

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 10 月 29 日 (29.10.2020)



(10) 国际公布号
WO 2020/215696 A1

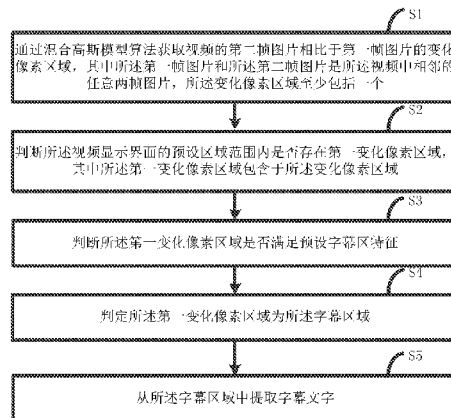
- (51) 国际专利分类号:
G06K 9/32 (2006.01)
- (21) 国际申请号: PCT/CN2019/118411
- (22) 国际申请日: 2019 年 11 月 14 日 (14.11.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910324978.6 2019 年 4 月 22 日 (22.04.2019) CN
- (71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
- (72) 发明人: 肖玉宾(XIAO, Yubin); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号

平安金融中心 23 楼, Guangdong 518000 (CN)。
喻红(YU, Hong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

- (74) 代理人: 深圳市明日今典知识产权代理有限公司(普通合伙)(SHENZHEN MINGRIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM (GENERAL)); 中国广东省深圳市南山区南头街道安乐社区关口二路 15 号智恒产业园 30 栋 4 层 406 室, Guangdong 518000 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,

(54) Title: METHOD FOR EXTRACTING VIDEO SUBTITLES, DEVICE, COMPUTER APPARATUS AND STORAGE MEDIUM

(54) 发明名称: 提取视频字幕的方法、装置、计算机设备及存储介质



- S1 Obtaining a change pixel area of a second frame of picture of a video relative to a first frame of picture through a mixed Gaussian model algorithm, wherein the first frame of picture and the second frame of picture are any two adjacent frames in the video, and the change pixel area includes at least one
- S2 Judging whether a first change pixel area is presented in a preset area range of the video display interface, wherein the first change pixel area is included in the change pixel area
- S3 Judging whether the first change pixel area meets preset subtitle area features
- S4 Determining that the first change pixel area is the subtitle area
- S5 Extracting the subtitle characters from the subtitle area

图 1

(57) Abstract: The present application discloses a method and a device for extracting video subtitles, a computer apparatus and a storage medium, wherein the method for extracting video subtitles comprises: obtaining a change pixel area of a second frame of picture of a video relative to a first frame of picture through a mixed Gaussian model algorithm; judging whether a first change pixel area exists in a preset area range of the video display interface; if the first change pixel area exists in the preset area range of the video display interface, judging whether the first change pixel area meets preset subtitle area features; if the first change pixel area meets the preset subtitle area features, determining that the first change pixel area is the subtitle area; extracting the subtitle characters from the subtitle area. Through the mixed Gaussian mixture model algorithm, the change pixel area of the second frame of picture relative to the first frame of picture is determined, and the subtitle area is determined according to the change pixel area, so that the subtitle extraction precision is improved.

PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本申请揭示了提取视频字幕的方法、装置、计算机设备及存储介质, 其中, 提取视频字幕的方法, 包括: 通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域; 判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域; 若所述视频显示界面的预设区域范围内存在所述第一变化像素区域, 则判断所述第一变化像素区域是否满足预设字幕区特征; 若所述第一变化像素区域满足预设字幕区特征, 则判定所述第一变化像素区域为所述字幕区域; 从字幕区域中提取字幕文字。通过混合高斯模型算法确定第二帧图片相比于第一帧图片的变化像素区域, 进而根据变化像素区域确定字幕区域, 提高字幕提取的精准度。

提取视频字幕的方法、装置、计算机设备及存储介质

本申请要求于2019年04月22日提交中国专利局、申请号为2019103249786, 发明名称为“提取视频字幕的方法、装置、计算机设备及存储介质”的中国专利申请的优先权, 其全部内容通过引用结合在本申请中。

5 技术领域

本申请涉及到计算机领域, 特别是涉及到提取视频字幕的方法、装置、计算机设备及存储介质。

背景技术

目前自动语音识别技术的训练需要大量的标注数据, 但是想要获取标注数据目前大多采用两类方法, 一类方法是投入大量人力去录音, 然后人工标注; 另一类是将已有的录音进行人工的听写标注, 标注数据的成本非常高, 且标注的质量并不高。若将现有视频中的音频数据制成标注数据可大大节省成本, 但音频数据制成标注数据是需要一一对应的文字内容对应, 而目前市场上绝大多数的视频字幕都是与视频合成为一体, 未对视频和字幕进行分离, 现有通过文本定位的方式从视频图片中进行文本定位并提取字幕, 识别过程复杂、识别效率较低。

技术问题

本申请的主要目的为提供提取视频字幕的方法, 旨在解决现有从视频中直接获取到音频对应的字幕信息时过程复杂且识别效率低的技术问题。

20 技术解决方案

本申请提出一种提取视频字幕的方法, 包括:

通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域, 其中所述第一帧图片和所述第二帧图片是所述视频中相邻的任意两帧图片, 所述变化像素区域至少包括一个;

25 判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域, 其中所述第一变化像素区域包含于所述变化像素区域;

若所述视频显示界面的预设区域范围内存在所述第一变化像素区域, 则判断所述第一变化像素区域是否满足预设字幕区特征;

30 若所述第一变化像素区域满足预设字幕区特征, 则判定所述第一变化像素区域为所述字幕区域;

从所述字幕区域中提取字幕文字。

本申请还提供了一种提取视频字幕的装置，包括：

第一获取模块，用于通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域，其中所述第一帧图片和所述第二帧图片是所述视频中相邻的任意两帧图片，所述变化像素区域至少包括一个；

5 第一判断模块，用于判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域，其中所述第一变化像素区域包含于所述变化像素区域；

