

(12) 发明专利申请

(10) 申请公布号 CN 102737480 A

(43) 申请公布日 2012. 10. 17

(21) 申请号 201210235097. 5

G10L 15/00(2006. 01)

(22) 申请日 2012. 07. 09

(71) 申请人 广州市浩云安防科技股份有限公司

地址 511400 广东省广州市番禺区桥南街市
良路金业花园东二街 16 号 101

(72) 发明人 曹江中 高传江

(74) 专利代理机构 深圳市科吉华烽知识产权事
务所 44248

代理人 胡吉科

(51) Int. Cl.

G08B 25/00(2006. 01)

H04N 7/18(2006. 01)

G10L 11/00(2006. 01)

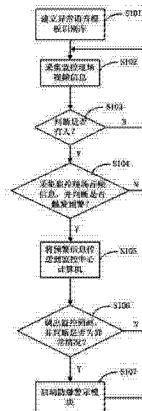
权利要求书 2 页 说明书 7 页 附图 4 页

(54) 发明名称

一种基于智能视频的异常语音监控系统及方
法

(57) 摘要

本发明涉及一种基于智能视频的异常语音监
控系统及方法，其系统包括监控区域现场装置，与
监控区域现场装置连接的智能监控分析装置，以
及分别与监控区域现场装置和智能监控分析装置
连接的监控中心计算机；监控区域现场装置包括
视频采集模块，被智能监控分析装置触发的语音
接收模块和被监控中心计算机触发的防爆警示模
块；智能监控分析装置包括分别与视频采集模块
和语音接收模块连接、其输出端与监控中心计算
机连接的中心处理器，与中心处理器连接的数据
存储模块，与中心处理器连接的视频处理模块，及
与中心处理器连接的声强检测模块。实施本发明
的基于智能视频的异常语音监控系统及方法，具
有以下有益效果：智能化较高、实时性较强、误报
率较低。



1. 一种基于智能视频的异常语音监控系统，其特征在于，包括监控区域现场装置，与所述监控区域现场装置连接的智能监控分析装置，以及分别与所述监控区域现场装置和智能监控分析装置连接的监控中心计算机；所述监控区域现场装置包括用于采集监控现场区域图像的视频采集模块，被所述智能监控分析装置触发并用于对现场声音进行监听的语音接收模块和被所述监控中心计算机触发并在启动时发出报警的防爆警示模块；所述智能监控分析装置包括其输入端分别与所述视频采集模块和语音接收模块连接、其输出端与所述监控中心计算机连接的中心处理器，与所述中心处理器连接并用于存储现场活动数据的数据存储模块，与所述中心处理器连接并用于根据具体动态情况检测到现场有人后启动所述语音接收模块的视频处理模块，以及与所述中心处理器连接并用于进行声强检测的声强检测模块。

2. 根据权利要求 1 所述的基于智能视频的异常语音监控系统，其特征在于，所述智能监控分析装置还包括被所述中心处理器触发并用于进行异常语音检测的异常语音识别模块。

3. 一种基于智能视频的异常语音监控系统的监控方法，其特征在于，包括如下步骤：

B) 所述视频采集模块采集监控现场区域图像，所述视频处理模块根据具体动态情况判断监控现场区域是否有人，如有人，执行步骤 C)；否则，执行步骤 E)；

C) 所述语音接收模块收集监控现场声音并进行判断是否触发预警，如触发预警，将预警信息传送到所述监控中心计算机，执行步骤 D)；否则，返回步骤 B)；

D) 所述监控中心计算机收到预警信息后，调出现场监控画面，并由值班人员判断是否为异常情况，如是，启动所述防爆警示模块；否则，执行步骤 E)；

E) 结束本次监控并返回步骤 B) 进行下一次监控。

4. 根据权利要求 3 所述的基于智能视频的异常语音监控系统的监控方法，其特征在于，在所述步骤 B) 之前还包括步骤：

A) 建立异常语音模板识别库。

5. 根据权利要求 4 所述的基于智能视频的异常语音监控系统的监控方法，其特征在于，所述异常语音模板识别库是依次经过语音样本采集、样本预处理、MFCC 特征提取和支持向量机训练步骤得到。

6. 根据权利要求 5 所述的基于智能视频的异常语音监控系统的监控方法，其特征在于，所述步骤 C) 进一步包括：

C1) 所述语音接收模块收集监控现场声音，将声压模拟信号经过增益放大、模数转换为数字信号并将所述数字信号传送到所述智能监控分析装置；

C2) 用所述中心处理器中的 FIR300Hz~3.4KHz 带通滤波器对所述数字信号进行滤波，提取人的 300Hz~3.4KHz 频率范围内的声音信号；

C3) 由 $P=20\log_{10}(x)$ 计算声强值；其中，P 为声强，其单位为分贝，x 为输入声压数据；

C4) 所述中心处理器依据得到的声强值波形特点，对非语音类声音进行滤除；

C5) 判断声强是否大于预设的声强触发阈值，如是，截取声音并执行步骤 C6)；否则，返回步骤 B)；

C6) 判断声强是否大于预设的声强预警阈值，如是，向监控中心计算机发出声强预警，并启动异常语音识别模块工作；否则，返回步骤 B)；

C7) 提取输入的可疑语音数据的 MFCC 特征参数；

C8) 判断所述提取的 MFCC 特征参数与所述异常语音模板识别库是否匹配，如匹配，向监控中心计算机发出异常语音预警；否则，返回步骤 B)。

7. 根据权利要求 6 所述的基于智能视频的异常语音监控系统的监控方法，其特征在于，所述步骤 C4) 中对非语音类声音进行滤除为用 FIR2.5KHz ~3KHz 带阻滤波器对车鸣笛声进行滤除，所述中心处理器通过检测连续数据长度直接将拍掌声、敲击声滤除。

