

DOI:10.6041/j.issn.1000-1298.2012.08.035

蛋鸡养殖过程中海量数据自动处理系统*

劳凤丹¹ 滕光辉² 余礼根² 李卓²

(1. 中国农业大学网络中心, 北京 100083; 2. 中国农业大学农业部设施农业工程重点开放实验室, 北京 100083)

【摘要】 采用分时变频的采集模式,应用Java媒体框架模型,设计并实现了蛋鸡养殖过程中海量数据自动处理系统。系统既可同步采集和保存关键的环境、图像及音频信息,又能减少存储空间、降低存储成本和加速查询速度。系统运行结果表明:依据采集设备及采集参数设置的不同,单采集点一天采集的音频及图像数据量为200~400 MB,一年采集的数据量为70~140 GB。系统同时提供了以采集时间戳为索引的多源信息同步查询及同步展示模块,可复原大部分生产过程场景。

关键词: 蛋鸡 养殖 图像 音频 海量数据

中图分类号: TP391; S818.9 **文献标识码:** A **文章编号:** 1000-1298(2012)08-0193-05

Automatic Processing System of Massive Data in Laying Hens Raise

Lao Fengdan¹ Teng Guanghui² Yu Ligen² Li Zhuo²

(1. Network Center, China Agricultural University, Beijing 100083, China

2. Key Laboratory of Ministry of Agriculture for Agricultural Engineering in Structure and Environment, China Agricultural University, Beijing 100083, China)

Abstract

For providing traceable farm production, a data automatic processing system for egg production was designed and implemented, which altered frequency of acquisition based on time. Java media framework was used as the main model to acquire and process massive image and sound datasets. The system could not only acquire and save synchronized environmental, image and sound data, but also significantly reduce storage space and costs, and speed up query speeds. Tests show that image and sound data in one day is only 200~400 MB, and collected data in one year is about 70~140 GB, yet contains most of the scenes of the production process.

Key words Laying hens, Raise, Image, Sound, Massive data

引言

现代畜禽生产不再只局限于以经济利益为目标,针对动物健康和福利、食品安全和质量、畜牧业可持续发展、畜牧业对环境的影响等方面的研究及关注度正在不断提高^[1-2]。在肉类需求不断增长的情况下,畜禽饲养过程中如何确保动物福利、减少环境影响、生产出安全的畜禽产品是当前畜禽生产的研究热点。

现代畜禽养殖过程存在几个问题:依靠专家观察畜禽行为进行动物福利评价不但成本高,且因观察时间短,具有主观性,所以不同专家的结论可能会不同^[1-7];辨别和避免饲养风险的解决方案需要庞大生产数据的支撑,但数据的收集几乎是依赖于手工方式,手工收集数据已无法适应大规模商业养殖的需求^[1];目前较为普及的视频自动监控方案只解决了实时监控的目的,但无法提供多年的历史数据及自动溯源查询功能;畜禽

收稿日期:2011-09-28 修回日期:2012-01-18

* 国家自然科学基金资助项目(31072066)和公益性行业(农业)科研专项经费资助项目(201003011)

作者简介:劳凤丹,高级工程师,博士生,主要从事计算机网络、农业信息化研究,E-mail: laofd@cau.edu.cn

通讯作者:滕光辉,教授,博士生导师,主要从事农业信息化、设施农业工程研究,E-mail: futong@cau.edu.cn

的福利状态通过体态行为(图像)和音频信息体现,二者具有同步相关性,但长期视频存储成本太高。

本研究针对蛋鸡养殖过程的海量生产数据,包括环境信息、图像信息和音频信息,研究自动实时数据采集及处理系统,既可同步采集和保存各类数据,又可减少存储空间、降低存储成本、加速查询速度和同步还原及展示多源信息。

1 系统组成

蛋鸡养殖过程中海量数据自动处理系统主要由环境变量采集模块、图像和音频同步采集模块、数据处理/存储模块、数据同步还原模块、Web 查询及展示模块组成,如图 1 所示。