第二判断模块，用于若所述视频显示界面的预设区域范围内存在第一变化像素区域，则判断所述第一变化像素区域是否满足预设字幕区特征；

判定模块，用于若所述第一变化像素区域满足预设字幕区特征，则判定所述
10 所述第一变化像素区域为所述字幕区域；

提取模块，用于从所述字幕区域中提取字幕文字。

本申请还提供了一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现上述方法的步骤。

本申请还提供了一种计算机可读存储介质，其上存储有计算机程序，所述
15 计算机程序被处理器执行时实现上述的方法的步骤。

有益效果

本申请通过将前一时间序列对应的第一帧图片，作为后一时间序列对应的第二帧图片的背景，以便通过混合高斯模型算法确定第二帧图片相比于第一帧图片的变化像素区域，进而根据变化像素区域确定字幕区域，以便从字幕区域
20 中提取字幕文字，实现字幕文字与视频显示界面的分离，提高字幕提取的精准度。通过字幕区域特有的宽高比特征作为预设字幕区特征。上述预设阈值为获取到字幕的最小宽高比，上述最小宽高比 r 的设定值范围为 r 大于等于视频宽的三分之一。以防 r 设置得太大会造成视频的一个帧图片中，满足条件的字幕区域太少，容易漏选； r 设置得太小会造成提取的字幕位置不准确，计算量增大，
25 且使定位字幕区域的误差增大。本申请的预设区域范围内指视频显示界面中靠近所述视频显示界面的底部边缘，占比所述视频高的四分之一区域，与位于中部区域的视频宽的三分之一区域的交界区域，通过预先设定选择的预设区域范围可极大地降低数据处理量，有利于快速且准确的定位到字幕区域。本申请采用现有的音频分离工具将视频中的音频提取出来并保存，并将字幕文字与切割
30 后的音频文件一一对应完成音频标注，上述标注数据可用于自动语音识别技术的训练时的样本数据，以降低现有人工标注数据的成本，且提高标注数据的质量。

量。

附图说明

图1 本申请一实施例的提取视频字幕的方法流程示意图；

图2本申请一实施例的提取视频字幕的装置结构示意图；

5 图3 本申请一实施例的计算机设备内部结构示意图。

本发明的最佳实施方式

参照图1，本实施例的提取视频字幕的方法，包括：

S1：通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域，其中所述第一帧图片和第二帧图片是所述视频中相邻的任意两帧
10 图片，所述变化像素区域至少包括一个。

本实施例的混合高斯模型算法是多个单模型的组合，提高数据分配的合理性。本实施例中视频的每帧图片中的每个像素由多个单模型描述：

$$P(p) = \{[w_i(x, y, t), u_i(x, y, t), \sigma_i(x, y, t)^2]\}, i = 1, 2, \dots, k$$
, k的值为3到5，表示混合高斯模型中单模型的个数， $w_i(x, y, t)$ 表示每个单模型的权重，满足

$$\sum_{i=1}^k w_i(x, y, t) = 1$$
, $u_i(x, y, t)$ 表示每个单模型的均值， $\sigma_i(x, y, t)^2$ 表示每个单模型对应的方差，上述权重、均值和方差共同确定一个单模型。本实施例中通过将视频的图像序列中的相邻的第一帧图片和第二帧图片输入到混合高斯模型算法中，若第一帧图片和第二帧图片在(x, y)处的像素值对于*i*=1, 2, ..., *k*满足

$$I(x, y, t) - u_i(x, y, t) \leq \lambda * \sigma_i(x, y, t)$$
, 则像素值与该单模型匹配，则判定该
20 像素值为背景，若不存在与该像素值匹配的单模型，则为前景，即视频内容。本实施例通过将相邻且时间早于第二帧图片的第一帧图片，作为第二帧图片的背景，以便确定第二帧图片相比于第一帧图片的变化像素区域，上述变化像素区域为包括差异像素点的区域。

S2：判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域，
25 其中所述第一变化像素区域包含于所述变化像素区域。

本实施例的预设区域范围包括现有字幕常设置的视频显示区域，比如，预设区域范围包括视频显示界面的底部区域的中间位置范围，可通过视频显示界面中的坐标数据，实现定位预设区域范围，以便提高获取字幕区域的精准性，降低数据处理过程中的计算量。本实施通过识别预设区域范围内存在的第一变
30 化像素区域，初步确定可能存在字幕区域。

S3: 若所述视频显示界面的预设区域范围内存在第一变化像素区域, 则判断所述第一变化像素区域是否满足预设字幕区特征。

本实施例通过将第一变化像素区域的特征与预设字幕区特征进行比较, 以便通过预设字幕区特征确定第一变化像素区域是否为字幕区域, 提高确定字幕区域的精准度。第一变化像素区域的特征与预设字幕区特征一致, 或处于预设差异范围之内, 则均认为第一变化像素区域满足预设字幕区特征, 则判定所述第一变化像素区域为所述字幕区域, 否则第一变化像素区域不是所述字幕区域。上述预设字幕区特征包括字幕区的高度值范围、字幕区的宽高比等。

S4: 若所述第一变化像素区域满足预设字幕区特征, 则判定所述第一变化像素区域为所述字幕区域。

本实施例的视频中第二帧图片相比于第一帧图片的变化像素区域中, 包括字幕区域的变化、视频图像变化等, 比如不同帧图像对应不同的字幕内容。本实施例的预设规则遵循现有视频中字幕区域的设置特点进行设定。比如现有字幕区域多设置于视频显示界面的底部区域中间位置, 且常以宽条状形态存在。本实施例首先通过混合高斯模型算法获取各帧图片对应的变化像素区域, 然后再从变化像素区域中确定字幕区域, 进而实现对字幕区域的字幕文字的提取, 可快速从视频文件中准确提取相对应的字幕文字, 以便将字幕文字进行二次处理, 比如标注音频、优化显示过程甚至制作文本训练样本等。上述字幕区域为字幕文字的图像映射区域, 根据不同的文字对应的映射像素不同, 进而区别不同字幕文字的字幕区域。

S5: 从所述字幕区域中提取字幕文字。

本实施例通过图片中文字识别技术, 从所述字幕区域中提取字幕文字, 实现字幕文字与视频显示界面的分离。以便对字幕文字实现进一步的优化处理。包括优化字幕文字的显示方式, 比如设置为3D显示状态、改变字幕文字的显示颜色, 优化字幕文字的动画显示效果等, 扩大字幕文字的使用范围。

进一步地, 所述判断所述第一变化像素区域是否满足预设字幕区特征的步骤S3, 包括:

S31: 计算所述第一变化像素区域的宽高比, 其中所述第一变化像素区域中沿所述视频时序播放方向为所述宽, 垂直于所述宽的方向为所述高, 所述宽高比为所述宽除以所述高。

S32: 判断所述宽高比是否大于预设阈值。

S33: 若所述宽高比大于预设阈值, 则判定所述第一变化像素区域满足所述预设字幕区特征。

S34: 若所述宽高比不大于预设阈值, 则判定所述第一变化像素区域不满足所述预设字幕区特征。

5 本实施例通过字幕区域特有的宽高比特征作为预设字幕区特征。上述预设阈值为获取到字幕的最小宽高比, 上述最小宽高比 r 的设定值范围为 r 大于等于视频宽的三分之一。以防 r 设置得太大会造成视频的一个帧图片中, 满足条件的字幕区域太少, 容易漏选; r 设置得太小会造成提取的字幕位置不准确, 计算量增大, 且使定位字幕区域的误差增大。

