[19]中华人民共和国国家知识产权局

[51]Int. Cl[7]
G06F 17/30

[12]发明专利申请公开说明书

[21]申请号 99804574.8

[43]公开日 2001年5月16日

[11]公开号 CN 1295690A

[22]申请日 1999.12.24 [21]申请号 99804574.8
[30]优先权
[32]1999.1.28 [33]US [31]60/117,658
[32]1999.8.9 [33]US [31]09/370,931
[86]国际申请 PCT/EP99/10408 1999.12.24
[87]国际公布 WO00/45291 英 2000.8.3
[85]进入国家阶段日期 2000.9.27
[71]申请人 皇家菲利浦电子有限公司
地址 荷兰艾恩德霍芬
[72]发明人 L·阿格尼霍特里 N·迪米特罗瓦
J·H·埃伦巴尔斯

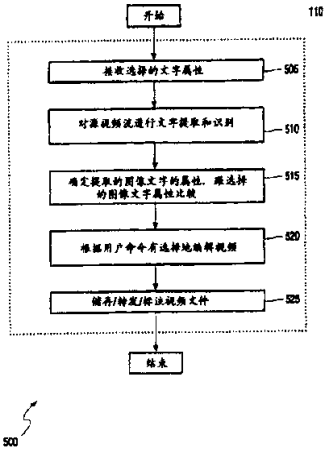
[74]专利代理机构 中国专利代理(香港)有限公司
代理人 邹光新 王忠忠

权利要求书3页 说明书16页 附图页数5页

[54]发明名称 用图像帧中检测到的文本信息分析视频内容的系统和方法

[57]摘要

公开了一种视频处理装置,用于视频文本分析系统,在视频流中 搜索一个或者多个用户选择的图像文本属性。这一视频处理装置包括 一个图像处理器,能够从图像帧中检测和提取图像文本,确定提取的 图像文本的属性,比较提取的图像文本属性和用户选择的图像文本属性,如果它们相同,就根据用户命令修改、传送和/或标注至少一部分 视频流。本发明采用用户选择的图像文本属性在视频剪辑文档中进行 搜索,以1)找出特定类型的事件的位置,比方说新闻节目或者体育 事件;2)找出特定人物或群体特写节目的位置;3)用名字来定位节目;4)储存或者去掉所有或者一些广告,或者根据出现在视频剪辑帧 里的图像文本,对视频剪辑的部分或者全部进行排序、编辑和储存操作。



ISSN1008-4274

权　利　要　求　书

1. 一种视频处理装置（110），用于能够分析图像帧中图像文本的系统（100），该装置能够在收到选择的至少一个图像文本属性的时候，对视频流进行搜索和过滤操作中的一项操作，该视频处理装置（110）包括:

一个图像处理器（120），能够接收包括多个图像帧（305、350）的第一个视频流，从这多个视频流（305、350）中检测和提取图像文本，确定所提取图像文本的至少一项属性，比较提取的至少一个图像文本属性和选择的至少一个图像文本属性，并在提取的至少一个图像文本属性和选择的至少一个图像文本属性相同的情况下，至少完成以下操作之一:

- 修改第一个视频流的至少一部分;
- 传送第一个视频流的至少一部分；和
- 标注第一个视频流的至少一部分。

2. 权利要求 1 的视频处理装置（110），其中提取的至少一个图像文本属性说明所述多个图像帧（305、350）中的所述图像文本是以下中的一个:

水平滚动;

垂直滚动；和

淡入淡出。

3. 权利要求 1 的视频处理装置（110），其中提取的至少一个图像文本属性说明所述多个图像帧（305、350）中的图像文本是以下文本中的一个:

一个人名；和

一个群体名。

4. 权利要求 1 的视频处理装置（110），其中提取的至少一个图像文本属性说明所述多个图像帧（305、350）中的所述图像文本是商业广告的一部分。

5. 权利要求 1 的视频处理装置（110），其中提取的至少一个图像文本属性说明所述多个图像帧（305、350）中所述图像文本是在以下情形之一中出现的:

节目开头；和

节目结尾。

6. 权利要求 1 的视频处理装置（110），其中提取的至少一个图像文本属性说明所述多个图像帧（305、350）中的所述图像文本是节目名的一部分。

7. 权利要求 1 的视频处理装置（110），其中提取的至少一个图像文本属性说明所述多个图像帧（305、350）中的所述图像文本是新闻节目的一部分。

8. 权利要求 1 的视频处理装置（110），其中提取的至少一个图像文本属性说明所述多个图像帧（305、350）中的所述图像文本是体育节目的一部分。

9. 一种图像文本分析系统（100），包括:

一个视频处理装置（110），能够在收到选择的至少一个图像文本属性的时候，完成搜索和过滤视频流操作中的一项操作，该视频处理装置（110）包括:

- 一个图像处理器（120），能够接收包括多个图像帧（305、350）的第一个视频流，从多个图像帧（305、350）中检测和提取图像文本，确定所提取的图像文本的至少一个属性，比较提取的至少一个图像文本属性和选择的至少一项图像文本属性，并在所提取的至少一个图像文本属性跟所述选择的至少一个图像文本属性相同的情况下，完成以下操作之一:

修改所述第一个视频流中的至少一部分;

传送所述第一个视频流中的至少一部分；和

标注所述第一个视频流的至少一部分;

一个显示监视器（185），用于显示第一个视频流中的至少一部分;和

一个用户输入装置（190）。

10. 收到所选至少一个图像文本属性的时候，进行搜索和过滤操作中一项操作的方法，用于能够分析图像帧中图像文本的系统，该方法包括以下步骤:

接收包括多个图像帧（305、350）的第一个视频流;

从这多个图像帧（305、350）中检测和提取图像文本;

确定提取的图像文本的至少一项属性;

比较提取的至少一个图像文本属性和选择的至少一个图像文本属性；和

根据在提取的至少一个图像文本属性跟选择的至少一个图像文本属性之间的匹配，进行以下操作中的至少一项操作：

- 修改第一个视频流中的至少一部分；
- 传送第一个视频流中的至少一部分；和
- 标注第一个视频流中的至少一部分。

11. 计算机能够执行的步骤，储存在计算机能够读出的存储媒介（140）里，用于能够分析图像帧（305、350）中图像文本的系统，在收到选择的至少一个图像文本属性的时候，对视频流完成搜索和过滤这两项操作中的一项操作，这些计算机可执行的步骤包括以下步骤：

接收包括多个图像帧（305、350）的第一个视频流；

从这多个图像帧（305、350）中检测和提取图像文本；

确定提取的图像文本的至少一项属性；

比较提取的至少一项图像文本属性和选择的至少一项图像文本属性；和

在提取的至少一项图像文本属性跟选择的至少一项图像文本属性相同的情况下，完成以下操作中的至少一项操作：

- 修改第一个视频流中的至少一部分；
- 传送第一个视频流中的至少一部分；和
- 标注第一个视频流中的至少一部分。

12. 权利要求 11 的计算机可读存储媒介（140），其中提取的至少一个图像文本属性说明，多个图像帧中的图像文本属性是水平滚动、垂直滚动、淡入淡出、缩放、波纹、飞行和闪烁中的一个。

说　明　书

用图像帧中检测到的文本信息分析视频内容的系统和方法

相关申请

本申请跟 1999 年 1 月 28 日提交，标题为“视频中文本信息检测和定位的方法和装置”的第 60/117658 号美国临时专利申请中公开的内容有关，该专利被共同转让给本发明的受让人。这里将这一相关临时专利申请的内容全部引入作为参考，就象它的内容就在本申请中一样。

技术领域

总的来说，本申请涉及视频处理系统，更具体地说，涉及一种系统，用于在检测到的视频内容中文本属性的基础之上，分析视频流，找出其特征。

发明背景

数字电视（DTV）的出现、因特网的普及以及象激光唱盘（CD）和数字化视频光盘（DVD）播放机这样的消费多媒体电子产品的引入，为消费者提供了大量多媒体信息。随着视频内容越来越容易获得以及访问这些视频内容的产品进入消费市场，对大量的多媒体数据进行搜索、编制索引和识别变得更加重要，更加富有挑战性。

