

基于多模板匹配的单人脸检测*

梁路宏 艾海舟 何克忠

(清华大学计算机科学与技术系智能技术与系统国家重点实验室, 北京 100084)

摘要 提出了一种基于多模板匹配的一般环境图象中单人脸的检测方法。首先用双眼模板初步筛选, 然后使用具有不同长宽比的多个人脸模板进行检测和定位, 最后利用人脸器官的边缘特征进行确认。采用以固定比率压缩输入图象的方法解决不同尺度人脸的检测问题。实验表明这种方法具有较高的速度和检测成功率。

关键词 模板匹配 人脸识别 人脸检测

0 引言

人脸识别的研究最近几年再度受到普遍重视, 它与指纹识别、视网膜识别等同属于生物特征识别技术范畴, 在诸如安全检查、保安管理等方面有着十分重要的应用价值。浏览各种文献可以发现, 对于人脸识别算法的研究已经相当深入^[1], 但对于如何检测人脸或输入人脸的研究近几年才受到重视。这是因为过去的人脸识别研究主要针对具有较强约束条件的人脸图象(如无背景), 往往假设人脸的位置已知或很容易获得; 近几年来, 电子商务等网络资源的利用使得人脸识别成为最有潜力的生物身份验证手段, 在这种应用背景下, 人脸检测作为一个单独的课题开始受到研究者的重视。此外, 人脸检测在基于内容的检索、多模态人机接口、智能监视等方面也有着重要的应用价值, 这使得人脸检测的研究倍受关注。从学术观点来看, 人脸检测的挑战性意义在于人脸是自然界中构成和细节变化最复杂的模式之一, 如果能够研究出一个成功的人脸检测系统, 那么就可以对其它相似特征或模式的检测问题提供有价值的参考^[2]。

本文所研究的是一般环境图象中单个正面端正人脸的检测问题。这种条件下的人脸检测的方法主要有模板匹配方法^[3]、可变形模板方法^[4]、镶嵌图方法^[5]、基于统计特征的方法^[6]以及基于人工神经网络的方法^[2]等。概括地说, 基于模板匹配的方法是在图象灰度的层次上直接比较目标模板和候选图象区域之间的相似性, 而基于特征的匹配方法是比较从图象中抽取的一定特征的相似性, 对人脸检测而言, 特

征包括眼睛、鼻、嘴等。模板匹配方法具有简单、直观的特点, 与特征匹配方法相比具有较强的适应性, 原因在于通常特征抽取依赖于图象的质量, 易受噪音影响而不稳定。

本文提出了一种基于多模板匹配的单人脸检测方法, 其中用到两种模板: 双眼模板和不同长宽比的人脸模板。在检测时首先使用双眼模板进行粗筛选, 然后使用不同长宽比的人脸模板确定出人脸区域的位置和范围, 最后利用人脸器官的边缘特征进行确认。为了解决图象中不同尺度人脸的检测问题, 采用了按照固定比率逐步缩小输入图象的方法进行多个尺度上的搜索。实验表明该方法在 162 幅单人头部图象(包括简单背景和复杂背景)中检测人脸的成功率可以达到 94.4%; 在 55 幅较清晰的单人半身图象中检测人脸的成功率也达到了 81.8%。

1 模板生成

我们在人脸检测中使用了 6 个模板: 一个用于初筛选的双眼模板, 5 个用于检测的不同长宽比的人脸模板。采用多个人脸模板是为了适应不同人脸部的不同长宽比。模板是通过多个样本取平均构造出来的, 如图 1 所示, 首先在选取的样本图象上用手工划出人脸的区域作为人脸样本。由于各个样本的尺度大小和灰度分布各不相同, 因此先对它们进行尺度和灰度分布标准化, 然后将所有样本取灰度平均并压缩到需要的尺度作为原始模板。拷贝原始模板的眼睛部分, 进行灰度分布标准化后作为双眼模板; 对原始模板分别按照 1:0.9、1:1、1:1.1、

