



(12)发明专利

(10)授权公告号 CN 104268594 B

(45)授权公告日 2017.12.19

(21)申请号 201410493959.3

(22)申请日 2014.09.24

(65)同一申请的已公布的文献号
申请公布号 CN 104268594 A

(43)申请公布日 2015.01.07

(73)专利权人 中安消技术有限公司
地址 100190 北京市海淀区科学院南路2号
融科资讯中心C座北楼17层12-13
专利权人 北京大学

(72)发明人 田永鸿 陈峥莹 王耀威 黄铁军

(74)专利代理机构 北京同达信恒知识产权代理
有限公司 11291

代理人 黄志华

(51)Int.Cl.

G06K 9/66(2006.01)

(56)对比文件

CN 102938058 A,2013.02.20,
CN 102164270 A,2011.08.24,
US 2006/0067562 A1,2006.03.30,
Yang Congl et.al.Sparse

Reconstruction Cost for Abnormal Event
Detection.《2011 IEEE CONFERENCE ON
COMPUTER VISION AND PATTERN RECOGNITION》
.2011,第3449-3456页.

独大为.拥挤场景下视频异常事件检测技术
研究.《中国优秀硕士学位论文全文库》.2014,
(第1期),第1138-1874页.

审查员 杨婷

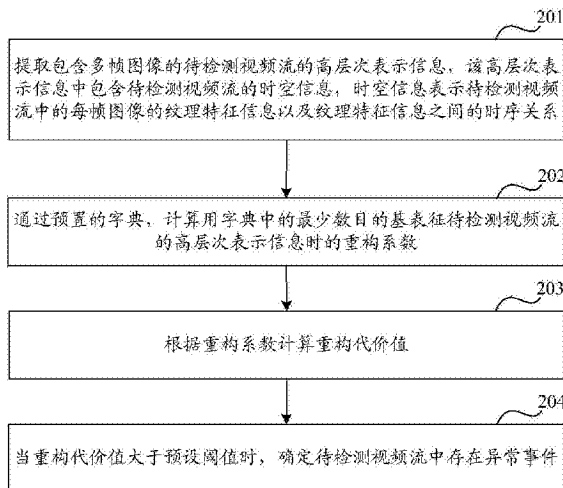
权利要求书3页 说明书12页 附图3页

(54)发明名称

一种视频异常事件检测方法及装置

(57)摘要

本发明提供一种视频异常事件检测方法和装置,所述方法包括:提取包含多帧图像的待检测视频流的高层次表示信息,该高层次表示信息中包含所述待检测视频流的时空信息;通过预置的字典,计算用所述字典中的最少数目的基表征所述待检测视频流的高层次表示信息时的重构系数;根据所述重构系数计算重构代价值;当所述重构代价值大于预设阈值时,确定所述待检测视频流中存在异常事件;当所述重构代价值小于或等于预设阈值时,确定所述待检测视频流中不存在异常事件。本发明提供的方法特征表达能力强,能够很好的描述异常事件,提高视频异常事件检测的效率和准确性。



1. 一种视频异常事件检测方法,其特征在于,所述方法包括:

提取包含多帧图像的待检测视频流的高层次表示信息,该高层次表示信息中包含所述待检测视频流的时空信息,所述时空信息表示所述待检测视频流中的每帧图像的纹理特征信息以及所述纹理特征信息之间的时序关系;

通过预置的字典,计算用所述字典中的最少数目的基表征所述待检测视频流的高层次表示信息时的重构系数,所述字典根据以下方式建立并不断更新:将不存在异常事件的视频流作为训练样本,利用卷积自编码器对所述训练样本的每帧图像进行降维处理,提取每帧图像的纹理特征,然后基于所述训练样本中每帧图像的时序关系,对提取的纹理特征进行时空建模,获取所述训练样本的高层次表示信息,其中,所述训练样本的高层次表示信息中包含所述训练样本的所有的正常特征;由所述训练样本的高层次表示信息构成一个特征池,对所述特征池进行优化,获得基的数量小于第一预设阈值,且重构误差小于第二预设阈值的字典,其中,所述字典中的每一个正常特征为一个基;

根据所述重构系数计算重构代价值;

当所述重构代价值大于预设阈值时,确定所述待检测视频流中存在异常事件。

2. 根据权利要求1所述的方法,其特征在于,所述字典中的每一个基对应一个权重系数,其中,根据权重公式确定所述权重系数的初始值,所述权重公式为:

$$w_i^0 = 1 - \frac{\beta_i^0}{\|\beta^0\|_1}, \quad \text{其中所述 } \beta_i^0 = \|X_i^0\|_2$$

其中,所述 w_i^0 表示字典中第 i 基的初始权重系数;所述 β_i^0 表示所述字典中每个基对应系数的累加和;所述 X_i^0 表示初始选择向量的第 i 维;所述 β^0 表示由 β_i^0 组成的矩阵;所述 $\|\beta^0\|_1$ 表示取L1范数;所述 $\|X_i^0\|_2$ 表示取L2范数。

3. 根据权利要求2所述的方法,其特征在于,所述通过预置的字典,计算用所述字典中的最少数目的基表征所述待检测视频流的高层次表示信息时的重构系数,包括:

通过预置的字典,根据重构系数求解公式计算用所述字典中的最少数目的基,表征所述待检测视频流的高层次表示信息时的重构系数;

其中,所述重构系数求解公式为:

$$x^* = \arg \min_x \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|Wx\|_1$$

其中, x^* 表示重构系数; y 表示所述待检测视频流的包含时空信息描述的高层次表示;所述 Φ 表示所述字典;所述 W 表示所述权重系数构成的矩阵, x 表示权重向量; λ 表示惩罚系数。

4. 根据权利要求2所述的方法,其特征在于,所述方法还包括:

当确定所述待检测视频流中不存在异常事件时,根据所述待检测视频流的重构系数和权重系数更新公式更新目标基的权重系数,其中,所述目标基为按照从大到小的顺序排列,排列靠前的前 K 个权重系数对应的基;

其中权重系数更新公式为:

$$w_i^{t+1} = 1 - \frac{\beta_i^t + x_i^*}{\|\beta^{t+1}\|_1}, \quad \text{其中, } \beta_i^t = \|X_i^t\|_2; \quad \beta_i^{t+1} = \|X_i^{t+1}\|_2$$

其中,所述 w_i^{t+1} 表示第i个基第t+1次迭代后的权重系数;所述 x_i^* 表示用所述字典表征所述待检测视频流时,第i个基对应的重构系数值;所述t表示第t次迭代;所述t+1表示第t+1次迭代,所述 β^{t+1} 表示由 β_i^{t+1} 构成的矩阵;所述X表示选择向量;所述 X_i 表示选择向量X中的第i个选择向量。

5. 一种视频异常事件检测装置,其特征在于,所述装置包括:

高层次表示提取模块,用于提取包含多帧图像的待检测视频流的高层次表示信息,该高层次表示信息中包含所述待检测视频流的时空信息,所述时空信息表示所述待检测视频流中的每帧图像的纹理特征信息以及所述纹理特征信息之间的时序关系;

