(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102332096 A
(43)申请公布日 2012.01.25

(21)申请号 201110315054.3

(22)申请日 2011.10.17

(71)申请人 中国科学院自动化研究所
地址 100190 中国北京市海淀区中关村东路95号

(72)发明人 刘成林 白博 殷飞

(74)专利代理机构 中科专利商标代理有限责任公司 11021
代理人 周国城

(51)Int.Cl.
G06K 9/20(2006.01)
G06K 9/32(2006.01)
G06K 9/46(2006.01)

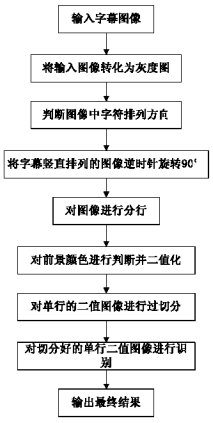
权利要求书 4 页 说明书 8 页 附图 5 页

(54)发明名称

一种视频字幕文本提取和识别的方法

(57)摘要

本发明公开了一种视频字幕文本提取和识别的方法，其步骤包括：输入视频中字幕区域的图像；将输入图像转化为灰度图；判断字幕区域中字符排列的方向；将竖直排列的字幕区域逆时针旋转90°得到水平字幕区域；对字幕区域进行分行得到单行字幕图像；对单行字幕区域图像自动判断前景颜色，得到单行字幕二值图像；对单行字幕二值图像进行过切分得到字符片段序列；对过切分后的单行字幕二值图像进行文本行识别。本方法能有效分割水平和竖直的视频字幕文本行，准确判断字符前景颜色并滤除噪声，并得到准确的字符切分与识别结果，可以适用于视频和图像内容编辑、索引与检索等多种用途。



CN 102332096 A

1. 一种视频字幕文本提取和识别的方法，其特征在于，该方法包括：

步骤 S1：输入视频中字幕区域的图像；

步骤 S2：将输入图像转化为灰度图；

步骤 S3：判断字幕区域中字符排列的方向；

步骤 S4：如果字幕区域中字符排列的方向为竖直排列，则将竖直排列的字幕区域逆时针旋转 90° 得到水平字幕区域；

步骤 S5：对字幕区域进行分行得到单行字幕图像；

步骤 S6：对单行字幕图像自动判断前景颜色，并得到真实的单行字幕二值图像；

步骤 S7：对单行字幕二值图像进行过切分得到字符片段序列；

步骤 S8：对过切分后的单行字幕二值图像进行文本行识别。

2. 如权利要求 1 所述的视频字幕文本提取和识别方法，其特征在于，步骤 S5 中对字幕区域进行分行具体包括如下步骤：

步骤 S51：利用 Sobel 算子求取字幕区域中每个像素点的边缘强度，得到字幕区域边缘图像；

步骤 S52：利用大津法 Otsu 对字幕区域边缘图像进行二值化得到二值图像；

步骤 S53：统计二值图像中每一像素行的有效边缘点数；从所述二值图像的起始行开始，按某一扫描方向逐行进行扫描，当某行中有效边缘点数超过阈值时，以该行为基准，按扫描方向的反方向倒退三行作为一个文本行的开始，然后按扫描方向跳跃 20 行，继续按扫描方向进行扫描；当某一行中所含有效边缘点数低于阈值时，以该行为基础，按扫描方向前进三行作为一个文本行的结束；重复以上过程，直至扫描完最后一行停止；最后得到了所有的文本行区域。

3. 如权利要求 1 所述的视频字幕文本提取和识别方法，其特征在于，步骤 S6 对单行的字幕区域自动判断字符前景颜色，并得到真实的单行字幕二值图像的具体过程包括如下步骤：

步骤 S61：对单行字幕图像进行局部二值化；

步骤 S62：计算两个全局阈值：高亮度阈值 ThH 和低亮度阈值 ThL；

步骤 S63：对于单行字幕图像中的每一个像素点，如果其局部二值化的输出为 1，并且本身灰度值高于高亮度阈值 ThH，则记为前景候选 1；如果其局部二值化的输出为 0，并且本身灰度值低于低亮度阈值 ThL，记为前景候选 2；其他不符合以上条件的像素点不作为前景候选；

步骤 S64：基于前景候选 1 和前景候选 2 分别生成二值图像，对每个二值图像分别进行去噪和是否为真实前景进行打分；分低的二值图像为最终的单行字幕二值图像。

4. 如权利要求 3 所述的视频字幕文本提取和识别方法，其特征在于，所述步骤 S64 中对两种前景候选的二值图像分别进行去噪和是否为真实前景进行打分的具体过程包括如下步骤：

步骤 S641：将所有对应前景候选 1 或前景候选 2、且距离单行字幕图像边缘距离大于 2 的像素点记为 1，其他像素点记为 0，生成二值图像；

步骤 S642：利用每个连通部件与背景的颜色对比度、几何形状、位置关系以及与字符的相似程度等信息，对步骤 S641 所得的二值图像进行去噪；

步骤 S643：对去噪后得到的二值图像进行形态打分，得到分值 M；

步骤 S644：对去噪后得到的二值图像进行笔画宽度一致性打分，得到分值 T；

步骤 S645：最终该二值图像的前景真实度分值为 TM = 0.6×T+0.4×M。

5. 如权利要求 4 所述的视频字幕文本提取和识别方法，其特征在于，所述步骤 S642 中利用每个连通部件与背景，即标记为 0 的像素点，的颜色对比度、几何形状、位置关系以及与字符的相似程度，对 S641 所得的二值图像进行去噪包括如下步骤：

步骤 S6421：从步骤 S641 得到的的二值图像中提取连通部件，顺序分析和处理每一个连通部件；

步骤 S6422：设第 i 个连通部件的 top_i 为所含像素点纵坐标的最小值，$bottom_i$ 为所含像素点纵坐标的最大值，$left_i$ 为所含像素点横坐标的最小值，$right_i$ 为所含像素点横坐标的最大值，单行字幕图像的行高为 h；

步骤 S6423：对第 i 个连通部件的每个边缘点，计算其与相邻背景点的灰度值差 δ，如果 δ > v，其中，v 为整行图像像素点灰度值的方差，则将该边缘点记为合理边缘点；

步骤 S6424：设第 i 个连通部件所含边缘点数为 EN_i，合理边缘点数为 REN_i，若满足条件 $\frac{REN_i}{EN_i} < 0.5$，则删除该连通部件；

步骤 S6425：设第 i 个连通部件的宽、高分别为 w_i、h_i，若满足如下条件之一：(1) $\max(w_i, h_i) < 0.2\times h \cap \min(w_i, h_i) < 0.1\times h$，(2) $w_i > 2\times h \cap h_i < 0.4\times h$，则删除该连通部件；

步骤 S6426：设第 i 个连通部件的垂直中心位置为 $CH_i = \frac{top_i + bottom_i}{2}$，若满足 $CH_i < 0.2\times h \cup CH_i > 0.8\times h$，则删除该连通部件；

步骤 S6427：设第 i 个连通部件的平均笔画宽度为 SW_i，所有连通部件笔画宽度的平均值为 SW，若第 i 个连通部件满足 $SW_i > 1.5\times SW \cap w_i < h$，则删除该连通部件。

6. 如权利要求 4 所述的视频字幕文本提取和识别方法，其特征在于，所述步骤 S643 中对去噪后得到的二值图像进行形态打分，得到分值 M 的具体过程包括如下步骤：

