(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106909220 A
(43)申请公布日 2017.06.30

(21)申请号 201710093165.1

(22)申请日 2017.02.21

(71)申请人 山东师范大学
地址 250014 山东省济南市文化东路88号

(72)发明人 孙建德 吴雪梅 李静 万文博 吴强

(74)专利代理机构 济南圣达知识产权代理有限公司 37221
代理人 李健康

(51)Int.Cl.
G06F 3/01(2006.01)
G06K 9/00(2006.01)
G06T 3/00(2006.01)

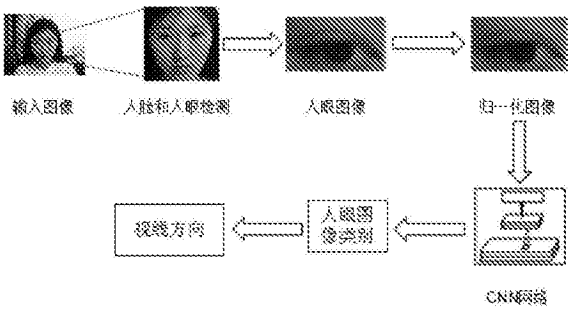
权利要求书2页 说明书4页 附图3页

(54)发明名称

一种适用于触控的视线交互方法

(57)摘要

本发明公开了一种适用于触控的视线交互方法，属于视频、多媒体信号处理技术领域。此方法将人眼区域图像输入到CNN中，通过卷积和下采样操作对图像进行特征提取和维度降低，改善了传统视线估计方法采用高清相机和外加红外设备构建眼球模型的局限性。图像采集阶段，通过对不同个体在不同位置不同时间的图像采集使得人眼数据库具备光照、肤色及形态的多样性，使得系统满足不同个体多种条件的要求。极大程度上提高了分类的准确性。



CN 106909220 A

1.一种适用于触控的视线交互方法，其特征在于：将不同个体在不同外部条件下注视同一屏幕位置块的若干人眼图像归为一类，用卷积神经网络CNN实现模式分类，从而通过人眼图像识别其对应的屏幕注视位置块，该方法包括如下步骤：

(1)人眼数据库的建立：在不同光照、不同时间、不同位置条件下对多个个体注视屏幕上随机出现的注视点的过程进行数据采集，对采集的图像进行人脸检测和人眼检测，获得人眼区域图像，建立人眼数据库并对人眼图像进行筛选；

(2)将人眼图像分为训练集和验证集，作为卷积神经网络的输入，根据实际应用需求对电脑屏幕进行分块，注视屏幕上同一分块的人眼图像视为一类，对人眼图像进行分类，训练卷积神经网络；

(3)进行视线估计时，将待分类人眼图像输入到训练好的卷积神经网络模型中，即可判断其类别，并对应得到相应的屏幕注视位置块，从而对视线方向进行估计。

2.根据权利要求1所述的适用于触控的视线交互方法，其特征在于：上述步骤(1)的具体实现步骤为：

a、人眼图像采集，其具体步骤如下：

(a)考虑到实际应用，首先将计算机屏幕分为若干大块，每个大块内部继续分成若干个小块，每个小块中间设置注视点；

(b)实验者坐在屏幕前方摄像头范围内，眼睛跟随注视点移动，同时摄像机采集人脸图像，为了防止人眼疲劳，相邻两个注视点设置时间间隔，此时摄像头不采集图像；

(c)对采集的图像进行筛选，提取可用样本数据；

b、对人脸图像进行人脸检测和人眼检测，并将人眼矩形归一化到统一大小用于训练卷积神经网络。

3.根据权利要求1所述的适用于触控的视线交互方法，其特征在于：上述步骤(2)的具体实现步骤为：

a、数据选择方面，将可用样本按一定比例分为训练集和验证集分别用于卷积神经网络训练和分类准确性的验证；

b、根据实际数据大小设置卷积神经网络结构模型为：

(a)卷积神经网络的卷积层：网络通过卷积层提取图像的深度特征，根据特征图大小选择相应的卷积核并对图像进行边缘扩充，卷积层第i层第j个特征图在(x,y)位置处的值为：

$$f_{ij}^{xy}=relu\left(b_{ij}+\sum_{n}\sum_{p=0}^{p_i-1}\sum_{q=0}^{q_i-1}w_{ijn}^{pq}f_{(i-1)n}^{(x+p)(y+q)}\right)$$