8. 根据权利要求 6 所述的基于智能视频的异常语音监控系统的监控方法，其特征在于，所述步骤 C5) 中的声强触发阈值为 70 分贝。

9. 根据权利要求 6 所述的基于智能视频的异常语音监控系统的监控方法，其特征在于，所述步骤 C6) 中的声强预警阈值为 105 分贝。

10. 根据权利要求 6 所述的基于智能视频的异常语音监控系统的监控方法，其特征在于，所述步骤 C8) 中的判断是通过支持向量机进行判断并得到识别结果。

一种基于智能视频的异常语音监控系统及方法

技术领域

[0001] 本发明涉及安防监控领域,更具体地说,涉及一种基于智能视频的异常语音监控系统及方法。

背景技术

[0002] 随着社会日新月异的发展,人民的安全意识在不断增强,对舒适、安全的活动环境期望越来越高,这也使得安防监控面临着更高难度的挑战。在公共安全领域,特别是在ATM (Automatic Teller Machine, 自动柜员机)室、垂直升降电梯和监狱等一些常容易发生抢劫、性骚扰或者打架等犯罪、暴力事件的区域更需要进行监控。

[0003] 目前,视频监控是常用的一种监控方法,但是普通的视频监控太过依赖人的作用,由于值班人员的时间、精力有限,常会出现疏忽漏报的情况,从而错失了预警的最佳时机,监控信息只能作为事后证据了。为了弥补这一缺点,一些有实力的安防公司研发了智能视频监控产品,提高了监控效率,减轻了监控值班人员的负担,但是这些智能监控产品只对视频信息进行处理,技术实现比较复杂,容易受阳光、灯光的光线变化影响,摄像头也无法监控到视角以外的区域,因此,在一些要求较高、环境较复杂的场景下,智能视频仍具有一定的局限性。

[0004] 语音识别技术发展到现在取得了很大的进步,开始从实验室走向社会,但是在普通安防监控领域的应用还比较少见。在抢劫、暴力犯罪场景中,常会伴有呼救声、叫喊等一些比较紧急的声音,这些异常的语音可以作为安防监控的重要信息加以利用。目前,语音识别技术虽然取得了长足的进步,一些研究者对视频和语音相结合的安防监控方法进行了探讨,取得了一定的成果,但是,由于语音容易受环境噪声的影响,鲁棒性比较差,一般的语音识别技术很难适应各种监控环境,造成智能化不高、实时性不强、误报率较高等问题。

发明内容

[0005] 本发明要解决的技术问题在于,针对现有技术的上述智能化不高、实时性不强、误报率较高的缺陷,提供一种智能化较高、实时性较强、误报率较低的基于智能视频的异常语音监控系统及方法。

[0006] 本发明解决其技术问题所采用的技术方案是:构造一种基于智能视频的异常语音监控系统,包括监控区域现场装置,与所述监控区域现场装置连接的智能监控分析装置,以及分别与所述监控区域现场装置和智能监控分析装置连接的监控中心计算机;所述监控区域现场装置包括用于采集监控现场区域图像的视频采集模块,被所述智能监控分析装置触发并用于对现场声音进行监听的语音接收模块和被所述监控中心计算机触发并在启动时发出报警的防爆警示模块;所述智能监控分析装置包括其输入端分别与所述视频采集模块和语音接收模块连接、其输出端与所述监控中心计算机连接的中心处理器,与所述中心处理器连接并用于存储现场活动数据的数据存储模块,与所述中心处理器连接并用于根据具体动态情况检测到现场有人后启动所述语音接收模块的视频处理模块,以及与所述中心处

理器连接并用于进行声强检测的声强检测模块。

[0007] 在本发明所述的基于智能视频的异常语音监控系统中，所述智能监控分析装置还包括被所述中心处理器触发并用于进行异常语音检测的异常语音识别模块。

[0008] 本发明还涉及一种采用上述基于智能视频的异常语音监控系统的监控方法，包括如下步骤：

B) 所述视频采集模块采集监控现场区域图像，所述视频处理模块根据具体动态情况判断监控现场区域是否有人，如有人，执行步骤 C)；否则，执行步骤 E)。

[0009] C) 所述语音接收模块收集监控现场声音并进行判断是否触发预警，如触发预警，将预警信息传送到所述监控中心计算机，执行步骤 D)；否则，返回步骤 B)。

[0010] D) 所述监控中心计算机收到预警信息后，调出现场监控画面，并由值班人员判断是否为异常情况，如是，启动所述防爆警示模块；否则，执行步骤 E)。

[0011] E) 结束本次监控并返回步骤 B) 进行下一次监控。

[0012] 在本发明所述的采用基于智能视频的异常语音监控系统的监控方法中，在所述步骤 B) 之前还包括步骤：

A) 建立异常语音模板识别库。

[0013] 在本发明所述的采用基于智能视频的异常语音监控系统的监控方法中，所述异常语音模板识别库是依次经过语音样本采集、样本预处理、MFCC（梅尔倒谱系数，Mel Frequency Cepstrum Coefficient，简写 MFCC）特征提取和支持向量机训练步骤得到。

