



(12)发明专利申请

(10)申请公布号 CN 107666853 A

(43)申请公布日 2018.02.06

(21)申请号 201680029350.8

(22)申请日 2016.05.09

(30)优先权数据

15168575.7 2015.05.21 EP

(85)PCT国际申请进入国家阶段日

2017.11.21

(86)PCT国际申请的申请数据

PCT/EP2016/060250 2016.05.09

(87)PCT国际申请的公布数据

W02016/184705 EN 2016.11.24

(71)申请人 皇家飞利浦有限公司

地址 荷兰艾恩德霍芬

(72)发明人 G·德哈恩 王文锦

(74)专利代理机构 永新专利商标代理有限公司

72002

代理人 李光颖 王英

(51)Int.Cl.

A61B 5/00(2006.01)

A61B 5/024(2006.01)

G06K 9/00(2006.01)

G06K 9/62(2006.01)

G06T 7/00(2017.01)

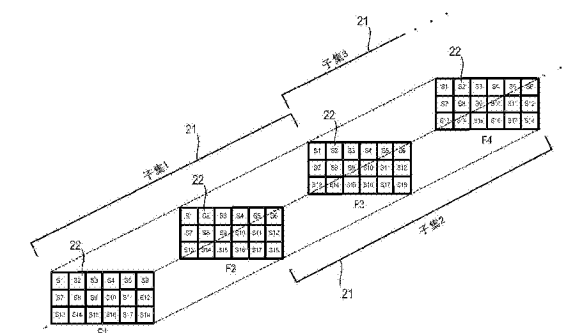
权利要求书2页 说明书18页 附图13页

(54)发明名称

根据视频序列确定搏动信号

(57)摘要

根据一方面,提供了一种用于根据视频序列确定搏动信号的装置,所述装置包括处理单元,所述处理单元被配置为:获得视频序列,所述视频序列包括多个图像帧;形成多个视频子序列,每个视频子序列包括来自图像帧的子集中的每个图像帧的帧分段,其中,每个图像帧被分成多个帧分段;针对根据来自图像帧的第一子集的帧分段形成的第一视频子序列,将针对所述第一视频子序列的代表值与针对根据来自图像帧的第二子集的帧分段形成的视频子序列的代表值进行比较;基于代表值的比较来将第一视频子序列与根据来自图像帧的第二子集的帧分段形成的第二视频子序列级联;并且根据级联的视频子序列确定搏动信号。



1. 一种用于根据视频序列确定搏动信号的装置,所述装置包括:

处理单元,其被配置为:

获得视频序列,所述视频序列包括多个图像帧;

形成多个视频子序列,每个视频子序列包括来自所述图像帧的子集中的每个图像帧的帧分段,其中,每个子集包括多个图像帧,其中,每个图像帧被划分成多个帧分段,其中,每个帧分段是所述图像帧中的近邻像素的组;

针对根据来自图像帧的第一子集的帧分段形成的第一视频子序列,将针对所述第一视频子序列的代表值与针对根据来自图像帧的第二子集的帧分段形成的视频子序列的代表值进行比较;

基于代表值的比较来将所述第一视频子序列与根据来自图像帧的所述第二子集的帧分段形成的第二视频子序列进行级联;并且

根据级联的视频子序列确定搏动信号。

2. 根据权利要求1所述的装置,所述处理单元还被配置为重复所述比较和所述级联以便将来自图像帧的多个子集中的每个子集的视频子序列进行级联。

3. 根据权利要求1或2所述的装置,其中,通过基于每个图像帧中的像素的颜色相似性和空间相似性将所述像素分组成帧分段来将所述图像帧划分成多个帧分段。

4. 根据权利要求1-3中的任一项所述的装置,其中,所述处理单元被配置为通过以下操作来形成所述多个视频子序列:针对每个视频子序列,从图像帧的所述子集中的每个图像帧选择帧分段,使得所述帧分段的色度之间的差异和/或所述视频子序列中的所述帧分段的位置之间的空间距离被最小化。

5. 根据权利要求1-4中的任一项所述的装置,其中,所述处理单元还被配置为通过对每个视频子序列中的所述帧分段中的像素的像素值进行平均来确定针对所述视频子序列的代表值。

6. 根据权利要求5所述的装置,其中,所述处理单元被配置为通过以下操作来对所述像素值进行平均:

对每个帧分段中的像素的所述像素值进行加权,其中,基于所述帧分段中的所述像素的空间位置和/或与所述帧分段的中心处或附近的像素或像素的组的颜色的差异来对所述像素值进行加权;并且

对帧分段中的像素的经加权的像素值进行平均。

7. 根据权利要求5或6所述的装置,其中,所述处理单元还被配置为通过对所述帧分段中的所述像素值的平均进行归一化来确定代表值。

8. 根据权利要求5-7中的任一项所述的装置,其中,所述处理单元被配置为通过确定每个视频子序列中的所述帧分段中的像素的所述像素值的所述平均之间的差异来确定针对所述视频子序列的代表值。

9. 一种用于根据视频序列确定搏动信号的方法,所述方法包括:

获得视频序列,所述视频序列包括多个图像帧;

形成多个视频子序列,每个视频子序列包括来自所述图像帧的子集中的每个图像帧的帧分段,其中,每个子集包括多个图像帧,其中,每个图像帧被划分成多个帧分段,其中,每个帧分段是所述图像帧中的近邻像素的组;

针对根据来自图像帧的第一子集的帧分段形成的第一视频子序列,将针对所述第一视频子序列的代表值与针对根据来自图像帧的第二子集的帧分段形成的视频子序列的代表值进行比较;

基于代表值的比较来将所述第一视频子序列与根据来自图像帧的所述第二子集的帧分段形成的第二视频子序列级联;并且

根据级联的视频子序列确定搏动信号。

10. 根据权利要求9所述的方法,所述方法还包括重复比较和级联的步骤,以便将来自图像帧的多个子集中的每个子集的视频子序列级联。

11. 根据权利要求10所述的方法,其中,通过基于像素的颜色相似性和空间相似性将所述像素分组成帧分段来将每个图像帧划分成多个帧分段。

12. 根据权利要求10或11所述的方法,其中,形成所述多个视频子序列的步骤包括:针对每个视频子序列,从图像帧的所述子集中的每个图像帧选择帧分段,使得所述帧分段的色度之间的差异和/或所述视频子序列中的所述帧分段的位置之间的空间距离被最小化。

13. 根据权利要求12所述的方法,所述方法还包括以下步骤:通过对每个视频子序列中的所述帧分段中的像素的像素值进行平均来确定针对所述视频子序列的代表值。

14. 根据权利要求13所述的方法,其中,对所述像素值进行平均的步骤包括:

对每个帧分段中的像素的所述像素值进行加权,其中,基于所述帧分段中的所述像素的空间位置和/或与所述帧分段的中心处或附近的像素或像素的组的颜色的差异来对所述像素值进行加权;并且

对帧分段中的像素的经加权的像素值进行平均。

15. 一种包括计算机可读介质的计算机程序产品,所述计算机可读介质具有在其中实现的计算机可读代码,所述计算机可读代码被配置为使得在由合适的计算机或处理器运行时令所述计算机或处理器执行根据权利要求9-14中的任一项所述的方法。

根据视频序列确定搏动信号

技术领域

[0001] 本发明涉及一种用于根据视频序列确定搏动信号的装置和方法。

背景技术

[0002] 近来,已经开发了用于执行远程光体积描记(远程PPG或rPPG)的技术。这些技术使得能够从使用成像单元(例如相机)捕获的图像帧的视频序列获得PPG信号。期望视频序列被处理并且rPPG信号被自动提取,使得对象可以被自动监测。然而,这需要在视频序列中自动识别活皮肤组织的区域。

[0003] 在视频中检测对象的任务作为计算机视觉的基本课题之一在过去几十年中已经得到了广泛的研究。给定包含对象的视频序列,目标是定位与对象的身体部分对应的区域。大多数现有的工作利用人的外观特征在受监督的训练机制中区分对象与背景。然而,这些方法的常见问题在于,它们的经训练的特征不是人类所独有的,与人类外观相似的任何特征都可能被错误分类。此外,受监督的方法通常限于先前已知的样本,并且在不可预测的样本出现时趋于失败,例如用前脸训练的面部检测器无法定位从侧面查看的面部,而用亮皮肤对象训练的皮肤分类器对于暗皮肤对象失败。其他类型的方法需要在视频序列中手动选择皮肤的区,该区随着时间被跟踪以补偿运动。然而,该技术清楚地需要手动输入,并且当存在实质运动时不容易正确地跟踪选定的区。

[0004] 基于rPPG技术的发展,已经观察到,与物理外观特征相比,不可见的生理特征(例如搏动)可以更好地区分视频序列中的人类与非人类。在自然环境中,仅活着的对象的皮肤组织呈现搏动性,因此不示出搏动信号的任何客体都可以安全地被分类到非人类类别中。这可以防止对与人类有相似外观的客体的错误检测,例如如图1中所示。

[0005] 图1提供了活组织检测技术应当如何成功操作的两个范例。在左手图像中,人面部和人工面部都正面呈现于相机,并且仅人面部应当被识别(尽管人工面部具有与人面部相似的物理外观特征),如虚线框和对应于活皮肤组织的区的轮廓指示的。在右手图像中,人面部和人工面部侧面呈现于相机,并且仅出人面部应当被识别。

[0006] 在G.Gibert和D.D'Alessandro和F.Lance的文章“Face detection method based on photoplethysmography”(10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp.449-453, (2013))中,设置了硬阈值以选择具有较高频谱能量的分割的局部区域(例如网格、三角形或体素)作为皮肤区域。在R.van Luijtelaaar、W.Wang、S.Stuijk和G.de Haan的文章“Automatic ROI detection for rPPG measurements”(ACCV2014, Singapore)中,预定义聚类参数被用于将与皮肤区域共享相似性的区域聚类。

[0007] 然而,当存在对象或皮肤区的显著移动时,这些方法仍然要努力提取有用的搏动信号。

[0008] 因此,目的是提供一种用于根据视频序列确定搏动信号的改进方法和装置。

发明内容

[0009] 根据第一方面,提供了一种用于根据视频序列确定搏动信号的方法,所述方法包括获得视频序列,所述视频序列包括多个图像帧;形成多个视频子序列,每个视频子序列包括来自图像帧的子集中的每个图像帧的帧分段,其中,每个子集包括多个图像帧,其中,每个图像帧被划分成多个帧分段,其中,每个帧分段是图像帧中的近邻像素的组;针对根据来自图像帧的第一子集的帧分段形成的第一视频子序列,将第一视频子序列的代表值与根据来自图像帧的第二子集的帧分段形成的视频子序列的代表值进行比较;基于代表值的比较将第一视频子序列与根据来自图像帧的第二子集的帧分段形成的第二视频子序列级联;并且根据级联的视频子序列确定搏动信号。

