

基于 Gabor 特征与 BP 神经网络的屏幕显示自动校验系统

向荣*, 周慧娟

(浙江大学 数字技术及仪器研究所, 杭州 310027)

(* 通信作者电子邮箱 xiangrong0302@gmail.com)

摘要:针对人工校验视频监控设备屏幕显示(OSD)效率低下、人力物力资源耗费大的问题,提出一种 OSD 自动校验系统,取代传统的人工校验方式。系统首先综合多种数理统计特征进行 OSD 定位,然后利用改进的 Otsu 算法进行精确字符分割并二值化,最后通过基于 Gabor 特征离线训练的改进型 BP 神经网络进行字符识别。实验结果表明,在确保 92.7% 识别率的前提下,该系统识别一帧 OSD 平均耗时 53 ms。

关键词:屏幕显示校验; Gabor 特征; BP 神经网络; 字符分割; 快速二值化

中图分类号: TP391.41 **文献标志码:** A

Automatic on-screen-display verification system based on Gabor features and BP neural network

XIANG Rong*, ZHOU Huijuan

(Institute of Advanced Digital Technologies and Instrument, Zhejiang University, Hangzhou Zhejiang 310027, China)

Abstract: To deal with low efficiency and long-time consumption in verifying OSD (On-Screen-Display) of video devices, this paper devised an automatic OSD verification system. The system consisted of three parts. OSD area location was achieved by synthesizing statistical characteristics. Single character was then segmented based on improved Otsu algorithm. Finally, Gabor features and improved BP neural network were used to recognize these characters. The experimental results show that this system costs average 53 ms per recognition of one frame with a recognition rate at 92.7%.

Key words: On Screen Display (OSD) verification; Gabor feature; Back Propagation (BP) neural network; character segmentation; quick binarying

0 引言

随着网络多媒体技术的发展,数字视频系统在银行、交通、治安、家电、手持式设备等领域取得了广泛的应用^[1-3],屏幕显示(On-Screen-Display, OSD)技术随之发展起来。此技术是利用在屏幕上显示特定的字或者图像来完成人机交互的过程,让使用视频的用户获得如监控的时间及位置等信息。而在视频终端设备如机顶盒、硬盘录像机(Digital Video Recorder, DVR)、网络视频录像机(Network Video Recorder, NVR)中,OSD 更是以界面菜单的方式起到关键性作用^[4]。

在监控系统或视频设备使用之前,要花费很大的人力物力对 OSD 的正确性进行校验。由于 OSD 在图像中的位置随机,背景复杂程度各异,很难采用传统的程序化方式进行校验。目前普遍采用人工校验的方式,但这样校验耗时耗力,而且很难实现 OSD 字符集的遍历校验。一旦校验疏忽,投产的视频设备出现 OSD 错误,会对产品信誉造成极大的负面效应,带来严重的经济损失。

目前国内外已公布 OSD 识别方面相关的研究成果很少。关于字体识别,文献中采用的 Gabor 特征是一种能够良好反映字形特征的处理方法^[5-7],近年来为很多识别系统所采用。通过分析多种 OSD 原理,本文提出了一种基于 Gabor 特征与改进型 BP 神经网络的 OSD 自动校验系统,以取代传统的人工校验方式。目前已经作为 OSD 自动校验的一套解决方案,为国内某视频监控公司实际使用。

1 OSD 原理

OSD 应用的终端多种多样,如数字电视机顶盒、网络摄像头、数字摄像头、视频电话等。在这些设备内部产生 OSD 的核心部件也是五花八门,先进精简指令集处理器(Advanced RISC Machine, ARM)、数字信号处理器(Digital Signal Processor, DSP)、现场可编程门阵列(Field-Programmable Gate Array, FPGA)或 OSD 专用芯片均可以作为 OSD 算法的执行部件^[8-12]。另外,随着 3D 技术的普及,现在还出现了 3D 显示效果的 OSD^[13]。但无论用什么硬件实现,实现效果如何,其叠加在最后视频上的原理均大同小异。图 1 为 OSD 叠加一般关键步骤流程。