[0014] 在本发明所述的采用基于智能视频的异常语音监控系统的监控方法中，所述步骤 C) 进一步包括：

C1) 所述语音接收模块收集监控现场声音，将声压模拟信号经过增益放大、模数转换为数字信号并将所述数字信号传送到所述智能监控分析装置；

C2) 用所述中心处理器中的 FIR(Finite Impulse Response, 有限长单位冲激响应) 300Hz ~ 3.4KHz 带通滤波器对所述数字信号进行滤波，提取人的 300Hz ~ 3.4KHz 频率范围内的声音信号；

C3) 由 $P=20\log_{10}(x)$ 计算声强值；其中，P 为声强，其单位为分贝，x 为输入声压数据；

C4) 所述中心处理器依据得到的声强值波形特点，对非语音类声音进行滤除；

C5) 判断声强是否大于预设的声强触发阈值，如是，截取声音并执行步骤 C6)；否则，返回步骤 B)；

C6) 判断声强是否大于预设的声强预警阈值，如是，向监控中心计算机发出声强预警，并启动异常语音识别模块工作；否则，返回步骤 B)；

C7) 提取输入的可疑语音数据的 MFCC 特征参数；

C8) 判断所述提取的 MFCC 特征参数与所述异常语音模板识别库是否匹配，如匹配，向监控中心计算机发出异常语音预警；否则，返回步骤 B)。

[0015] 在本发明所述的采用基于智能视频的异常语音监控系统的监控方法中，所述步骤 C4) 中对非语音类声音进行滤除为用 FIR 2.5KHz ~ 3KHz 带阻滤波器对车鸣笛声进行滤除，所述中心处理器通过检测连续数据长度并将拍掌声、敲击声滤除。

[0016] 在本发明所述的采用基于智能视频的异常语音监控系统的监控方法中，所述步骤 C5) 中的声强触发阈值为 70 分贝。

[0017] 在本发明所述的采用基于智能视频的异常语音监控系统的监控方法中,所述步骤C6)中的声强预警阈值为105分贝。

[0018] 在本发明所述的采用基于智能视频的异常语音监控系统的监控方法中,所述步骤C8)中的判断是通过支持向量机进行判断并得到识别结果。

[0019] 实施本发明的基于智能视频的异常语音监控系统及方法,具有以下有益效果:由于利用视频处理模块检测到现场有人之后才启动语音接收装置,避免了监控区域在无人状态下因噪声影响而发生误报;由于语音分析包含声强检测模块与异常语音识别模块,声强检测模块只检测语音声强,对汽车鸣笛、拍掌和一些敲击等各种环境噪声进行了滤除,减少了外界噪声的干扰;由于先进行声强检测,对监控区域监听可疑的呼喊声和尖叫声进行声强预警,以防识别错误而造成漏报;触发声强预警后再进行异常语音检测,具有声强预警和异常语音预警双重预警的功能,信息更丰富,提高报警正确率,最大程度上减少漏报和误报情况的发生;双重预警为监控值班人员提供更多有用信息,提高监控效率和预警性能;所以其智能化较高、实时性较强、误报率较低。

附图说明

[0020] 图1是本发明基于智能视频的异常语音监控系统及方法实施例中系统的结构示意图。

[0021] 图2是所述实施例中智能监控分析装置的结构示意图。

[0022] 图3是所述实施例中方法的流程图。

[0023] 图4是所述实施例中监控现场声音是否触发预警判断的流程图。

[0024] 图5是所述实施例中异常语音检测的原理图框图。

具体实施方式

[0025] 为了便于本领域的普通技术人员能够理解并实施本发明,下面将结合附图对本发明实施例作进一步说明。

[0026] 在本发明基于智能视频的异常语音监控系统及方法实施例中,其系统的结构示意图如图1所示,在图1中,该系统包括监控区域现场装置1、智能监控分析装置2和监控中心计算机3;其中,智能监控分析装置2与监控区域现场装置1连接,监控中心计算机3分别与监控区域现场装置1和智能监控分析装置2连接。监控区域现场装置1包括视频采集模块11、语音接收模块12和防爆警示模块13,其中,视频采集模块11用于采集监控现场区域图像,本实施例中的视频采集模块11为摄像头;语音接收模块12被智能监控分析装置2触发并用于对现场声音进行监听,本实施例中的语音接收模块12采用的是拾音器;防爆警示模块13被监控中心计算机3触发并在启动时发出报警,以威慑犯罪分子、及时制止现场犯罪行为,本实施例中的防爆警示模块13可采用射灯、报警喇叭或声光报警设备。监控中心计算机3接收现场发送的视频、语音信息以及智能监控分析装置2的分析结果,由值班人员根据具体情况作出相关处理。

