



(12)发明专利申请

(10)申请公布号 CN 110392893 A

(43)申请公布日 2019.10.29

(21)申请号 201880012755.X

(22)申请日 2018.02.19

(30)优先权数据

17156741.5 2017.02.17 EP

(85)PCT国际申请进入国家阶段日

2019.08.19

(86)PCT国际申请的申请数据

PCT/EP2018/054032 2018.02.19

(87)PCT国际申请的公布数据

W02018/150026 EN 2018.08.23

(71)申请人 考吉森公司

地址 意大利那不勒斯

(72)发明人 C·E·里金德斯

(74)专利代理机构 上海专利商标事务所有限公司 31100

代理人 蔡悦 陈斌

(51)Int.Cl.

G06K 9/52(2006.01)

G06K 9/62(2006.01)

权利要求书1页 说明书17页 附图20页

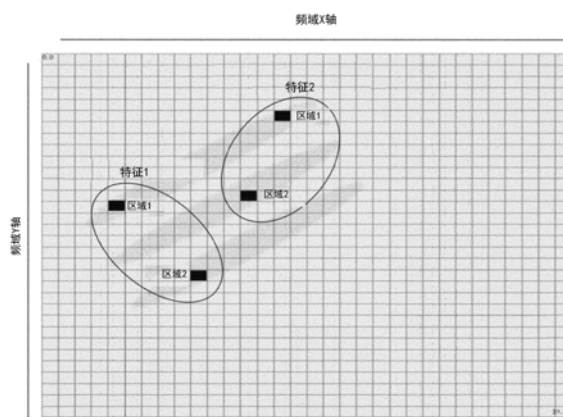
(54)发明名称

用于内容检测的图像处理方法

(57)摘要

一种用于内容检测的方法,所述方法由电子处理单元基于图像或数字视频图像流来执行,所述图像由所述视频流的单个帧或帧序列来定义,目的是增强并接着隔离表示要标识的内容的频域信号,并相对于所述图像或所述视频流内的所述内容减少或忽略频域噪声,包括以下步骤:从所述视频流的对应单个帧或对应帧序列获得数字图像或数字图像序列,所有的数字图像在空间域中被定义;选择一对或多对稀疏区域,每对稀疏区域覆盖所述单个帧的至少一部分或所述帧序列的至少两个帧,每对稀疏区域生成特征,每个区域由两个空间数据序列定义;通过以下方式将所选择的特征变换成频域数据:对于每个区域,通过L变换的2D变型来组合所述两个空间数据序列,改变每个区域的频域数据的传递函数、形状和方向,因而为所述特征的每一者生成经标准化的复向量;组合所有所述经标准化的复向量以定义要标识的所述内容的模型,能用作给内容

分类器的输入。优点在于,当前检测方法被赋予更好的输入数据、更快的计算,以及尤其是生成捕捉视频流中的时间数据的模型的大大的增加的能力。



1. 一种用于内容检测的方法,所述方法由电子处理单元基于图像或数字视频图像流来执行,所述图像由所述视频流的单个帧或帧序列来定义,目的是增强并接着隔离表示要标识的内容的频域信号,并相对于所述图像或所述视频流内的所述内容减少或忽略频域噪声,包括以下步骤:

- 从所述视频流的对应单个帧或对应帧序列获得数字图像或数字图像序列,所有的所述数字图像在空间域中被定义;

- 选择一对或多对稀疏区域,每对稀疏区域覆盖所述单个帧的至少一部分或所述帧序列的至少两个帧,每对稀疏区域生成特征,每个区域由两个空间数据序列定义;

- 通过以下方式将所选择的特征变换成频域数据:对于每个区域,通过L变换的2D变型来组合所述两个空间数据序列,改变每个区域的所述频域数据的传递函数、形状和方向,因而为所述特征的每一者生成经标准化的复向量;

- 组合所有所述经标准化的复向量以定义要标识的所述内容的模型,能用作给内容分类器的输入。

2. 根据权利要求1所述的用于内容检测的方法,其特征在于,将所选择的特征变换成频域数据的步骤使用来自不同数量/选择的帧的空间数据。

3. 根据权利要求1所述的内容检测方法,其特征在于,搜索逻辑被用在全输入图像上以生成其中所述稀疏区域被标识的输入帧。

4. 根据权利要求1所述的内容检测方法,其特征在于,所述稀疏区域被分组在一起,要么可能彼此部分交叠要么并排放置,以增加局部分辨率。

5. 根据权利要求1所述的内容检测方法,其特征在于,关于所述两个轴的计算是并行执行的。

6. 根据权利要求1所述的内容检测方法,其特征在于,在所述变换步骤中,首先1D Görtzel计算按行来执行,并且随后结果被用于第二步骤,其中1D Görtzel计算按列来执行,或反之亦然。

7. 根据权利要求1所述的内容检测方法,其特征在于,对于一对中的每个稀疏区域,不同的目标频率被选择。

8. 根据权利要求1所述的内容检测方法,其特征在于,用于到频域的变换步骤的输入单元仅在需要该值的索引的位置周围被取得。

9. 根据权利要求1所述的内容检测方法,其特征在于,所述索引的结果通过分开计算所述索引处的行和列的1D输出并接着将这组合成单个值来获取。

10. 根据权利要求1所述的内容检测方法,其特征在于,所述传递函数针对稀疏区域的每个输入来被分开选择,使得第一输入和第二输入具有不同的离散传递函数设置。

用于内容检测的图像处理方法

[0001] 发明的背景描述

1. 发明领域

[0002] 本公开涉及一种用于内容的检测的方法,该方法旨在被构建在提供有用于数字化图像的合适装置的任何类型的设备中,可能是常见设备。该方法对于创建在数字视频流(但不限于数字视频流)中使用的时间模型特别有用。

[0003] 内容意指感兴趣要去检测的任何对象。然后,内容的概念不仅限于对象(即通过图像可见的物理项目),而且还包括按论点或种类选择的对象族,例如表达暴力、展示裸露、显示体育活动、人群中的面部、标识车辆以及按种类或大小选择它们、为自动驾驶车辆系统辨别行人、骑车人和交通信号、识别地点或风景的图像,等等。在本发明中尤其可以涉及包括检测特定内容的步骤的任何领域。

[0004] 常见设备意指商业上易于获得的电子设备,如智能电话、平板、膝上型设备或具有合适的数字摄像机的任何便携式或手持式设备。另一方面,该设备可以由一个或多个相机表示,可能被组织成网络,被要么链接到计算机要么链接到服务器以进行图像计算。此外,可以在离线过程中通过在硬件中运行的软件来对通过合适硬件(包括存储存储器、RAM存储器和至少微处理器)存储的数字图像或视频简单地实现该检测方法。

[0005] 本公开还涉及一种用于操作至少提供有产生视频流或一系列数字图像的数字相机的设备或系统的方法,以通过该设备或系统的相机和处理器两者结合存储在可由所操作的设备或系统访问的存储器设备中的至少一个分类器模型来获得对象检测。

[0006] 2. 现有技术的描述

[0007] 在图像处理的工程应用中,近似愈发不再能够接受。此类应用的示例可以在诊断和创建仿真模型时或在仿真器中找到。需要适当地捕捉相对于信号具有高水平噪声的高度非线性、瞬态和动态效应。该应用的示例是仅使用最少量的信息的对象检测。可以针对部分遮挡的对象或远离拍摄相机的对象,或者甚至具有许多不同表示的对象和快速移动的对象,或者最终其组合来请求这样的检测。

[0008] 图像处理中的内容和对象检测技术正在各种上下文中广泛应用。作为示例并且没有限制的目的,这种识别被用在社交网络上的人脸标记技术、用于识别手势的软件、用于检测行人、骑车人和车辆的汽车软件、用于识别身体移动的软件、用于增强现实和具有3D效果的屏幕的人脸检测技术、用于增强现实的对象识别、使用头部取向或眼睛取向跟踪的界面、用于安全系统的对象跟踪技术以及最终注视跟踪技术中。

[0009] 对于这种对象检测产品仍然存在特定的硬件解决方案;例如,注视跟踪和姿势识别是通常需要集成红外投影仪或用于立体视觉的额外相机以获得所需精度的技术。

[0010] 存在所有这些对象检测技术迁移到下一代交互式接口和操作系统的清晰的未来趋势。例如,可在其上表示此类技术的设备不仅是智能电话、平板、可穿戴硬件(诸如交互式眼镜),而且还有未来家庭、办公室或公共场所中的任何种类的交互式对象。这些设备可被提供用于特殊用途,诸如交互式电视或智能家居,并且它们还可被用在汽车安全系统、医疗

保健、广告、安全相机网络、物联网,以及许多其他可能的用途中。

[0011] 这种技术基本上可以被集成到任何设备或网络连接设备中,其中可重编程硬件被使用并且其中摄像机输入可被添加。

[0012] 向设备添加额外的硬件,纯粹是为了帮助对象检测算法的运行,意味着更高的成本和额外的电池消耗。然后,创建微型硬件需要额外的研发成本,目前最先进的硬件通常仍然过大而无法被集成到大部分消费者电子设备中。

[0013] 除了硬件成本之外,在很大程度上阻碍在例如移动硬件平台上大规模使用对象检测技术的原因在于,所需的对象检测计算过慢以至于无法跟上相机的帧率,或者在这么做时使用过多的可用处理能力。

[0014] 因此,在下一代操作系统和设备上的实现在大规模体量上变得可行之前,这种对象检测技术首先将需要能够比当前现有技术快得多地处理图像的纯软件解决方案。

[0015] 这种向纯软件解决方案的迁移也得益于连续的相机技术改进,这些技术带来越来越高的帧率、更好的运动处理、更有效的色彩高亮显示、更敏锐地适应对比度、更智能地适应光线变化以及越来越高的屏幕分辨率。这种趋势将进一步提高用于对象检测的纯软件解决方案的有效性。

[0016] 内容检测使用尽可能少的处理能力的必要性旨在节省电池寿命以及针对实时使用的需要这两者。实时对象检测算法应当产生具有更自然的用户接口的设备。在实时运行时,对象检测算法还需要在后台运行,而不限制在前台运行的主进程。

[0017] 此外,应该注意,随着输入图像大小的增加,所需的计算量可能呈指数增长。视频帧率的上升也意味着对于对象检测算法而言在下一个视频输入帧到达之前完成计算的时间更短。

[0018] 因此,愈发得高的视频帧率和不断增长的输入图像质量的副作用在于,当前的现有技术对象检测算法将愈发需要对输入图像进行下采样,以返回到可接受的处理速度,因而失去较高质量输入图像中的大部分额外信息。

[0019] 这种下采样由此否定了在用于对象检测的输入中具有这种高清晰度图像的大部分优点。另外,内容检测方法在完全移动环境中是必需的,这意味着它们可以在日常情况下被使用而不需要在对象和相机之间相对较少的移动并且不需要恒定照明条件。

[0020] 使这些针对内容检测的挑战更加复杂的事实在于,需要被捕捉的内容愈发仅在视频流的时间数据中可见。各示例是暴力的检测、行人意图的检测、对安全相机的实况馈送上的可疑行为的检测,等等。这意味着视频流的两个或更多个图像帧需要在单个模型中交叉引用。目前的方法主要基于对静态图像的训练。换言之,视频被处理为一系列静态图像,而不是真正处理时间数据。在必须交叉引用多个帧以处理单个分类模型时增加的复杂度和处理开销将是明显的。

[0021] 依靠继续提高处理能力以减少给定算法对处理能力的相对使用也不是有效的,因为应用(例如,游戏)会缩放以使用最大的处理能力,因此总是给算法(诸如要在后台运行的对象检测)留出最小的量。

[0022] 鉴于上述情况,在对象识别和对象跟踪技术的计算机视觉文献中公开了许多方法。在内容检测中,所有已知方法使用从数字图像中提取的若干种特征,进而基于用于提取这些特征的多种方法。用于从数字图像中选择和提取特征的方法是对象检测技术的根源。

使用这些所选特征的不同方式也是已知的：在大部分情形中，这些特征被用于训练分类器或被用于在分类器中的已知特征组之间进行相关。

[0023] 对象检测方法的训练阶段通常需要很长时间，因为其需要输入相当大量的数据来创建合理且负担得起的模型。

[0024] 利用这样的经训练的模型，用于对象检测的通用过程可被概述为：输入图像、特征提取、特征分析和最终对象分类。

[0025] Viola-Jones是最常用的对象检测框架之一。该检测框架所采用的特征普遍地涉及简单矩形区域内的图像像素的总和，其通常通过灰度滤波器被转换以略微简化预见的计算。

[0026] 此类方法被称为弱分类器，其需要大量的特征（在数千个的量级上）或甚至这样大量的特征的多个级联以获得可接受的准确度水平，并且这些大量的特征涉及使用若干搜索框和多次重复的计算。

[0027] 使用许多特征的方法（诸如Viola-Jones）以及使用树状级联的技术（诸如AdaBoost，“自适应提升”的缩写）具有固有的非线性性能，因为算法检测的稳健性与这些算法所需的计算次数之间没有线性相关性。对于检测性能中的每个步骤，模型在复杂度方面增加。为了允许在实时应用中使用此类算法，因为常规设备专用于算法的处理能力的量通常受到限制，相比之下，无法针对算法的所需性能水平来预测所需的计算量，所以此类技术需要从模型中提取解空间，以便在使用完整模型之前被用于快速检测估计。

[0028] 其他方法（诸如SIFT, Scale-invariant（尺度不变）特征变换的缩写）检测和描述图像中局部特征的模型。

[0029] 对于数字图像中的任何对象，可以提取对象上的感兴趣的点以提供该对象的模型描述。SIFT特征是局部的并且基于特定兴趣点处的对象的外观，并且对于图像缩放和旋转是不变的。它们对于照明、噪声的变化和视点中的微小变化也很稳健。通过SIFT特征集的对象描述对于部分遮挡也是稳健的。

[0030] 然而，此类技术可以是计算上非常密集的，并且为了使这些技术实时地以高帧率工作，通常需要限制特征的数量和特征的复杂度，因而妨碍模型的有效性。

[0031] 通常被用于检测人类头部取向或人脸上的情绪捕捉的其他方法从数字化框架中提取几何点。它们依赖于相关性或者依赖于这些点与预定的2D或3D形状模型之间的对应关系。这些方法的示例被称为活动形状模型（ASM）和活动外观模型（AAM）。

[0032] ASM方法沿着关于当前模型点位置的轮廓搜索以更新对象形状的当前估计。在实践中，ASM技术试图将一组模型点与受形状的统计模型约束的图像相匹配。

[0033] 相反，AAM技术对图像数据进行采样，并且使用模型和样本之间的差异来更新外观模型参数。与ASM的不同之处在于，AAM寻求将模型点的位置和对象纹理的表示两者与图像相匹配。

[0034] 上面提到的方法的特征通常在空间域内被提取。仅使用空间域图像信息内的计算的方法具有供显著速度改进的较小范围。

[0035] 属于现有技术的用于内容检测的方法还可以在频域中使用所谓的过滤内核（简称内核）来首先强调要被分类并接着被识别的对象的辨别特征。

[0036] 接着，在使用内核之后提取的特征可以是经变换的图像中的像素的求和或频域中

各单元的求和的结果。

[0037] 大部分此类方法使用均匀分布在频域中的内核的预定义系列。这种技术的示例是Gabor库的使用。每种内核设置在空间域中导致已由内核变换的不同图像。在此类方法中提取的特征通常是经变换的图像内的像素强度的最高值的坐标。通常,在此类方法中使用大约四十种不同的Gabor滤波器。

[0038] 由于全频域计算被执行,并且需要每次从空间域到频域并接着再从频域返回到空间域的变换,因此此类方法还需要大量的计算。

[0039] 此外,此类方法通常使用FFT的变体之一(傅立叶变换的加速计算的快速版本),其中DCT是其离散版本。FFT要求对输入图像进行重新采样,因为输入处的图像的像素宽度和像素高度都需要具有2的幂的大小。