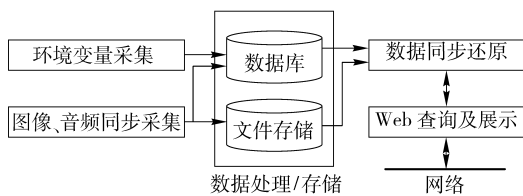


图 1 系统体系结构

Fig. 1 System architecture

环境变量采集模块定期自动采集温度、湿度、二氧化碳、氨气、风速、光照等传感器的数据,以采集时间为索引存入数据库;图像、音频同步采集模块使用 Java 媒体框架(Java media framework, 简称 JMF)构建,实现对媒体数据的获取;数据处理/存储模块负责以媒体文件形式存储压缩处理后的图像及音频数据,且相关描述信息被同步存储到数据库中;Web 查询及展示模块接受用户查询某一日期(或时刻)生产数据信息的请求,并提交给同步还原模块;数据同步还原模块响应查询请求,对数据处理及存储模块进行同步数据调度,并把结果返回,供用户 Web 方式查看。

1.1 开发工具

为了保证系统可跨平台,增加稳定性和规范性,软件部分的算法、协议、框架的具体实现选用 Java 语言开发。

基于使图像及音频信息采集与具体硬件平台无关的构想,未使用摄像机的 SDK 开发包,而是选取了具有先进媒体处理能力的 JMF 媒体框架来采集音频、视频信号。JMF 是一个把音频、视频和其他基于时间的媒体结合到 Java 程序的应用程序接口,能支持绝大多数的媒体类型。通过利用 Java 平台的优势,JMF 为访问底层的媒体框架提供了一个通用、跨平台和与硬件设备无关的 Java API^[8-9],符合开发需求。

1.2 数据采集/存储方案

在蛋鸡的整个养殖周期,不作任何处理地直接保留所有的录像信息可行性不高,具体表现为:所需存储空间大,存储资金成本高;同一天的不同时段,蛋鸡活动频繁度差异显著,例如蛋鸡在夜间很平静,而在白天的进食阶段活动频繁,24 h 不间断的录制没有必要;媒体文件太大会导致查询不方便,网络访问速度慢。长时间录制一个视频文件,中间出现故障导致视频数据损坏或丢失概率也会升高。

基于以上分析,本系统采取如下解决方案:

(1)采用分时变频的采集模式,音频、图像的采集频率随蛋鸡活动频繁度变化。夜间 21:00 至次日凌晨 5:00 每 1 h 采集 1 min;白天 5:00 ~ 21:00 在喂料等关键活动时间点,每 1 min 采集 20 s 数据(共计 2 h),非关键时间点每 10 min 采集 1 min 数据。

(2)音频连续采集,图像间隔抓拍,音频和图像压缩后有序分开保存。采集阶段,在连续录制声音的同时,每隔 2 s 抓拍一幅图片,进一步大幅度地缩减所需的存储空间。每次采集的数据文件以时间戳命名,合理压缩后被分别保存到文件服务器对应存储路径。

(3)选择合适的音频和图像文件的压缩技术及压缩率。为了能在 Web 上流畅地播放采集的图像和音频文件,节省存储空间,媒体文件必须要足够小,但是,压缩后的图像和音频品质也必须要达到进一步的动物行为识别研究需求。本研究选取了目前网络上主流和高效的压缩算法。

利用 MP3 技术(MPEG audio layer3)压缩音频,它能够在降低文件容量情况下保持较好音质。

利用 JPEG 标准(joint photographic experts group)压缩图像,它用有损压缩方式去除冗余的图像数据,在获得极高的压缩率的同时能展现丰富生动的图像。JPEG 虽然损失了图像的一部分高频谱线,但绝大部分有用信息都保存了下来^[10-11]。

2 关键技术

数据采集是蛋鸡养殖过程中海量数据自动处理系统的首要环节,是数据处理、存储和同步还原、同步播放的基础;溯源查询结果准确与否取决于所采集的数据能否同步;采集数据存储模式及同步还原方法决定了海量数据搜索的速度,也影响用户的查询体验。图 2 为系统图像音频同步采集/处理流程图。