步骤 S6431：对 M 进行初始化，M = 0；

步骤 S6432：统计步骤 S642 去噪处理后所有剩余连通部件的平均高度 averageh、平均宽度 averagew、最大高度 maxh、最大宽度 maxw、所有剩余连通部件的总跨度 length，如果满足如下条件之一：(1) 剩余连通部件的数目为 0，(2) $averagew < 0.3\times h$，(3) $averageh < 0.3\times h$，(4) $maxh < 0.5\times h$，(5) $maxw < 0.5\times h$，其中，h 为单行字幕图像的行高，则该二值图像的形态打分 M 为 1000；

步骤 S6433：若 M ≠ 1000，估计二值图像中整行字的上边缘 ET，下边缘 EB，有效连通部件的数目 usefulNum，有效连通部件所含像素点数目的均值 averageNum，平均字符宽度 averageWid；

步骤 S6434：如果该二值图像满足如下条件之一：(1) $usefulNum < 0.5\times \frac{length}{averageWid}$，(2) $usefulNum > 2\times \frac{length}{averageWid}$，则该二值图像的形态打分 M 为 100；

步骤 S6435：若 M ≠ 1000 且 M ≠ 100，该二值图像的形态打分 M 为：

$$M=\frac{\sum_{i=1}^{usefulNum}\left|blackNum_i-averageNum\right|}{averageNum\times usefulNum},$$

其中，$blackNum_i$ 为第 i 个满足条件 $w_i > 0.3\times h \cap w_i < 0.9\times h \cap h_i > 0.3\times h \cap h_i < 0.9\times h$ 的连通部件所含像素点个数，w_i、h_i 分别为第 i 个连通部件的宽和高。

7. 如权利要求 4 所述的视频字幕文本提取和识别方法，其特征在于，所述步骤 S644 中分值 T 的计算方法为：

$$T=\frac{1}{N}\sum_{i=1}^{N}\left|SW_i-SW\right|,$$

其中，N 为步骤 S642 处理后所有剩余连通部件的数目，SW_i 为第 i 个连通部件的笔画宽度，SW 为所有连通部件笔画宽度的平均值。

8. 如权利要求 1 所述的视频字幕文本提取和识别方法，其特征在于，所述步骤 S7 中对单行字幕二值图像进行过切分的具体过程包括如下步骤：

步骤 S71：计算单行字幕二值图像的行高 averageWid，当作平均字符高度和平均字符宽度。

步骤 S72：计算单行字幕二值图像在 X 轴上的投影，将每一个投影为 0 的区间所在 X 位置作为候选切分点；

步骤 S73：对于每个投影不为 0 的区间，如果其宽度超过平均字符宽度的 0.8 倍，则对其进行过切分，在区间中寻找新的候选切分点，在新的候选切分点处再将该区间分成多个投影不为 0 的区间；

步骤 S74：每个投影不为 0 的区间的二值图像看作一个字符片段，将所有字符片段按从左到右的顺序排序。

9. 如权利要求 8 所述的视频字幕文本提取和识别方法，其特征在于，所述步骤 S73 中对于投影不为 0 的区间进行过切分具体包括如下步骤：

步骤 S731：计算该区间中每一像素列的切分置信度：

$$Conf_i=\frac{(ET-FV_i)^2+(EB-LV_i)^2}{(ET-EB+1)^2}\times\left\{1+\exp\left[20\times\left(0.25-\frac{n_i}{ET-EB+1}\right)\right]\right\},$$

其中，$Conf_i$ 为第 i 列像素的切分置信度，FV_i 为第 i 列所有前景像素点纵坐标的最小值，LV_i 为第 i 列所有前景像素点纵坐标的最大值，n_i 为第 i 列前景像素点数，ET 为单行字幕图像的上边缘估计，EB 为下边缘估计；

步骤 S732：从该区间的左端开始，沿文本行方向，每隔 0.5×averageWid 得到一个假设切分点，记为 C_j；在以 C_j 为中心，0.15×averageWid 为半径的区域内寻找最终切分置信度最大的列为切分位置；其中，最终切分置信度计算如下：设第 i 列的切分置信度为 $Conf_i$，到 C_j 的水平距离为 d_i，则第 i 列的最终切分置信度为：

$$FC_i=Conf_i\times\left(1-\left(\frac{d_i}{averageWid}\right)^2\right)。$$

10. 如权利要求 1 所述的视频字幕文本提取和识别方法，其特征在于，所述步骤 S8 中对过切分后的单行字幕二值图像进行文本识别的具体过程包括如下步骤：

步骤 S81：从左至右顺序考虑每一个字符片段，将字符片段分别与右边相邻的 0

个、1 个、2 个、3 个片段合并，合并后构成的图像前景像素左、右、上、下边界分别表示为 left，right，top，bottom，若该图像前景像素满足如下条件，则构成一个候选字符：(1) (right-left) ＜ 1.5×averageWid，(2)max(right-lef，bottom-top) ＞ 0.6×averageWid，其中，averageWid 为单行字幕二值图像的平均字符宽度；将所有候选字符存储在一个候选切分网格中，其中每个节点对应一个候选切分位置，从起始节点到终止节点的每一条路径，即候选切分路径，表示文本行的一种切分方式，路径上的每一条边表示一个候选字符；

步骤 S82：用一个字符分类器对每个候选字符进行识别，得到几个分数最大的候选类别以及对应的分数；

步骤 S83：对每一条候选切分路径，组合不同候选字符的多个候选类别，得到多条候选切分识别路径；对每一条候选切分识别路径进行评价，得到该路径的评价分数。

步骤 S84：用动态规划算法搜索所有的候选切分识别路径，分数最高的路径给出了最终的文本行字符切分和识别结果。

一种视频字幕文本提取和识别的方法

技术领域

[0001] 本发明属于模式识别与计算机视觉领域，特别是涉及视频图像中的文本检测与识别的处理方法。

背景技术

[0002] 视频作为一种最为流行的媒体形式，通过网络和电视广泛传播。为了使用户更方便、快捷地寻找到感兴趣的视频内容，视频检索与分类逐渐成为模式识别与计算机视觉领域研究的热点。在这其中，视频中的文本信息，特别是字幕信息对于视频的检索以及分类效果最为显著。这是因为：(1) 视频中的文本信息与视频的当前内容密切相关；(2) 视频中的字符有非常明显的视觉特征，便于提取；(3) 字符识别 (OCR) 技术相对目前的语音识别和图像分类技术更为准确和成熟。因此，视频中的文本检测与识别引起了广泛的兴趣。

[0003] 视频中的文本检测与识别过程主要包括以下三个步骤：(1) 文本检测与定位；(2) 文本提取；(3) 字符识别。其中针对步骤 (1) 的研究较多，涌现出了许多适用的方法和技术，如中国知识产权局 2005 年 8 月 24 日公开的公开号为 1658227 的专利（“检测视频文本的方法和装置”）主要根据帧间图像变化检测文本区域。针对步骤 (2)(3) 的技术相对较少，如 2008 年 1 月 30 日公开的公开号为 101115151 的专利（“一种视频字幕提取的方法”）根据颜色判断文字极性并通过局部二值化提取文字，然后用 OCR 软件进行识别。总的来说，现有的视频字幕文字提取和识别的技术还不够完善，主要体现在：对于复杂多变的背景，特别是与前景颜色相似甚至相同的背景无法处理；对于多变的字体，如：宋体、黑体、楷体等的适应性不强；字符识别采用常规的 OCR 方法，对字符切分和字符图像噪声、低分辨率考虑不足。