重构系数计算模块,用于通过预置的字典,计算用所述字典中的最少数目的基表征所述待检测视频流的高层次表示信息时的重构系数,所述字典根据以下方式建立并不断更新:将不存在异常事件的视频流作为训练样本,利用卷积自编码器对所述训练样本的每帧图像进行降维处理,提取每帧图像的纹理特征,然后基于所述训练样本中每帧图像的时序关系,对提取的纹理特征进行时空建模,获取所述训练样本的高层次表示信息,其中,所述训练样本的高层次表示信息中包含所述训练样本的所有的正常特征,去除所述训练样本的高层次表示信息中的冗余的正常特征后生成字典,其中,所述字典中的每一个正常特征为一个基;

重构代价值计算模块,用于根据所述重构系数计算重构代价值;

判定模块,用于当所述重构代价值大于预设阈值时,确定所述待检测视频流中存在异常事件;

其中,所述装置还包括:

字典优化模块,用于由训练样本的高层次表示信息构成一个特征池,对所述特征池进行优化,获得基的数量小于第一预设阈值,且重构误差小于第二预设阈值的字典。

6. 根据权利要求5所述的装置,其特征在于,所述字典中的每一个基对应一个权重系数,其中,根据权重公式确定所述权重系数的初始值,所述权重公式为:

$$w_i^0 = 1 - \frac{\beta_i^0}{\|\beta^0\|_1}, \quad \text{其中所述 } \beta_i^0 = \|X_i^0\|_2$$

其中,所述 w_i^0 表示字典中第i基的初始权重系数;所述 β_i^0 表示所述字典中每个基对应系数的累加和;所述 X_i^0 表示初始选择向量的第i维;所述 β^0 表示由 β_i^0 组成的矩阵;所述 $\|\beta^0\|_1$ 表示取L1范数;所述 $\|X_i^0\|_2$ 表示取L2范数。

7. 根据权利要求6所述的装置,其特征在于,所述重构系数计算模块用于通过预置的字典,根据重构系数求解公式计算用所述字典中的最少数目的基,表征所述待检测视频流的高层次表示信息时的重构系数;

其中,所述重构系数求解公式为:

$$x^* = \arg \min_x \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|Wx\|_1$$

其中, x^* 表示重构系数;y表示所述待检测视频流的包含时空信息描述的高层次表示;所述 Φ 表示所述字典;所述W表示所述权重系数构成的矩阵,x表示权重向量; λ 表示惩罚系数。

8. 根据权利要求7所述的装置,其特征在于,所述装置还包括:

更新模块,用于当确定所述待检测视频流中不存在异常事件时,根据所述待检测视频流的重构系数和权重系数更新公式更新目标基的权重系数,其中,所述目标基为按照从大到小的顺序排列,排列靠前的前K个权重系数对应的基;

其中权重系数更新公式为:

$$w_i^{t+1} = 1 - \frac{\beta_i^t + x_i^*}{\|\beta^{t+1}\|_1}, \quad \text{其中, } \beta_i^t = \|X_i^t\|_2; \quad \beta_i^{t+1} = \|X_i^{t+1}\|_2$$

其中,所述 w_i^{t+1} 表示第i个基第t+1次迭代后的权重系数;所述 x_i^* 表示用所述字典表征所述待检测视频流时,第i个基对应的重构系数值;所述t表示第t次迭代;所述t+1表示第t+1次迭代,所述 β^{t+1} 表示由 β_i^{t+1} 构成的矩阵;所述X表示选择向量;所述 X_i 表示选择向量X中的第i个选择向量。

一种视频异常事件检测方法及装置

技术领域

[0001] 本发明涉及视频分析领域,尤其涉及一种视频异常事件检测方法及装置。

背景技术

[0002] 随着视频监控设备的日益普及和人们对安防工作的日益重视,对于监控视频的分析,尤其是对视频中异常事件的自动检测的需求越来越迫切。现有技术中主要通过以下两种方法进行视频异常事件的检测:

[0003] 基于目标轨迹的检测方法:需要先进行目标的检测和跟踪,获取目标的运动轨迹,然后将目标的运动轨迹和预定义的异常事件类型进行匹配,判断目标是否异常。

[0004] 基于局部的低层次表示的检测方法:该方法中,不需要进行目标的检测和跟踪,需要预定义一些表示正常的低层次表示(low-level representation),和表示异常的低层次表示。该低层次表示例如边缘、角、点,一般为根据经验手工设计的特征。用该些预定义的低层次表示进行事件的描述以便于进行异常事件检测。例如提取待检测视频中的低层次表示信息,然后用预定义的低层次表示描述该信息,当用表示异常的低层次表示描述该信息时则确定检测出异常事件。

[0005] 然而,上述基于目标轨迹的检测方法具有很大的局限性:该方法进行异常检测的结果很大程度上依赖于跟踪获得的运动轨迹,但轨迹跟踪技术仍然不够成熟,不能提供可靠准确的运动轨迹信息;另外,需要根据目标的特征检测出目标,而目标的特征对噪声和遮掩十分敏感,若在拥挤场景下,这一方法因目标受遮掩,而影响准确和快速的确定目标;因此,该方法进行异常事件检测的效率和准确性较低。

[0006] 基于局部的低层次表示的检测方法虽然适用于拥挤场景下的异常事件检测,因其预定义的低层次表示需要手工设计,造成特征描述存在着精度不够、表达能力不够强等问题,因此该方法难以很好地描述视频中的异常事件,导致异常事件检测的效率和准确性较低。

发明内容

[0007] 本发明的目的是提供一种视频异常事件检测方法和装置,以克服相关技术中视频异常事件检测的效率和准确性较低的问题。

[0008] 本发明提供一种视频异常事件检测方法,包括:

[0009] 提取包含多帧图像的待检测视频流的高层次表示信息,该高层次表示(high-level representation)信息中包含所述待检测视频流的时空信息(spatial-temporal information),所述时空信息表示所述待检测视频流中的每帧图像的纹理特征信息以及所述纹理特征信息之间的时序关系;

[0010] 通过预置的字典,计算用所述字典中的最少数目的基表征所述待检测视频流的高层次表示信息时的重构系数,所述字典根据以下方式建立并不断更新:将不存在异常事件的视频流作为训练样本,对所述训练样本的每帧图像进行卷积编码,提取每帧图像的纹理

特征,然后基于所述训练样本中每帧图像的时序关系,对提取的纹理特征进行时空建模,获取所述训练样本的高层次表示信息,其中,所述训练样本的高层次表示信息中包含所述训练样本的所有的正常特征,去除所述训练样本的高层次表示信息中的冗余的正常特征后生成字典,其中,所述字典中的每一个正常特征为一个基;

[0011] 根据所述重构系数计算重构代价值;

[0012] 当所述重构代价值大于预设阈值时,确定所述待检测视频流中存在异常事件。

[0013] 所述去除所述训练样本的高层次表示信息中的冗余的正常特征后生成字典包括:

[0014] 由训练样本的高层次表示信息构成一个特征池,对所述特征池进行优化,获得基的数量小于第一预设阈值,且重构误差小于第二预设阈值的字典。

[0015] 本发明实施例中经过字典优化,可以获得一个能够表达训练样本的最小尺寸的字典。

[0016] 所述字典中的每一个基对应一个权重系数,其中,根据权重公式确定所述权重系数的初始值,所述权重公式为:

$$[0017] \quad w_i^0 = 1 - \frac{\beta_i^0}{\|\beta^0\|_1}, \quad \text{其中所述 } \beta_i^0 = \|X_i^0\|_2$$