自动处理模块已实现了完全自动化,每次采集会顺次执行:采集、ffmpeg 抽取所需数据及压缩保存、删除中间文件、关闭采集等流程,无需人为干预。

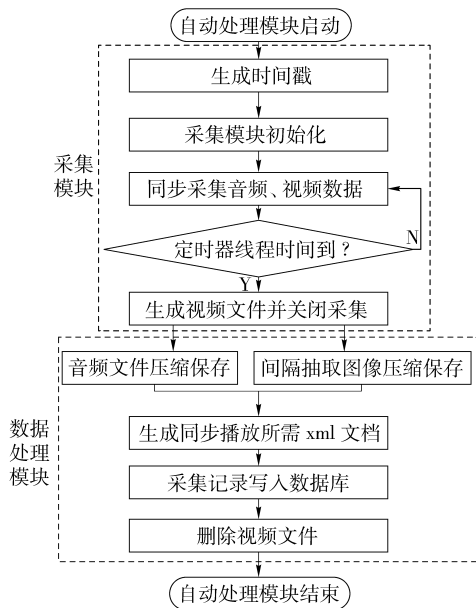


图 2 图像、音频同步采集/处理方法流程图

Fig. 2 Synchronous acquisition and process flow chart of image and sound

它可依据部署的平台不同分别由任务计划 (Windows 平台) 或 crond 守护进程 (Linux 平台) 按照部署设置的采集间隔周期性自动执行。

实现音频、图像的同时采集 (同步还原) 以及压缩保存是借助于跨平台的开源音频、视频方案 ffmpeg 实现的。ffmpeg 提供了格式转换以及压缩音频、视频的完整解决方案, 包含了先进的音频/视频编、解码库 libavcodec。系统首先获取的是视频, 再通过程序调用 ffmpeg 专门工具间隔抽取得到图像集, 此种采集比定时单张采集图像所获得的图像集与音频的同步更精确, 且容易实现同步压缩。

2.1 图像及音频同步采集方法

摄像头和传声器分别接入采集服务器, 作为 JMF 的图像和音频的捕获设备。采集音频和图像的重点是实现两者的同步采集。

在每次采集开始时, 程序根据当前系统时间首先生成索引时间戳, 用以标记音频和图像文件。例如某次开始采集的时间是: 2011-08-11 14:02, 则根据程序所定义的时间格式, 时间戳变量 timeformat 被赋值为: 201108111402。

JMF 采集基本过程如下:

查询设备管理器 (CaptureDeviceManager), 定位需要使用的媒体采集设备并得到设备的 CaptureDeviceInfo 实例; 用实例的 getLocator 方法获得媒体定位器 (MediaLocator), 并通过它创建原始采集数据源 (DataSource); 指定音、视频文件的输出格式 (Format) 及定义要采集的媒体内容类型

(FileTypeDescriptor); 用数据源创建处理器 (Processor); 启动处理器开始捕获媒体数据, 通过程序创建的数据接收器 (DataSink) 将音频流和视频流同步保存到本地文件。

以下为启动采集的函数 start 的部分关键代码:

```

private void start() {
    DataSource source = processor.getDataOutput();
    MediaLocator d = new MediaLocator(savePath);
    try {
        // 创建数据接收器:
        fileWriter = Manager.createDataSink(source, d);
        fileWriter.open();
        fileWriter.start();
    } catch (Exception e) { e.printStackTrace(); }
    // 开始采集:
    processor.start();
    player.start();
    // 调用定时关闭采集的线程实例:
    autoclose.start();
}
  
```

定时关闭采集 autoclose 线程的功能主要是控制采集的时间, 采集时间到, 则停止采集, 并停止写视频流文件。

2.2 图像、音频及环境变量的同步保存方法

按照 1.2 节的分时变频采集模式, 1 个鸡舍 (假设单一摄像头、单一传声器的情况) 1 d 所采集到的图像和音频文件的数量为 4 384 个, 则 1 年采集生成的文件数量超过 160 万个, 随着时间、鸡舍数量、采集设备数量的增多, 文件数量的增长是可观的。如果把所有采集数据堆放在同一个目录中, 打开目录速度会非常慢。本系统数据处理模块设计声音 (包括 xml 文件) 和图像自动实现多级目录分开存放, 存储目录分别为 (以 Windows 为例)

E:\video\sound\<年>\<月>\<日>\<鸡舍号>\

E:\video\img\<年>\<月>\<日>\<鸡舍号>\

<> 内表示的是当前目录所代表的含义。例如: 2011 年 8 月 1 日采集到的 2 号鸡舍的声音文件 (及每一次采集所生成的 xml 文件) 自动保存到 E:\video\sound\2011\8\1\2\ 目录下, 图像文件自动保存到 E:\video\img\2011\8\1\2\ 目录下。

系统使用的抓图命令为

```

cmd = "cmd /c start d:\ffmpeg.exe -i " + avi + "
-r 0.2 -s 640 * 480 -y -f image2 " + pname + "
-%03d.jpg";
  
```

命令中, avi 表示本次采集获得的 avi 文件名; pname 表示抽取的图像系列的名称前缀; %03d 表示生成的图像文件名包含 3 位流水号。avi 和 pname 变量都包含绝对路径,且在采集过程开始前已被程序算出并自动赋值。