[0027] 图2是上述智能监控分析装置2的结构示意图。图2中,智能监控分析装置2包括中心处理器21、数据存储模块25、视频处理模块22、声强检测模块23和异常语音识别模块24;其中,中心处理器21的输入端分别与视频采集模块11和语音接收模块12连接、其输出端与数据存储模块25连接,数据存储模块25与中心处理器21连接,中心处理器21与声强检测模块23连接,声强检测模块23与中心处理器21连接,中心处理器21与异常语音识别模块24连接,异常语音识别模块24与中心处理器21连接。

出端与监控中心计算机 3 连接,数据存储模块 25 与中心处理器 21 连接并用于存储现场活动数据,视频处理模块 22 与中心处理器 21 连接并用于根据监控现场区域具体动态情况检测到现场有人后启动上述语音接收模块 12,声强检测模块 23 与中心处理器 21 连接并用于进行声强检测,异常语音识别模块 24 被中心处理器 21 触发并用于进行异常语音检测。硬件采用的是 DSP(数字信号处理,Digital Signal Processing,简称 DSP) 嵌入式系统,特点是系统性能好,运行稳定,成本低且易于批量化生产。值得一提的是,监控中心计算机 3 可访问、下载或删除数据存储模块 25 存储的现场活动数据,方便数据管理。应当注意的是,为了减少系统误报,该异常语音监控系统利用视频处理模块 22 检测到现场有人之后才启动语音接收模块 12。视频处理模块 22 采用连续帧间差分法检测现场是否有人,具体来讲,就是通过分析帧差二值化图像的连通区域以及检测目标动态变化时间特征等方式实现,本实施例中,将背景差分法和帧间差分法结合,将动态图像中连续两帧差图像和背景差图像直接进行与操作,再将结果进行二值化处理得到运动结果,这样就加大目标信息的权重,使检测更准确。用帧间差分法检测监控现场是否有异常物体运动,结合背景差分法则可准确检测是否有人。

[0028] 本发明还涉及一种采用上述系统的监控方法,该方法的流程图如图 3 所示,在图 3 中,其方法包括如下步骤:

步骤 S101 建立异常语音模板识别库:本实施例中,监控工作之前首先要建立异常语音模板识别库,语音样本采集包括男声和女声,录音人数多于几十人比较适宜,每人录 5 个词左右,录音内容包括人遇到紧急情况发出类似“啊~”、“救命啊”、“抢劫啦”、“来人啊”等等这样的呼救声、叫喊声,还包括一些人平常交流的话语,把异常语音和正常说话的语音作为正负样本用训练器进行训练,得到异常语音模板识别库。异常语音模板识别库可以自定义录音得到,灵活性比较强,为了对非特定人语音进行识别,当然录音的人数越多越好。本实施例中,通过支持向量机进行训练得到异常语音模板识别库,关于本步骤中如何具体建立异常语音模板识别库请参见图 5,稍后也会详细描述。

[0029] 步骤 S102 采集监控现场视频信息:本步骤中,视频采集模块采集监控区域现场的图像。

[0030] 步骤 S103 判断是否有人? 本步骤中,视频处理模块 22 根据监控现场区域具体动态情况检测现场是否有人? 视频处理模块 22 采用连续帧间差分法检测现场是否有人,具体来讲,就是通过分析帧差二值化图像的连通区域以及检测目标动态变化时间特征等方式实现,本实施例中,将背景差分法和帧间差分法结合,将动态图像中连续两帧差图像和背景差图像直接进行与操作,再将结果进行二值化处理得到运动结果,这样就加大目标信息的权重,使检测更准确。用帧间差分法检测监控现场是否有异常物体运动,结合背景差分法则可准确检测是否有人。如果判断结果为有人,执行步骤 S104;否则,返回步骤 S102。

[0031] 步骤 S104 采集监控现场音频信息,并判断是否触发预警? 本步骤中,语音接收模块收集监控现场声音,并进行判断是否触发预警,如果触发预警,将预警信息传送到监控中心计算机;否则,返回步骤 S102。关于本步骤中如何具体的判断是否触发预警,请参见图 4,稍后也会加以描述。

[0032] 步骤 S105 将预警信息传送到监控中心计算机:本步骤中,上述步骤 S104 判断为触发预警,则执行本步骤,即将预警信息传送到监控中心计算机。

[0033] 步骤 S106 调出监控画面，并判断是否为异常情况？本步骤中，监控中心计算机收到预警信息后，调出现场监控画面，供值班人员分析处理并判断是否为异常情况，如果出现异常情况，执行步骤 S107；否则，返回步骤 S102。

[0034] 步骤 S107 启动防爆警示模块：如果上述步骤 S106 判断为出现异常情况，则执行本步骤即启动防爆警示模块 13，发出报警，用于威慑犯罪分子，通知相关安保人员并向警方报警，值得一提的是，执行完本步骤返回步骤 S102 继续进行下一次监控工作。

[0035] 对于本实施例而言，上述监控现场声音是否触发预警的判断步骤具体如图 4 所示，其包括：

步骤 S401 将声压模拟信号进行增益放大、模数转换后传送到智能监控分析装置：本步骤中，语音接收模块收集监控现场声音，将声压模拟信号经过增益放大、模数转换为数字信号后，并将该数字信号传送到智能监控分析装置中。

[0036] 步骤 S402 FIR 300Hz~3.4KHz 带通滤波，提取 300Hz~3.4KHz 声音：本步骤中，用中心处理器中的 FIR300Hz~3.4KHz 带通滤波器对数字信号进行滤波，提取人的 300Hz~3.4KHz 频率范围内的声音信号。