[0040] 在许多对象识别方法中,存在进一步的分类步骤,其中所提取的特征需要是有差别的,允许针对特定对象来被识别。这一方面主要用于人脸识别,其中个体被识别。用于对象识别的方法的示例是使用本征脸(eigenface)的主成分分析、线性判别分析和使用Fisherface算法的弹性束图匹配。

[0041] 与上述类似的用于对象检测的方法也可以与用于对象跟踪的方法相组合。在对象跟踪中,遵循对象在屏幕上的移动动态。用于此的常用方法是例如Mean-Shift(均值漂移)算法和模板匹配。此类算法具有良好的实时使用性能。然而,此类方法不生成对象的通用模型,而是学习对象的小的子样本。

[0042] 如上面提到的,常用的内容检测方法已对图像中的典型对象(例如检测在相机前摆姿势的人的面部)实现了良好的检测率。然而,它们在常见设备上的实时应用中的使用仍然有限,因为检测任务的难度极大地增加,因为在现实世界使用中,可能存在大量的照明条件,要检测的对象的多得多的未预期版本以及还有通常仅对要识别的对象的部分可见性。由于实现最终检测所要求的方法检测训练的复杂度,以及还有这些方法为实现必要的稳健性所需的计算次数,因此所有这些可能的偶然事件限制了许多常用的对象检测方法的使用。

[0043] 以上并非是对用于对象检测的图像的离线处理的限制,其中检测所经过的时间不是主要约束。然而,在实时应用中,对象检测必须至少以视频输入帧率工作。

[0044] 总而言之,以上作为现有技术所描述的方法仍未在无约束的现实世界实时应用中被大规模使用,因为根据利用此类方法的当前处理能力,难以在对象检测的稳健性和速度之间实现可接受的折衷。而且,重要的是,现有技术中描述的方法已大部分被设计成用于处理静止图像,而非仅可在交叉引用帧序列是动态视频流时才被找到的对时间数据的处理。

[0045] US 2009/0238466公开了一种聚焦于面部标识的对象识别方法,其中静态图像的边缘通过像素分类和一致性测试来被确定。

发明内容

[0046] 用于内容检测和图像处理的本发明的方法可以被应用于其中数据可以被变换为频域的情况。其不适用于基于空间域中的像素信息的直接分析而不将图像数据变换到频域的方法。一般而言,本发明的方法可被概述为如下:具有多个经标准化的复向量的模型包含变换到频域的稀疏数据的表示。可以从在频域中从多个图像帧序列中提取的特征来创建这

些经标准化的向量,使得该方法对于必须捕捉视频流中的时间数据的模型而言尤其有效。

[0047] 在该过程中,数个特征首先被选择,并且之后,在离线学习阶段(其中每个特征的特征参数(诸如目标频率)、变换内的底层传递函数以及频率变换的方向和顺序被优化)中,最佳参数设置被选择。这些特征可以跨多个帧散布,因而捕捉视频流中的时间进程。该组特征被组合成频域中的经标准化的复向量的单个空间。这组经标准化的复向量因此描述了频域中用于在视频或视频流中要检测的内容的模型。

[0048] 更深入详细地,用于在频域中提取和使用特征的方法包括以下步骤:

[0049] • 获得视频流中通过空间域中的数据定义的数字图像或数字图像序列;

[0050] • 选择用于变换到频域中的一对或多对稀疏区域,每个稀疏区域覆盖频域的一部分,并且可能覆盖视频流的序列中的两个或更多个帧;

[0051] • 使用L变换的2D变型将图像数据的总频域数据的稀疏部分转移到频域;

[0052] • 为每个区域的每个输入应用目标频率、传递函数以及给2D L变换的输入的的形状和方向的优化:

[0053] • 将每对区域转换成特征,该特征是经标准化的复向量;

[0054] • 将所有经创建的向量组合在一起成为包含所有经标准化的复向量的单个模型;

以及

[0055] • 使用经标准化的复向量模型来创建给分类器的输入数据。

[0056] 对于对象检测领域的专家而言清楚的是,存在作为数字图像输入使用的若干可能性:

[0057] • 整个图像被变换,

[0058] • 图像的一些片段被变换。

[0059] 还将清楚的是,数字输入的大小不改变此处所公开的权利要求。

[0060] 鉴于上述,根据本发明的用于内容检测的图像处理方法在所附权利要求1中被定义。

[0061] 在从属权利要求中限定了导致附加优点的方法的进一步细节。

[0062] 因此,本方法能够大大提高内容检测和图像处理方法的处理速度,同时提高其准确度。本方法在提高时间模型的有效性方面尤其有效。

[0063] 如上面所阐述的,对于能够提高图像处理中的内容检测技术的速度的新方法存在巨大需求。这种速度增加应使内容检测足够快,以便在最新一代移动设备和其他类似设备内部的后台中工作。

[0064] 除了在最新一代移动设备上使用外,屏息以待诸如可穿戴硬件等未来趋势,算法需要能够在主要个人计算机操作系统和移动操作系统之外工作,并因而能够在处理器上进行编程以及对诸如现场可编程门阵列等硬件进行重新编程。这些方法还需要由算法构建,这些算法可以利用个人计算机和移动设备(诸如多核和强大的图形处理单元(GPU))上的最新硬件开发。

[0065] 在这一点上,使用频域中的计算自然地使其本身响应于出于若干原因对更快计算的需求。该方法允许这种并行处理。本领域的专家众所周知,FFT的许多变体不适合在处理器上编程。

[0066] 重要的是要强调所描述的方法在需要创建时间模型时是如何特别有利的,其需要

捕捉仅能够在视频流中的帧的动态和时间序列中找到的模型。换言之，在视频流仅被视作静止帧序列时，内容无法被捕捉。

[0067] 傅立叶的标准频域理论表明，任何信号（在我们的情形中是数字图像）都可以被表示为一系列正弦曲线的求和。在图像处理的情形中，这些是跨图像的亮度中的正弦变化。正弦函数可以编码：

[0068] • 空间频率

[0069] • 幅度

[0070] • 相位

[0071] 空间频率是跨亮度随其调制的空间的频率。

[0072] 正弦曲线的幅度对应于其对比度，或者图像的最暗和最亮峰之间的差异。相位表示波相对于原点如何被移位。

[0073] 傅立叶变换不单单编码单个正弦曲线，而且编码通过从零直到“奈奎斯特频率”的空间频率范围的整个系列的正弦曲线，“奈奎斯特频率”意指可以在数字图像中被编码的最高空间频率，其与分辨率、或像素总数相关。

[0074] 傅立叶变换同时编码存在于图像中的所有空间频率。

[0075] 奈奎斯特频率是离散信号处理系统（在我们的情形中是数字图像）的采样率的1/2。被用于图像处理的傅立叶变换背后的底层原理在于，每个像素影响所有频率，并且每个频率影响所有像素。

[0076] 此处所公开的方法背后的直觉在于，空间域中所述对比度的位置由频域中的正弦曲线编码。空间域中的某些对比度仅受频域中某些正弦曲线的影响。对于由空间域中的对比度的组合定义的形状，这意味着该形状通过频域中的位置信息的唯一组合来被捕捉。并且这意味着通过频域中位置信息的唯一改变来捕捉移动。

[0077] 因此，我们可以将对频域中特定信息或信息改变的捕捉用于特定对象的检测。

[0078] 在频域中工作允许更容易的计算，因为利用传递函数的乘法和利用矩阵的其他类似计算是简单的分量级乘法，与在其中它是两个函数之间的卷积（其在计算上更昂贵）的空间域中不同。

[0079] 因此，频域允许对用于对象检测的稀疏的、小的、部分频域信息的计算上简单的使用。在下文中将解释和详述稀疏区域以及频域中的传递函数、特征和格式。

[0080] 完全工作在频域中而不需要在频域计算之后的将图像变换回空间域的计算，还允许在执行到频域的变换的数学选择中的增加了的灵活性。

附图说明

[0081] 通过下面对参考附图给出的优选实施例的描述，本方法的其他特征和优点将变得更加明显，在附图中：

[0082] • 图1示出了在图像的空间域中，索引是怎样指示位置，而图像内的频率是怎样指示像素强度的正弦变化的，并且这在频域内正好相反；

[0083] • 图2示出了将怎样通过编码频域中位置的波的变化来捕捉给定对象的空间域中的移动信息；

[0084] • 图3示出了在频域中怎样需要数个波来捕捉足够的位置和形状信息以对图像内

的对象进行分类；

- [0085] • 图4示出了如何仅需要频域信息的稀疏部分来捕捉频域中的正弦信息；
- [0086] • 图5和图6示出了 Görtzel 的完整2D实现实际上将怎样是许多1D Görtzel 计算的组合；
- [0087] • 图7示出了完整2D实现将怎样在计算上是不理想的；
- [0088] • 图8示出了怎样可以更快地做出完整2D实现,即使仍然不理想；
- [0089] • 图9示出了如果通过分开计算索引处的行和列的1D输出并接着将这组合成单个值来获取索引的结果,则这是怎样有效的；
- [0090] • 图10示出了如何更快地进行计算;用于变换到频域的输入单元仅在需要该值的索引的位置周围被取得；
- [0091] • 图11示出了区域的输入是怎样不限于仅行和列作为输入的,而可以是任何自由形状；
- [0092] • 图12示出了每个区域如何具有可以是任何自由形状的两个输入；
- [0093] • 图13示出了两个区域如何形成特征,该特征利用标准化变成给出两个区域之间的信息移位的复向量；
- [0094] • 图14示出了每个索引如何具有目标频率、变换的底层传递函数以及定义频域变换的输入的方向和序列的特定特征形状；
- [0095] • 图15示出了怎样通过顺序移动输入的每个索引来优化频域变换的输入形状；
- [0096] • 图16示出了输入图像如何可以是任何形状,不限于方形和矩形；
- [0097] • 图17示出了区域和特征可以如何交叉引用视频流中的帧序列中的数据；
- [0098] • 图18示出了定义每个索引的频域变换序列的形状不限于单帧,而是可以交叉引用视频流中的多个帧；
- [0099] • 图19解说了该方法允许选择被用来生成模型的频域变换的帧序列的自由度,从而允许真实的时间数据被用于内容的检测；
- [0100] • 图20示出了频域中的优化如何在非常大的多维噪声区域中具有成为清晰信号的非常大的优势,从而允许不仅监督学习而且还有无监督学习的目前逻辑类型；
- [0101] • 图21示出了流程图,该流程图是所描述的方法的可能的优化逻辑的示例。

具体实施方式

[0102] 在下文中,将参考附图详述根据本发明的方法的实施例。

[0103] 显而易见的是,本文中参考视频流(即一连串具有特定速率的图像帧系列)所描述的内容,也适用于等同于视频流的帧的任何一连串单个图像,并且适用于等同于单个帧的一个单一图像。根据本发明的方法的目标是增强并接着隔离表示要标识的内容的频域信号,并相对于图像或视频流内的内容减少或忽略频域噪声。

[0104] 在第一步中,可以在完整输入图像上使用搜索逻辑以生成用于该方法的计算的输入帧。搜索逻辑可以例如是整个图像或图像的子集。应该清楚的是,许多类型的搜索逻辑是可能的,但是从此处所公开的方法的观点来看,该计算或权利要求并不改变,只是用于变换的图像输入改变。还将清楚的是,单个帧可以具有用于多个计算的多个输入,该多个计算的每一者如权利要求中所描述进行处理。

[0105] 然后在频域中提取从帧或帧序列获得的一个或多个输入。帧的频域数据内的数据接着被处理以检测内容。此处所描述的方法使所使用的分类开放,相反,在该方法中强调的是被用于选择的分类的数据的质量和类型中的改进。

[0106] 如上面提到的,所描述的方法在检测特别或甚至排他地在视频流的时间数据中的对象和过程时尤其有效。此处将解释可以怎样将多个帧组合成单个输入以进行检测。

[0107] 当已检测到对象或者搜索逻辑的重复未能在图像内找到对象时,检测将移动到视频流的下一帧或帧序列。应该清楚的是,搜索逻辑可以基于哪个对象在前一帧中被找到而被致使是自适应的。

[0108] 在已知技术中,在频域中处理图像通常使用快速傅立叶变换(FFT)的变型来完成,但是本方法既不使用FFT或其变体(例如离散余弦变换(DCT)),也不使用离散傅立叶变换(DFT)。

[0109] 然而,为了突出常规图像处理和本方法之间的差异,本文中给出了FFT和DFT的一般概览。

[0110] FFT被用在广泛的应用中,诸如图像分析、图像重构和图像压缩、文本识别等等。

[0111] FFT的主要原理遵循自离散傅立叶变换(DFT)。由于DFT需要大量的计算,因此存在寻求加速该过程的其他类型的变换。快速傅立叶变换(FFT)是这些中最有建树的。利用DFT,计算的数量与 N^2 相关,其中 N 是输入矩阵的长度。

[0112] FFT算法依赖于标准DFT涉及大量冗余计算的事实。

[0113] 通过将样本序列划分成子序列来计算FFT;在每个阶段,需要 $N/2$ 次复数乘法来组合前一级的结果。

[0114] 由于存在 $\log(N)$ 级,因此复数乘法的数目需要在 N 点DFT上进行评估,其中FFT近似为 $N \cdot \log(N)$ 。

[0115] 频率数对应于空间域图像中的像素数,即空间和频域中的图像具有相同的大小。

[0116] 如上面提到的,存在FFT的许多变型。FFT在图像处理方面也有其局限性。例如,在用于FFT的输入中使用的图像的边需要具有像素中的长度,其是2的幂。

[0117] 另一限制是需要给定特定频率的结果之前计算完整的FFT。换言之,FFT无法被转换用于稀疏计算,因为必须在可以获得单个频率的值之前计算完整的FFT。FFT的复杂结构也不允许在可重新编程的硬件和多核处理器上的简单编码实现。另外,由于首先需要计算完整的FFT以获得单个结果,因此还需要更多地使用设备上的存储器。

[0118] 可以提供诸如修剪的快速傅立叶变换之类的方法,但是它们需要大量复杂的代码以获得速度和存储器使用方面的相对小的增益,同时仍然难以在可重新编程的硬件上实现。

[0119] 在空间域中,值通常是像素的光强度,其范围从0到255。相同图像的傅立叶域值具有比空间域中的图像大得多的范围。

[0120] 傅立叶变换产生复数值输出图像,其可以用两个图像显示,要么通过实部和虚部要么通过幅度和相位。在图像处理中,通常仅显示傅立叶变换的幅度,因为其包含空间域图像的几何结构的大部分信息。然而,为了在频域中的一些处理之后将傅立叶图像重新变换到正确的空间域中,必须保留傅立叶图像的幅度和相位两者。

[0121] 在根据本公开的方法中,使得对所有计算而言排他地使用频域中的信息成为可

能。由于不需要保留所有信息以返回空间域,因此存在若干优点。

[0122] 首先,缺少从频域返回空间域的额外步骤加速了整体计算。

[0123] 其次,由于不需要将频域数据转换回正确的空间域图像,因此较小的稀疏区域可以被使用。这是因为不需要具有将允许图像数据被转换回空间域的频域数据而没有图像质量和信息的大的损失。频域中的稀疏区域本身不一定包含足够的频域信息来重新创建空间图像。但是它们包含足够的信息以供分类。

[0124] 第三,额外的计算可以被执行以移除FFT中常见的混叠并且其他计算可以被执行以更好地准备数据以用于频域内的分类。

[0125] 第四,在如FFT和DCT之类的方法中存在的其他限制被移除。例如,FFT和DCT的频率变换沿着图像的行和列完成,并且总是在单个帧内。在该方法中,频域变换的方向可以是任何排列,其中对于传递函数具有更大的自由度,并且变换甚至在视频序列的帧之间交叉。

[0126] 在图1中,表示了图像的空间域中,索引是怎样指示位置,而图像内的频率是怎样指示像素强度中的正弦变化的。在频域内正好相反,索引示出频率,而正弦波则包含位置数据。