[0010] 在一些实施例中,该方法还包括以下步骤:针对第二视频子序列,将第二视频子序列的代表值与根据来自图像帧的第三子集的帧分段形成的视频子序列的代表值进行比较;并且基于代表值的比较将第一视频子序列和第二视频子序列与根据来自图像帧的第三子集的帧分段形成的第三视频子序列级联。

[0011] 在一些实施例中,该方法还包括重复比较和级联的步骤,以便级联来自多个图像帧子集中的每个的视频子序列。

[0012] 在一些实施例中,图像帧的每个子集包括视频序列中的连续图像帧的相应的集合。

[0013] 在一些实施例中,每个图像帧在多于一个子集中,使得图像帧的子集交叠。

[0014] 在一些实施例中,每个子集包括与另一个子集公共的至少一个与图像帧。

[0015] 在备选实施例中,视频序列中的每个图像帧处于图像帧的相应的子集中。

[0016] 在一些实施例中,图像帧的每个子集包括视频序列中的三个图像帧的相应的集合。

[0017] 在一些实施例中,图像帧的第一子集和图像帧的第二子集包括在视频序列中邻近的图像帧。

[0018] 在一些实施例中,每个帧分段是图像帧中的近邻像素的组。

[0019] 在一些实施例中,帧分段具有相同的形状。在备选实施例中,通过基于像素的颜色相似性和空间相似性将像素分组成帧分段来将每个图像帧划分成多个帧分段。

[0020] 在一些实施例中,每个图像帧独立于其他图像帧被划分成多个帧分段。在备选实施例中,基于图像帧和子集中的其他图像帧的内容将子集中的每个图像帧分成多个帧分段。

[0021] 在一些实施例中,形成多个视频子序列的步骤包括根据图像帧子集中的图像帧中的对应空间位置的帧分段形成每个视频子序列。

[0022] 在备选实施例中,形成多个视频子序列的步骤包括,针对每个视频子序列,从图像帧子集中的每个图像帧中选择帧分段,使得在视频子序列中的帧分段之间的色能和/或空间距离能量被最小化。

[0023] 在一些实施例中,该方法还包括确定每个视频子序列的代表值的步骤。

[0024] 在一些实施例中,确定每个视频子序列的代表值的步骤可以包括将视频子序列中的帧分段中的像素的像素值进行平均。

[0025] 在一些实施例中,对像素值进行平均的步骤可以包括对每个帧分段中的像素的像素值进行加权,其中,基于帧分段中的像素的空间位置和/或与帧分段的中心处或附近的像素或像素的组的颜色差异对像素值进行加权;并且对帧分段中的像素的加权像素值进行平均。

[0026] 在一些实施例中,确定代表值的步骤还包括对帧分段中的像素值的平均进行归一化。

[0027] 在一些实施例中,确定每个视频子序列的代表值的步骤包括确定视频子序列中的帧分段中的像素的像素值的平均之间的差异。

[0028] 在一些实施例中,比较的步骤包括将第二视频子序列识别为根据来自与第一视频子序列相似或最相似的图像帧的第二子集的帧分段形成的视频子序列。

[0029] 在一些实施例中,比较的步骤包括将第二视频子序列识别为根据来自在空间距离和/或代表值中与第一视频子序列相似或最相似的图像帧的第二子集的帧分段形成的视频子序列。

[0030] 在一些实施例中,确定搏动信号的步骤包括根据级联的视频子序列的代表值确定搏动信号。

[0031] 在一些实施例中,针对被划分为第一多个帧分段的图像帧执行形成、确定代表值、比较、级联和确定搏动信号的步骤;并且其中,所述方法还包括以下步骤:当图像帧被划分成第二多个帧分段时,针对视频序列中的图像帧重复形成、确定代表值、比较、级联和确定搏动信号的步骤,所述第二多个帧分段包括与所述第一多个帧分段不同数量的帧分段。

[0032] 在一些实施例中,该方法还包括以下步骤:针对根据来自图像帧的第一子集的帧分段形成的其他视频子序列重复比较、级联和确定搏动信号的步骤以确定另外的搏动信号;并且分析搏动信号以识别视频序列中活皮肤组织的区域。

[0033] 在一些实施例中,该方法还包括根据与视频序列中的所识别的活皮肤组织的区相关联的一个或多个搏动信号确定一个或多个生理特性的步骤。

[0034] 根据第二方面,提供了一种计算机程序产品,其包括具有实现在其中的计算机可读代码的计算机可读介质,所述计算机可读代码被配置为使得在由合适的计算机或处理器运行时,令所述计算机或处理器执行上述任何方法。

[0035] 根据第三方面,提供了一种用于根据视频序列确定搏动信号的装置,所述装置包括处理单元,所述处理单元被配置为:获得视频序列,所述视频序列包括多个图像帧;形成多个视频子序列,每个视频子序列包括来自图像帧的子集中的每个图像帧的帧分段,其中,每个子集包括多个图像帧,其中,每个图像帧被划分成多个帧分段,其中,每个帧分段是图像帧中的近邻像素的组;针对根据来自图像帧的第一子集的帧分段形成的第一视频子序列,将第一视频子序列的代表值与根据来自图像帧的第二子集的帧分段形成的视频子序列的代表值进行比较;基于代表值的比较将第一视频子序列与根据来自图像帧的第二子集的帧分段形成的第二视频子序列级联;并且根据级联的视频子序列确定搏动信号。

[0036] 在一些实施例中,处理单元还被配置为,针对第二视频子序列,将第二视频子序列的代表值与根据来自图像帧的第三子集的帧分段形成的视频子序列的代表值进行比较;并且基于代表值的比较将第一视频子序列和第二视频子序列与根据来自图像帧的第三子集的帧分段形成的第三视频子序列级联。

- [0037] 在一些实施例中,处理单元还被配置为重复比较和级联以便级联来自图像帧的多个子集中的每个的视频子序列。
- [0038] 在一些实施例中,图像帧的每个子集包括视频序列中连续图像帧的相应的集合。
- [0039] 在一些实施例中,每个图像帧在多于一个子集中,使得图像帧的子集交叠。
- [0040] 在一些实施例中,每个子集包括与另一个子集共有的至少一个图像帧。
- [0041] 在备选实施例中,视频序列中的每个图像帧处于图像帧的相应的子集中。
- [0042] 在一些实施例中,图像帧的每个子集包括视频序列中的三个图像帧的相应的集合。
- [0043] 在一些实施例中,图像帧的第一子集和图像帧的第二子集包括在视频序列中邻近的图像帧。
- [0044] 在一些实施例中,每个帧分段是图像帧中的近邻像素的组。
- [0045] 在一些实施例中,帧分段具有相同的形状。在备选实施例中,通过基于像素的颜色相似性和空间相似性将像素分组成帧分段来将每个图像帧划分成多个帧分段。
- [0046] 在一些实施例中,每个图像帧独立于其他图像帧被划分成多个帧分段。在备选实施例中,基于图像帧和子集中的其他图像帧的内容将子集中的每个图像帧划分成多个帧分段。
- [0047] 在一些实施例中,处理单元被配置为通过根据图像帧子集中的图像帧中的对应的空间位置中的帧分段形成每个视频子序列来形成多个视频子序列。
- [0048] 在备选实施例中,处理单元被配置为通过针对每个视频子序列从图像帧的子集中的每个图像帧选择帧分段使得视频子序列中的帧分段之间的色能和/或空间距离能量被最小化来形成多个视频子序列。
- [0049] 在一些实施例中,处理单元还被配置为确定每个视频子序列的代表值。
- [0050] 在一些实施例中,处理单元被配置为能够通过对视频子序列中的帧分段中的像素的像素值进行平均来确定每个视频子序列的代表值。
- [0051] 在一些实施例中,处理单元被配置为通过对每个帧分段中的像素的像素值进行加权来平均像素值,其中,基于帧分段中的像素的空间位置和/或与帧分段的中心处或附近的像素或像素的组的颜色的差异对像素值进行加权;并且对帧分段中的像素的加权的像素值进行平均。
- [0052] 在一些实施例中,处理单元被配置为通过归一化帧分段中的像素值的平均来确定代表值。
- [0053] 在一些实施例中,处理单元被配置为通过确定视频子序列中的帧分段中的像素的像素值的平均之间的差异来确定每个视频子序列的代表值。
- [0054] 在一些实施例中,处理单元被配置为通过将第二视频子序列识别为根据来自与第一视频子序列相似或最相似的图像帧的第二子集的帧分段形成的视频子序列来进行比较。
- [0055] 在一些实施例中,处理单元被配置为通过将第二视频子序列识别为根据来自在空间距离和/或代表值中与第一视频子序列相似或最相似的图像帧的第二子集的帧分段形成的视频子序列来进行比较。
- [0056] 在一些实施例中,处理单元被配置为通过从级联的视频子序列的代表值确定搏动信号来确定搏动信号。

[0057] 在一些实施例中,处理单元被配置为针对被划分成第一多个帧分段的图像帧形成、确定代表值、比较、级联并确定搏动信号;并且处理单元还被配置为,当图像帧被划分成第二多个帧分段时,针对视频序列中的图像帧重复形成、确定代表值、比较、级联和确定搏动信号,第二多个帧分段包括与第一多个帧分段不同数量的帧分段。

[0058] 在一些实施例中,处理单元还被配置为,针对根据来自图像帧的第一子集的帧分段形成的其他视频子序列重复比较、级联并确定搏动信号,以确定另外的搏动信号;并且分析搏动信号以识别视频序列中活皮肤组织的区。

[0059] 在一些实施例中,处理单元还被配置为根据与视频序列中的所识别的活皮肤组织的区相关联的一个或多个搏动信号确定一个或多个生理特性。

附图说明

[0060] 为了更好地理解本发明,并且更清楚地显示如何实施其,现在将仅通过范例参考附图,其中:

[0061] 图1示出了活皮肤组织检测技术的期望操作;

[0062] 图2是根据本发明的实施例的装置的框图;

[0063] 图3是图示根据本发明的实施例的方法的流程图;

[0064] 图4(a)-(d)示出了能够如何从视频序列获得搏动信号;

[0065] 图4(a)图示了视频序列如何由一系列图像帧组成;

[0066] 图4(b)图示了图像帧中的每个如何被划分成多个帧分段;

[0067] 图4(c)图示了如何使用图像帧内的相同空间位置中的帧分段形成两个视频子序列;

[0068] 图4(d)图示了如此形成的两个视频子序列的示范性搏动信号;

[0069] 图5是图示示范性体素搏动光谱方法中的处理阶段的图。

[0070] 图6图示了三个不同尺度的图像帧的分割;

[0071] 图7图示了成对相似性和所得到的相似性矩阵的四种不同量度;

[0072] 图8示出了使用增量稀疏PCA的相似性矩阵分解的范例;并且

[0073] 图9图示了特征向量到分级体素的投影以及指示视频序列的哪些部分对应于活皮肤组织的融合图。

具体实施方式

[0074] 根据本发明的实施例的可以用于根据视频序列确定搏动信号的装置2在图2中示出。在另外的实施例中,装置2可以用于根据所确定的搏动信号识别活皮肤组织。装置2包括在一段时间上捕获视频序列的成像单元4。成像单元4可以是或包括能够被用于rPPG测量的相机,例如RGB相机。成像单元4将包括多个图像帧的视频序列提供到处理单元6。

