

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06K 9/60 (2006.01)



[12] 发明专利说明书

专利号 ZL 00805011.2

[45] 授权公告日 2006 年 9 月 20 日

[11] 授权公告号 CN 1276384C

[22] 申请日 2000.10.27 [21] 申请号 00805011.2

[30] 优先权

[32] 1999.11.17 [33] US [31] 09/441,943

[86] 国际申请 PCT/EP2000/010730 2000.10.27

[87] 国际公布 WO2001/037212 英 2001.5.25

[85] 进入国家阶段日期 2001.9.14

[71] 专利权人 皇家飞利浦电子有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 L·阿尼霍特里 N·迪米特罗瓦

J·H·埃伦巴尔斯

审查员 齐 霁

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 吴立明 王忠忠

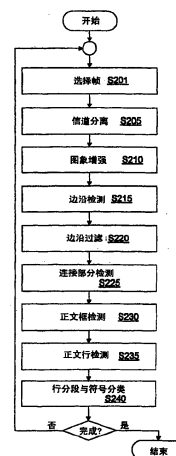
权利要求书 3 页 说明书 19 页 附图 12 页

[54] 发明名称

视频流可分类符号分离的方法与系统

[57] 摘要

一种用于分类符号的图象处理设备与方法依赖于隔离符号区的连接部分技术。该设备与方法从边沿检测过滤器的作用所导出的图象中构成连接部分。来自过滤后的图象的连接部分的构成定义字符形状的边沿。作为结果,实质上减少了定义各连接部分所必须连接的象素数目并从而提高了处理速度。主要对视频流中的正文讨论了本方法的应用。



1. 一种用于分类包含符号的图象数据流中的符号的设备，包括：
图象数据存储单元（140），具有一个连接来从所述图象数据流（170）中捕捉数据的输入，并具有一个输出（175）；

5 图象处理器（120），被连接至所述图象数据存储单元的输出，该图象处理器包括：

用于检测在所述图象数据存储单元中存储的、所述捕捉到的图象数据的至少一部分中的边沿的边沿检测器；

10 用于基于所述检测到的边沿构成包含边沿象素与非边沿象素的边沿图象的边沿图象构成器；

用于从所述边沿象素定义至少一个连接部分的连接部分定义器，所述连接部分基本上是只包含边沿象素的邻接区；

15 用于把其邻域包含多于规定数目的边沿象素的非边沿象素标记为边沿象素的边沿象素标记器，由此闭合只包含边沿象素的分开的邻接区之间的小空隙；

用于定义所述部分中的子图象的子图象定义器，所述子图象的高度与宽度基本上与所述至少一个连接部分的最大高度与宽度相同；以及

20 用于分类所述子图象的子图象分类器，由此识别一个符号集与所述子图象之间的最佳匹配。

2. 一种用于分类包含符号的图象数据流中的符号的设备，包括：
图象数据存储单元（140），具有一个连接来从所述图象数据流（170）捕捉数据的输入，并具有一个输出（175）；

25 该图象处理器（120），被连接到所述图象数据存储单元的输出上，该图象处理器包括：

用于过滤所述图象数据存储单元中存储的所述图象数据的至少一部分的阈值边沿过滤器，所述阈值边沿过滤器区分边界象素的邻接集合，该边界象素的邻接集合限定了非边界象素的至少一个邻接集合的范围；

30 用于找出包围边界象素的所述邻接集合的最小包络的包络查找器；

用于响应非边界象素的规定邻域中的边界象素的数目把非边界象

素标记为边界象素的边界象素标记器，从而使得在只包含边界象素的、分开的邻接区之间的空隙被闭合；

用于在对应于所述包络的所述部分中定义子图象的子图象定义器；以及

- 5 用于分类所述子图象的子图象分类器，由此识别在一个符号集与所述子图象之间的最佳匹配。

3. 一种用于分类包含符号的图象数据流中的符号的方法，包括下述步骤：

- 10 把一个边沿检测过滤器施加到一个图象上，以便识别所述图象中的边沿象素；

从所述边沿象素构成连接部分；

把其邻域包含多于规定数目的边沿象素的非边沿象素标记为边沿象素，由此闭合在只包含边沿象素的、分开的邻接区之间的小空隙；

- 15 定义符号划界区，每个符号划界区均包含所述连接部分中对应的一个连接部分；以及

分类所述图象的子图象，每个子图象均与所述符号划界区中相应的一个符号划界区相重合。

4. 一种用于分类包含符号的图象数据流中的文本的方法，包括下述步骤：

- 20 把一个边沿检测过滤器施加到一个图象上，以便识别所述图象中的边沿象素；

从边沿象素构成连接部分；

把其邻域包含多于规定数目的边沿象素的非边沿象素标记为边沿象素，由此闭合在只包含边沿象素的、分开的邻接区之间的小空隙；

- 25 响应所述构成步骤的结果定义文本区，使得所述文本区中的每一文本区均划出所述连接部分中的多个连接部分的边界；

响应在所述连接部分之间的空间划分所述文本区，以致识别所述图象的单个字符区；以及

- 30 分类所述图象的子图象，每个所述子图象均与所述单个字符区中相应的一个单个字符区相重合。

5. 一种用于分类包含符号的图象数据流中的符号的设备，包括：

图象处理器（120），可连接到所述图象数据流（180）上来从其

抽取图象，该图象处理器包括：

用于过滤所述图象以便识别所述图象中的边沿象素的边沿检测过滤器；

用于从所述边沿象素构成连接部分的连接部分构成器；

5 用于把其邻域包含多于规定数目的边沿象素的非边沿象素标记为边沿象素的边沿象素标记器，由此闭合在只包含边沿象素的、分开的邻接区之间的小空隙；

用于定义符号区的符号区定义器，每个所述符号区均包含所述连接部分中的一个连接部分；和

10 用于分类所述图象的子图象的子图象分类器，每个所述子图象均与所述符号区中相应的一个符号区相重合。

6. 一种用于分类包含符号的图象数据流中的文本的设备，包括：

图象处理器（120），可连接到图象数据流（180）上以便抽取图象，该图象处理器包括：

15 用于过滤所述图象以便识别所述图象中的边沿象素的边沿检测器；

用于从施加了所述边沿检测过滤器而得出的边沿象素构成连接部分的连接部分构成器；

20 用于把其邻域中包含多于规定数目的边沿象素的非边沿象素标记为边沿象素的边沿象素标记器，由此闭合在只包含边沿象素的、分开的邻接区之间的小空隙；

用于响应所述构成步骤的结果定义文本区的文本区定义器，从而使得每个所述文本区均划出多个所述连接部分的边界；

25 用于响应在所述连接部分之间的断裂划分所述文本区的文本区划分器，以便识别所述图象的单个字符区；以及

用于分类所述图象的子图象的子图象分类器，每个所述子图象均与所述单个字符区中相应的一个单个字符区相重合。

30 7. 如权利要求6所述的设备，其中，所述图象处理器还包括用于生成所述子图象的投影的子图象投影生成器，以及所述文本区划分器被操作用于响应所述投影划分所述文本区。

视频流可分类符号
分离的方法与系统

5 相关申请的交叉引用

本发明与共同转让给本发明的受让人的名为“利用视频帧中检测到的正文分析视频内容的系统与方法”的1999年8月9日提交的美国专利申请序号09/370,931中所描述的发明相关,现通过引用将其全文结合在此如同在这里全面陈述的一般。本发明还与名为
10 “检测与定位视频中的正文的方法与装置”的共同转让给本发明的受让人的1999年1月28日提交的美国临时专利申请号60/117,658中所公开的发明相关。这里为一切目的将该相关临时专利申请的公开通过引用结合在此,如同在这里全面陈述的一般。本发明还与共同转让给本发明的受让人的名为“用作用在神经网络上的形状特征的
15 的符号分类”的与本发明同时提交的申请中所公开的发明相关。通过引用为一切目的将该相关临时专利申请的公开结合在此,如同在这里全面陈述的一般。

背景

20 本发明涉及识别数字化图象中的模式的系统,而更具体地涉及分离视频数据流中诸如正文字符等符号的这些系统。

实时广播、模拟磁带及数字视频对教育、娱乐及许多多媒体应用是重要的。随着视频收集量的规模以百万小时计,需要插入视频数据的技术以便能更高效地使用与访问这一资料。已提出了各式各样的这种增强的使用。例如,正文与声音识别的使用能导向原始视
25 频的摘要的建立及索引视频内容的关键的自动生成。另一应用范围有赖于广播(或多道播放等)视频数据流中正文与/或其它符号的快速实时分类。例如,正文识别可用于诸如视频内容索引等任何适当目的。

已利用各种正文识别技术来识别数字化模式。最普通的实例是文件
30 光学字符识别(OCR)。所有这些技术的通用模型为从图象中导出输入矢量,该输入矢量刻划原始模式的特征。将该矢量映射到固定数目或范围的符号类别之一上来“识别”该图象。例如,位图图象的象

素值可用作输入矢量而对应分类集可以是诸如英文字母表等字母表。没有特定的模式识别技术已达到普适的统治地位。各识别问题具有其本身的难点集合：分类集的大小、输入矢量的大小、要求的速度与精度、及其它问题。同时，可靠性是呼唤几乎每一应用领域中的改进的领域。

5 作为上述不足的后果，模式识别是不断积极研究的领域，各种应用根据它们各自的诸如实用性与可行性标准受到不同程度的关注。也许这些技术中最成熟的是对正文字符的模式识别的应用，即光学字符识别（OCR）。这一技术由于将印刷品转换成计算机可读的字符的客观需要与实用性而得到开发。从实用观点上，印刷文件提供相对清楚与坚实的数据源。这些文件通常以对均匀的背景高对比度模式集合为特征并可用高分辨率存储。例如，印刷文件可在任意分辨率上扫描以形成印刷字符的二进制图象。同时，存在着对模式识别的这一应用的明显需求，由于从文件到基于计算机的文本免除了键盘打字的劳动，实现经济的数据存储，允许搜索文件，等。

10 由于随着执行符号或字符分类而来的困难，一些应用领域受到的关注不够。例如，至少由于下述因素而视频流中的模式识别是困难的领域。视频流中的字符趋向于在空间上不均匀（有时在时间上变化的）背景上显出、带有低分辨率、及低对比度。因此识别视频流中的字符是困难的并且没有已知的可靠方法。此外，对于某些应用，如至少在上述相关申请中所公开的，高度希望快速识别速度。