* 本文研究受清华大学青年科学基金及国家“863”高科技计划(No. 980311096)资助

收稿日期: 1999-03-04; 收到修改稿日期: 1999-07-26

1:1.2和1:1.3的宽长比变形,进行灰度分布标准化 化后作为人脸模板。

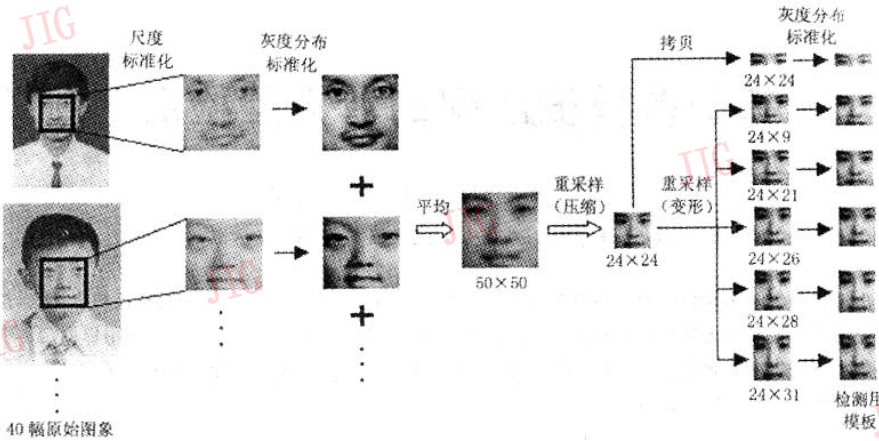


图1 模板生成的过程

在模板生成中最主要的工作是图象的尺度变换和灰度分布标准化。尺度变换采用了基于线性插值的重采样方法,这里不再具体讨论。我们考虑了表征图象灰度分布的两个参量——灰度的平均值和方差,将它们调整到给定值,以消除由于采集时光照等因素对人脸图象的影响。对于模板匹配,我们认为这种灰度分布标准化方法比常用的直方图均衡化方法更有效,原因在于它将均值和方差这两个关键数字特征一致化了。

将图象看作一个二维矩阵 $D[W][H]$, 其中 W 和 H 为图象的宽和高, 则该图象的灰度平均值可表示为:

$$\bar{\mu} = \frac{1}{W \cdot H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} D[i][j] \quad (1)$$

而灰度分布的方差可表示为:

$$\sigma^2 = \frac{1}{W \cdot H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} (D[i][j] - \bar{\mu})^2 \quad (2)$$

对于输入的第一个样本图象, 为了将其灰度平均值和方差变换到事先设定的灰度平均值 μ_0 和方差 σ_0 , 对样本中的每个象素点的灰度值进行如下的变换:

$$\hat{D}[i][j] = \frac{\sigma_0}{\sigma} (D[i][j] - \bar{\mu}) + \mu_0 \quad (3)$$

$0 \leq i < W, 0 \leq j < H$

公式(3)是一种近似算法, 效果参见图1。实验中我们手工提取了40个人脸样本, 通过以上方法得到一个尺度为 24×9 像素的双眼模板和尺度分别为 24×21 、 24×24 、 24×26 、 24×28 、 24×31 像素的5个人脸模板。为了保证采集和求平均时人脸样本之间能够对准, 我们在构造模板的过程中首先将样本标准化到 50×50 , 求平均后再压缩到需要的大小。

2 检测方法

假设人脸模板的灰度矩阵为 $T[M][N]$, 灰度均值为 μ_T , 均方差为 σ_T , 输入图象区域的灰度矩阵为 $R[M][N]$, 灰度均值为 μ_R , 均方差为 σ_R , 那么它们之间的相关系数 $r(T, R)$ 和对应象素灰度值的平均偏差 $d(T, R)$ 分别为:

$$r(T, R) = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (T[i][j] - \mu_T)(R[i][j] - \mu_R)}{M \cdot N \cdot \sigma_T \cdot \sigma_R} \quad (4)$$

$$d(T, R) = \sqrt{\frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (T[i][j] - R[i][j])^2}{M \cdot N}} \quad (5)$$