[0018] 其中,所述 w_i^0 表示字典中第 i 基的初始权重系数;所述 β_i^0 表示所述字典中每个基对应系数的累加和;所述 X_i^0 表示初始选择向量的第 i 维;所述 β^0 表示由 β_i^0 组成的矩阵;所述 $\|\beta^0\|_1$ 表示取L1范数;所述 $\|X_i^0\|_2$ 表示取L2范数。

[0019] 所述通过预置的字典,计算用所述字典中的最少数目的基表征所述待检测视频流的高层次表示信息时的重构系数,包括:

[0020] 通过预置的字典,根据重构系数求解公式计算用所述字典中的最少数目的基,表征所述待检测视频流的高层次表示信息时的重构系数;

[0021] 其中,所述重构系数求解公式为:

$$[0022] \quad x^* = \arg \min_x \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|Wx\|_1$$

[0023] 其中, x^* 表示重构系数; y 表示所述待检测视频流的包含时空信息描述的高层次表示;所述 Φ 表示所述字典;所述 W 表示所述权重系数构成的矩阵, x 表示权重向量; λ 表示惩罚系数。

[0024] 所述方法还包括:

[0025] 当确定所述待检测视频流中不存在异常事件时,根据所述待检测视频流的重构系数和权重系数更新公式更新目标基的权重系数,其中,所述目标基为按照从大到小的顺序排列,排列靠前的前 K 个权重系数对应的基;

[0026] 其中权重系数更新公式为:

$$[0027] \quad w_i^{t+1} = 1 - \frac{\beta_i^t + x_i^*}{\|\beta^{t+1}\|_1}, \quad \text{其中, } \beta_i^t = \|X_i^t\|_2, \quad \beta_i^{t+1} = \|X_i^{t+1}\|_2$$

[0028] 其中,所述 w_i^{t+1} 表示第 i 个基第 $t+1$ 次迭代后的权重系数;所述 x_i^* 表示用所述字典表征所述待检测视频流时,第 i 个基对应的重构系数值;所述 t 表示第 t 次迭代;所述 $t+1$ 表示第

t+1次迭代,所述 β^{t+1} 表示由 β_i^{t+1} 构成的矩阵;所述X表示选择向量;所述 X_i 表示选择向量X中的第i个选择向量。

[0029] 本发明实施例中通过更新字典中的基的权重系数可以使字典不断更新,提高字典的表达能力,从而提高检测异常事件的准确性。

[0030] 本发明还提供一种视频异常事件检测装置,所述装置包括:

[0031] 高层次表示提取模块,用于提取包含多帧图像的待检测视频流的高层次表示信息,该高层次表示信息中包含所述待检测视频流的时空信息,所述时空信息表示所述待检测视频流中的每帧图像的纹理特征信息以及所述纹理特征信息之间的时序关系;

[0032] 重构系数计算模块,用于通过预置的字典,计算用所述字典中的最少数目的基表征所述待检测视频流的高层次表示信息时的重构系数,所述字典根据以下方式建立并不断更新:将不存在异常事件的视频流作为训练样本,对所述训练样本的每帧图像进行卷积编码,提取每帧图像的纹理特征,然后基于所述训练样本中每帧图像的时序关系,对提取的纹理特征进行时空建模,获取所述训练样本的高层次表示信息,其中,所述训练样本的高层次表示信息中包含所述训练样本的所有的正常特征,去除所述训练样本的高层次表示信息中的冗余的正常特征后生成字典,其中,所述字典中的每一个正常特征为一个基;

[0033] 重构代价值计算模块,用于根据所述重构系数计算重构代价值;

[0034] 判定模块,用于当所述重构代价值大于预设阈值时,确定所述待检测视频流中存在异常事件。

[0035] 所述装置还包括:

[0036] 字典优化模块,用于由训练样本的高层次表示信息构成一个特征池,对所述特征池进行优化,获得基的数量小于第一预设阈值,且重构误差小于第二预设阈值的字典。

[0037] 所述字典中的每一个基对应一个权重系数,其中,根据权重公式确定所述权重系数的初始值,所述权重公式为:

$$[0038] \quad w_i^0 = 1 - \frac{\beta_i^0}{\|\beta^0\|_1} \quad \text{其中所述 } \beta_i^0 = \|X_i^0\|_2$$

[0039] 其中,所述 w_i^0 表示字典中第i基的初始权重系数;所述 β_i^0 表示所述字典中每个基对应系数的累加和;所述 X_i^0 表示初始选择向量的第i维;所述 β^0 表示由 β_i^0 组成的矩阵;所述 $\|\beta^0\|_1$ 表示取L1范数;所述 $\|X_i^0\|_2$ 表示取L2范数。

[0040] 所述重构系数计算模块用于通过预置的字典,根据重构系数求解公式计算用所述字典中的最少数目的基,表征所述待检测视频流的高层次表示信息时的重构系数;

[0041] 其中,所述重构系数求解公式为:

$$[0042] \quad x^* = \arg \min_x \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|Wx\|_1$$

[0043] 其中, x^* 表示重构系数;y表示所述待检测视频流的包含时空信息描述的高层次表示;所述 Φ 表示所述字典;所述W表示所述权重系数构成的矩阵,x表示权重向量; λ 表示惩罚系数。

[0044] 所述装置还包括:

[0045] 更新模块,用于当确定所述待检测视频流中不存在异常事件时,根据所述待检测视频流的重构系数和权重系数更新公式更新目标基的权重系数,其中,所述目标基为按照从大到小的顺序排列,排列靠前的前K个权重系数对应的基;

[0046] 其中权重系数更新公式为:

$$[0047] \quad w_i^{t+1} = 1 - \frac{\beta_i^t + x_i^*}{\|\beta^{t+1}\|_1}, \quad \text{其中, } \beta_i^t = \|X_i^t\|_2; \quad \beta_i^{t+1} = \|X_i^{t+1}\|_2$$

[0048] 其中,所述 w_i^{t+1} 表示第i个基第t+1次迭代后的权重系数;所述 x_i^* 表示用所述字典表征所述待检测视频流时,第i个基对应的重构系数值;所述t表示第t次迭代;所述t+1表示第t+1次迭代,所述 β^{t+1} 表示由 β_i^{t+1} 构成的矩阵;所述X表示选择向量;所述 X_i 表示选择向量X中的第i个选择向量。

[0049] 本发明至少具有以下有益效果:通过提取训练样本的高层次表示信息,该高层次表示信息中描述了训练样本的时空信息;在异常事件检测阶段,通过基于高层次表示信息建立的字典,计算重构代价值,将重构代价值大于预设阈值的待检测视频流确定为包含异常事件的视频流。由此可见,由于高层次表示信息中包含了时空信息,使得本发明实施例中获得的高层次表示信息,相对于现有技术的低层次表示表达能力强,能够很好的描述异常事件。从而,通过本发明实施例提供的方法能够提高视频异常事件检测的效率和准确性。

