Software Engineering and Applications 软件工程与应用, 2016, 5(2), 114-121
Published Online April 2016 in Hans. http://www.hanspub.org/journal/sea
http://dx.doi.org/10.12677/sea.2016.52013

Hans汉斯

Video Key Frame Recognition Technology Research Based on the Content

Qian Liu, Gui'e Luo, Xianru Liu*

Central South University, Changsha Hunan
Email: *748799183@qq.com

Received: Mar. 17th, 2016; accepted: Apr. 2nd, 2016; published: Apr. 5th, 2016

Abstract

The key frame extraction is vital for automatic segmentation of video; the standard of extracted key frame directly affects the final result. Based on the analysis and research about project video, the paper proposes template matching, histogram comparison, the key frame detection and key frame recognition algorithm, and for the recognition accuracy and recognition rate of template matching and histogram comparison, the paper also does the comparison. In the recognition, the key frames adopted a new character segmentation algorithm. Finally, the paper summarizes the advantages of this method and good results have been achieved.

Keywords

Key Frame, Template Matching, Character Segmentation

基于内容的项目视频关键帧识别技术研究

刘　倩，罗桂娥，刘献如*

中南大学，湖南　长沙
Email: *748799183@qq.com

收稿日期：2016年3月17日；录用日期：2016年4月2日；发布日期：2016年4月5日

摘　要

关键帧的提取对于视频的自动切分至关重要，提取的好坏直接影响最终的结果，本文通过对项目视频的

*通讯作者。

文章引用: 刘倩, 罗桂娥, 刘献如. 基于内容的项目视频关键帧识别技术研究[J]. 软件工程与应用, 2016, 5(2): 114-121.
http://dx.doi.org/10.12677/sea.2016.52013

分析研究，提出了模板匹配，直方图比较，关键帧检测与关键帧识别的算法，并比较了模板匹配与直方图对比的识别准确率和识别速率，其中在关键帧识别中，采用了新的字符分割的算法。最后，总结了本文方法的优越性，并且取得了很好的效果。

关键词

关键帧，模板匹配，字符分割

1. 引言

随着计算机技术、多媒体技术[1]和网络技术的不断发展，图像和视频资源[2]日益丰富，从这些海量图像、视频中获取感兴趣[3]的信息已经成为当前多媒体信息技术研究的热点。为加快国家科技管理信息系统建设，各类计划等实现全流程在线信息化管理[4]；实现全过程痕迹管理，做到“可申诉、可查询、可追溯”[5]。在研究、借鉴网易视频、百度网盘、微软 skydrive 等大数据[6]、大文件视频等管理模式基础上，结合国家科技信息管理的新要求以及信息中心自身建设的特点和要求，科技部提出了对各类视频数据进行有效管理的要求，具体要求实现对相关视频文件按需求进行切分、标注、归档、查询和服务等功能。目前，这一项工作主要由工作人员通过利用相关软件手动拖动视频进行观察、估计、切分等工作，这不仅效率低下且造成人力资源的浪费。针对此问题，本项目拟设计并开发一个基于内容的视频切分系统[7]，该系统可根据需求能对一种或多路高清视频进行自动解码、识别关键帧、切分等多项功能。在目前的条件下本文针对特定的视频类型来设计特定的、既简单又实用的视频提取算法，下面本文就介绍一种针对项目视频的识别关键帧而设计的一种关键帧提取算法。

通过反复观察此项目视频之后，发现在视频中，如果出现时间表的文字时，一定在视频帧的右上角，这些视频帧最大限度地反映了报告人作报告所用的时间，具有极强的代表性，正是我们要提取的关键帧. 因此，在对项目视频进行分析和对项目视频的关键帧进行提取时，考虑到了时间表信息的利用。为了对关键帧进行识别，主要分为两个步骤，分别是关键帧检测和关键帧识别，本文分别对这两个内容进行了描述。

2. 关键帧检测

关键帧提取[8]的方法有很多种，目前比较典型的有基于镜头的方法，帧平均法和直方图平均法[9]，基于运动的分析法方法[10]，基于聚类的方法[11]。其中基于镜头的方法主要是将镜头检测中得到的镜头中的首帧(或尾帧)作为镜头关键帧。帧平均法是从镜头中取所有帧在某个位置上像素值的平均值，然后将镜头中该点位置的像素值最接近平均值的帧作为关键帧；直方图平均法则是将镜头中所有帧的统计直方图取平均，然后选择与该平均直方图最接近的帧作为关键帧。基于运动的分析法方法通过光流分析来计算镜头中的运动量，在运动量局部最小值处选取关键帧，它反映了视频数据的静止。基于聚类的方法[12]可以通过对视频帧进行聚类来选取关键帧。算法主要分为三个步骤，首先是特征提取阶段，这里的特征主要是帧间直方图的差别，第一阶段提取的特征作为第二阶段的输入进行聚类，第三阶段即是关键帧的选取。

