

# 基于 Hadoop 云计算平台的车牌识别

侯向宁,刘华春

(成都理工大学 工程技术学院,四川 乐山 614007)

**摘要:**针对单机车牌识别系统易受硬件的限制,导致其硬件升级困难,最新的优秀识别算法难以应用以及极端环境下的识别率不理想等困境,提出了基于 Hadoop 云计算平台的车牌识别系统。分别设计了基于 HSV 空间模型的颜色分割与 Sobel 算子相结合的车牌定位算法,以及基于神经网络的字符识别算法。在研究 Hadoop 云计算平台架构的基础上,为了将车牌识别算法应用于 Hadoop 云计算平台,对 MapReduce 数据流及 Hadoop 云计算平台的系统架构进行了设计。通过搭建 Hadoop 云计算平台,将车牌识别算法成功应用于 Hadoop 云计算平台,并对基于 Hadoop 云计算平台的车牌识别系统进行了相关测试。实验结果表明,Hadoop 云计算平台在识别海量车牌数据时,由于其良好的分布式运算特点,使车牌识别的效率和性能远胜于单机。

**关键词:**Hadoop;云计算;车牌识别;MapReduce

**中图分类号:**TP391

**文献标识码:**A

**文章编号:**1673-629X(2018)08-0201-04

**doi:**10.3969/j.issn.1673-629X.2018.08.042

## License Plate Recognition Based on Hadoop Cloud Computing Platform

HOU Xiang-ning, LIU Hua-chun

(School of Engineering & Technology, Chengdu University of Technology, Leshan 614007, China)

**Abstract:** The traditional license plate recognition system is facing many difficulties, such as difficult upgrading limited by hardware, hard application of the latest perfect recognition algorithm and poor recognition rate in extreme environment and so on. Therefore, we put forward a license plate recognition system based on Hadoop cloud computing platform and design license plate localization algorithm based on HSV color space model of segmentation and the Sobel operator, as well as the character recognition algorithm based on neural network. On the basis of research on Hadoop cloud computing platform architecture, we design the data flow of MapReduce and system architecture of Hadoop cloud computing platform in order to apply license plate recognition algorithm into Hadoop cloud computing platform. Through constructing the Hadoop cloud computing platform, license plate recognition algorithm has been successfully applied to the Hadoop cloud computing platform. Experiment shows that the efficiency and performance of Hadoop cloud computing platform is far better than the traditional single machine because of its distributed computing.

**Key words:** Hadoop; cloud computing; license plate recognition; MapReduce

## 0 引言

随着国内机动车辆规模的不断扩大,机动车流量呈爆发式增长,城市交通面临巨大的挑战。车牌识别系统在进行交通管理、处理交通事故、追查被盗车辆以及肇事取证等方面起着非常重要的作用。传统的基于终端的车牌识别系统因受硬件的限制,易导致出现错误、漏识、升级困难,以及一些当前最新的识别算法难以应用等问题。此外,实时性较差、极端环境下的识别率也很不理想。Hadoop 云计算平台具有强大的海量数据分发、存储以及对大数据进行并行运算的能力,是

大数据处理领域的首选。文中提出的车牌识别系统基于 Hadoop 云计算平台,充分利用 Hadoop 云计算平台的各种优势,在实时性、准确性和高效性方面进一步提高车牌识别的效率。

## 1 Hadoop 云计算平台

Hadoop 是对大数据进行分布式处理的软件框架,具有高效分发数据、可靠存储 PB 级数据,便于部署在低成本的普通机群,快速并行处理数据等诸多特点。这些优点使得 Hadoop 在高效性、可靠性以及经济性

收稿日期:2017-08-18

修回日期:2017-12-28

网络出版时间:2018-04-28

基金项目:四川省教育自然科学重点项目(12ZA200);成都理工大学工程技术学院青年科学基金(C122016006)

作者简介:侯向宁(1972-),男,硕士,讲师,研究方向为图形图像处理、机器学习与云计算。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20180427.1630.024.html>

上都无与伦比。HDFS 和 MapReduce 是 Hadoop 框架的核心,其中,HDFS 负责对海量数据的存储和管理,MapReduce 负责对海量数据进行相关运算。