[0050] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本发明。

附图说明

[0051] 图1为本发明实施例中视频异常事件检测模型的示意图;

[0052] 图2为本发明实施例中视频异常事件检测方法的示例性流程图;

[0053] 图3为本发明实施例中条件受限玻尔兹曼机的示意图;

[0054] 图4为本发明实施例中定义的基结构的示意图;

[0055] 图5为本发明实施例中视频异常事件检测装置的示意图;

[0056] 图6为本发明实施例中视频异常事件检测装置的另一示意图。

具体实施方式

[0057] 以下结合说明书附图对本发明的优选实施例进行说明,应当理解,此处所描述的优选实施例仅用于说明和解释本发明,并不用于限定本发明,并且在不冲突的情况下,本发明中的实施例及实施例中的特征可以相互组合。

[0058] 人类视觉皮层接收场所具有的特性使得人眼能够稀疏地捕获自然场景中的关键信息。换句话说,视觉系统只耗费了最少数量的神经元来感知信息。基于这一发现,稀疏表示理论通过将事件的特征提取出来组成一个字典,该字典中含有基作为表征特征的基本单元(可以理解为神经元)。进行检测时,是利用字典中的基(可以理解为激活的神经元)来表征待检测的样本。这个重构的过程,可以通过求解重构系数来完成。

[0059] 本发明实施例基于稀疏重构理论,提供一种视频异常事件检测方法及装置,本发明实施例提供的视频异常事件检测方法的模型,如图1所示:在特征提取阶段,将正常视频

流(即不存在异常事件的视频流)作为训练样本,通过卷积自编码器和CRBM(Conditional Restricted Boltzmann Machines,条件受限玻尔兹曼机)提取训练样本的高层次表示(high-level representation)信息,例如人脸识别中的鼻子特征、眼睛特征,该高层次表示信息中描述了训练样本的时空信息(spatial-temporal information),该时空信息用于检测时序上的异常事件和/或空间上的异常事件,该时序上的异常事件例如通过多帧图像辨别出正在行走的人摔倒,空间上的异常事件如在一帧图像中辨别出机动车占据了人行道;在字典选择阶段,将获得的包含时空信息描述的高层次表示信息作为一个特征池,对该特征池进行优化,获得一个能够表达训练样本的最小尺寸的字典,并为字典中的基引入对应的权重系数;在异常事件检测阶段,结合权重系数计算重构系数,并通过重构系数计算重构代价值,将重构代价值大于预设阈值的待检测视频流确定为包含异常事件的视频流;此外,在确定待检测视频流为正常事件时,还可以自动更新字典中的基的权重系数,使得字典能够不断的优化。由此可见,本发明实施例中获得的高层次表示信息,相对于现有技术的低层次表示信息表达能力强,能够很好的描述异常事件。从而,通过本发明实施例提供的方法能够提高视频异常事件检测的效率和准确性。下面对本发明实施例中视频异常事件检测方法进行详细说明。

[0060] 实施例一

[0061] 如图2所示,为本发明实施例中视频异常事件检测方法的示例性流程图,该方法包括以下步骤:

[0062] 步骤201:提取包含多帧图像的待检测视频流的高层次表示信息,该高层次表示信息中包含待检测视频流的时空信息,时空信息表示待检测视频流中的每帧图像的纹理特征信息以及纹理特征信息之间的时序关系。

[0063] 步骤202:通过预置的字典,计算用字典中的最少数目的基表征待检测视频流的高层次表示信息时的重构系数。

[0064] 字典根据以下方式建立并不断更新:将不存在异常事件的视频流作为训练样本,对训练样本的每帧图像进行卷积编码,提取每帧图像的纹理特征,然后基于训练样本中每帧图像的时序关系,对提取的纹理特征进行时空建模,获取训练样本的高层次表示信息。其中,训练样本的高层次表示信息中包含训练样本的所有的正常特征,去除训练样本的高层次表示信息中的冗余的正常特征后生成字典。其中,字典中的每一个正常特征为一个基。

[0065] 其中,在一个实施例中,步骤202中的训练样本即包括初始阶段生成字典时所使用的训练样本,也可以包括在模型生成后进行异常检测阶段中确定的用于更新字典的视频流,该用于更新字典的视频流可以为确定为正常事件的视频流。

[0066] 步骤203:根据重构系数计算重构代价值。

[0067] 步骤204:当重构代价值大于预设阈值时,确定待检测视频流中存在异常事件。

[0068] 其中,在一个实施例中,当重构代价值小于或等于预设阈值时,确定待检测视频流中不存在异常事件。

[0069] 上述步骤201-204说明了本发明实施例中进行视频异常事件检测的模型如何进行异常事件检测的,下面对本发明实施例中进行视频异常事件检测的模型的其他方面详细说明:在该模型中,建立字典需要两个阶段,一是特征提取阶段,二是字典优化选择阶段;当建立字典后,可以为字典中的每个基引入一个权重系数,并在检测阶段,结合权重系数计算待

检测视频流的重构系数；此外，该模型中当确定待检测视频流为正常事件时，还可以更新字典的基的权重系数。以下，对该模型的各个部分进行详细说明。

[0070] 一、特征提取阶段：

[0071] 该阶段主要完成对训练样本的高层次表示的提取，包括以下步骤A1-A2：

[0072] 步骤A1：将不存在异常事件的视频流数据作为训练样本，并将训练样本输入给卷积自编码器；针对每一个训练样本，通过卷积自编码器进行降维处理得到训练样本的纹理特征；并按照在训练样本中的时间先后顺序，将降维处理后的结果输入给CRBM。

[0073] 其中，在一个实施例中，将训练样本中每一帧图像预先划分为预设数量的特征图。例如将9*9像素点的图像，划分为3*3像素点的9块区域，每一区域为一个特征图。

[0074] 其中，在一个实施例中，可以通过贪婪逐层训练的方法训练卷积自编码器；其中，在训练的过程中，可以使用反向传播算法来进行最优化的求解，即求解卷积自编码器的模型参数。

[0075] 步骤A2：CRBM根据步骤A1中的降维处理后的结果进行学习，以获得训练样本的高层次表示信息，训练样本的高层次表示信息中包含了训练样本的时空信息。

[0076] 其中，经过步骤A2即实现对训练样本的纹理特征进行时空建模。

[0077] 下面对特征提取进行详细说明：

[0078] 其中，在步骤A1中：

[0079] 卷积自编码器是一种深度学习技术，本发明实施例中的卷积自编码器是通过使用全卷积操作实现对训练样本的自编码，在解码部分通过使用有效卷积操作。本发明实施例中的卷积自编码器可以包含输入层、隐含层和输出层；输入层用于引入训练样本；隐含层用于对输入层的数据进行降维处理，以便于能够得到训练样本的纹理特征；输出层用于得到最终的纹理特征并可用于对训练样本的重构。卷积自编码器使得输入的图像，在经过卷积自编码器降维处理后能够重构，并且重构的图像等同于输入的图像。

[0080] 其中，在编码部分使用全卷积操作，记为 $*$ ；在解码部分使用有效卷积操作，记为 $*$ 。其输入和重构输出关系为：

$$[0081] \quad h_j = \sigma(x * W_j + b_j) \quad (1)$$

$$[0082] \quad \hat{x} = \sigma\left(\sum_{j \in H} h_j * V_j + c\right) \quad (2)$$