[0075] 处理单元6控制装置2的操作,并且可以包括一个或多个处理器、多核处理器或处理模块以用于实施本文描述的活皮肤组织识别技术。在一些实施例中,处理单元6可以被实施为多个处理模块,其中了,每个模块被配置为执行本文描述的活皮肤组织识别技术的特定部分或步骤。

[0076] 装置2还包括用于存储计算机可读程序代码的存储器单元8,所述计算机可读程序

代码可以由处理单元6运行以执行根据本发明的方法。存储器单元8还可以用于存储或缓冲在处理单元6的处理之前、期间和之后的来自成像单元4的视频序列以及处理的任何中间产物。

[0077] 将认识到,在一些实施例中,装置2可以包括具有集成或分离的成像单元4的通用计算机(例如,台式PC)或具有集成或分离的成像单元4的便携式计算设备(例如,膝上型计算机、平板电脑或智能电话)。在一些实施例中,装置2可以专用于以下目的:根据视频序列确定搏动信号,使用搏动信号识别视频序列中的活皮肤组织,和/或用于根据被识别为对应于活皮肤组织的视频序列的区中提取的rPPG信号测量对象的生理特征。

[0078] 在实际的实施方式中,装置2可以包括除了图2所示和上述的那些之外的其他部件或另外的部件,诸如,允许对象启用和/或操作装置2的用户接口,以及电源,例如电池或与主电源的连接,以用于为装置2供电。用户接口可以包括允许对象与装置2交互并控制装置2的一个或多个部件。作为范例,一个或多个用户接口部件可以包括开关、按钮或其他控制模块,以用于启动和停用装置2和/或搏动确定过程。用户接口部件还可以或者备选地包括用于向对象提供关于装置2的操作的信息的显示器或其他视觉指示器(诸如灯)。同样,用户接口部件可以包括用于向对象提供关于装置2的操作的听得见反馈的音频源。

[0079] 图3中的流程图图示了根据实施例的根据视频序列确定搏动信号的方法。

[0080] 在步骤101中,成像单元4获得视频序列。视频序列由一系列图像帧组成。图4(a)中示出了一系列图像帧20。

[0081] 接下来,根据图像帧20形成多个视频子序列(步骤103)。每个视频子序列包括来自视频序列中的图像帧20的子集中的两个或更多个图像帧20中的每个的帧分段。

[0082] 每个视频子序列优选地根据两个或更多个连续图像帧20中的每个的帧分段形成(并且因此图像帧20的每个子集优选地包括连续的图像帧20)。

[0083] 图4(a)图示了来自分别被标记F1、F2、F3、F4的视频序列的四个连续图像帧20。形成图像帧的子集21,在该范例中,每个子集包括三个连续的图像帧20。此外,在该范例中,子集21被形成使得子集21交叠(即,图像帧20中的一些或全部能够是多于一个子集21的部分)。因此,图4(a)示出了:包括图像帧F1、F2和F3的第一子集21,子集1;包括图像帧F2、F3和F4的第二子集21,子集2;以及包括图像帧F3、F4和F5的第三子集21,子集3,等等。因此,每个子集21与先前邻近子集21交叠两个图像帧20。

[0084] 将认识到,在备选实施例中,子集21可以包括不同数量的图像帧20,和/或子集21之间的交叠的程度可以是不同的。例如,子集21可以包括三个图像帧20,但是每个子集21可以仅与先前的邻近子集21交叠一个图像帧20。备选地,在子集21中的图像帧20之间可以不存在交叠(即,在这种情况下,每个图像帧20将仅是单个子集21的部分)。

[0085] 每个图像帧20被划分成多个帧分段22。每个帧分段22是图像帧20中的邻近像素的组。在图4(b)中图示了示范性分割,其中,每个帧20被划分成相等尺寸的方形。在备选实施例中,分段可以是不同的形状,例如三角形。在其他(优选)备选实施例中,分段22的形状可以由图像帧20中的图像来确定(例如,形状的边界可以跟随图像帧中的不同颜色之间的边界)。然而,在每个实施例中,将认识到,每个帧分段22包括每个图像帧20中的空间相关(即,近邻)像素的组。在优选实施例中,帧分段22在本领域中也称为“超像素”,例如,如Achanta等人的“SLIC Superpixels Compared to State-of-the-art Superpixel

Methods” (IEEE Transactions on Pattern Analysis & Machine Intelligence 2012, 第34卷、第11期、第2274-2282页 (2012年11月)) 中描述的。

[0086] 在图4 (b) 中, 每个图像帧20被划分成十八个帧分段22, 并且为了便于参考, 每个帧分段22被给予相应的标号: S1、...、S18。

[0087] 在优选的“超像素”实施例 (其中, 分段22的像素/形状的分组由图像帧的内容确定) 中, 可以通过基于像素的颜色相似性和空间相似性对图像帧20中的像素进行分组来确定帧分段22。以这种方式, 具有相似或一致的颜色近邻或紧密近邻像素将被分组在一起成为单个帧分段22。

[0088] 在上述实施例中, 图像帧20仅仅基于该图像帧20的分析被划分为帧分段22。然而, 在一些实施例中, 能够基于对该图像帧20和图像帧的子集21中的其它图像帧20的分析来将特定图像帧20划分为帧分段22。换言之, 上述图像帧20的基于空间和/或颜色的分割被扩展到时间域中, 使得时间域中的像素共享外观 (例如, 颜色) 和空间相似性被分组在一起。到时间域的该扩展可以用于导致将相同的帧分割模式应用于子集21中的每个图像帧, 或者应用于子集21中的具有相应分段模式的每个图像帧。

[0089] 如上所述, 根据来自子集中的图像帧20中的每个的帧分段形成视频子序列。这些视频子序列在本文中也称为“体素”。因此, 参考图4 (b), 根据来自子集1中的图像帧20中的每个的帧分段22形成多个视频子序列, 根据来自子集2中的图像帧20中的每个的帧分段22形成多个视频子序列, 等等。

[0090] 在图示的实施例中, 每个图像帧20中的每个帧分段22是相应的视频子序列23的部分, 尽管在其它实施例中不一定是这种情况 (即, 一些帧分段22不是任何视频子序列23的部分)。

[0091] 在图4 (c) 中图示的一些实施例中, 每个视频子序列23使用子集21中的每个图像帧20内的相同空间位置中的帧分段22形成。例如, 一个视频子序列23可以根据子集1中的图像帧20的左上角的帧分段S1形成, 另一个可以根据子集1中的图像帧20的左下角的帧分段S13形成, 等等。同样, 另外的视频子序列23使用子集2中的图像帧20中的帧分段22形成。图4 (c) 图示了根据本实施例的根据第一和第二子集21形成的子序列23。

[0092] 然而, 在优选实施例 (当通过根据颜色相似性和空间相似性对像素进行分组来形成帧分段22时, 这是特别优选的) 中, 为了改进该方法对视频序列的内容的变化 (例如, 由于对象移动) 的鲁棒性, 可以通过从彼此一致的子集21中的图像帧20中选择帧分段22来形成视频子序列23 (例如, 在图像帧20内的空间位置基本一致, 并且颜色大致一致)。这可以导致视频子序列“卷绕”通过图像帧20, 使得视频子序列23包含针对对象的特定部分的帧分段22 (例如, 当对象在视频序列中从左向右移动时, 特定的视频子序列23可以由每个图像帧20中由于空间和颜色相似性而对应于对象的脸颊的帧分段22形成)。对于图像帧20中的特定帧分段22, 形成视频子序列23的一种优选方式是从特定的帧分段22中识别下一图像帧20中的具有最小色能 (即, 最小色度差) 和空间距离能量 (即, 最小空间距离) 的帧分段22。将理解, 色能是指基于帧分段22和下一个图像帧20中的帧分段22中的像素的色度值的能量函数, 并且因此使形成视频子序列23的色能最小化可以包括, 对于特定帧分段22, 识别下一图像帧20中针对考虑的帧分段22具有最小色能的帧分段22。将认识到, 与所考虑的帧分段22中的像素的色度相比, 帧分段22中的像素的色度越不同, 色能越高, 并且因此帧分段22被选择用

于该视频子序列23的可能性越低。还将理解,空间距离能量是指基于图像帧20中的帧分段22的空间位置和下一图像帧20中的帧分段22的空间位置的能量函数,并且因此使形成视频子序列23的空间距离能量最小化可以包括,对于特定帧分段22,识别下一图像帧20中的提供所考虑的帧分段22中的最小空间距离的帧分段22。应当理解,从下一图像帧20中的帧分段22的位置到所考虑的帧分段22的位置的距离越大,空间距离能量越高,并且帧分段22将被选择用于该视频子序列23的可能性越低。在一些实施例中,可以在将RGB像素值投影到子空间(例如UV)之后考虑空间距离能量。在备选方法中,还能够使用在时刻 $t-1$ 处的体素23的中心在时刻 t 初始化新体素。在体素23在时间上交叠(即,交叠图像帧)的情况下,体素23的最后帧分段22中的中心确定下一体素23的第一帧分段22中的中心。

[0093] 将图像帧20划分成上述多个帧分段22以第一尺度或分辨率提供视频子序列,其中,尺度或分辨率由每个图像帧中的帧分段22的数量指示(例如图4中的十八个),或者由每个帧分段22的尺寸来指示(例如,依据每帧分段22的图像像素的数量)。在一些实施例中,图像帧20还可以被划分成具有与第一多个不同的尺寸/分辨率的第二多个帧分段22,另外的多个视频子序列23根据这些分段22形成。多个分辨率的该使用的优点在于其能实现视频序列中的对象的尺度不变检测,并且其改进视频序列中的多个对象的检测。

[0094] 一旦形成视频子序列23,该方法进行到识别可以级联或链接在一起以允许确定搏动信号的视频子序列23。该方法利用以下事实:特定对象的活皮肤组织中的搏动信号在皮肤的不同位置处或多或少相同(假定考虑归一化颜色值),这意味着来自视频序列的(略微)不同的空间部分的视频子序列23可以被级联在一起。

[0095] 因此,在步骤105中,将根据来自图像帧20的第一子集21的帧分段22形成的第一视频子序列23与根据来自图像帧20的第二子集21的帧分段形成的视频子序列23进行比较。第二子集21优选地是与第一子集21邻近的子集。

[0096] 该比较优选地包括将第一视频子序列23的至少一个代表值与第二子集21中的视频子序列23的相应代表值进行比较。视频子序列的代表值基于或被导出自构成视频子序列23的帧分段22的内容(例如,像素值)。在一些实施例中,每个视频子序列23可以具有多于一个代表值。在这种情况下,步骤105可以包括将特定视频子序列23的代表值中的多个代表值与其他视频子序列23的代表值中的多个代表值进行比较。