[0004] 针对上述未解决的技术问题，本发明提出了一种有效的用于视频中字幕区域文本提取与识别的方法。

发明内容

[0005] 本发明的目的是为了克服视频中文本的背景颜色复杂、前景颜色的不确定性、字体多变性和字符切分的不确定性，从而提出了一种对字体、背景颜色具有鲁棒性、并且可以自动判断前景颜色的文本提取和识别方法，对视频字幕文本能够实现快速、准确的提取、切分与识别。

[0006] 本发明提出的一种视频字幕文本提取和识别的方法采用的技术方案为：

[0007] 步骤 S1：输入视频中字幕区域的图像；

[0008] 步骤 S2：将输入图像转化为灰度图；

[0009] 步骤 S3：判断字幕区域中字符排列的方向；

[0010] 步骤 S4：如果字幕区域中字符排列的方向为竖直排列，则将竖直排列的字幕区域逆时针旋转 90° 得到水平字幕区域；

[0011] 步骤 S5：对字幕区域进行分行得到单行字幕图像；

[0012] 步骤 S6 :对单行字幕图像自动判断前景颜色,并得到真实的单行字幕二值图像 ;

[0013] 步骤 S7 :对单行字幕二值图像进行过切分得到字符片段序列 ;

[0014] 步骤 S8 :对过切分后的单行字幕二值图像进行文本行识别。

[0015] 本发明提出的视频字幕文本提取与识别方法的有益效果为 :

[0016] (1) 本发明提出的方法能同时处理水平和竖直的视频字幕文本行 ;

[0017] (2) 本发明提出的方法能够对字幕图像区域进行自动分行,对多文本行的字幕图像进行处理与识别 ;

[0018] (3) 本发明提出的方法通过对单行字幕图像的连通部件进行分析,自动确定字符前景颜色并滤除噪声连通部件,从而得到清晰的字符前景二值化图像 ;

[0019] (4) 本发明提出的方法对文本行图像进行字符过切分得到候选切分方式,并结合候选字符的识别分数和语言上下文模型对候选切分方式进行评价,同时得到字符切分和识别结果,避免了字符宽度变化和间隔不均匀以及字符间笔画粘连造成的切分错误。

[0020] 综合说来,本发明的视频字幕文字提取和识别方法能有效分割字幕文本行,准确判断字符前景颜色并滤除噪声,并得到准确的字符切分与识别结果,可以适用于视频和图像内容编辑、索引与检索等多种用途。

附图说明

[0021] 图 1 是本发明提出的视频字幕文字提取和识别方法流程图。

[0022] 图 2 是对字幕图像进行分行的流程图。

[0023] 图 3 是对单行字幕图像进行自动判断前景并二值化的流程图。

[0024] 图 4 是对二值单行字幕图像进行过切分的流程图。

[0025] 图 5 是本发明方法的实现结构图。

[0026] 图 6 是对单行字图像进行二值化和去噪后的图像示例。

[0027] 图 7 是对二值单行字幕图像进行过切分的图像示例。

[0028] 图 8 是对过切分后单行二值图像进行识别中候选切分网格生成的图像示例。

[0029] 图 9 是一幅水平字幕区域图像的识别结果示例。

[0030] 图 10 是一幅竖直字幕区域图像的识别结果示例。

具体实施方式

[0031] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明进一步详细说明。

[0032] 本发明可在个人电脑、服务器等计算设备上实现。

[0033] 本发明采用的技术方案为 :将视频中的字幕区域进行分行,对每一行自动判断前景颜色并生成二值图像,对二值图像中的字符进行切分与识别,得到最终的文本识别结果。其中对于视频中字幕的定位不是本发明的内容,假设已经用别的方法定位得到了字幕区域。

[0034] 参照图 1,本发明提出的视频字幕文字提取和识别方法具体包括以下几个主要步骤 :

[0035] 步骤 S1 :输入视频中字幕区域的图像 ;

[0036] 步骤 S2：将输入的图像转化为灰度图；

[0037] 将输入图像转化为灰度图的过程是：设输入图像中的每一个像素点的 R、G、B 值分别为 r、g、b，则变换后的灰度图中该像素点的灰度值为 gray = 0.299×r+0.587×g+0.114×b。

[0038] 步骤 S3：判断字幕区域中字符排列的方向；

[0039] 对字幕区域中字符排列方向进行判断的过程是：比较输入图像的长和宽，当输入图像的长大于等于宽时，认为字幕是水平排列；当宽大于长时，认为字幕是竖直排列。

[0040] 步骤 S4：如果字幕区域中字符排列的方向为竖直排列，则将竖直排列的字幕区域逆时针旋转 90° 得到水平字幕区域；

[0041] 设原图像的宽、高分别为 W_1、H_1，$P_1(x, y)$ 为原图像横坐标为 x、纵坐标 y 的点所对应的灰度值；则旋转后的图像的宽、高分别为 $W_2 = H_1$、$H_2 = W_1$，经过逆时针旋转 90°，$P_2(x, y) = P_1(y, x)$ 为旋转后图像横坐标为 x、纵坐标 y 的点所对应的灰度值。

[0042] 步骤 S5：对字幕区域进行分行得到单行字幕图像；

[0043] 参照图 2，对字幕区域进行分行的具体过程包括如下步骤：

[0044] 步骤 S51：利用 Sobel 算子求取字幕区域中每个像素点的边缘强度，得到字幕区域边缘图像；

[0045] 步骤 S52：利用大津法 (Otsu) 对字幕区域边缘图像进行二值化得到二值图像，边缘图像中边缘强度超过阈值的像素点记为有效边缘点，置为 1，否则置为 0；

[0046] 步骤 S53：统计二值图像中每一行（指像素行，以下同）的有效边缘点数；设二值图像的高为 H，有效边缘点总数为 T，设有效边缘点数阈值为 $TH = \frac{T}{2 \times H}$。从二值图像的起始行开始，按某一扫描方向逐行进行扫描，优选地，将二值图像平均分为上下等高的两个部分，分别统计上下两部分的有效边缘点个数，如果上半部分的有效边缘点数多，则以最上一行为起始行，扫描方向为从上至下；否则，以最下一行为起始行，扫描方向为从下至上。当某行中的有效边缘点数超过阈值 TH 时，则以该行为基准，按扫描方向的反方向倒退三行作为一个文本行的开始，然后按扫描方向跳跃 20 行，继续按扫描方向进行扫描；当某一行中所含的有效边缘点数低于阈值 TH 时，则以该行为基础，按扫描方向前进三行作为一个文本行的结束；重复以上过程，直至扫描完最后一行停止；最后就会得到所有的文本行区域。取每个文本行区域的原始灰度图像，即单行字幕图像，进行下面的前景颜色判断和二值化。

[0047] 步骤 S6：对单行字幕图像自动判断前景颜色，并得到真实的单行字幕二值图像；

[0048] 参照图 3，对单行的字幕图像自动判断前景颜色，并得到真实的单行字幕二值图像的具体过程包括如下步骤：

[0049] 步骤 S61：对单行字幕图像进行局部二值化，局部二值化窗口为边长等于三分之一图像高度的正方形，在窗口内的阈值用大津法 (Otsu) 计算，如果窗口中心点的灰度值大于阈值，则二值化输出为 1，低于或等于阈值则二值化输出为 0；

[0050] 步骤 S62：计算两个全局的阈值：高亮度阈值 ThH 和低亮度阈值 ThL，如果单行字幕图像所有像素点的平均灰度值为 m，方差为 v，则高亮度阈值为 ThH = m+0.3×v，低亮度阈值为 ThL = m-0.3×v；