[0083] 在公式(1)中， h_j 表示第j个特征图； σ 表示sigmoid函数（神经元的非线性作用函数）； x 表示输入； W_j 表示编码阶段第j个特征图的卷积核参数； b_j 表示编码阶段第j个特征图的偏置值。

[0084] 在公式(2)中， \hat{x} 表示重构输出； σ 表示sigmoid函数； h_j 为第j个特征图； V_j 表示解码阶段对应第j个特征图的卷积核参数； c 表示输出图像的偏置值。

[0085] 其中，在一个实施例中，根据上述公式(1)和(2)可以得到卷积自编码器的损失函数如公式(3)所示，该损失函数用于描述输入和重构输出之间的误差，损失函数越小代表误差越小，重构得到的图像与原始输入图像越相近：

$$[0086] \quad L(\theta) = \frac{1}{2N} \sum_{n=1}^N (x_n - \hat{x}_n)^2 \quad (3)$$

[0087] 其中,在公式(3)中, $L(\theta)$ 表示损失代价; θ 表示自编码器的模型参数; N 表示训练样本的数量; x_n 表示第 n 个训练样本; \hat{x}_n 表示第 n 个训练样本的重构输出。

[0088] 其中,在一个实施例中,卷积自编码器中还可以包括次采样层,次采样层通过采样操作实现局部的变换不变性。它是一种特殊的卷积层,由于其卷积操作(用 \otimes 表示)时感受野没有重叠,所以能实现降采样的功能。较佳的,可以使用均值次采样方法,即计算局部感受野内像素值的平均值实现降采样,如公式(4)所示:

$$[0089] \quad s_i = \frac{1}{k^2} [x_i \otimes 1(k)] \quad (4)$$

[0090] 在公式(4)中, s_i 表示第 i 个训练样本的局部感受野内像素值的平均值; k 表示感受野尺寸; x_i 表示第 i 个训练样本; $1(k)$ 代表 $k \times k$ 的全一矩阵。

[0091] 其中,在步骤A2中:

[0092] 在一个实施例中,通过训练来调整CRBM的模型参数,使之能够输出包含时空信息描述的高层次表示。

[0093] CRBM是RBM(Restricted Boltzmann Machines,受限玻尔兹曼机)中的一种。RBM是一种两向的无向图模型。在该模型中,包括可视层单元和隐含层单元,并且限制单元之间的连接只存在于不同层的单元之间,相同层的单元之间则不存在连接。在可视层单元和隐含层单元之间使用不同的条件分布假设,便可以获得不同的RBM模型。加入隐式信息后的RBM被称为CRBM。

[0094] 其中,在一个实施例中,CRBM的网络结构可以如图3所示:图中描述的是2阶的CRBM,其中阶数可以视应用情况而改变,本发明对此不做限定。

[0095] 在图3所示的CRBM的网络中包含了两类有向的连接:

[0096] (1)如标号1所示,从过去 n 个时刻的可视层单元指向当前可视层单元,可看作一种自回归(autoregressive)形式的连接。

[0097] (2)如标号2所示,从过去时刻的可视层单元指向当前隐含层单元,说明历史信息还可以通过隐含层影响动态模型。

[0098] 在图3所示的CRBM的模型中,给定 $t, t-1, \dots, t-n$ 时刻的数据,其 t 时刻的隐含层单元仍然是条件独立的,这说明增加了有向连接后的条件受限玻尔兹曼机并没有增加其推断和学习的复杂性。其中,在一个实施例中,可以使用对比散度算法来训练CRBM。需要说明的是任何可以用于训练CRBM的现有方法均适用于本发明实施例,本发明对此不做限定。

[0099] 二、字典优化选择阶段:

[0100] 该阶段用于生产字典,具体可执行为:由所有训练样本的高层次表示信息构成一个特征池,对特征池进行优化,获得基的数量小于第一预设阈值,且重构误差小于第二预设阈值的字典。

[0101] 由此可见,经过字典优化选择阶段,可以获得一个能够表达训练样本的最小尺寸的字典。

[0102] 其中,在一个实施例中,该阶段可以具体通过以下过程实现:执行步骤A2,获得的特征池,可以记为 $B = [b_1, b_2, \dots, b_k] \in R^{m \times k}$,其中 $b_i \in R^m$ 为一个训练样本的特征, k 表示训练样本的数目, R 表示维度。由于该特征池中包含了冗余的特征和噪声,需要从 B 中选择出一个尺寸最小的最优子集,可以记为 $B' \in R^{m \times n} (n < k)$ 。由该子集组成字典,使得该字典能够使本

发明实施例提供的异常事件检测模型更具有鲁棒性,并且能够提高稀疏重构时的计算效率。

[0103] 可以通过求解以下优化问题,求得最优子集 B' ,该优化问题如公式(5)所示:

$$[0104] \quad \arg \min_x : \frac{1}{2} \|B - BX\|_F^2 + \lambda_1 \|X\|_1 \quad (5)$$

[0105] 其中,在公式(5)中: B 表示特征池; X 表示选择向量; λ_1 表示惩罚系数(penalty)。其中, $\frac{1}{2} \|B - BX\|_F^2$ 表示取F范数的平方,代表重构误差,即选择出的最优子集 B' 与特征池 B 之间的差异程度。 F 表示取F范数, $\|X\|_1$ 代表稀疏正则化,以便于使解 X 具有稀疏特性。

[0106] 为使求得的解 X 具有稀疏一致性,即使得解 X 中包含更多元素为“0”的行,可以引入“组稀疏”(group sparsity)的限制,将上述优化问题改为如公式(6)所示的优化问题:

$$[0107] \quad \arg \min_x : \frac{1}{2} \|B - BX\|_F^2 + \lambda_1 \|X\|_{2,1} \quad (6)$$

[0108] 其中,在公式(6)中: B 表示特征池; X 表示选择向量; λ_1 表示惩罚系数。其中,在公式(6)中, $\|X\|_{2,1}$ 定义为如公式(7)所示:

$$[0109] \quad \|X\|_{2,1} := \sum_{i=1}^k \|X\|_2 \quad (7)$$

[0110] 其中,在公式(7)中, X 表示选择向量; $\|X\|_2$ 表示取L2范数。

[0111] 通过根据公式(6)和(7)得到解 X 后,便可以获得最优子集 B' :即为 X 中非0行所对应的特征的集合,并将该最优子集作为字典。

[0112] 其中,在一个实施例中,可以用一个 $m \times m$ 的单位阵对得到的字典进行扩展,即字典 $\Phi = [B', I_{m \times m}] \in R^{m \times D}$,其中 m 为特征的维度, $D = n + m$, $R^{m \times D}$ 表示构成一个 m 行 D 列的矩阵。这样的扩展可以使字典适应视频中可能会发生的形变或者非预期情况等,从而提高模型的鲁棒性。

[0113] 其中,在一个实施例中,可以先限定重构误差为小于第二预设阈值中的一固定值,根据该限定条件,通过上述公式(7)求解得到 X ,当通过 X 得到的最优子集 B' 中基的个数小于预设的第一阈值时,则确定该 B' 为最终的字典。否则,可以继续调整重构误差的值,重复求得字典,直至最终得到的字典的基的数量小于第一预设阈值,且重构误差小于第二预设阈值为止。