10 进一步地, 所述判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域的步骤S2之前, 包括:

S20a: 获取所述视频的视频宽和视频高, 其中所述视频显示界面中沿所述视频时序播放方向为所述视频宽, 垂直于所述视频宽的方向为所述视频高。

15 S20b: 设定所述预设值等于第一预设值, 设定靠近所述视频显示界面的底部边缘, 占比第二预设值的区域范围为所述预设区域范围。

 本实施例的预设区域范围内指视频显示界面中靠近所述视频显示界面的底部边缘, 占比所述视频高的四分之一区域, 与位于中部区域的视频宽的三分之一区域的交界区域, 即上述第一预设值为所述视频宽的三分之一, 第二预设值为所述视频高的四分之一。通过预先设定选择的预设区域范围可极大地降低
20 数据处理量, 有利于快速且准确的定位到字幕区域。

 进一步地, 所述从所述字幕区域中提取字幕文字的步骤S5, 包括:

S51: 将所述字幕区域从所述第二帧图片中切割分离。

S52: 将分离后的所述字幕区域通过图像文字识别算法识别出所述字幕文字。

25 S53: 将所述字幕文字复制到预设文件中。

S54: 通过预设格式标注所述字幕文字并存储。

 本实施例通过将字幕区域从第二帧图片中切割分离, 进行单独存储, 以便精准地处理字幕区域。通过将按照视频时序依次获得的各帧图片中的字幕区域按照顺序依次输入到OCR (optical character recognition) 文字识别模型中
30 进行文字识别。OCR文字识别是指电子设备 (例如扫描仪或数码相机) 检查纸上打印的字符, 然后用字符识别方法将字符形状翻译成计算机文字的过程; 通

过对字幕区域对应的文本资料进行扫描,然后对字幕区域对应的图像文件进行分析处理,获取文字及版面信息的过程。本实施例的预设格式包括视频的视频文件名、第二帧图片的帧索引、字幕的文字内容、视频总帧数及视频宽高尺寸等。上述预设文本为按照字幕所在帧图片的时序,依次存储的字幕文字内容。

5 进一步地,所述预设格式至少包括所述视频的视频文件名和第二帧图片对应的帧索引,所述通过预设格式标注所述字幕文字并存储的步骤S54之后,包括:

 S55: 根据所述视频的视频文件名和所述第二帧图片对应的帧索引,判断所述预设文件中是否存在与第一字幕文字具有相同标注信息的第二字幕文字,其中所述第一字幕文字和所述第二字幕文字,分别包含于所述预设文件中所有所述字幕文字。

 S56: 若所述预设文件中存在与第一字幕文字具有相同标注信息的第二字幕文字,则判断所述第一字幕文字和所述第二字幕文字的文字内容是否相同。

 S57: 若所述第一字幕文字和所述第二字幕文字的文字内容相同,则删除所述所述第一字幕文字或所述第二字幕文字。

 本实施例的视频文件名为当前视频的文件名,比如AVI.123等等;上述帧索引指帧图片处于所有帧中的排序,比如按照时序位于第三帧的图片。本实施例通过标注格式中标注内容,包括视频的视频文件名和所述第二帧图片的帧索引,初步判断是否出现重复字幕文字,若标注内容不同,则第一字幕文字和第二字幕文字的文字内容不相同;若标注内容相同,再判断具体的文字内容是否相同,若不相同,则第一字幕文字和第二字幕文字的文字内容不相同。通过逐步判断的方式,以便通过标注信息初步判断字幕文字是否相同,以便节省重复调用字符识别方法将字符形状翻译成计算机文字的过程,以节省流程,加快响应效率。本实施例通过识别变化像素区域,避免了连续多帧图片具有相同字幕时重复提取字幕文字的情况,且通过上述的逐步判断方式剔除由于视频背景干扰,而导致重复提取的字幕文字,以净化预设文件中的字幕文字。

 进一步地,从所述字幕区域中提取字幕文字的步骤S5之后,包括:

 S6: 获取所述第二字幕文字的起始时间和终止时间。

 S7: 确定所述起始时间和终止时间范围内对应的第一音频文件。

30 S8: 将所述第一音频文件通过音频截取工具从所述视频对应的音频文件中截取分离。

S9: 将所述第二字幕文字与所述第一音频文件一一对应进行音频标注。

本实施例通过遍历视频中所有的变化像素区域 $a_1, a_2, a_3, \dots, a_n$, 计算各变化像素区域的宽高比 (w/h) 是否大于设定 r , 如果大于设定 r , 则从当前帧的图片中切割对应的字幕区域, 并将当前帧的帧索引换算成对应的时间[帧索引*(1/视频帧率) 就得到当前帧在视频中的时间位置], 并缓存该时间点的字幕区域, 将本次缓存的字幕区域与上一时序缓存的字幕区域进行像素对比, 差异小于预设阈值时, 则上一次缓存的时间位置与当前新的缓存时间位置作为上一时序对应帧的字幕区域的时间间隔, 并将时间间隔与上一时序对应帧的字幕区域的字幕文字关联保存。本实施例采用现有的音频分离工具将视频中的音频提取出来并保存, 并将字幕文字与切割后的音频文件一一对应完成音频标注, 上述标注数据可用于自动语音识别技术的训练时的样本数据, 以降低现有人工标注数据的成本, 且提高标注数据的质量。

进一步地, 所述通过预设格式标注所述字幕文字并存储的步骤S54之后, 还包括:

S541: 根据所述视频的视频文件名和所述第二帧图片的帧索引, 判断所述预设文件中是否存在与第一帧索引对应的空字幕文件, 其中所述第一帧索引包含于所述预设文件中所有帧索引中;

S542: 若存在, 则将所述第一帧索引对应的帧图片, 输入文本定位模型;

S543: 判断根据所述文本定位模型是否提取到所述第一帧索引对应的帧图片的指定字幕文字;

S544: 若提取到, 则将所述指定字幕文字补充到所述预设文件中所述第一帧索引对应位置;

S545: 若未提取到, 则在所述预设文件中所述第一帧索引对应位置标记为空字幕。

本实施例通过帧索引与字幕文字的对应关系, 查找遗漏提取的字幕文字, 以保证整个视频文件中的字幕文字的完整性。当查找到第一帧索引对应空字幕文件, 即无对应的字幕文字与其相对应, 则判定存在遗漏提取, 则将第一帧索引对应的帧图片, 输入文本定位模型, 以根据文本定位模型进行字幕文字定位与提取。上述文本定位模型为CTPN, CTPN结合CNN与LSTM深度网络, CTPN是从Faster R-CNN改进而来, 能有效的检测出复杂场景的横向分布的文字, 比如识别视频图片中的文字, 虽然识别过程复杂、识别效率较低, 但识别精度高, 可

