文章编号:1671-4598(2015)05-1754-04 DOI:10.16526/j.cnki.11—4762/tp.2015.05.091 中图分类号:TP3 文献标识码:A

视频文字信息检查工具的设计与实现

文 毅[1],龚 飞[2],党静雅[2],邢更力[3]
(1.中国航天科工二院,北京 100854;
2.北京航天测控技术有限公司,北京 100041;
3.公安部第一研究所,北京 100048)

摘要:针对某些特定视频中,画面文字信息经常包含较为敏感文字信息,导致信息泄露,设计实现了一种视频画面中的文字的检测识别系统,对视频画面中的文字标语、文字条幅,新闻画面中的文字导语等信息进行识别与比对;采用基于双阈值的视频镜头分割算法,根据颜色直方图信息提取关键帧,采用最大稳定极值区域算法提取图像中稳定区域,通过聚类和级联分类器实现文字区域提取,最后将文字区域分割后进行OCR识别,实验表明系统针对复杂背景中的文字能够达到较高的检测识别率。

关键词:双阈值;颜色直方图;最大稳定极值区域;均值聚类;分类器

Design and Implementation of a Checking Tool for Video Text

Wen Yi[1], Gong Fei[2], Dang Jingya[2], Xing Gengli[3]
(1. The Second Academy of China Aerospace, Beijing 100854, China;
2. Beijing Aerospace Measurement and Control Technology Co., LTD., Beijing 100041, China;
3. First Research Institute of The Ministry of Public Security of PRC, Beijing 100048, China)

Abstract: This article introduced a text extraction and recognition system for video content, which aimed to the text banners, texts of the news images in the videos. Presents a algorithms based on adaptive dual-threshold which could divide the video into segments by shots. Then the key frames are extracted based on the color histogram information of the frame. Use Maximally Stable Extremal Regions (MSER) for detecting text character candidates in natural scene images. Then a combination of two classifiers is applied to filter non-text meaningful groups. At last, We put the regions of text into OCR system for recognition. The system that design in this paper achieved high recognition rate.

Keywords: dual-threshold; color histogram information; MSER; perceptual organization clustering; classifiers

0 引言

当前社会已经进入信息化时代,伴随着计算机技术,通讯技术,以及多媒体技术飞速发展,以图像、声音和视频为主的多媒体信息迅速成为信息交流与服务的主流,我们身边存在大量的视频和图片信息,如何在这互联网这个庞大的数据库中实现对海量视频信息中文字的快速检索,这是当前人们普遍关注的一个重要课题。

对于视频和图像本身,往往包含了许多文字信息,但是这些文字是以像素的形式存在与视频和图像中,例如新闻视频中的导语以及字幕,视频中的文字标语,文字条幅等,如果能够准确提取视频中的文字信息,将文字信息与内部的关键信息库中的文字进行比对,这样就能够更好的判断这个视频的种类和属性,能够为视频的审查提供重要的依据。

本文设计了一种快速检测视频中文字信息的方法,首先对视频进行关键帧提取,通过对视频的关键帧进行文字检测与识别,提取视频图像中的文字信息,为视频性质的判定提供依据。

1 相关研究背景

当前,互联网上的信息呈几何级的增长,每天互联网上都有数以百万计的视频更新,为了满足对视频信息进行快速的检查的需要,随之开发了很多的关键帧提取算法,大大减少了视频的冗余信息,只需要对视频的关键帧进行分析,就可以获取需要的视频信息,这样就为视频信息的甄别和分析减少了运算量,大大节省了运算时间。

视频关键帧提取作为视频信息提取的基础,对后续的分析起着决定性的作用,目前针对关键帧提取的算法主要有基于镜头边界的方法[1]、基于聚类分析的方法[2]、基于运动矢量分析的方法[3],基于内容分析的方法。