其中，relu(·)为修正线性单元，其公式为：g(x)＝max(0,x)，b_{ij}是第i层第j个特征图的偏置，n指与当前特征图连接的上一层特征图的集合，p_i,q_i分别指第i层卷积核的长、宽，w_{ijn}^{pq}是与上层特征图连接的卷积核在(p,q)处的值；

(b)卷积神经网络的采样层：网络通过采样层对图像进行降维处理，采样层第i层第j个特征图表示为

$f_{ij}=f(\beta_{ij}\mathrm{down}(f_{(i-1)j})+b_{ij})$

其中，β_{ij}和b_{ij}分别是第i层第j个特征图的乘子偏差和附加偏差，down(·)是降采样函数，采用最大池化，在下采样层的后面连接LRN层，LRN层模仿生物神经系统的侧抑制机制，对局部神经元的活动创建竞争机制，使得响应比较大的值相对更大，提高模型的泛化能力。

(c)卷积神经网络的输出层:网络通过内积层实现全连接,最后输出类别号。

一种适用于触控的视线交互方法

技术领域

[0001] 本发明涉及一种适用于触控的视线交互方法，属于视频、多媒体信号处理技术领域。

技术背景

[0002] 人机交互技术是目前用户界面研究中发展得最快的领域之一。目前比较普遍的人机交互方式包括键盘、鼠标、摄像头等传统交互方式。然而，随着人工智能技术的不断发展，很多非接触式的交互方式应运而生，如手势识别等。相对于传统的接触式交互方式，非接触式交互方式具有安全、卫生等优点，被越来越多的应用于生活的方方面面。

[0003] 本发明涉及的视线交互方法是非接触式人机交互的一种。所谓视线交互顾名思义就是通过估计人眼注视方向来实现人机交互的技术，其核心技术为视线估计方法。视线交互在虚拟现实、发展心理学、医学诊断、感知分析、商业广告测评等领域的应用日益广泛。同时，视线交互对于重度瘫痪但视觉良好的病人来说也是一个很好的辅助，他们可以通过眼球的运动来表达自己的意愿与需求，并控制相应的系统来满足自己的需求。

[0004] 传统的基于模型的视线估计技术通过高分辨率相机和红外光源获取瞳孔中心和角膜反射点的具体位置建立眼球的数学模型，并通过一定的映射关系得到人眼注视点的位置。这种方法对硬件设备要求较高、易受光照环境的影响并且需要定标过程，为视线交互设立了一个较高的门槛。而基于人眼图像的视线估计采用机器学习的方法，首先建立人眼图像数据库，利用神经网络对人眼图像特征进行降维和特征提取，在人眼图像和注视方向之间建立对应关系，从而实现根据人眼图像的注视方向估计。但是，此类方法以获得准确的注视方向为目标，在机器学习的训练过程中，往往采用回归的方法；这使得这类方法训练过程复杂，网络参数调节难度大，并且所获得的注视方向精度比较差。同时，这类方法中所使用的人眼图像数据库大多不是在注视屏幕的场景下采集的，因此，在基于屏幕的触控应用中实现并不理想。