[0097] 图4(d)图示了步骤105的范例。因此,根据第一子集21(包括图像帧F1、F2和F3的子集1)中的图像帧20中的帧分段22形成的第一视频子序列24,并且尤其是根据图像帧F1、F2和F3中的分段S9形成的子序列24与根据在第二子集21(包括图像帧F2、F3和F4的子集2)中的图像帧20中的帧分段22形成的视频子序列25进行比较。在一些实施例中,可以将第一视频子序列24与根据图像帧20的第二子集21形成的所有视频子序列25进行比较。在备选实施例中,可以将第一视频子序列24与根据图像帧20的第二子集21形成的一些视频子序列25进行比较。例如,第一视频子序列24可以仅与在图像帧20中的相同和/或相似空间位置的视频子序列25进行比较。这在图4(d)中由仅与帧F2、F3和F4中的标记的视频子序列25(即包括分段S2、S3、S4、S8、S9、S10、S14、S15和S16的视频子序列25)比较的第一视频子序列24图示。

[0098] 步骤105中的比较旨在识别第一视频子序列24可以与其级联的视频子序列25,并且这样一来比较旨在识别具有与第一视频子序列24的代表值相似或最相似的代表值的视频子序列25。在一些实施例中,视频子序列25还应当图像帧20中与第一视频子序列24处

于相似的空间位置中(例如,如图4(d)图示的)。

[0099] 将认识到,当比较代表值以决定要级联哪些子序列时,在子序列的时间上存在交叠的实施例(例如,如图4中的子集1和子集2)提供了益处,即能够比较相同帧(在平均的情况下)中的每个子序列的代表值或根据相同帧对导出(在(归一化)差异的情况下)的每个子序列的代表值。

[0100] 在比较步骤之前,可以确定针对每个视频子序列23的代表值。在图4中,根据视频序列中的图像帧Y到Z中的帧分段SX形成的特定视频子序列23的代表值被指代为 $V_{SX:Y \rightarrow Z}$ 。用于图4(c)中的每个视频子序列23的代表值V的注释被示出在视频子序列本身下面。

[0101] 在一些实施例中,可以根据视频子序列23中的所有帧分段22中的像素的像素值(例如RGB值)的平均来确定每个视频子序列23的代表值。在备选实施例(其中,每个视频子序列23可以具有多于一个代表值)中,每个值可以是视频子序列23中的相应帧分段22中的像素的像素值的平均。

[0102] 在一些实施例中,当确定平均时,可以基于帧分段22中的像素的空间位置和/或与帧分段22的中心处或附近的像素或像素的组的颜色的差异对像素值进行加权,并且确定加权值的平均。例如,可以基于从像素到空间边界的距离、和/或从像素到帧分段22的中心的距离来对像素值进行加权。优选地,加权导致靠近分段22边界的像素被较少加权,因为它们比靠近分段22的中间的像素更不可靠,这是由于近邻分段22之间的抖动伪迹。

[0103] 在一些实施例中,上面确定的像素值的平均被归一化,并且归一化平均被用作代表值。平均可以通过将平均除以视频子序列23中的所有帧分段22中的所有像素的像素值的平均(或如果需要为加权平均)来归一化。备选地,平均可以通过将平均除以形成视频子序列23的子集21中的所有图像帧20中的所有像素的像素值的平均(或如果需要为加权平均)来归一化。

[0104] 在一些(或另外的)实施例中,可以根据视频子序列23中的每个帧分段的像素的像素值的平均(或加权平均)之间的差异来确定每个视频子序列23的代表值。在一些实施例中,可以使用子序列23中的所有帧分段22上的像素值的平均(或加权平均)来对差异进行归一化。备选地,可以使用在子序列23中的所有帧分段22上的像素值的平均(或加权平均)的对差异进行归一化。归一化差异的使用是有利的,因为其允许差异值(即视频子序列)被级联,并且甚至当表示不同视频子序列的平均值显著不同时计算搏动信号。当归一化差异被级联时,级联的差异值可以在导出搏动信号之前或之后被积分(下面的步骤109)。

[0105] 返回到图3,在步骤107中,基于第一视频子序列24和来自第二子集21的视频子序列25之间的比较,第一视频子序列24与来自第二子集21的与第一视频子序列24充分相似(例如,代表值在彼此的特定量之内)或最相似(该视频子序列25在本文中被称为第二视频子序列25)的视频子序列25进行级联。

[0106] 这种级联如图4(e)所示,其中,第一视频子序列24(根据F1、F2、F3中的分段S9形成)与第二视频子序列25(根据F2、F3、F4中的分段S10形成)进行级联。

[0107] 一旦第一视频子序列24已经与根据来自图像帧20的第二子集21的帧分段形成的第二视频子序列25级联,可以针对第二视频子序列25重复步骤105和107以识别根据来自图像帧20的第三子集21的帧分段(即,F3、F4、F5)形成的第三视频子序列26。因此,第三视频子序列26可以与第二视频子序列25级联,如图4(e)所示。可以重复步骤105和107,以从要与子

序列24、25、26级联的图像帧20的另外的子集21识别视频子序列。

[0108] 一旦视频子序列23已经被级联或链接在一起,则根据视频子序列确定搏动信号(步骤109)。搏动信号表示级联序列中的帧分段22和/或视频子序列23的颜色或颜色的变化。

[0109] 在一些实施例中,步骤109可以包括根据级联的视频子序列24、25、26中的每个的代表值来确定搏动信号。因此,在一些实施例中,搏动信号可以根据代表值本身形成,例如,如图4(f)所示。因此,代表值 $V_{S9:1-3}$, $V_{S10:2-4}$ 和 $V_{S...:3-5}$ 分别在时间 t_{F1} 、 t_{F2} 、 t_{F3} (与每个视频子序列24、25、26中表示的第一图像帧的定时相对应)提供搏动信号27的值。

[0110] 在该范例中,形成搏动信号27的视频子序列24、25、26包含活皮肤组织的区,因此根据该级联的视频子序列集合确定的搏动信号27将呈现PPG信号的典型特性(即幅度变化与由于心脏跳动引起的对象皮肤中的血液灌注的变化一致)。如果级联的视频子序列集合不包含活皮肤组织的区,则根据该集合确定的搏动信号将不会呈现PPG信号的典型的特性(并且,在视频序列中的环境照明没有变化的情况下,该集合的搏动信号通常可以对应于噪声信号)。

[0111] 将认识到,在根据级联视频子序列23的代表值形成搏动信号27的实施例中,步骤107和109可以被有效地视为一个步骤,因为第一视频子序列24的该或一个代表值与第二视频子序列25的代表值级联。

[0112] 在备选实施例中,可以使用用于从图像/图像帧/帧分段序列中提取搏动信号的技术来分析级联的视频子序列24、25。用于根据视频序列确定搏动信号的各种技术在本领域是已知的并且在此将不必进行详细描述。然而,在下面给出的体素搏动光谱方法的描述中提到了一些示范性技术。

[0113] 在一些实施例中,在形成级联的视频子序列集合并确定搏动信号27之后,该方法可以返回到步骤105,并且对于根据第一子集,子集1中的图像帧20形成的视频子序列23中的不同的一个重复。以这种方式,将为视频序列的不同(空间)部分确定多个搏动信号27,其中一些搏动信号可以包括活皮肤组织的区,而其他可能不包括活皮肤组织的区。将认识到,由于来自不同子集21的视频子序列23彼此进行比较(即,通过代表值的比较)的方式,特定视频子序列23能够有助于若干不同的搏动信号(或者实际上特定视频子序列23可能不对任何搏动信号27有贡献)。

[0114] 一旦已经确定或提取了搏动信号27(或多个搏动信号),则可以分析搏动信号27以识别视频序列中的活皮肤组织的区。在视频子序列23根据具有多个分辨率的帧分段形成的情况下,分析可以包括一起分析在不同分辨率下获得的搏动信号,以识别活皮肤组织的区。在一些情况下,搏动信号27可以基于相似性(例如空间、时间、颜色和/或频率相似性)聚类在一起,并且从这些聚类识别活皮肤组织的区。

[0115] 在一些实施例中,可以分析一个或多个搏动信号27以确定哪个(如果有的话)呈现活皮肤组织的特性。在一些实施例中,确定(一个或多个)搏动信号的频率特性,并且将(一个或多个)搏动信号的确定的频率特性与从活皮肤组织的区获得的搏动信号的典型频率特性进行比较。例如,固定频率阈值或频带(例如对应于典型心跳/搏动频率)可以用于确定周期性搏动信号是否对应于活皮肤组织。

[0116] 在备选实施例中,基于搏动信号的相似性,搏动信号27被空间聚类。用于实施这一

点的合适的聚类算法是基于密度的噪声应用空间聚类 (DBSCAN), 或者在上文引用的“Face detection method based on photoplethysmography”中提出的聚类。

[0117] 然而, 在一些实施例中, 可以确定每个搏动信号27与其他搏动信号27的成对相似性。亦即, 对于每个搏动信号, 确定与其它搏动信号中的每个的相似性的量度。这些成对相似性量度然后被分析以识别活皮肤组织的区。

[0118] 成对相似性量度优选地包括或是基于频率的成对相似性量度。这是有利的, 因为对应于同一特定对象的活皮肤组织的区的不同帧分段22应当呈现在相关中具有相似(或相同) 频率峰值指数、相位或低熵的搏动信号。

[0119] 搏动信号和其他搏动信号之一的成对相似性的量度可以是搏动信号与其它搏动信号之一的频谱的至少部分(其在本文中被称为频谱峰) 之间的相关性的量度、搏动信号与其它搏动信号之一的频谱的至少部分(其在本文中称为频谱相位) 之间的归一化互相关的量度、两个搏动信号的频谱的至少部分(这里称为频谱熵) 之间的相关性的正则性的量度, 和/或两个搏动信号的内积的结果(其可以任选地在计算内积之前被滤波)。这些成对相似性量度的进一步细节可以在下面的体素搏动光谱(VPS) 方法的描述中发现。

[0120] 本领域技术人员将认识到可以确定并使用上面呈现的成对相似性的备选或另外的量度来识别活皮肤组织的区。

[0121] 在一些实施例中, 可以针对每个搏动信号与其他搏动信号中的每个来确定成对相似性(例如, 频谱峰值和频谱相位) 的多个量度, 并且这些量度被组合以形成表示每个搏动信号与其他搏动信号中的每个的成对相似性的距离度量。

[0122] 在一些实施例中, 可以确定视频序列的相似性矩阵, 并且可以根据相似性矩阵识别活皮肤组织的区。可以通过组合为搏动信号确定的成对相似性(或距离度量) 来形成相似性矩阵。相似性矩阵是跨两个或更多个视频子序列的相似搏动信号相互相关的矩阵。相似性矩阵的使用是有利的, 因为其不需要预定义任何参数(例如基于肤色或聚类的参数)。

[0123] 一旦确定了相似性矩阵, 通过执行相似性矩阵的矩阵分解来识别活皮肤组织的区。在一些实施例中, 矩阵分解可以通过奇异值分解(SVD)、QR分解、稀疏SVD、增量SVD、主成分分析PCA或独立成分分析ICA。这些技术在本领域中是公知的, 并且将不在此详细描述。在下面更详细描述的优选实施例中, 相似性矩阵使用增量稀疏PCA进行分解。