许多出版物中都介绍了为视频信号编制索引和分类的系统和方法，包括：M. Abdel-Mottaleb 等的“CONIVAS：基于内容的图像和视频访问系统”，ACM 多媒体论文集，第 427～428 页，波士顿，1996 年；S-F Chang 等等的“视频 Q：基于内容利用视觉线索的自动视频搜索系统”，ACM 多媒体论文集，第 313～324 页，西雅图，1994 年；M. Christel 等等的“信息数字视频库”，ACM 评论，第 38 卷，第 4 期，第 57～58 页，1995 年；N. Dimitrova 等等的“消费装置中的视频内容管理”，IEEE 知识和数据工程学报，1998 年 11 月；U. Gargi 等等“在数字视频数据库中为文本事件编制索引”，模式识别国际会议，布里斯班，第 916～918 页，1998 年 8 月；M. K. Mandal 等等的“利用矩和小波的图像索引编制”，IEEE 消费电子学报，第 42 卷，第 3 期，1996 年 8 月；以及 S. Pfeiffer 等等的“数字运动的自动摘要提取”，视觉通信和图像表示杂志，第 7 卷，第 4 期，第 345～353

页，1996。

在视频流中检测广告也是一个非常活跃的研究领域。见 R. Lienhart 等等“关于电视广告的检测和识别”，IEEE 多媒体计算和系统国际会议论文集，第 509～516 页，1997；以及 T. McGee 等等“对电视节目进行分析以识别和剔除非情节片断”，SPIE 图像和视频数据库中的存储和读取会议，San Jose，1999 年 1 月。

文件图像中的文本识别在本领域中众所周知。文件扫描仪和有关的光学字符识别（OCR）软件俯拾即是，大家也十分了解。然而，图像帧中的文本检测和识别却是少见的难题，跟印刷文件相比，需要完全不同的方法。印刷文件中的文本常常仅限于均匀背景（普通纸）上的单色字符，通常只需要简单的阈值处理算法将文本跟背景分离。相反，按比例缩小的视频图像中的字符带有很多的噪声分量，包括无控制的照明状态。还有，背景会频繁地移动，文本字符会有不同的颜色、大小和字体。

Ohya 等等在 1994 年 2 月 IEEE 模式分析和机器智能学报第 16 卷第 214～224 页上的文章“在场景图像中的识别字符”，介绍了如何用本地阈值处理提取字符，以及通过在相邻区域之间评估灰度级差别来检测包含字符的图像区域。Ohya 等等还公开了合并具有相近和相似灰度级的检测到的区域，从而产生字符模式候选对象的方法。

A. Hauptmann 等等在 1995 年秋季 AAAI 语言和视觉集成计算模型学术讨论会上的文章“视频片断的文本、语音和视觉：信息媒体计划”中，介绍了如何利用视频文本的空间环境和高对比度特性来合并相互邻近，具有水平和垂直边缘的区域从而检测文本。R. Lienhart 和 F. Suber 在 1996 年 1 月的 SPIE 图像和视频处理会议上的文章“视频索引的自动文本识别”，讨论了在视频图像中减少颜色数量的一种非线性红、绿、蓝（RGB）颜色系统。随后的分裂和合并过程产生了具有相似颜色的均匀片断。Lienhart 和 Suber 采用各种试探方法来检测均匀区域中的字符，包括前景字符、单色或者硬字符、尺寸受限字符和跟周围区域相比具有高对比度的字符。

1998 年 11 月 12 日 IEEE 模式识别论文集第 31 卷第 2055～2076 页上 A. K. Jain 和 B. Yu 的文章“图像和图像帧的自动文本定位”介绍了如何利用多值图像分解对文本定位，并将图像分成多个真实前

景和背景图像。J-C. Shim 等等在 1998 年模式识别国际会议论文集第 618~620 页上的文章“基于内容的注释和检索的自动视频文本提取”，介绍了如何用广义的区域标注算法寻找均匀区域以及分段和提取文本。识别出来的前景图像被分成组，以确定文本的颜色和位置。

其它有用的字符分段算法在 K. V. Mardia 等等 1998 年 IEEE 模式分析和机器智能学报第 10 卷第 919~927 页上的文章“图像分段的空间阈值处理方法”，以及 A. Perez 等等在 1987 年 IEEE 模式分析和机器智能第 9 卷第 742~751 页上的文章“图像分段的迭代阈值处理方法”中有介绍。

然而，现有技术中的文本识别系统没有将视频内容中检测到的文本的非语义属性考虑在内。现有技术系统简单地识别图像文本的语义内容，并根据该语义内容为视频剪辑编制索引。图像文本的其它属性，比方说在帧内的物理位置、持续时间、运动和/或节目中的临时位置，都被忽略了。另外，还没有做出过任何努力利用视频内容来识别和编辑视频剪辑。

因此在这一领域中需要一种改进的视频处理系统，它使用户能够搜索整个视频剪辑文档，有选择地储存和/或编辑包含图像文本属性符合用户选择的图像文本属性的所有或部分视频剪辑。

发明简述

为了解决现有技术中的上述缺陷，本发明公开一种视频处理装置，用于在视频流中搜索或者过滤出用户选择的一个或者更多的图像文本属性。一般而言，在视频流中进行“搜索”指的是根据用户定义的输入进行搜索，其中“过滤”一般是指一个自动过程，需要很少的用户输入，或者不需要用户输入。然而，在这一说明中，“搜索”和“过滤”可以互换使用。图像处理器从视频剪辑中检测和提取视频文本，确定提取的视频文本的有关属性，并将提取出来的图像文本属性跟用户选择的图像文本属性进行比较。如果它们相同，这一视频处理装置就可以修改、传送、标注，否则根据用户命令识别至少一部分视频流。这一视频处理装置用用户选择的图像文本属性来搜索整个视频剪辑文档，以 1）找出特定类型事件的位置，比方说新闻节目或者体育事件；2）找出描写特定人物或群体的节目的位置；3）按照名字找出节目的位置；4）储存或者去掉所有或者一些广告，否则根据出现在视频剪辑

帧中的图像文本对所有或者部分视频剪辑进行分类、编辑和储存。

本发明的主要目的是提供一种视频处理装置，用于能够分析图像帧中图像文本的系统，能够根据收到的选中的至少一个图像文本属性，搜索和/或过滤视频流。在一个示例性的实施方案里，这一视频处理装置包括一个图像处理器，能够接收包括多个图像帧的第一个视频流，从多个图像帧中检测和提取图像文本，确定提取出来的图像文本的至少一个属性，将提取出来的至少一个图像文本属性跟选中的至少一个图像文本属性进行比较，并且，如果在提取出来的至少一个图像文本属性跟选中的至少一个图像文本属性相同的情况下，执行 1）根据第一个用户命令修改第一个视频流的至少一部分；2）根据第二个用户命令传送第一个视频流的至少一部分；和 3）根据第三个用户命令为第一个视频流的至少一部分做标记，这三项操作中的至少一项。

根据本发明的一个示例性实施方案，提取出来的这至少一个图像文本属性说明多个图像帧中的图像文本属性是：水平滚动；垂直滚动；淡入淡出、特技效果和动画效果中的一个。

根据本发明的一个实施方案，提取出来的这至少一个图像文本属性说明多个图像帧中的图像文本属性是：一个人的名字；一群人的名字中的一个。

根据本发明的另一个实施方案，提取出来的这至少一个图像文本属性说明多个图像帧中的图像文本是商业广告的一部分。

根据本发明的再一个实施方案，提取出来的这至少一个图像文本属性说明多个图像帧中的图像文本是出现在：节目开头；和节目结尾的文本。

根据本发明的又一个实施方案，提取出来的至少一个图像文本属性说明这多个图像帧中的图像文本是节目名的一部分。

根据本发明的一个实施方案，提取出来的这至少一个图像文本属性说明这多个图像帧的图像文本是新闻节目的一部分。

根据本发明的另一个实施方案，提取出来的这至少一个图像文本属性说明这多个图像帧的图像文本是体育节目的一部分。

前面已经大致地概括了本发明的特征和技术优点，从而使本领域里的技术人员能够更好地理解本发明的以下详细介绍。本发明的其它特征和优点将在下面介绍，它们构成本发明权利要求的主体。本领域