$r(T, R)$ 越大表示模板与输入图象区域的匹配程度越高; 而 $d(T, R)$ 正相反。将它们综合起来作为匹配程度的度量:

$$D(T, R) = r(T, R) + \frac{\alpha}{1 + d(T, R)} \quad (6)$$

其中 α 为权重系数, 我们取经验值 $\alpha = 35.0$, 它是根据实验数据使前后两部分的量值相当而取定的平均当量。

作为辅助手段, 我们采用人脸器官边缘特征的分布规则作为进一步确认的方法, 排除一些错检, 提高检测率。首先对待检测区域使用 Sobel 算子获得梯度图象, 然后进行灰度分布标准化, 最后二值化得到边缘图象。由于人脸边缘图象的细节受光照、面部表情等因素的影响很大, 我们采用统计的方法来提提高鲁棒性: 根据人脸部器官分布特点, 将边缘图象划分成 3×3 个相同大小的块, 统计整个图象和每一块

中边缘点所占的比例,如图 2 所示。

使用比较弱的经验规则,仅对双眼、鼻子和嘴的边缘统计特征进行检验。设 3×3 块中边缘点的比例为 $R[m,n]$ (其中 $0 \leq m \leq 2, 0 \leq n \leq 2$), 编号排列如图

2(e), 整个区域中边缘点所占比例为 \bar{R} , 若该区域是人脸则应满足规则: (1) $0.16 < \bar{R} < 0.30$; (2) $R[0,0] > 0.7\bar{R}$ 且 $R[0,2] > 0.7\bar{R}$; (3) $R[1,1] > 0$; (4) $R[2,1] > 0.5\bar{R}$ 。



图 2 利用边缘特征判别

3 算法实现

算法的基本思路是:搜索输入图象中所有可能为人脸的区域,认为满足一定条件且与模板匹配程度最高的区域是人脸。为了检测不同尺度的人脸,采用按固定比率(如 1/1.1 倍)反复压缩输入图象,在压缩后的图象中进行检查的方法;为了检测不同长宽比的人脸,采用前述 5 种不同长宽比的模板进行匹配。

假设输入图象中最多只有一个完整的人脸(正面、端正),可能的人脸宽度在 24 象素(人脸的最小分辨尺度为 16×16 ^[7],考虑到模板匹配的准确性,取最小宽度为 24)到图象宽度的 1/2 之间,人脸的位置未知。设输入图象为 I_S (宽 W 、高 H); 双眼模板为 T_{Eyes} , 5 个人脸模板为 $T_{Face}[i](1 \leq i \leq 5)$, 其宽度为 24、高度分别为 $H_{Face} = \{21, 24, 26, 28, 31\}$ 。算法如下:

- ① 初始化最大匹配度 $D_{max} = 0$, 设定当前图象 $I_C = I_S$;
- ② 对图象 I_C 使用 Sobel 算子得到梯度图象 I_{grad} , 令当前图象的宽 $w = W$ 、高 $h = H$;
- ③ 设当前扫描点 (x, y) 为图象起始点 $(0, 0)$;
- ④ 检测区域 $R_{Eyes}(x, y, x + 24, y + 9)$ 是否为双眼, 若不是则转⑦;
- ⑤ 检测相应区域是否为人脸, 是则求出当前区域与人脸模板的匹配度 D , 否则转⑦;
- ⑥ 若 $D > D_{max}$, 记录当前区域的位置和大小(换算为输入图象 I_S 中的坐标)并令 $D_{max} = D$;
- ⑦ 若 $x + 24 < W, x = x + 1$, 转④;
- ⑧ 若 $y + 31 < H, x = 0, y = y + 1$, 转④;
- ⑨ 若 $w > 48$, 对图象 I_C 使用线性插值重采样法缩小 1/1.1 倍, 修改 w 和 h , 转③;
- ⑩ 左右颠倒输入图象, 设定为当前图象 I_C , 重复②—⑨;