本文采用模板匹配和直方图比较两种方法，通过对项目视频进行大量的分析发现，不包含时间表帧和包含时间表帧的区别的最为行之有效的方法就是提取时间表本身的特征，包含时间表的帧和不包含时间表的帧空间结构模型是固定的，如图 1 所示为视频帧的空间结构模型。

2.1. 模板匹配法

由于本项目需要实现大量视频的自动切分，各类视频数据量非常多，实现视频的自动切分不仅识

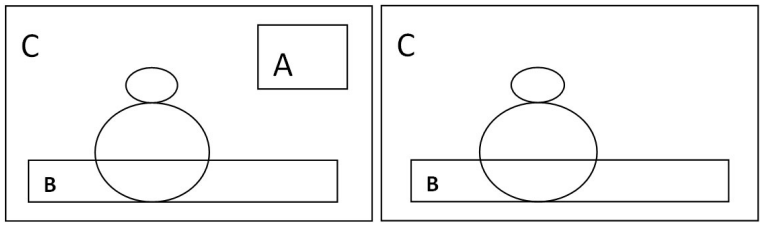

Figure 1. The spatial structure model of video frames
图 1. 视频帧的空间结构模型

别的准确率要高，而且识别的时间也必须提高。因此本文采用了改进的模板匹配[13]方法不仅能够提高识别的准确率，而且也提高了识别的效率。模板匹配方法是通过在输入图像上滑动图像块对实际的图像块和输入图像进行匹配，算法的检测速度比较快，可以对视频流进行实时检测，通过多次分析统计项目视频中的时间表，逐步对算法进行改进，并利用改进后的算法对项目视频中固定区域(位于屏幕右上方十二分之一范围内)的图片进行了检测，下面对优化后的基于帧图像序列模板提取的模板匹配方法做简要介绍。

根据图 1 所示空间结构可知，时间表出现的位置是一定的，因此要设定一个局部区域来进行检测，为了提高检测的准确性，局部区域的选择要满足两个条件：

1) 位置要准，所选的区域一定是时间表可能出现的区域；

2) 范围不能大，否则直接导致检测的时间会延长，而且准确率也会下降。

基于上述条件，选择了图 1 空间结构图所示的“局部 A”区域。图中所示的“局部 A”区域相对帧于整帧图像的尺寸如式(2-1)所示：

$$w = a * W,\ h = b * H \tag{2-1}$$

其中，H——整帧图像的高度；W——整帧图像的宽度；h——局部区域的高度；w——局部区域的宽度；a，b——尺寸系数。

对于不同项目视频，尺寸系数略有差异，可以根据具体情况来调整尺寸系数的大小。对于本文中所采用的项目视频，本文用 $a = 0.25$，$b = 1/3$。

得到了视频中的区域，算法开始用模板匹配的方法进行检测，多次分析视频数据，提取时间表本身作为固定的模板对项目视频进行检测。具体实现过程为：首先，读取视频，在视频帧序列中选取一幅有时间帧的图像，根据上述区域分割的方法提取出时间表，作为模板匹配方法的模板，然后连续读入视频序列，对视频序列中所有帧进行固定区域检测，将提取出的模板与视频序列中所有的帧进行模板匹配，检测视频帧中是否包含时间表帧，最后，将检测的结果进行输出并将包含时间表的帧提取出来。时间表的具体提取流程图如图 2 所示。

2.2. 直方图比较法

直方图[14]中的数值都是统计而来，描述了该图像中关于颜色的数量特征，可以反映图像颜色的统计分布和基本色调。本文采用了直方图比较的方法实现两幅图像的对比。要比较两个直方图(H1 和 H2)，首先必须要选择一个衡量直方图相似度的对比标准 $d(H_1, H_2)$。Opencv 中提供了函数 cvCompareHist()用于对比两个直方图的相似性。该函数提供了四种比较标准来计算直方图的相似性。

a) 相关(CV_COMP_CORREL)，如式(2-2)所示。

$$d_{correl}(H_1, H_2) = \frac{\sum_i H_1(i) \cdot H_2(i)}{\sqrt{\sum_i H_1'^2(i) \cdot H_2'^2(i)}} \tag{2-2}$$