通常从图像中提取文字都需要首先定位包含文字的图像区域,但文字在字体、大小、对齐方式和排列上变化多端,文字背景复杂,图像分辨率低,而且许多应用场合还要求算法具有一定处理速度,这些都使得从图像中有效地提取文字变得非常困难。文字提取主要分为文字检测和文字分割两大部分,针对每一部分国内外的研究者们都进行了大量的研究和实验,设计了一些卓有成效的系统和算法。

提取文字区域的方法主要有以下3种:基于边缘提取,基于纹理提取[4],基于连通区域提取[5]。基于连通区域方法首先对图像进行自适应分割,对字符颜色层提取连通分量.然后提取连通分量的特征,并用分类器过滤非字符连通分量,最后对候选的字符连通分量根据其位置和颜色层进行聚类分析来定位文本区域,这种方法对于背景复杂的区域,以及文字仿射变化具有很高的适应性[6]。

收稿日期:2015-01-09; **修回日期**:2015-03-12。

作者简介:文 毅(1988-),男 湖南益阳人,硕士研究生,主要从事视频特征提取方向的研究。

2　系统设计与实现

2.1　总体设计

视频检查工具需要扫描大量的视频文件和图片，包括各种会议视频，以及宣传报道视频，有的视频时间长，数据量大，同时视频图像的复杂度高，给视频关键帧提取及文字检测带来了很大的挑战，针对被检测视频以及图像的特征，本文有针对性地设计运用了 MSER、聚类、级联分类等算法，围绕速率与效率相结合的目标，解决了视频检查工具设计过程中一些常见的问题。系统总体设计如图 1 所示。

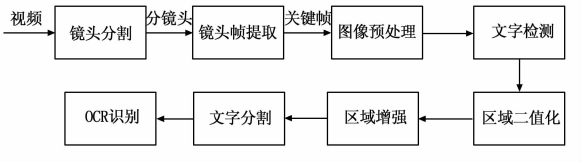

图 1　视频文字提取总体框图

视频检查工具由镜头分割，镜头帧提取、文字检测、区域二值化、文字分割、OCR 识别等部分构成，我们将在下文重点介绍关键帧提取、文字检测、文字识别部分。

2.2　关键帧提取

关键帧提取分为两步完成，首先是镜头检测分割，然后是镜头帧提取。镜头变化检测方法分为两类：镜头突变（切变）检测和镜头渐变检测。镜头的突变即摄像机的停机或者镜头的切换；镜头的渐变即在两个镜头之间加入了编辑效果：淡入淡出，叠画和划像等，使得镜头内的场景是平缓变化而不是急剧变化。

我们研究了镜头变换的各种类型及其表现，发现大部分的镜头检测算法只注重了切变镜头的检测，对于渐变镜头的识别往往做得并不好。通过分析现有的像素域和压缩域镜头检测算法的优劣，采用了一种自适应的基于双阈值的关键帧提取算法，算法采用 HSV 颜色模型，根据颜色直方图特性选取关键帧，选出的关键帧不仅具有代表性，而且在速度方面能够达到大容量视频快速提取的要求。

采用 HSV 颜色模型，将 HSV 特征合成一个新的特征矢量 $L=8H+4S+V$，对于图像帧，其直方图的特征矢量为 $P(p_1,p_2,\cdots,p_L)$，其中 $p_1,p_2,\cdots,p_L$ 为归一化概率值 $0\leqslant p_1,p_2,\cdots,p_L\leqslant 1$。利用直方图相交算法进行图像相似性度量，计算公式如下：

$$Sim(A,B)=\sum_{i=0}^{L}\min(a_i,b_i) \tag{1}$$

$Sim(A,B)$ 越大，则 A,B 两帧的相似度越大[7]。

自适应双阈值法是通过 T_H 和 T_L 两个阈值（其中 $T_H>T_L$）来检测镜头，切变镜头可以直接使用 T_L 检测，而渐变镜头则需要 T_H 和 T_L 共同检测。设第 i 帧和第 $i+1$ 帧的帧间差异为 D_i，计算从首帧开始的累计帧间差异