### 1.1 HDFS

HDFS (Hadoop distributed file system) 是用 Java 开发的能够运行在通用机器上的具有高容错性、高吞吐量的分布式文件系统。HDFS 基于 M/S 模式,其体系结构由一个 NameNode 节点和若干 DataNode 节点构成。NameNode 即主控服务器,其职责是维护 HDFS 命名空间,协调客户端对文件的访问及操作,记录文件数据在每个 DataNode 节点上的位置和副本信息<sup>[1-3]</sup>。DataNode 是数据存储节点,在 NameNode 的调度下,负责客户端的读写请求以及本节点的存储管理<sup>[4-5]</sup>。

### 1.2 MapReduce

MapReduce 是处理海量数据的分布式编程模型和并行计算框架<sup>[6]</sup>。编程模型便于开发人员高效地开发应用程序,并行计算框架为应用程序在上千通用机器组成的集群上并行处理 TB 级的海量数据提供了有效保障。

MapReduce 分布式计算模型的核心是 Map 和 Reduce 函数,依据分治法思想,Map 将任务分割成多个子任务,将数据映射为<key, value>键值对,Reduce 将这些具有相同 key 的所有 value 进行规约合并,重复这个过程直到产生所需要的最终结果。MapReduce 分布式编程模型和并行计算框架实现了数据分布存储、任务调度、容错处理等底层机制,软件开发人员只需实现 Map 和 Reduce 函数中的业务逻辑,极大地提高了软件开发的效率。

## 2 车牌识别

### 2.1 车牌定位

车牌定位在车牌识别系统中起着至关重要的作用,车牌定位的准确与否,直接影响到后续的字符分割和字符识别的效率,从而影响整个系统的速度和识别率。从广义的角度,车牌定位可以分为基于灰度图像的边缘检测和基于彩色图像的色彩分割两类<sup>[7]</sup>:基于灰度图像的边缘检测的优点在于计算量小、速度快、实时性好;缺点是在复杂背景下定位的效果不理想。基于颜色分割定位的优点是在复杂背景下车牌定位的效果很好;缺点是计算量较大、速度不够快,当光照不均,车牌与车身颜色相近时效果不理想。

在众多边缘检测算法中,Sobel 算子较为常用,原因在于 Sobel 算子计算量小,抗噪性较好,边缘检测效果好。与 RGB 空间相比,HSV 颜色空间较符合对颜色的主观体验<sup>[8]</sup>。因此,文中采用基于 HSV 空间模型的颜色分割与 Sobel 算子相结合的定位算法,使得车

牌定位的准确性和实时性都得到了较大提升。算法具体流程如图 1 所示。

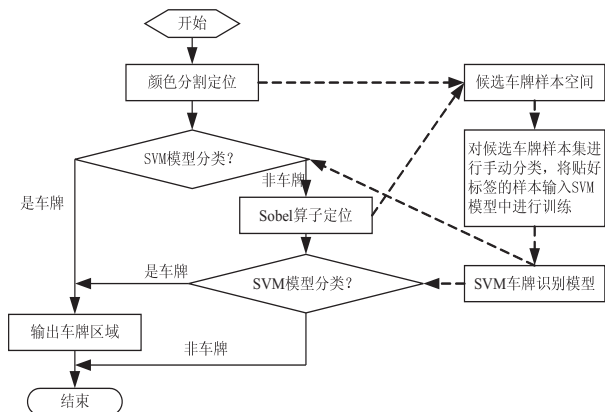


图 1 基于 HSV 空间模型的颜色分割与 Sobel 算子相结合的车牌定位流程

### 2.2 字符分割与识别

字符分割的流程是首先对已定位的车牌进行颜色判断,然后对蓝、黄车牌分别用不同的阈值进行二值化。为了取得良好的车牌字符分割效果,对二值化后的车牌去除边框及铆钉。然后取字符轮廓求其外接矩形,将外接矩形所在的单个字符区域提取出来,最后对单个字符区域进行归一化,以便送入识别模型进行识别。

在众多字符识别方法中,基于神经网络的识别算法自适应和自学习能力突出,其并行分布处理使得算法实时性较好,即使对模糊、扭曲的字符也能识别,具有很好的容错性<sup>[9-10]</sup>。