在许多出版物中已描述了索引与分类视频的系统与方法，其中包含：M. Abdel-Mattaleb 等人的“CONIVAS：基于内容的图象与视频访问系统”，ACM 多媒体学报，427-428 页，Boston（1996）；
25 S-F. Chang 等人的“Video Q：使用视觉提示的自动化的基于内容的视频搜索系统”，ACM 多媒体学报 313-324 页，Seattle（1994）；
M. Christel 等人的“信息媒体数字视频库”，ACM 通讯，卷 38，第 4 期 57-58 页（1995）；N. Dimitrova 等人的“消费者设备中的视频内容管理”，IEEE 知识与数据工程会报（1998 年 11 月）；
30 U. Gargi 等人的“数字视频数据库中索引正文事件”，模式识别国际会议，Brisbane，916-918 页（1998 年 8 月）；M. K. Mandal 等人的“使用矩与小波的图象索引”，IEEE 消费者电子设备会报，

卷 42, 第三期 (1996 年 8 月); 以及 S. Pfeiffer 等人的“自动提取数字运动”, 视觉通信与图象表示期刊, 卷 7, 第 4 期, 345 - 353 页 (1996)。

5 在 Ohya 等人的“在场景图象中识别字符” IEEE 模式分析与机器智能会报, 卷 16, 214-224 页 (1994 年 2 月), 中描述了用局部阈值法抽取字符及用估算邻接区之间的灰度级差检测包含字符的图象区。Ohya 等人进一步公开了检测到的具有紧密接近性与类似灰度级的区的合并, 以便生成字符模式候选。

10 在 A. Hauptmann 等人的“用于视频分段的正文、语音与视觉: 信息媒体项目”, AAAI 综合语言与视觉的计算模型 1995 年秋季讨论会, 中描述了使用视频正文的空间上下文与高对比度特征合并带有互相紧密靠近的水平与垂直边的区以便检测正文。在“用于视频索引的自动正文识别”, SPIE 图象与视频处理会议 (1996 年 1 月), 中 R. Lienhart 与 F. Suber 讨论了用于减少视频图象中的彩色数目的非线性彩色系统。该对比文件描述了产生具有类似颜色的同类段的分裂与合并过程。Lienhart 与 Suber 使用各种试探法来检测同类区中的字符, 其中包含前景字符、单色或刚接字符、大小受限制的字符、及与周围区相比具有高对比度的字符。

20 在 A. K. Jain 与 B. Yu 的“图象与视频帧中的自动正文定位”, IEEE 模式识别学报, 2055-2076, 卷 31 (1998 年 11 月 12 日), 中描述了用于定位正文及将图象分成多个实前景与背景图象的多值图象分解的使用。在“用于基于内容的注释与检索的从视频自动抽取正文”, 模式识别国际会议会报, 618-620 页 (1998), 中 J-C. Shim 等人描述了使用广义区域标记算法来找同类区及分段与抽取正文。为了确
25 定正文的颜色与位置, 聚集了识别出的前景图象。

K. V. Mardia 等人在“图象分段的空间阈值方法”, IEEE 模式分析与机器智能会报, 卷 10, 919-927 页 (1988) 及 A. Perez 等人在“图象分段的迭代阈值方法”, IEEE 模式分析与机器智能会报, 卷 9, 742-751 页 (1987) 中描述了其它有用的图象分段算法。

30 在数字化位图中定位正文的各种技术是已知的。用于二进制化字符数据来形成能以白底黑色为特征的图象及用于在位图图象上执行字符识别的技术也是已知的。视频流中的正文与其它模式从容易分类

的可预测的、大的、与清楚的延伸到即使在原则上也只包含没有来自
辅助文字数据的协助来分类的不充分信息的粗糙的、短暂的、不可预
测地朝向与定位的。还正在进行研究来提高识别速度以及精度。因
此，本技术的当前状态中存在着改进空间，尤其是当诸如视频流数据
5 等应用将当前技术使用到最大限度时。

发明概述

简言之，用于分类符号的图象处理设备与方法依赖于用于隔离符号
区的连接部分技术。本设备与方法从应用程序边沿检测过滤器导出的
图象构成连接的部分。从经过过滤的图象构成连接的部分定义字符形
10 状的边沿。连接的部分是根据诸如面积、高度、宽度、高宽比等阈值
标准过滤的。结果，充分地减少了定义各连接部分必须连接的象素数
目并从而提高了处理速度。本方法的应用主要是相对于视频流中的正
文讨论的。

视频流中正文的字符分类器采用向后传播神经网络 (BPNN)，其
15 特征空间是从尺寸、平移、与旋转不变的与形状相关的特征导出的。
用上述技术精确隔离字符区能使这些特征空间可以实现。这些特征空
间的实例包含正则与不变矩及从细的阈值字符的 Delaunay 三角剖分
导出的角直方图。由于视频流中的字符的低劣分辨率，这些特征空间
提供对作为分类器的 BPNN 的良好匹配。

20 检测与分类出现在视频流中的正文的能力具有许多用途。例如，
可根据从这一正文中导出的分类特征化与索引视频序列及其部分。这
能导向索引、增强的搜索能力、注释特征、等。此外，视频流中的正
文识别允许提出上下文敏感的特征，例如到达响应广播视频流中网址
的出现所生成的网站的可调用的链路。

25 视频中的正文提出来自发展得很好但仍在成熟中的技术的文件
OCR 的非常不同的问题集合。文件中的正文趋向于单色而高质量的。
视频中，缩小的场景图象可包含噪声与不受控制的亮度。出现在
视频中的字符可具有不同颜色、尺寸、字体、朝向、粗细，背景
可以是复杂而随时变化的等。并且，视频符号识别的许多应用要求
30 高速度。

本发明所采用的用于分类视频正文的技术采用符号隔离的精确
高速技术。然后利用符号位图生成作用在 BPNN 上的与形状相关的特

征矢量。该特征矢量提供对于整体图象形状的较大强调而对于上面提出的可变性问题相对地不敏感。在隔离字符区的技术中，根据检测到的边沿定义连接部分结构。由于边沿检测产生比二进制化符号所占用的整个区总体上少得多的象素，生成连接部分的过程能够快得多。特征空间的选择也增进识别速度。利用模拟的 BPNN，输入矢量的大小能严重影响吞吐量。对于来自选择的特征空间的所使用的部分的选择性是非常重要的。当然，通过组合诸如矩与线段特征等不同特征的混合物，可构成复杂的特征空间。并且，当所选择的特征共用计算步骤时可实现计算经济性。

10 将要参照下面的示例性附图结合某些较佳实施例描述本发明，以便更全面地理解它。对于附图的参照，强调所示出的细节是用示例方式的并且只是为了本发明的较佳实施例的示例性讨论的目的，并且是为了提供认为是本发明的原理与概念方面的最有用与容易理解的描述而提出的。因此，并未试图展示比基本理解本发明所必需的更详细的发明的结构细节，利用附图所作的描述使熟悉本技术的人员明白在
15 在实践中可以如何实施本发明的多种形式。

附图说明

图 1 为展示可用于实现本发明的设备的图。

图 2 为展示按照本发明的实施例的字符分类方法的流程图。

20 图 3A 与 3B 示出包含可按照本发明的实施例分类的信息的视频屏幕中的正文区。

图 4A 示出来自视频帧的捕捉到的数字图象的正文段的外观。

图 4B 示出边沿检测过滤后的正文段。

图 4C 示出在边沿检测中或之前的若干级过滤的效果，注意这些
25 不是实际显示中间结果而是为了说明与本发明相关的概念的目的示出的。

图 5A 与 5B 示出按照本发明的实施例的边沿过滤的效果。

图 5C 示出能用在本发明中的缺口闭合算法的实例。

图 6A - 6D 示出按照本发明的实施例的正文线分段的技术。

30 图 7A 与 7B 为展示按照本发明的实施例的过滤过程建立与管理连接的部分的技术的流程图。

图 8 为展示按照本发明的实施例的字符分类方法的流程图。

图 9A - 9D 示出过滤分段字符来导出特征矢量前体。

图 10A 与 10B 示出在按照本发明的实施例的字符分类过程中的图象过滤步骤中的 Delaunay 三角剖分与 Voronoy 图阶段。

图 11A 与 11B 示出按照本发明的实施例的角直方图型特征空间。

5 最佳实施例详述

参见图 1，图象正文分析系统 100 采用视频处理设备 110、视频源 180、及有可能监视器 185 来接收视频输入并生成与存储嵌入其中的字符信息。视频处理设备 110 接收视频图象，分析帧，隔离正文区与字符区，并按照下面详细讨论的过程分类正文与/或字符区。视频是从视频源 180 施加的。视频源 180 可以是任何视频数据源，其中包括带模 - 数转换器 (ADC) 的 VCR、带数字化视频的盘、带 ADC 的有线盒、DVD 或 CD-ROM 驱动器、数字视频家用系统 (DVHS)、数字录像机 (DVR)、硬盘驱动器 (HDD) 等。视频源 180 能提供少量短剪辑或多个剪辑，其中包括较长长度的数字化视频图象。视频源 180 可提供诸如 MPEG - 2、MJPEG 等任何模拟或数字格式的视频数据。

视频处理设备 110 可包含图象处理器 120、RAM 130、存储器 140、用户 I/O 卡 150、视频卡 160、I/O 缓冲器 170、及处理器总线 175。处理器总线 175 在视频处理设备 110 的各元件之间传输数据。RAM 130 进一步包括图象正文工作空间 132 及正文分析控制器 134。图象处理器 120 提供对视频处理设备 110 的总体控制并执行图象正文分析系统 100 所需的图象处理，其中包含根据系统选择的与用户选择的属性的视频帧中的正文分析。它还包含实现编辑过程，处理供在监视器 185 上显示与/或存储在存储器 140 中的数字化视频图象，并在图象正文分析系统 100 的各种元件之间传输数据。对图象处理器 120 的要求与能力是本技术中众所周知的，除了本发明所需要的以外，无须详尽描述。

RAM 130 提供随机存取存储器供临时存储视频处理设备 110 生成的数据，这些数据不是系统中的部件提供的。RAM 130 包含用于图象正文工作空间 130 及正文分析控制器 134 的存储器，以及图象处理器 120 与相关设备所需的其它存储器。图象正文工作空间 132 表示在正文分析进程中在其中临时存储与特定视频剪辑相关的视频图象的 RAM 130 部分。图象正文工作空间 130 允许修改帧的拷贝而不影响原始数