$$M_n=\sum_{i=1}^{n}D_i \tag{2}$$

当 $D_n\leqslant T_L$ 且 $M_n>T_H$，我们认为找到了渐变镜头的最后一帧 f_n。其中 T_H 和 T_L 由以下公式确定：

$$G=\frac{1}{w}\sum_{i=1}^{w}D_i$$

$$R=\frac{1}{w}\sum_{i=1}^{w}(D_i-G)^2 \tag{3}$$

其中：G 为帧间差异均值，R 为帧间方差；

$T_H=Q_H(G-nR)$，其中 $n=1\sim 3$，$Q_H=0.8\sim 1$；

$T_L=Q_LG$，其中 $Q_L=0.3\sim 0.5$。

根据双阈值法实现了镜头分割，通过设置阈值，我们可以限定镜头检测的门槛，对于较长的镜头，可以根据双阈值的切割方法分成多个子镜头，我们需要在镜头中定位关键帧，传统的方法是取镜头的首帧或者尾帧，但是这种取法往往不具有代表性，容易漏掉镜头中的关键信息[8]，本文提出了一种基于颜色直方图的选取关键帧的方法。

首先计算镜头中所有帧图像平均直方图特征值

$$\overline{P}=\frac{1}{n}\sum_{k=1}^{n}P_K \tag{4}$$

将得到的平均特征矢量与镜头中的每一帧的特征矢量比较，利用直方图相交算法进行图像相似性度量，计算方法如下：

$$Sim(\overline{P},P_K)=\sum_{i=0}^{L}\min(P_{K_i},\overline{P}_i) \tag{5}$$

取与平均帧最相似的帧作为该镜头的关键帧。

2.3　文字区域提取

针对文字的检测，我们提出了一种基于区域的方法，通过一种自底向上的以相似性证据来提取区域，相似性证据包括多种因素如颜色，尺寸，或者笔画宽度等，为了能够得到有意义的类似文字区域，比如说段落、文本行，或者单个字符，我们主要分为以下三步进行文字提取：区域分解、知觉组织分析、线性分类。

区域分解是通过求最大稳定极值区域（MSER）获得疑似文字区域，最大稳定极值区域具备可区分、仿射不变性、稳定等特性，能够很好地解决复杂背景下的文字识别问题。MSER 算法利用边界亮度的梯度特点建立一个极值特性稳定的区域树[9]。这种边界颜色差异特性是所有文本区域共有的特征，通常文本区域都与背景有明显的亮度对比度。最大稳定极值区域树中的一些非文字区域通过一些文字特征，如尺寸，文字纵横比，笔画宽度，区域孔洞等来排除。

对所有获得的最大稳定极值区域采用知觉组织聚类算法，首先通过检查不同的特征子空间生成一定数量的可能的假设区域，然后通过分析保留最有意义区域，最后那些有效聚类通过证据积累结合，使用简单并且计算量较小的特征来描述文字区域与提取的 MSER 区域的相似性，这些特征包括几何特征，如边界区域面积，像素数量，边界周长等[10]；区域中像素的亮度和颜色均值（Lab 域）；边界的亮度和颜色均值；笔画宽度；边界梯度均值等。为了从特征子空间中找出有意义的文字区域，我们使用亥姆霍兹原则来判断，亥姆霍兹原则为自动检测随机性偏差的一个统计方法提供了理论依据。假设 n 为元素总数，分组 G 中有 k 个区域有相同特征，计算以下二项式：

$$B_G(k,n,p)=\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i} \tag{6}$$

P 为单个节点有上述特征的可能性，是组成区域的样本集的特征占整个特征子空间的比值[11]。

我们利用这种衡量标准在每一个特征子空间的区域树中评

估所有分组假设的有效性，在区域合并中，利用上式计算区域树的每一个节点的特征二项式。对于聚类 A 的每个祖先 B，以及每个后继元 C，$B_B(k,n,p) > B_A(k,n,p)$ 且 $B_C(k,n,p) > B_A(k,n,p)$ 成立时，选取 A 为最大均值聚类，这样的话没有区域可以同时属于两个不同的分组。