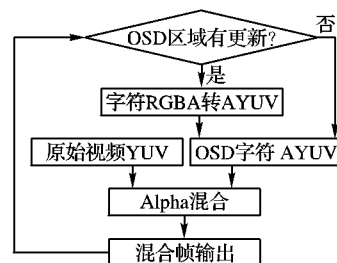


图1 OSD 叠加流程

显示功能模块在获知 OSD 区域有更新之后,通过编码查表(如图2)的方式,获取待叠加区域的点阵 RGBA 信息,经过 RGBA 到 AYUV 的格式转换,最终将包含 OSD 信息的 AYUV 信息与原始视频 YUV 数据进行 Alpha 混合,最终显示在终端

收稿日期: 2012-12-03; 修回日期: 2012-12-29。 基金项目: 国家 863 计划项目(2010AA09Z104)。

作者简介: 向荣(1988-), 男, 湖北嘉鱼人, 硕士研究生, 主要研究方向: 视频监控系统、图像处理、模式识别; 周慧娟(1987-), 女, 福建莆田人, 硕士研究生, 主要研究方向: 视频监控系统、嵌入式系统 Web。

设备的界面上。在嵌入式系统中,字符的点阵数据通常存放在 asc. dat 与 hzk. dat 文件中。前者包括了 ISO 8859-1 规定的 256 个字符,后者包括了 GB2312 的所有汉字,这些点阵信息可以作为神经网络训练阶段的样本。

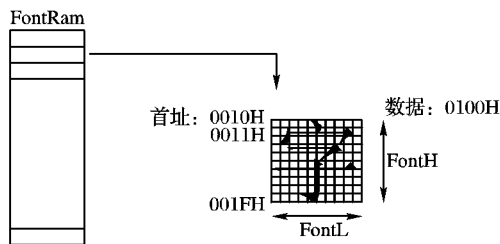


图2 OSD 字符点阵索引示意图

2 校验系统整体方案设计

由以上分析可知,虽然 OSD 设备与显示原理多种多样,但最终显示的结果均为原始码流上叠加 OSD 的字符串。因此,识别 OSD 的时候不必拘泥于 OSD 如何产生,而可以直接考虑最终显示的码流中的 OSD 特点。图 3 为针对已经应用于高速公路、安防等领域的某视频监控系统的 OSD 校验方案框架。终端 OSD 的产生为 DSP 的点阵式 OSD 与视频码流叠加而成,通过图中组网方式,可提取网络中的视频码流,从而获取到校验系统的识别目标。

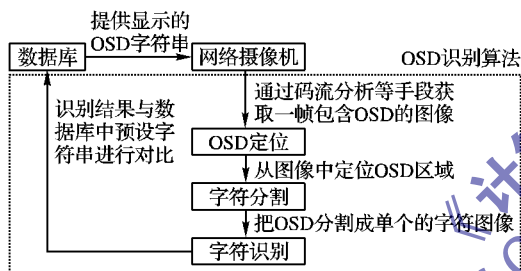


图3 OSD 识别系统框架

视频监控系统的 OSD 内容一般由提供当前监控时刻的时间 OSD 与提供当前监控位置的厂名 OSD 组成,厂名 OSD 的内容通常来源于监控网络中的数据库。网络摄像机获取数据库中的厂名字符串,然后与原始码流叠加发送到网络中的接收端(比如警卫室中监视器等)。因此,校验系统可以截取合成后的 OSD 码流中的一帧进行识别,依次进行 OSD 定位、字符分割和基于 Gabor 特征与改进 BP 神经网络的字符识别,从而识别出 OSD 的字符信息。最后,OSD 自动校验系统将识别结果与数据库中的预设字符串进行对比,从而达到验证的目的。

3 OSD 识别算法设计与实现

3.1 多特征 OSD 定位

在图像识别中,定位是最为关键的步骤,定位的准确与否直接影响后面的分割和识别的结果。考虑到 OSD 区域的字符有丰富的边缘纹理信息,本文采用图像边缘特征提取的方式,从截取的一帧图像中定位出 OSD 的字符区域。首先对截取的一帧图像预处理与去噪,得到高信噪比的灰度图像,步骤如下:1)灰度化;2)Gauss-Laplace 边缘检测;3)高斯低通滤波去噪。