据，从而稍后可以恢复原始数据。

在本发明的一个实施例中，正文分析控制器 134 表示专用于存储在系统或用户定义的正文属性的基础上执行视频图象分析的图象处理器 120 所执行的应用程序的 RAM 130 的部分。正文分析控制器 134
5 可执行众所周知的编辑技术，诸如场景之间的逐渐改变与边界检测，以及与本发明关联的视频正文识别的新颖技术。正文分析控制器 134 也可体现为 CD-ROM、计算机软盘、或可加载到存储器 140 中的可拆卸的盘端口或诸如视频源 180 中的任何地方的其它存储介质上的程序。

10 存储器 140 包括一或多个盘系统，其中包括可拆卸的（磁或光）盘，用于永久性存储包含需要的视频与音频数据在内的程序及其它数据。取决于系统要求，可将存储器 140 配置成与向和自视频源 180 以及系统的其余部分传输视频与音频数据的一或多条双向总线接口。必要时，存储器 140 能在视频速率上传输数据。将存储器 140 的大小定
15 为为了包含正文属性分析在内的编辑目的对若干分钟视频提供足够的存储器。取决于特定应用及图象处理器 120 的能力，可将存储器 140 配置成提供存储大量视频剪辑的容量。

用户 I/O 卡 150 可将各种用户设备（未示出）接口到图象正文分析系统 100 的其余部分上。用户 I/O 卡 150 将从用户设备接收的数据
20 转换成接口总线 175 的格式供传输到图象处理器 120 或 RAM 130 供图象处理器 120 以后访问。用户 I/O 卡 150 还传输数据给诸如打印机（未示出）等用户输出设备。视频卡 160 提供监视器 185 与视频处理设备 110 的其余部分之间通过数据总线 175 的接口。

I/O 缓冲器 170 通过总线 175 连接在视频源 180 与图象正文分析
25 系统 100 的其余部分之间。如上面讨论的，视频源 180 至少具有一条双向总线来与 I/O 缓冲器 170 接口。I/O 缓冲器 170 在要求的视频图象传输速率上向/自视频源 180 传输数据。在视频处理设备 110 内，I/O 缓冲器 170 将从视频源 180 接收的数据根据要求传输给存储器 140、图象处理器 120、或 RAM 130。同时传输视频数据给图象处理器
30 120 提供了在接收它们时便显示视频图象的手段。

现在参见图 2、3A 与 3B，视频处理设备 110 或任何其它适当设备能在诸如图 3A 与 3B 中所示的包含正文的视频序列上执行正文抽取与

识别操作（如图 2 中概述的）100。将单个帧 305 提交给图 2 中概述的过程而得出诸如 310、315、360、365、370、及 375 等单个正文区的隔离。注意可将该过程作用在综合的多个帧的整体上以减少背景的复杂性及提高正文的清楚性。这便是，当若干后面的帧包含相同的正文区时（而当正文区包含诸如类似的频谱密度函数等相同的信号特征时便能识别这一点），便能将多个接连的帧综合（如加以平均）。这趋向于使正文区更清楚并导致正文更好地与背景隔开。当背景是移动图象时，必须用这一过程减少背景的复杂性。注意当如在现代电视中为移动画面增强进行了时间积分时，也可从源获得这一信号平均的某些好处。从而，对于下面的讨论，在“单个”帧上的操作的概念并不限于单个“帧抓取”而在其上面进行图象分析的“帧”可能是一或多个接连的视频帧的组合。

起初，图象处理器 120 可分离视频图象的一或多帧的彩色及存储减少彩色的图象供在抽取正文中使用。在本发明的一个实施例中，图象处理器 120 使用红绿蓝（RGB）彩色空间模型来隔离象素的红色分量。图 4A 中示出帧中正文部分的可能外观的实例。红色分量通常是用于检测视频正文中占统治地位的白、黄、与黑色最有用的。这便是，对于覆盖（重叠）正文，隔离出的红色帧提供公共正文彩色的鲜明的高对比度边沿。当前的方法也可用于抽取不是覆盖在视频上而是诸如停留在广告牌或路牌上的膜片序列等作为其实际部分的正文。在这一情况中，红色帧可能不是最好用的。在这些情况中，灰度（alpha 信道）可提供最佳起点。注意在本发明的其它实施例中，图象处理器 120 可使用各种彩色模型，诸如灰度图象或 YIQ 视频帧的 Y 分量、等。

将隔离的帧图象存储在图象正文工作空间 132 中。然后在步骤 S210 中，在执行进一步的处理之前，可使捕捉到的图象锐化。例如，可采用下述 3×3 屏蔽：

$$\begin{array}{ccc} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{array}$$

其中各象素是其本身的 8 倍加上其各相邻象素的负值之和。上面的位图过滤器（或“屏蔽”）的矩阵表示是本技术中的普通符号。存在着许多本技术中已知的这种派生的过滤器。本发明设想采用各式各样的

不同技术中任何一种用于隔离正文区。上面只是一个非常简单的例子。过滤步骤可包含多遍，例如沿一个维度的梯度检测后随沿另一维度的梯度检测（同时在各自的正交方向上加以平顺）随后将两种过滤结果相加。在步骤 S210 中，可采用例如 R. C. Gonzalez 与 R. E. Woods 在“数字图象处理”，Addison-Wesley 出版公司（1992），中所描述的中值滤波器减少随机噪声。

边沿检测可采用另一边沿过滤器。通过这一过滤器，可以（及最好）放大锐化（红色、灰度、等）图象中的边沿并利用例如下述边沿屏蔽衰减非边沿：

$$\begin{array}{ccc} 10 & -1 & -1 & -1 \\ & -1 & 12 & -1 \\ & -1 & -1 & -1 \end{array}$$

其中各象素再一次是作用在其本身及相邻象素上的上述各自的系数（权重）之和。图 4C 中，示出前面的过滤步骤的结果。边沿过滤原始图象 163 得出差分图象 164，然后边沿增强该图象得出最终图象 165，将其提交给下面的过滤。

在步骤 S215 中，应用阈值边沿过滤器或“边沿检测器”。如果 $Edge_{m,n}$ 表示 $M \times N$ 边沿图象的 m, n 象素，而 $F_{m,n}$ 为从步骤 S210 得出的增强图象，下面的等式可用于边沿检测：

20 式 1

$$Edge_{m,n} = \sum_{i=-1}^1 \sum_{j=-1}^1 w_{i,j} F_{m+i,n+j} < L_{edge}$$

其中 $0 < m < M$ 及 $0 < n < N$ ，并且 L_{edge} 为一可以是也可不是常量的阈值。值 $w_{i,j}$ 为来自边沿屏蔽的权重。在边沿检测进程中可忽略不计最外面的象素。注意，在这一阈值运算中也可蕴含应用锐化过滤器。

边沿阈值 L_{edge} 为预定的阈值，它可以是固定值或变量值。使用固定阈值可导致过度的花白噪声并导致正文周围的固定边沿中的不连续性。已知的开放（如腐蚀后再膨胀）方法导致丢失部分正文。带有可变阈值的自适应边沿过滤器改善这些趋势并且是在采用静态阈值上的极大改进。

在步骤 S220 中，在调节边沿检测阈值的一种模式中，在用边沿检测器作用第一固定阈值之后，降低在固定阈值步骤中识别出的边沿象素的任何邻接（在规定的容差内）象素的局部阈值，并重新作用过滤器。在另一模式中，通过将平顺函数（假定结果是以大于 2 的象素深度存储的）作用在阈值步骤的结果上然后再一次与阈值比较便可容易地完成最后一操作。这会导致标记为非边沿的象素变成标记为边沿的。象素的阈值降低程度最好取决于标记为边沿的相邻象素的数目。它后面的原理是当相邻的象素为边沿时，当前象素较为可能是边沿。从降低它们的局部阈值得出的边沿象素不用于计算相邻象素的降低的阈值。

作为替代，可与低通加权函数一起使用固定阈值以保证将被强边沿象素（具有高梯度的象素）包围的单个或少数非边沿象素标记为边沿象素。事实上，上面描述的所有步骤 S210-S220 可用式 1 形式的单一数值运算来描述，但在求和上具有较宽的范围。将它们分成独立的步骤不应认为是必要的或限制性的，并可依赖于计算设备与软件的特点以及其它考虑。

一旦检测出了字符边沿，图象处理器 120 便执行初级边沿过滤以消除不包含正文或在其中不能可靠地检测到正文的图象区。例如，带有极少量边、非常低的边沿密度（单位面积中边沿象素数）、或边沿象素低聚集度（即它们并不构成长距离结构，如噪声）的帧可从进一步处理中排除。

图象处理器 120 可在不同的级上执行边沿过滤。例如可在帧级或亚帧级上执行边沿过滤。在帧级上，如果一帧的合理的一部分以上呈现为由边沿组成，图象处理器 120 可忽略该帧。作为替代，能应用诸如频谱分析等过滤函数来确定该帧是否可能具有太多的边沿。这可从该帧的强边沿对象的高密度得出。前提是过度复杂的帧包含高比例的非字符细节并且通过字符分类来过滤它是不相称地繁重的。

当采用帧级过滤时，图象处理器 120 维护边沿计数器来确定该图象帧中的边沿象素数目。然而，这可能导致跳过或忽略包含清晰的正文的帧，诸如带有噪声部分以及带有清晰的正文部分的帧。为了避免排除这些图象帧或亚帧，图象处理器 120 可在亚帧级上执行边沿过滤。为了这样做，图象处理器 120 可将帧分成较小的区。为了做到这

一点，例如，图象处理器 120 可将帧分成三组象素列及三组象素行。

接着，图象处理器 120 确定各亚帧中的边沿数目并相应地设定其相关计数器。如果一个亚帧具有比预定数目多的边，处理器可舍弃该亚帧。可根据处理图象区所需的时间量或它们相对于象素密度的大小会使识别精度低于要求的最小值的概率来设定每一区的预定最大边沿计数。可利用较大数目的亚帧来保证不丢失被识别为不能判读的区包围的较小的清楚正文区。

接着在步骤 S225 中，图象处理器 120 在前一步骤中生成的边沿上执行连接部分 (CC) 分析。这一分析在规定的容差范围内组合所有连接的边沿象素。这便是，将邻接另一边沿象素或在其一定距离以内的每一个边沿象素与该象素合并在一起。最终，这一合并过程定义结构或各具有连接或接近连接的边沿象素集的连接部分。这一点的动机在于假定各正文字符区对应于单一 CC。取决于图象捕捉的分辨率、向上抽样（在原始图象上用内插加上的象素的比例）或向下抽样（从原始图象消除的象素的比例）的程度，可将容差范围设定为任何适当的值。