⑪ 输出检测结果, 结束。

其中步骤④、⑤的详细过程见图 3。

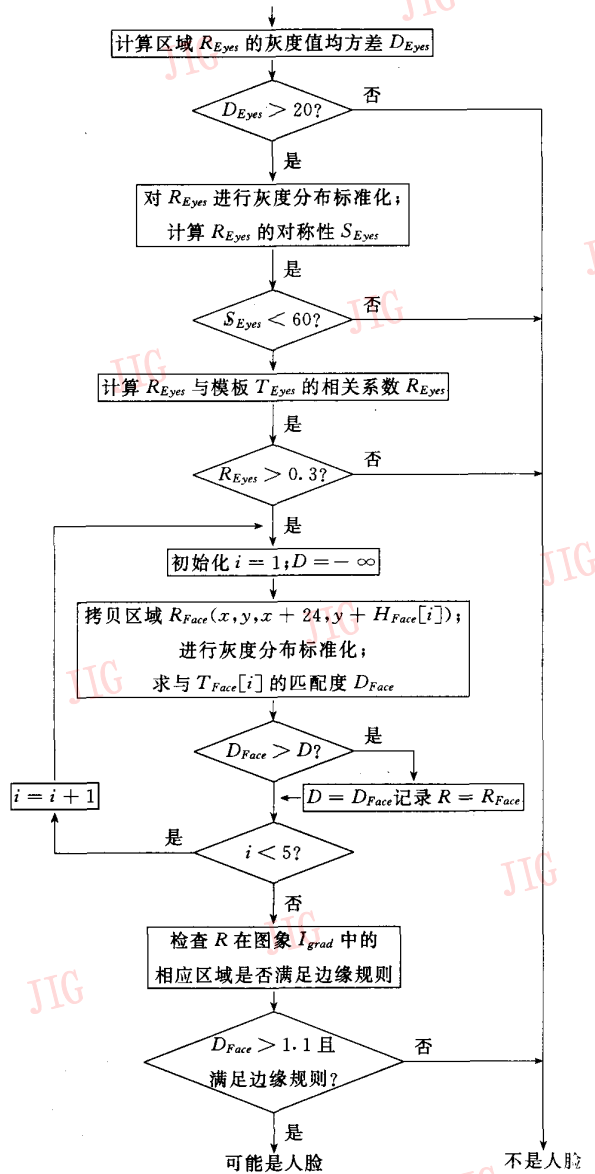


图 3 步骤④、⑤的流程

图3中双眼区域的对称性检查是基于一般人脸(正面、端正)的双眼左右对称这一先验知识进行的,度量指标 $S_{E_{yes}}$ 是区域中所有左右相对的象素间灰度值的平均偏差。在使用双眼模板匹配前先对可能为双眼的区域进行灰度值均方差和对称性的检查,可以提前“淘汰”不可能为双眼的区域,从而提高检测速度。

我们使用的双眼和人脸模板是左右非对称的,因此在算法的步骤⑩中左右颠倒输入图象后再进行次检测,以防止漏检。

4 实验结果

我们使用了两类包含单个人脸(正面、端正)的图象:头部图象共162幅,包括30幅MIT图象库中的简单和复杂背景图象、33幅摄像机采集的图象、61幅扫描的证件照以及38幅其它来源的照片;半身图象共55幅,包括13幅摄像机采集的图象、32幅扫描的普通照片和从Internet上下载的艺术照。

我们采用4种不同的方法对上述两类图象进行检测:(1)使用相关系数(公式(4))衡量匹配程度;(2)使用平均偏差(公式(6))衡量匹配程度;(3)使用相关系数和平均偏差的加权和(公式(7))衡量匹配程度;(4)使用加权和衡量匹配程度并用边缘特征确认。其中方法(4)完全采用图3所示流程,方法(1)~(3)仅采用了图3流程的一部分。我们假设已知人脸的大致尺度,将检测的尺度范围限定为图象中人脸实际尺度的约70%到130%之间。这样在检测时仅需在4~6种尺度上进行搜索,减小了计算量,同时一定程度上降低了检测的难度。

使用不同方法检测的正确率见表1。每幅图象检测所用时间从6秒(头部图象)到35秒(半身图象)不等(Pentium II 350 CPU,128MB内存)。可以看到,我们使用的相关系数与平均偏差加权和的方法优于单独使用相关系数或平均偏差,而且使用边缘特征确认可以提高检测正确率。总体而言,我们提出的算法是有效的。