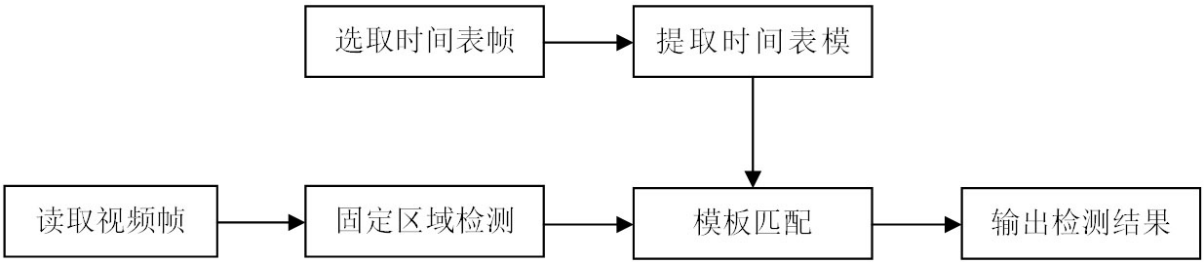

Figure 2. The extraction of schedule steps
图 2. 时间表的提取步骤

其中，$H_k'(i)=H_k(I)-(1/N)\left(\sum_i H_k(j)\right)$ 且 N 等于直方图中 bin 的个数。

b) 卡方(CV_COMP_CHISQR)，如式(2-3)所示。

$$d_{\text{chi-square}}(H_1,H_2)=\sum_i\frac{\left(H_1(i)-H_2(i)\right)^2}{H_1(i)+H_2(i)} \tag{2-3}$$

c) 相交(CV_COMP_INTERSECT)，如式(2-4)所示。

$$d_{\text{interection}}(H_1,H_2)=\sum_i\min\left(H_1(i),H_2(i)\right) \tag{2-4}$$

d) Bhattacharyya 距离(CV_COMP_BHATTACHARYYA)，如式(2-5)所示。

$$d_{\text{Bhattacharyya}}(H_1,H_2)=\sqrt{1-\sum_i\frac{\sqrt{H_1(i)\cdot H_2(i)}}{\sum_i H_1(i)\cdot\sum_j H_2(i)}} \tag{2-5}$$

通过大量的实验可以验证，使用 Bhattacharyya 方法的效果最好。对 Bhattacharyya 方法的直方图匹配，低分数表示好匹配，而高分数表示坏的匹配。完全匹配是 0，完全不匹配是 1。本文所采用的直方图具体步骤如下。

1) 计算视频中提取到的模板图片的直方图 hist1。

2) 读取视频帧，截取帧上与模板图片相同大小的位置，并计算其直方图。

3) 采用 Bhattacharyya 直方图比较的方法比较每一帧图片与模板图片的直方图相似性。

4) 统计结果设定出阈值 T，通过对所有项目视频做大量的实验分析，可以将本项目视频的阈值 T 设定为 0.6，将直方图比较结果与阈值 T 进行对比，若小于 T，则包含时间表，否则，此视频帧不包含时间表。

2.3. 实验结果对比

通过上述两种方法，为了验证算法的准确性与高效性，本文选取了多个项目中的视频作为实验对象，其中包括多人做 ppt 时的报告视频。采用识别的准确率与识别效率作为评判标准，模板匹配具体结果如表 1 所示，直方图比较具体结果如表 2 所示。

由表可以看出，模板匹配的方法匹配的准确率较直方图匹配的方法要高一些，且平均识别时间也较少，可以看出本文提出的模板匹配方法的优越性和高效性。

3. 关键帧识别

当提取到时间表后，对时间表进行分析观察可知，时间表具有以下几个特征：

1) 所有的时间表都具有规则的形状，通常出现在帧右上角 1/12 的位置；

2) 时间表中字符的大小通常为高占 16 个像素，宽占 30 个像素，字符间有一定的间隙。

Table 1. Template matching method performance metrics
表 1. 模板匹配方法性能指标

项目视频	总帧数	检测出时间表数	实际时间表数	查全率	查准率	识别时间
视频 1	263	157	157	100%	100%	42.19
视频 2	1020	883	884	98%	99%	43.04
视频 3	1283	785	783	100%	98%	43.88

Table 2. Histogram comparison method performance metrics
表 2. 直方图比较方法性能指标

项目视频	总帧数	检测出时间表数	实际时间表数	查全率	查准率	识别时间(ms)
视频 1	263	106	106	100%	100%	59.38
视频 2	1020	873	884	96%	96%	62.57
视频 3	1283	768	783	98%	97%	64.09