[0051] 步骤 S63：对于单行字幕图像中的每一个像素点，如果其局部二值化的输出为 1，

并且本身灰度值高于高亮度阈值 ThH，则记为前景候选 1；如果其局部二值化的输出为 0，并且本身灰度值低于低亮度阈值 ThL，则记为前景候选 2；其他不符合以上条件的像素点均不作为前景候选；

[0052] 步骤 S64：基于前景候选 1 和前景候选 2 分别生成二值图像，对每个二值图像分别进行去噪和是否为真实前景进行打分，分值为 TM；取得分 (TM) 低的前景二值图像为最终的单行字幕二值图像。

[0053] 所述步骤 S64 中对两种前景候选的二值图像分别进行去噪，并对每个像素是否为真实前景进行打分的具体过程包括如下步骤：

[0054] 步骤 S641：将所有对应当前前景候选（前景候选 1 或前景候选 2），且距离单行字幕图像边缘距离大于 2 的像素点记为 1，其他像素点记为 0，生成二值图像；

[0055] 步骤 S642：对所得二值图像的前景像素，即标记为 1 的像素点提取连通部件，进而利用每个连通部件与背景，即标记为 0 的像素点，的颜色对比度、几何形状、位置关系以及与字符的相似程度等信息，对步骤 S641 所得的二值图像进行去噪；

[0056] 利用每个连通部件与背景的颜色对比度、几何形状、位置关系以及与字符的相似程度等信息，对 S641 所得的二值图像进行去噪的具体过程包括如下步骤：

[0057] 步骤 S6421：从生成的二值图像中提取连通部件，顺序分析和处理每一个连通部件；

[0058] 步骤 S6422：设第 i 个连通部件的 top_i 为所含像素点纵坐标的最小值，$bottom_i$ 为所含像素点纵坐标的最大值，$left_i$ 为所含像素点横坐标的最小值，$right_i$ 为所含像素点横坐标的最大值，设单行字幕图像的行高为 h；

[0059] 步骤 S6423：对第 i 个连通部件的每个边缘点，计算其与相邻背景点的灰度值差 δ，如果 $\delta > v$（v 为整行单行字幕灰度图像所有像素点灰度值的方差），则将该边缘点记为合理边缘点；

[0060] 步骤 S6424：设第 i 个连通部件所含边缘点数为 EN_i，合理边缘点数为 REN_i，若满足条件 $\frac{REN_i}{EN_i} < 0.5$，则删除该连通部件；

[0061] 步骤 S6425：设第 i 个连通部件的宽、高分别为 w_i、h_i，若满足如下条件之一：(1) $\max(w_i, h_i) < 0.2\times h \cap \min(w_i, h_i) < 0.1\times h$，(2) $w_i > 2\times h \cap h_i < 0.4\times h$，则删除该连通部件；

[0062] 步骤 S6426：设第 i 个连通部件的垂直中心位置为 $CH_i = \frac{top_i + bottom_i}{2}$，若满足 $CH_i < 0.2\times h \cup CH_i > 0.8\times h$，则删除该连通部件；

[0063] 步骤 S6427：设第 i 个连通部件的平均笔画宽度为 SW_i，其计算方法如下：设连通部件所含像素点的个数为 N_i，边缘点个数为 C_i，则笔画宽度 $SW_i = \frac{N_i}{2\times C_i}$，设所有连通部件笔画宽度的平均值为 SW，若第 i 个连通部件满足 $SW_i > 1.5\times SW \cap w_i < h$，则删除该连通部件。

[0064] 步骤 S643：对去噪后得到的二值图像进行形态打分，得到分值 M；

[0065] 对去噪后得到的二值图像进行形态打分，得到分值 M 的具体过程包括如下步骤：

[0066] 步骤 S6431：对 M 进行初始化，M = 0；

[0067] 步骤 S6432：统计步骤 S642 去噪处理后所有剩余连通部件的平均高度 averageh、平均宽度 averagew、最大高度 maxh、最大宽度 maxw、所有剩余连通部件的总跨度 length，其中，length = max(right$_i$)-min(left$_i$)，right$_i$ 为第 i 个连通部件中所有像素点横坐标的最大值，left$_i$ 为第 i 个连通部件中所有像素点横坐标的最小值，如果满足如下条件之一：(1) 剩余连通部件的数目为 0，(2)averagew < 0.3×h，(3)averageh < 0.3×h，(4)maxh < 0.5×h，(5)maxw < 0.5×h，该二值图像的形态打分分值 M 为 1000；

[0068] 步骤 S6433：若 M ≠ 1000，估计二值图像中整行字的上边缘 ET，下边缘 EB，有效连通部件的数目 usefulNum，有效连通部件所含像素点数目的均值 averageNum，平均字符宽度 averageWid，计算方法如下：ET 为所有满足 top$_i$ < 0.3×h 的连通部件的 top$_i$ 的平均值，EB 为所有满足 bottom$_i$ > 0.7×h 的连通部件的 bottom$_i$ 的平均值，usefulNum 为二值图像中满足条件 h$_i$ > 0.3×h ∩ h$_i$ < 0.9×h 的连通部件的数目，averageNum 为二值图像中满足条件 h$_i$ > 0.3×h ∩ h$_i$ < 0.9×h 的连通部件所含像素点数目的均值，averageWid 为满足条件 hi > 0.5×h ∩ h$_i$ < h 的连通部件的 hi 的均值；

[0069] 步骤 S6434：如果该二值图像满足如下条件之一：(1) $usefulNum < 0.5\times\frac{length}{averageWid}$，(2) $usefulNum > 2\times\frac{length}{averageWid}$，则形态打分分值 M 为 100；

[0070] 步骤 S6435：若 M ≠ 1000 且 M ≠ 100，形态打分分值 M 的计算方法如下：设 blackNum$_i$ 为第 i 个满足条件 w$_i$ > 0.3×h ∩ w$_i$ < 0.9×h ∩ h$_i$ > 0.3×h ∩ h$_i$ < 0.9×h 的连通部件所含像素点个数，$M=\frac{\sum_{i=1}^{usefulNum}|blackNum_i-averageNum|}{averageNum\times usefulNum}$，其中 blackNum$_i$ 为第 i 个连通部件中所含像素点的个数。

[0071] 步骤 S644：对去噪后得到的二值图像进行笔画宽度一致性打分，得到分值 T；

[0072] 对去噪后得到的二值图像进行笔画宽度一致性打分，得到 T 的计算方法如下：设步骤 S642 处理后所有剩余连通部件的数目为 N，第 i 个连通部件的笔画宽度为 SW$_i$，所有连通部件笔画宽度的平均值为 SW，则$T=\frac{1}{N}\sum_{i=1}^{N}|SW_i-SW|$。

[0073] 步骤 S645：最终该二值图像的前景真实度分值为 TM = 0.6×T+0.4×M；

[0074] 步骤 S7：对单行字幕二值图像进行过切分得到字符片段序列；

[0075] 参照图 4，对单行字幕二值图像进行过切分的方法具体包括如下步骤：

[0076] 步骤 S71：计算单行字幕二值图像的行高，当作平均字符高度和平均字符宽度，记为 averageWid，averageWid = EB-ET。

[0077] 步骤 S72：计算单行字幕二值图像在 X 轴上的投影（每一像素列的前景点个数）；投影为 0 的连续像素列构成一个投影为 0 的区间，投影不为 0 的连续像素列构成一个投影不为 0 的区间；将每一个投影为 0 的区间所在 X 轴的位置作为候选切分点（相邻字符可在候选切分点分隔开）；