里的技术人员应当明白，他们能够很容易地利用这里公开的概念和具体实施方案，以此为基础改进或者设计其它结构，实现本发明的目的。本领域里的技术人员还应当明白，广义地说这种等价的结构没有偏离本发明的实质和范围。

在发明详述之前，首先为这一专利文献从头到为都要使用的某些词和术语下一个定义可能会更好：术语“包含（include）”和“包括（comprise）”，以及它们的派生词，指的是包括（inclusion）而没有限制；术语“或者”（“or”）的意思是包括（inclusive），意思是和/或；术语“跟……有关（associated with）”和“跟它有关（associated therewith）”及其派生词的意思是包括、被包括在内、相互连、包含、包含在内、连接到……或、跟……连接、耦合到……或与……相耦合、可以跟……连通、跟……协作、交叉、并列、接近、捆绑到……或与……相捆绑、具有，拥有、等等；术语“处理器”或者“控制器”指的是控制至少一项操作的任意装置、系统或者它们的一部分，这种装置可以用硬件、固件或者软件来实现，或者用它们中的至少两样的组合来实现。应当指出，跟任意特定控制器有关的功能可以是集中式的或者分布式的，而不管是在本地或者远处。另外，术语“视频剪辑”可以指一个视频片段、视频序列、视频内容等等。某些词和短语的定义在整个这一专利文献中给出。本领域里的普通技术人员应当明白，在许多情况下，如果不是多数情况下，这些定义适用于定义的这些词和短语现有的以及将来的用法。

附图简述

为了更全面地理解本发明及其优点，下面参考以下说明和附图，其中相似的数字表示相似的对象，这些附图中：

图 1 说明的是本发明一个实施方案中的一种示例性图像文本分析系统；

图 2 中的流程图说明的是按照本发明一个实施方案中图 1 所示示例性视频处理装置的文本提取和识别操作过程；

图 3A 和 3B 画出了包含具有图 1 所示视频处理装置能够识别的所选属性的图像文本的图像帧；

图 4 给出了一个示例性的图像文本属性表，包括本发明一个实施方案中系统定义的和用户定义的图像文本属性；和

图 5 中的流程图说明了本发明一个实施方案中图 1 所示视频处理装置的图像文本属性分析操作。

发明详述

下面讨论的图 1~5，以及这一专利文献中用于介绍本发明的原理的各种实施方案，都是用于进行说明的，无论如何都不应当理解为是要限制本发明的范围。本领域里的技术人员会明白，本发明的原理可以用任何合适的图像文本分析系统来实现。

图 1 画出了本发明一个实施方案的示例性图像文本分析系统 100。图像文本分析系统 100 包括视频处理装置 110、视频源 180、监视器 185 和用户装置 190。视频处理装置 110 提供装置，供分析接收到的视频图像使用。这包括完成本发明的过程，通过这些过程提取出视频文本，根据系统或者用户定义的文本属性进行分析和分类。

视频源 180 提供视频剪辑文档供视频处理装置 110 搜索。视频源 180 可以是天线、磁带录像机（VTR）、数字化视频光盘（DVD）播放机/录像机、视盘播放机/录像机或者能够储存和传送有或者没有音频的数字视频图像的类似装置。视频源 180 能够提供一些短剪辑或者多个剪辑，包括更长的数字化视频图像。视频源 180 可以包括任何已知格式的模拟或数字视频数据，比方说 MPEG-2、MJPEG 等等。

监视器 185 提供显示视频图像的装置，还可能配备了音频装置，如果需要的话。用户装置 190 表示一种或者多种外围设备，可以被图像文本分析系统 100 的用户操作，将用户输入提供给这一系统。典型的外围用户输入设备包括计算机鼠标、键盘、光笔、游戏操纵杆、触摸表（a touch-table）和有关的摄像头，和/或能够选择用来输入、选择和/或操作数据，包括所有或者部分显示的视频图像，的任何其它装置。用户装置 190 能够选择所需要的视频文本识别属性，输入给视频处理装置 110。用户装置 190 可能还包括输出装置，比方说彩色打印机，产生某一图像、帧或者剪辑的硬拷贝。

视频处理装置 110 包括图像处理器 120、RAM 130、存储器 140、用户 I/O 卡 150、视频卡 160、I/O 缓冲器 170 和处理器总线 175。处理器总线 175 在视频处理装置 110 的各单元之间传送数据。RAM 130 还包括图像文本工作空间 132 和文本分析控制器 134。

图像处理器 120 为视频处理装置 110 提供总的控制，并进行图像

文本分析系统 100 所需要的图像处理，包括在系统选择的和用户选择的属性的基础之上，分析图像帧内的文本。这还包括完成编辑过程、处理数字视频图像，显示在监视器 185 上，和/或储存在存储器 140 中，并在图像文本分析系统 100 的各种单元之间传送数据。图像处理器 120 的需要和能力在本领域里众所周知，不需要更详细地介绍，本发明所需要的除外。

RAM 130 为视频处理装置 110 产生的临时数据提供随机存取存储器，否则就不由该系统内的部件提供。RAM 130 包括图像文本工作空间 132 和文本分析控制器 134 存储器，以及图像处理器 120 和有关装置需要的其它存储器。图像文本工作空间 132 表示那一部分 RAM 130，在文本分析过程中，跟某一图像帧有关的视频图像临时储存在那里。图像文本工作空间 132 为在一个视频剪辑内修改帧而不影响原始数据提供装置，这样在晚些时候能够恢复原始数据。

在本发明的一个实施方案里，文本分析控制器 134 表示那部分 RAM 130，它专用于储存图像处理器 120 执行的应用程序，这一应用程序根据系统或者用户定义的文本属性分析视频图像。文本分析控制器 134 可以采用著名的编辑技术，比方说场景之间的变形或者边界检测，以及跟本发明有关的视频文本分析新技术。编辑控制器 134 还可以作为 CD-ROM、计算机磁盘或者能够载入存储器 140 或者视频源 180 中的那种可拆卸盘端口的其它存储媒介上的一个程序。

存储器 140 包括一个或者多个盘系统，包括可拆卸盘（磁盘或者光盘），用于永久性地储存节目和其它数据，包括所需要的视频和音频数据。根据系统需要，存储器 140 跟一个或者多个双向总线连接，传送视频和音频数据给视频源 180 和系统的其余部分以及从视频源 180 和系统的其余部分接收视频和音频数据。需要的话，存储器 140 能够以视频速率传送数据。存储器 140 的容量足以储存几分钟的视频数据供编辑，包括文本属性分析。根据应用的具体情况和图像处理器 120 的能力，存储器 140 可以配置成能够储存大量的视频剪辑。

用户 I/O 卡 150 为用户装置 190 和图像文本分析系统 100 的其它部分提供接口装置。用户 I/O 卡 150 将从用户装置 190 收到的数据转换成接口总线 175 的格式，传送给图像处理器 120 或者 RAM 130，供图像处理器 120 随后访问。用户 I/O 卡 150 还传送数据给打印机这样