为了对视频实现准确的切分，必须找到视频开始与结束的关键帧，本文中即找到开始时间与结束时间，所以为了对时间表进行识别，必须把时间表中的字符一个一个的分割出来，然后在对字符识别。本文采用的方法为先对时间表图像进行灰度化，然后再用特定的方法分割字符，最后把分割好的字符提取出来。本文采用重置像素的方法，对时间表图像直接分割。

3.1. 字符分割

字符分割是一种将一行包含文字的图像分割为单个字符图像的技术。由于本文处理的时间表结构完整间隔明显，因此，可以依靠先验知识，对字符进行分割，字符的固定模式为：N_1N_2:N_3N_4 其中 N_1,N_2,N_3,N_4 分别为 0 到 9 的数字，“:”为固定模式不需要识别。如图 3 为在视频中提取出的时间表。

字符分割的算法有很多种，本文采用对图片重置像素的方法，如图 3 所示，时间表中的字符为固定的模式，因此对时间表中的每一个数字所在的位置进行就能够把每一个字符都分割出来。设 width 为图 3 中时间表的宽度，则每一个字符分割的结果如图 4 所示。

其中，当宽度大于 width/4+10 时，将其像素值设为 255，分割出第一个字符(a)。

当宽度小于 width/4 且大于 width/2-6 时，将其像素值设为 255，分割出第二个字符(b)。

当宽度小于 width/2 + 10 且大于(width/4)∗3 + 6 时，将其像素值设为 255，分割出第三个字符(c)。

当宽度小于(width/4)∗3+6 时，将其像素值设为 255，分割出第四个字符(d)。

3.2. 字符识别

本文运用了模板匹配的方法对字符进行识别，模板匹配方法的基本原理：模板匹配方法是实现离散输入模式分类的有效途径之一，其实质是度量输入模式与样本之间的某种相似性，取相似性最大者为输入模式所属类别。它根据字符的直观形象抽取特征，采用归一化相关匹配法进行识别，即是将输入字符与标准字符在一个分类器中进行匹配，若 I 表示图像，T——模板，R——结果，Z——归一化结果，x、y 表示图像中的某个点，相关匹配法是采用模板和图像之间的乘法操作，所以匹配结果中较大的数表示匹配程度较高，0 表示最坏的匹配效果。如式(3-1)，式(3-2)，式(3-3)所示。

$$R_{\text{ccorr}}(x,y)=\sum_{x',y'}\left[T(x',y')\cdot I(x+x',y+y')\right]^2 \tag{3-1}$$

Figure 3. The timetable extracted from video
图 3. 视频中提取出的时间表

(a) (b) (c) (d)

Figure 4. Character of the segmentation
图 4. 分割出的字符

$$T(x',y')=T(x',y')-\frac{1}{(\omega\cdot h)\sum_{x'',y''}T(x'',y'')} \tag{3-2}$$

$$I'(x+x',y+y')=I(x+x',y+y')-\frac{1}{(\omega\cdot h)\sum_{x'',y''}T(x'',y'')} \tag{3-3}$$

对于一种匹配方法有其归一化的形式，根据 Rodgers[Rodgers88]的描述，归一化最早由 Galton[Galton]提出，此方法可以减少模板和图像上光线变化所产生的影响，归一化系数如式(3-4)所示：

$$Z(x,y)=\sqrt{\sum_{x',y'}T(x',y')^2\cdot\sum_{x',y'}I(x+x',y+y')^2} \tag{3-4}$$

其中，则T为模板图像，I为原图，x、y表示图像中的某个点，x',y'为模板图像上的某个点，由以上内容可知，归一化的相关匹配法计算结果为式(3-5)。

$$R_{\text{ccorr\_normed}}(x,y)=\frac{R_{\text{ccorr}}(x,y)}{Z(x,y)} \tag{3-5}$$

其中，$R_{\text{ccorr}}(x,y)$为模板图像与原图像的相关性值，$Z(x,y)$为归一化值，$R_{\text{ccorr_normed}}(x,y)$表示归一化相关匹配结果。

本文采用的字符识别的算法具体步骤如图 5 所示。

根据上述所介绍的本文所采用的模板匹配的字符识别方法，为使得模板与待识别的字符具有最大的相似性，本文所建造了 10 个模板。为使它们与待识别的字符不管在字体还是笔画的粗细上都具有极大的相似性，本文使用 windows 自带的软件画图中通过反复的实验并调用 opencv 库得到数字模板，且所有模板的尺寸大小同样为 62 × 85 像素。