[0078] 步骤 S73：对于每个投影不为 0 的区间，如果其宽度超过平均字符宽度的 0.8 倍，则对其进行过切分，在区间中寻找新的候选切分点，在新的候选切分点处再将该区间分成

多个投影不为 0 的区间；

[0079] 对于投影不为 0 的区间进行过切分的过程具体包括如下步骤：

[0080] 步骤 S731：计算该区间中每一列（指像素列）的切分置信度，第 i 列的切分置信度 $Conf_i$ 计算方法如下：设第 i 列所有前景像素点纵坐标的最小值为 FV_i，最大值为 LV_i，该列前景像素点数为 n_i，则

[0081] $$Conf_i = \frac{(ET-FV_i)^2+(EB-LV_i)^2}{(ET-EB+1)^2} \times \left\{1+\exp\left[20\times\left(0.25-\frac{n_i}{ET-EB+1}\right)\right]\right\},$$

[0082] 其中，ET 为单行字幕图像的上边缘估计，EB 为下边缘估计，已在步骤 S6432 中描述；

[0083] 步骤 S732：从该区间的左端开始，沿文本行方向，每隔 0.5×averageWid 得到一个假设切分点，记为 C_j；在以 C_j 为中心，0.15×averageWid 为半径的区域内寻找最终切分置信度最大的列为切分位置；最终切分置信度计算如下：设第 i 列的切分置信度 $Conf_i$，到 C_j 的水平距离为 d_i，则第 i 列的最终切分置信度为：

[0084] $$FC_i = Conf_i \times (1-(\frac{d_i}{averageWid})^2)；$$

[0085] 步骤 S74：将每个最终分出的投影不为 0 的区间的二值图像看作一个字符片段，将所有字符片段按从左到右的顺序进行排序。

[0086] 步骤 S8：对过切分后的单行字幕图像进行文本行识别。

[0087] 对过切分后的单行字幕图像进行文本识别的目的是同时确定字幕中各个字符的最终切分位置和类别，即同时得到字符切分和识别结果，其具体过程包括如下步骤：

[0088] 步骤 S81：从左至右顺序考虑每一个字符片段，将字符片段分别与右边相邻的 0 个、1 个、2 个、3 个片段合并，合并后构成的图像前景像素左、右、上、下边界分别表示为 left，right，top，bottom，若该图像前景像素满足如下条件，则构成一个候选字符：(1)(right-left) ＜ 1.5×averageWid，(2)max(right-lef,bottom-top) ＞ 0.6×averageWid；将所有候选字符存储在一个候选切分网格中，其中每个节点对应一个候选切分位置，从起始节点（对应文字行的开始位置）到终止节点（对应文字行的结束位置）的每一条路径（称为候选切分路径）表示文本行的一种切分方式，路径上每一条边表示一个候选字符；

[0089] 步骤 S82：用一个字符分类器对每个候选字符进行识别，得到几个（比如 10 个）分数最大的候选类别以及对应的分数；

[0090] 字符分类器从候选字符图像中提取特征，表示为特征矢量 x_i，用一个统计分类器（比如最近原型分类器）对特征矢量进行分类，具体地，计算特征矢量到每一类别集 c_i（类别集是事先指定的，包括常用汉字和英文字母、数字）原型的最近距离 $d_i = d(x_i, c_i)$，选择距离最近的 10 个类别，将其距离通过函数 $P(c_i|\mathbf{x}_i)=\frac{1}{1+e^{\alpha(d_i-\tau)}}$ 转换为概率置信度，即字符识别分数；其中参数 τ 为字符分类器训练样本集上每类样本到本类别距离 d(x,c) 的均值，α 经验性地设为 2/τ。

[0091] 步骤 S83：对每一条候选切分路径，组合不同候选字符的多个候选类别，得到多条候选切分识别路径，该路径中同时包括候选字符及每个候选字符对应的类别；对每一条候选切分识别路径进行评价，得到该路径的评价分数。

[0092] 所述步骤 S83 中对于候选切分识别路径的评价具体为结合候选类别的分数和统计语言模型（通常用 Bi-gram）给出路径的评价分数：设候选切分路径 X 上有 n 个候选字符，对应的候选类别依次为 $C = c_1c_2\cdots c_n$，候选切分识别路径的分数为

[0093] $$f(X,C)=\sum_{i=1}^{n}[k_i \log P(c_i|\mathbf{x}_i)+\lambda \log P(c_i|c_{i-1})],$$

[0094] 其中，k_i 为构成候选字符（其对应的特征矢量为 x_i）的字符片段个数，$P(c_i|c_{i-1})$ 为事先得到的统计语言模型 Bi-gram，λ 为经验设定的权值（0 到 1 之间）。

[0095] 步骤 S84：用动态规划算法搜索所有的候选切分识别路径，分数最高的路径给出的文本行字符切分和识别结果即为最终处理结果。

[0096] 其中，字符分类器的特征提取和分类器设计在模式识别领域有很多公开的具体方法，因而不是本发明的主要内容，代表性的方法可参考文献：

[0097] [1]C.-L.Liu, K.Nakashima, H.Sako, H.Fujisawa, Handwritten digit recognition:Investigation of normalization and feature extraction techniques, Pattern Recognition,37(2):265-279,2004.

[0098] [2]X.-B.Jin, C.-L.Liu, X.Hou, Regularized margin-based conditional log-likelihood loss for prototype learning, Pattern Recognition,43(7): 2428-2438,2010.

[0099] 参照图 5，本发明主要包括以下四个模块：字幕区域分行模块 105、字符前景判断和二值化模块 106、过切分模块 107 和文本行识别模块 108，其他模块均为辅助的输入 / 输出或控制模块，其中：

[0100] 字幕图像输入模块 101，用于获取字幕区域图像，字幕区域由其他字幕定位技术对字幕进行检测和定位得到，或者假定视频图像中一个固定区域为字幕区域。

[0101] 灰度图像转换模块 102，利用公式 gray = 0.299×r+0.587×g+0.114×b，将彩色图像转化成灰度图像。

[0102] 字符排列方向判断模块 103，通过比较输入图像的长、宽，来确定字符的排列方向，当输入图像的长度大于等于宽度时，判定字幕为水平方向；反之，则判定为垂直方向。

[0103] 字幕旋转模块 104，用于将字符垂直排列的图像转化成字符水平排列的图像。

[0104] 字幕区域分行模块 105，用于将字符水平排列的灰度图像进行细分，当该图像含有多行文本时，将其拆分成多个单行文本图像；当该图像只包含一行文本时，对该行文本进行位置修正，得到垂直方向上字符位置居中、外围包含 3 个像素单纯背景的文本图像。

[0105] 字符前景判断和二值化模块 106，对单行的字幕图像自动判断前景颜色，并得到真实的单行字幕二值图像。

[0106] 过切分模块 107，对二值化后的单行字幕图像进行过切分得到字符片段序列。

[0107] 文本行识别模块 108，对过切分后的单行字幕图像进行文本行识别。

[0108] 本发明的具体实施效果如图 6 至图 10 所示。

[0109] 参考图 6，左边三个水平字幕行图像从上至下依次为：彩色字幕区域图像、局部二值化之后的图像、字符前景二值化图像；右边三个垂直字幕行图像从左至右分别为：彩色字幕区域图像、局部二值化之后的图像、字符前景二值化图像。