Java 程序调用 process 执行抓图命令

```
process p = Runtime.getRuntime().exec(cmd);
```

按照采集先后顺序,依次生成例如: 201108111402 - 001. jpg、201108111402 - 002. jpg、... 的单次采集图像文件系列。

在采集及处理模块程序执行过程中,程序会同步在数据库的采集记录表中生成一条保存了采集日期、鸡舍号、xml 文件名、MP3 文件名、采集时间、图像数量、图像采集间隔等关键信息的记录,以加快 Web 查询速度。

2.3 图像、音频及环境变量的同步播放方法

本系统使用专用的 Flash 播放器来实现图像和音频的 Web 同步播放,选用的是 jw flv player 中的 imagerotator 播放器。每次采集及处理过程会自动生成一个播放序列;背景音乐为采集的 MP3 音频文件,图像文件集合及播放顺序则由数据处理模块按照 imagerotator 规定格式所生成的 xml 文件来定义。只需在 Web 调用时设置图像系列的播放时间间隔与采集图像的时间间隔一致,则可实现图像和音频数据的同步播放。以下给出 Web 页面调用播放器的部分关键代码

```
//调用 Flash 播放器并设定其大小为 640 × 480
s1 = new SWFObject("imagerotator.swf",
"rotator","640","480","7");
//定义当前播放文件为变量 xml 指定的文件
s1.addVariable("file","<% = xml% >");
//定义显示下一幅图像的时间间隔为 2 s,此
//时间间隔与图像采集间隔时间需一致
s1.addVariable("rotatetime","2");
//定义播放图像系列时的背景声音
s1.addVariable("audio","<% = mp3% >");
//定义图像按 xml 顺序依次播放
s1.addVariable("shuffle","false");
```

因环境参数的变化相对缓慢,程序提供了可根据需要调节采集时间间隔的功能。本系统目前设置每隔 3 min 采集一次环境数据。在 Web 查询显示中,根据播放的音频文件的时间戳,系统会查找数据库中时间差异最小的时刻所采集的环境数据,并同步展示。

3 系统测试

设置的 JMF 视频采集参数为:分辨率为 640 × 480,使用 RGB 模式,24 位的编码格式,抽取并通过 ffmpeg 工具压缩后图像大小为 20 ~ 40 KB(依据摄像头规格、垫料、光照、鸡个体的不同图像大小略有差异)。设置的音频采样频率为 44 100 Hz,声道为立体声,采样大小为 16 位,经由 ffmpeg 压缩,20 s 生成的 MP3 音频文件约为 370 KB。依据如前所叙的分时变频采集方案,1 d 采集到的音频和图像数据所需存储空间为 226 ~ 384 MB,1 年的数据量为 83 ~ 140 GB,1 个养殖周期的数据量也仅约为 200 GB。图 3 为本系统的 Web 查询及多源数据同步展示界面。



图 3 多源数据同步查询及展示界面

Fig. 3 Synchronous query and display interface

4 结论

(1) 实现了一种可溯源海量数据自动处理系统,所提出的分时变频的采集模式既可同步保存蛋鸡养殖过程的关键数据,又可极大节省存储空间和降低存储成本。

(2) 所采用的数据库与文件系统有机结合的存储模式加快了海量数据同步检索的速度。系统将环境信息及媒体采集的索引信息保存到数据库,而音频和图像数据则被有规律地以文件形式保存,既利用了数据库处理结构化数据的优势,又兼顾了文件系统处理非结构化音频、视频数据的优势。