发明内容

[0005] 基于触控的人机交互是现在消费电子产品中最常见的人机交互方式，它通过触发在有限面积的屏幕上具有一定面积的触控按钮来激活相应的功能。本发明参考触控技术的特点，针对传统的基于眼球模型和基于人眼图像的视线估计方法中难以应用于触控应用的缺点，本发明提供了一种适用于触控的视线交互方法。该方法将注视屏幕分块，以屏幕中的块模拟触控界面中的按钮，创新性地将注视屏幕不同块的人眼图像与其相应的区域位置之间的关系进行建模，借助人眼图像数据库的建立和卷积神经网络的应用，根据人眼图像估计出人眼所注视的屏幕上的块，实现基于视线交互的触控。在这个过程中，由于不需要估计准确的注视方向，只需要估计注视的块，因此在机器学习中需要采用模式分类方法，而不是回归的方法。实验证明，该方法在不同光照条件和不同个体条件下均能达到较高的识别度。

[0006] 该方法的实现需要一个普通网络摄像头，以及个人计算机一台。本发明采用的技

术方案如下。

[0007] 一种适用于触控的视线交互方法，其特征在于：将不同个体在不同外部条件下注视同一屏幕位置块的若干人眼图像归为一类，用卷积神经网络CNN实现模式分类，从而通过人眼图像识别其对应的屏幕注视位置块，该方法包括如下步骤：

[0008] (1)人眼数据库的建立：在不同光照、不同时间、不同位置条件下对多个个体注视屏幕上随机出现的注视点的过程进行数据采集，对采集的图像进行人脸检测和人眼检测，获得人眼区域图像，建立人眼数据库并对人眼图像进行筛选；

[0009] (2)将人眼图像分为训练集和验证集，作为卷积神经网络的输入，根据实际应用需求对电脑屏幕进行分块，注视屏幕上同一分块的人眼图像视为一类，对人眼图像进行分类，训练卷积神经网络；

[0010] (3)进行视线估计时，将待分类人眼图像输入到训练好的卷积神经网络模型中，即可判断其类别，并对应得到相应的屏幕注视位置块，从而对视线方向进行估计。

[0011] 更具体地，上述步骤(1)的具体实现步骤为：

[0012] a、人眼图像采集，其具体步骤如下：

[0013] (a)考虑到实际应用，首先将计算机屏幕分为若干大块，每个大块内部继续分成若干个小块，每个小块中间设置注视点；

[0014] (b)实验者坐在屏幕前方摄像头范围内，眼睛跟随注视点移动，同时摄像机采集人脸图像，为了防止人眼疲劳，相邻两个注视点设置时间间隔，此时摄像头不采集图像；

[0015] (c)对采集的图像进行筛选，提取可用样本数据；

[0016] b、对人脸图像进行人脸检测和人眼检测，并将人眼矩形归一化到统一大小用于训练卷积神经网络。

[0017] 更具体地，上述步骤(2)的具体实现步骤为：

[0018] a、数据选择方面，将可用样本按一定比例分为训练集和验证集分别用于卷积神经网络训练和分类准确性的验证；

[0019] b、根据实际数据大小设置卷积神经网络结构模型为：

[0020] (a)卷积神经网络的卷积层：网络通过卷积层提取图像的深度特征，根据特征图大小选择相应的卷积核并对图像进行边缘扩充，卷积层第i层第j个特征图在(x,y)位置处的值为：

[0021] $$f_{ij}^{xy}=relu\left(b_{ij}+\sum_{n}\sum_{p=0}^{p_i-1}\sum_{q=0}^{q_i-1}w_{ijn}^{pq}f_{(i-1)n}^{(x+p)(y+q)}\right)$$

[0022] 其中，relu(·)为修正线性单元，其公式为：g(x)＝max(0,x)，b_{ij}是第i层第j个特征图的偏置，n指与当前特征图连接的上一层特征图的集合，p_i，q_i分别指第i层卷积核的长、宽，w_{ijn}^{pq}是与上层特征图连接的卷积核在(p,q)处的值；

[0023] (b)卷积神经网络的采样层：网络通过采样层对图像进行降维处理，采样层第i层第j个特征图表示为

[0024] $f_{ij}=f(\beta_{ij}down(f_{(i-1)j})+b_{ij})$

[0025] 其中，β_{ij}和b_{ij}分别是第i层第j个特征图的乘子偏差和附加偏差，down(·)是降采样函数，采用最大池化，在下采样层的后面连接LRN层，LRN层模仿生物神经系统的侧抑制机制，对局部神经元的活动创建竞争机制，使得响应比较大的值相对更大，提高模型的泛化能