[0127] 以相同的方式,给定对象的空间域中的移动信息将通过在频域中的位置进行编码的波的改变来捕捉。这用图2中的眼睛移动图像示意性地示出。

[0128] 上面给出的示例当然是为了概念的解说而简化的。实际上,以与在空间域中需要捕捉像素强度中的变化的许多频率来绘制图像相同的方式,在频域中需要数个波来捕捉足够的位置和形状信息以对图像内的对象进行分类。这在图3内被表示。

[0129] 如上面提到的,频域内的每个索引潜在地影响空间域中的所有像素。因此,与利用从空间域提取的特征将对象进行分类相比,在频域中需要相对较少的特征来将对象进行分类。在本文中所描述的对象检测方法中,公开了一种技术以在频域中找到捕捉空间域中的特定类型的形状的信息的最小部分组合。这种信息组合尤其也可以是视频流中的帧序列,目的是捕捉在单独考虑序列的每个静止图像时未被找到的时间和动态信息。

[0130] 这里和以下描述中,稀疏区域意指覆盖频域的一部分的信息选择。应将每个区域视为特定的频域信息。图4示出了频域中稀疏特征的布局示例。必须注意每个特征是从一对区域创建的,并且频域网格的大小仅用于解说的目的,而可以具有许多其他大小,如稍后将解说的。图4示出了捕捉空间域中的位置和移动的可能的频域正弦对比度的示例,也以叠加示出。图4所示的是,仅需要捕捉频域正弦对比度的一部分以检测空间域中的移动类型或形状,这是稀疏区域所做的。

[0131] 稀疏区域可以被分组在一起,或者可以彼此部分交叠或者并排放置,以增加局部分辨率。

[0132] 对从所述稀疏区域导出的频率值的计算被指示为稀疏计算。

[0133] 由于不需要将图像转换回空间域,且并非所有频域信息是必需的,因此对于使用除DFT或FFT之外的其他方法来将图像转换到空间域中的可能性是开放的。

[0134] 根据本方法,一对或多对稀疏区域被选择,每个稀疏区域覆盖单个帧的至少一部分,或者在帧序列的情形中,覆盖序列的至少两个帧。

[0135] 如上面提到的,每对稀疏区域生成特征,并且每个稀疏区域由两个空间数据序列定义。接着,根据本方法,所述经选择的特征通过以下方式被变换成频域数据:对于每个稀

疏区域,通过L变换的2D变型来组合所述两个空间数据序列,改变每个区域的频域数据的传递函数、形状和方向,因而为所述特征的每一者生成经标准化的复向量。

[0136] 因此,可以使用诸如从 G \ddot{o} ertzel 算法导出的二维变换等其他方法来执行变换,其中关于目标频率、在变换中使用的传递函数以及定义变换的输入的环路的形状和方向具有相当大的设计自由度。如将在本公开中进一步解释的,所使用的方法与 G \ddot{o} ertzel 算法非常不同,并因此所使用的描述在于其是L变换的2D变型。

[0137] 如之前提到的,该方法的优点在于,其可以以更灵活的方式来被稀疏地并行设置,以供在可重新编程的处理器或GPU上使用,同时使用最小量的存储器。在下文中,首先描述了 G \ddot{o} ertzel 变换背后的理论。之后详述了该方法的实现,以及针对图像处理中2D情形的扩展和可以被使用的各种设计选项。

[0138] 当单个正弦频调的检测和测量中的频谱分析不得被执行时,无限脉冲响应(IIR)滤波器结构被使用。

[0139] 频谱能量的标准方法是离散傅立叶变换(DFT),通常使用快速傅立叶变换(FFT)或离散余弦变换(DCT)算法来实现。

[0140] 然而,存在仅需要在N点DFT的N-箱(N-bin)中心频率的子集上进行频谱分析的应用。用于在1D中计算稀疏FFT结果的流行且高效的技术是 G \ddot{o} ertzel 算法,其使用IIR滤波器实现来基于N个输入时间样本计算单个复DFT频谱箱值。

[0141] 这个过程最常见的应用是检测单个连续波正弦频调的存在。作为1D计算, G \ddot{o} ertzel 算法不旨在被用于其中图像是2D的图像处理。

[0142] G \ddot{o} ertzel 算法基于计算长度为N的信号 $\{x[n]\}$ 的k分量的概念

$$[0143] \quad X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi k \frac{n}{N}} \quad (\text{等式 1})$$

[0144] 用 $e^{j2\pi k \frac{N}{N}} = 1$ 乘以这个等式(1)的右边,我们得到:

$$[0145] \quad X[k] = e^{j2\pi k \frac{N}{N}} \sum_{n=0}^{N-1} x[n] e^{-j2\pi k \frac{n}{N}} \quad (\text{等式 2})$$

[0146] 其可以被写为

$$[0147] \quad X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi k \frac{n-N}{N}} \quad (\text{等式 3})$$

[0148] (3)的右边可以被看作信号 $\{x[n]\}$ 和 $\{h_k[n]\}$ 的离散线性卷积,其中:

$$[0149] \quad h_k[l] = e^{j2\pi k \frac{l}{N}} u[l]$$

[0150] 事实上,如果 $\{y_k[n]\}$ 表示该卷积的结果,那么我们有:

$$[0151] \quad y_k[m] = \sum_{n=-\infty}^{\infty} x[n] h_k[m-n] \quad (\text{等式 4})$$

[0152]

[0153] 其可以被改写为

$$[0154] \quad y_k[m] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi k \frac{m-n}{N}} u[m-n] \quad (\text{等式 5})$$

[0155] 卷积被定义为两个函数的在一个被反转和移位之后的乘积的积分。由此,其是一种特殊的积分变换。

[0156] 卷积定理表明,在适当的条件下,卷积的傅立叶变换是傅立叶变换的逐点乘积。换言之,一个域(例如,时域)中的卷积等于另一域(例如,频域)中的逐点乘法。

[0157] 将(3)与(5)进行比较,明显的是,所需的 $X[k]$ 是卷积的第 N 个样本:

$$[0158] \quad X[k] = y_k[N] \quad (\text{等式 6})$$

[0159] 对于 $k=0, \dots, N-1$ 。这意味着所需的值可以被获得作为具有脉冲响应 $\{h_k[n]\}$ 的IIR线性系统的时间 N 中的输出样本。

[0160] 现在将导出该系统的传递函数 $H_k(z)$;这是其脉冲响应的 L 变换:

$$H_k(z) = \sum_{n=-\infty}^{\infty} h_k[n] z^{-n} \quad (\text{等式 7})$$

$$= \sum_{n=-\infty}^{\infty} e^{j2\pi k \frac{n}{N}} u[n] z^{-n} \quad (\text{等式 8})$$

[0161]

$$= \sum_{n=0}^{\infty} e^{j2\pi k \frac{n}{N}} z^{-n} \quad (\text{等式 9})$$

$$= \sum_0^{\infty} (e^{j2\pi k \frac{1}{N}} z^{-1})^n \quad (\text{等式 10})$$

[0162] 该几何级数是收敛的,并且其求和等于传递函数:

$$[0163] \quad H_k(z) = \frac{1}{1 - e^{j\frac{2\pi k}{N}} z^{-1}} \quad (\text{等式 11})$$

[0164] 这给出以下差分方程:

$$[0165] \quad y_k[n] = x[n] + e^{j\frac{2\pi k}{N}} y_k[n-1] \text{ 其中 } y_k[-1] = 0$$

[0166] (等式 12)

[0167] 等式(12)涉及与复数相乘,并且每个复数乘法导致四个实数乘法和四个实数加法。为了避免复杂的乘法,该函数可以乘以复共轭极点并被简化为如下:

$$[0168] \quad H_k(z) = \frac{1 - e^{-j\frac{2\pi k}{N}} z^{-1}}{1 - 2 \cos\left(\frac{2\pi k}{N}\right) z^{-1} + z^{-2}}$$

[0169] (等式 13)

[0170] 二阶的该IIR的差分方程是:

$$[0171] \quad y_k[n] = x[n] - x[n-1] e^{-j\frac{2\pi k}{N}} + 2 \cos\left(\frac{2\pi k}{N}\right) y_k[n-1] - y_k[n-2] \quad (\text{等式 14})$$

[0172] 并且这样的结构可以使用状态变量来描述:

[0173]

$$s[n] = x[n] - x[n-1]e^{-j\frac{2\pi k}{N}} + 2\cos\left(\frac{2\pi k}{N}\right)s[n-1] - s[n-2] \quad (\text{等式 15})$$

[0174] 并且我们令 $s[-1]=s[-2]=0$

$$y_k[n] = X(k) = s[n] - s[n-1]e^{-j\frac{2\pi k}{N}} \quad (\text{等式 16})$$

$$y_k[n] = s[n] - e^{-j\frac{2\pi}{N}k}s[n-1] \quad (\text{等式 17})$$

$$= A - Be^{-j\theta} \quad (\text{等式 18})$$

$$= [A - B \cos\theta] + j B \sin\theta \quad (\text{等式 19})$$

$$A = s[n]$$

$$B = s[n-1]$$

$$\theta = \frac{2\pi k}{N}$$

[0182] 事实上，Göertzel算法执行单个1D DFT系数的计算。与DFT相比，它具有若干优点，并且出于这一原因，其有时被用在1D应用中。

[0183] Göertzel算法在仅需要很少的频谱分量的值而不是整个频谱时的情况中是有利的。一示例是识别按下具有特定音频脉冲的按钮。在这样的情形中，该算法可以显著地更快速。

[0184] 将FFT算法用于计算DFT分量的效率很大程度上由信号长度 N (N 必须是2的幂) 确定。相比而言，在Göertzel算法的情形中， N 可以是任意的，并且计算复杂度不变。可以在任意时刻启动计算，不必像在FFT的情形中那样等待整个数据块。因而，从存储器容量的角度来看，Göertzel算法的要求较低，并且其可以以非常低的等待时间执行。因此，Göertzel算法不需要以位反转顺序对输入或输出数据进行任何重新排序。

[0185] 1D Göertzel 算法

[0186] 1D Göertzel 的算法具有非常基本的结构。我们可以从等式 (17) 开始。

[0187] 一些中间处理在每个样本中被完成。与FFT一样，我们用样本块工作。

[0188] 若干设置被需要以初始化1D Göertzel 的计算：

[0189] 1. 采样速率

[0190] 2. 块大小， N

[0191] 3. 目标频率

[0192] 一旦采样速率和块大小被选择，便存在计算所需常数的五步过程：

[0193] 常数 k 、 w 、 \cosine 、 $sine$ 和 $coeff$ 定义如下： $k = \text{int}\left(0.5 + \frac{N * \text{target} - \text{freq}}{\text{sample} - \text{rate}}\right)$ [0194] $w = (2\pi/N) * k$ [0195] $\cosine = \cos w$ [0196] $sine = \sin w$ [0197] $coeff = 2 * \cosine$ (等式20)

[0198] 对于每样本处理，三个变量被使用： S_0 、 S_1 和 S_2 。 S_1 仅仅是最后一次迭代时 S_0 的值。 S_2 是两个迭代步骤之前(或者换言之，在 S_1 之前的一次迭代)的 S_0 的值。必须在每个样本块的开

头将 S_1 和 S_2 初始化为零。

[0199] 对于矩阵 $[n \times m]$ 的每个列(行),计算以下三个等式:

[0200] $\{S_0 = \text{coeff} * S_1 - S_2 + \text{sample}$

[0201] $S_2 = S_1$

[0202] $S_1 = S_0$

[0203] $\text{real} = (S_1 - S_2 * \text{cosine})$

[0204] $\text{imag} = (S_2 * \text{sine})$

[0205] $\text{magnitude}^2 = \text{real}^2 + \text{imag}^2$ (等式21)

[0206] 这是1D Görtzel 算法的基本版本。如上面提到的,其给出与1D DFT 相同的结果。

[0207] 也可以使用1D Görtzel 的需要比基本版本更少的计算的版本,代价是相位信息,这意味着不计算变换的实部和虚部两者。将清楚的是,优选计算实部和虚部两者,并且更快的选项仅用于其中处理器开销被非常严格地限制的情形。

[0208] 在更快的版本中,每样本处理是相同的,但块处理的结束是不同的。不是计算实数和虚数分量并接着将这些转换成相对平方幅度,而是直接计算以下而无需其中还计算实数和虚数分量的基本版本的步骤:

[0209]

$$\text{magnitude} = S_1^2 + S_2^2 - S_1 * S_2 * \text{coeff} \quad (\text{等式 } 22)$$

[0210] L变换版本的2D实现

[0211] 应当再次注意, Görtzel 算法的这个通用版本是为1D计算定义的。在图像处理中,这种计算是不够的,因为用于将图像变换到频域中的计算需要在两个维度上完成:X和Y。另外,尽管 Görtzel 的

[0212] 1D实现等同于1D DFT,但对于2D而言,这将不是真的。因此, Görtzel 算法似乎不是对象检测和图像处理的候选者。Görtzel 的另一限制在于,没有太多能力来针对特定信号进行调谐和优化。

[0213] 然而,本公开从1D Görtzel 算法的原理开始描述了藉此利用2D实现将图像转换到频域的方法,但是将它们改变到可以在2D中被称为全新方法的程度,该全新方法在本文中描述为L变换的2D变型。另外,由于该方法中的计算完全在频域中,而不需要返回到空间域,因此不要求2D计算等同于2D DFT。

[0214] 图5和图6示出了 Görtzel 的完整2D实现实际上将怎样是许多1D Görtzel 计算的组合。一选项将首先对图5的行进行各种1D计算,接着将这些结果用于第二步,其中所有1D Görtzel 计算针对各列完成,如图6中那样。替代地,首先可以计算各列,接着是各行。

[0215] 即使此处所描述的方法可以使用这样的实现,但由于若干原因其不是优选方式。首先,行的计算必须等待列的计算完成,或反之亦然。

[0216] 意味着并行处理将是不可能的。其次,计算仍然不会真正稀疏。图7解说了这点。在该图中,针对2D实现示出了所需的计算,其中需要索引 (i, j) 中的频域值。在图7中,示出了该选项,其中首先计算各行,接着计算各列。1D计算将首先计算索引 i 处每行的值。在这之后,列的1D计算可以

[0217] 被计算完成以获得索引j处的值。将清楚的是,在计算上这并不理想。对于频域变换领域的专家而言将清楚的是, Görtzel 的2D实现将以这样的方式改变数据,即原始图像无法在返回空间域时被重新创建。然而,如之前所阐述的,该方法通过仅使用频域日期来描述数据的分类。因此,此处所描述的方法中的驱动程序要具有尽可能快地为分类器生成最佳可能输入的计算,而不是驱动程序作为空间域数据。

[0218] 接下来,本公开将描述在速度和检测方面具有用于分类器的经优化数据的一系列选项。具体而言,将描述如何最佳地捕捉视频流中的时间数据。

[0219] 这些尤其是:

[0220] • 使用多个特征,每个特征使用两个区域;

[0221] • 选择由特征和区域覆盖的视频流序列中的帧数;

[0222] • 为每个区域选择不同的目标频率;

[0223] • 具有给每个区域的两个输入,每个输入是频域变换;

[0224] • 具有给每个输入的可变核滤波器,其可以针对变换的实部和虚部两者被优化;

[0225] • 具有给定义每个索引的频域变换的环路的输入的可变序列和形状;

[0226] • 使用每个特征的特征对来为每个特征生成经标准化复向量;以及

[0227] • 最后,将所有经标准化的复向量以单一格式组合在一起;

[0228] 可以使这些计算比图7中的示例更稀疏。一种方式如图8所示,其中用于变换到频域的输入单元仅在需要该值的索引的位置周围被取得。然而,这仍然需要各行等待各列的结果,或反之亦然。优点是单位量的长度可以变成输入参数,从而允许特征之间的更多区分,同时还可以捕捉被分类的对象的细节。