有效补充通过混合高斯模型算法遗漏的字幕区域,提高整个视频文件中的字幕文字的完整性。

本实施例通过将前一时间序列对应的第一帧图片,作为后一时间序列对应的第二帧图片的背景,以便通过混合高斯模型算法确定第二帧图片相比于第一帧图片的变化像素区域,进而根据变化像素区域确定字幕区域,以便从字幕区域中提取字幕文字,实现字幕文字与视频显示界面的分离,提高字幕提取的精准度。通过字幕区域特有的宽高比特征作为预设字幕区特征。上述预设阈值为获取到字幕的最小宽高比,上述最小宽高比 r 的设定值范围为 r 大于等于视频宽的三分之一。以防 r 设置得太大会造成视频的一个帧图片中,满足条件的字幕区域太少,容易漏选; r 设置得太小会造成提取的字幕位置不准确,计算量增大,且使定位字幕区域的误差增大。本实施例的预设区域范围内指视频显示界面中靠近所述视频显示界面的底部边缘,占比所述视频高的四分之一区域,与位于中部区域的视频宽的三分之一区域的交界区域,通过预先设定选择的预设区域范围可极大地降低数据处理量,有利于快速且准确的定位到字幕区域。

本实施例采用现有的音频分离工具将视频中的音频提取出来并保存,并将字幕文字与切割后的音频文件一一对应完成音频标注,上述标注数据可用于自动语音识别技术的训练时的样本数据,以降低现有人工标注数据的成本,且提高标注数据的质量。

参照图2,本实施例的提取视频字幕的装置,包括:

第一获取模块1,用于通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域,其中所述第一帧图片和第二帧图片是所述视频中相邻的任意两帧图片,所述变化像素区域至少包括一个。

本实施例的混合高斯模型算法是多个单模型的组合,提高数据分配的合理性。本实施例中视频的每帧图片中的每个像素由多个单模型描述:

$P(p) = \{[w_i(x, y, t), u_i(x, y, t), \sigma_i(x, y, t)^2]\}, i = 1, 2, \dots, k$, k 的值为3到5,表示混合高斯模型中单模型的个数, $w_i(x, y, t)$ 表示每个单模型的权重,满足 $\sum_{i=1}^K w_i(x, y, t) = 1$, $u_i(x, y, t)$ 表示每个单模型的均值, $\sigma_i(x, y, t)^2$ 表示每个单模型对应的方差,上述权重、均值和方差共同确定一个单模型。本实施例中通过将视频的图像序列中的相邻的第一帧图片和第二帧图片输入到混合高斯模型算法中,若第一帧图片和第二帧图片在 (x, y) 处的像素值对于 $i = 1, 2, \dots, k$ 满足

$I(x, y, t) - u_i(x, y, t) \leq \lambda * \sigma_i(x, y, t)$ ，则像素值与该单模型匹配，则判定该像素值为背景，若不存在与该像素值匹配的单模型，则为前景，即视频内容。本实施例通过将相邻且时间早于第二帧图片的第一帧图片，作为第二帧图片的背景，以便确定第二帧图片相比于第一帧图片的变化像素区域，上述变化像素区域为包括差异像素点的区域。

第一判断模块2，用于判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域，其中所述第一变化像素区域包含于所述变化像素区域。

本实施例的预设区域范围包括现有字幕常设置的视频显示区域，比如，预设区域范围包括视频显示界面的底部区域的中间位置范围，可通过视频显示界面中的坐标数据，实现定位预设区域范围，以便提高获取字幕区域的精准性，降低数据处理过程中的计算量。本实施通过识别预设区域范围内存在的第一变化像素区域，初步确定可能存在字幕区域。

第二判断模块3，用于若所述视频显示界面的预设区域范围内存在第一变化像素区域，则判断所述第一变化像素区域是否满足预设字幕区特征。

本实施例通过将第一变化像素区域的特征与预设字幕区特征进行比较，以便通过预设字幕区特征确定第一变化像素区域是否为字幕区域，提高确定字幕区域的精准度。第一变化像素区域的特征与预设字幕区特征一致，或处于预设差异范围之内，则均认为第一变化像素区域满足预设字幕区特征，则判定所述第一变化像素区域为所述字幕区域，否则第一变化像素区域不是所述字幕区域。上述预设字幕区特征包括字幕区的高度值范围、字幕区的宽高比等。

判定模块4，用于若所述第一变化像素区域满足预设字幕区特征，则判定所述第一变化像素区域为所述字幕区域。

本实施例的视频中第二帧图片相比于第一帧图片的变化像素区域中，包括字幕区域的变化、视频图像变化等，比如不同帧图像对应不同的字幕内容。本实施例的预设规则遵循现有视频中字幕区域的设置特点进行设定。比如现有字幕区域多设置于视频显示界面的底部区域中间位置，且常以宽条状形态存在。本实施例首先通过混合高斯模型算法获取各帧图片对应的变化像素区域，然后再从变化像素区域中确定字幕区域，进而实现对字幕区域的字幕文字的提取，可快速从视频文件中准确提取相对应的字幕文字，以便将字幕文字进行二次处理，比如标注音频、优化显示过程甚至制作文本训练样本等。上述字幕区域为字幕文字的图像映射区域，根据不同的文字对应的映射像素不同，进而区别不

同字幕文字的字幕区域。

提取模块5，用于从所述字幕区域中提取字幕文字。

本实施例通过图片中文字识别技术，从所述字幕区域中提取字幕文字，实现字幕文字与视频显示界面的分离。以便对字幕文字实现进一步的优化处理。

- 5 包括优化字幕文字的显示方式，比如设置为3D显示状态、改变字幕文字的显示颜色，优化字幕文字的动画显示效果等，扩大字幕文字的使用范围。

进一步地，所述第二判断模块，包括：

- 10 计算单元，用于计算所述第一变化像素区域的宽高比，其中所述第一变化像素区域中沿所述视频时序播放方向为所述宽，垂直于所述宽的方向为所述高，所述宽高比为所述宽除以所述高。