[0124] 分解可以包括将相似性矩阵因子化(分解) 成正交基, 以发现属于相同对象的视频序列的部分。该因子化导致不同种类的相似性被分离到独立方向中。这导致属于相同的对象的帧分段在相同方向上聚类。

[0125] 一旦已经在视频序列中识别出活皮肤组织的区, 可以根据视频序列确定对象(具有识别的活皮肤组织的对象) 的一个或多个生理特性。

[0126] 在一些实施例中, (一个或多个) 生理特性可以根据与识别出的活皮肤组织的区相关联的一个或多个搏动信号27确定。在这种情况下, 可以个体地分析一个或多个搏动信号27以确定生理特征并且生理特性被组合(例如被平均) 以给出对象的生理特征的总体量度。备选地, 可以将一个或多个搏动信号27组合(例如平均) 以给出单个搏动信号, 并且分析搏动信号以确定生理特性。

[0127] 在其他实施例中, 可以重新处理视频序列以从被识别为活皮肤组织的区提取一个或多个搏动信号, 并且处理(一个或多个) 该搏动信号以确定生理特性。

[0128] 根据视频序列导出的(一个或多个)搏动信号相似于使用PPG传感器获得的信号,因此对象的一个或多个生理特性可以包括可以根据PPG信号或对象的血液灌注(或血液灌注的变化)的其他量度导出的任何特性,例如心率、心率变异性、搏动间隔、呼吸率、呼吸信号、SpO2值(即血液的动脉血氧水平)等等。因此用于根据PPG信号导出这种生理特性(例如用于确定心率的频域中的峰值检测)的本领域已知的技术可以用于根据从被识别为活皮肤组织的区获得的搏动信号导出生理特性的值。

[0129] 本文呈现的技术的特定实施例在下文中被描述,并且在本文中被称为体素搏动光谱(VPS)方法。

[0130] 体素搏动光谱(VPS)方法

[0131] 基于相机的搏动提取

[0132] 在人类心血管系统中,传播穿过全身的血液搏动会改变皮肤组织中的血液体积。由于血液中的血红蛋白的光学吸收跨光谱范围变化,因此检测皮肤反射的颜色变化可以显示搏动率。最近的远程光体积描记(rPPG)技术使得能够使用常规RGB相机检测人皮肤上的搏动诱导的颜色变化。已经提出了盲源分离方法(例如基于PCA和基于ICA)将时间RGB信号进行因子化以发现搏动。还已经提出了基于色度的rPPG方法,以在标准化肤色假设下,将搏动定义为RGB通道的线性组合,这是处理真实挑战(例如不同对象皮肤)的最准确的rPPG方法之一。

[0133] 基于搏动的感兴趣区域检测

[0134] 鉴于人类搏动可以通过视频序列中的rPPG进行测量的事实,因此可以使用搏动信号来辅助对象检测,即通过定位其活皮肤组织来检测活的对象。现有技术提出了一种基于将视频切割成固定的刚体网格以用于局部搏动提取的搏动信号的面部检测方法。其设置了一个硬阈值以发现具有高频谱能量的网格,并将其标记为面部区域。其限于静态面部需要被放置在距相机预定距离处的视频。本文描述的VPS方法不受这些限制的影响。在另一种现有技术中,提出了一种用于增强rPPG监测的面部的最佳感兴趣区域(RoI)选择方法。然而,RoI被约束到预定义的面部界标,这不是对于对象检测的一般解决方案,即其不能检测在视频中可能可见的其他身体部分(例如,手)。相反,本文描述的VPS方法不能做出这样的假设,并且可以利用脉动的血液体积检测所有身体部分。

[0135] VPS方法的概述如图5所示,其采用包括图像帧20的输入视频序列并输出对象RoI 52(即,对应于活皮肤组织的视频序列的区)。与图3所示的方法一致,将视频序列分割为多个子序列(分级体素阶段54-步骤103),根据子序列确定搏动信号(搏动提取阶段56-步骤105-109),相似性矩阵被确定和分析以识别RoI(光谱分析阶段58)。下面将详细讨论每个阶段。

[0136] 分级体素

[0137] 给定没有关于对象(或关于一般内容)的任何先前信息的视频序列,首先将视频序列分割成密集的局部区域,其中,可以独立地测量搏动(这是图5中的分级体素阶段54)。尽管视频序列可以被切割成固定的刚体网格,但这意味着对象尺寸由网格几何结构量化,当对象小或存在身体运动时,这种方法会发生问题或失败。因此,在VPS方法中优选使用用于搏动提取的优越视频分割方法,其被称为“分级体素”。分级体素包括多个尺度(其中,尺度决定每个图像帧20被划分成的聚类/分段的数量)的时空相干聚类(帧分段),其中,在时间

域中共享外观和空间相似性的图像帧20中的像素被分组在一起。尽管多个尺度是优选的并且用在VPS方法的以下描述中,但是能够在VPS方法中使用单个尺度。

[0138] 从一个尺度开始,构造体素(包括来自连续图像帧的子集(例如2、3、4等)的帧分段的视频子序列)被定义为最小化在短间隔 $T \in \{2n+1, n \in \mathbb{N}^+\}$ 内的时间上邻近的超像素/帧分段之间的色能 E_c (即最小化帧分段的色度的差异) 和空间距离能量 E_s (即最小化帧分段之间的空间距离的差异) 为以下项的流程:

$$[0139] \quad \arg \min \left(\sum_{t=\frac{T-1}{2}}^{t+\frac{T-1}{2}} \sum_{p \in P(t)} (1-\lambda) E_c^t(p, k) + \lambda E_s^t(p, k) \right) \quad (1)$$

[0140] 其中, $p \in P(t)$ 是第 t 帧中的像素的集合。 p 的表示是4维特征向量 (x, y, u, v) , 其中, (x, y) 和 (u, v) 分别是图像平面和色度平面(例如YUV空间的UV平面, YUV空间是皮肤分割的经验空间)中的坐标。 k 均值聚类被执行以将像素分配给 k 个聚类, 以便在 T 期间最小化总能量。 λ 是控制两个能量项之间的平衡的参数。

[0141] 此外, 通过在等式1中同时初始化不同的 k , 将单尺度体素扩展到多个尺度, 其中, 每个尺度是独立的聚类。考虑到单独的尺度中的体素具有不同的分辨率和能量变化, 第 i 个尺度的 λ_i 基于 t 处的其自身的能量变化而自适应自调谐为:

$$[0142] \quad \lambda_i^t = \log(k) \sqrt{\frac{\sigma(\phi(u_i^t)) \cdot \sigma(\phi(v_i^t))}{\sigma(\phi(x_i^t)) \cdot \sigma(\phi(y_i^t))}} \quad (2)$$

[0143] 其中, $\sigma(\cdot)$ 指代标准差运算符; $\phi(\cdot)$ 表示聚类均值的集合; $\log(k)$ 控制体素紧凑度, 即具有较高分辨率(较大 k) 的体素应更加紧凑。不同尺度的 λ 的实时调谐避免了易失性和闪烁聚类, 从而保留细粒度分割。

[0144] 如上所述, 使用分级体素有四个益处, 如图6图示的, 其示出了包括活对象和人工面部的两种不同图像的三种不同分辨率/尺度 ($k=16, k=36; k=64$)。在图6中, 第一图像60对应于图1中的左手图像, 并且示出了面对相机的人面部和人工面部。第二图像62对应于图1中的右手图像, 并且示出了人面部和人工面部侧面呈现给相机。首先, 使用分级体素建立了时空相干“管”(视频子序列)以用于搏动测量。其次, 其使得能够在视频中进行尺度不变的对象检测。第三, 其保持对象形状的高边界回忆, 如图像中的形状之后的帧分段的边界所指示的。第四, 其创建皮肤区域的统计学观察, 因为从具有不同分辨率的体素测量的搏动具有不同的量化质量。

[0145] 搏动提取

[0146] 本节描述图5中的搏动提取阶段56。在层次结构中的每个体素(即, 视频子序列)被假设为并行搏动提取中的独立搏动传感器。在优选实施例中, 基于色度的方法(CHROM)用于搏动测量, 其描述于G.de Haan和V.Jeanne的“Robust pulse rate from chrominance-based rPPG”(TBME, 60(1):2878-2886, 2013)中。然而, 本领域技术人员将会认识到可以用于确定针对体素的搏动信号的其他技术。例如, 如上所述, 基于PCA或ICA的技术可用于从体素提取搏动信号, 或可以将如在G.de Haan和A.van Leest的“Improved motion robustness of remote-PPG by using the blood volume pulse signature”(Physiol.Meas.35 1913, 2014)中描述的“PBV”方法用于从RGB轨迹中提取搏动。与使用所有像素的空间平均的RGB来导出局部区域/聚类中的搏动信号的CHROM不同, 通过基于它们

到体素边界的距离来对它们加权来组合体素中的像素RGB值,即,接近于体素边界的像素由于近邻体素之间的偶然抖动伪迹而不太可靠,并且因此应该被加权较少。假设从像素k到体素边界的最近的欧几里德距离为 d_k ,在t处第i个尺度中第j个体素的平均RGB被组合为:

$$[0147] \quad (\bar{R}_{ij}^t, \bar{G}_{ij}^t, \bar{B}_{ij}^t) = \frac{\sum_{k=0}^N (d_k \cdot (\bar{R}_{ijk}^t, \bar{G}_{ijk}^t, \bar{B}_{ijk}^t))}{\sum_{k=0}^N d_k} \quad (3)$$

[0148] 其中,N指代第j个体素中的像素的数量。在恒定的照明环境中,人类皮肤组织示出相同的相对PPG幅度,但体素的色差导致搏动幅度的变化。因此与CHROM不同,使用体素中的平均RGB的时间导数,即 $dC_{ij}^t = C_{ij}^t - C_{ij}^{t-1}$, $C \in \{\bar{R}, \bar{G}, \bar{B}\}$,来导出其色度信号。在间隔T(在等式1中定义)中,归一化色度导数被计算为:

$$[0149] \quad \left\{ \begin{array}{l} \overrightarrow{dX}_{ij}^T = 3 \frac{\overrightarrow{dR}_{ij}^T}{\sum_{t=0}^T dR_{ij}^t} - 2 \frac{\overrightarrow{dG}_{ij}^T}{\sum_{t=0}^T dG_{ij}^t} \\ \overrightarrow{dY}_{ij}^T = 1.5 \frac{\overrightarrow{dR}_{ij}^T}{\sum_{t=0}^T dR_{ij}^t} + \frac{\overrightarrow{dG}_{ij}^T}{\sum_{t=0}^T dG_{ij}^t} - 1.5 \frac{\overrightarrow{dB}_{ij}^T}{\sum_{t=0}^T dB_{ij}^t} \end{array} \right\} \quad (4)$$