收集 3 000 个归一化字符区域作为样本,首先对这些样本进行手工分类,贴好标签后,用其中 2 100 个作为训练样本,将其输入人工神经网络(ANN)的 MLP 模型中进行训练,反复迭代直至模型收敛。训练完成后,将剩余 900 个样本输入模型进行测试,进而对参数进一步调优。基于神经网络的字符识别的具体流程如图 2 所示。

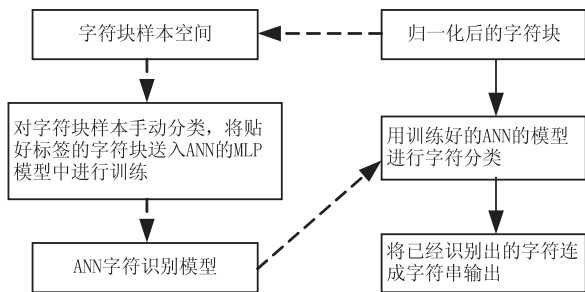


图 2 基于神经网络的字符识别流程

## 3 基于 Hadoop 云平台的车牌识别

### 3.1 MapReduce 流程设计

MapReduce 数据流设计如图 3 所示。

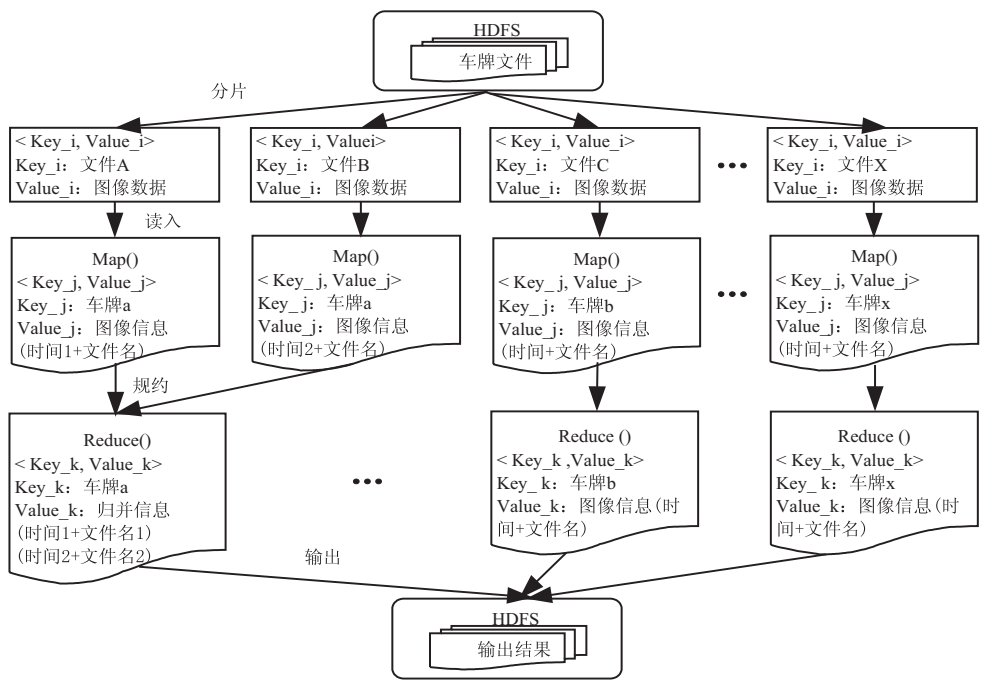


图 3 MapReduce 数据流设计

Map/Reduce 数据流设计是 MapReduce 流程设计的核心,要对车牌图像数据进行存取、运算及处理,首先要对 Map/Reduce 相应的输入、输出 key/value 进行合理的设置<sup>[11]</sup>。MapReduce 框架将存储在 HDFS 上的车牌文件以<Key\_i, Value\_i>进行分片,其中,Key\_i 是车牌文件名,类型是 Text, Value\_i 是要识别的车牌图像,类型是 Image<sup>[12-13]</sup>。接着 MapReduce 框架将分片<Key\_i, Value\_i>输入 Map 函数进行车牌定位、车牌分割、字符识别等操作,处理完后中间结果是<Key\_j, Value\_j>,其中,Key\_j 是识别出来的车牌号码, Value\_j 是诸如图像文件名及拍摄时间等相关信息,类型均为 Text。然后 MapReduce 框架对具有相同 Key\_j 的值 Value\_j 进行组合,以<Key\_j, list( Value\_j )>的形式输入给 Reduce 函数,Reduce 函数进行规约得到<Key\_k, Value\_k><sup>[14]</sup>,其中,Key\_k 是识别出来的车牌号码, Value\_k 是经过规约后的诸如图像文件名及拍摄时间等相关信息的汇总,类型均为 Text,最终交由 HDFS 输出相关结果<sup>[15-16]</sup>。