[0229] 更有效的计算如图9所示。这里,通过分开计算索引处的行和列的1D输出并接着将这组合成单个值来获取索引的结果。除了速度中的增益之外,这种方式的最大优点在于,在首先计算行的1D或者列的1D的情况下是没有区别的,因此可以并行计算行和列的值。通过如图10所示限制输入的长度可以实现甚至更大的速度增益,其中在这种情形中,只需要计算单个行和列输入。

[0230] 如果考虑跟着在给定区域索引中得到结果的2个输入甚至不需要沿着行和列或甚至是毗邻的单元(如图11所示),则用频域输入数据训练分类器的自由度的量变得甚至更大。

[0231] 在以下描述中,通常每区域显示单个输入,以使图更加示意性。但是应该强调对于每个区域存在两个输入,如图12所示。

[0232] 频域变换针对每个输入完成,从而给出实数和虚数。如所提到的,每个区域具有两个输入,并且在下文中描述了将它们组合成单个经标准化的复数值的方式。这也在图13中示出。首先,区域1的两个实输入被组合:

[0233]

$$\text{Real}(\text{Zone1}_{\text{Feature}_i}) = \sqrt{\text{Real}(\text{Input1}_{\text{Zone1}})^2 + \text{Imag}(\text{Input1}_{\text{Zone1}})^2} + \sqrt{\text{Real}(\text{Input2}_{\text{Zone1}})^2 + \text{Imag}(\text{Input2}_{\text{Zone1}})^2} \quad (\text{等式 } 23)$$

[0234] 区域1的两个虚输入被组合以给出相位:

[0235]

$$\text{Imag}(\text{Zone1}_{\text{Feature}_i}) = \frac{\text{Imag}(\text{Input1}_{\text{Zone1}})}{\text{Real}(\text{Input1}_{\text{Zone1}})} + \frac{\text{Imag}(\text{Input2}_{\text{Zone1}})}{\text{Real}(\text{Input2}_{\text{Zone1}})} \quad (\text{等式 } 24)$$

[0236] 对区域2的两个实和虚输入重复相同的过程:

[0237]

$$Real(\text{Zone2}_{\text{Feature}_i}) = \sqrt{Real(\text{Input1}_{\text{Zone2}})^2 + Imag(\text{Input1}_{\text{Zone2}})^2} + \sqrt{Real(\text{Input2}_{\text{Zone2}})^2 + Imag(\text{Input2}_{\text{Zone2}})^2} \quad (\text{等式 25})$$

[0238]

$$Imag(\text{Zone2}_{\text{Feature}_i}) = \left/ \frac{Imag(\text{Input1}_{\text{Zone2}})}{Real(\text{Input1}_{\text{Zone2}})} \right. + \left/ \frac{Imag(\text{Input2}_{\text{Zone2}})}{Real(\text{Input2}_{\text{Zone2}})} \right. \quad (\text{等式 26})$$

[0239] 接下来,区域1和区域2的实数值的结果在标准化中被组合:

$$Real(\text{Feature}_i) = \left| \frac{Real(\text{Zone2}) - Real(\text{Zone1})}{Real(\text{Zone1}) + Real(\text{Zone2})} \right| \quad (\text{等式 27})$$

[0241] 这还针对区域1和区域2的虚数值完成:

$$Imag(\text{Feature}_i) = \left| \frac{Imag(\text{Zone2}) - Imag(\text{Zone1})}{Imag(\text{Zone1}) + Imag(\text{Zone2})} \right| \quad (\text{等式 28})$$

[0243] 以这种方式,形成1个特征的每对区域给出经标准化的复向量,如图13所示:

$$\vec{V}_{\text{Feature}_i} = Real(\text{Feature}_i) + jImag(\text{Feature}_i) \quad (\text{等式 29})$$

[0245] 可以使用大量这样的经标准化的复向量来构建模型:

$$\vec{V}_{\text{Feature}} = \{ \vec{V}_{\text{Feature}_1}, \vec{V}_{\text{Feature}_2}, \dots, \vec{V}_{\text{Feature}_n} \} \quad (\text{等式 30})$$

[0247] 这种经标准化的复向量的格式正是此处描述的方法给予的给分类器的新输入类型的输入。对本领域专家而言将清楚的是,该格式允许将概率论和量子物理学的所有数学应用于分类。

[0248] 与使空间域中的像素作为给分类器的输入相比,在所描述的方法中获得索引中的值所需的频域计算的数量将在很大程度上被减少。以这种方式获得的值仍将与频域中的形状信息强相关,同时还允许大量控制减少如混叠和周期信号等影响。这些影响的减少是重要的,因为一个关键目标是在每个索引中具有唯一结果。这里需要再次注意,对于该方法,不需要具有在空间域中重建图像的所有数据。目标是对频域信息的捕捉,其稀疏地编码被检测对象的位置和移动。

[0249] 清楚的是,当与直接使用空间域中的像素输入相比时,所创建的频域空间非常强烈地减少了分类器需要处理的参数量。当所描述的方法被用来捕捉只能在视频流中的帧序列中检测到的时间信息时,该优点被指数地放大。

[0250] 返回到沿每个索引的1D计算(其不需要沿着行或列,而是自由的),我们可以按如下方式将它们重写为:

$$k = (Int) \left(0.5 - \frac{N - \text{Target_frequency}}{\text{Sample_rate}} \right) \quad (\text{等式 31})$$

$$\omega = \left(\frac{2\pi k}{N} \right) \quad (\text{等式 32})$$

$$\text{coeff} = 2 \cdot \text{digital_filter} \quad (\text{等式 33})$$

$$[0254] \quad \text{digital_filter} = \sin(A\omega + b\pi) + j \sin\left(A\omega + (2b + 1)\frac{\pi}{2}\right) \quad (\text{等式 34})$$

$$[0255] \quad S_0 = \text{coeff} \cdot S_1 - S_2 + \text{sample} \quad (\text{等式 35})$$

$$[0256] \quad S_1 = S_0$$

$$[0257] \quad S_2 = S_1$$

$$[0258] \quad \text{Real} = (S_1 - S_2 \cdot \sin(A\omega + b)) \quad (\text{等式 36})$$

$$[0259] \quad \text{Imag} = \left(S_2 \cdot \sin\left(A\omega + b\frac{\pi}{2}\right)\right) \quad (\text{等式 37})$$

[0260] 我们看到在这种方法中,对于频域变换的实部和虚部两者,是怎样以利用参数a和b来致使传递函数可调谐的。

[0261] 应当注意,对于区域的每个输入,作为数字滤波器工作的这些传递函数具有可以针对每一者来分开地选择的选项,这意味着第一输入和第二输入可以具有不同的离散数字传递函数设置。

[0262] 如之前提到的,该方法的优点之一在于,所描述的许多选项还用于调谐在用作给分类器的输入之前要被清理的频域数据时的大量自由度。

[0263] 这在图14中示出。我们看到对每个输入而言是如何存在选择目标频率、核环路中使用的输入的方向和顺序以及核环路中使用的传递函数的自由度的。

[0264] 图15示出了可如何在优化阶段调整给核内环的输入的方向和顺序。

[0265] 将清楚的是,此时该方法中的计算与使用L变换(等式7)创建Goertzel算法的理论非常不同。

[0266] 其实质上也与Z变换不同,Z变换与L变换相连。为了描述起见,我们在这里称这种方法中使用的变换是L变换的2D变型。

[0267] 本领域技术人员将清楚,通过该方法,仅需要计算索引中包含特征的频域值。而在使用FFT的情形中,必须计算频域中的所有值,因为FFT无法被稀疏地计算。同样重要的是,再次强调该方法不具有像FFT那样的图像输入大小的限制。实际上,甚至没有具有方形或矩形输入图像大小的限制,如图16所示。

[0268] 接下来,描述了将该方法应用于从视频帧序列中提取时间信息。

[0269] 图17示出了形成特征的每对区域不一定必须位于视频序列的同一帧中。而图18示出了核内环的每个输入的方向和顺序如何可以采用不限于视频帧序列的单个帧的路径。专家们将清楚,如果像素被直接用作来自空间域的输入,则与本文中所描述的方法相比,当在视频序列的各帧之间交叉引用像素时要求解的排列训练起来将慢得多且难得多。

[0270] 图19解说了该方法允许选择被用来生成模型的频域变换的帧序列的自由度,从而允许真实的时间数据被用于内容的检测;

[0271] 我们还在图20中看到,与像素的空间域相比,在频域中进行工作如何具有大的优势。而在空间域中,每个像素具有0到255之间的输入,而没有太多可能性来推断像素的良好性,相比之下在频域中,频域中的搜索空间最大部分是由大量噪声组成的,伴随着信号清晰地凸显出。因此,本文中所描述的方法还具有基于信号质量的更有效的优化逻辑的潜力,其潜在地也可以以无监督的方式完成。

[0272] 图21示出了当使用该方法创建检测视频帧序列中的时间效应的模型时可用于所

描述的方法的训练的可能的更高级别图。图21示出了序列中所使用的帧数如何是变量,并接着对于每个特征的每个索引(亦如图14所示),目标频率被优化,底层环路的序列被选择并且在环路内使用的底层传递函数被优化。这是针对每个环路的实部和虚部完成的。优化的输出将是具有多个经标准化的复向量的格式,其可以被用于检测和分类视频帧序列中的时间效益的尤其有效的输入,但是该方法不限于这样的实现。