[0150] 其中, $(dR_{ij}^t, dG_{ij}^t, dB_{ij}^t)$ 指代两个图像帧之间的体素中的RGB的时间导数。在每个间隔中估计的色度导数被线性组合成搏动导数并进一步积分。然后,将不同的搏动间隔交叠添加到具有长度L的完整搏动信号 $\vec{S}_{ij}^L$ 上。此过程解释为:

$$[0151] \quad \vec{S}_{ij}^L = \sum_{t=0}^{L-T+1} \vec{S}_{ij}^{t+T} + w \cdot \text{csum}(\overrightarrow{dX}_{ij}^T - \frac{\sigma(\overrightarrow{dX}_{ij}^T)}{\sigma(\overrightarrow{dY}_{ij}^T)} \overrightarrow{dY}_{ij}^T) \quad (5)$$

[0152] 其中, $\text{csum}(\cdot)$ 指代时间导数信号的累积和;w是用于平滑交叠加法的Hanning窗口。因此,并行提取的搏动信号 $\vec{S}_{ij}^L$ (从第i个尺度中的第j个体素)集中并归一化为:

$$[0153] \quad \vec{S}_{ij}^L = \frac{\vec{S}_{ij}^L - \mu(\vec{S}_{ij}^L)}{\sigma(\vec{S}_{ij}^L)} \quad (6)$$

[0154] 其中, $\mu(\cdot)$ 指代平均运算。请注意,搏动信号是此方法中使用的唯一特征。不使用其他外观特征(如颜色或纹理)。

[0155] 频谱分析

[0156] 本节描述了频谱分析阶段58,其包括三个子阶段,形成相似性矩阵,在相似性矩阵上执行增量稀疏PCA,并且使用分级融合来识别视频序列中的活皮肤组织的区。

[0157] 已经注意到,从属于相同对象的皮肤区域提取的搏动信号在诸如相位和频率的许多方面共享相似性,而从非皮肤区域(例如背景)提取的搏动信号是无相关性的随机噪声。因此,在从分级体素提取搏动信号之后,搏动信号的成对相似性被用于发现活着的对象。这也适用于视频序列中多个对象的情况,因为从不同对象测量的搏动也可以在相位和频率上进行区分。

[0158] 相似性矩阵-在该步骤中,创建相似性矩阵 $\Sigma = (D, C)$,以基于测量的搏动互连分级体素。在 Σ 中,对角线轨迹中的条目D包含不同尺度的所有体素;剩下的条目C指代任何一对体素之间的成对连接。为了建立这样的相似性矩阵,需要定义用于测量搏动相似性的距离度量。最常用的距离度量,即L1和L2距离,不适用于搏动特征。然而,与其他外观特征(例

如, Haar和HOG) 相比, 搏动特征中的基本和独特的特性是其包含周期性。已经注意到, 来自相同对象的搏动信号示出以下关系: (1) 它们具有相似的频率, 并且因此它们的互相关呈现出显著的频谱峰值; (2) 它们没有显著的相移; (3) 它们的频率相关性是规则的和较少紊乱的; 并且 (4) 如果将搏动信号考虑为多维向量, 则两个相似向量之间的夹角小。因此, 用于建立搏动信号的相似性矩阵的优选距离度量强调上述连接, 并且包括四个不同的测量结果:

[0159] 频谱峰值-在频域中, 定义了用于体素通信的搏动率频带 $f \in [40, 240]$ BPM (每分钟搏动), 这对于包括新生儿和运动对象在内的健康对象是宽的范围。两个互相关的搏动信号的频谱峰值定义为:

$$[0160] \quad F = \arg \max_{f \in [40, 240]} ((\mathcal{F}(\vec{S}_{ij}^L) \circ \mathcal{F}(\vec{S}_{i'j'}^L))^*) \quad (7)$$

[0161] 其中, $\circ$ 指代逐元素积; $*$ 是共轭; $\mathcal{F}(\cdot)$ 表示快速傅里叶变换 (FFT)。

[0162] 频谱相位-两个相似的搏动信号也处于相同的相位中, 因此它们的归一化互相关应该在时间域中示出强响应为:

$$[0163] \quad p = \max(\mathcal{F}^{-1}(\text{NCC})) \quad (8)$$

[0164] 其中,

$$[0165] \quad \text{NCC} = \frac{(\mathcal{F}(\vec{S}_{ij}^L) \circ \mathcal{F}(\vec{S}_{i'j'}^L))^*}{\|(\mathcal{F}(\vec{S}_{ij}^L) \circ \mathcal{F}(\vec{S}_{i'j'}^L))^*\|_2} \quad (9)$$

[0166] 其中, $\|\cdot\|_2$ 是 L2 范数; $\mathcal{F}^{-1}(\cdot)$ 指代逆 FFT。

[0167] 频谱熵-术语“熵”用于测量两个搏动信号之间的相关性的正则性:

$$[0168] \quad E = \frac{\sum_{f=40}^{240} \text{NCC}(f) \log(\text{NCC}(f))}{\log(240-40)} \quad (10)$$

[0169] 其中, E 的解释与其他测量结果一致, 即较大的 E 指代更好的相关性。

[0170] 内积-在时间域中, 我们使用内积来测量两个搏动信号之间的余弦角为:

$$[0171] \quad I = \left\langle \frac{\vec{S}_{ij}^L}{\|\vec{S}_{ij}^L\|_2}, \frac{\vec{S}_{i'j'}^L}{\|\vec{S}_{i'j'}^L\|_2} \right\rangle \quad (11)$$

[0172] 其中 $\langle, \rangle$ 指代内积运算。

[0173] 最后, 将这四个测量值归一化到范围 [0; 1] 并与高斯内核融合在一起为:

$$[0174] \quad \Sigma = 1 - \exp\left(-\frac{(F \circ P \circ E \circ I)^2}{2\sigma_{I,F,P,E}^2}\right) \quad (12)$$

[0175] 其中, $\sigma_{I,F,P,E}$ 表示四个矩阵之间的逐条目标准差。应当注意, 四个测量结果彼此不完全独立, 测量结果之间的冗余性有助于减少相似性估计中的不确定性。图 7 示出了四个测量结果的范例以及针对两个视频序列的它们的融合相似性矩阵 $\Sigma_{64,66}$, 一个包括单个活的对象, 并且另一个包括两个活的对象。具有较高能量的条目表示层次中相似体素的索引。

[0176] 在本文使用的距离度量中, 两个良好对准的搏动信号在互相关期间示出升高的频率能量, 其可以有效地抑制噪声条目 (例如, 没有搏动的体素)。相比之下, 先前的距离度量是不能增强比较中的相似条目之间的连接的所有客观测量结果。最后, 层次结构中的所有

体素在相似性矩阵中相互连接。可以将在体素中检测活的任务重新公式化为发现相似性矩阵的子空间分区,使得在相同的子空间中的条目具有相同的相似性方向。

[0177] 增量稀疏矩阵分解-相似性矩阵 Σ 64、66可以解释为 $\lambda_1 x_1 x_1^T + \lambda_2 x_2 x_2^T + \dots \lambda_n x_n x_n^T$ 的线性组合,其中, $x_i \in X$ 是多维空间中的一组正交向量。为了找到属于同一对象的体素,将矩阵分解技术用于将 Σ 因子化为 X ,其中,不同的对象被分离成不同的特征向量。由于 Σ 是具有许多零条目的稀疏矩阵(例如,指向背景的体素没有相似性),通过寻找表达力与数据可解释性之间的折衷而应用稀疏PCA来将 Σ 分解为 X 。

[0178] 稀疏PCA在Y.Zhang,A.d'Aspremont和L.Ghaoui的“Sparse PCA:Convex relaxations,algorithms and applications”(International Series in Operations Research&Management Science Volume 166:915-940,2012)中有描述。稀疏PCA通过优化以下非凸目标函数来找到 Σ 中具有最大方差的第一个稀疏特征向量:

[0179] $\arg\max_X (X^T \Sigma X)$ 经受 $\|X\|_2=1, \|X\|_1 \leq n$ (13)

[0180] 其中, $\|\cdot\|_2$ 是L1范数; $n>0$ 控制 X 的基数。然而,计算具有最大方差的稀疏特征向量是一个组合问题,数值难以求解,因此在利用 l_1 惩罚的半定松弛提升流程后,等式13中的非凸秩约束下降:

[0181] $\arg\max_{\hat{\Sigma}} \text{Tr}(\Sigma \hat{\Sigma}) - \rho \|\hat{\Sigma}\|_1$ 经受 $\text{Tr}(\hat{\Sigma}) = 1, \hat{\Sigma} \succeq 0$ (14)

[0182] 其中, $\text{Tr}(\cdot)$ 指代矩阵轨迹操作; $\rho>0$ 控制稀疏度; $\hat{\Sigma} = XX^T$ 是由第一个主要特征向量近似的对称矩阵。在这一点上,可以使用名为混合条件梯度平滑 (HCGS) 的算法来求解等式14。HCGS在A.Argyriou,M.Signoretto和J.Suykens的“Hybrid conditional gradient-smoothing algorithms with applications to sparse and low rank regularization”(ArXiv preprint:1404.3591,2014)中有描述。HCGS的优点是使用条件梯度方法的凸松弛中的快速收敛。

[0183] 然而在实践中, Σ 可以在多个对象的情况下由多个稀疏特征基组成,而等式14仅促进第一个主要特征向量的稀疏性。为了解决这个问题,通过使用Hotelling的缩小利用先前的稀疏特征向量 $x_1, x_2, \dots, x_{i-1}$ 顺序缩小 Σ 来估计后续的稀疏特征向量 x_i 为:

[0184] $\Sigma_i = \Sigma_{i-1} - (x_i^T \Sigma_{i-1} x_i) x_i x_i^T, i \in [1, m]$ (15)

[0185] 其中,

[0186] $m = \arg\max_i \left(\frac{x_{i-1}^T \Sigma_{i-1} x_{i-1}}{1 + x_i^T \Sigma_i x_i} \right)$ (16)

[0187] 其中,可以通过HCGS中的幂迭代导出 $x_i | \in X$; m 是自动发现的最具表现力的特征向量的数量,其也意味着视频序列中的对象的数量,即, m 通常在最大特征值间隙处发现。

[0188] 图8示出了使用图7所示的两个相似性矩阵64、66的增量稀疏PCA的相似性矩阵分解的范例,其中,相似的体素在所选择的特征向量中被因子化到相同的方向上。因此,图8示出了从相似性矩阵因子化和选择的特征基;原始 Σ 64、66中的有噪声的条目在 $\hat{\Sigma}$ 68、70中消除;特征值(分别如曲线图72和74所示)分别清楚地示出了 $\hat{\Sigma}$ 中最有代表性的特征向量的数量(分别如曲线图76和78所示)。

