年生,1998 年毕业于长春科技大学,讲师,现为北京工业大学信号与信息处理专业博士研究生. 研究方向为基于内容的图象检索、压缩域视频/图象处理、图象编码等.

沈兰荪 1938 年生,教授,博士生导师. 主要研究领域为图象编码与传输、VLSI 实时信号处理、光谱信号检测等智能化信息处理领域.