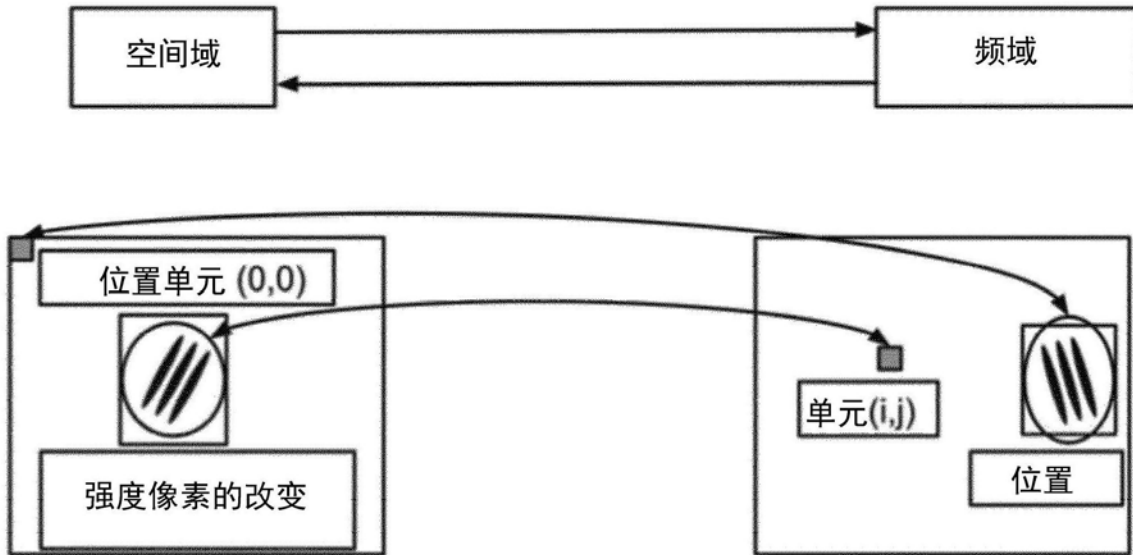


图1

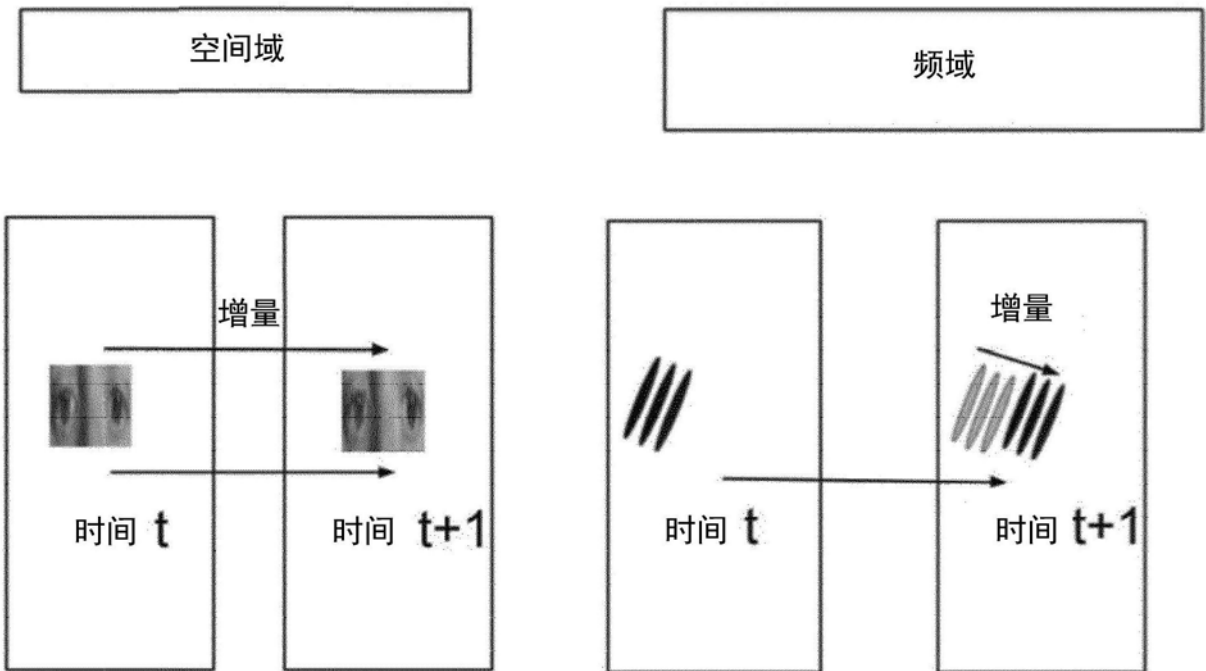


图2

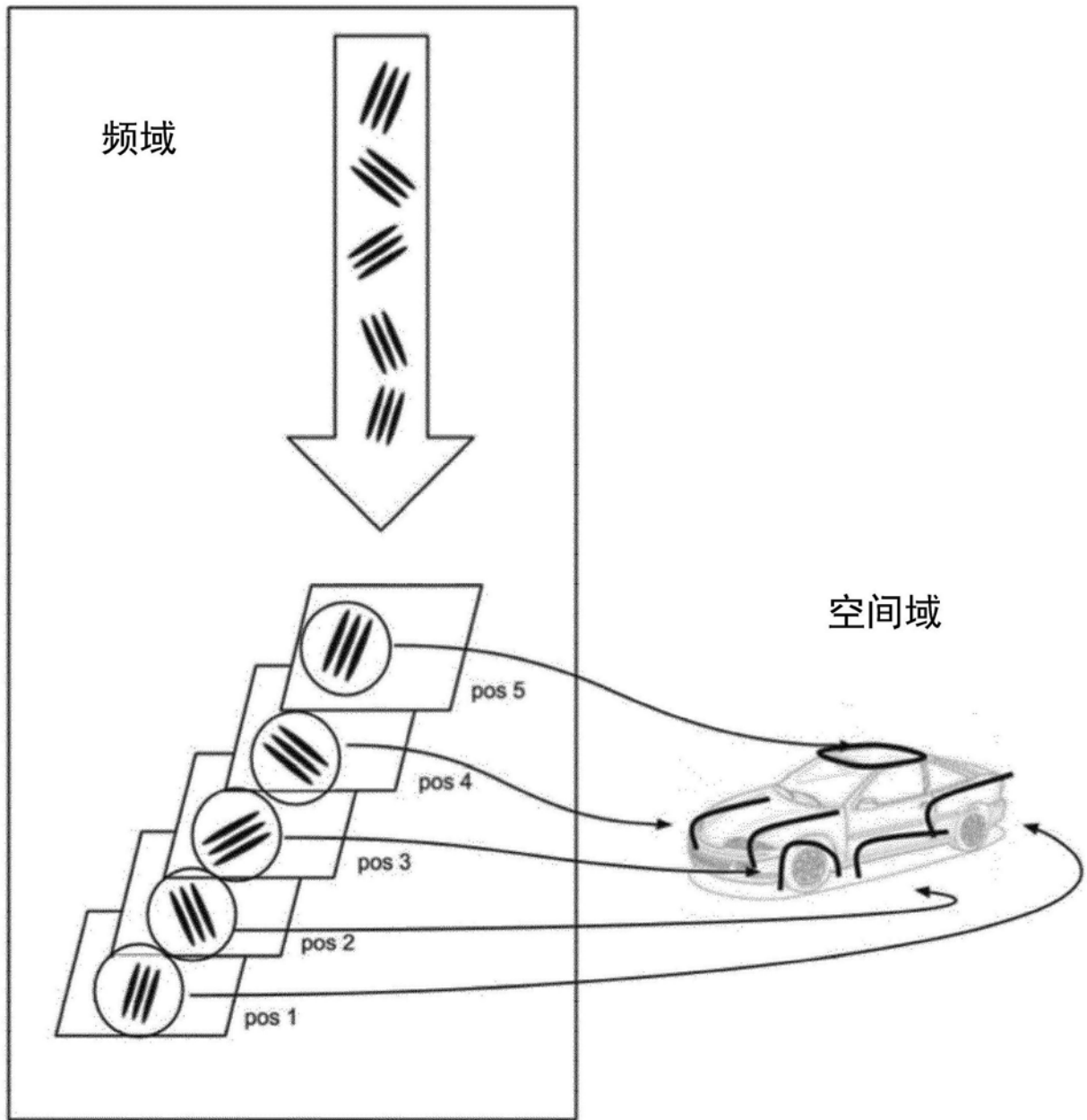


图3

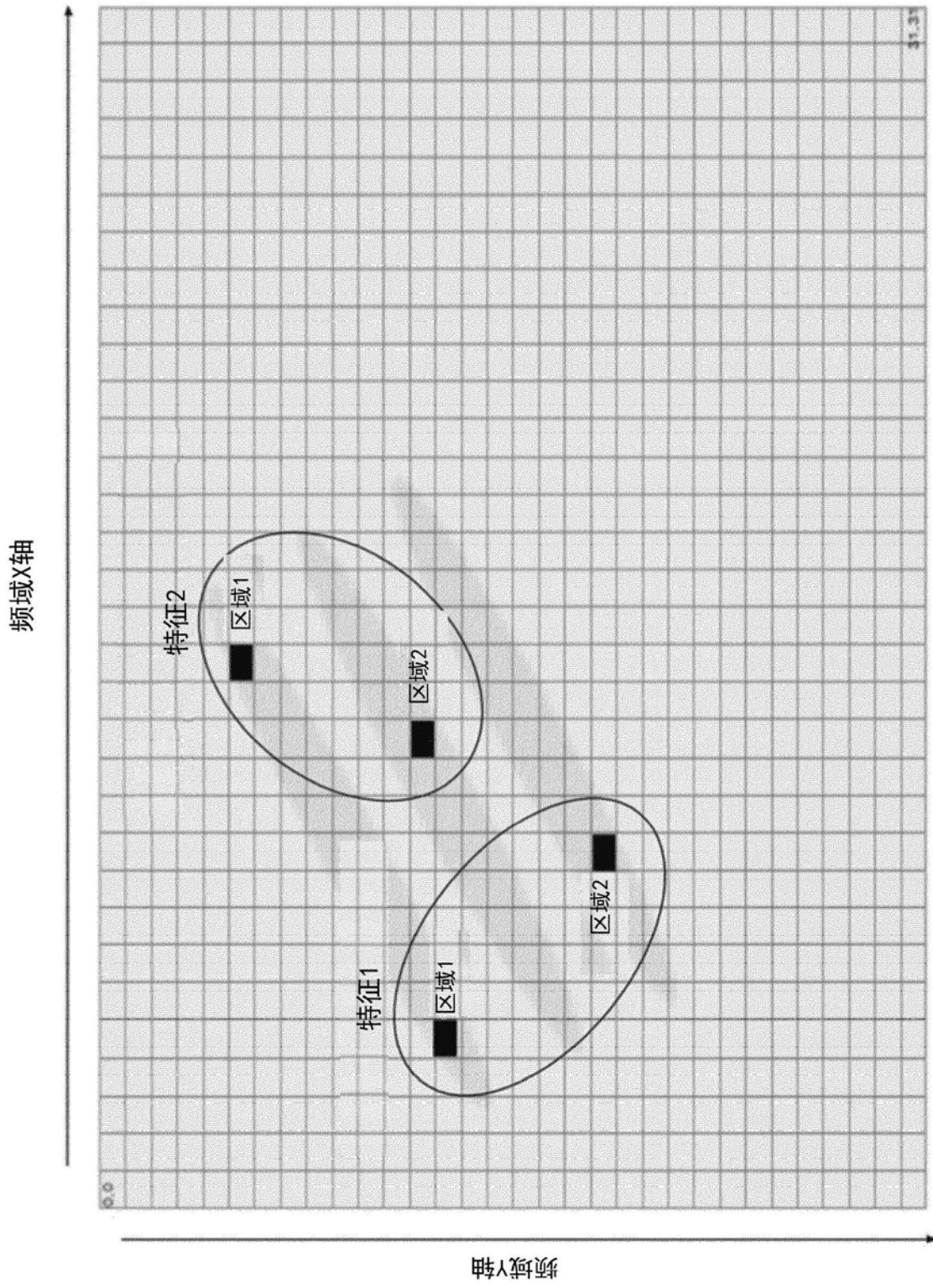


图4

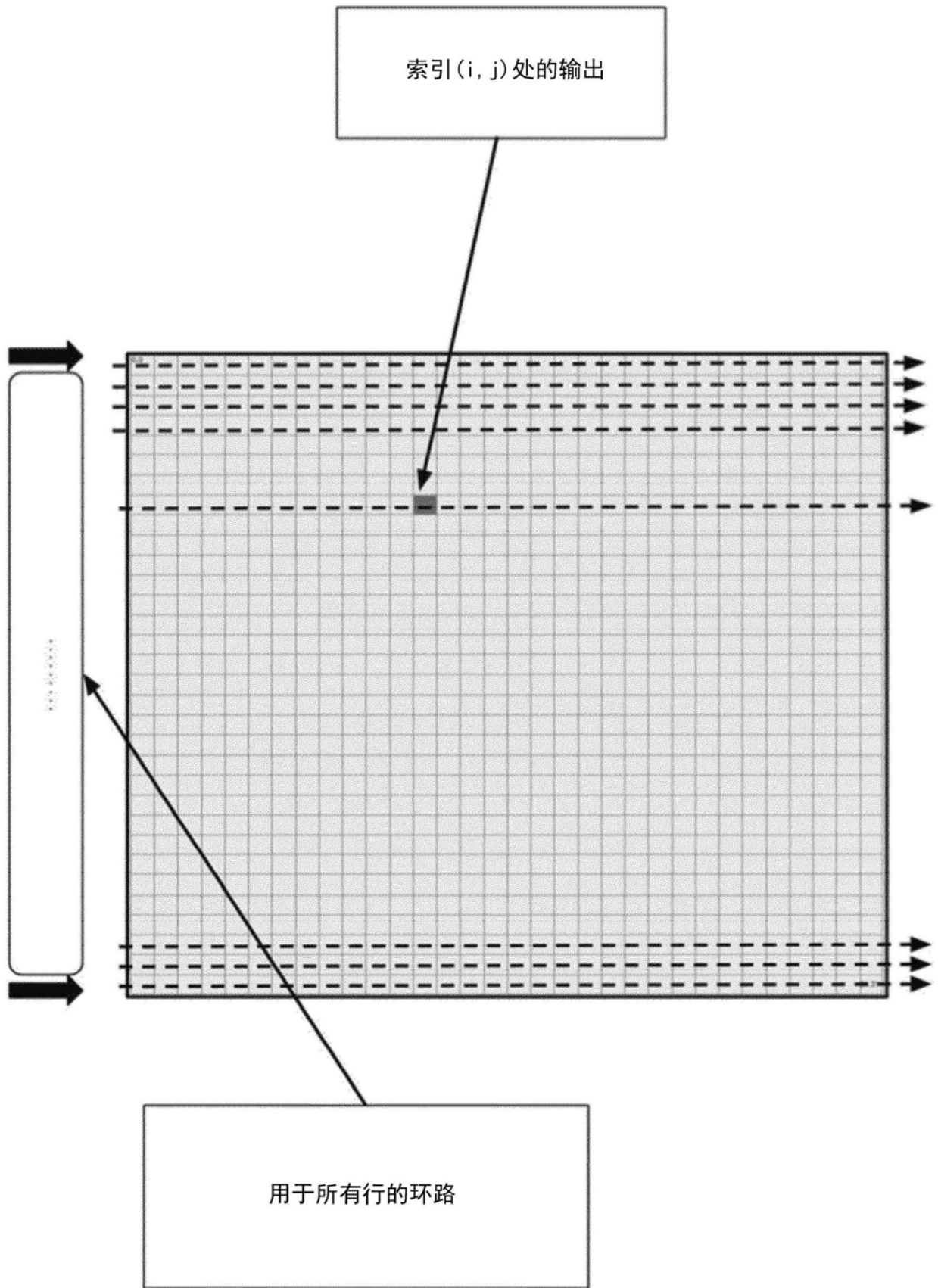


图5