(3) 实现了通过时间戳同步检索多源数据的接口。通过 Flash 播放器实现了各个采集时间段内音频、图像和环境参数的 Web 同步查询及展示。

(4) 目前 Web 查询及展示模块还未实现连续的多个采集片段数据自动连续播放的功能。每播放完一个采集片段,需手工点击下一个采集片段的时间戳才可继续播放。

参 考 文 献

- 1 Berckmans D. Perspectives of precision livestock farming[C]//Biennial Conference of the Australian Society for Engineering in Agriculture(SEAg), Brisbane: SEAg, 2009.
- 2 王玲, 王超峰, 郑奎. 基于 ARM 的蛋鸡养殖场网络视频监控系统设计[J]. 农业机械学报, 2012, 43(2): 186 ~ 191.
Wang Ling, Wang Chaofeng, Zheng Kui. Design and implementation of network video surveillance system for laying hens farm based on ARM[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(2): 186 ~ 191. (in Chinese)
- 3 Cangar O, Leroy T, Guarino M. Automatic real-time monitoring of locomotion and posture behavior of pregnant cows prior to calving using online image analysis[J]. Computers and Electronics in Agriculture, 2008, 64(1): 53 ~ 60.
- 4 Xue Xinwei, Henderson Thomas C. Feature fusion for basic behavior unit segmentation from video sequences[J]. Robotics and Autonomous Systems, 2009, 57(3): 239 ~ 248.
- 5 Aydin A, Cangar O, Ozcan S E, et al. Application of a fully automatic analysis tool to assess the activity of broiler chickens with different gait scores[J]. Computers and Electronics in Agriculture, 2010, 73(2): 194 ~ 199.
- 6 Ahmad P, Anette W, Claudia B, et al. Automatic monitoring of turkeys: a vision-based approach to detect and analyse the behaviour of turkeys in transport cages based on ellipse fitting[C]//ASABE-7th World Congress on Computers in Agriculture Conference Proceedings, 2009.
- 7 陈长喜, 张宏福, 飞颢经纬. 肉鸡产业技术体系生产监测与产品质量可追溯平台设计[J]. 农业机械学报, 2010, 41(8): 100 ~ 106.
Chen Changxi, Zhang Hongfu, Feixie Jingwei. Traceability platform design of production monitoring and products quality for broilers industry technology system[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(8): 100 ~ 106. (in Chinese)
- 8 陆颖瑜, 林海, 张永林. 基于 Java 与 JMF 的视频监控系统[J]. 计算机工程与设计, 2007, 28(11): 2 740 ~ 2 742, 2 761.
Lu Yingyu, Lin Hai, Zhang Yonglin. Video surveillance system based on Java and JMF[J]. Computer Engineering and Design, 2007, 28(11): 2 740 ~ 2 742, 2 761. (in Chinese)
- 9 王从局. 基于 JMF 的流媒体传输实践[J]. 中国民航飞行学院学报, 2011, 22(4): 53 ~ 56.
Wang Congju. Transmission practice based on JMF streaming media[J]. Journal of Civil Aviation Flight University of China, 2011, 22(4): 53 ~ 56. (in Chinese)
- 10 熊迎军, 沈明霞, 孙玉文, 等. 农田图像采集与无线传输系统设计[J]. 农业机械学报, 2011, 42(3): 184 ~ 187.
Xiong Yingjun, Shen Mingxia, Sun Yuwen, et al. Design on system of acquisition and wireless transmission for farm land image[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(3): 184 ~ 187. (in Chinese)
- 11 刘龙申, 沈明霞, 孙玉文, 等. 基于 FPGA 的农田图像采集与 3G 无线传输系统设计[J]. 农业机械学报, 2011, 42(12): 186 ~ 190.
Liu Longshen, Shen Mingxia, Sun Yuwen, et al. Acquisition system and wireless transmission by 3G for farmland image based on FPGA[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(12): 186 ~ 190. (in Chinese)

(上接第 234 页)

- 6 张建华. 单程平面磨削淬硬层预测及其摩擦磨损性能研究[D]. 济南: 山东大学, 2008: 19 ~ 102.
Zhang Jianhua. Hardening layer forecast and frictional wear investigation for single-pass plane grinding-hardening technology [D]. Ji'nan: Shandong University, 2008: 19 ~ 102. (in Chinese)
- 7 Lavine A S, Malkin S, Jen T. Thermal aspects of grinding with CBN wheels[J]. CIRP Annals-manufacturing Technology, 1989, 38(1): 557 ~ 560.
- 8 王霖, 葛培琪, 秦勇. 基于有限元法的湿式磨削温度场分析[J]. 机械工程学报, 2002, 38(9): 155 ~ 158.
Wang Lin, Ge Peiqi, Qin Yong. Wet grinding temperature analysis based on FE [J]. Chinese Journal of Mechanical Engineering, 2002, 38(9): 155 ~ 158. (in Chinese)
- 9 张建华, 葛培琪, 张磊. 磨削淬火技术的温度场仿真研究[J]. 武汉理工大学学报, 2008, 32(1): 78 ~ 80.
Zhang Jianhua, Ge Peiqi, Zhang Lei. Temperature simulation for grinding-hardening technology [J]. Journal of Wuhan Polytechnics University, 2008, 32(1): 78 ~ 80. (in Chinese)
- 10 张朝晖. ANSYS 热分析教程与实例解析[M]. 北京: 中国铁道出版社, 2007.