[0037] 步骤 S403 计算声强值：本步骤中，由公式 $P=20\log_{10}(x)$ 计算声强值；其中，P 为声强，其单位为分贝，x 为输入声压数据。

[0038] 步骤 S404 依据声强值波形特点，对非语音类声音进行滤除：在本步骤中，中心处理器依据得到的声强值波形特点，对非语音类声音进行滤除；具体来讲，本步骤中，针对不同情况的非语音类声音采用不同的方法进行滤除，由于车鸣笛声频率主要集中在 2.5KHz~3KHz 频段内，再加上人在该频段内的分量较少，所以本步骤用 FIR2.5KHz~3KHz 带阻滤波器对车鸣笛声进行滤除，也即阻止 2.5KHz~3KHz 频段内声音通过；对于拍掌声、敲击声音等一些断断续续的声音，本步骤中不需要滤波器进行滤除，中心处理器直接根据这些声音的特点将其识别出来，进而进行滤除，具体来讲，拍掌声、敲击声音分布在 0~4KHz 频率范围内（抽样率为 8000，即模数转换时，每秒钟从模拟信号中提取 8000 个数据），其中 1.2KHz 以下分量较多，此频段跟人说话的频段有重叠，这时，用滤波器不能进行滤除，但可根据拍掌声、敲击声音不连续的特点，通过检测连续数据长度将拍掌声、敲击声音进行滤除，本实施例中，一般检测到的语音声强值大于 70 分贝的连续数据长度为 0.5s，而检测到的拍掌声、敲击声音的连续数据长度小于 0.1s，中心处理器根据这个区别，可将拍掌声、敲击声音进行滤除。本步骤中，对于非语音类声音与正常语音（人的说话声）叠加在一起的情况，因为这时非语音类声音与正常语音的频率在同一频段，要具体情况具体分析；如果非语音类声音的能量很小，这样不影响结果，可以不考虑，但如果非语音类声音的能量跟正常语音的差不多，这时只能将其识别为正常语音处理，但不会影响声强预警。

[0039] 步骤 S405 判断声强是否大于预设的声强触发阈值？本步骤中，预设的声强触发阈值为 70 分贝，如果现场声音声强大于 70 分贝，则截取声音并对声音进行检测；如果声强不超过 70 分贝，则不可能有异常情况出现，这时不需要对声音进行检测，直接返回步骤 S102。

[0040] 步骤 S406 截取声音：因为检测的声音是连续的，本步骤把有人说话的这一段声音截取下来。

[0041] 步骤 S407 判断声强是否大于预设的声强预警阈值？本步骤中，预设的声强预警

阈值为 105 分贝,如果现场声音声强大于 105 分贝,执行步骤 S408 ;否则,返回步骤 S102。

[0042] 步骤 S408 向监控中心计算机发出声强预警,并启动异常语音识别模块工作:本步骤中,如果上述步骤 S407 中的判断结果为声强大于预设的声强预警阈值,则向监控中心计算机发出声强预警,并启动异常语音识别模块工作,本实施例中,声强预警为橙色预警,即为一般预警。值得一提的是,触发声强预警后才检测异常语音,此方法丰富了预警信息。

[0043] 步骤 S409 提取输入语音数据的 MFCC 特征参数:本步骤中,提取输入的可疑语音数据的 MFCC 特征参数。

[0044] 步骤 S410 判断提取的 MFCC 特征参数是否与异常语音模板库匹配? 本步骤中,判断提取的 MFCC 特征参数与异常语音模板识别库是否匹配,如果匹配,执行步骤 S411 ;否则,返回步骤 S102。值得一提的是,本步骤中是通过支持向量机进行判断并得到识别结果,识别时提取输入语音的 MFCC 特征,然后与异常语音模板识别库进行匹配得到输出结果,该方法可实现非特定人语音识别。

[0045] 步骤 S411 向监控中心计算机发出异常语音预警:如果上述步骤 S410 判断结果为匹配,则执行本步骤向监控中心计算机发出异常语音预警,同时数据存储模块储存相关视频数据和语音数据;本实施例中的异常语音预警为红色预警,即为紧急预警。

[0046] 图 5 是本实施例中异常语音检测的原理图框图。本实施例中的异常语音模板识别库是依次经过语音样本采集、样本预处理、MFCC 特征提取和支持向量机训练步骤得到。在图 5 中,语音样本采集得到的训练数据依次经过训练数据预处理 1001、训练数据特征提取 1002 和模板训练 1003 得到异常语音模板识别库 1004 ;其中,训练数据预处理 1001 包括 FIR 300Hz~3. 4KHz 带通滤波、归一化、端点检测、预加重、分帧和加汉明窗, FIR 300Hz~3. 4KHz 带通滤波把 300Hz~3. 4KHz 频率范围的声音提取出来,将 300Hz~3. 4KHz 频率以外的一些背景噪声进行滤除;因为每次说话时,声音的大小都是不同的,即使对于同一句话,其声音的大小都是不同的,所以进行了归一化处理,有助于后续处理数据的方便;端点检测就是检测声音开始与结束的位置。训练数据特征提取 1002 就是提取训练数据的 MFCC 特征,MFCC 特征提取包括如下步骤:

快速傅立叶变换(Fast Fourier Transformation, FFT):将时域信号变换成为信号的功率谱;

三角窗滤波:用一组 Mel (梅尔)频标上线性分布的三角窗滤波器(共 24 个三角窗滤波器),对信号的功率谱滤波,每一个三角窗滤波器覆盖的范围都近似于人耳的一个临界带宽,以此来模拟人耳的掩蔽效应;

求对数:三角窗滤波器组的输出求取对数,可以得到近似于对数变换的结果;

离散余弦变换(Discrete Cosine Transformation, DCT):去除各维信号之间的相关性,将信号映射到低维空间;

谱加权:由于倒谱的低阶参数易受说话人特性、信道特性等的影响,而高阶参数的分辨能力比较低,所以需要进行谱加权,抑制其低阶和高阶参数;

倒谱均值减(Cepstrum Mean Subtraction, CMS):CMS 可以有效地减小语音输入信道对特征参数的影响;

差分参数:大量实验表明,在语音特征中加入表征语音动态特性的差分参数,能够提高系统的识别性能;本实施例中也用到了 MFCC 参数的一阶差分参数和二阶差分参数;

短时能量：语音的短时能量也是重要的特征参数，本实施例中采用了语音的短时归一化对数能量及其一阶差分、二阶差分参数。

[0047] 值得一提的是，由于训练只需要一维固定长度的数据，而 MFCC 特征提取出来的是二维矩阵数据，所以将提取的二维特征数据转成一维数据后再进行后续的模板训练 1003，本实施例中的模板训练 1003 采用支持向量机训练得到异常语音模板识别库 1004。

[0048] 图 5 中，监测数据依次经过检测数据预处理 2001、声强检测 2002、声强预警 2003 和监测数据特征提取 2004，由异常语音识别 2005 判断提取的监测数据的特征参数是否与异常语音模板识别库 1004 匹配；其中，检测数据预处理 2001 包括 FIR 300Hz~3.4KHz 带通滤波、声音截取、长度检测、归一化、端点检测、预加重、分帧和加汉明窗；由于检测的声音是连续的，所以使用声音截取把有人说话的这一段声音截取下来；值得一提的是，由于检测数据是实际要检测的声音数据，其长度是不定的，所以通过长度检测去掉过长或过短的声音，只保留 0.4s~3s 的声音长度，而上述训练数据的长度是通过人工挑选的，所以不用进行长度检测；声强检测用于检测声强是否超过预设的声强触发阈值，如果超过预设的声强触发阈值，则进一步检测声强是否超过预设的声强预警阈值，否则不用进行后续步骤；如果声强超过预设的声强预警阈值，则进行声强预警 2003，并启动异常语音识别模块工作；监测数据特征提取 2004 就是提取监测数据的 MFCC 特征，同样将提取的二维特征数据转成一维数据；异常语音识别 2005 判断提取的监测数据的 MFCC 特征参数是否与异常语音模板识别库 1004 匹配，如果匹配，则进行异常语音预警 2006；否则不进行动作。

[0049] 总之，在本实施例中，视频采集模块 11 采集现场图像，视频处理模块 22 对监控区域是否有人进行检测，若有人，则启动语音接收模块 12 对现场声音进行监听，智能监控分析装置 2 将采集的声音数据进行处理和判断，如果有异常语音触发了预警，一方面将相关语音数据存储，另一方面向监控中心计算机 3 发出预警信息，调出现场监控画面，由监控值班人员做出启动防爆警示模块 13、通知安保人员等相关预警处理。本实施例使用双重预警即声强预警和异常语音预警（异常情况预警），声强预警是环境中声音强度达到一定程序就会发出预警，异常语音预警是检测环境中如果出现异常声音（如：救命啊，来人啊等）就会发出预警。在现实当中，如果出现异常声音，这些声音的特点都是比较紧急，而且比较大声，声强预警能够检测到这种情况。双重预警就是能通过声强预警检测到上述特点的前提下，再进行异常情况检测，这样就大大降低了误报率。由于结合智能视频分析的优势，利用声强检测和语音识别技术，建立了一种立体式的安全监控方法，其具有智能化较高、实时性较强和误报率较低等优点。该监控系统发挥了视频监控和异常语音监控的各自优势，将“视”和“听”的功能相结合，建立起一个立体式的安防系统。该监控系统以异常语音监控为主，视频监控为辅，可以作为独立的系统对安防区域进行监控，也可以和其它智能监控系统结合使用。

[0050] 以上所述实施例仅表达了本发明的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变形和改进，这些都属于本发明的保护范围。因此，本发明专利的保护范围应以所附权利要求为准。