[0114] 当然,需要说明的是,还可以采用其他的方法达到优化特征池获得字典的目的,本发明对此不做限定。

[0115] 其中,在一个实施例中,为了处理不同场景下发生的局部异常事件(Local abnormal events, LAE)和全局异常事件(Global abnormal events, GAE),本发明实施例还定义了不同的基结构,以便于根据不同的场景和异常事件选择不同的基结构来构建模型。如图4所示,图中的每一个小方块代表一个单元,其中基结构的一个单元对应于训练样本中每一帧图像的预先划分的一块区域。例如,需要检测与空间纹理相关的异常事件时(如非机动车道出现机动车),可以选择图4中标记为(a)所示的基结构,该基结构中每一个单元按照其对应于训练样本的图像中的空间位置关系排列;需要检测与时间序列相关的异常事件时

(如异常的动作序列),可以选择图4中标记为(b)所示的基结构,该基结构中每一个单元按照其对应训练样本的时序关系排列。若需要同时检测与空间和时间信息相关的异常事件,可以选择图4中标记为(c)所示的基结构,该基结构中不同单元间之间的空间位置关系和时序关系,与各单元在训练样本对应的区域在训练样本中空间位置关系和时序位置关系对应。

[0116] 其中在一个实施例中,训练样本的每一帧可以切分为若干块,每一块对应一个单元,可以将对应每个单元训练得到的高层次表示按照基结构进行拼接,便可以得到一个用于稀疏表达的基特征。

[0117] 三、结合权重系数计算重构系数

[0118] 其中,在一个实施例中,字典中的每一个基对应一个权重系数,权重系数的初始值可以根据权重系数公式确定,其中,权重公式,如公式(8)所示:

$$[0119] \quad w_i^0 = 1 - \frac{\beta_i^0}{\|\beta^0\|_1}, \quad \text{其中 } \beta_i^0 = \|X_i^0\|_2 \quad (8)$$

[0120] 其中, w_i^0 表示字典中第*i*基的初始权重系数; β_i^0 表示字典中每个基对应系数的累加和; X_i^0 表示初始选择向量的第*i*维; β^0 表示由 β_i^0 组成的矩阵; $\|\beta^0\|_1$ 表示取L1范数; $\|X_i^0\|_2$ 表示取L2范数。

[0121] 当字典中的基具有权重系数时,步骤202计算重构系数的步骤可以执行为:通过预置的字典,根据重构系数求解公式计算用字典中的最少数目的基,表征待检测视频流的高层次表示信息时的重构系数;

[0122] 其中,重构系数求解公式如公式(9)所示:

$$[0123] \quad x^* = \arg \min_x \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|Wx\|_1 \quad (9)$$

[0124] 其中, x^* 表示重构系数; y 表示待检测视频流的包含时空信息描述的高层次表示; Φ 表示字典; W 表示权重系数构成的矩阵, x 表示权重向量; λ 表示惩罚系数。

[0125] 根据公式(9)求得重构系数之后,可以进一步根据公式(10)求得重构代价值:

$$[0126] \quad S_w = \frac{1}{2} \|y - \Phi x^*\|_2^2 + \lambda \|Wx^*\|_1 \quad (10)$$

[0127] 其中,在公式(10)中, S_w 表示重构代价值; y 表示待检测视频流的包含时空信息描述的高层次表示; Φ 表示字典; x^* 表示重构系数; λ 表示惩罚系数; W 表示权重系数矩阵。

[0128] 四、权重系数更新

[0129] 其中,在一个实施例中,当确定待检测视频流中不存在异常事件时,即确定待检测视频流为正常事件时,还可以根据该待检测视频流的重构系数更新字典中的基的权重系数,该方法包括:根据待检测视频流的重构系数和权重系数更新公式更新目标基的权重系数,其中,目标基为按照从大到小的顺序排列,取前*K*个权重系数对应的基;

[0130] 其中权重系数更新公式,如公式(11)所示:

$$[0131] \quad w_i^{t+1} = 1 - \frac{\beta_i^t + x_i^*}{\|\beta^{t+1}\|_1}, \quad \text{其中, } \beta_i^t = \|X_i^t\|_2; \quad \beta_i^{t+1} = \|X_i^{t+1}\|_2 \quad (11)$$

[0132] 其中, w_i^{t+1} 表示第 i 个基第 $t+1$ 次迭代后的权重系数; x_i^* 表示用字典表征待检测视频流时, 第 i 个基对应的重构系数值; t 表示第 t 次迭代; $t+1$ 表示第 $t+1$ 次迭代, β^{t+1} 表示由 β_i^{t+1} 构成的矩阵; X 表示选择向量; X_i 表示选择向量 X 中的第 i 个选择向量。

[0133] 其中, 在一个实施例中, 基的权重系数的自我更新可以遵循以下规则:

[0134] (1) 更新时间: 当待检测视频流判定为正常事件, 即确定待检测视频流的重构系数大于预设阈值时进行更新。

[0135] (2) 更新对象: 按照从大到小的顺序, 将重构系数排序靠前的 K 个值所对应的基的权重系数进行更新; 当然, 也可以更新所有的基的权重系数, 本发明对此不做限定。其中 K 可以是预先设定的固定的值, 例如 5, 也可以根据场景因素动态确定, 该因素例如是需要检测的异常类型, 对计算速度的要求等等。其中, K 值越大更新后的字典越能够准确的进行重构, 使得用字典检测视频异常事件越准确。

[0136] (3) 更新方式: 通过权重系数更新公式进行更新。

[0137] 通过更新字典中的基的权重系数可以使字典不断更新, 提高字典的表达能力, 从而提高检测异常事件的准确性。

[0138] 综上, 本发明实施例中通过卷积自编码器和条件受限玻尔兹曼机的级联, 能够获得高层次表示信息, 相对于现有技术的低层次表示表达能力强, 能够很好的描述异常事件。从而, 通过本发明实施例提供的方法能够提高视频异常事件检测的效率和准确性。

[0139] 基于相同的构思, 本发明实施例中还提供一种视频异常事件检测装置, 如图 5 所示, 为该装置的示意图, 该装置包括:

[0140] 高层次表示提取模块 501, 用于提取包含多帧图像的待检测视频流的高层次表示信息, 该高层次表示信息中包含待检测视频流的时空信息, 时空信息表示待检测视频流中的每帧图像的纹理特征信息以及纹理特征信息之间的时序关系。

[0141] 重构系数计算模块 502, 用于通过预置的字典, 计算用字典中的最少数目的基表征待检测视频流的高层次表示信息时的重构系数。字典根据以下方式建立并不断更新: 将不存在异常事件的视频流作为训练样本, 对训练样本的每帧图像进行卷积编码, 提取每帧图像的纹理特征, 然后基于训练样本中每帧图像的时序关系, 对提取的纹理特征进行时空建模, 获取训练样本的高层次表示信息, 其中, 训练样本的高层次表示信息中包含训练样本的所有的正常特征, 去除训练样本的高层次表示信息中的冗余的正常特征后生成字典, 其中, 字典中的每一个正常特征为一个基。