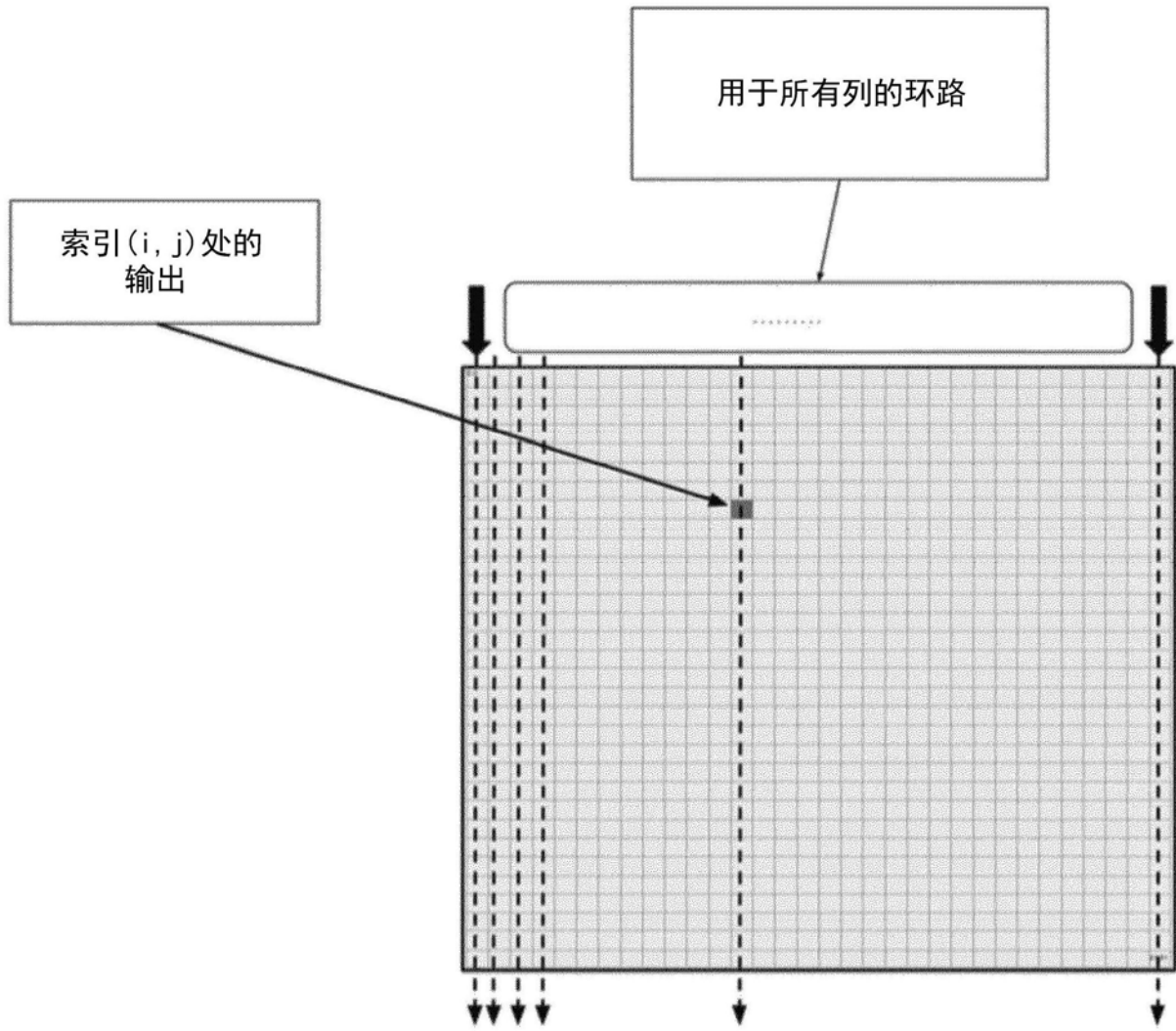


图6

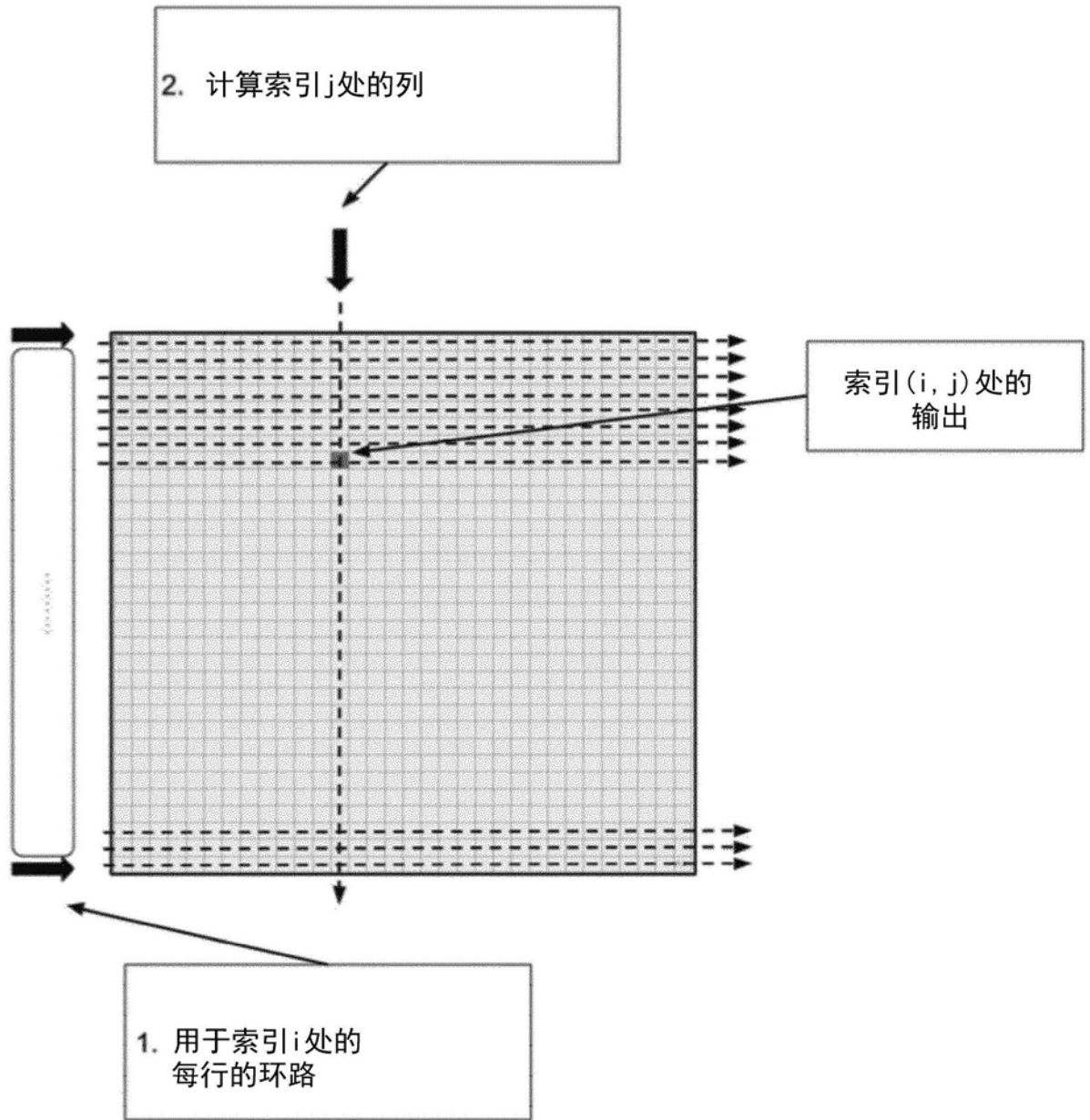


图7

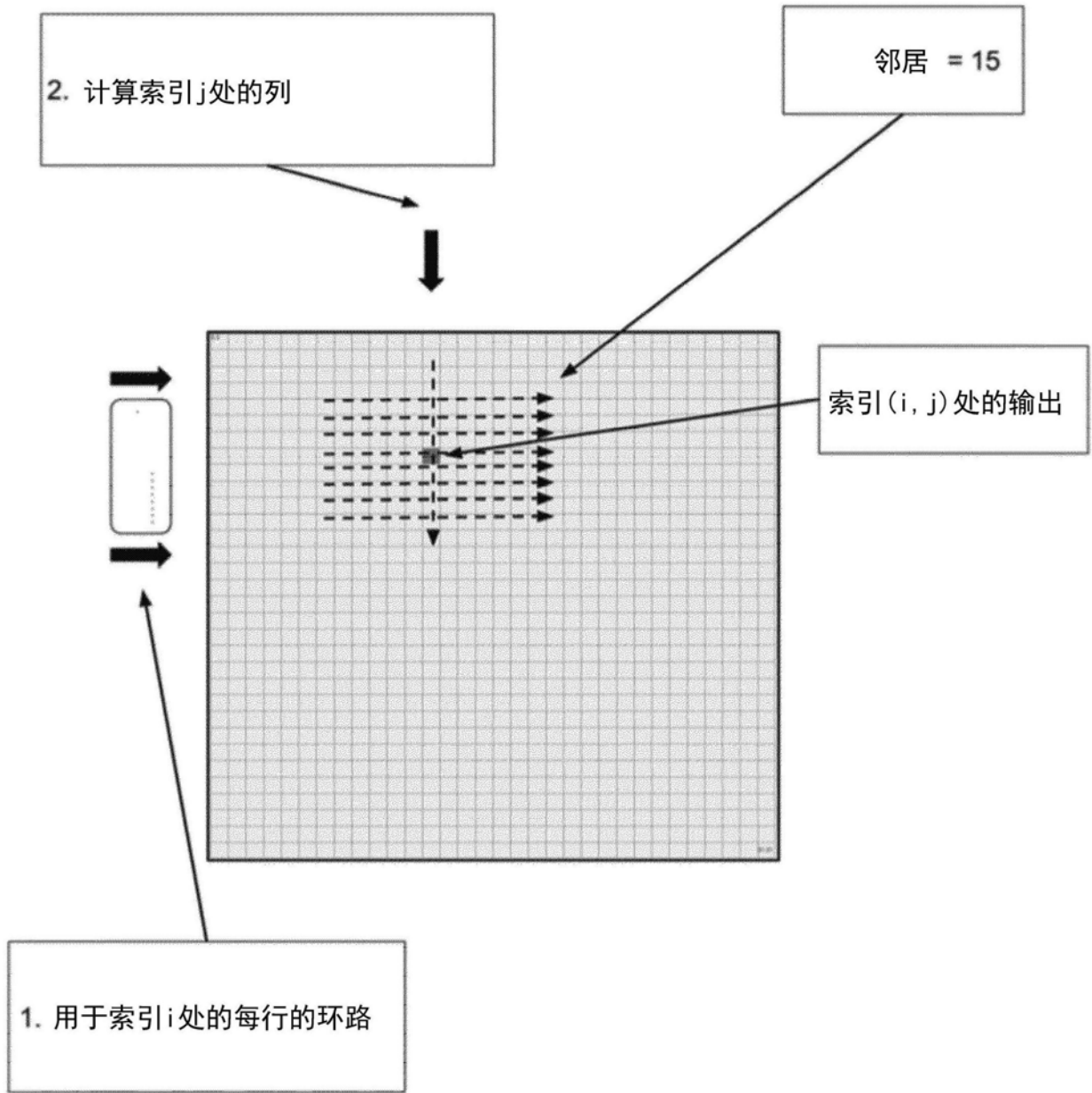


图8

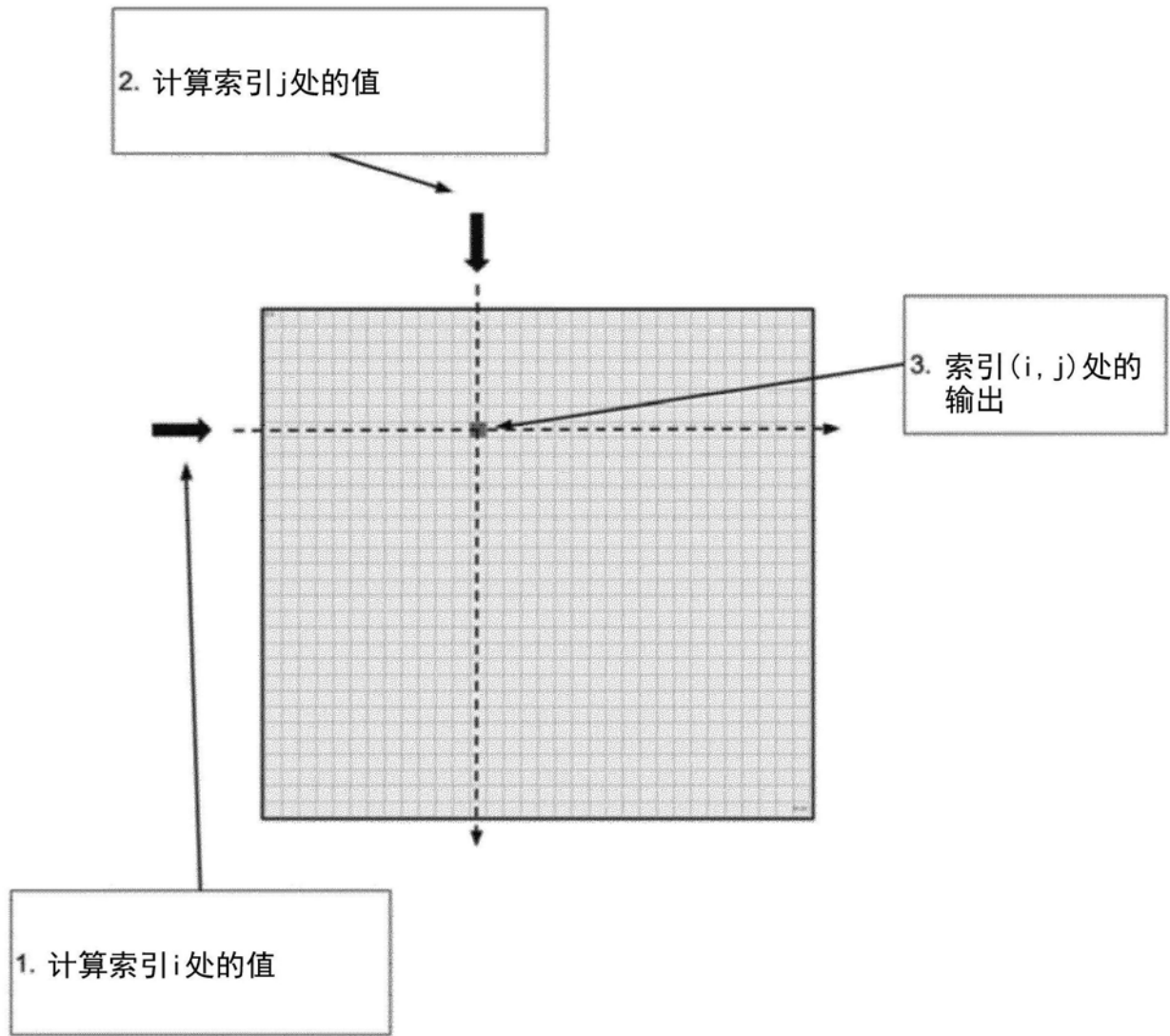


图9

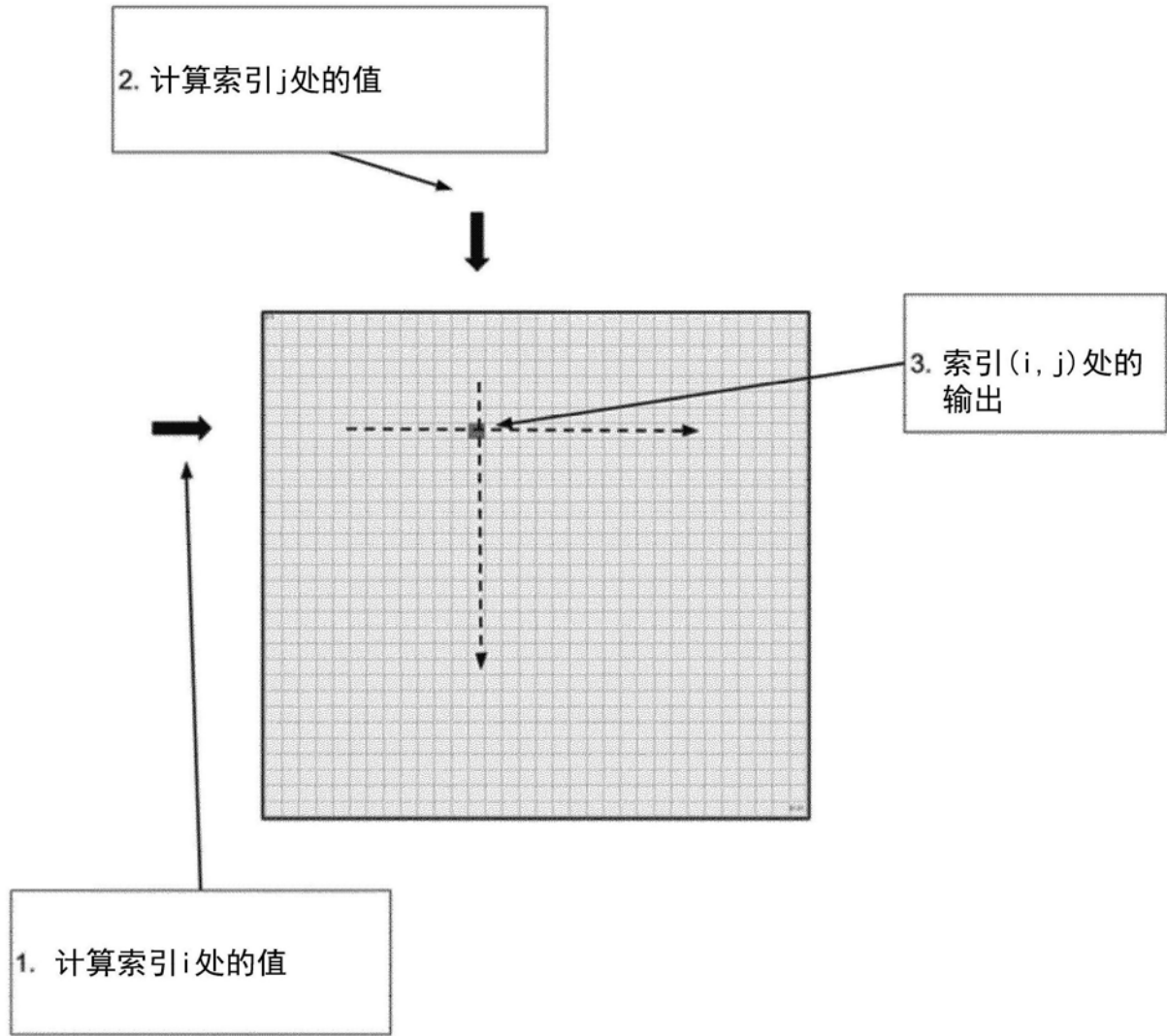


图10

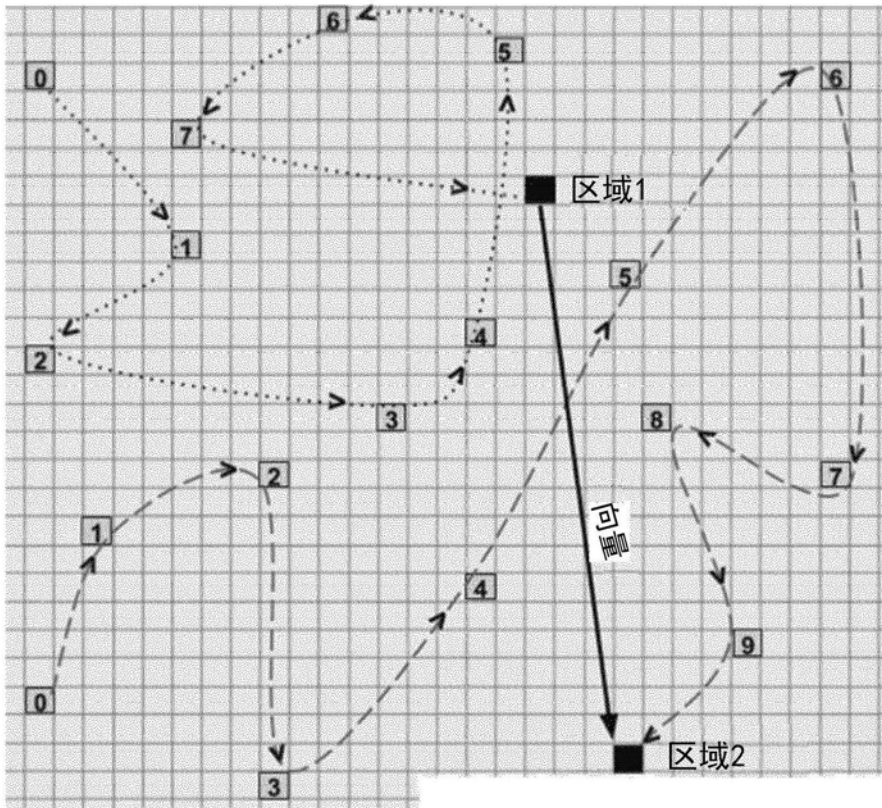