去噪后,统计水平方向的投影,再计算逐行之间的投影梯度,得到第 i 行与上一行像素的投影像素和差的绝对值,取梯

度极值记为可能的上下边界,以上过程可以参考图 4。

由于上下边界检测结果可能不止一对,需要遍历所有的上下组合从而定位所有的 OSD 字符区域。在定位 OSD 左右区域的时候,本文综合了字符高度的先验知识与统计平均的思想,采取了不同于其他文献的多数理统计特征定位法:

- 1)垂直投影分量信息:首先统计边缘检测结果中每一组可能的上下边界之间的垂直方向投影;
- 2)字块模板统计信息:然后以上下边界之间的距离为假定字符高度进行逐个字块的边缘信息和统计;
- 3)邻域边缘统计信息:为了避免空格对字符右边界的误判,相邻 5 个字块的边缘信息和进行中值平均;
- 4)左右边界极值信息:最后采用极值法判断左右边界,从而尽可能准确地截取到每一个 OSD 所在区域。

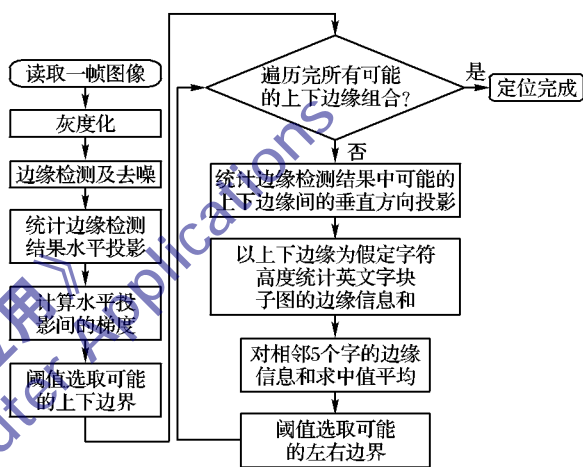


图4 OSD 定位总流程

3.2 精确字符分割

当 OSD 字符定位完成后,需要遍历所有可能的字符区域,将区域中的字单独分割开来识别。利用字与字之间的间隙的垂直投影特点,进行逐字分割。在分割过程中,考虑到视频流中图像不可能出现水平或者垂直的畸变,可以结合理想情况下英文字符占 1 个单位宽度、汉字占 2 个单位宽度的先验知识进行精确分割。具体步骤为:

- 1)垂直方向边缘检测;
- 2)统计边缘检测后的每一列字符像素和;
- 3)综合考虑像素和的极小值与字符的先验知识分割边缘;
- 4)改进的 Otsu 算法二值化每一个字块子图;
- 5)归一化。

传统的分割方式直接利用图像中的字符像素的投影,对噪点粘连干扰非常敏感。本文为了最大限度减小这种干扰,采用了特殊的仅垂直方向的边缘检测,将拉普拉斯算子水平方向的加权叠加到垂直方向上,可得垂直边缘检测的算子为:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & -2 & 0 \\ 0 & 4 & 0 \\ 0 & -2 & 0 \end{bmatrix} \quad (1)$$

仅做垂直方向的边缘检测还可以有效避免一般边缘检测算法时字符边缘间的粘连对边界识别的影响,从而更加准确地切分到字符的边界。边缘检测后,垂直计算前景像素和,取极小值为字符的边缘,图 5 为反色后的投影示意图。

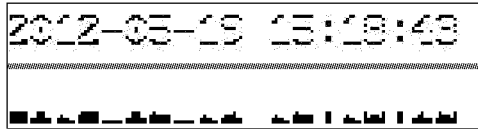


图5 垂直投影示意图

针对参考文献中提到的分割常见难点,如标点符号分割过小、汉字过分割、空格分割干扰等,本文采用以下三个辅助分割方法:1)从左至右检测下一个分割点时,自上一个分割点统计投影和,用以判定当前字符为汉字还是非汉字符号;2)如果出现标点符号或者空格字符,此时零点或者极小值可能有多,本文直接用字符宽度先验值作为分割结果,从而可有效避免标点符号如“:”分割过小的情况;3)对每个字块内的前景色(字符的颜色)采用统计方差阈值的方式来判断当前字块是否为空格字符,对于空格字符不进行二值化处理,避免空格字符误识别情况。经过以上投影后的补充处理,可以极大提高分割的精确性,在定位正确的情况下,分割准确率接近 100%。