第一判断单元，用于判断所述宽高比是否大于预设阈值。

第一判定单元，用于若所述宽高比大于预设阈值，则判定所述第一变化像素区域满足所述预设字幕区特征。

- 15 第二判定单元，用于若所述宽高比不大于预设阈值，则判定所述第一变化像素区域不满足所述预设字幕区特征。

- 20 本实施例通过字幕区域特有的宽高比特征作为预设字幕区特征。上述预设阈值为获取到字幕的最小宽高比，上述最小宽高比 r 的设定值范围为 r 大于等于视频宽的三分之一。以防 r 设置得太大会造成视频的一个帧图片中，满足条件的字幕区域太少，容易漏选； r 设置得太小会造成提取的字幕位置不准确，计算量增大，且使定位字幕区域的误差增大。

进一步地，提取视频字幕的装置，包括：

第二获取模块，用于获取所述视频的视频宽和视频高，其中所述视频显示界面中沿所述视频时序播放方向为所述视频宽，垂直于所述视频宽的方向为所述视频高。

- 25 设定模块，用于设定所述预设值等于第一预设值，设定靠近所述视频显示界面的底部边缘，占比第二预设值的区域范围为所述预设区域范围。

- 30 本实施例的预设区域范围内指视频显示界面中靠近所述视频显示界面的底部边缘，占比所述视频高的四分之一区域，与位于中部区域的视频宽的三分之一区域的交界区域，即上述第一预设值为所述视频宽的三分之一，第二预设值为所述视频高的四分之一。通过预先设定选择的预设区域范围可极大地降低数据处理量，有利于快速且准确的定位到字幕区域。

进一步地，所述提取模块5，包括：

分离单元，用于将所述字幕区域从所述第二帧图片中切割分离。

识别单元，用于将分离后的所述字幕区域通过图像文字识别算法识别出所述字幕文字。

5 赋值单元，用于将所述字幕文字复制到预设文件中。

标注单元，用于通过预设格式标注所述字幕文字并存储。

本实施例通过将字幕区域从第二帧图片中切割分离，进行单独存储，以便精准地处理字幕区域。通过将按照视频时序依次获得的各帧图片中的字幕区域按照顺序依次输入到OCR (optical character recognition) 文字识别模型中
10 进行文字识别。OCR文字识别是指电子设备（例如扫描仪或数码相机）检查纸上打印的字符，然后用字符识别方法将字符形状翻译成计算机文字的过程；通过对字幕区域对应的文本资料进行扫描，然后对字幕区域对应的图像文件进行分析处理，获取文字及版面信息的过程。本实施例的预设格式包括视频的视频文件名、第二帧图片的帧索引、字幕的文字内容、视频总帧数及视频宽高尺寸
15 等。上述预设文本为按照字幕所在帧图片的时序，依次存储的字幕文字内容。

进一步地，所述预设格式至少包括所述视频的视频文件名和第二帧图片对应的帧索引，所述提取模块5，包括：

第二判断单元，用于根据所述视频的视频文件名和所述第二帧图片对应的帧索引，判断所述预设文件中是否存在与第一字幕文字具有相同标注信息的第
20 二字幕文字，其中所述第一字幕文字和所述第二字幕文字，分别包含于所述预设文件中所有所述字幕文字。

第三判断单元，用于若所述预设文件中存在与第一字幕文字具有相同标注信息的第二字幕文字，则判断所述第一字幕文字和所述第二字幕文字的文字内容是否相同。

25 删除单元，用于若所述第一字幕文字和所述第二字幕文字的文字内容相同，则删除所述所述第一字幕文字或所述第二字幕文字。

本实施例的视频文件名为当前视频的文件名，比如AVI.123等等；上述帧索引指帧图片处于所有帧中的排序，比如按照时序位于第三帧的图片。本实施例通过标注格式中标注内容，包括视频的视频文件名和所述第二帧图片的帧索引，初步判断是否出现重复字幕文字，若标注内容不同，则第一字幕文字和
30 第二字幕文字的文字内容不相同；若标注内容相同，再判断具体的文字内容是否

相同，若不相同，则第一字幕文字和第二字幕文字的文字内容不相同。通过逐步判断的方式，以便通过标注信息初步判断字幕文字是否相同，以便节省重复调用字符识别方法将字符形状翻译成计算机文字的过程，以节省流程，加快响应效率。本实施例通过识别变化像素区域，避免了连续多帧图片具有相同字幕时重复提取字幕文字的情况，且通过上述的逐步判断方式剔除由于视频背景干扰，而导致重复提取的字幕文字，以净化预设文件中的字幕文字。

进一步地，本申请另一实施例的提取视频字幕的装置，包括：

第三获取模块，用于获取所述第二字幕文字的起始时间和终止时间。

确定模块，用于确定所述起始时间和终止时间范围内对应的第一音频文件。

截取模块，用于将所述第一音频文件通过音频截取工具从所述视频对应的音频文件中截取分离。

标注模块，用于将所述第二字幕文字与所述第一音频文件一一对应进行音频标注。

本实施例通过遍历视频中所有的变化像素区域 $a_1, a_2, a_3, \dots, a_n$ ，计算各变化像素区域的宽高比(w/h)是否大于设定 r ，如果大于设定 r ，则从当前帧的图片中切割对应的字幕区域，并将当前帧的帧索引换算成对应的时间[帧索引*(1/视频帧率) 就得到当前帧在视频中的时间位置]，并缓存该时间点的字幕区域，将本次缓存的字幕区域与上一时序缓存的字幕区域进行像素对比，差异小于预设阈值时，则上一次缓存的时间位置与当前新的缓存时间位置作为上一时序对应帧的字幕区域的时间间隔，并将时间间隔与上一时序对应帧的字幕区域的字幕文字关联保存。本实施例采用现有的音频分离工具将视频中的音频提取出来并保存，并将字幕文字与切割后的音频文件一一对应完成音频标注，上述标注数据可用于自动语音识别技术的训练时的样本数据，以降低现有人工标注数据的成本，且提高标注数据的质量。

进一步地，所述提取模块5，还包括：

第四判断单元，用于根据所述视频的视频文件名和所述第二帧图片的帧索引，判断所述预设文件中是否存在与第一帧索引对应的空字幕文件，其中所述第一帧索引包含于所述预设文件中所有帧索引中；