[0110] 参考图 7，左边和右边三个字幕分图像从上至下分别为：彩色字幕区域图像、字符

前景二值化图像、字符过切分效果图（垂直白线表示候选切分位置）。

[0111] 图 8 为候选切分网格，每一条折线表示一种切分路径，加粗的切分路径表示最终的字符切分结果。

[0112] 图 9 是一个水平字幕行图像文字提取和识别的完整过程：第一行为输入彩色图像，第二行为两种候选前景色生成的二值图像，第三行为过切分效果图，第四行为最终的文本行识别结果。

[0113] 图 10 是一个垂直字幕行图像文字提取和识别的完整过程：第一列为输入彩色图像，第二列为行分割结果，第三列分别为两种候选前景色生成的二值图像，第四列为过切分效果图，第五列为最终的文本行识别结果。

[0114] 以上所述的具体实施例，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施例而已，并不用于限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

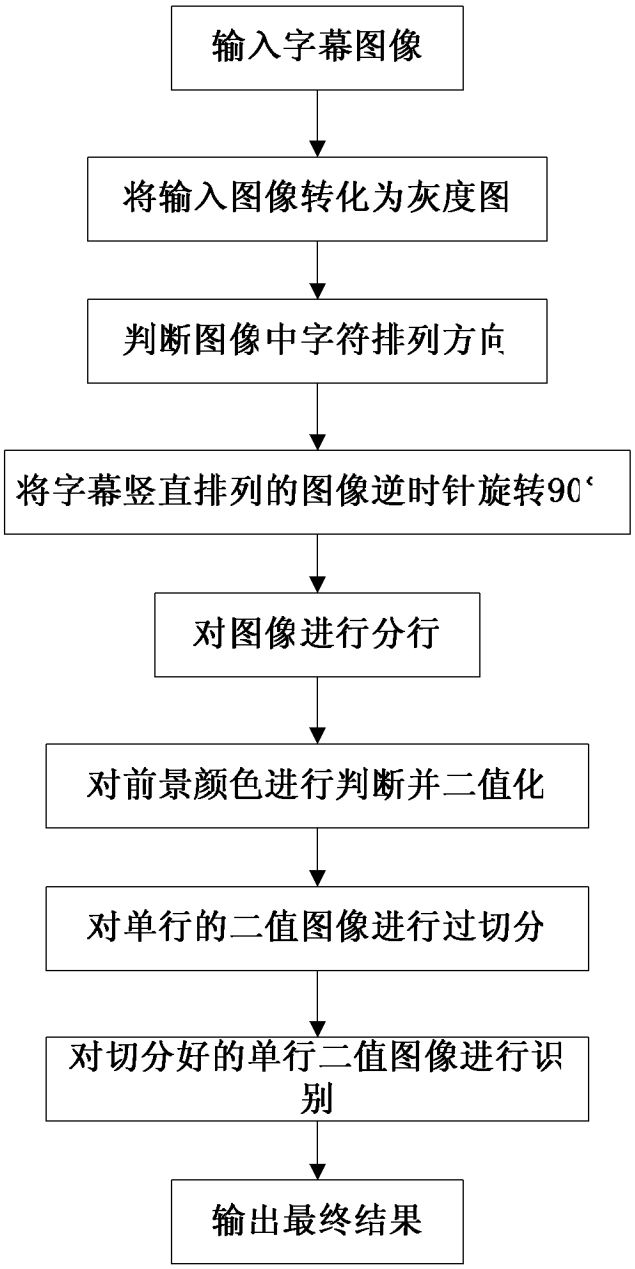

输入字幕图像
将输入图像转化为灰度图
判断图像中字符排列方向
将字幕竖直排列的图像逆时针旋转90°
对图像进行分行
对前景颜色进行判断并二值化
对单行的二值图像进行过切分
对切分好的单行二值图像进行识别
输出最终结果