图11

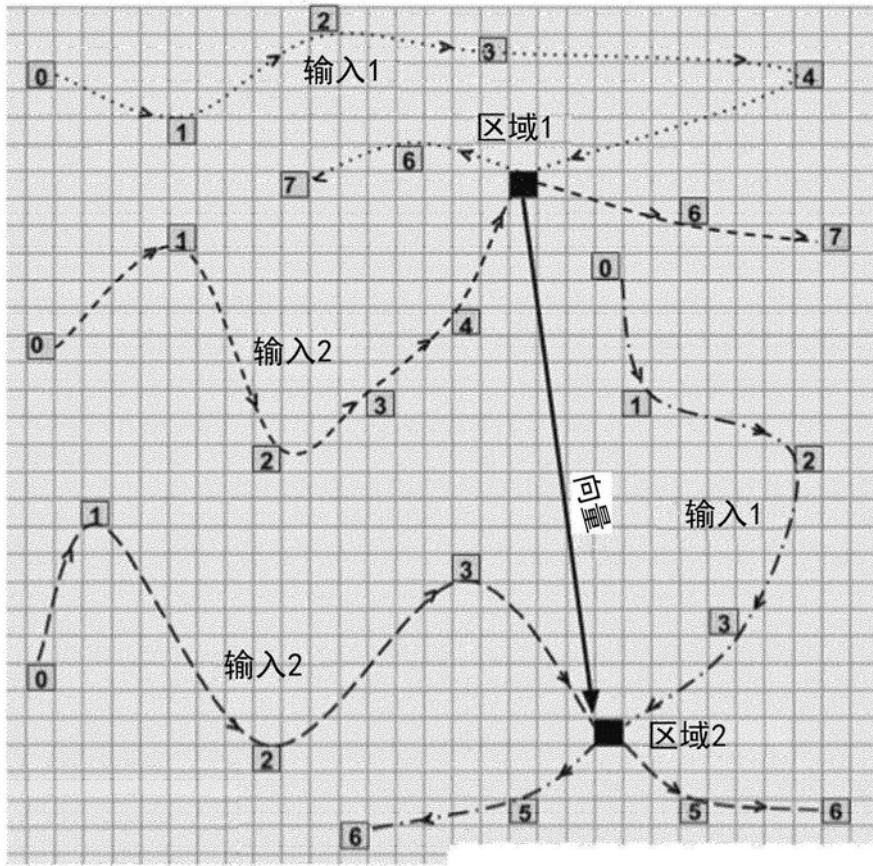


图12

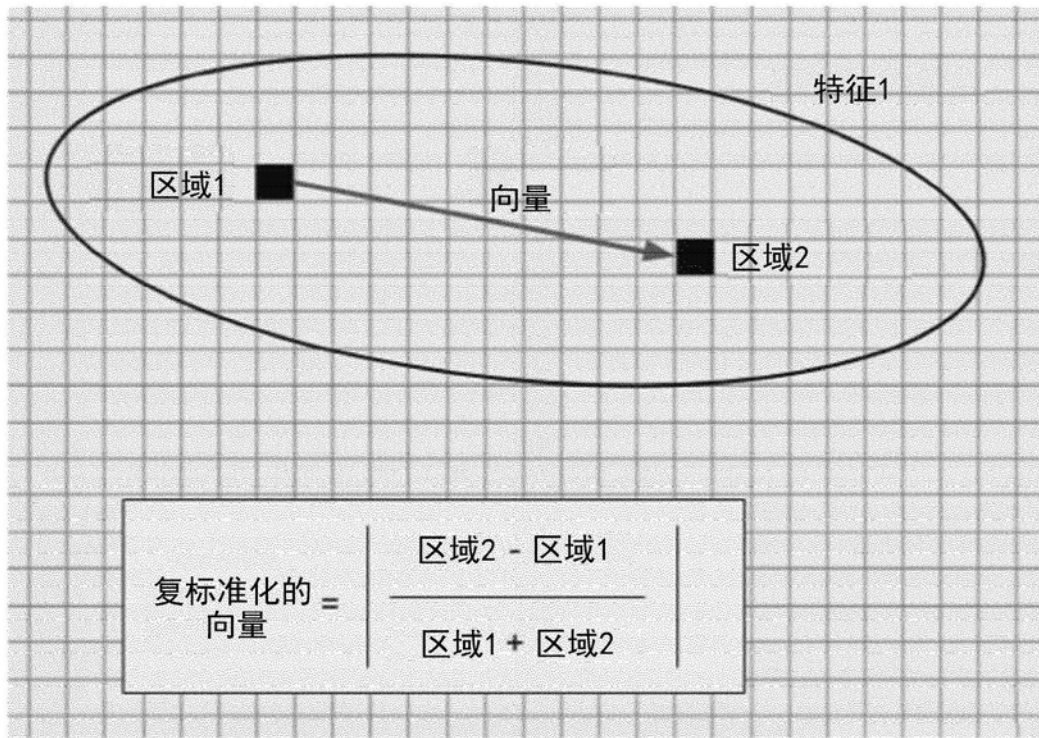


图13

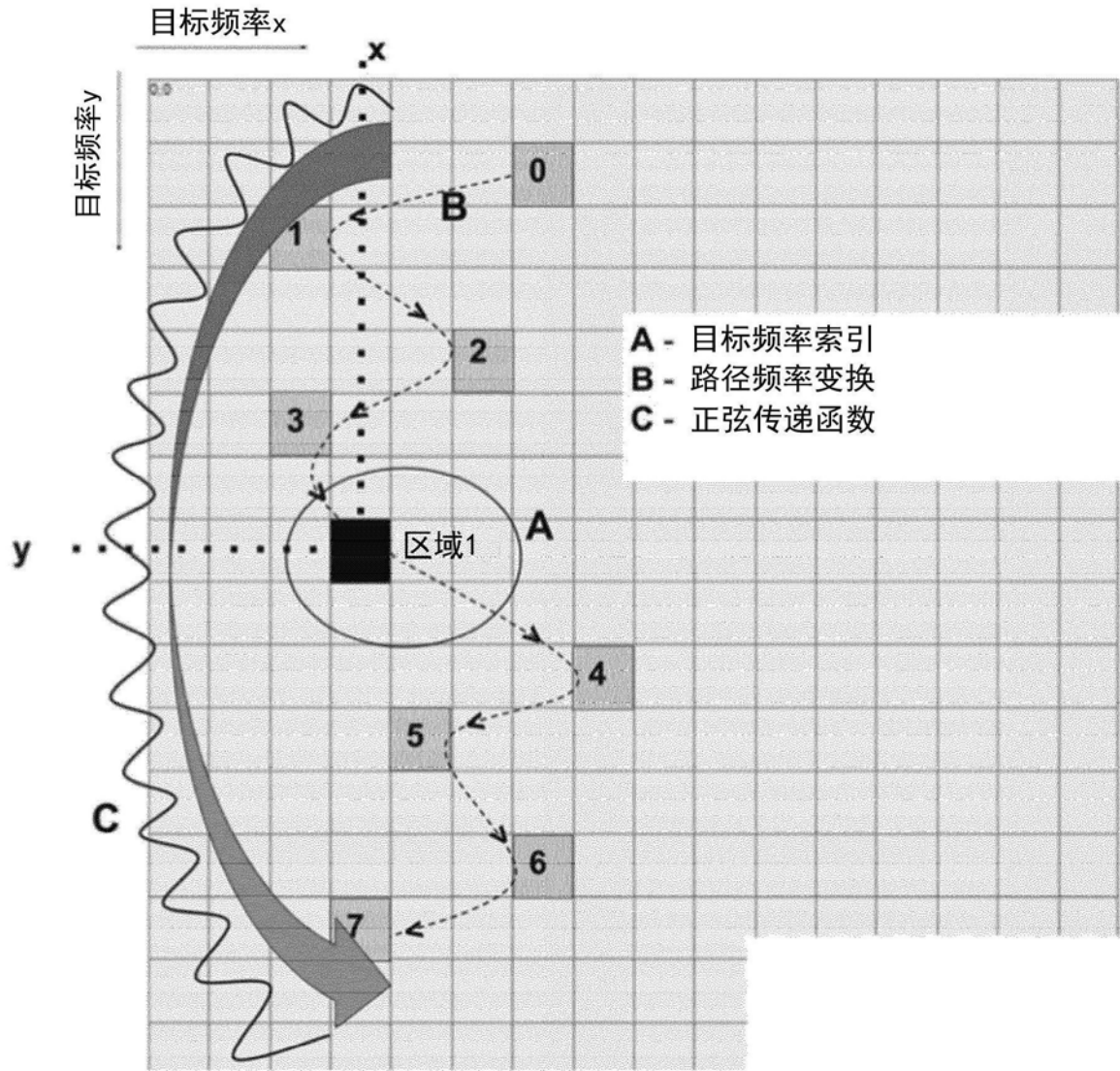


图14

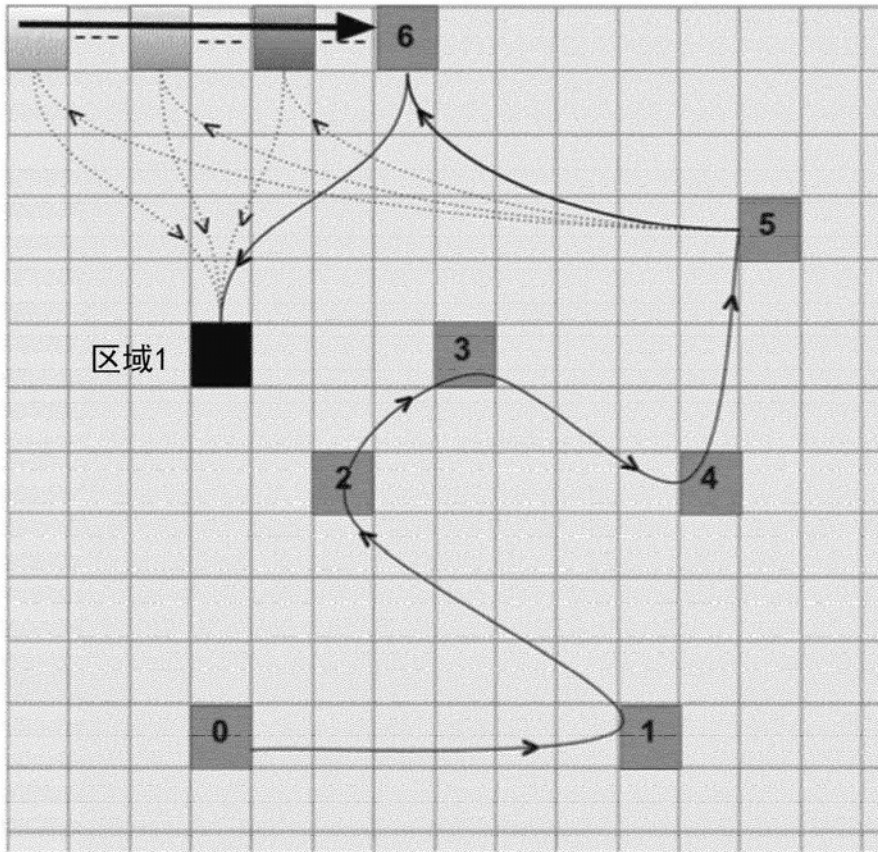


图15

任何图像形状可被用作输入

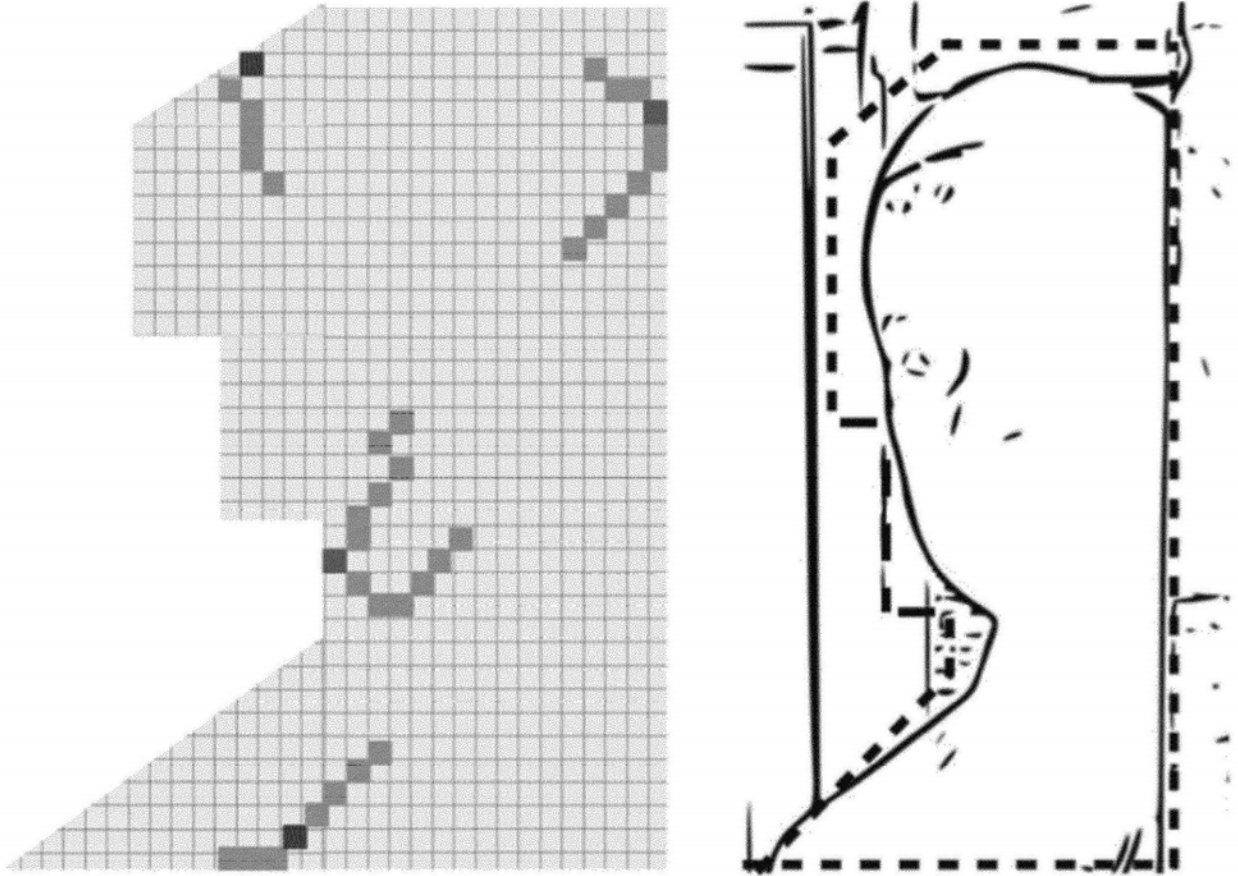