的用户输出设备。视频卡 160 通过数据总线 175 在监视器 185 和视频处理装置 110 之间提供接口。

I/O 缓冲器 170 通过总线 175 在视频源 180 和图像文本分析系统 100 之间提供接口。如上所述，视频源 180 至少有一条双向总线，用于连接 I/O 缓冲器 170。I/O 缓冲器 170 以需要的视频图像传输速率在它跟视频源 180 之间传送数据。在视频处理装置 110 内，I/O 缓冲器 170 根据需要将从视频源 180 收到的数据传送给存储器 140、图像处理器 120 或者 RAM 130。同时传送视频数据给图像处理器 120 提供了按照收到的方式显示视频图像的一种手段。

图 2 描述了一个流程图 200，它说明根据本发明的一个实施方案，视频处理装置 110 随后进行的文本提取和识别操作。文本提取是针对一个一个图像帧进行的，将 M×N 帧的原点（0，0）作为左上角。帧内的像素用（x，y）坐标表示，其中 x 表示像素的列（0~N），y 表示是第几行（0~M）的像素。

通道分离（步骤 205）

一开始，图像处理器 120 分离视频图像一帧或者多帧的颜色，并储存减少了颜色的图像供文本提取时使用。在本发明的一个实施方案里，图像处理器 120 用红-绿-蓝（RGB）颜色空间模型来隔离图像的红色分量。红色分量在检测白色、黄色和黑色时更加有用，这些颜色是视频文本采用的主要颜色。隔离出来的红色帧提供了为这些频繁使用的文本颜色提供了尖锐的高对比度边缘。隔离出来的红色帧图像储存在图像文本工作空间 132 里。在本发明的其它实施方案里，图像处理器 120 可以使用其它的颜色空间模型，比方说灰度级图像或者 YIQ 图像帧的 Y 分量。

图像增强（步骤 210）：

进行进一步的处理之前，捕获的红色帧用下面的 3×3 掩码增强：

```
-1  -1  -1
-1   8  -1
-1  -1  -1
```

另外，用一个中值滤波器去掉黑白点相间的噪声（随机噪声），比方说用 R. C. Gonzalez 和 R. E. Woods 在“数字图像处理”中介绍的那种，该书于 1992 年由 Addison-Wesley 出版公司出版。

边缘检测（步骤 215）：

用以下掩码检测增强红色图像中的文本字符边缘:

-1 -1 -1

-1 12 -1

-1 -1 -1

其中矩阵中的数字是边缘算子的权。

如果 EDGE 表示 M×N 边缘图像，那就可以用以下等式进行边缘检测:

$$EDGE_{m,n} = \sum_{i=1}^{1}\sum_{j=1}^{1} w_{i,j} F_{m+i,n+j} < \text{边缘阈值}$$

其中 0<m<M，0<n<N。$w_{i,j}$ 值是边缘掩码的权，$F_{x+i,y+j}$ 表示图像“F”的一个像素。在边缘检测过程中，帧的顶部和底部行以及左边和右边列的像素（也就是最外层的像素）被忽略。

边缘阈值是一个预先确定的阈值，可以是固定的，也可以是变化的。采用固定的阈值会出现大量黑白点相间的噪声，随后需要删除这些噪声点。还有，用固定阈值会导致文本周围的固定边缘断断续续，出现分裂了的字符。使用已知的开孔方法（例如先侵蚀然后再膨胀）会导致沿着黑白点相间的噪声的文本部分丢失。因此，采用自适应阈值是对采用静态阈值的一种改进。

对于一个像素，如果将部分或者所有相邻像素都标为边缘，就为当前像素降低这一阈值以便将它标为边缘。当前像素的阈值能否降低取决于标为边缘的相邻像素的个数。相邻像素是边缘这一事实增加了当前像素是边缘的概率。采用更低的边缘阈值来计算相邻像素的降低了的阈值。这一点保证了当这些像素不是边缘时它们不被标为边缘。这一个过程可以反过来，如果它被边缘像素包围，那么它就是一个边缘像素。

边缘过滤（步骤 220）：

一旦检测到字符边缘，图像处理器 120 就进行初步的边缘过滤，以去掉可能不包含文本或者其中的文本无法可靠地检测的图像区域。图像处理器 120 可以在不同的级别上进行边缘过滤。例如，边缘过滤可以在帧一级或者子帧一级进行。

在帧一级，如果帧中看起来包括边缘的部分超出合理的比例，图

像处理器 120 就忽略或者滤掉这一帧，这种情况的出现可能是因为帧中有高密度的对象。一旦一帧被过滤掉，文本分析就进入到输入的下一帧。在帧一级进行过滤时，图像处理器 120 维持一个边缘计数器，记录这一图像帧中边缘点的个数。但这样做会导致图像某些清洁区的文本被丢失，还可能导致假否定。

为了解决这些问题，图像处理器 120 可以在子帧一级进行边缘过滤。在“过分拥挤”的帧内找到文本，图像处理器 120 将帧分成更小的区域也就是子帧。在本发明一个示例性实施方案里，图像处理器 120 将和帧分成三列像素和三行像素，得到 6 个子帧。

图像处理器 120 指定一个子帧计数器，用于对图像每一个子部分进行边缘计数。在这一示例性实施方案里，图像的三个垂直（列）子帧用三个计数器。每一个垂直子帧都覆盖帧的三分之一区域。同理，图像的三个水平（行）子帧用三个计数器。每一个水平子帧同样覆盖帧区域的三分之一。

然后，图像处理器 120 检查每一个子帧区域，以确定子帧中的边缘像素个数，用它的计数器反映这一数字。可以用更多的子帧来产生更小的子帧区域，以便获得更多的清洁区域，在比三分之一个图像更小的区域里包含文本。

字符检测（步骤 225）：

下一步，图像处理器对前一步骤产生的边缘进行连通分量（CC）分析。假定每一个文本字符都有一个连通的分量或者它的一部分。图像处理器 120 将相隔某一距离的边缘像素点合并起来（比方说 8 像素近邻），成为单独一个连通分量结构。这一连通分量结构包含了互相连接在一起的像素的位置。这一结构还包含了最左边、最右边、顶部和底部的像素，以及这一结构的中心点的值（用 x-和 y-轴坐标来描述）。

连通分量结构还包含构成连通分量的多个像素点的个数。像素点个数表示这一连通分量区域的面积。预先确定的系统和/或用户阈值规定了连通分量区域的面积、高度和宽度的最大和最小值，以便判断哪些连通分量应当通过下一步处理。超出阈值标准的连通分量被过滤掉。

文本框检测（步骤 230）：

图像处理器 120 根据左下方像素的位置将前一步骤中通过了判别

式的连通分量按上升顺序排序。图像处理器根据（x，y）坐标位置排序，它表示像素的绝对位置，用 y 乘以列大小再加上 x 表示。排序以后的这一列连通分量被遍历，然后将连通分量合并起来一起形成文本框。

图像处理器 120 将第一个连通分量，连通分量（1），叫做第一个框，并作为初始或者当前框供分析使用。图像处理器 120 测试每一个随后的连通分量（i），看它最底部的像素距离当前文本框最底部的像素是否在预定可接受像素行阈值以内。如果连通分量（i）距离当前框在几行以内（例如 2 行），那就很可能当前文本框和连通分量（i）属于文本的同一行。行差阈值可以是固定的或者变化的，视需要而定。例如，阈值可以是当前文本框高度的一部分。

为了防止将图像中相隔太远的连通分量合并到一起，进行第二次测试，看连通分量（i）跟文本框的列距离是不是小于一个列阈值。这一可变阈值是连通分量（i）宽度的倍数。如果文本框和连通分量（i）相隔很近，图像处理器 120 就将连通分量（i）跟当前文本框合并。如果连通分量（i）不满足跟当前文本框合并的判据，就从连通分量（i）开始一个新的文本框，作为它的第一个分量，并继续遍历。这一过程会导致图像中一行文本出现多个文本框。

图像处理器 120 对初始字符合并过程产生的所有文本框进行第二级合并。这一次会将由于连通分量合并判据过于严格或者由于边缘检测不良，使同一个字符出现多个连通分量，从而被错误地理解为不同行文本的文本框合并起来。