当我们在每个特征子空间找到最大均值聚类 P^i 其中 $i \in N$，聚类集合 $Q = \{P^1, P^2, P^3, \cdots P^N\}$ 用来计算每组区域属于哪个分组，定义共生矩阵

$$\boldsymbol{D}(i,j) = \frac{m_{ij}}{N} \tag{7}$$

其中 i，j 是指区域 i，j 在 Q 中被归类到相同的最大均值聚类的次数，共生矩阵 $\boldsymbol{D}$ 用于区域的最终聚类分析中作为相异度矩阵，应用同样的层次聚类过程再做一次处理。

在聚类过程中，不止文字区域被判定为有意义区域，同样有些有类似文字纹理的区域也被判定为有意义区域，算法得到的聚类算是相对比较准确的，几乎所有的文本区域已经在聚类结果中。为了更进一步去掉极少数的非文本区域，采用两级分类器进行文字区域筛选。首先将组内的每一个区域用一个实值 Adaboost 分类器计算一个是为文字特征的可能性数值，这些特征包括文字的笔画宽度，面积，周长，孔洞数量等；然后将计算得到的数值用到第二级 Adaboost 分类器中判定是否为文字区域，在这一级中也是使用相同的特征，但是针对整个分组，而不是单个独立区域[12]。两个分类器都使用合成的单独的文字图片进行训练。经过级联分类器之后，非文字区域都被排除，只剩下最终的文字区域，将文字区域进行二值化处理，都到黑底白字的文字区域。

2.4 文字识别

对视频帧进行预处理，以及文本区域定位提取之后，得到了只包含二值化后黑底白字的图像，通过对识别出文字区域的边缘像素点进行线性拟合，计算出文字行的上下边界和左右边界，对文字区域块进行提取，将文字区域块送入 OCR 模块中进行文字检测

系统采用 Hideaki Goto 的中文识别 OCR，将经过分割的最终二值化矩形块输入 OCR 模块进行识别。Hideaki Goto OCR 是一个开源 OCR，能够识别中英文，可以直接通过内存输入识别文字图像或者二值化的点阵，输入格式必须是黑底白字的文字图像，输出为识别的文字，对于规整的文字块，识别率可以达到 95%以上。

3 实验结果与分析

根据系统的设计要求，采用的 wmv、avi 的视频格式的文件进行视频关键帧提取与文字信息识别，针对新闻类和会议类两种视频，包括会议场景的短片，以及新闻播放画面等特定的场景。

实验采用新闻视频和会议视频各 50 段，部分会议视频为数码相机拍摄的画面，新闻画面我们采用中央电视台新闻频道的新闻片段，使用文字检测的检出率和错误率两个指标来衡量算法的性能：

检出率＝（检出文字/文字总数）×100%

正确率＝（正确检出数/总检出数）×100%

本文的算法可以有效地从关键帧中提取文字区域，将文字部分从场景中提取出来。经过测试，平均文字检出率为 92.4%，正确率为 91.3%，部分实验结果如图 2 所示：

(a) 原视频帧 (b) 文字检测后二值话图

(c) 原视频帧 (d) 文字检测后二值话图

(e) 原视频帧 (f) 文字检测后二值话图

图 2 部分关键帧识别文字提取结果

表 1 为部分视频检测结果，从表中结果可以看出，分辨率越高所需要的检测时间越；同时，因为会议视频场景变化较小，使用关键帧提取信息可以去掉更多的冗余帧，检测时间也大大缩短。