图 1

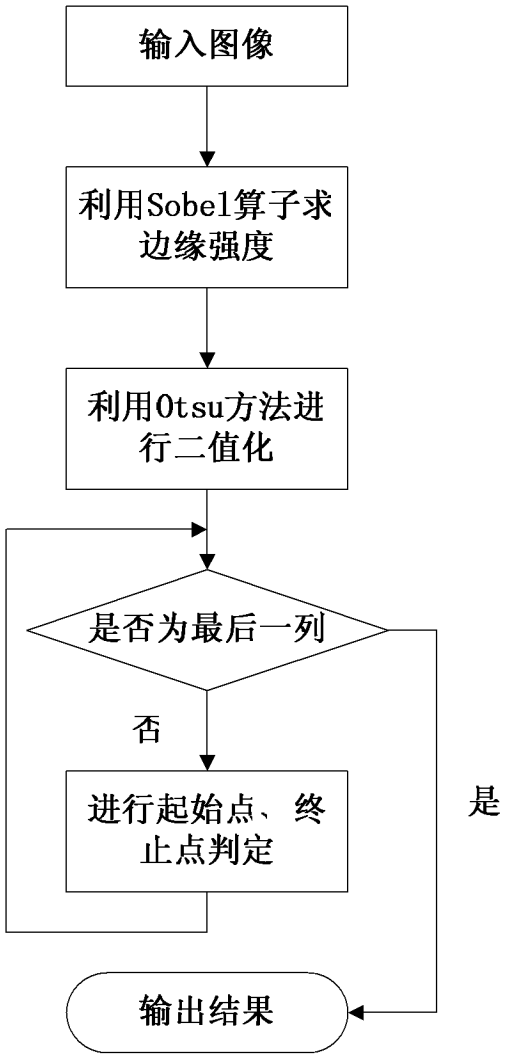

输入图像
利用Sobel算子求边缘强度
利用Otsu方法进行二值化
是否为最后一列
否
进行起始点、终止点判定
是
输出结果

图 2

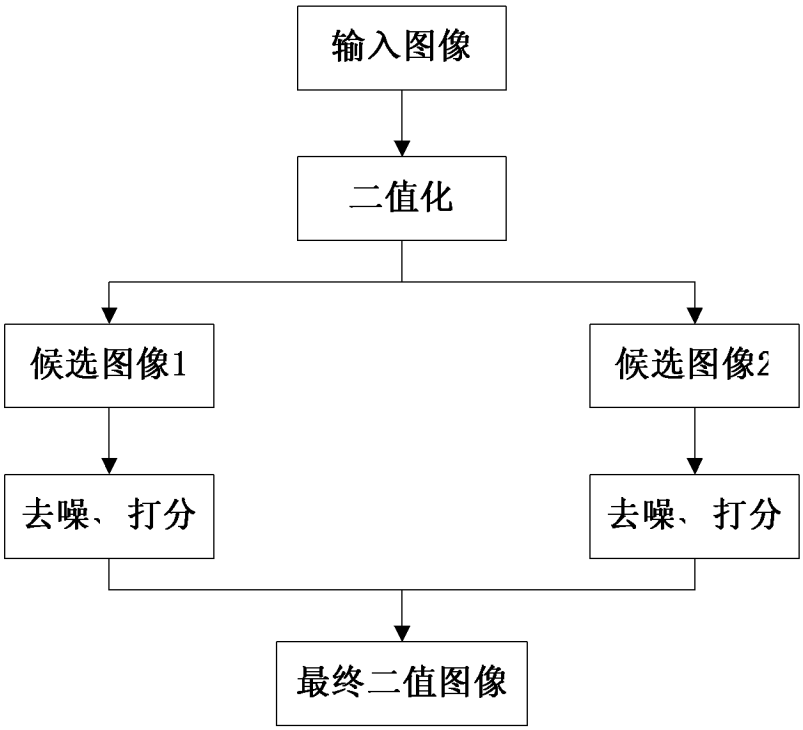


图 3

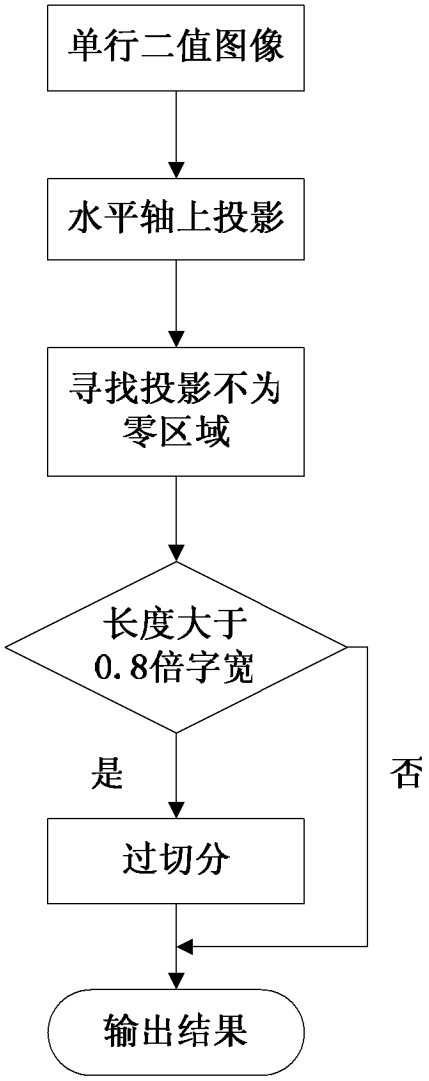

单行二值图像
水平轴上投影
寻找投影不为零区域
长度大于0.8倍字宽
是
否
过切分
输出结果

图 4

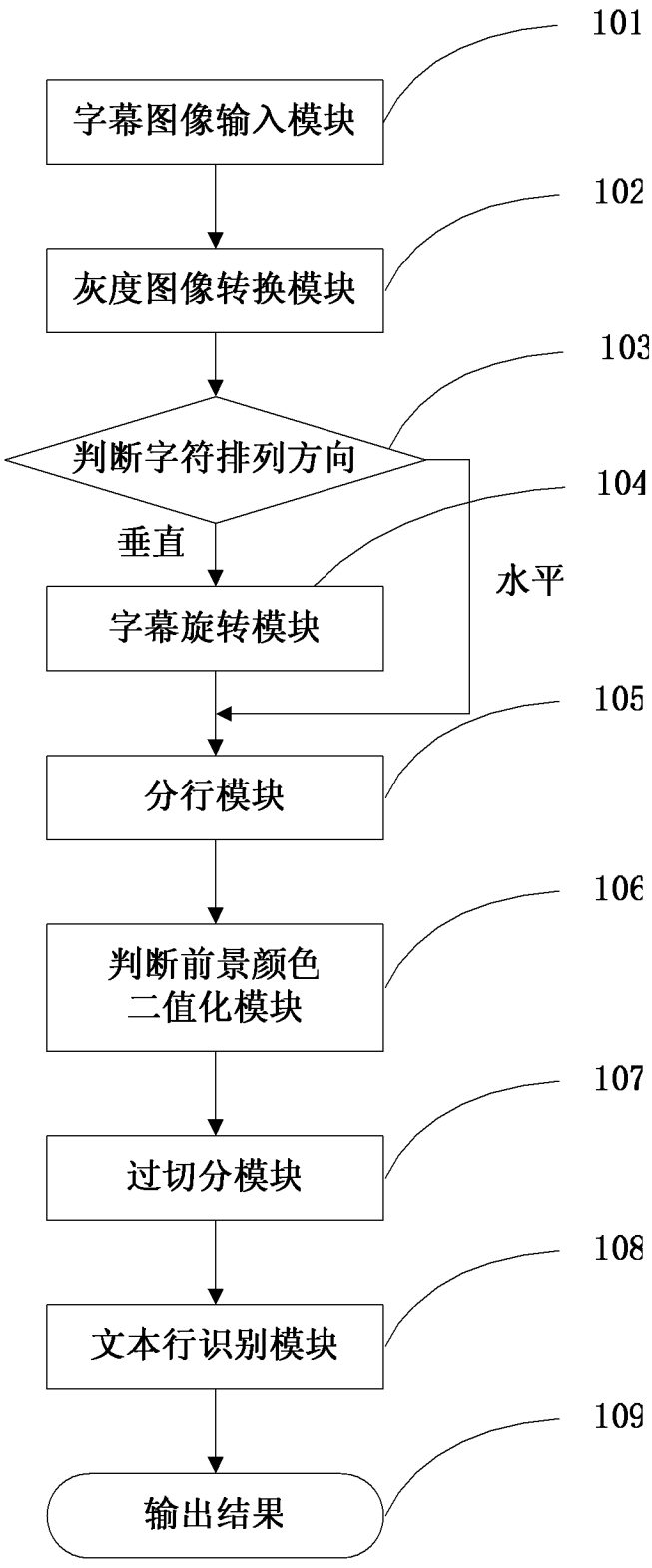

101
字幕图像输入模块
102
灰度图像转换模块
103
判断字符排列方向
垂直
水平
104
字幕旋转模块
105
分行模块
106
判断前景颜色二值化模块
107
过切分模块
108
文本行识别模块
109
输出结果