图16

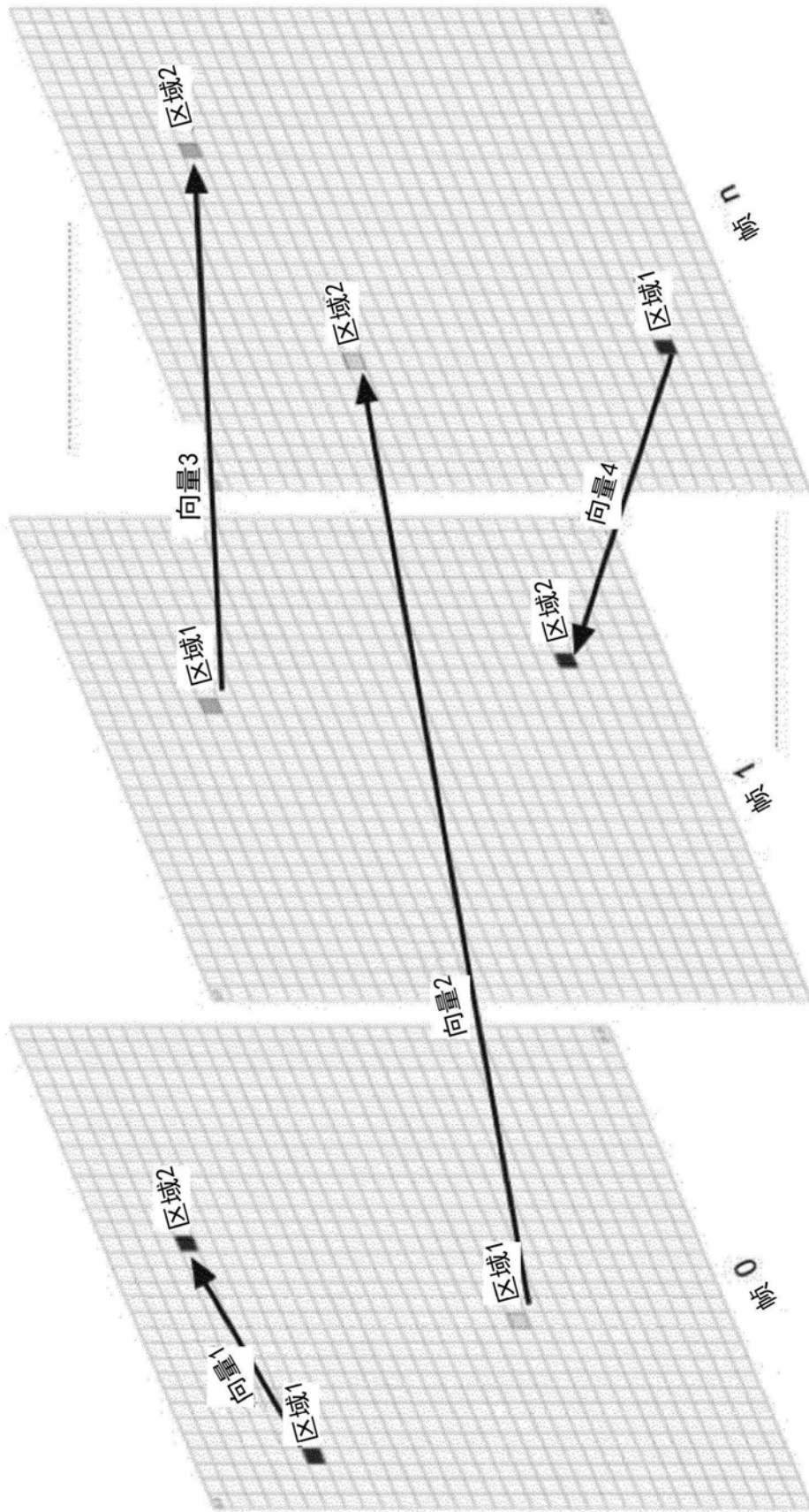


图17

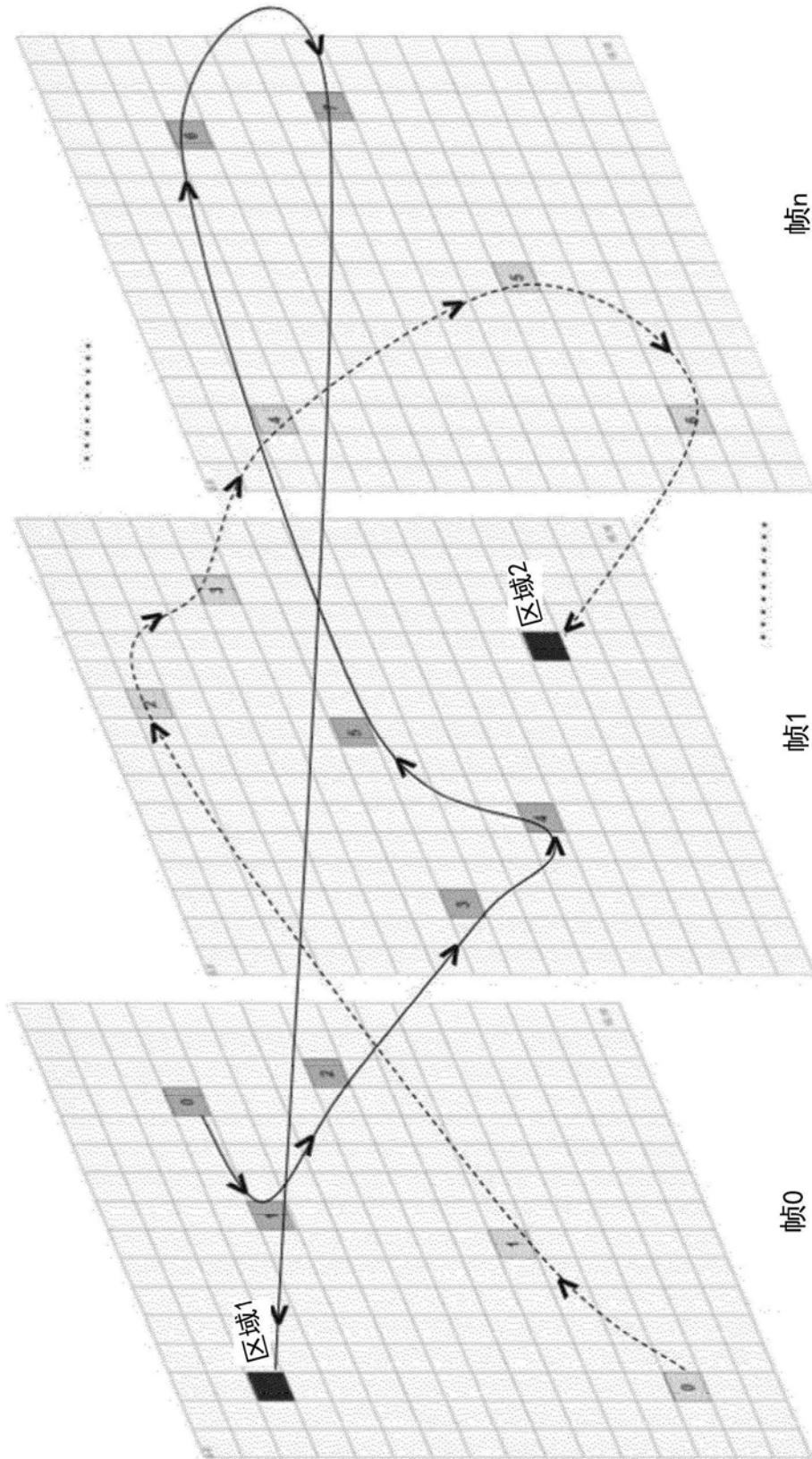


图18

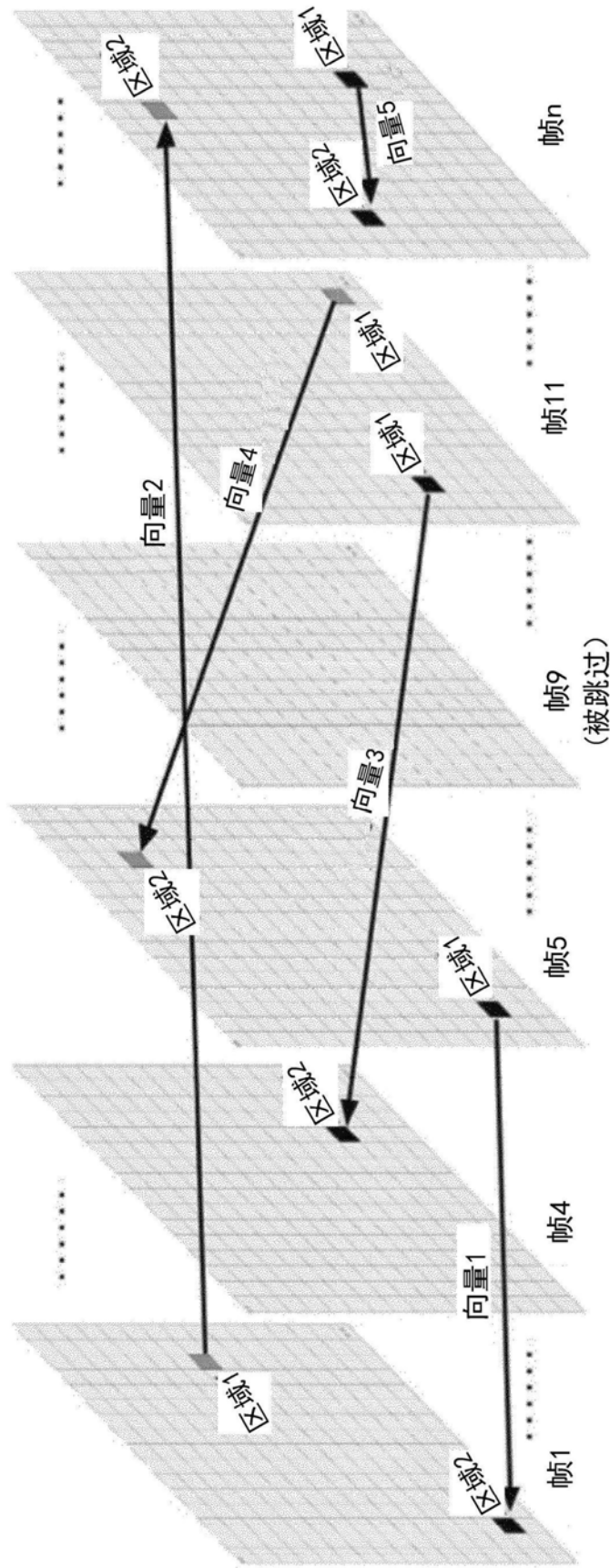


图19

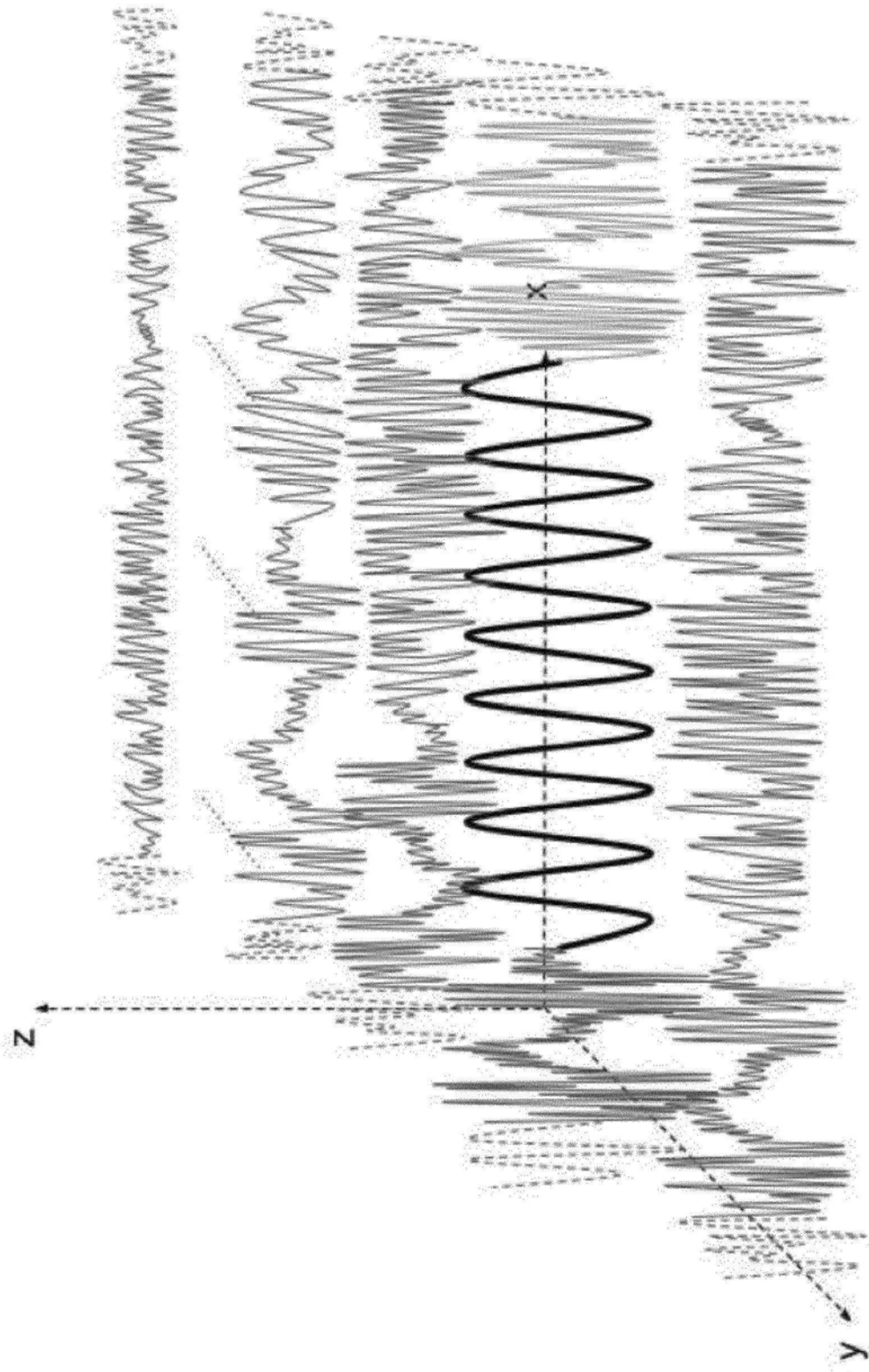


图20

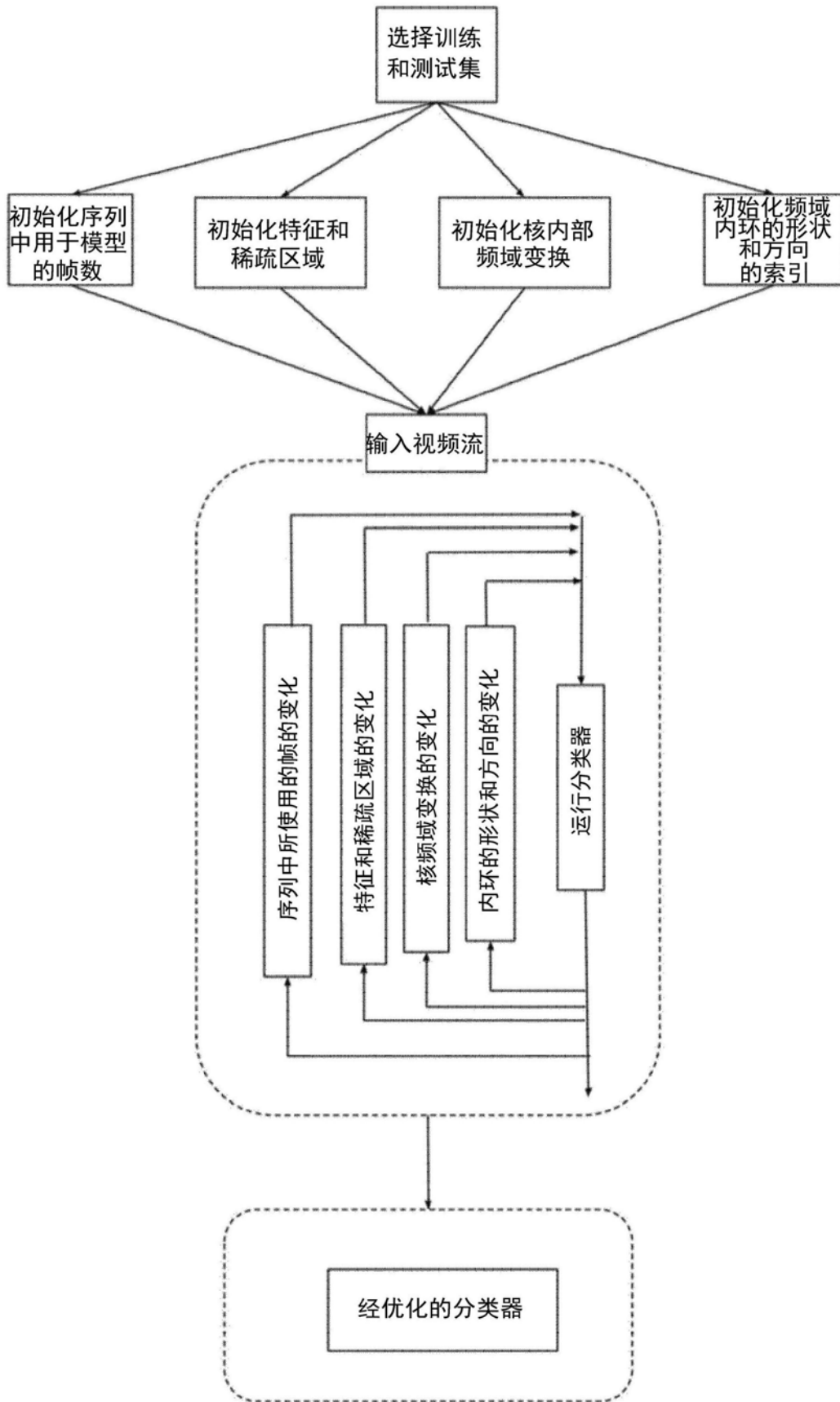


图21