表1 使用不同方法的检测正确率

照片类型	方法			
	相关系数	平均偏差	加权和	加权和+边缘确认
头部图象 (111人,162幅)	88.9% (正确144幅、错误18幅)	88.9% (正确144幅、错误18幅)	91.4% (正确148幅、错误14幅)	94.4% (正确153幅、错误8幅、漏检1幅)
半身图象 (29人,55幅)	74.5% (正确41幅、错误14幅)	76.4% (正确42幅、错误13幅)	78.2% (正确43幅、错误12幅)	81.8% (正确45幅、错误10幅)

图4是部分头部图象的检测结果,从图中可以看到,该方法对于戴眼镜的情况(左1、左3)仍有一定适应性,尽管检测用的模板没有考虑戴眼镜及有胡须等特殊情况。

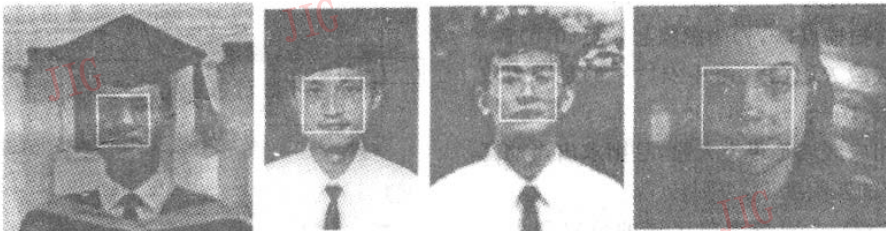


图4 头部图象中检测的部分结果

图5是对MIT复杂背景人脸图象库中的头部图象检测的部分结果。除最后一幅检测错误外,均能得到正确的结果,由此可见该算法在复杂背景的情况下仍然是有效的。



图5 MIT复杂背景人脸图象中检测的结果

图 6 是部分半身照检测的结果。可以看到,本文的方法对于人脸在图象中相对较小且背景比较复杂情况下的检测仍然有效。但总体而言,这种情况下的

检测正确率比头部图象低。这主要是因为图象的范围越大,背景越复杂,则出现类似于人脸的模式的可能性越大。

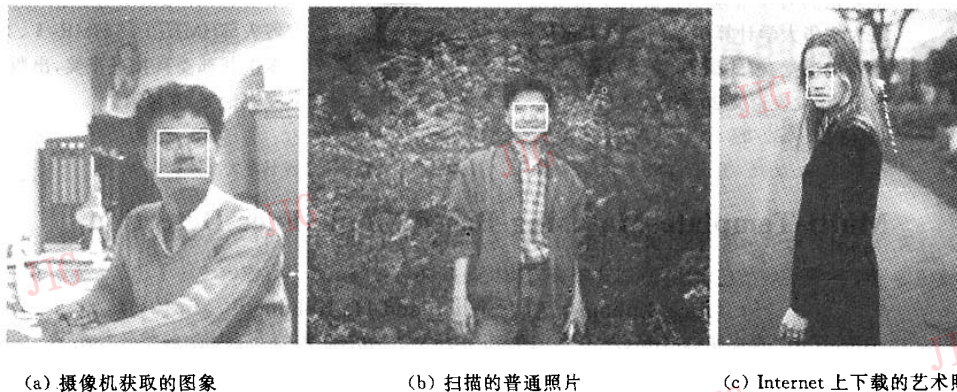


图 6 单人生活照中检测的部分结果

6 总 结

我们主要研究了静态图象中单人脸的检测算法,由于算法主要是依靠双眼模板和脸部模板来检测、定位的,因此受模板类型的限制,该算法的适用范围是正面端正的单人脸,所谓正面端正的含义是指双眼在图中均可见而且应大致在同一水平线上。这种约束对于人脸的自动识别系统以及某些基于内容的检索系统是可以接受的,因而这种方法有很大的应用价值。

我们针对满足上述限定条件的大量人脸图象进行了实验。对于类似证件照的头部图象(包括简单背景和复杂背景),检测的正确率达到 94.4%;对于质量较好的半身图象,检测的正确率也可达到 81.8%。实验结果是令人满意的。

我们的算法也存在一些缺陷。首先是对于人脸的背景较多且复杂情况下检测的错误率较高。这一方面是由于自然界的物体千变万化,在人脸图象中很可能存在某一或几方面特征类似人脸的模式;另一方面是由于人脸图象(模板)本身不能直接反映人脸的本质特征,由模板代表的人脸模式的紧致性并不好。另外,我们的算法虽然采取了一些提高检测速度的措施,但仍有许多方面没有考虑,如多尺度搜索中不同尺度图象间已有信息的利用等。这些都有待于将来进一步研究。