图像处理器 120 按照一组条件将每一个框跟它后面的文本框进行比较。两个文本框的多个测试条件是:

a）一个框跟另一个框底部的距离在行差阈值以内。还有，这两个框水平方向的距离小于基于第一个框中平均字符宽度的可变阈值。

b）这两个框中每一个框的中心都位于另一个文本框的区域以内，或者

c）这些文本框相互重叠。

如果满足上述条件中的任意条件，图像处理器 120 就从文本框清单中删去第二个框，并将它合并到第一个框中去。图像处理器重复这一过程，直到所有文本框都两两一起测试过，并且尽可能地合并到一

起。

文本行检测和增强（步骤 235）：

图像处理器 120 接收在步骤 235 中获得的文本框作为文本行，只要它们符合面积、宽度和高度要求。对于每一个文本框，图像处理器 120 都要提取对应于原始图像文本框的子图像。然后，图像处理器 120 修改提取出来的子图像，从而使文本作为白色背景上的黑色前景出现。这一修改是必要的，这样这一步骤的输出就能由光学字符识别（OCR）应用程序处理。

图像处理器 120 通过计算文本框中像素的平均灰度级值（AvgFG）来修改文本框图像。同时计算文本框周围区域（例如 5 个像素）的平均灰度级值（AvgBG）。在这一文本框内，图像处理器 120 将超过 AvgFG 的所有像素标为白色，将低于 AvgFG 的所有像素标为黑色。同时计算标为白色的像素的平均灰度级 Avg1 和标为黑色的像素的平均灰度级 Avg2。

一旦文本框被转换成黑白（二进制）图像，图像处理器 120 就将 Avg1 和 Avg2 跟 AvgBG 进行比较。平均值更接近 AvgBG 的区域被当作背景，其它区域被当作前景（或者文本）。例如，如果黑色区域的平均值更接近 AvgBG，这一黑色区域就转换成白色，反之亦然。这样做保证了文本总是黑色，跟 OCR 程序的输入相容。

然后，图像处理器 120 将提取出来的帧文本存入图像文本工作空间 132，这一过程针对下一帧从步骤 205 继续下去。这一序列继续下去，直到将文本从选定视频剪辑的所有帧提取出来。一旦视频图像文本已经被识别和提取出来，图像处理器 120 就会根据系统选定的或者用户选定的文本属性，进一步隔离和分析提取出来的文本。可以根据一帧或者一组帧内的文本，对视频序列或者剪辑进行搜索或编制索引。例如，在一帧中出现“旅馆”这个词意味着很可能会在当前帧或者以后的帧中出现旅馆的画面。存在街道名称和门牌号码说明有可能出现城市场景。在搜索有公路的帧的时候，广告牌文本的识别会很有帮助。

还可能利用文本来分析和标注具体的图像帧。例如，图像处理器 120 可以识别和提取作为某一新闻节目的一部分广播的，或者某一新闻记者播发的实况新报道。图像处理器 120 可以通过搜索图像文本中的关键字来做到这一点，包括“实况”、节目名（例如“晚间新闻 4

（NEWS4 AT NITE）”）或者记者的名字（例如“卡尔·琼斯”）。

通过识别文本框，图像处理器 120 能够根据节目名、主持人名字、制片人名字、导演名字或者任何其它名字，为图像帧和剪辑编制索引。这包括根据音乐组的名字（例如 MTV 和 VH-1 视频剪辑）、现场访问节目主持人或者嘉宾的名字等等，对图像进行识别和分类。更进一步，体育节目可以通过提取频繁出现在屏幕上的个人成绩表而检索到。例如，图像处理器 120 可以搜索得分、队名、运动员名字等等。股票市场信息可以被提取出来并编制索引。天气信息也可以根据屏幕上的文本信息提取出来并编制索引。

比赛过程中滚动的“纸带”、现场访问、新闻等等可以通过检查对应于选定属性的图像文本而识别出来。屏幕纸带发布最新天气状态、股票市场行情信息，以及其它信息，它们能够通过属性分析而识别和标注，并提取出来供将来使用。提取的纸带信息可以看成独立于广播节目本身的信息，因为它跟它出现的节目视频没有关系。

需要识别视频片断的关键帧时，分析跟帧的其余内容有关的叠印字幕会很有用。这种叠印信息是视频、音频和文本（transcript）（闭路字幕或者电视文本广播））信息的补充。这一文本可以跟镜头检测算法一起使用，用于编辑视频索引，产生其中有文本的关键帧。有文本的关键帧更有意义，比没有文本的帧更能代表视频序列。例如，有足球比赛分数的关键帧比没有分数的帧更有用。此外，检测广告中文本的位置和大小可以跟其它属性一起使用，从而可靠地检测广告。

此外，视频文本分析为分析其余图像视频内容并对它们进行分类提供了一种新的手段。场景文本和它的属性提供了视频内容的线索。这有几个优点，包括分析文本时需要的计算机存储器容量和时间要少得多，而不是花很多时间和存储器容量去分析视频的视觉内容这一事实。例如，检测滚动文本可以识别节目的开头和结尾。于是，需要在有一组多个节目的视频内找到节目边界的时候，这一属性可能非常有用。

图 3A 和 3B 给出了示例性的图像帧 305 和 350，它们包含有图 1 中的视频处理装置 110 能够识别的选定属性的图像文本。在所有视频剪辑中，图像帧 305 和 350 中的图像文本不一定要同时出现在屏幕上。给出了图像帧 305 和 350 以说明图像文本的所选属性，比方说滚动、

位置、淡入淡出、短暂出现和关键字。在介绍本发明工作过程的时候为了简洁和清楚起见，不同类型节目的图像文本被组合成图像帧 305 和 350。

图像帧 305 表示从一个电视节目图像帧中提取出来的文本。在这种情况下，系统/用户已经选择了区分水平滚动文本和垂直滚动文本的属性，比方说跟节目片头字幕或者帧底部的信息纸带行有关的文本。通过识别在一系列帧中相同的文本来检测滚动属性，除非文本的位置从一帧到另一帧会不断地缓慢偏移。此外，即使对于不滚动的节目片头字幕，图像处理器 120 仍然能够通过识别只在屏幕上短暂出现的一系列文本消息，以及选择进一步识别文本中的关键字，比方说“制片人”、“导演”、“主演”、“演员表”等等，来识别节目片头字幕。

利用选择的垂直滚动属性，图像处理器 120 已经识别出了片头字幕文本行 310，它是虚线矩形框说明的一个向上滚动的文本区。利用选择的水平滚动属性，图像处理器 120 已经识别出了滚动着预告文本消息（315 的帧底部，它是如图所示矩形框内的一则天气消息，其中的文本朝观众的左侧滚动。

图像帧 350 包含图像文本的其它实例，这些图像文本有很容易识别的特定属性。例如，图像帧 350 左上角的个人成绩表文本 355 有三行文本。第一行说明是哪一个台或者哪一个网络，其余两行显示比赛得分。图像处理器 120 通过识别屏幕中具有类似于个人成绩表文本 355 属性的体育得分，可以识别体育节目。多数分数通常都是在屏幕的一个角落上显示的，数值数据（也就是每一个队的得分总和）跟个人成绩表在垂直方向对齐。

类似地，广告文本 360 有跟广告商有关的电话号码的关键字属性(例如“1-800-”)，广告文本 365 有跟广告商有关的因特网地址的关键字属性（例如www.[公司名].com）。此外，广告文本 360 和广告文本 365 都有另外一个文本属性，也就是说它们都位于视频图像 350 中心附近，该文本属性可以用于识别商业广告。多数其它类型的文本都位于屏幕的底部或者角落里。最后，文本区域 370 有一个关键字属性（也就是“新闻”），它说明了这一帧是新闻节目的一部分。文本区域 375 有另一个关键字属性（也就是“实况”），它说明显示的文本帧是新闻节目的一部分。