参见图 4B，对应于连接的字符的 CC 之间的偶然的间隙或断裂可作为带有固定阈值的边沿检测的后果出现。例如，能发生在 171 或 172 上所示的断裂。上述边沿检测方案的使用有助于保证这些断裂的 CC 部分的合并。从图 5A 与 5B 左方字符中的断裂开始，CC 合并法得出将断裂 174、175 与 176 中的点识别为边沿点并将其分别合并到 181 与 182 处的单一连接的部分结构中。除了上面描述的特定方法以外，可用各种机制完成连接区中的“不良”断裂的闭合。例如，可在腐蚀或变细之后应用膨胀。为了避免增加边沿象素的总面积的后果，在检测连接部分之前膨胀后可随即变细。并且，可增加从式 1 的应用得出的二进制化阈值图象的灰度深度，然后可作用平顺函数及再一次执行阈值法（式 1）。存在着许多能用来完成所要求的闭合效果的图象处理技术。又另一替代方法是诸如在图 5C 中所示当象素基本上被连接的系列中的边沿象素包围时便将其标记为边沿。这便是，所示的 24 种情况中的每一种是带有 8 个相邻象素的一个象素。在这些情况的每一种中，邻接象素具有连接的系列中的 5 个或以上边沿象素。当然，连接的系列中的数目可以改变或在该组上增加特殊情况。此外，矩阵的

大小可以增加。有利于被诸如对图 5C 定义的算法标记为边沿的象素类型是认为较少可能是连续的断裂部分的那些象素。通过闭合（膨胀然后腐蚀）或通过屏蔽中使用较少的锐化或对阈值（应用式 1）进行预处理能得到类似结果。

- 5 CC 为确定为构成没有非边沿象素将其分成两部分的连接的系列的象素集合。构成各 CC 的一张表，它包含该结构中最左、最右、最上、及最下象素的坐标，连同诸如结构中心的坐标等该结构的位置的指示。同时存储的能有构成连接部分结构的象素的数目。注意象素计数表示特定连接部分结构的面积。可利用预定的系统与/或用户阈值
- 10 来定义连接部分结构的面积、高度与宽度的最大与最小极限以便确定将哪些连接部分结构向前传递到下一处理阶段。最后的步骤为过滤器确定 CC 是否有资格作为字符。可利用其它直观推断来组合太小的 CC 来使它们本身符合直观推断或分裂太大的 CC。

15 在步骤 S230 中，图象处理器 120 根据左下方象素的位置按升序排序满足前一步骤中的标准的连接部分。图象处理器 120 根据象素坐标排序。通过连接部分的排序的表来确定哪些 CC 构成正文块（“框”）。

20 图象处理器 120 将第一 CC 分配给第一个框并作为初始或当前分析的框。图象处理器 120 测试各后续的 CC 来判明其最下方象素是否位于与第一 CC 的对应象素相同的水平线（或靠近的水平线）上。即

25 如果其垂直位置靠近当前 CC 的垂直位置则将其加到当前正文框上。如果是，便假定属于同一正文行。垂直坐标差阈值可以是固定的或可变的。最好，第二 CC 的水平坐标的接近度是 CC 的高度的函数。还测试候选新增加者到当前正文框的水平距离来判明它是否位于可接受的范围内。

30 如果 CC 不符合与当前正文框合并的标准，生成带有将失败的 CC 标记为其第一元素的新的正文框。这一过程可对图象中的单行正文得出多个正文框。当系列中的下一个连接部分具有实质上不同的垂直坐标或低于最后的 CC 的水平坐标时，可在水平横截线的端点上闭合当前正文框并开始一新框。

35 对于各框，这时图象处理器 120 对初始字符合并过程所建立的各正文框执行第二级合并。这将可能已错误地解释为分开的正文的行并因而放置在分开的框中的正文框合并。这可能由严格的连接部分合并

标准或由于低劣的边沿检测导致，从而使同一字符得出多个 CC。

图象处理器 120 将各框与它后面的正文框用一组条件对比。两个正文框的多个测试条件为：

5 a) 一个框的底部在另一个的规定的垂直间隔之内，该间隔对应于预期的行间隔。并且，两个框之间的水平间隔小于基于第一框中的字符平均宽度的可变阈值。

b) 框之一的中心位于另一正文框的区域内，或

c) 第一框的顶部与第二正文框的底部重叠且一个框的左或右边分别在另一框的左或右边的少数象素之内。

10 如果满足上述条件中任何一项，图象处理器 120 从正文框表中删除第二框并将其合并到第一框中。图象处理器 120 重复这一过程直到互相相对测试了所有正文框并尽可能组合了为止。

在步骤 S235 中，如果框符合规定的面积、宽度与高度的制约，图象处理器 120 接受从步骤 235 得出的正文框作为正文行。对于各该
15 正文框，图象处理器 120 抽取对应于来自原始图象的正文框的子图象。然后，图象处理器 120 二进制化子图象准备字符识别。这便是，将彩色深度降低到 2，将阈值设定为保证将字符与背景适当地隔离的值。这是难题，它可包含若干步骤，诸如综合多帧来简化复杂背景。

用于二进制化图象的阈值可确定如下。图象处理器 120 通过计算
20 正文框中的象素的平均灰度值 (AvgFG) 来修正正文框图象。将其用作二进制化图象的阈值。同时计算的有正文框周围的区 (例如 5 个象素) 的平均灰度值 (AvgBG)。通过将 AvgFG 以上的任何象素标记为白并将 AvgFG 以下的任何象素标记为黑来二进制化子图象。计算标记为白的象素的平均值 Avg1，以及标记为黑的象素的平均值 Avg2。

25 一旦将正文框转换成黑白(二进制)图象，图象处理器 120 将 Avg1 及 Avg2 对 AvgBG 进行比较。将具有接近 AvgBG 的平均值分配为背景而将另一区分配为前景 (或正文)。例如，如果黑区平均值更接近 AvgBG，便将黑区转换成白色并将白区转换成黑色。这保证正文永远是一致的值供输入到 OCR 程序中。随后图象处理器 120 将抽取的帧正文
30 存储在图象正文工作空间 132 中并且进程在进程步骤 205 在下一帧上继续进行。注意在用局部阈值法之前，可执行超分辨率步骤来增强正文分辨率。

下面，在能进行分类以前必须隔离单个字符区。为了隔离正文行中的单个字符区，可用各种直观推断，例如字符高宽比、高度与宽度的上限与阈值等。这些直观推断通常落入各种维度特征的容许值的预测类别中。

- 5 由于原始正文中缺乏清楚性，连接部分有可能不能与字符对应。参见图 6A-6D，如果 CC 划分失败，可利用另一工具沿水平线划分字符。一个实例是作为水平坐标的函数并且其值是与当前正文框内所包含的并与该 X 坐标重合的垂直列中的前景象素的数目（也有可能如示出的灰度值）成比例的垂直投影 425。这便是，在其上面累计象素的垂直列不超过正文框的大小，因此只用这一方法测定字符的当前行。
- 10 也可用窗口函数 425 加权这一“灰度”垂直投影 425，窗口的宽度与序列中下一字符的预期宽度成正比。420 上示出用窗口函数 425 加重的结果。可用最小投影值来定义字符的左与右边。

- 参见图 7A，用于隔离字符区的方法从第一 CC 开始并顺序进行通
- 15 过正文框。在步骤 S310 上开始，选择第一或下一 CC。在步骤 S312 上对照维度直观推断测试所选择的 CC 来判定该 CC 是否符合它们。CC 上的直观推断测试可能指示该 CC 不大可能是整个字符或者它太大而有可能包含一个以上字符。如果在步骤 314 中发现该 CC 太大，在步骤 S316 中应用划分字符的替代方法，例如上述灰度投影。如果在步
- 20 骤 S322 中发现该 CC 太小，则在步骤 S318 中对照直观推断测试下一 CC。如果在步骤 S320 中这显示后面的 CC 也太小，则在步骤 S326 中将当前与后面的 CC 合并而流程返回到步骤 S310 直到隔离了所有字符区为止。如果后面的 CC 并不太小，则在步骤 S324 中丢弃当前 CC 而流程进行到步骤 S310。

- 25 参见图 7B，另一划分字符的方法保留直观推断失败的选择字符区并试图分类这些选择对象。在分类时，选择达到最高置信度级的选择对象。然后相应地对待其它字符区。例如如果用高置信度测定分类对应于两个合并的 CC 的图象，不再将对应于用来与第一 CC 合并的字段作为独立的字符字段对待。在步骤 S330 中，选择第一或下一 CC。在
- 30 步骤 S332 上，对照维度直观推断测试选择的 CC 来判定该 CC 是否符合它们。如果在步骤 S334 中发现该 CC 太大，在步骤 S336 中应用划分字符的替代方法。如果在步骤 S338 中发现该 CC 太小，保留当前 CC

及与下一 CC 组合的当前 CC 两者作为供两者择一的字符字段。当将这些字符字段提交给下述分类时，利用置信度测定来选择两者之一。然后流程返回到步骤 S310 直到隔离了所有字符区为止。如果步骤 S336 的分裂操作产生低置信度测定，则保留超大的与断裂的字段供在分类中用作选择对象，并用分类结果在选择对象之间进行选择。

5 注意 不一定非将与字符重合的区定义为用直线组成的。它们可以是橡胶带类型的边界区（任意边数的凸多边形）或正交凸直线多边形（直线组成的多边形其中连接内部两点的每一条水平或垂直线段全部位于内部）或基本上封闭预期的符号或字符的所关心的特征的任何其它适当的形状。

10 还注意可完全省略正文框的形成而直接用连接部分来识别候选字符区。然而，在这一情况中，预期会有较大数目的连接部分将超出将它们映射（分类）到其中的特定符号集合。还注意从上面的描述中很清楚可将上面的技术一般性地应用在符号分类上而限于文字字符分类。

15 参见图 8，一旦隔离了所有字符区（用步骤 S405 概括），便可顺序分类字符。下面在步骤 S410 中，选择第一或顺序的字符区。在步骤 S415 中，将原始图象的一部分（或其红色部分）提交给一些适当的图象分析来为特征分析作准备。例如，可将图象二进位化（用阈

20 值），得出灰度图象，二进位化与变细，等。预处理根据所使用的特征空间变化。

参见图 9A - 9D，例如，特征空间可利用某些特征点（如下面描述的）。特征点是可用构架字符识别的并从正规的视频字符（图 9A）导出它们，可将图象二进位化（图 9B）然后变细（图 9C）。然后可作为变细后的字符 460、470 的角点 465、弯点 466、交叉点 467 及端点

25 468 导出特征点（图 9D，465 - 468）。这种图象处理很适用于下述角直方图特征空间。可能需要较低级的图象处理来计算大小不变矩。注意同样可用其它特征点定义系统。