力。

[0026] (c)卷积神经网络的输出层:网络通过内积层实现全连接,最后输出类别号。

[0027] 上述基于卷积神经网络的分类机制为将已知类别人眼图像输入到配置好的CNN网络中计算网络损失,直至损失不断降低并趋于稳定。网络训练完成后将验证集输入网络中计算分为每一类的概率值,概率值最高者即为最终的类别。

[0028] 基于卷积神经网络的视线估计技术直接以人眼图像作为输入,通过学习的方式提取特征进行分类,避免了人工特征提取和眼球建模阶段,将高维人眼图像转换为低维特征图提升了分类性能,降低了实验的复杂度。通过对不同人眼图像进行训练增加了训练集的多样性,使得分类结果适用于不同个体。多分类任务中较高的准确性验证了本发明具有一定的实用价值,为视线估计技术提供了新途径。

附图说明

[0029] 图1是系统硬件图;

[0030] 图2是及数据采集过程图;

[0031] 图3是基于CNN的视线估计系统框图;

[0032] 图4是人脸区域和人眼区域提取示意图;

[0033] 图5是卷积神经网络结构图;

[0034] 图6是6分类问题训练集损失和验证集准确性曲线图;

[0035] 图7是54分类问题训练集损失和验证集准确性曲线图;

[0036] 图8是网络各层特征图。

具体实施方式

[0037] 下面结合附图对发明作进一步描述:

[0038] 本发明提供了一种基于卷积神经网络的视线估计新方法。依据本发明提出的方法,首先进行人脸数据的采集。实验所需的硬件设施包括一台普通个人电脑和一个网络摄像头。实验过程中摄像头放置于计算机屏幕上方中间。图1给出了系统的硬件图,包括一台个人电脑和一个网络摄像头。实验中首先将屏幕分为6大块,每个大块继而分为9小块并在每个小块内设置注视点。如图2,实验过程中测试者坐在屏幕前方50~60cm处,保持头部不动,眼睛随屏幕上注视点的移动而移动。同时,摄像头采集人脸图像。图3给出整个视线估计系统的流程图。以流程图所示,我们以数据库中某个个体为例对发明的基于卷积神经网络的视线估计方法的实现做详细介绍。本发明包含如下具体步骤:

[0039] 1、提取人脸区域和人眼区域

[0040] (1)图4给出人脸区域提取过程。图4(a)为在原始彩色图像上检测人脸区域。实验采用Haar特征进行人脸检测。人脸的Haar特征分类器就是一个XML文件,该文件中会描述人脸的Haar特征值,通过人脸不同区域的特征值确定人脸不同区域。

[0041] (2)在提取出的人脸区域内进一步使用Haar人眼检测提取左右眼。图4(b)为在检测到的人脸区域检测人眼,检测到的人眼图像如图4(c)所示。

[0042] (3)将提取出的人眼图像进行尺寸归一化处理,根据实验需要将人眼矩形区域归一化为40*72像素。

[0043] 2、训练集和验证集处理

[0044] 实验选定上述步骤(3)中的181440幅左眼图像，按一定比例选择训练集和验证集。将图像和对应的初始标签转换为LMDB格式用于卷积神经网络的训练。

[0045] 3、构建CNN网络

[0046] CNN网络训练采用Caffe深度学习框架。Caffe是一个清晰而高效的深度学习框架。Caffe是纯粹的C++/CUDA架构，支持命令行、Python和MATLAB接口；可以在CPU和GPU直接无缝切换。实验采用GPU训练，极大地节省了训练时间。图5给出用于分类的卷积神经网络结构图。包括若干卷积层、最大池层、全连接层实现全连接以及其他若干激励函数等。各层层数图5也给出了一个示例。实际运用时各层层数可由实验确定。