本文改进了 Otsu 算法^[14],在与传统 Otsu 算法效果基本一致的情况下,使耗时降低为原来的 1/8。假设图像灰度均值为:

$$M = r_1 \cdot g_1 + r_2 \cdot g_2; g_1 < M < g_2 \quad (2)$$

方差计算如下:

$$C^2 = r_1(g_1 - M)^2 + r_2(g_2 - M)^2 \quad (3)$$

对于黑底白字图片, r_1 为背景像素比例, g_1 为背景灰度, r_2 为字符像素比例, g_2 为字符灰度。为把白字从黑底中分割出来,选取粗阈值。

$$T_1 = M + \sqrt{r_1/r_2} \cdot C \quad (4)$$

式(4)中字符像素点所占比例 r_1/r_2 未知,但可以通过字符点阵数据进行估算,经统计该比例约为 0.2。然后利用整个区域的均值和方差进行 M 与 C 的近似。计算公式如下:

$$M = \sum_{j=0}^H \sum_{i=0}^W g(i, j) / (H \times W) \quad (5)$$

$$C^2 = \sum_{j=0}^H \sum_{i=0}^W [g(i, j) - M]^2 / (H \times W) \quad (6)$$

其中: H 为图像高度像素值, W 为图像宽度像素值, $g(i, j)$ 为图像 i 列 j 行像素的灰度值。采用 T_1 作为可能的二值化分割阈值,上下放宽 10 个灰度值进行最大类间方差处理得到准确的分割阈值。

最后,对分割好的字符进行归一化的处理,归一化尺寸对英文为 8×16 , 中文为 16×16 。同时,采用较小的尺寸也可以提高识别的处理速度。

3.3 字符识别

Gabor 特征具有与人类大脑皮层单细胞的二维反射区相同的特征,即能够捕捉对应于空间频率、空间位置及方向选择性的局部结构特征,利用它较强的分类鉴别能力,对分割字符进行逐一识别^[15-17]。

用二维 Gabor 滤波器进行纹理特性分析,当加窗函数为高斯函数时其复数形式为:

$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) [\exp(i\omega x') - \exp(-\omega^2 \sigma^2 / 2)] \quad (7)$$

其中: $x' = x \cos \theta + y \sin \theta$, $y' = -x \sin \theta + y \cos \theta$; ω 为径向

中心角频率; θ 为滤波器的方向; σ 为高斯包络在 X 和 Y 方向上的均方差; φ 为倍频程带宽,一般取 $1 \sim 1.5$, 本文取 1.0。 σ 与 ω 的关系如下:

$$\sigma = \sqrt{2 \ln 2} \left(\frac{2^\varphi + 1}{2^\varphi - 1} \right) / \omega \quad (8)$$

则由 ω 可确定 σ 。由于汉字具有横、竖、撇、捺基本笔划的形状特点,可选取 $0, \pi/8, \pi/4, 3\pi/8, \pi/2, 5\pi/8, 3\pi/4, 7\pi/8$ 八个方向进行 Gabor 特征提取。图 6 为字符为“2”的八向 Gabor 特征。

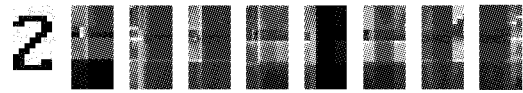


图6 “2”及其八向 Gabor 特征

获取 Gabor 特征后,采用三层改进型 BP 神经网络进行特征训练与字符识别,其结构如图 7 所示。

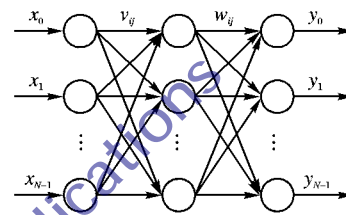


图7 三层 BP 神经网络

输入层到输出层激活函数均采用 Sigmoid 函数:

$$f(x) = 1 / (1 + e^{-x}) \quad (9)$$

由于 OSD 生成时分为 ASCII 字符型和汉字型,则可通过两个子网络分开识别,根据分割的图像大小为 8×16 , Gabor 特征选取 8 个方向可知, ASCII 字符型的输入层神经元个数为 $8 \times 8 \times 16$, 输出层的神经元个数与 ASCII 的非控制通信字符与 del 键,即 $128 - 32 - 1 = 95$ 个。同理,汉字型的输入层神经元个数为 $8 \times 16 \times 16$, 输出层的汉字个数为视频系统常用汉字标点符号 32 个与视频系统常用字 130 个。对于隐含层神经元个数的选取可参考以下经验公式:

$$i = \sqrt{n + m} + a \quad (10)$$

其中: i 为隐含层神经元个数, n 为输入层神经元个数, m 为输出层神经元个数, a 为 $1 \sim 10$ 的整数。通过大量的实验,最后选取 ASCII 字符型的隐含层神经元个数为 47, 汉字型个数为 75。具体参数设置如表 1 所示,利用这些参数建立神经网络进行离线训练,对于每种字符选取 20 个不同字体随机背景的归一化图片。

表1 各神经网络的参数配置表

识别对象	输入节点数	隐层节点数	输出节点数	学习速率	动量因子	误差目标值	迭代次数
汉字	2048	75	162	0.3	0.5	0.01	5000
ASCII 字符	1024	47	95	0.3	0.5	0.01	5000

4 实验结果与分析

本文针对业界常用三种视频分辨率: 352×288 , 720×576 , 1920×1080 对不同应用场景下的随机实时视频数据共 120 帧进行了识别实验,图 8 给出了实验所用的部分样本。

4.1 识别算法分析比较

传统字符识别算法有像素点匹配和 Gabor 特征 + 欧氏距离法。前者将要匹配的字符样本直接与每一个模板字符样本

逐一对比,找出相差像素个数最少的那个模板即为识别字符;后者通过计算要匹配的字符的多方向 Gabor 特征的欧氏距离,选出最小距离的即为待识别文本。

本文做了用 Gabor 特征与改进 BP 神经网络识别与传统识别方法的对比实验。在字符分割完成的基础上,分别采用像素点匹配、Gabor 特征 + 欧氏距离、Gabor 特征 + 改进 BP 神经网络三种方法进行上述三种分辨率下,共 2614 个字符的识别比较,结果如表 2。

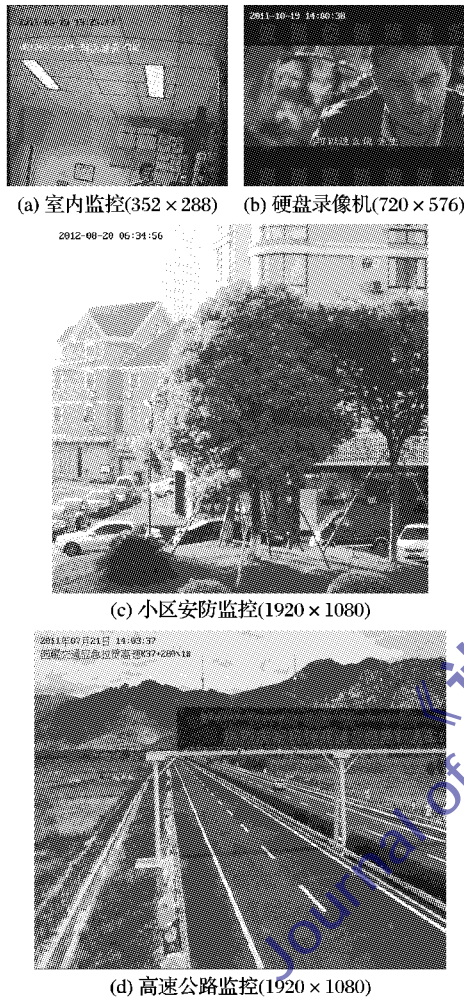


图 8 实验所用的部分 OSD 识别样本

表 2 识别算法对比结果

识别方法	待识别 文字总数	正确识别 文字数	识别 率/%
像素点匹配	2614	1929	73.8
Gabor 特征 + 欧氏距离	2614	2316	88.6
Gabor 特征 + 改进 BP 神经网络	2614	2423	92.7