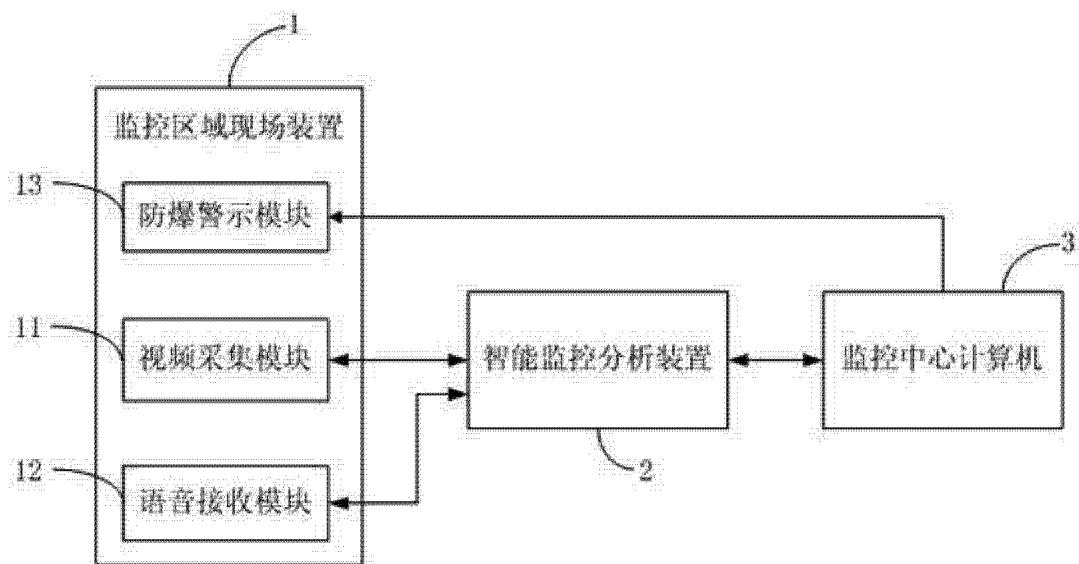
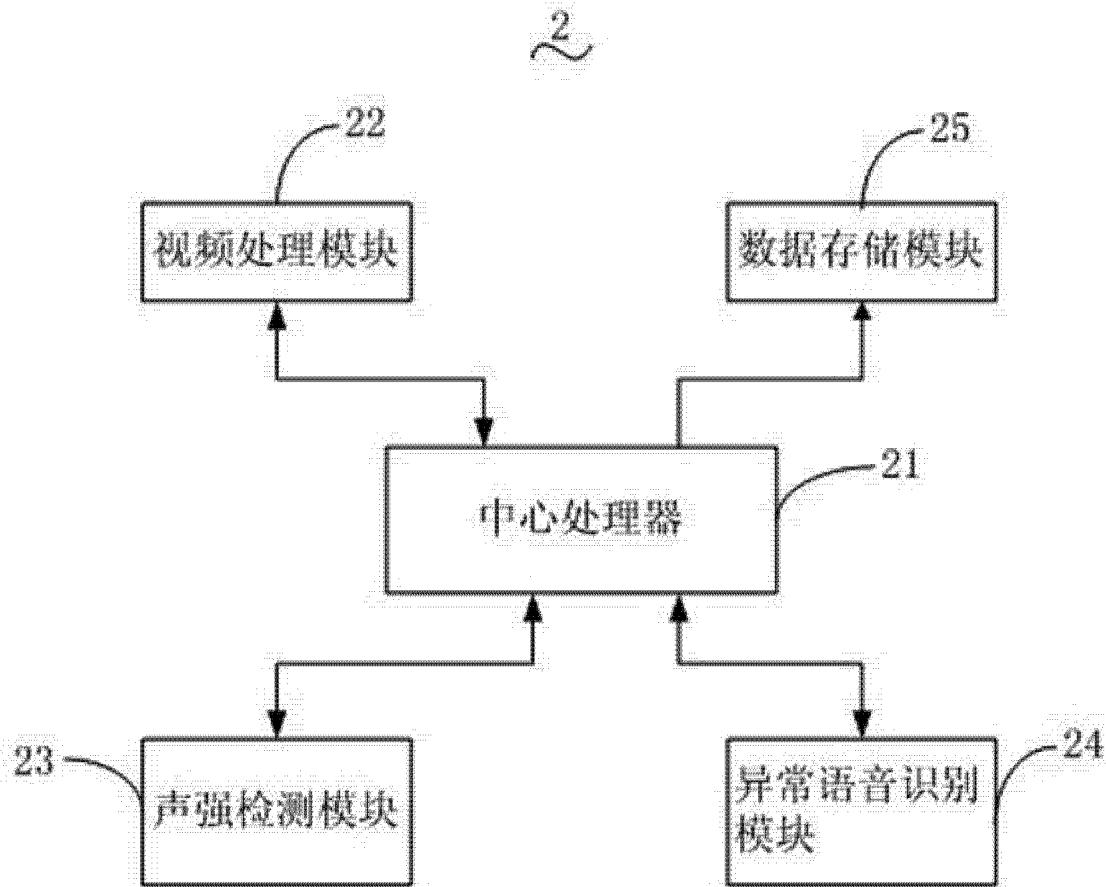