表 1 部分视频识别结果

序号	视频 1	视频 2	视频 3	视频 4
视频类型	会议	会议	报道素材	新闻
时长	35 分 11 秒	36 分 35 秒	8 分 21 秒	7 分 18 秒
分辨率	1920 * 1080	1024 * 768	720 * 576	1024 * 768
检测用时	22s	17s	9s	8s
检出率/(%)	90.7	94.1	93.4	91.3
正确率/(%)	88.3	92.4	96.6	90.9

从图 2 和表 1 中的数据中可以看出，视频文字信息检查工具的文字检出率，能够很好地将文字从背景中提取出来。但是在检测过程中，对于有些类似于文字的最大稳定极值区域，分类器还是会出现错误划分的情况，针对此情况，我们将对聚类算法和级联分类器算法进行优化。

4 结束语

本文设计的关键帧提取算法能够在保留视频原有信息完整度的同时，大大降低图像处理算法的运算量，基于最大稳定极值区域的文字区域算法具有良好的仿射不变性，对于背景复杂，形态不规整的文字也能够有效识别。

视频文字信息检查工具能够对涉密系统内部视频文件、内部宣传报道图像资料中的文字信息进行提取与识别，判断视频或者其他影像资料中有无敏感信息泄露或者其他违规情况，为视频审查提供重要依据。视频文字信息检查工具对完善涉密系统内部资料管理，促进传统管理模式的转变，防止涉密信息泄露具有重要意义。

参考文献：

[1] Zhuang Y. T, Rui Y, Huang T. S et al Adaptive Key Frame Extraction Using Unsupervised clusting [A]. Prov of IEEE Int Conf

on Image Processing [C]. 1998.
[2] 马小勇，谢　萍，张宪民. 视频帧中提取文字区域的算法 [J]. 计算机工程，2003 (9)：155－157.
[3] Jeannin S. Jasinschi R. Mition. descriptors for content－based video representation [J]. Signal Processing：Image Communication 2000，16 (1－2).：59－85.
[4] Gao J，Yang J，An adaptive algorithm for text detection from natural scenes [J]. Computer Vision and Pattern Recognition，2001，2.
[5] Liu T，Zhang H J. Automatic video scene extraction by shot grouping [A]. in proc ICPR [C]. Barcelona，2000.
[6] Jain A K，Yu B. Automatic text location in images and videoframes [J]. Pattern Recognition，1998，31 (12)：2055－2076.
[7] 李　艳. 基于内容的视频分析与检索方法的研究 [D]. 西安：西安电子科技大学，2007.
[8] 邓　婵. 视频摘要关键技术研究 [D]. 长沙：中南大学，2012.
[9] Zhang Y，Pui Y，Huang T S，et al. Adaptive key frame extraction using unsupervised clustering [J]. In Proc，ICIP [C]. Chicago，1998 (1)：866－870.
[10] Lluis，Gomez，Dimosthenis，et al. Multi－script Text Extraction from Natural Scenes [J]. IEEE，2013，(1341)：21－29.
[11] Chen H，Tsai S，Schroth G，Robust text detection in natural images with edge－enhanced maximally stable extremal regions [A]. in Proc. ICIP [C]. 2011.
[12] Coates A，Carpenter B，Case C，Text detection and character recognition in sceneimages with unsupervised feature learning [A]. in Proc. ICDAR [C]. 2011.