由表 2 可以发现,Gabor 特征比像素点的直接匹配更能体现每一个字符的特殊性,而采用 BP 神经网络对 Gabor 特征进行分类能进一步提高识别正确率。

4.2 自动化测试

为了验证本文系统的校验效率,在不同分辨率下进行了自动化校验的时耗测试。共测试 120 帧,总用时 6365 ms。表 3 为各分辨率下,一帧 OSD 的模块校验平均时耗。由表 3 可见,时耗与分辨率大小密切相关。总的来说,本文系统的平均时耗为 53 ms,在长时间自动化校验任务中非常理想。

表 3 自动校验系统时耗分析 ms

测试分辨率	定位	分割	识别	总耗时
352 × 288	<1	<1	<1	<3
720 × 576	12	9	7	28
1920 × 1080	43	36	33	112

5 结语

本文提出了一种基于 Gabor 特征与改进型 BP 神经网络的 OSD 自动校验系统,在局部强光照、噪点、花纹等较为恶劣的背景干扰下,所采用多数统计特征定位方法有效,仅作垂直边缘检测能够更加精确地进行字符分割,改进的 Otsu 二值化算法大幅度提高了二值化的处理速度,Gabor 特征与 BP 神经网络识别算法比传统算法识别率高,实验结果表明在识别率达到 92.7% 时,平均识别一帧耗时 53 ms。

参考文献:

- [1] 桂丹. 高速公路视频监控系统及 OSD 的设计与实现[D]. 武汉: 武汉理工大学, 2010.
- [2] 白杨. 基于 Nios II 的红外热像仪 OSD 显示与功能实现[D]. 南京: 南京理工大学, 2010.
- [3] 费莉梅. 网络视频录像机 OSD 子系统软件设计[D]. 杭州: 浙江大学, 2012.
- [4] 任晓, 邱军, 顾成成. 基于 Linux 平台的实时视频 OSD 设计与实现[J]. 计算机应用与软件, 2010, 27(10): 185 - 187.
- [5] 戴维. 基于 Gabor 特征与 SVM 的汉字识别系统研究与实现[D]. 上海: 上海交通大学, 2008.
- [6] 赵英男. Gabor 滤波器在车辆检测和车型识别中的应用研究[D]. 南京: 南京理工大学, 2004.
- [7] 邓洪波. 基于 Gabor 特征的人脸表情识别和手写体汉字识别[D]. 广州: 华南理工大学, 2006.
- [8] 曾德瑞, 普玉伟, 杨明武. 数字视频 ASIC 中 OSD 系统的实现[J]. 器件与应用, 2011, 35(23): 36 - 39.
- [9] 黄燕群, 李利品, 王爽英. 基于 FPGA 的 OSD 设计[J]. 液晶与显示, 2010, 25(3): 429 - 433.
- [10] 林国政, 沈小林. 基于 ARM 的嵌入式视频监控系统设计[J]. 电子测试, 2011(3): 55 - 59.
- [11] 郑世程, 于忠臣. 基于 BES7456 芯片的 OSD 的应用平台设计[J]. 单片机与嵌入式系统应用, 2011, 11(6): 35 - 37.
- [12] 黄益贵, 王汝传. 基于 DM6446 的 P2P 机顶盒 OSD 研究与实现[J]. 计算机技术与应用, 2010, 20(4): 220 - 223.
- [13] 江维, 罗才华. 高清数字电视机顶盒 3D OSD 显示框架的设计与实现[J]. 有线电视技术, 2012(1): 72 - 75.
- [14] 张引. 基于空间分布的最大类间方差牌照图像二值化算法[J]. 浙江大学学报: 工学版, 2001, 35(3): 272 - 275.
- [15] HUANG L L, SHIMIZU A, KOBATAKE H. Classification-based face detection using Gabor filter features[C]// Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition. Washington, DC: IEEE Computer Society, 2004: 397 - 402.
- [16] FASEL I R, BARTLETT M S, MOVELLAN J R. A comparison of Gabor filter methods for automatic detection of facial landmarks[C]// Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition. Washington, DC: IEEE Computer Society, 2002: 231 - 235.
- [17] KYRKI V, KAMARAINEN J K, KÄLVIAINEN H. Simple Gabor feature space for invariant object recognition[J]. Pattern Recognition Letters, 2004, 25(3): 311 - 318.