[0189] 事实上,一些固有的(例如搏动率变化)和非本征(例如亮度变化)因素可能偶尔改

变后续帧中的相似性矩阵,这导致从每个单帧估计的稀疏特征向量的不稳定性。为了解决这个问题,使用增量子空间更新来使 $x_i \in X$ 平滑地适应于时间域中的实时变化。基本上,将时变相似性矩阵 $\hat{\Sigma}_{\text{新}}$ 考虑为新的观测结果,并使用来自的不同帧的多个观测结果 $[\hat{\Sigma}_{\text{旧}}, \hat{\Sigma}_{\text{新}}]$ 将子空间模型丰富为:

$$[0190] \quad [U, D, V] = \text{SVD}([\hat{\Sigma}_{\text{旧}}, \hat{\Sigma}_{\text{新}}]) \quad (17)$$

[0191] 其中, $\text{SVD}(\cdot)$ 指代奇异值分解; U 和 D 是递增更新的特征向量和特征值。在下面的算法1中示出了用于从时变相似性矩阵中递增地估计多个稀疏特征向量的示范性算法:

[0192] 输入: 相似性矩阵 $\Sigma \in \mathbb{R}^{n \times n}$ 、特征向量 U 、特征值 D

[0193] 1: $\rho = 0.25$ (稀疏性), $N = 100$ (迭代次数)

[0194] 2: 针对 $k = 1, 2 \dots N$, 进行

$$[0195] \quad 3: \beta_k = \frac{1}{n} \text{Tr}(\Sigma \Sigma_k) + \frac{k}{n} \|\Sigma_k\|_1$$

$$[0196] \quad 4: Z_k = \Sigma - \frac{\sqrt{k}\rho}{2\sqrt{2}} + \frac{\sqrt{k}\rho}{2\sqrt{2}} \text{sign}(\Sigma_k) \odot \left(\left| \Sigma_k \right| - \frac{2\sqrt{2}}{n\sqrt{k}} \right)$$

[0197] 5: $X = \{x_1, x_2 \dots x_m\} \leftarrow Z_k$ 的多个特征向量, 其中, m 是由等式15和等式16确定的

$$[0198] \quad 6: \hat{\Sigma} = XX^T$$

$$[0199] \quad 7: \Sigma_{k+1} = \left(1 - \frac{2}{k+1}\right) \Sigma_k + \frac{2}{k+1} \hat{\Sigma}$$

[0200] 8: 如果 $\frac{|\beta_k - \beta_{k-1}|}{\beta_k} < 10^{-3}$, 则

[0201] 9: 终止

[0202] 10: 结束条件

[0203] 11: 循环结束

[0204] 12: 如果 $U, D = 0$, 则

$$[0205] \quad 13: [U, D, V] = \text{SVD}(\hat{\Sigma})$$

[0206] 14: 否则

$$[0207] \quad 15: [U, \hat{\Sigma}'] R = \text{QR}([U D, \hat{\Sigma}]) \leftarrow \text{通过QR分解求解}$$

$$[0208] \quad 16: [U', D', V'] = \text{SVD}(R)$$

[0209] 17: $U'_m \in U', D'_m \in D' \leftarrow$ 选择前 m 个特征向量和特征值, 其中, m 是通过等式16确定的

$$[0210] \quad 18: U = \text{sign}(U) \odot |[U', \Sigma'_{\text{新}}] U'_m|, D = D'_m \leftarrow \text{更新子空间模型}$$

[0211] 19: 结束条件

[0212] 输出: 经更新的 U 和 D

[0213] 算法1

[0214] 分级融合-通过将估计的稀疏特征向量76、78投影到分级体素上, 获得多个尺度的基于体素的人类目标图, 其中, 每个尺度具有针对对象的不同的量化描述。包含来自图7和图8的两个对象的视频序列的该投影在图9中被图示。特征向量78不仅决定子空间中的对象方向(符号), 而且还决定对应皮肤区域的搏动(幅度), 即前额和脸颊在投影中示出相对高的搏动。最后一步是将多个目标图融合为单个输出。由于分级测量创建了皮肤区域的统计

观察的事实,因此基本思想是利用该冗余来导出单个输出,对于该单个输出,所有尺度具有最高的一致性。在这个意义上,分级融合被计算为多尺度目标图之间的能量最小化:

$$[0215] \quad \arg \min_{\hat{o}} (\gamma E_1 + (1 - \gamma) E_2) \quad (18)$$

[0216] 其中

$$[0217] \quad \begin{cases} E_1 = \sum_i \sum_j \|o_{ij}, \hat{o}\|_2 \\ E_2 = \sum_i (\sum_j \|o_{ij}, o_{i-1,j}\|_2 + \sum_j \|o_{ij}, o_{i+1,j}\|_2) \\ o_{ij} \subseteq o_{i-1,j} \quad o_{i+1,j} \subseteq o_{ij} \end{cases} \quad (19)$$

[0218] 其中, o_{ij} 对应于由特征向量元素确定的第 i 个尺度中的第 j 个体素的目标值; $\hat{o}$ 指代融合目标图; γ 控制两个能量项之间的平衡。在等式19中, E_1 使输入和输出之间的能量最小化,而 E_2 使不同尺度的空间交叠体素之间的能量最小化,即隐式树结构。图9示出了在具有两个活的对象的视频序列中的融合结果的范例,其中,为识别的对象中的每个提供单独的输出80、82。

[0219] 上述方法和VPS方法的优选实施方案提供了一种用于识别视频序列中的活皮肤组织的改进的方法和装置。特别地,与常规技术相比,该方法提供了改进的活皮肤组织检测率,其检测仅基于使用搏动信号来检测活组织。这些改进被获得,而不管对象的尺度(即与成像单元4的距离)、姿势、位置、皮肤、可见身体部分或运动、或对象在视频序列中的背景、对象是否从成像单元4部分闭塞、视频序列中是否存在人工面部或身体部分,或者视频序列中是否存在多个活的对象。

[0220] 通过研究附图、公开内容以及权利要求书,本领域技术人员在实践所要求保护的本发明时能够理解和实现对所公开的实施例的变型。在权利要求书中,“包括”一词不排除其他元件或步骤,并且词语“一”或“一个”不排除多个。单个处理器或其他单元可以履行权利要求书中所记载的若干项目的功能。尽管在互不相同的从属权利要求中记载了特定措施,但是这并不指示不能有利地使用这些措施的组合。计算机程序可以被存储/分布在合适的介质上,例如与其他硬件一起提供或作为其他硬件的部分提供的光学存储介质或固态介质,但计算机程序也可以以其他形式来分布,例如经由因特网或者其他有线或无线电信系统分布。权利要求中的任何附图标记不应被解释为对范围的限制。