(上接第 1753 页)

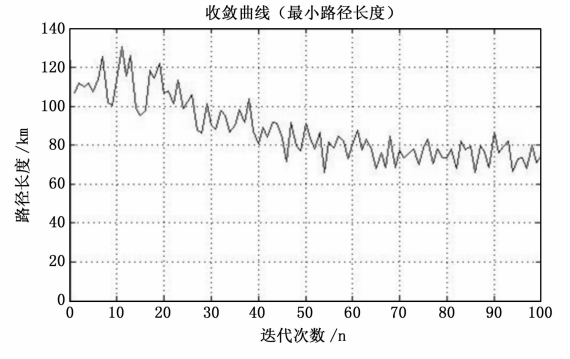

图 4　改进蚁群算法收的敛曲线

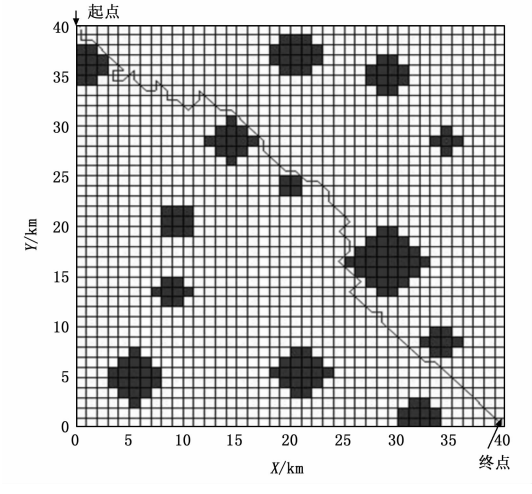

图 5　基本蚁群算法最优航路

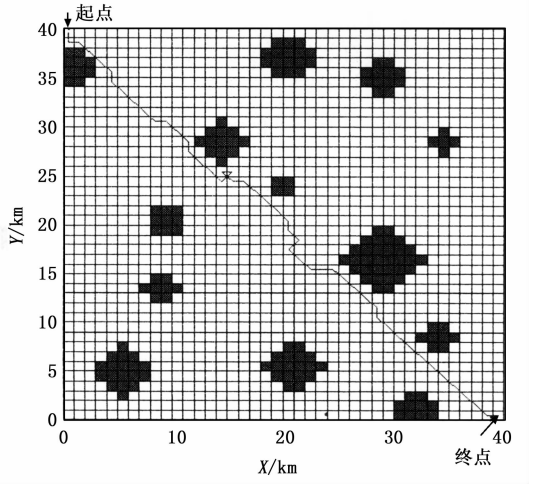

图 6　改进蚁群算法最优航路

4　结论

本文讨论了蚁群算法在无人机航路规划中的应用，针对基本蚁群算法在航路规划时容易陷入局部收敛，导致算法停滞的问题。基于此提出了一种改进的蚁群算法，通过引入多种群并行搜索策略和导引因子，动态的交换各种群信息素，并对每个种群的挥发系数进行自适应调整，扩大了搜索的全局性。通过仿真分析，本文提出的改进算法和基本蚁群算法相比，能有效地避免算法陷入局部最优，并得到更好的规划路径。

参考文献：

[1] 王　俊，周树道，朱国涛. 无人机航迹规划常用算法 [J]. 火力与指挥控制，2012，37 (8)：5－8.
[2] 胡中华，赵　敏，姚　敏. 引入导引因子蚁群算法的无人机二维航迹规划 [J]. 中国机械工程，2011，22 (3)：322－325.
[3] 唐必伟，方　群，朱战霞. 基于改进蚁群算法的无人飞行器二维航迹规划 [J]. 西北工业大学学报，2013，31 (5)：683－688.
[4] 赵开新，魏　勇，王东署. 改进蚁群算法在移动机器人路径规划中的研究 [J]. 计算机测量与控制，2014，22 (1)：3725－3727.
[5] 杜占玮，杨永健，孙永雄. 基于互信息的混合 蚁群算法及其在旅行商问题上的应用 [J]. 东南大学学报，2011，41 (3)：478－481.
[6] 张　臻，王光磊. 基于改进蚁群算法的飞行器航 迹规划 [J]. 指挥信息系统与技术，2011，2 (3)：30－34.
[7] 柴毅哲，杨任农，马明杰，等. 基于改进蚁群算法的可规避威胁源最优航线规划 [J]. 空军工程大学学报 (自然科学版)，2015，16 (2)：23－27.