输入单元，用于所述预设文件中存在与第一帧索引对应的空字幕文件，则将所述第一帧索引对应的帧图片，输入文本定位模型；

第五判断单元，用于判断根据所述文本定位模型是否提取到所述第一帧索

引对应的帧图片的指定字幕文字；

补充单元,用于若根据所述文本定位模型提取到所述第一帧索引对应的帧图片的指定字幕文字,则将所述指定字幕文字补充到所述预设文件中所述第一帧索引对应位置；

- 5 标记单元,用于若根据所述文本定位模型未提取到所述第一帧索引对应的帧图片的指定字幕文字,则在所述预设文件中所述第一帧索引对应位置标记为空字幕。

本实施例通过帧索引与字幕文字的对应关系,查找遗漏提取的字幕文字,以保证整个视频文件中的字幕文字的完整性。当查找到第一帧索引对应空字幕文件,即无对应的字幕文字与其相对应,则判定存在遗漏提取,则将第一帧索引对应的帧图片,输入文本定位模型,以根据文本定位模型进行字幕文字定位与提取。上述文本定位模型为CTPN,CTPN结合CNN与LSTM深度网络,CTPN是从Faster R-CNN改进而来,能有效的检测出复杂场景的横向分布的文字,比如识别视频图片中的文字,虽然识别过程复杂、识别效率较低,但识别精度
10 高,可有效补充通过混合高斯模型算法遗漏的字幕区域,提高整个视频文件中的字幕文字的完整性。

参照图3,本申请实施例中还提供一种计算机设备,该计算机设备可以是服务器,其内部结构可以如图3所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中,该计算机设计的处理器用于提供
20 计算和控制能力。该计算机设备的存储器包括可读存储介质、内存储器。该可读存储介质存储有操作系统、计算机可读指令和数据库。上述可读存储介质包括非易失性可读存储介质和易失性可读存储介质。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储提取视频字幕等数据。该计算机设备的网络接口用于与外部的终端通过
25 网络连接通信。该计算机可读指令在执行时,执行如上述各方法的实施例的流程。本领域技术人员可以理解,图3中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定。

本申请一实施例还提供一种计算机可读存储介质,其上存储有计算机可读指令,该计算机可读指令在执行时,执行如上述各方法的实施例的流程。上述
30 可读存储介质包括非易失性可读存储介质和易失性可读存储介质。以上所述仅为本申请的优选实施例,并非因此限制本申请的专利范围,凡是利用本申请说

说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本申请的专利保护范围内。

权利要求书

1、一种提取视频字幕的方法，其特征在于，包括：

通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域，其中所述第一帧图片和所述第二帧图片是所述视频中相邻的任意两
5 帧图片，所述变化像素区域至少包括一个；

判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域，其中所述第一变化像素区域包含于所述变化像素区域；

若所述视频显示界面的预设区域范围内存在所述第一变化像素区域，则判断所述第一变化像素区域是否满足预设字幕区特征；

10 若所述第一变化像素区域满足预设字幕区特征，则判定所述第一变化像素区域为所述字幕区域；

从所述字幕区域中提取字幕文字。

2、根据权利要求 1 所述的提取视频字幕的方法，其特征在于，所述判断所述第一变化像素区域是否满足预设字幕区特征的步骤，包括：

15 计算所述第一变化像素区域的宽高比，其中所述第一变化像素区域中沿所述视频时序播放方向为所述宽，垂直于所述宽的方向为所述高，所述宽高比为所述宽除以所述高；

判断所述宽高比是否大于预设阈值；

20 若所述宽高比大于预设阈值，则判定所述第一变化像素区域满足所述预设字幕区特征；

若所述宽高比不大于预设阈值，则判定所述第一变化像素区域不满足所述预设字幕区特征。

25 3、根据权利要求 2 所述的提取视频字幕的方法，其特征在于，所述判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域的步骤之前，包括：

获取所述视频的视频宽和视频高，其中所述视频显示界面中沿所述视频时序播放方向为所述视频宽，垂直于所述视频宽的方向为所述视频高；

设定所述预设值等于第一预设值，设定靠近所述视频显示界面的底部边缘，占比第二预设值的区域范围为所述预设区域范围。

30 4、根据权利要求 1 所述的提取视频字幕的方法，其特征在于，所述从所述字幕区域中提取字幕文字的步骤，包括：

将所述字幕区域从所述第二帧图片中切割分离；

将分离后的所述字幕区域通过图像文字识别算法识别出所述字幕文字；

将所述字幕文字复制到预设文件中；

通过预设格式标注所述字幕文字并存储。

- 5 5、根据权利要求4所述的提取视频字幕的方法，其特征在于，所述预设格式至少包括所述视频的视频文件名和所述第二帧图片对应的帧索引，所述通过预设格式标注所述字幕文字并存储的步骤之后，包括：

10 根据所述视频的视频文件名和所述第二帧图片对应的帧索引，判断所述预设文件中是否存在与第一字幕文字具有相同标注信息的第二字幕文字，其中所述第一字幕文字和所述第二字幕文字，分别包含于所述预设文件中所有所述字幕文字中；

若所述预设文件中存在与第一字幕文字具有相同标注信息的第二字幕文字，则判断所述第一字幕文字和所述第二字幕文字的文字内容是否相同；

15 若所述第一字幕文字和所述第二字幕文字的文字内容相同，则删除所述所述第一字幕文字或所述第二字幕文字。

6、根据权利要求4所述的提取视频字幕的方法，其特征在于，所述通过预设格式标注所述字幕文字并存储的步骤之后，还包括：

20 根据所述视频的视频文件名和所述第二帧图片的帧索引，判断所述预设文件中是否存在与第一帧索引对应的空字幕文件，其中所述第一帧索引包含于所述预设文件中所有帧索引中；

若存在，则将所述第一帧索引对应的帧图片，输入文本定位模型；

判断根据所述文本定位模型是否提取到所述第一帧索引对应的帧图片的指定字幕文字；

25 若提取到，则将所述指定字幕文字补充到所述预设文件中所述第一帧索引对应位置；

若未提取到，则在所述预设文件中所述第一帧索引对应位置标记为空字幕。

7、根据权利要求1所述的提取视频字幕的方法，其特征在于，所述从所述字幕区域中提取字幕文字的步骤之后，包括：

获取所述第二字幕文字的起始时间和终止时间；

30 确定所述起始时间和终止时间范围内对应的第一音频文件；

将所述第一音频文件通过音频截取工具从所述视频对应的音频文件中截

取分离；

将所述第二字幕文字与所述第一音频文件一一对应进行音频标注。

8、一种提取视频字幕的装置，其特征在于，包括：

5 第一获取模块，用于通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域，其中所述第一帧图片和所述第二帧图片是所述视频中相邻的任意两帧图片，所述变化像素区域至少包括一个；

第一判断模块，用于判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域，其中所述第一变化像素区域包含于所述变化像素区域；

10 第二判断模块，用于若所述视频显示界面的预设区域范围内存在第一变化像素区域，则判断所述第一变化像素区域是否满足预设字幕区特征；

判定模块，用于若所述第一变化像素区域满足预设字幕区特征，则判定所述第一变化像素区域为所述字幕区域；

提取模块，用于从所述字幕区域中提取字幕文字。

15 9、根据权利要求8所述的迁徙数据的装置，其特征在于，所述第二判断模块，包括：

计算单元，用于计算所述第一变化像素区域的宽高比，其中所述第一变化像素区域中沿所述视频时序播放方向为所述宽，垂直于所述宽的方向为所述高，所述宽高比为所述宽除以所述高；