参 考 文 献

- 1 Chellappa R, Wilson C L, Sirohey S. Human and machine recognition of faces: A survey. In: Proc of the IEEE, 1995, 83 (5): 705~740.
- 2 Sung K, Poggio T. Example-based learning for view-based human face detection, IEEE Trans Pattern Analysis and Machine Intelligence, 1998, 20(1): 39~51.
- 3 Brunelli R, Poggio T. Face recognition: Features versus templates, IEEE Trans Pattern Analysis and Machine Intelligence, 1993, 15(10): 1042~1052.
- 4 Yuille A, Hallinan P, Cohen D. Feature extraction from faces using deformable templates. Int J Computer Vision, 1992, 8(2): 99~111.
- 5 Yang G Z, Huang T S. Human face detection in a complex background. Pattern Recognition, 1994, 27(1): 53~63.
- 6 Rowley H A, Baluja S, Kanade T. Neural network-based face detection. IEEE Trans Pattern Analysis and Machine Intelligence, 1998, 20(1): 23~38.
- 7 Samal A, Iyengar P A. Automatic recognition and analysis of human faces and facial expression: A survey. Pattern Recognition Society, 1992, 25(1): 65~67.



梁路宏 1997 年获得清华大学计算机系学士学位,现为清华大学计算机系博士生。研究方向是计算机视觉与模式识别。



艾海舟 1985、1988、1991年分别获得清华大学计算机应用专业学士、硕士、博士学位,毕业后留校任教,1994—1996年为比利时布鲁塞尔自由大学博士后,现为清华大学计算机系副教。主要从事计算机视觉、模式识别、移动机器人技术的研究,发表论文20余篇。



何克忠 清华大学计算机系教授,1962年毕业于清华大学自动控制系。主要从事计算机控制、智能移动机器人技术的研究,先后主持研制了智能移动机器人 THMR-I、THMR-II、THMR-III,多次获部委科技进步奖,出版专著3部,发表论文40余篇。

Multi-Template-Matching-Based Single Face Detection

Liang Luhong, Ai Haizhou and He Kehong

(Computer Department of Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, Beijing 100084)

Abstract This paper develops a multi-template-matching-based method for single face detection in general background images. Firstly two-eyes-in-whole template is used for determining the potential location of the face roughly. Secondly 5 whole face templates with different width-height ratios are used for searching and localization. Finally some edge features are used for verifying the detection results. Input image is resampled via constant ratio in order to deal with the problem of multi-scale detection. Experiment results show that it is relatively fast and rather successful.

Keywords Template matching, Face recognition, Face detection

Creative 推介 Sound Blaster Live! Digital 数码版

1999年9月27日 Creative Technology Ltd. (NASDAQ:CREAF), 在个人娱乐多媒体解决方案领域中的领导者, 宣布 Sound Blaster® Live! Digital 数码版声卡将于今天面世。此款声卡集各项优势于一身, 让网络音乐爱好者可以自如地以 CD 音质创作、编排与播放 MP3 音乐及其它音频文件。Sound Blaster Live! 数码版还有一项别具一格的视频动画技术—LAVA!™ (Live Audio Visual Animation), 其实时产生的动画对 MP3 音乐和其它网络音乐文件起到了很好的视听丰富作用。

Sound Blaster Live! 数码版是基于去年一经推出便屡获殊荣、成为业界标准的 Sound Blaster Live!。Sound Blaster Live! 数码版采用强劲的 EMU10K1™ 音频处理器, 使得不论是音乐爱好者或是音乐家都可以方便地以数码到数码处理方式创作音乐。创作出的音质可媲美录音室, 同时也能播放 MP3、MIDI、WAV、Internet 及 CD 音频文件。Sound Blaster Live! 数码版的用户们也可以轻松地他们的 CD 藏品中抓取音轨并转换为 MP3 格式的文件, 在 PC 上存储和播放。