图 4 给出了存储器 140 中示例性的图像文本属性表 400，它包括本发明一个实施方案中系统定义的和用户定义的图像文本属性。表 400 中每一个系统/用户定义的属性分类都对应于一个属性文件，它们可以是固定的或者变化的，就象图 1 所示图像文本分析系统的特定实施方案所确定的一样。

广告属性 405 代表商业广告文本的特性，这些文本可以从一个文件里取出来供查阅。跟广告内容有关的属性可以包括特定尺寸或者位置范围以内的文本、短暂出现的文本、显示的电话号码、邮寄地址、因特网地址和广告内象“大减价销售”、“厂家折扣”之类的关键字。

节目名属性 410 为系统/用户提供了隔离视频剪辑的手段，这些视频剪辑中出现的文本说明了它属于哪一类节目。节目名属性 410 又一次包括大小和位置这样的属性，以及实际的节目名，比方说“Seinfeld”。节目名属性 410 可以说明图像处理器 120 只应当在视频剪辑中已经识别过的片断（比方说开头）中寻找节目名，以便删除节目中在不同节目中出现的广告。

节目类型属性 415 包括说明某类节目（体育、新闻、音乐视频等等）的文本属性。这些类型的节目可以按照前面介绍的方式，通过搜索个人成绩表属性来识别，这些属性包括体育代表团关键字（例如 NBA、NHL）、新闻节目关键字（例如“新闻”、“天气”、“实况”）或者音乐视频关键字（例如“制片人”、“录制人”）。

人名属性 420 包括说明某一个人（“约翰·史密斯”）的文本，可以跟其它文本属性（比方说新闻节目名、体育组织名称等等）一起使用。公司名属性 425 提供了检查视频剪辑中是否存在某一公司名的一种手段。例如，图像处理器 120 可以在包围棒球场的广告牌上找到某一公司的名字。属性选择可以包括以前介绍过的文本特征，用来搜索公司名的节目类型，具体公司名的显示特性，某一新闻展览中产品上显示的公司名等等。

事件属性 430 指的是某类事件的文本属性，比方说保龄球（Super Bowl）或者白宫简报。在这一点上，事件属性 430 跟节目类型属性或者人名属性非常相似。

文本效果属性 435 提供一组标准的文本特性，可以用于选择和显示。文本效果属性 435 可以包括水平和垂直滚动、缩放（也就是缩小

或者放大）、闪烁、波浪形（或者波纹）、剥离、扰乱、飞行、动画和实况文本这样的文本效果。

网络徽标属性 440 指的是跟网络标识徽标有关的文本属性。这些属性包括网络名称和徽标，供比较文本和最可能出现徽标的主帧区域时使用。网络常常将它们的徽标淡轮廓线（或者水印）跟节目的屏幕图像叠印在一起。

文本外观属性 445 指的是图像文本的一个或者多个特定特征，比方说文本颜色、字体类型、文本高度、文本宽度或者文本位置。对于文本高度、文本宽度或者文本位置这种情形，尺寸或者位置可以采用绝对量（例如具体数量的像素或者具体范围的像素）或者用相对量（例如屏幕尺寸的具体百分比或者百分比范围）给出。

图 5 给出了流程图 500，它说明的是本发明一个实施方案中示例性视频处理装置 110 的图像文本属性分析操作。一组标准文本属性可以由文本分析控制器 134 在系统初始化的时候存入存储器 140 和/或修改或者通过用户装置 190 输入。这样，在默认方式下，图像处理器 120 就可以从存储器 140，或者通过用户装置 190 的具体输入，接收选择的文本属性（步骤 505）。

启动了视频文本分析以后，图像处理器 120 检测、提取和储存选择的图像帧中的文本，就象参考图 2 更详细地介绍过的那样（步骤 510）。提取出来的文本的文本属性被确定，并存入图像工作空间 132。然后，需要的时候，将提取的图像文本跟选择的属性进行比较，结果存入图像文本工作空间 132 和/或存储器 140（步骤 515）。

根据具体应用的情况，跟选择的属性相同的视频图像文本可以响应用户命令通过已知的编辑过程进行修改（步骤 520）。这一编辑可以包括，例如，清除所有广告，或者，删除节目只保留广告。然后，得到的视频文件和有关的分析过的文本可以做上标记，供检索用，存入存储器 140，和/或转给内部或者外部存储器，供以后使用（步骤 525）。

虽然详细地介绍了本发明，但是，本领域里的技术人员应当明白，他们能够进行各种修改、替换和更改，而不会偏离本发明广义形式的实质和范围。

说 明 书 附 图

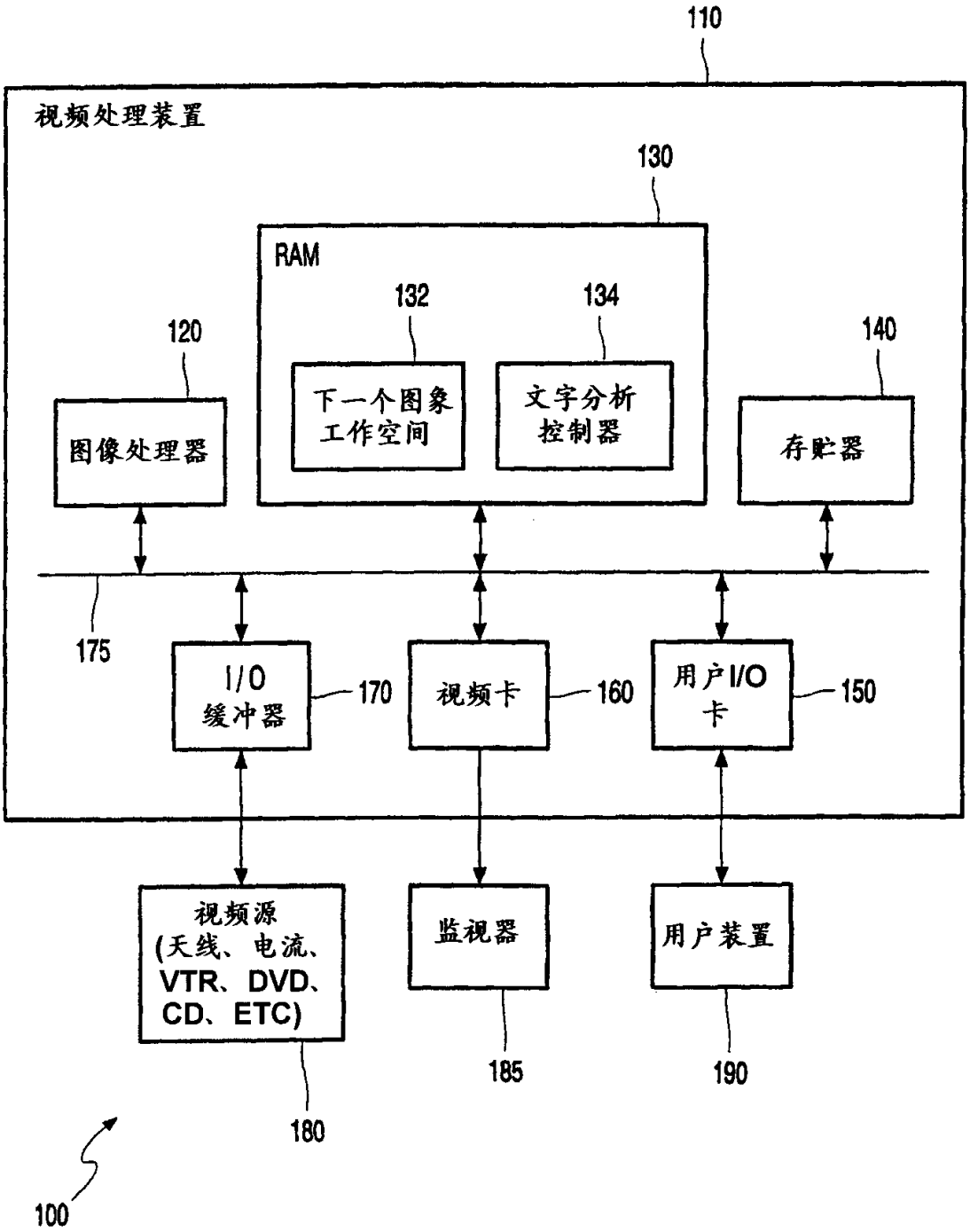

110
视频处理装置
130
RAM
132
134
120
140
下一个图象
工作空间
文字分析
控制器
图像处理器
存贮器
175
I/O
缓冲器
170
视频卡
160
用户I/O
卡
150
视频源
(天线、电流、
VTR、DVD、
CD、ETC)
监视器
用户装置
185
190
180
100