第一判断单元，用于判断所述宽高比是否大于预设阈值；

20 第一判定单元，用于若所述宽高比大于预设阈值，则判定所述第一变化像素区域满足所述预设字幕区特征；

第二判定单元，用于若所述宽高比不大于预设阈值，则判定所述第一变化像素区域不满足所述预设字幕区特征。

10、根据权利要求9所述的迁徙数据的装置，其特征在于，包括：

25 第二获取模块，用于获取所述视频的视频宽和视频高，其中所述视频显示界面中沿所述视频时序播放方向为所述视频宽，垂直于所述视频宽的方向为所述视频高；

设定模块，用于设定所述预设值等于第一预设值，设定靠近所述视频显示界面的底部边缘，占比第二预设值的区域范围为所述预设区域范围。

30 11、根据权利要求8所述的迁徙数据的装置，其特征在于，所述提取模块，包括：

分离单元, 用于将所述字幕区域从所述第二帧图片中切割分离;

识别单元, 用于将分离后的所述字幕区域通过图像文字识别算法识别出所述字幕文字;

赋值单元, 用于将所述字幕文字复制到预设文件中;

5 标注单元, 用于通过预设格式标注所述字幕文字并存储。

12、根据权利要求 11 所述的迁徙数据的装置, 其特征在于, 所述预设格式至少包括所述视频的视频文件名和第二帧图片对应的帧索引, 所述提取模块, 包括:

10 第二判断单元, 用于根据所述视频的视频文件名和所述第二帧图片对应的帧索引, 判断所述预设文件中是否存在与第一字幕文字具有相同标注信息的第二字幕文字, 其中所述第一字幕文字和所述第二字幕文字, 分别包含于所述预设文件中所有所述字幕文字中;

15 第三判断单元, 用于若所述预设文件中存在与第一字幕文字具有相同标注信息的第二字幕文字, 则判断所述第一字幕文字和所述第二字幕文字的文字内容是否相同;

删除单元, 用于若所述第一字幕文字和所述第二字幕文字的文字内容相同, 则删除所述所述第一字幕文字或所述第二字幕文字。

13、根据权利要求 11 所述的迁徙数据的装置, 其特征在于, 所述提取模块, 包括:

20 第四判断单元, 用于根据所述视频的视频文件名和所述第二帧图片的帧索引, 判断所述预设文件中是否存在与第一帧索引对应的空字幕文件, 其中所述第一帧索引包含于所述预设文件中所有帧索引中;

输入单元, 用于所述预设文件中存在与第一帧索引对应的空字幕文件, 则将所述第一帧索引对应的帧图片, 输入文本定位模型;

25 第五判断单元, 用于判断根据所述文本定位模型是否提取到所述第一帧索引对应的帧图片的指定字幕文字;

补充单元, 用于若根据所述文本定位模型提取到所述第一帧索引对应的帧图片的指定字幕文字, 则将所述指定字幕文字补充到所述预设文件中所述第一帧索引对应位置;

30 标记单元, 用于若根据所述文本定位模型未提取到所述第一帧索引对应的帧图片的指定字幕文字, 则在所述预设文件中所述第一帧索引对应位置标记为

空字幕。

14、根据权利要求 8 所述的迁徙数据的装置，其特征在于，包括：

第三获取模块，用于获取所述第二字幕文字的起始时间和终止时间；

确定模块，用于确定所述起始时间和终止时间范围内对应的第一音频文件；

5 截取模块，用于将所述第一音频文件通过音频截取工具从所述视频对应的音频文件中截取分离；

标注模块，用于将所述第二字幕文字与所述第一音频文件一一对应进行音频标注。

15、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现提取视频字幕的方法，提取视频字幕的方法，包括：

通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域，其中所述第一帧图片和所述第二帧图片是所述视频中相邻的任意两帧图片，所述变化像素区域至少包括一个；

15 判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域，其中所述第一变化像素区域包含于所述变化像素区域；

若所述视频显示界面的预设区域范围内存在所述第一变化像素区域，则判断所述第一变化像素区域是否满足预设字幕区特征；

20 若所述第一变化像素区域满足预设字幕区特征，则判定所述第一变化像素区域为所述字幕区域；

从所述字幕区域中提取字幕文字。

16、根据权利要求 15 所述的计算机设备，其特征在于，所述判断所述第一变化像素区域是否满足预设字幕区特征的步骤，包括：

25 计算所述第一变化像素区域的宽高比，其中所述第一变化像素区域中沿所述视频时序播放方向为所述宽，垂直于所述宽的方向为所述高，所述宽高比为所述宽除以所述高；

判断所述宽高比是否大于预设阈值；

若所述宽高比大于预设阈值，则判定所述第一变化像素区域满足所述预设字幕区特征；

30 若所述宽高比不大于预设阈值，则判定所述第一变化像素区域不满足所述预设字幕区特征。

17、根据权利要求 16 所述的计算机设备，其特征在于，所述判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域的步骤之前，包括：

获取所述视频的视频宽和视频高，其中所述视频显示界面中沿所述视频时序播放方向为所述视频宽，垂直于所述视频宽的方向为所述视频高；

5 设定所述预设值等于第一预设值，设定靠近所述视频显示界面的底部边缘，占比第二预设值的区域范围为所述预设区域范围。

18、一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现提取视频字幕的方法，提取视频字幕的方法包括：

10 通过混合高斯模型算法获取视频的第二帧图片相比于第一帧图片的变化像素区域，其中所述第一帧图片和所述第二帧图片是所述视频中相邻的任意两帧图片，所述变化像素区域至少包括一个；