[0047] (1)CNN的卷积层：网络通过卷积层提取图像的深度特征。根据特征图大小选择相应的卷积核并对图像进行边缘扩充。卷积层第i层第j个特征图在(x,y)位置处的值为：

[0048] $$f_{ij}^{xy}=relu\left(b_{ij}+\sum_{n}\sum_{p=0}^{p_i-1}\sum_{q=0}^{q_i-1}w_{ijn}^{pq}f_{(i-1)n}^{(x+p)(y+q)}\right)$$

[0049] 其中，relu(·)为修正线性单元(Rectified linear unit)，其公式为：g(x)＝max(0,x)。b_{ij}是第i层第j个特征图的偏置，n指与当前特征图连接的上一层特征图的集合，p_i，q_i分别指第i层卷积核的长、宽，w_{ijn}^{pq}是与上层特征图连接的卷积核在(p,q)处的值。

[0050] (2)CNN的采样层：网络通过采样层对图像进行降维处理，采样层第i层第j个特征图表示为：

[0051] $f_{ij}=f(\beta_{ij}down(f_{(i-1)j})+b_{ij})$

[0052] 其中，β_{ij}和b_{ij}分别是第i层第j个特征图的乘子偏差和附加偏差，down(·)是降采样函数，采用最大池化。在下采样层的后面连接LRN层。LRN层模仿生物神经系统的侧抑制机制，对局部神经元的活动创建竞争机制，使得响应比较大的值相对更大，提高模型的泛化能力。

[0053] (3)CNN的输出层：全连接层的特征维数分别选择256、256，最后设置类别数分别实现6大类和54小类的分类任务。

[0054] 4、配置网络模型以及其他所需参数，分别进行6类问题和54类分类网络的训练。

[0055] (1)对训练过程和验证过程的输出损失和准确性进行分析，确定网络收敛情况和分类准确性。图6图7分别给出6分类问题和54类分类问题对应的训练集和验证集的损失曲线图以及验证集分类准确性图。图6和图7中，点画线表示训练时损失函数的值，三角画线表示测试时损失函数的值，星画线表示测试时得到的实现估计准确率y。随着迭代次数的增加，损失不断降低直至收敛，此时验证集分类准确性达到最高直至稳定。通过调整网络配置参数确定最优网络即为我们需要的分类网络。

[0056] (2)对卷积神经网络各层可视化，以6分类问题的某张人眼图像为例，CNN网络各特征图如图8所示。图8(a)给出各卷积层和下采样层处理后输出的特征图及相应的卷积权重，图8(b)给出经过全连接层处理后输出的特征及其直方图以及最后的所属类别。通过最后的概率图可知，概率最高的类别即为输入图像所属类别。

[0057] (3)待分类人眼图像只需输入到得到的多分类网络模型中，即可判断其类别。将人眼图像对应的类别与屏幕位置对应即可确定此人眼图像对应的大体视线方向。实验结果表明，对图像进行6分类分类准确性可达93％，54类分类准确性达83％。