图 1



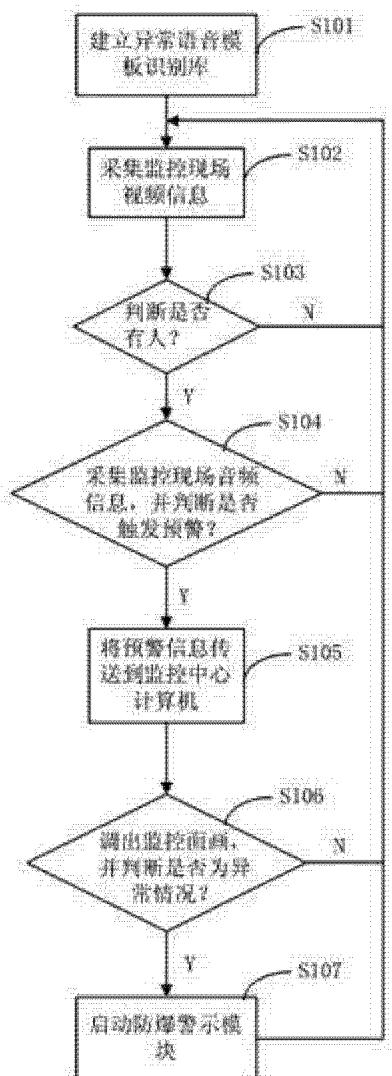


图 3

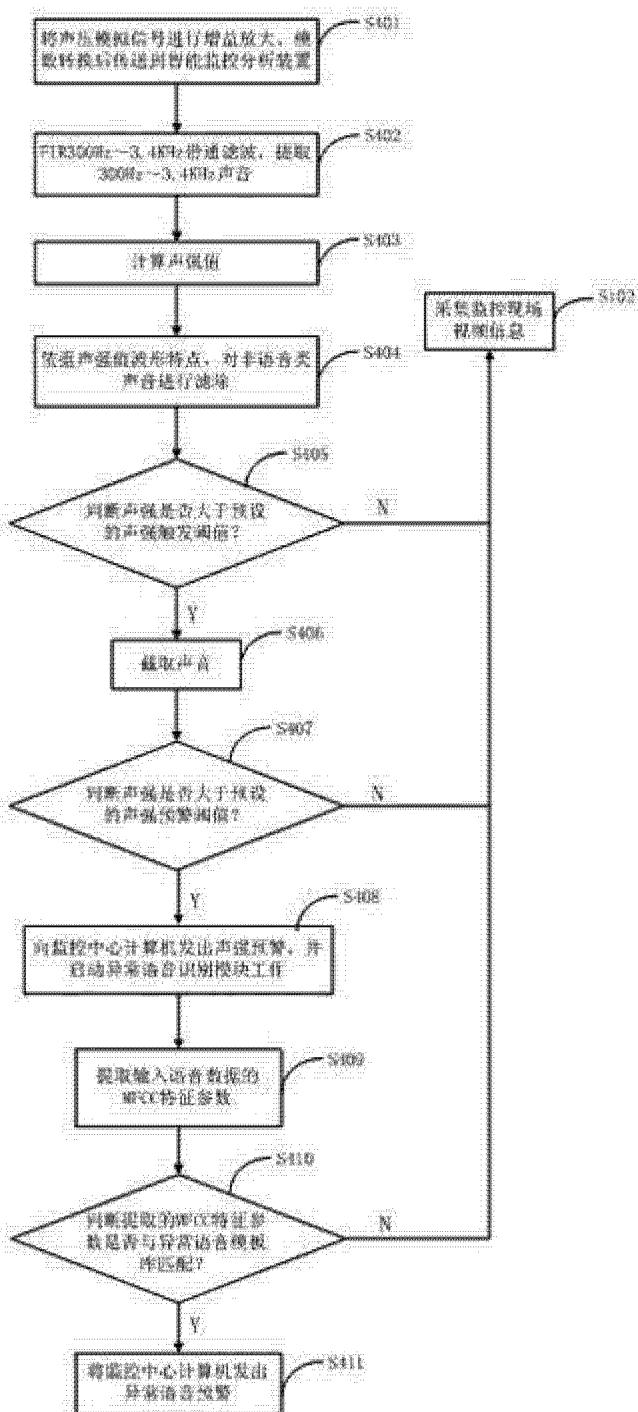


图 4

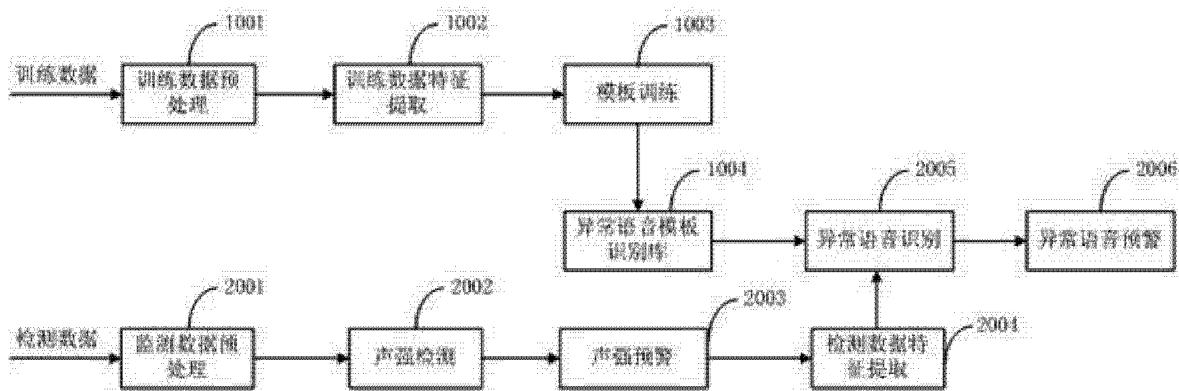


图 5