判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域，其中所述第一变化像素区域包含于所述变化像素区域；

15 若所述视频显示界面的预设区域范围内存在所述第一变化像素区域，则判断所述第一变化像素区域是否满足预设字幕区特征；

若所述第一变化像素区域满足预设字幕区特征，则判定所述第一变化像素区域为所述字幕区域；

从所述字幕区域中提取字幕文字。

20 19、根据权利要求 18 所述的计算机可读存储介质，其特征在于，所述判断所述第一变化像素区域是否满足预设字幕区特征的步骤，包括：

计算所述第一变化像素区域的宽高比，其中所述第一变化像素区域中沿所述视频时序播放方向为所述宽，垂直于所述宽的方向为所述高，所述宽高比为所述宽除以所述高；

25 判断所述宽高比是否大于预设阈值；

若所述宽高比大于预设阈值，则判定所述第一变化像素区域满足所述预设字幕区特征；

若所述宽高比不大于预设阈值，则判定所述第一变化像素区域不满足所述预设字幕区特征。

30 20、根据权利要求 19 所述的计算机可读存储介质，其特征在于，所述判断所述视频显示界面的预设区域范围内是否存在第一变化像素区域的步骤之

前，包括：

获取所述视频的视频宽和视频高，其中所述视频显示界面中沿所述视频时序播放方向为所述视频宽，垂直于所述视频宽的方向为所述视频高；

- 5 设定所述预设值等于第一预设值，设定靠近所述视频显示界面的底部边缘，占比第二预设值的区域范围为所述预设区域范围。

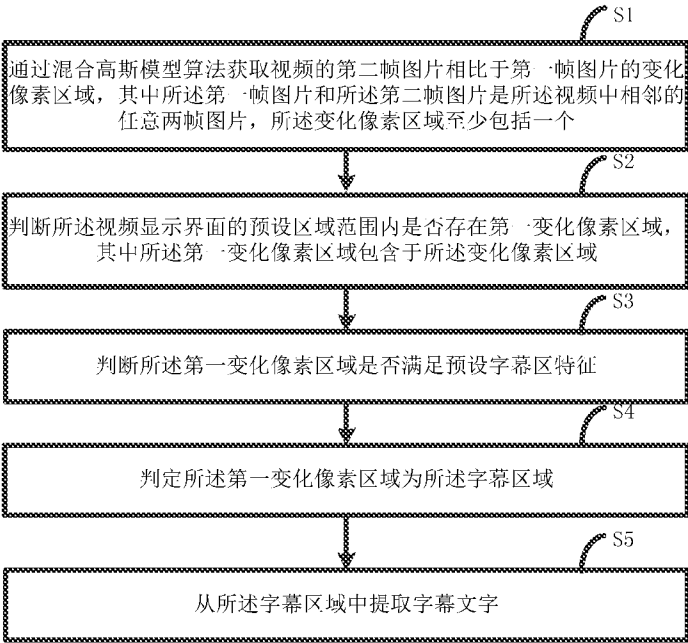


图 1

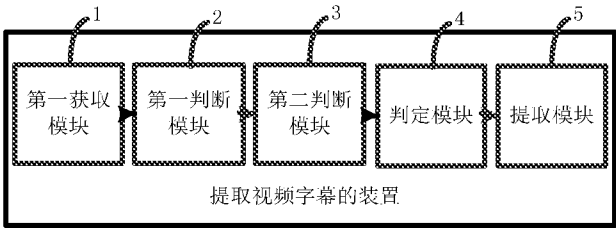


图 2

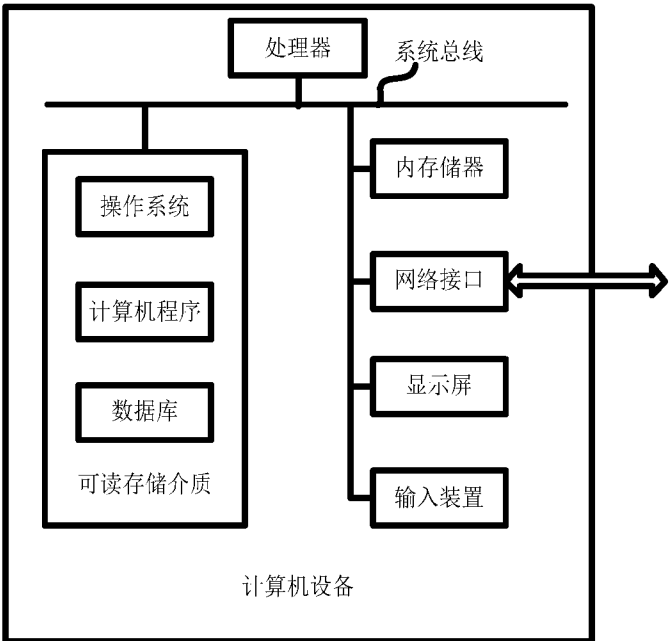


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/118411

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/32(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K; G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT; CNKI; WPI; EPODOC: 视频, 图像, 图象, 图片, 文字, 字幕, 文本, 提取, 检测, 定位, 分离, 识别, 帧, 变化, 改变, 差, video, image, picture, text, subtitle, caption, detect+, extract+, identif+, separat+, change, difference

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110197177 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 03 September 2019 (2019-09-03) description, paragraphs [0052]-[0160]	1-20
Y	CN 102208023 A (ZHEJIANG UNIVERSITY) 05 October 2011 (2011-10-05) description, paragraphs [0016] and [0017]	1-20
Y	CN 102802074 A (HISENSE GROUP CO., LTD.) 28 November 2012 (2012-11-28) description, paragraphs [0085]-[0126]	1-20
A	US 2016360123 A1 (SONIC IP, INC.) 08 December 2016 (2016-12-08) entire document	1-20

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

10 January 2020

Date of mailing of the international search report

06 February 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/118411

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	110197177	A	03 September 2019	None	
CN	102208023	A	05 October 2011	None	
CN	102802074	A	28 November 2012	None	
US	2016360123	A1	08 December 2016	None	

国际检索报告

国际申请号

PCT/CN2019/118411

A. 主题的分类

G06K 9/32 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06K; G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNPAT; CNKI; WPI; EPDOC: 视频, 图像, 图象, 图片, 文字, 字幕, 文本, 提取, 检测, 定位, 分离, 识别, 帧, 变化, 改变, 差, video, image, picture, text, subtitle, caption, detect+, extract+, identif+, separat+, change, difference

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 110197177 A (平安科技深圳有限公司) 2019年 9月 3日 (2019 - 09 - 03) 说明书第[0052]-[0160]段	1-20
Y	CN 102208023 A (浙江大学) 2011年 10月 5日 (2011 - 10 - 05) 说明书第[0016]-[0017]段	1-20
Y	CN 102802074 A (海信集团有限公司) 2012年 11月 28日 (2012 - 11 - 28) 说明书第[0085]-[0126]段	1-20
A	US 2016360123 A1 (SONIC IP, INC.) 2016年 12月 8日 (2016 - 12 - 08) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 1月 10日

国际检索报告邮寄日期

2020年 2月 6日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

陈希元

电话号码 86-(10)-53961594

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/118411

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110197177	A	2019年 9月 3日	无	
CN	102208023	A	2011年 10月 5日	无	
CN	102802074	A	2012年 11月 28日	无	
US	2016360123	A1	2016年 12月 8日	无	