再参见图 8，可将原始字符提交给各种不同分析来定义可作用在

30 适当地训练的向后传播神经网络（BPNN）的输入上的特征矢量。对于采用大小不变矩的技术，可使用不变细或变细的字符。在步骤 S420 中，用适当的图象分析生成选择的特征矢量。可使用各式各样的图象

分析。已为与本专利相关的应用定义了若干不同特征空间。下面详细描述的定义的特征空间是大小与旋转不变的并认为特别适用于使用 BPNN 分类器的视频字符分类。

从变细后的字符的特征点导出第一特征空间，如图 9A-9D 所示。

- 5 参见图 10A 与 10B, 首先从特征点 12 导出 Delaunay 三角剖分(图 10A) 或 Voronoy 图(图 10B)。图象处理器 120 执行三角剖分, 然后为各三角形 1-6 生成内角的清单。然后它利用这一清单生成角的直方图, 如图 11A 中所示。该直方图简单地表示由三角剖分定义的三角形 1-6 的集合中给定大小范围的角 A、B 与 C 的频率。注意其它三角剖分法
10 或多边形生成法也能使用。例如, 参见图 10B, 可利用 Voronoy 多边形 17 与 18 定义一组角 A'、B'、与 C', 各与 Voronoy 图的顶点 14 关联。得出的角直方图起到从其导出特征点的特定字符的特征矢量的作用。

- 可在上面的特征空间上加上其它大小与旋转不变特征, 例如水平
15 线的数目、交叉点的数目、端点的数目、孔、拐点、中点、等。角直方图的另一变型为只使用各三角形的两个最大(或最小)内角。角直方图的又另一变型为使用二维角直方图而不是一维角直方图。例如, 参见图 11B, 各三角形的最大(或最小)角对定义 Delaunay 三角剖分(或 Voronoy 图的各项点)中各三角形的有序的对(用大小排序)。
20 各有序对的第一元素用于矩阵的第一维而第二元素用于矩阵的第二维。以这一方式, 保存角之间的相关性作为利用 BPNN 分类器训练与分类的信息。

- 认为特别适用于视频字符 BPNN 分类器的又另一特征空间是大小
25 不变矩的阵列。这些矩用下面的等式定义。虽然在各种情况中可使用大量单独的矩, 本申请中选择了特定的少数几种。象素位置与质心重合的象数下标 i , j 用下式给出:

$$\bar{i} = \frac{\sum_{i=1}^n \sum_{j=1}^m iB[i][j]}{A}$$

$$\bar{j} = \frac{\sum_{i=1}^n \sum_{j=1}^m jB[i][j]}{A}$$

其中当阈值图象的第 i, j 象素为前景象素时 $B[i][j]$ 为 1, 而否则为 0, 而 A 是由下式给出的前景象素的聚集区:

5

$$A = \sum_{i=1}^n \sum_{j=1}^m B[i][j].$$

平移不变矩由下式给出:

$$\eta_{p,q} = \frac{M_{p,q}}{M_{0,0}^\gamma}$$

10 其中 $M_{p,q}$ 为由下式给出的字符图象的第 p, q 个原始矩:

$$M_{k,j} = \sum_{i=1}^n \sum_{j=1}^m (i-\bar{i})^k (j-\bar{j})^l B[i][j]$$

和

$$\gamma = 1 + \frac{p+q}{2}.$$

15

为输入到 BPNN 选择的不变矩为:

$$\varphi_1 = \eta_{2,0} + \eta_{0,2};$$

$$\varphi_2 = 4\eta_{1,1}^2 + (\eta_{2,0} - \eta_{0,2})^2;$$

$$\varphi_3 = (3\eta_{3,0} - \eta_{1,2})^2 + (3\eta_{2,1} - \eta_{0,3})^2;$$

$$\varphi_4 = (\eta_{3,0} - \eta_{1,2})^2 + (\eta_{2,1} - \eta_{0,3})^2;$$

$$\varphi_5 = (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} - \eta_{0,3})[3(\eta_{3,0} - \eta_{1,2})^2 - 3(\eta_{2,1} - \eta_{0,3})^2] + (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} - \eta_{1,2})[(\eta_{3,0} - \eta_{1,2})^2 - 3(\eta_{2,1} - \eta_{0,3})^2]$$

5 和

$$\varphi_6 = (\eta_{2,0} - \eta_{0,2})[(\eta_{3,0} - 3\eta_{1,2})^2 - (\eta_{2,1} - \eta_{0,3})^2] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{2,1} - \eta_{0,3})$$

再参见图 8，在步骤 S2425 中，将各特征矢量作用在输出各种候选类并取决于输入有希望输出非常强的候选者的受过训练的 BPNN 上。如果存在多个候选字符，通过将 BPNN 输出的概率与推测的语言与上下文的使用频率数据组合，可在步骤 S430 中作出最佳猜测。这一数据可从不同类型的资料搜集，例如，电视广告转录文本、印刷资料、从因特网流送或下载的文件。一种组合方式是用与使用频率统计关联的对应概率来加权 BPNN 输出的概率。

15 对于熟悉本技术者显而易见本发明不限于上面的示例性实施例的细节，而可以以其它特定形式实施本发明而不脱离其精神或主要属性。例如，上面提出的正文分析描述了对水平对齐的正文的偏爱。很