图 1

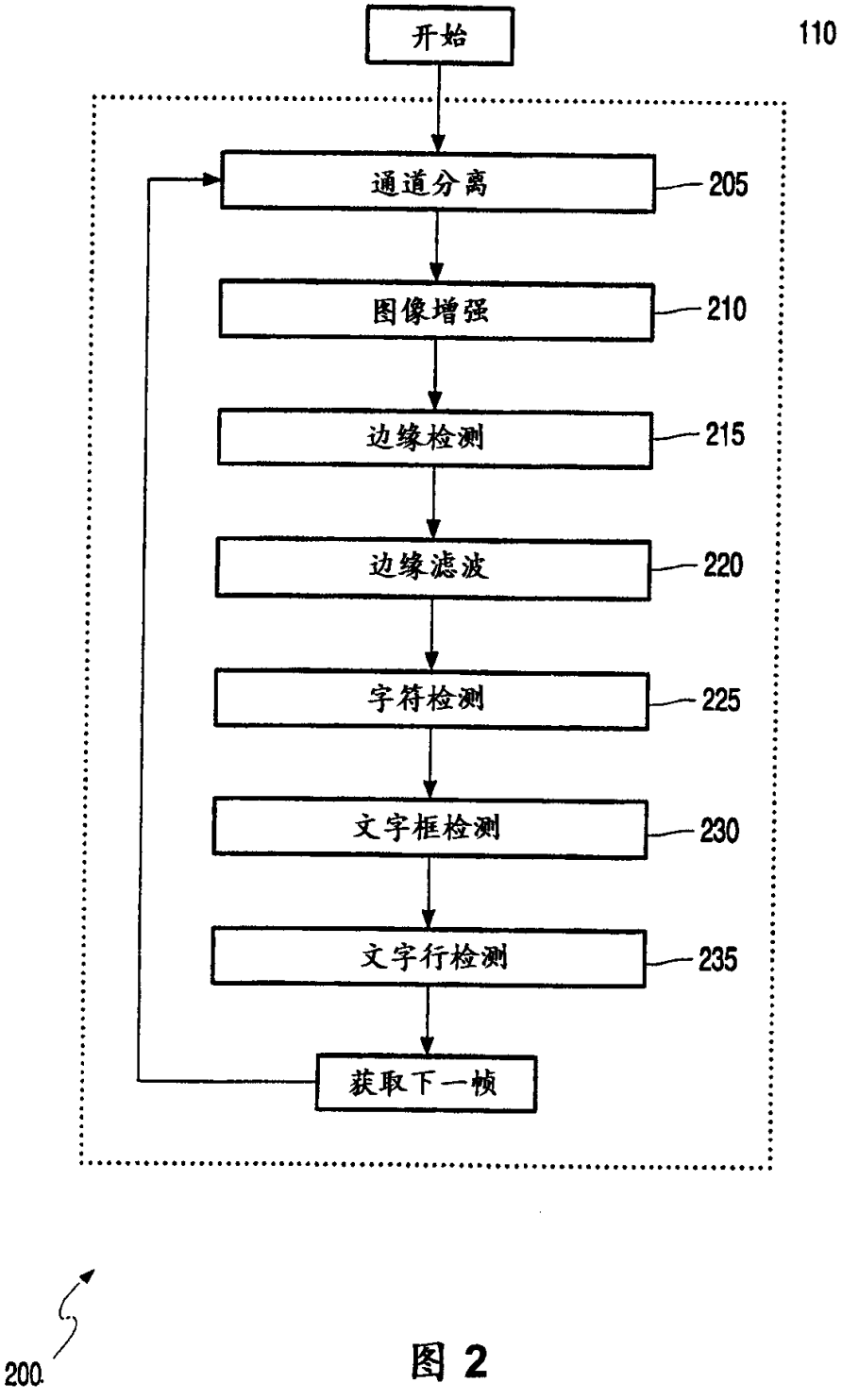

开始
110
通道分离
205
图像增强
210
边缘检测
215
边缘滤波
220
字符检测
225
文字框检测
230
文字行检测
235
获取下一帧
200