3.2 系统架构

Hadoop 云平台的系统架构如图 4 所示。为方便用户使用 Web 浏览器对 Hadoop 云计算平台进行操作,需要架设 Tomcat 作为应用服务器,因为 Tomcat 与 Hadoop 均由 Apache 基金会开发,且二者结合近乎完美。配置好 Tomcat 服务器之后,用户就可以通过 Web 浏览器导入车牌文件,Tomcat 服务器接收到车牌文件后调用车牌识别模块并将需要识别的车牌图像文件传给 Hadoop 云计算平台,通过 HDFS 将数据分片输入 MapReduce 进行数据识别,最后由 Tomcat 服务器将识别结

果反馈给用户。

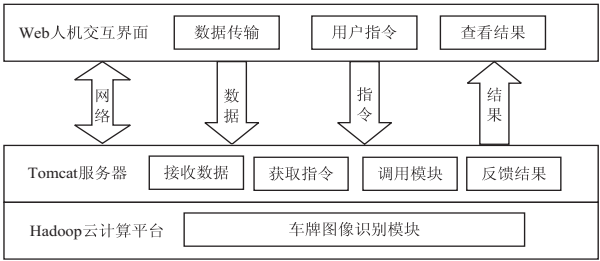


图 4 Hadoop 云计算平台系统架构

4 实验

4.1 实验环境搭建

文中采用伪分布式搭建 Hadoop 云计算平台,优点如下:第一,节省硬件资源,仅需一台内存至少 8 G 的主机就可享受跟完全分布式一样的环境;第二,操作系统安装简单,只需装一台虚拟机,后续只需复制虚拟机的镜像;第三,节省了重复繁杂配置的无谓耗时, Slave<sub>1</sub>配置好之后, Slave<sub>2</sub>和 Slave<sub>3</sub>的配置只需复制 Slave<sub>1</sub>的镜像就可以了;第四,由于三个 Slave 具有相同的软硬件环境,屏蔽了完全分布式方式下软硬件间的差异,便于最终实验结果的对比和分析;最后,伪分布式便于开发人员进行调试和学习。伪分布式方式下各节点的配置如表 1 所示。

4.2 实验结果与分析

实验的目的之一是要观察单机及分布式环境下车牌识别数量与系统耗时之间的关系;另外,更重要的是观察分布式环境中不同节点数量的识别速度。实验设置了 7 组数据,各组需要识别的车牌图像数分别为

100,200,300,500,1 000,2 000 和 3 000。伪分布式环境下节点的个数依次为 1,2,3。最终实验结果如图 5 所示。

表 1 Hadoop 集群节点配置

主机名	节点类型	IP	CPU 核心数	内存
Master	NameNode	192.168.80.100	2	2
Slave <sub>1</sub>	DataNode	192.168.80.101	2	2
Slave <sub>2</sub>	DataNode	192.168.80.102	2	2
Slave <sub>3</sub>	DataNode	192.168.80.103	2	2