明显相同方法可应用在诸如垂直对齐的正文、沿曲线的正文等其它对齐方式。

因此在所有方面都认为这些实施例是示例性而非限制性的，发明精神是由所附权利要求而非上面的描述指明的，因此旨在将凡是进入

5 权利要求的意义与等效范围内的所有改变都包括在此。

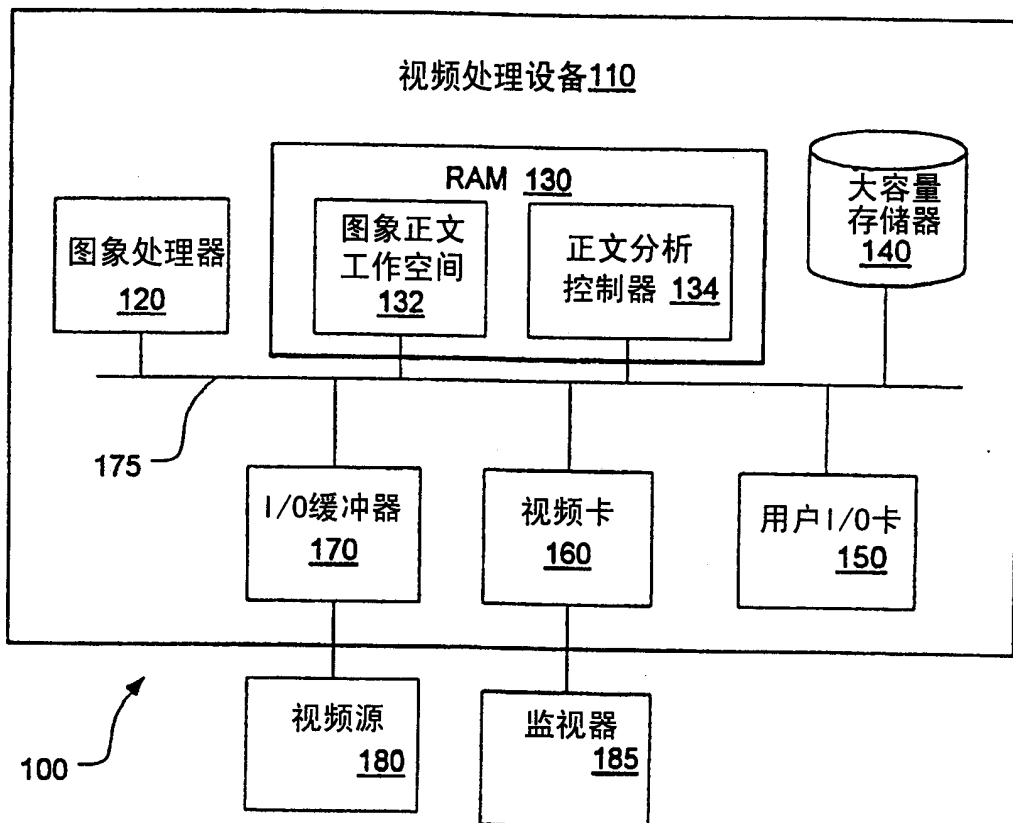


图 1

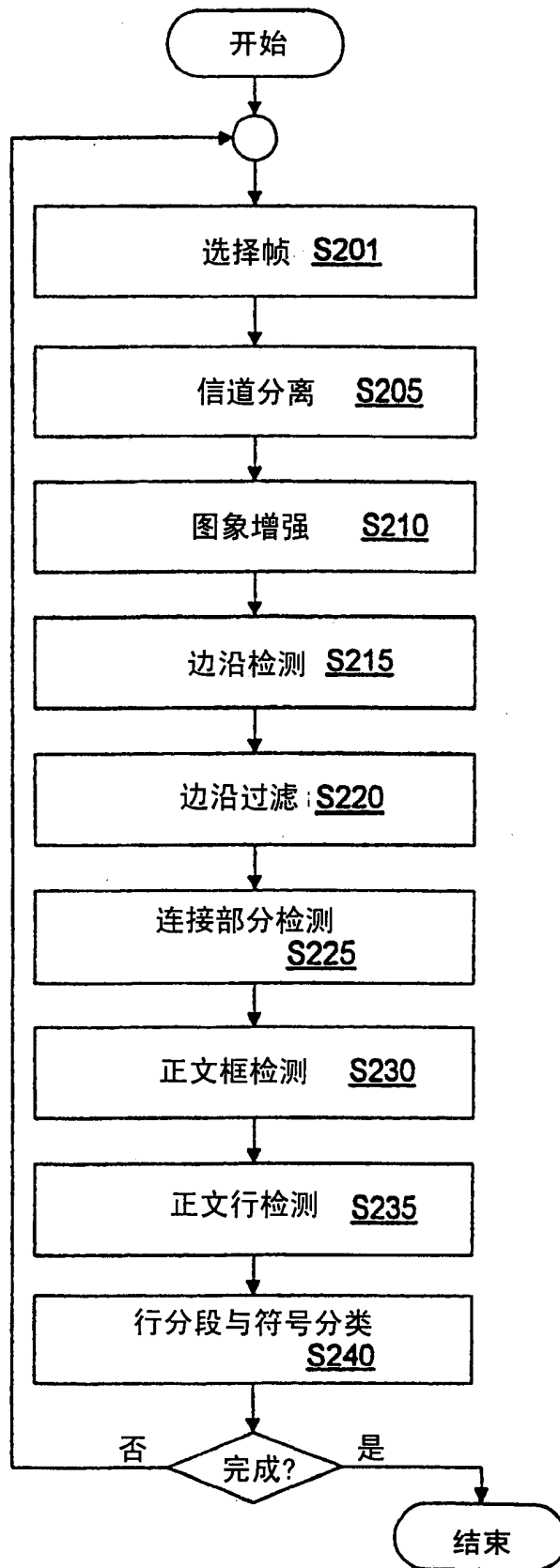


图 2

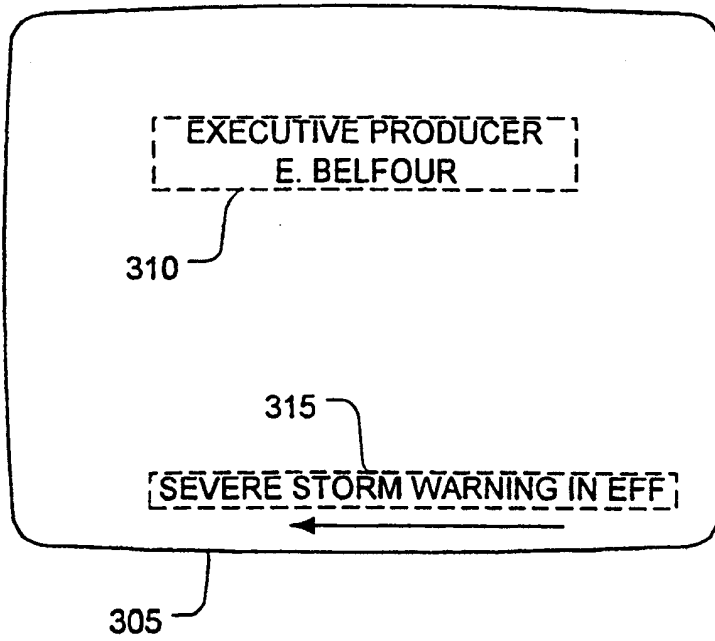


图 3A

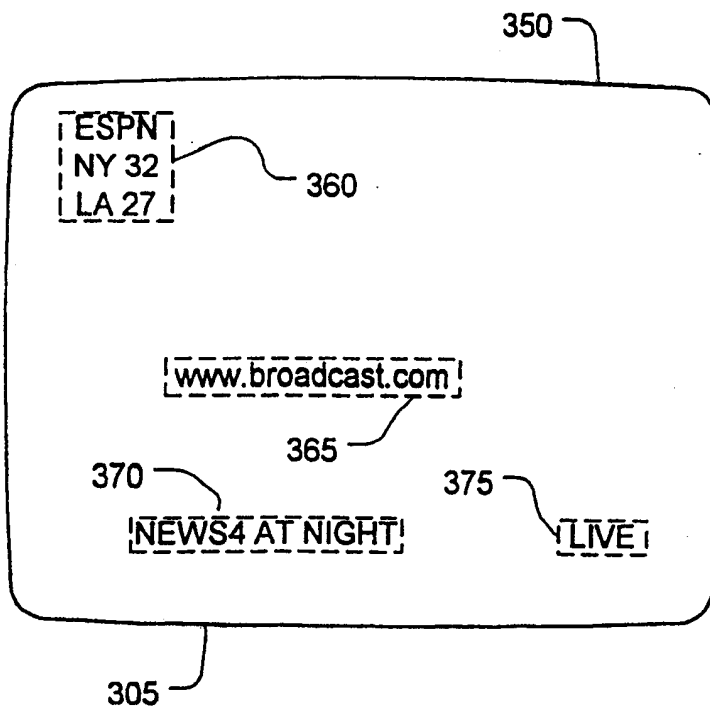


图 3B