图 2

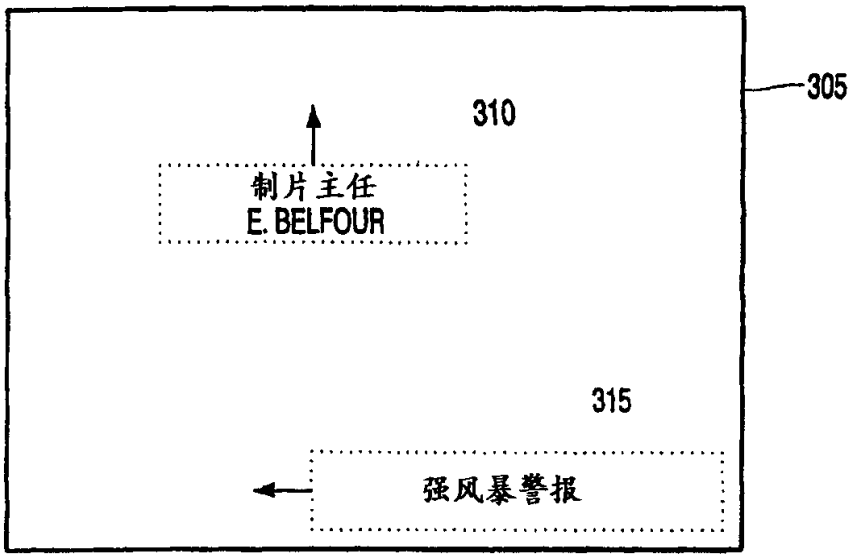

305
310
制片主任
E. BELFOUR
315
强风暴警报

图 3A

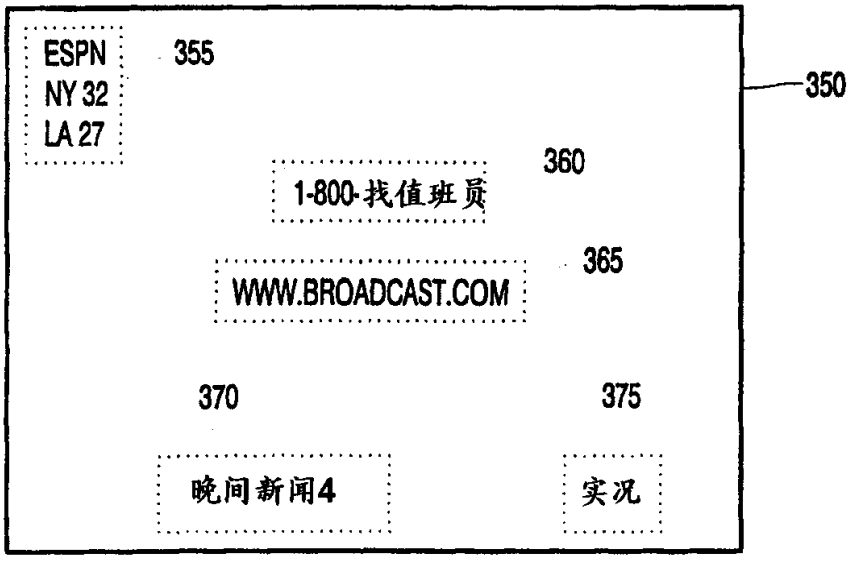

ESPN
NY 32
LA 27
355
350
360
1-800-找值班员
365
WWW.BROADCAST.COM
370
375
晚间新闻4
实况

图 3B

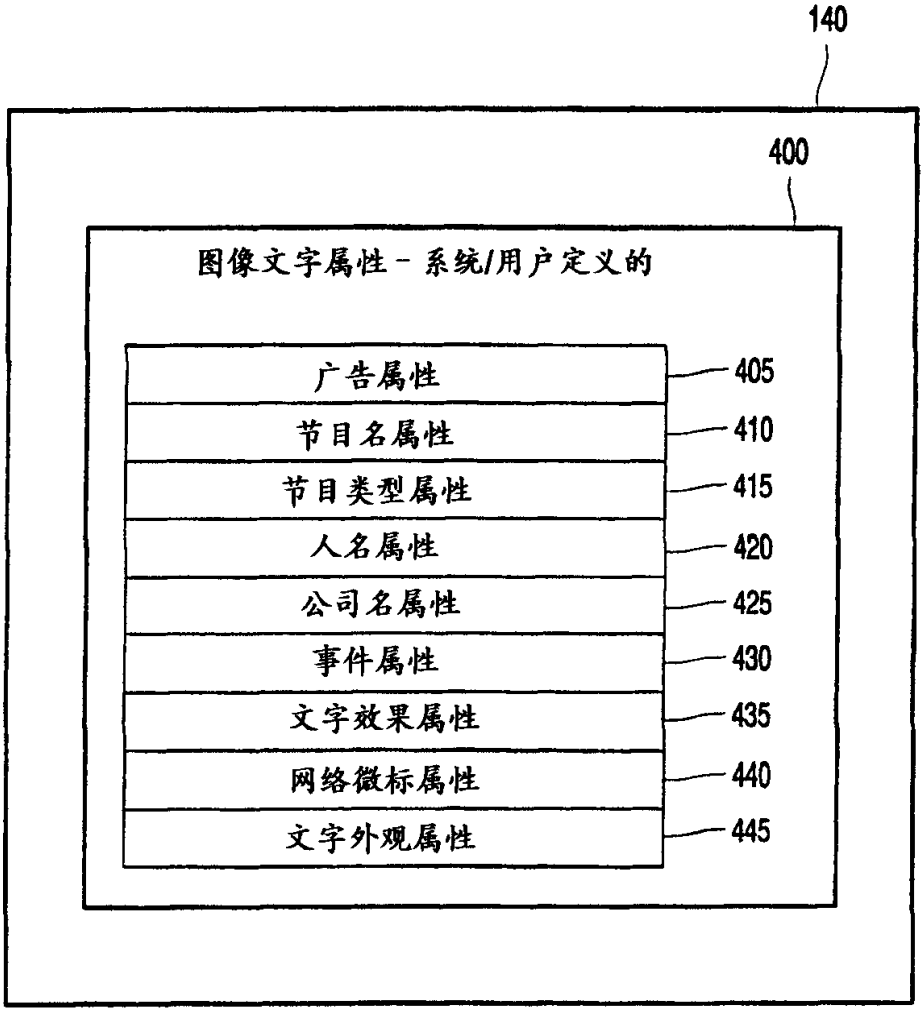

140
400
图像文字属性 - 系统/用户定义的
广告属性
405
节目名属性
410
节目类型属性
415
人名属性
420
公司名属性
425
事件属性
430
文字效果属性
435
网络微标属性
440
文字外观属性
445

图 4

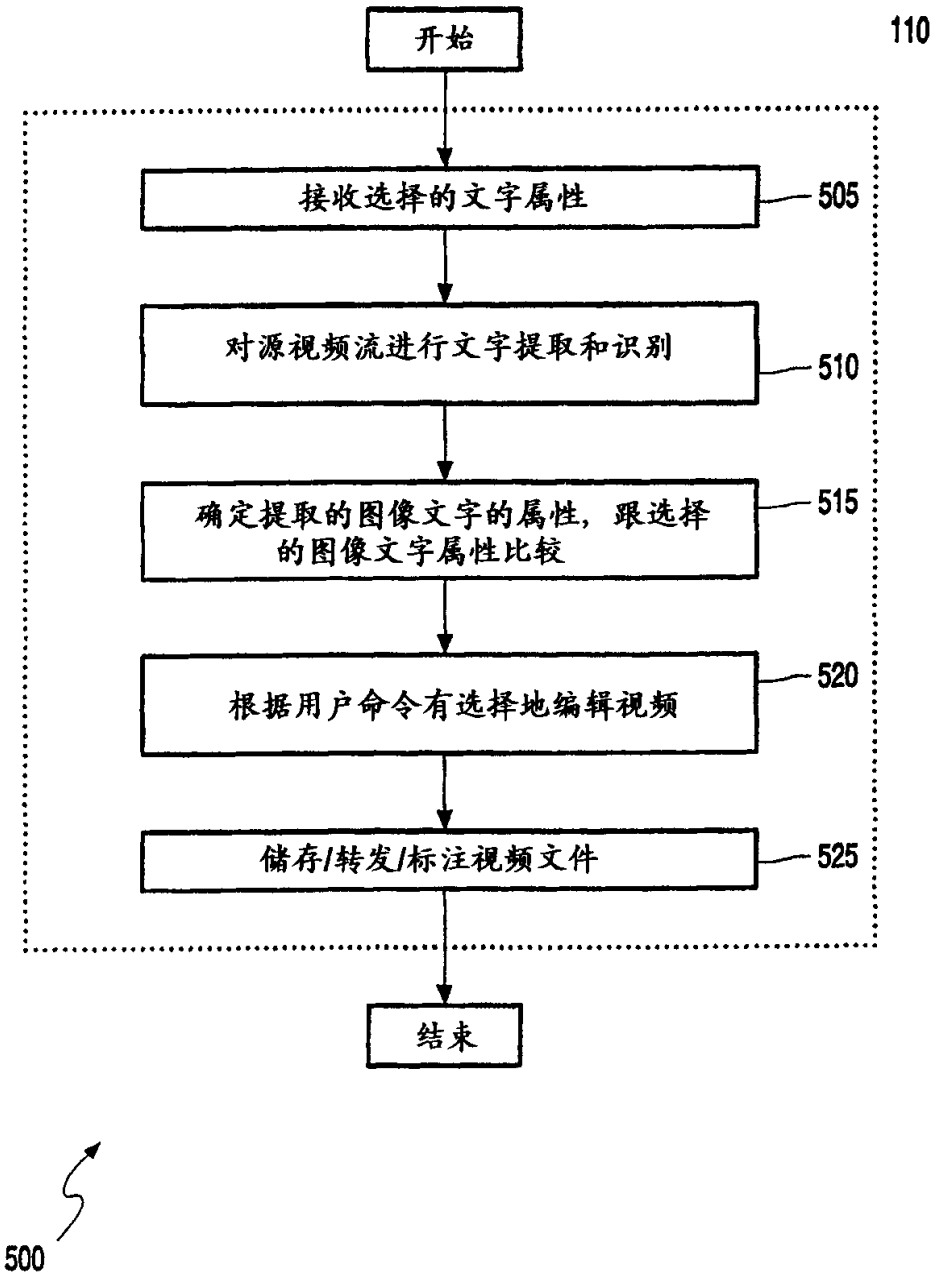

开始
110
接收选择的文字属性
505
对源视频流进行文字提取和识别
510
确定提取的图像文字的属性，跟选择的图像文字属性比较
515
根据用户命令有选择地编辑视频
520
储存/转发/标注视频文件
525
结束
500

图 5