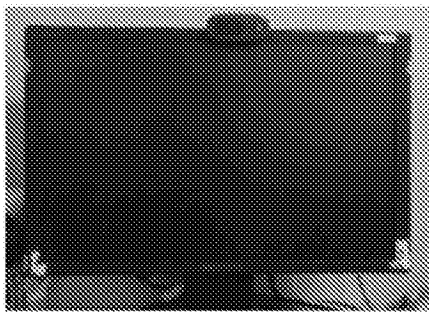


图1



图2

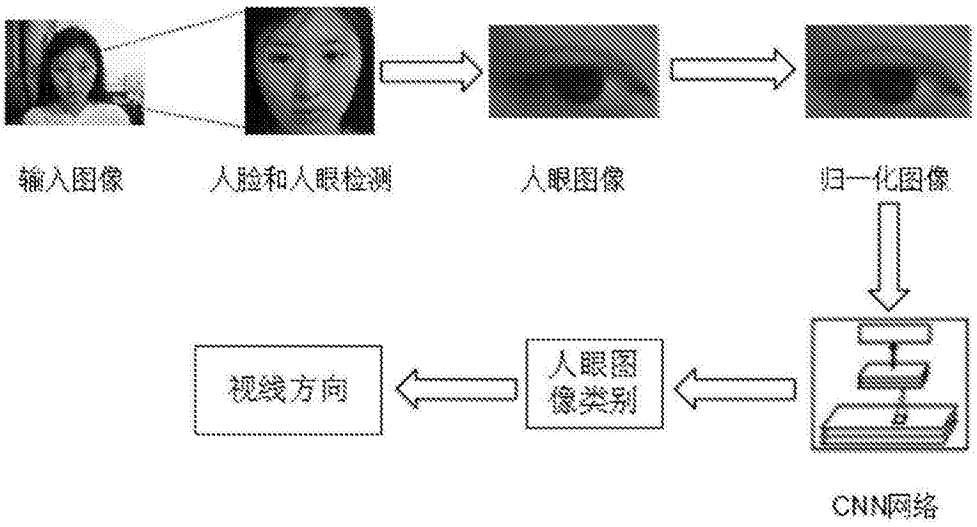


图3

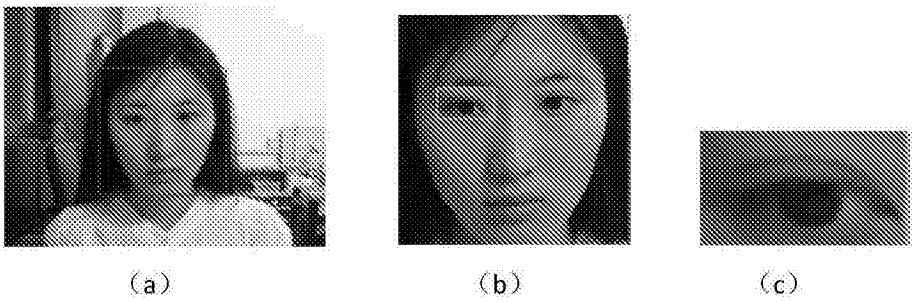

(a)
(b)
(c)

图4

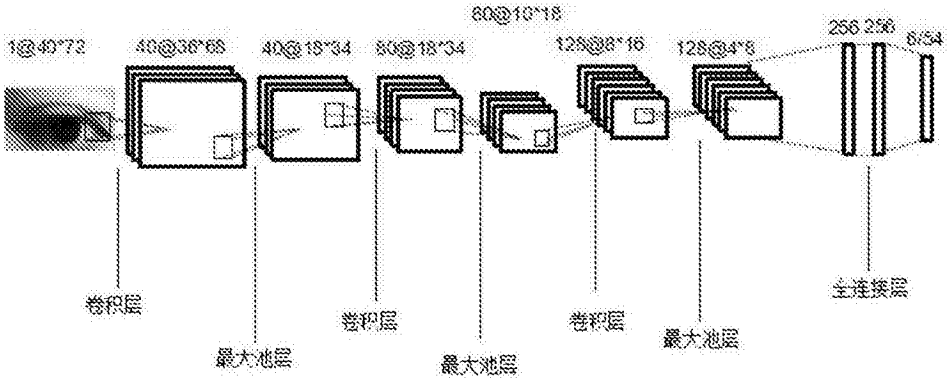

1@40*72
40@36*68
40@18*34
80@18*34
80@10*18
128@8*16
128@4*8
256 256
8*34
卷积层
最大池层
卷积层
最大池层
卷积层
最大池层
全连接层