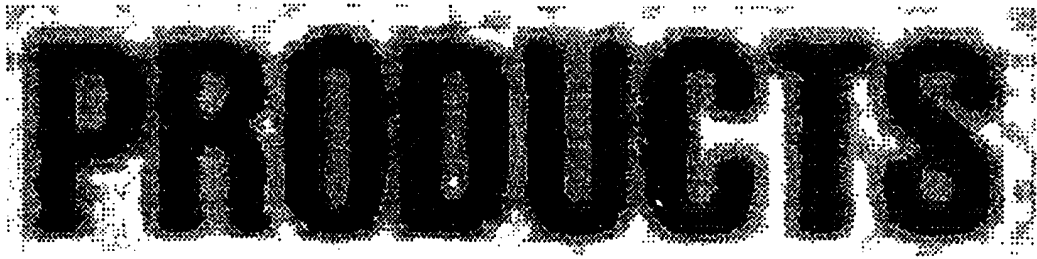


图 4A

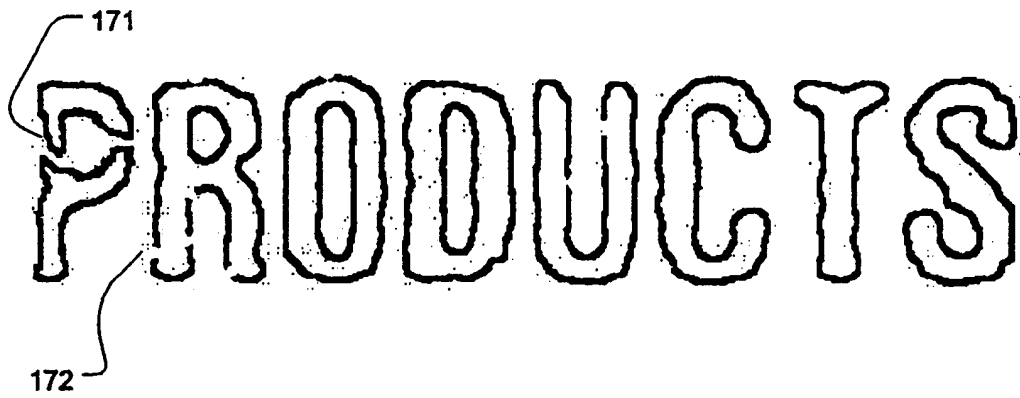


图 4B

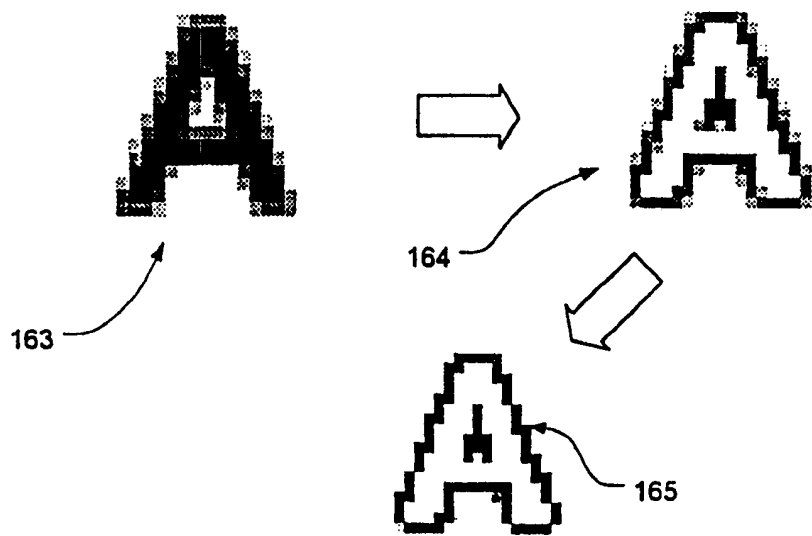


图 4C

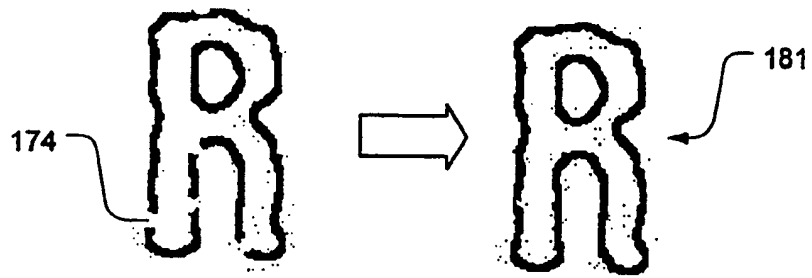


图 5A

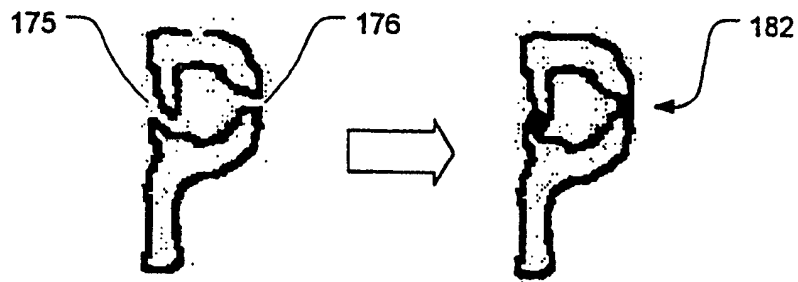


图 5B

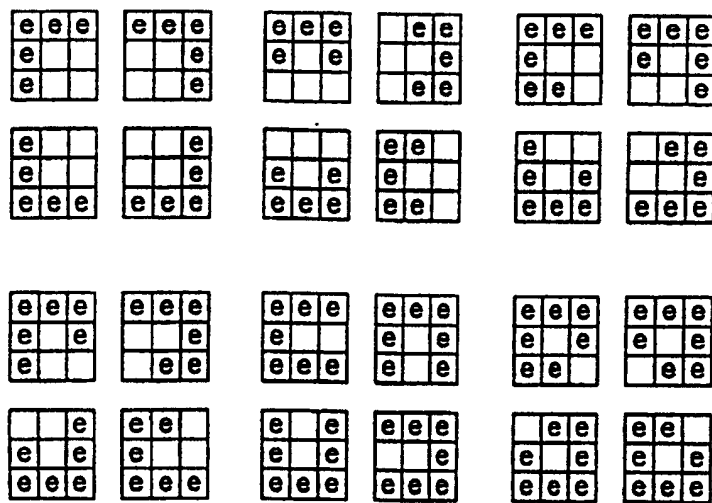
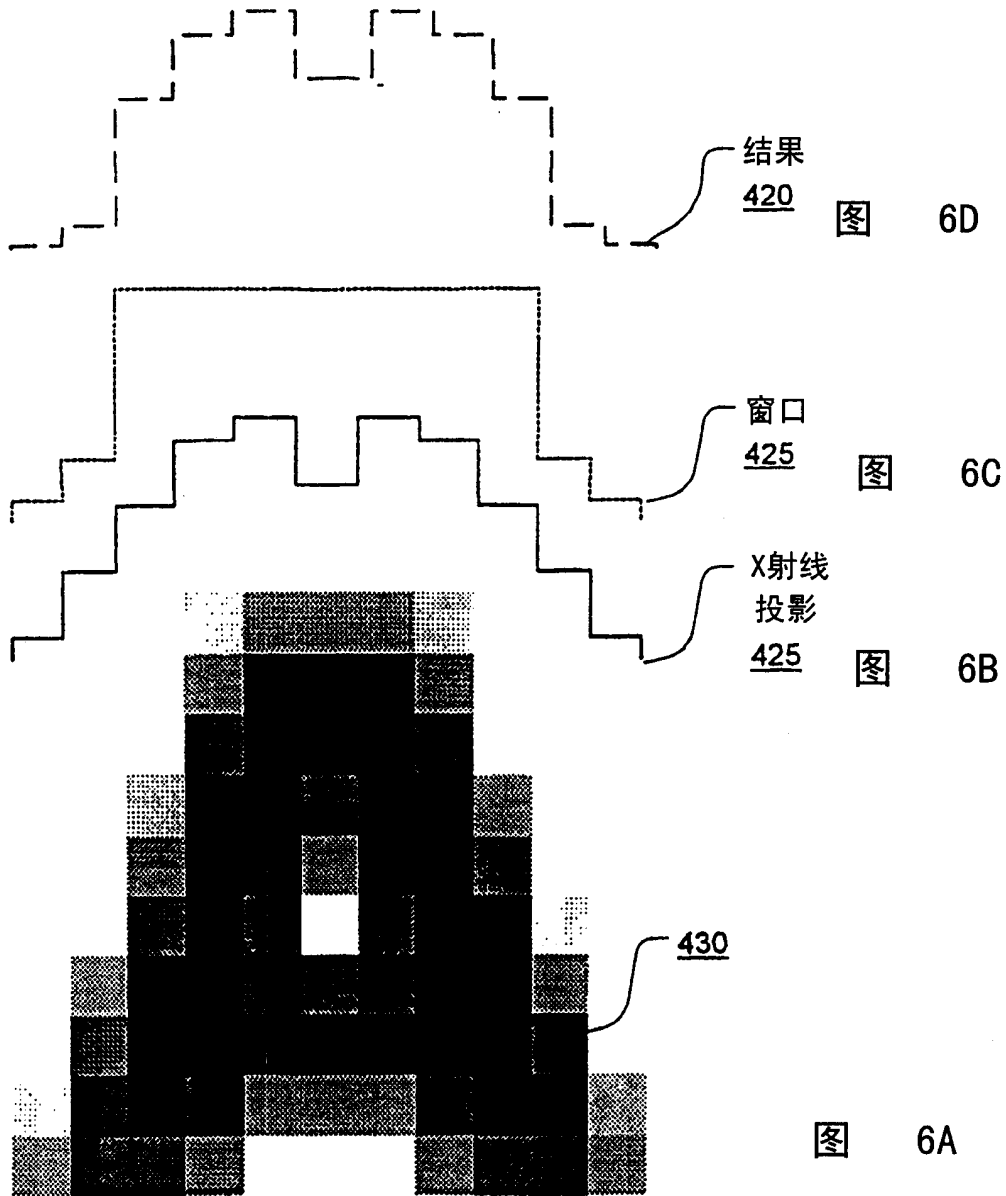