[0142] 重构代价值计算模块 503, 用于根据重构系数计算重构代价值。

[0143] 判定模块 504, 用于当重构代价值大于预设阈值时, 确定待检测视频流中存在异常事件。

[0144] 其中, 在一个实施例中, 如图 6 所示, 该装置还包括:

[0145] 字典优化模块 505, 用于由训练样本的高层次表示信息构成一个特征池, 对特征池进行优化, 获得基的数量小于第一预设阈值, 且重构误差小于第二预设阈值的字典。

[0146] 其中, 在一个实施例中, 字典中的每一个基对应一个权重系数, 其中, 根据权重公式确定权重系数的初始值, 权重公式为:

$$[0147] \quad w_i^0 = 1 - \frac{\beta_i^0}{\|\beta^0\|_1}, \quad \text{其中 } \beta_i^0 = \|X_i^0\|_2$$

[0148] 其中, w_i^0 表示字典中第 i 基的初始权重系数; β_i^0 表示字典中每个基对应系数的累加和; X_i^0 表示初始选择向量的第 i 维; β^0 表示由 β_i^0 组成的矩阵; $\|\beta^0\|_1$ 表示取 L1 范数; $\|X_i^0\|_2$ 表示取 L2 范数。

[0149] 其中, 在一个实施例中, 重构系数计算模块 502 用于通过预置的字典, 根据重构系数求解公式计算用字典中的最少数目的基, 表征待检测视频流的高层次表示信息时的重构系数;

[0150] 其中, 重构系数求解公式为:

$$[0151] \quad x^* = \arg \min_x \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|Wx\|_1$$

[0152] 其中, x^* 表示重构系数; y 表示待检测视频流的包含时空信息描述的高层次表示; Φ 表示字典; W 表示权重系数构成的矩阵, x 表示权重向量; λ 表示惩罚系数。

[0153] 其中, 在一个实施例中, 如图 6 所示, 该装置还包括:

[0154] 更新模块 506, 用于当确定待检测视频流中不存在异常事件时, 根据待检测视频流的重构系数和权重系数更新公式更新目标基的权重系数, 其中, 目标基为按照从大到小的顺序排列, 排列靠前的前 K 个权重系数对应的基;

[0155] 其中权重系数更新公式为:

$$[0156] \quad w_i^{t+1} = 1 - \frac{\beta_i^t + x_i^*}{\|\beta^{t+1}\|_1}, \quad \text{其中, } \beta_i^t = \|X_i^t\|_2; \quad \beta_i^{t+1} = \|X_i^{t+1}\|_2$$

[0157] 其中, w_i^{t+1} 表示第 i 个基第 $t+1$ 次迭代后的权重系数; x_i^* 表示用字典表征待检测视频流时, 第 i 个基对应的重构系数值; t 表示第 t 次迭代; $t+1$ 表示第 $t+1$ 次迭代, β^{t+1} 表示由 β_i^{t+1} 构成的矩阵; X 表示选择向量; X_i 表示选择向量 X 中的第 i 个选择向量。

[0158] 关于上述实施例中的装置, 其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述, 此处将不做详细阐述说明。

[0159] 本领域内的技术人员应明白, 本发明的实施例可提供为方法、系统、或计算机程序产品。因此, 本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且, 本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质 (包括但不限于磁盘存储器、CD-ROM、光学存储器等) 上实施的计算机程序产品的形式。

[0160] 本发明是参照根据本发明实施例的方法、设备 (系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器, 使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0161] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特

定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0162] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0163] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