4. 实验结果及分析

为了验证算法的准确性，我们选取了多个项目中的视频作为实验对象，其中包括多人做 ppt 时的报告视频。对于关键帧检测的结果如图 6 所示。

检测出包含时间表的帧之后，再对时间表帧进行识别，根据上述方法之后，识别结果如图 7 所示。

由图可知每一帧中所显示时间是多少，如此确定了起始时间和终止时间就能在视频中找到符合要求的帧并提取出来，完成了预先想要的关键帧提取目的。

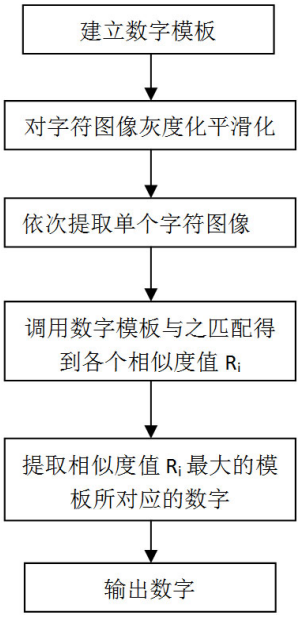

Figure 5. Character recognition algorithm steps
图 5. 字符识别的算法步骤

Figure 6. The frame containing a timetable by detected
图 6. 检测到的包含时间表的帧

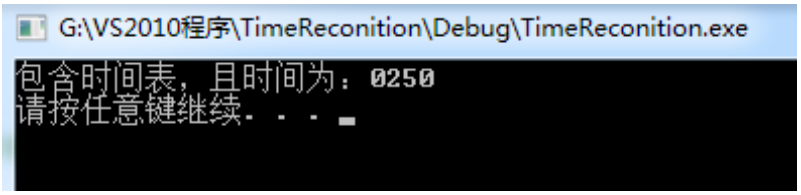

Figure 7. The identifying time
图 7. 识别出的时间

5. 结论

本文提出了一种对项目视频的时间表识别并提取的方法，该方法是基于模板匹配的方法来进行关键帧检测与识别，并且通过提取局部区域来进行优化加速，极大地减少了关键帧提取的时间，提高了效率。

参考文献 (References)

[1] 孙君顶. 基于内容的图像检索技术研究[D]: [博士学位论文]. 西安: 西安电子科技大学, 2005.

[2] Hu, W. and Xie, N.H. (2011) A Survey on Visual Content-Based Video Indexing and Retrieval. *IEEE Transactions on Systems, Man, and Cybernetics—Part C*: *Applications and Reviews*, **41**, 797-819. http://dx.doi.org/10.1109/TSMCC.2011.2109710

[3] 冈萨雷斯. 数字图像处理[M]. 北京: 电子工业出版社, 2005.

[4] 章毓晋. 图像处理和分析(图像工程上册) [M]. 北京: 清华大学出版社, 2004.

[5] 马永波. 基于内容的视频分割与检索技术研究[D]: [硕士学位论文]. 长春: 长春工业大学, 2007.

[6] 黎洪松. 数字视频处理[M]. 北京: 北京邮电大学出版社, 2006.

[7] Narasimha, R., Savakis, A. and Rao, R.M. (2004) A Neaural Network Approach to Key Frame Extraction. *Proceedings of SPIE-IS&T Electronic Imaging Storage and retrieval Methods and Applications for Multimedia*, **53**, 439-447.

[8] 刘洋. 基于内容的视频检索关键技术研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2008.

[9] 侯海珍. 基于内容的视频检索方法研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2009.

[10] 滑勇. 基于关键帧的视频内容检索问题的研究[D]: [硕士学位论文]. 大连: 大连理工大学, 2005: 18-20.

[11] 陆伟艳, 夏定元, 刘毅. 基于内容的视频检索的关键帧提取[J]. 微计算机信息, 2007, 23(33): 298-300.

[12] 周政, 刘俊义, 马林华, 等. 视频内容分析技术[J]. 计算机工程与设计, 2008, 29(7): 1766-1769.

[13] 季春. 基于内容的视频检索中的关键帧提取技术[J]. 情报杂志, 2006(11): 116-119.

[14] Liu, T.M. and Zhang, M. (2003) A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model. *IEEE Transactions on Circuits and Systems for Video Technology*, **13**, 1006-1013. http://dx.doi.org/10.1109/TCSVT.2003.816521