LAVA! 使网络音乐呈现出全新形象, 使用者可通过它创造出引人入胜的视觉拼接效果, 使网络音乐更具活力, 例如伴随着 MP3 音乐同步播放的 3D 图象。LAVA! 播放器最近于 Woodstock 99 摇滚音乐节上首度推出, 它可根据音乐的不同频率做出回应、实时产生出图象, 从而创造出了即时 3D 音乐景象。

LAVA! 占用的硬盘空间不足 100KB。使用者可在 8 种不同的模板中进行选择, 然后对 LAVA! 文件进行自定义设置, 许多元素如乐队标志、封套设计及说明文字都可自行编辑。在应用到 MP3 文件后, 集音乐图象于一身的一体化文件就可以发给亲朋好友或放置在某个站点供下载或观看。LAVA! 由 Creative-Emu 技术中心开发, 同时采用了技术先进的音频分析和硬件图形加速, 从而诞生出了可应用于网络音乐应用程序的全新视觉语言。

Sound Blaster Live! Digital 数码版的主要特性

- EMU10K1 数码信号处理器: 完全 32 位的音频信号处理能力, 提供极为纯净的声音和极为精确的 3D 定位音效。数码输出插孔可用于 Creative FourPointSurround™ FPS2000D 数码音箱系统的纯数码式连接。

- Creative 数码音频中心软件: 用户使用它可以对 MP3 文件进行编码、解码或归档处理, 还可以从个人 CD 藏品中将大量的 CD 音轨转换为 MP3 文件并归类编排目录。数码音频中心软件还捆绑有下列软件: Mixman®, CakewalkExpress®, Sound Forge®, Media Ring Talk®99。

(下转第 834 页)



陈晓峰 博士,毕业于中科院遥感应用研究所遥感与地理信息专业。主要从事遥感图象处理和地理信息工作。曾参加过国家重点工程“西藏—两江两河环境动态监测与评价”研究项目。曾获军内科技进步二等奖,三等奖。近年在国内外发表论文近10篇,专著1本。



刘建华 工程师,获遥感与摄影测量专业硕士学位。从事遥感信息分析、处理工作。

Construction of a Neural Network on Remote Sensing Image Classification

Zhang Jianbao

(Department of Computer, Northwest University, Xi'an 710068)

Chen Xiaofeng, Liu Jianhua

(Information Advanced Technology Institute, P. O. Box 2869, Beijing 100085)

Abstract Radial basis function neural network (RBFNN) and multilayer perceptron neural network (MPNN) are very similar with topologic structure. They play very similar roles in providing techniques for approximating arbitrary nonlinear function mappings between multi-dimensional space. Two models are compared in this paper. This paper describes a method to construct an RBFNN classifier by a fast clustering algorithm and computes the optimal weights between the middle layer and output layer statistically. We applied the proposed method to construct a RBFNN classifier for remote sensing image classification. The experiment showed that the method could construct classifier quickly and the performance of the classifier was better than others.

Keywords Radial basis function neural network (RBFNN), Multilayer perceptron neural network (MPNN), Remote sensing image classification, Linear discriminant function

(上接第830页)

• CD音质的声音:具有一个编码器,可产生超卓音质的MP3文件供下载到便携式MP3装置,如Creative出品的NO-MAD数码音频播放器。Creative的8点插值专利技术有效地改善了回放音质。

• 集成软件包:包括所有必须的软件供创建、编排MP3\MIDI\WAV和Internet音频文件之用。

• Environmental Audio™环境音效支持:用户可通过此功能使MP3文件在编码播放时具有更多的效果如“音乐厅”或“隧道”等等。使用预置功能,用户可随时添加或调用想要的效果,以丰富视听。

• LAVA!;Sound Blaster Live! 数码版的用户使用LAVA!可以将视频动画结合到他们的音频文件中去,制作出“即时音乐录象”共享给朋友或作歌曲和乐队的宣传之用。

• Live! Ware™支持:Creative在线资源为硬件提供不断的升级——新的功能、新的格式和特性,可用于音乐创作、播放和玩游戏。

价格和购买办法:Sound Blaster Live! Digital 数码版现在可以通过Creative广泛的发行和零售网络购买,估价为¥690.00元零售。