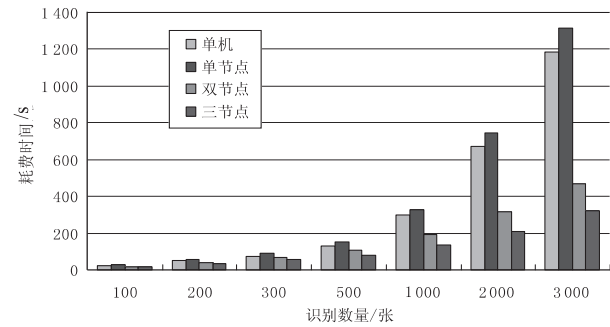


图 5 识别数量与耗时

实验结果显示,在识别数量较小的情况下,单机与单节点、双节点和三节点的速度相差无几,说明 Hadoop 云计算平台的优势还没有发挥出来,这主要因为分布式环境下 master 与 slaves 之间要进行必要的通信以及数据交换,这无疑增大了系统开销,尤其在识别数量较少的情况下,这种开销不容忽视。随着识别数量的增加,由于单机没有额外的系统开销,因而比单节点的耗时略小;然而单机与双节点、三节点的耗时差异越来越大,这时 Hadoop 云计算平台并行计算的优势越来越突出,当识别数量达到 3 000 时,分布式下三节点的识别速度比单机将近快了 3.6 倍,这主要是因为海量数据冲击下分布式环境中各节点的负载较均衡,内存消耗比单机的要小很多,正是单机内存的消耗在一定程度上拖累了它的效率。实验结果表明,Hadoop 云计算平台在识别海量车牌数据时,其效率和性能远远超越单机,具有很好的应用前景。

5 结束语

在分析车牌识别流程的基础上,设计了基于 HSV 空间模型的颜色分割与 Sobel 算子相结合的车牌定位算法,以及基于神经网络的字符识别算法。在研究 Hadoop 云计算平台架构的基础上,为了将车牌识别算法应用于 Hadoop 云计算平台,对 MapReduce 数据流及 Hadoop 云计算平台的系统架构进行了设计。通过搭建 Hadoop 云计算平台环境,对基于 Hadoop 云计算

平台的车牌识别系统进行了相关测试。实验结果表明,Hadoop 云计算平台在识别海量车牌数据时,其效率和性能远远超越单机,具有很好的应用前景。

参考文献:

[1] 赵鑫,石龙,张建光,等. HDFS 在智慧博物馆文件存储系统中的应用[J]. 衡水学院学报,2017,19(4):6-10.

[2] HAN Yongqi,ZHANG Yun,YU Shui. Research of cloud storage based on Hadoop distributed file system[J]. Applied Mechanics and Materials,2014,513-517:2472-2475.

[3] 朱媛媛,王晓京. 基于 GE 码的 HDFS 优化方案[J]. 计算机应用,2013,33(3):730-733.

[4] 余琦,凌捷. 基于 HDFS 的云存储安全技术研究[J]. 计算机工程与设计,2013,34(8):2700-2705.

[5] 廖彬,于炯,张陶,等. 基于分布式文件系统 HDFS 的节能算法[J]. 计算机学报,2013,36(5):1047-1064.

[6] DEAN J,GHEMAWAT S. MapReduce:a flexible data processing tool[J]. Communications of the ACM,2010,53(1):72-77.

[7] 李晓宾,李淑珍,张晓. 基于色彩校正与分层块匹配的图像配准算法[J]. 计算机工程与设计,2016,37(12):3349-3354.

[8] 陈亦欣,叶锋,肖锋,等. 基于 HSV 空间和形状特征的交通标志检测识别研究[J]. 江汉大学学报:自然科学版,2016,44(2):119-125.

[9] 尹龙,尹东,张荣,等. 一种扭曲粘连字符验证码识别方法[J]. 模式识别与人工智能,2014,27(3):235-241.

[10] 康健新. 基于图像的车牌识别系统的设计和实现[D]. 长春:吉林大学,2014.

[11] PAN Wumin,HA Libai. Study of Map-Reduce over Hadoop based cloud computing environment[J]. Applied Mechanics and Materials,2014,509:175-181.

[12] ZHU Haodong,SHEN Zhen,SHANG Li,et al. Parallel image texture feature extraction under Hadoop cloud platform[C]//10th international conference on intelligent computing theory. Taiyuan, China: Springer International Publishing: 2014:459-465.

[13] SMITH C,ALBARGHOUTH A. MapReduce program synthesis[J]. ACM SIGPLAN Notices,2016,51(6):326-340.

[14] 王旭仁,姚叶鹏,冉春风,等. 一种并行 LDA 主题模型建立方法研究[J]. 北京理工大学学报,2013,33(6):590-593.

[15] GUO Wenming,ALHAM N K,LIU Yang,et al. A resource aware mapreduce based parallel SVM for large scale image classifications[J]. Neural Processing Letters,2016,44(1):161-184.

[16] 杨婉婧,邢洪嘉,曹建芳. 基于 MapReduce 的 GA-BP 神经网络算法并行化设计及实现[J]. 软件导刊,2017,16(7):40-44.