图5

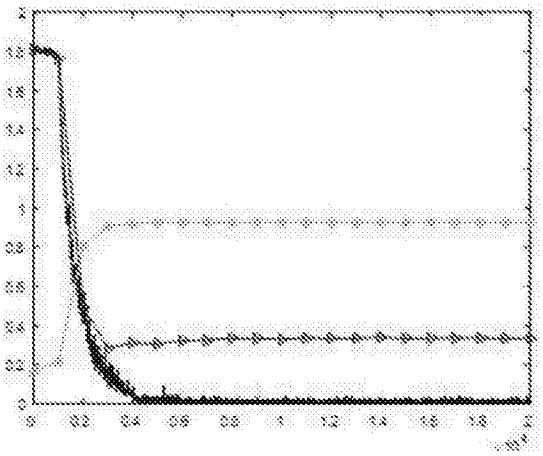


图6

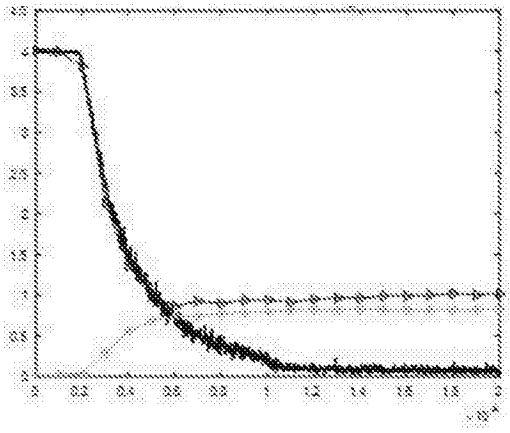
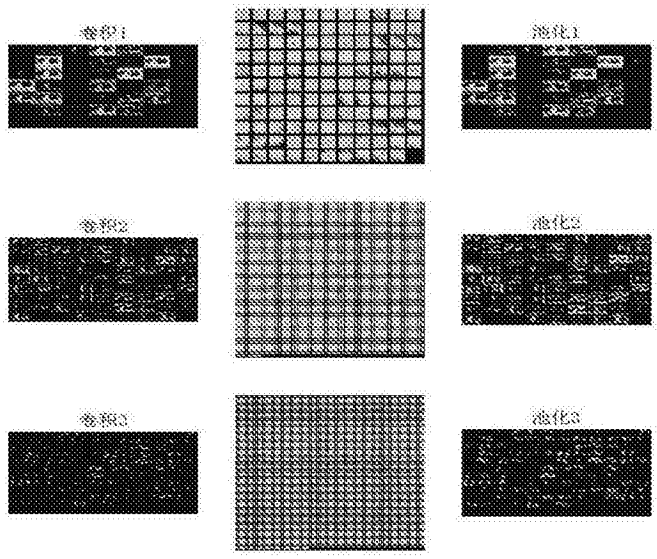
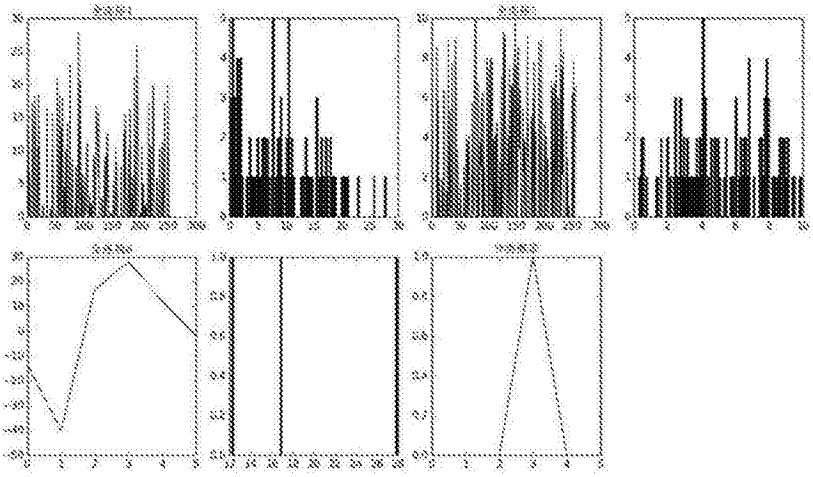


图7



（a）



（b）

图8