[0164] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

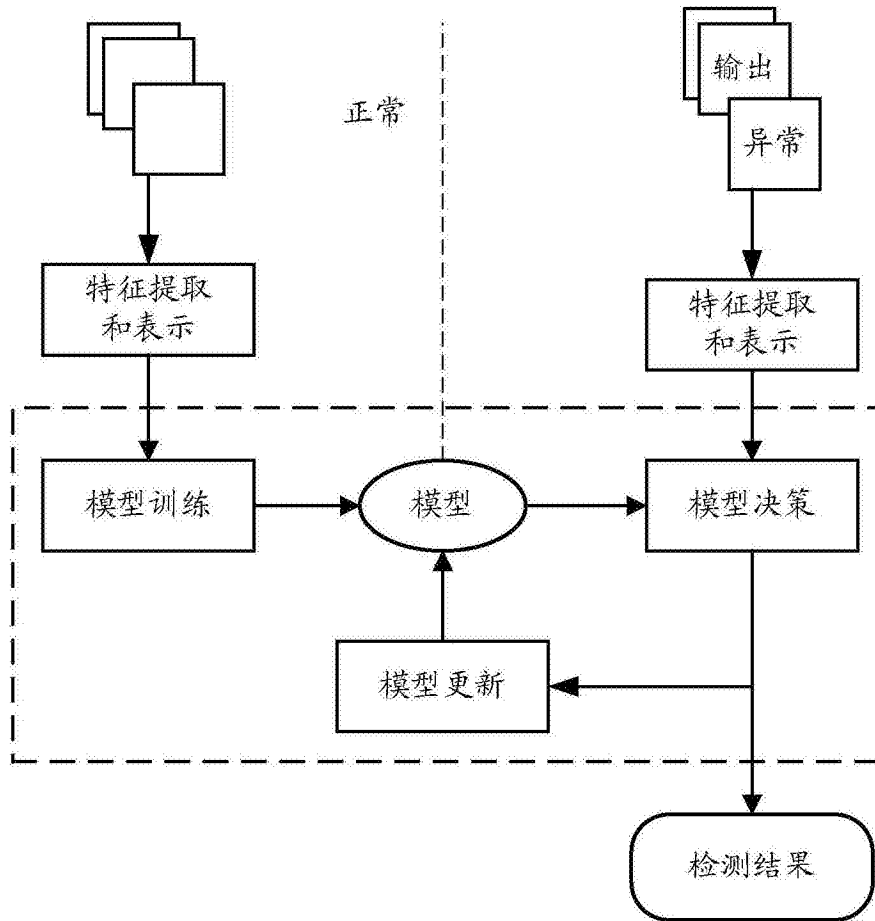


图1

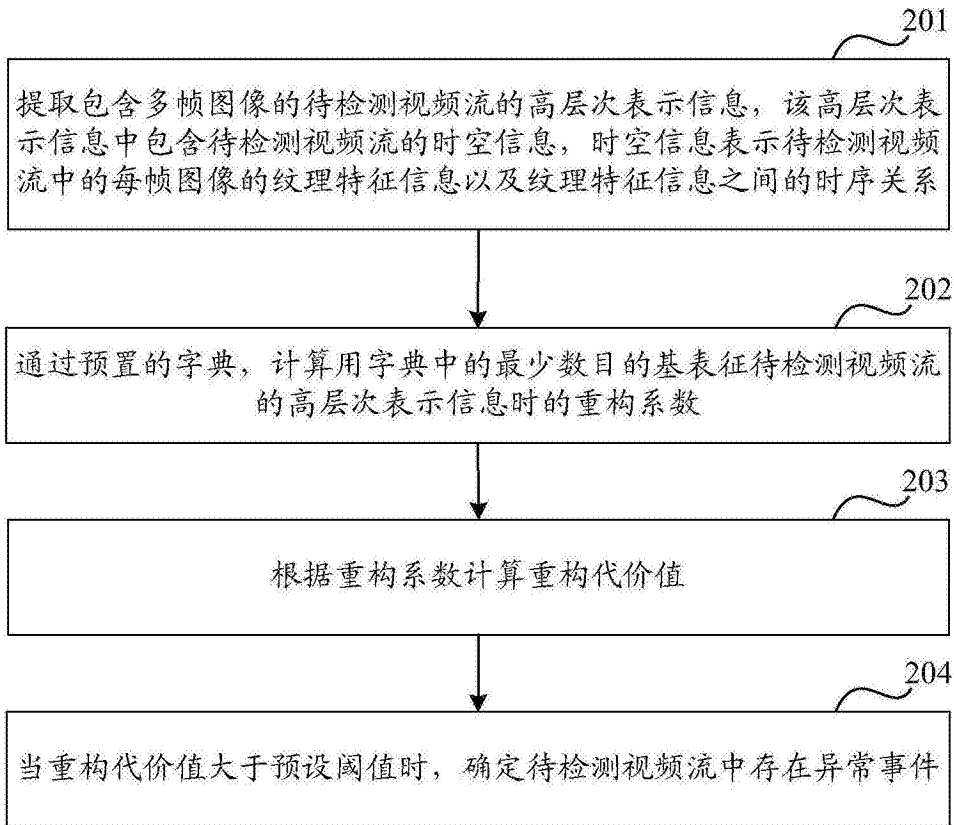


图2

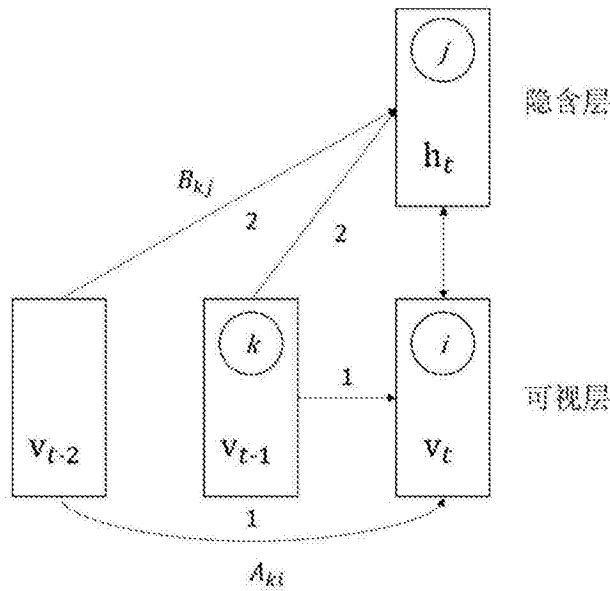


图3

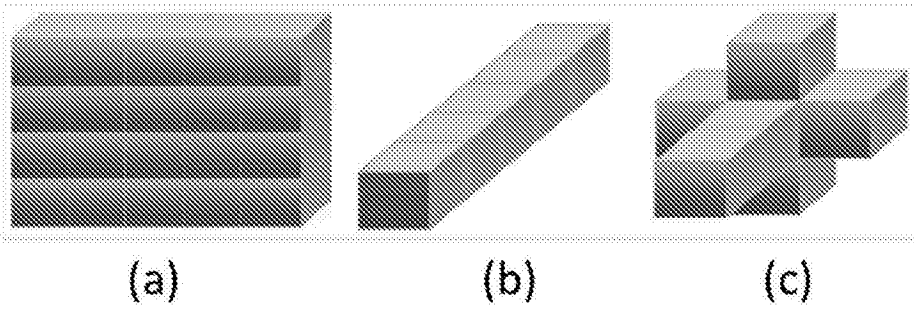


图4

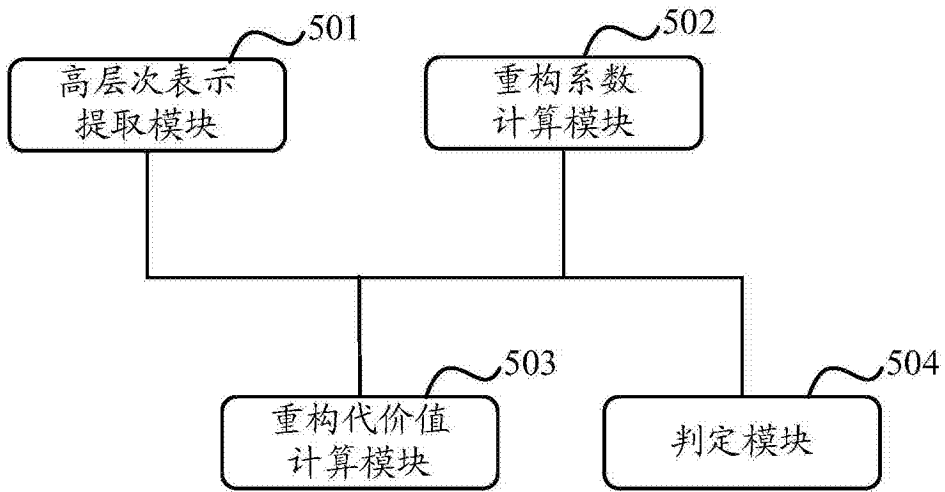


图5

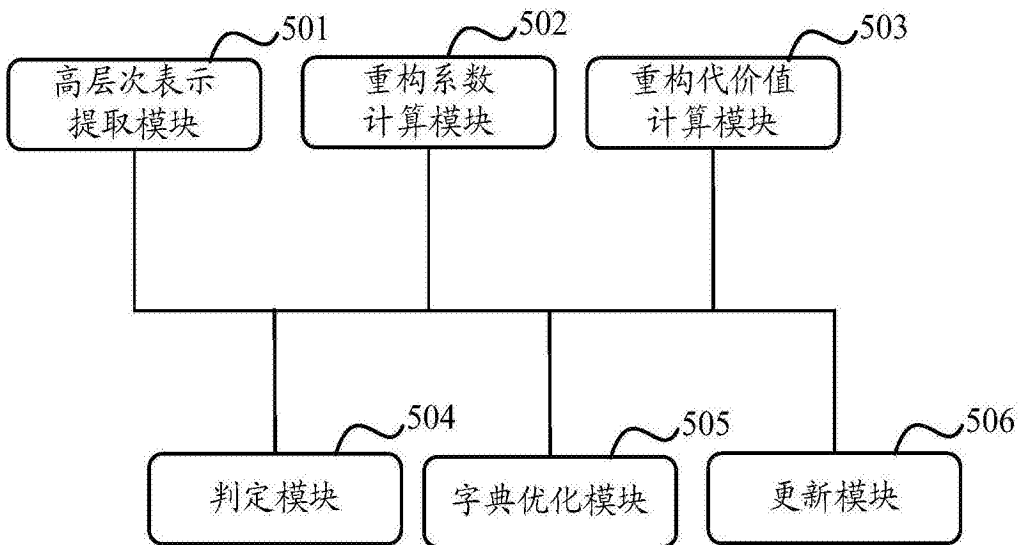


图6