图 5C



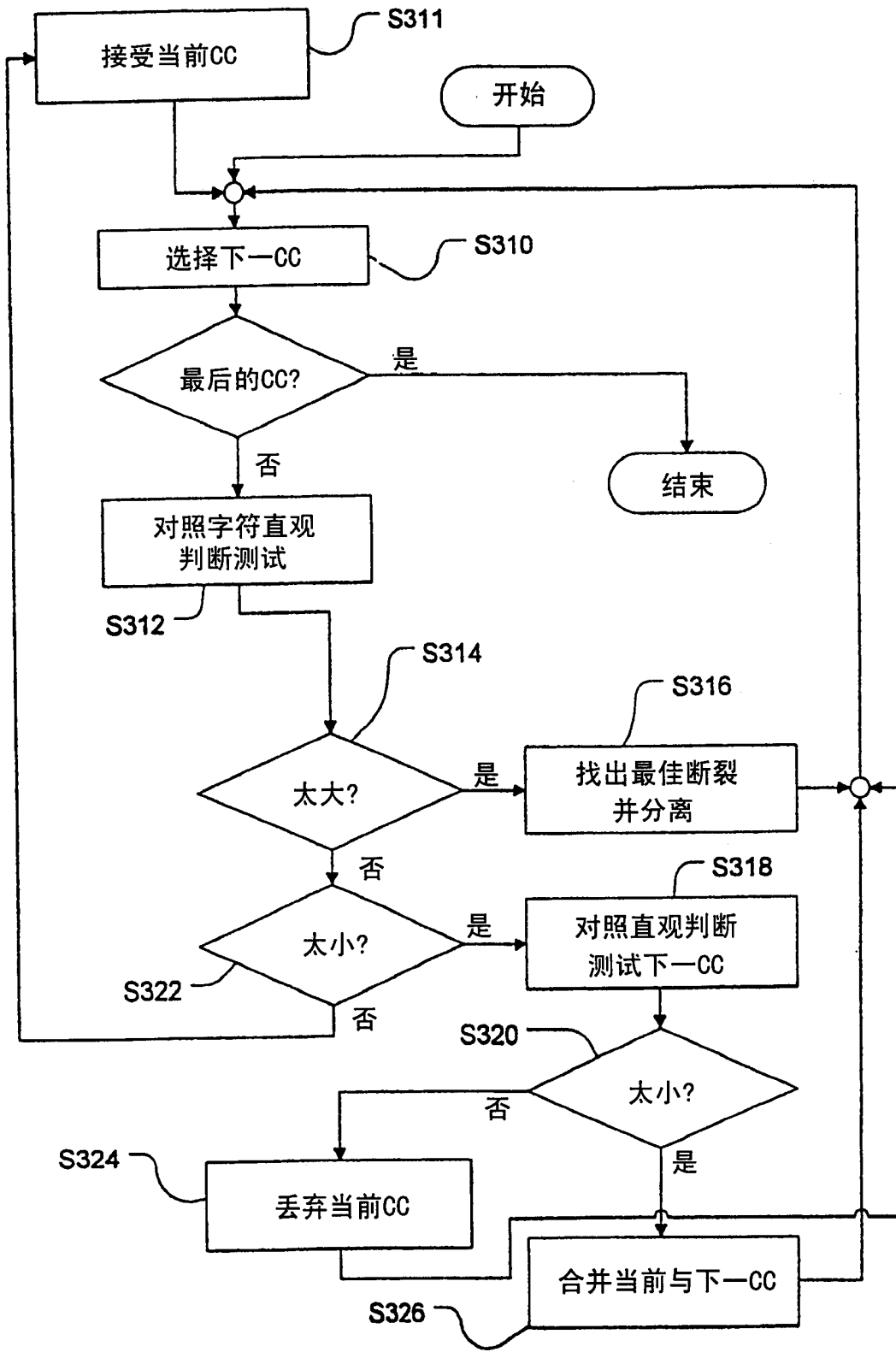


图 7A

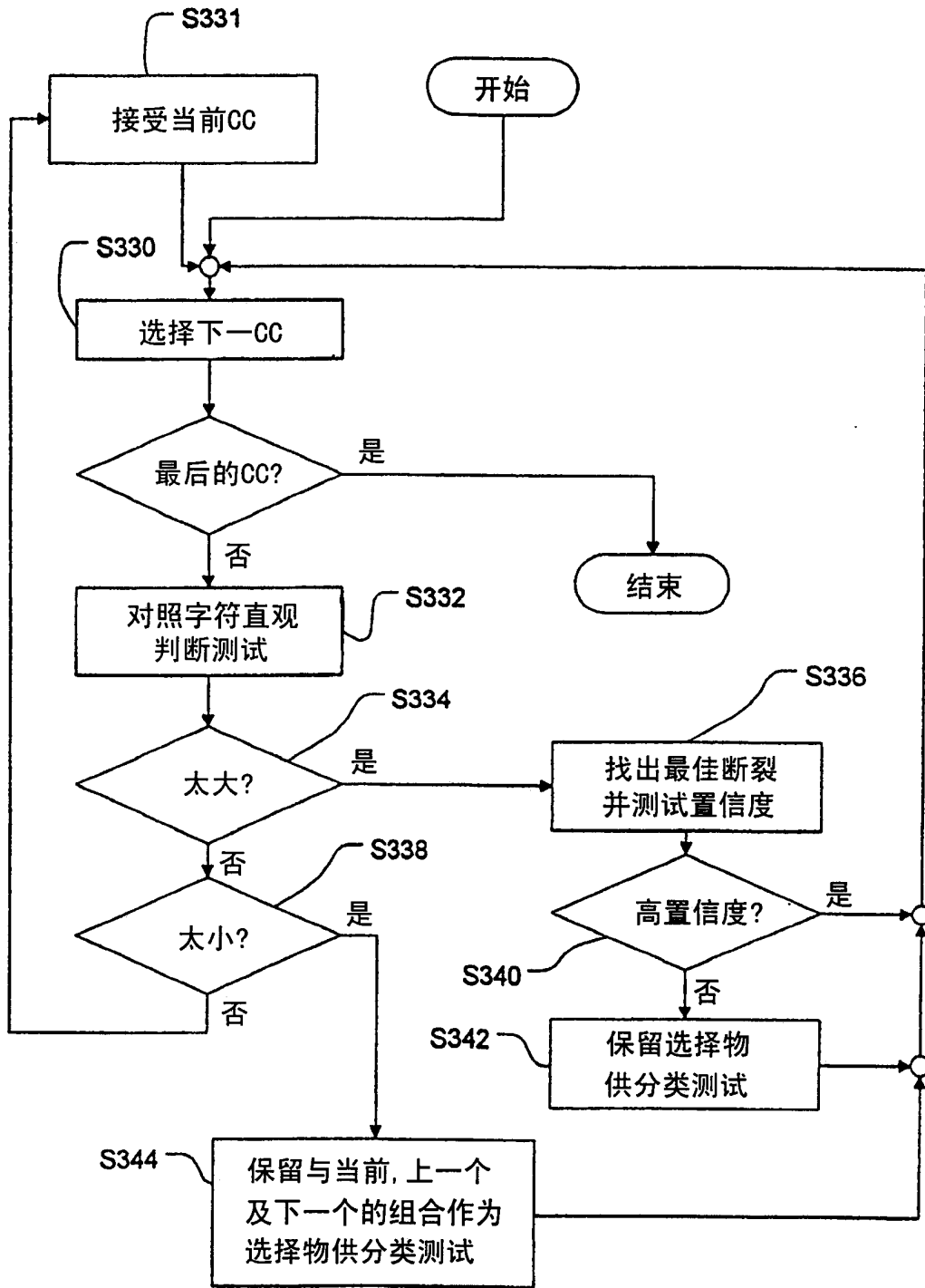


图 7B

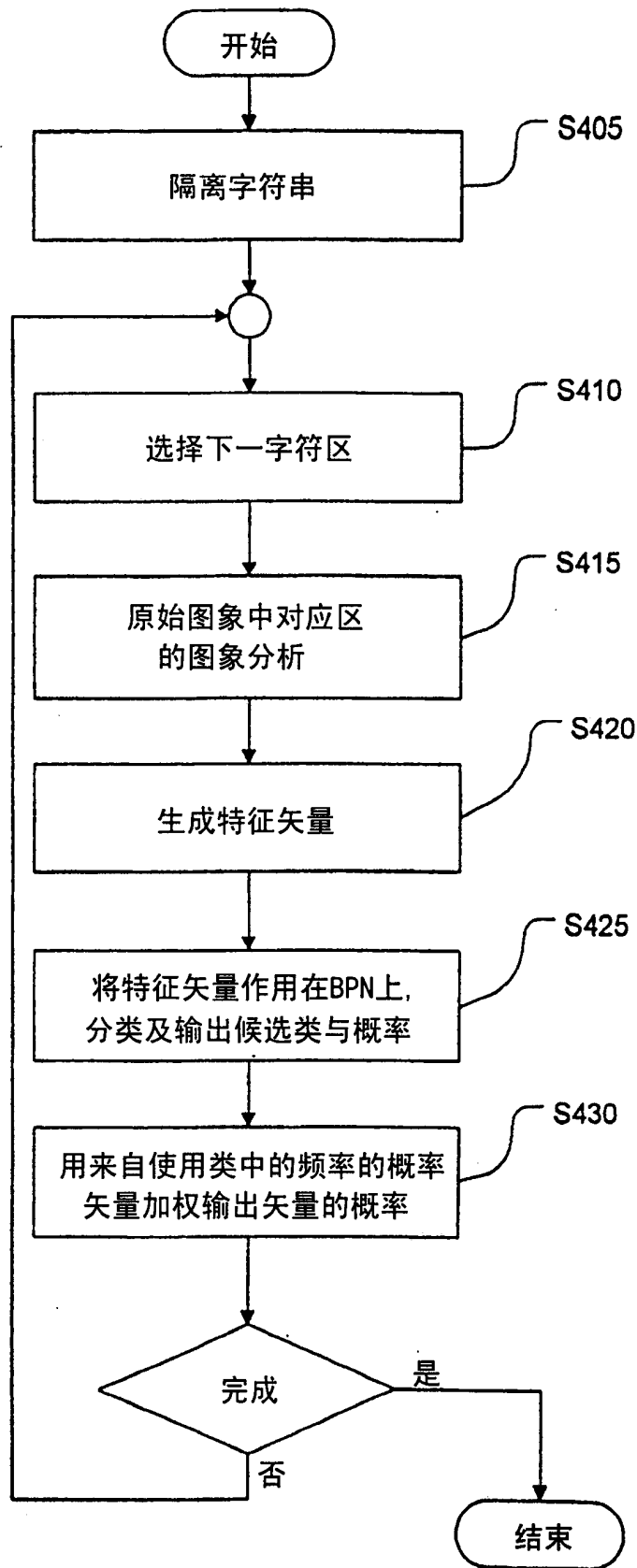
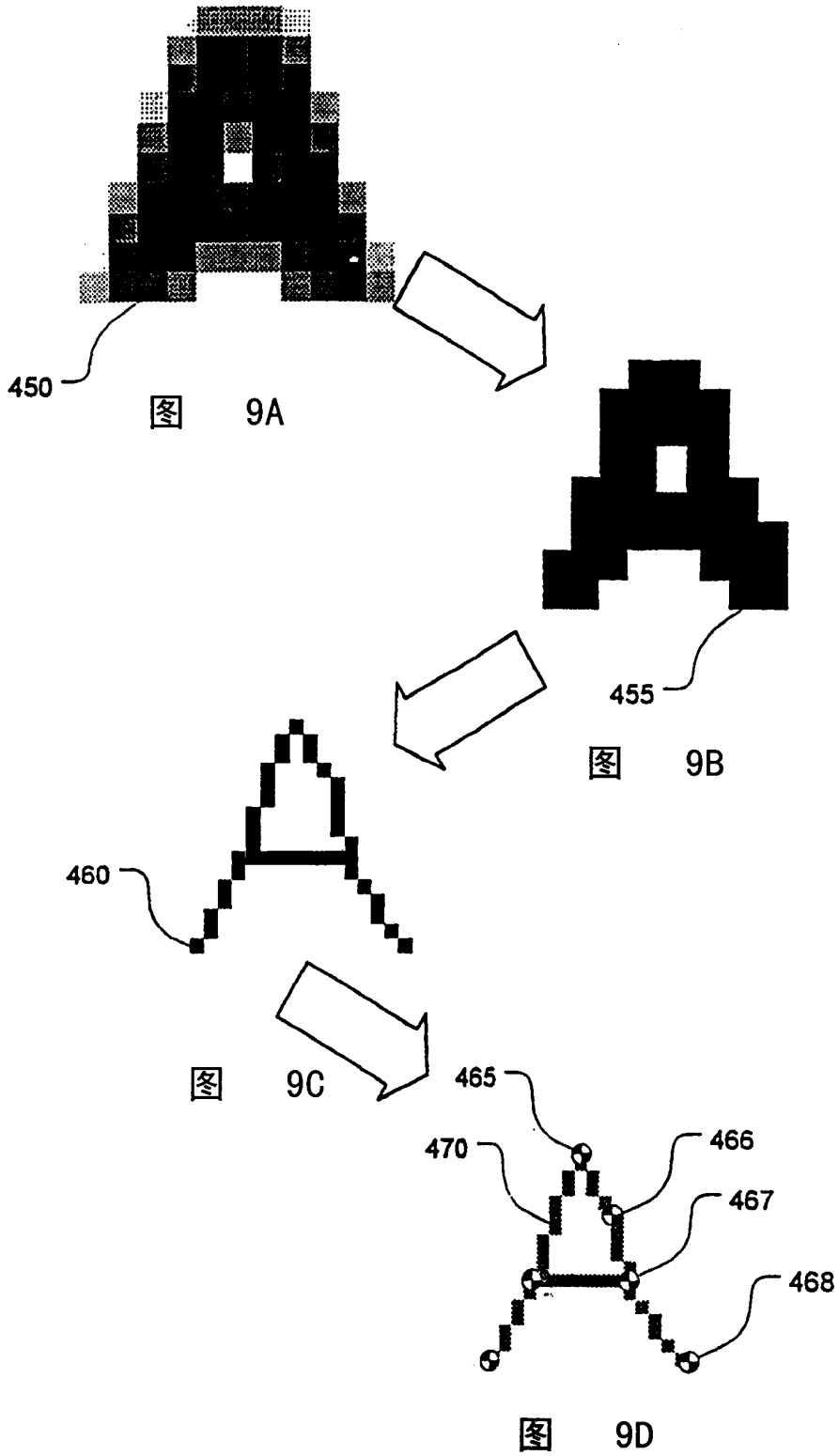


图 8



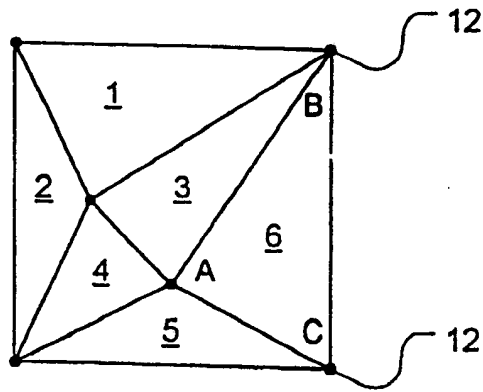


图 10A

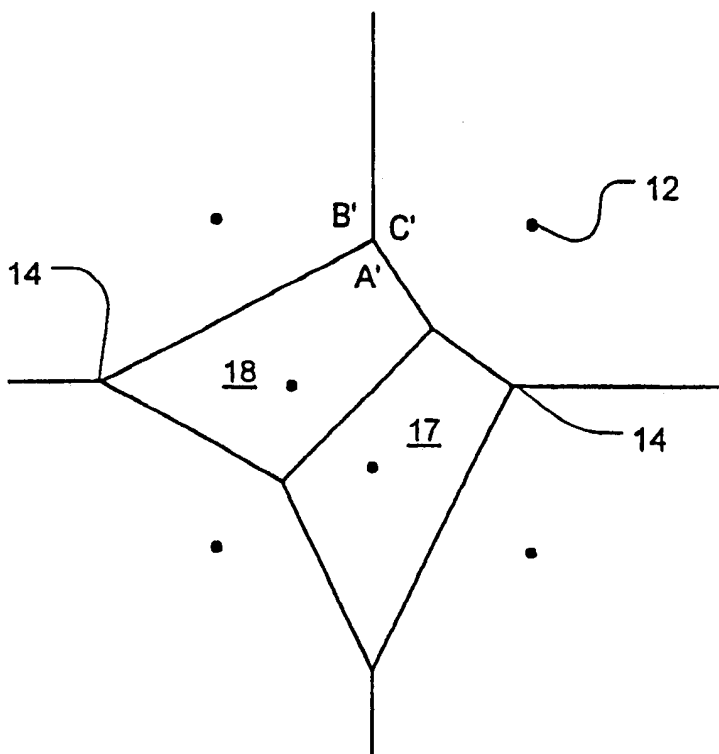


图 10B

4	1	0	2	● ● ●	1
0-5	5-10	10-15	15-20		175-180

图 11A

150-180	0	0	0	0	● ● ●	0
120-150	2	0	0	1	● ● ●	1
90-120	0	0	0	0	● ● ●	0
60-90	0	0	0	1	● ● ●	0
30-60	1	0	0	0	● ● ●	0
0-30	1	1	0	0	● ● ●	0
	0-5	5-10	10-15	15-20		175-180

图 11B