图 5

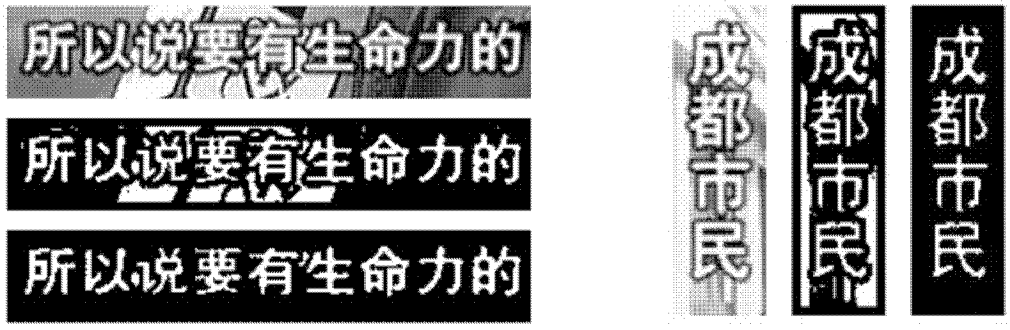

所以说要有生命力的
所以说要有生命力的
所以说要有生命力的
成都市民
成都市民
成都市民

图 6

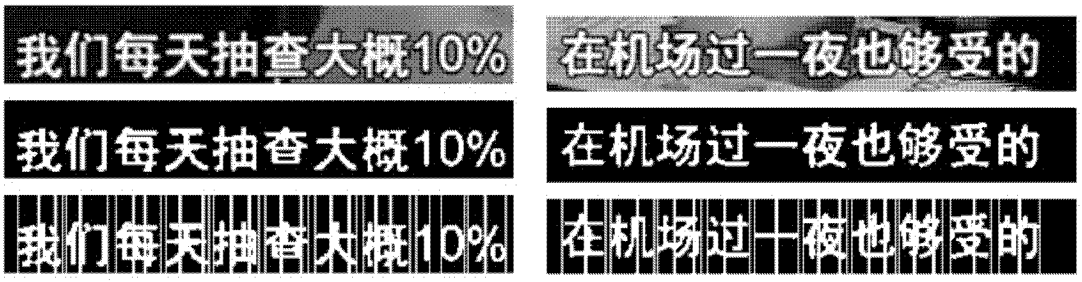

我们每天抽查大概10%
我们每天抽查大概10%
我们每天抽查大概10%
在机场过一夜也够受的
在机场过一夜也够受的
在机场过一夜也够受的

图 7

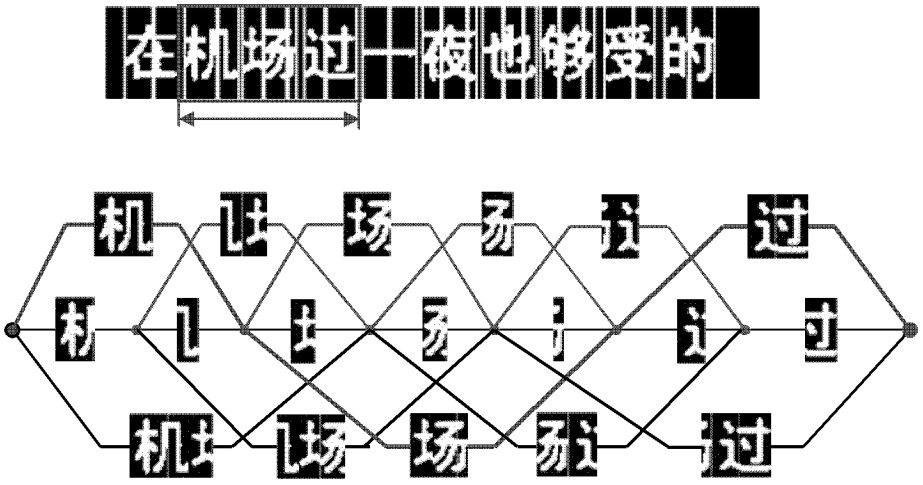

在机场过一夜也够受的

图 8

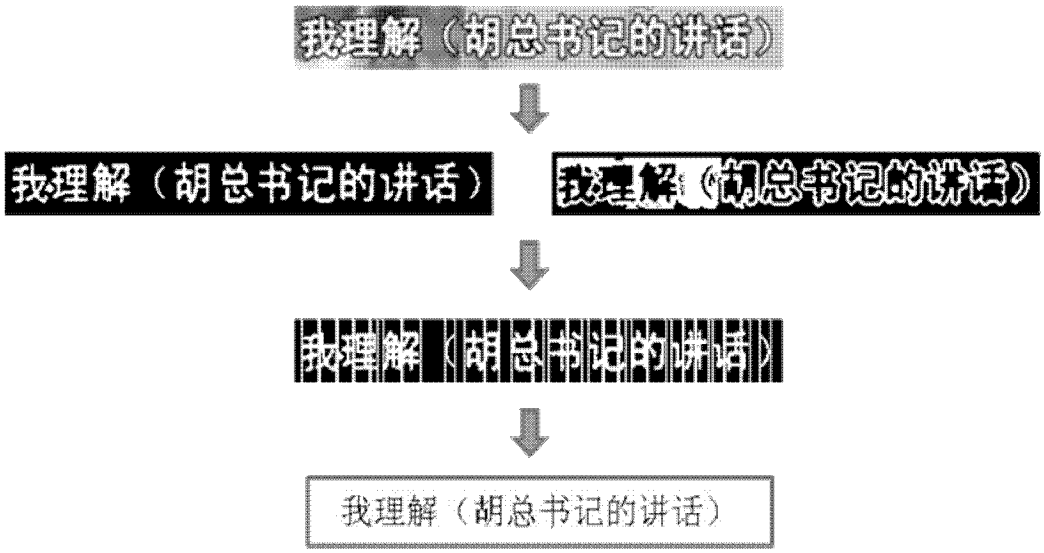

我理解（胡总书记的讲话）
我理解（胡总书记的讲话）
我理解（胡总书记的讲话）
我理解（胡总书记的讲话）
我理解（胡总书记的讲话）

图 9

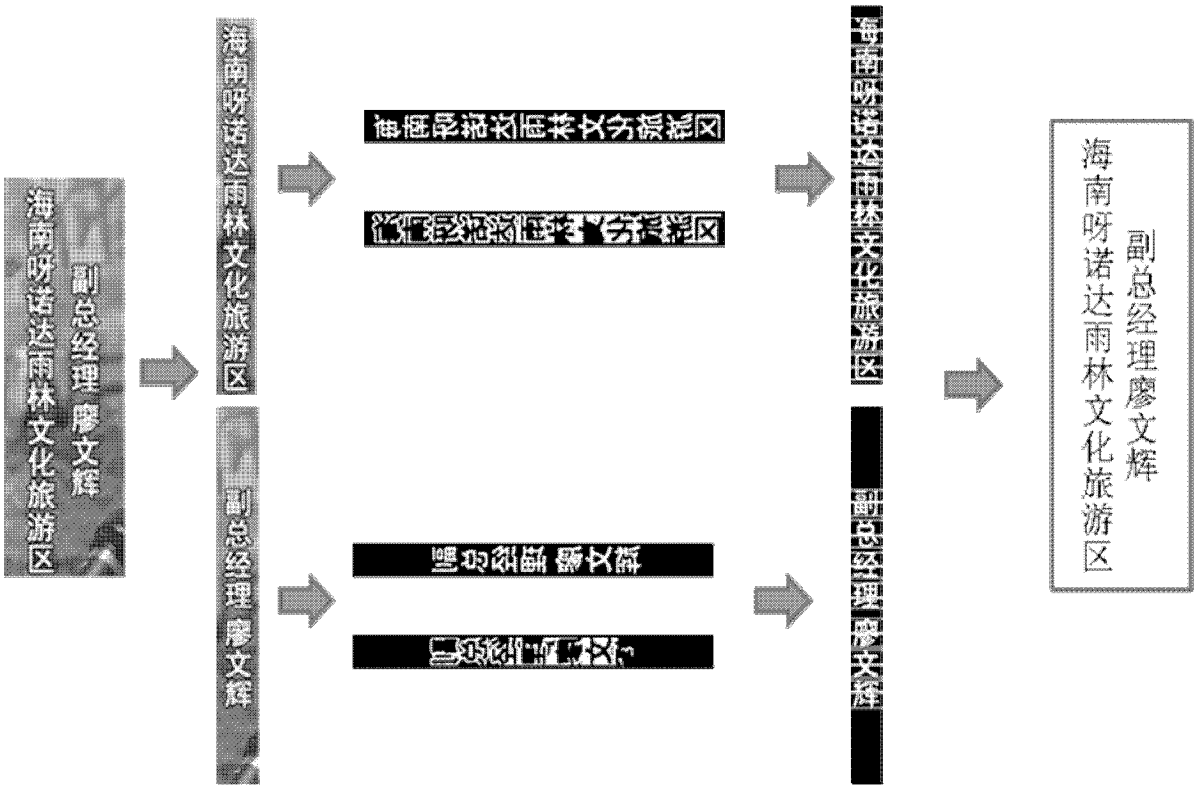

海南呀诺达雨林文化旅游区
副总经理 廖文辉
海南呀诺达雨林文化旅游区
副总经理 廖文辉
副总经理 廖文辉
海南呀诺达雨林文化旅游区

图 10