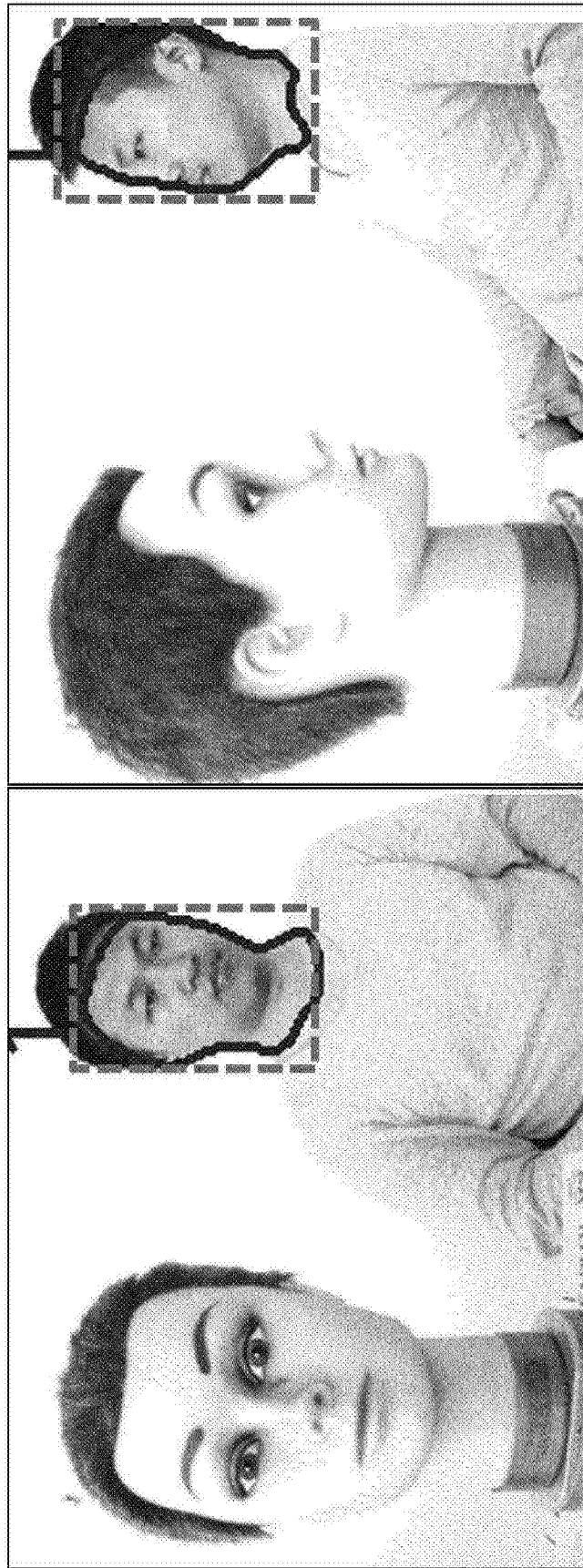


图1

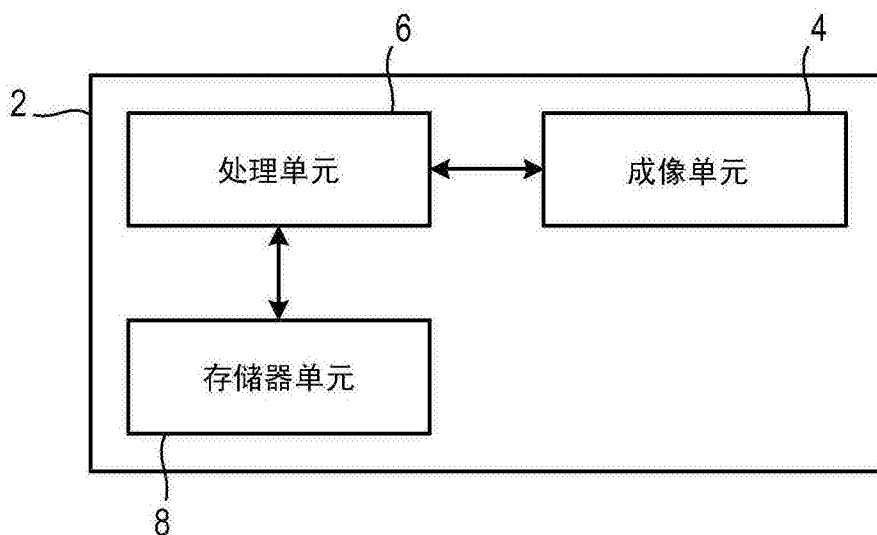


图2

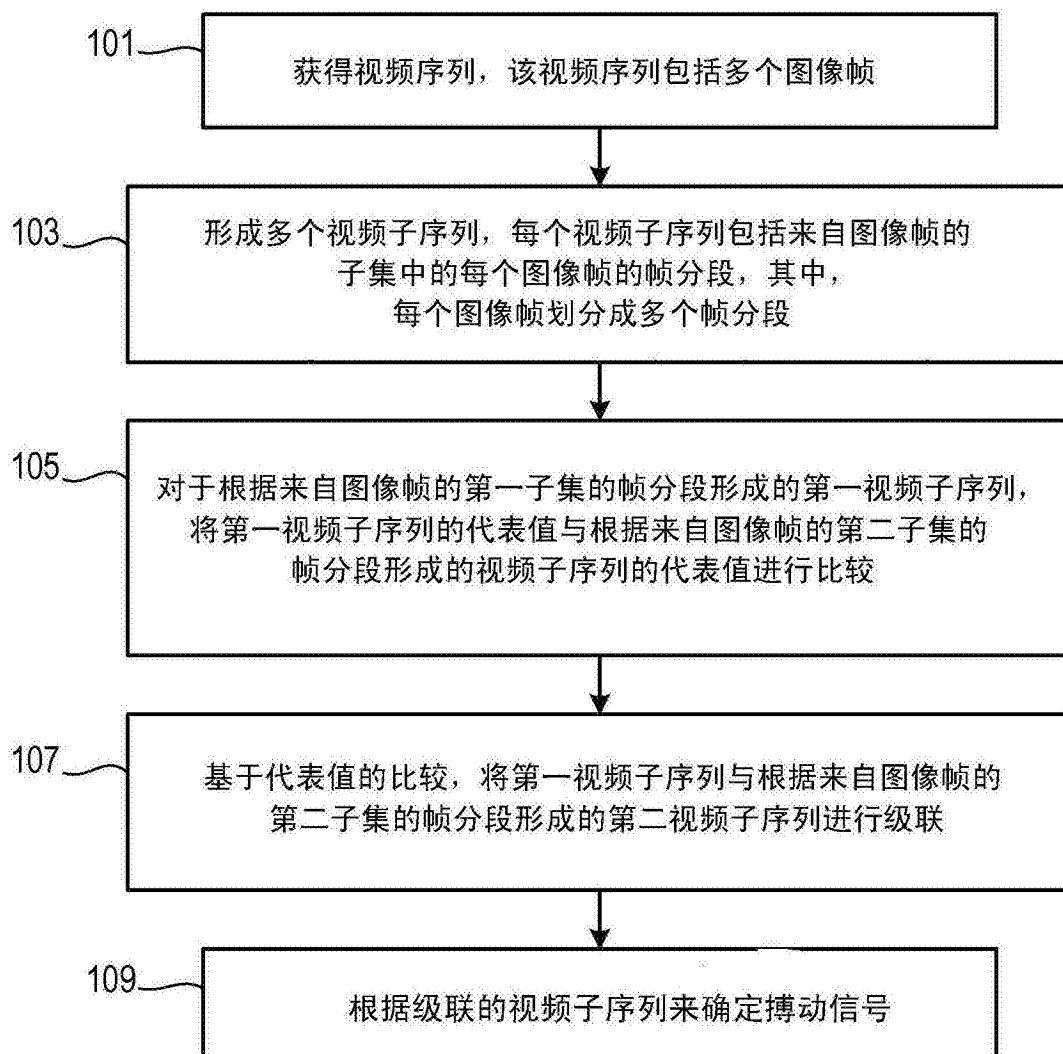


图3

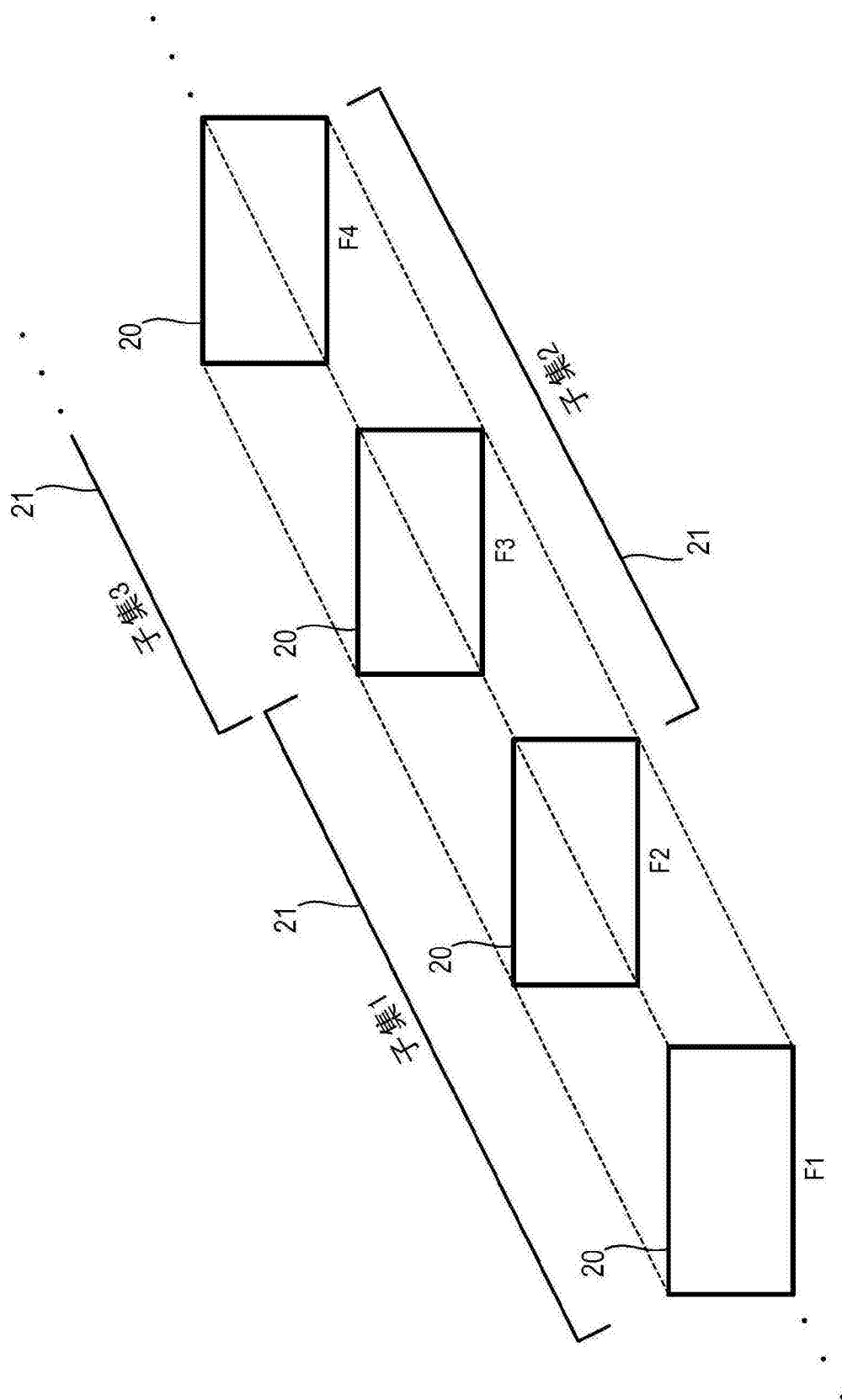


图4(a)

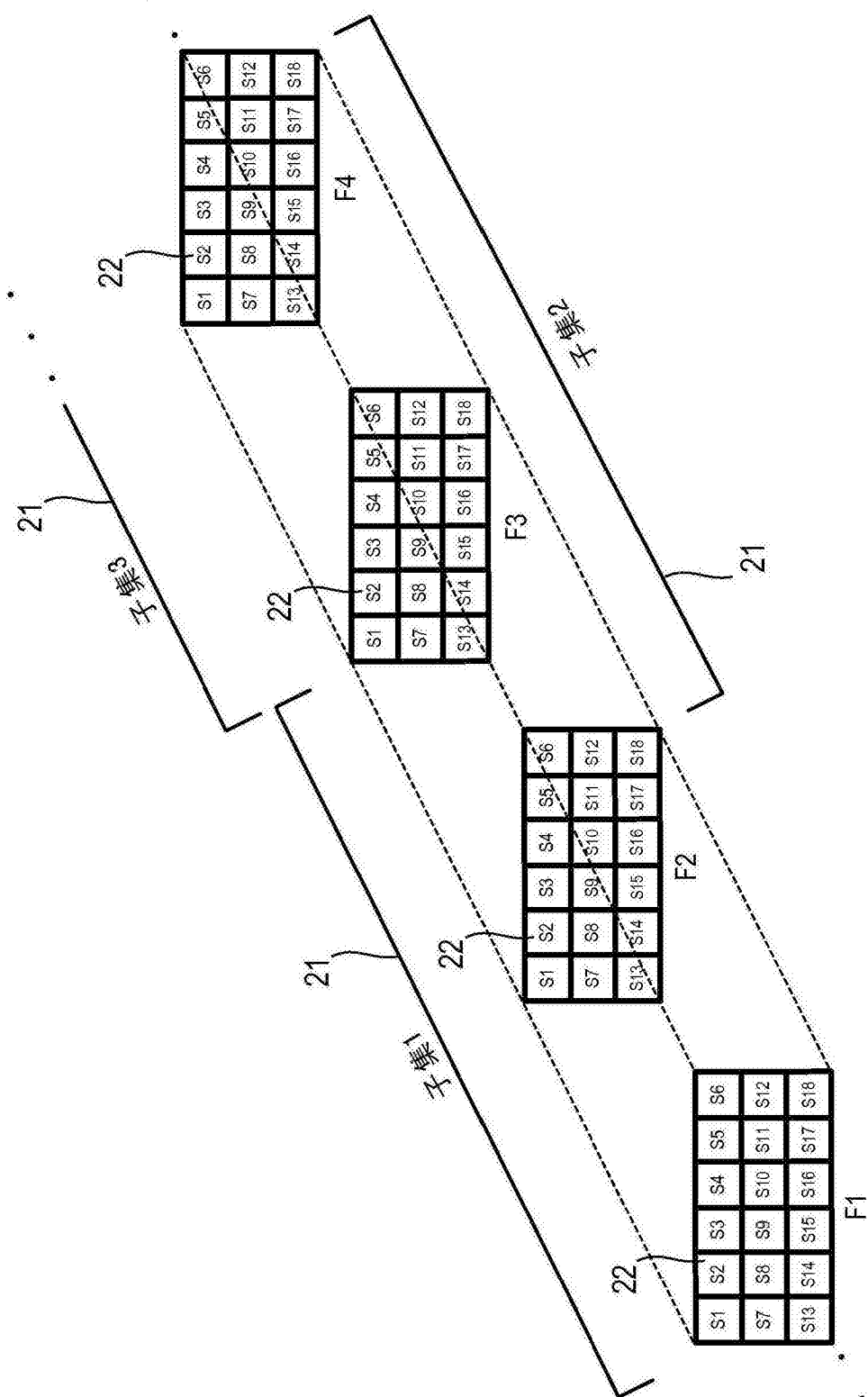


图4 (b)

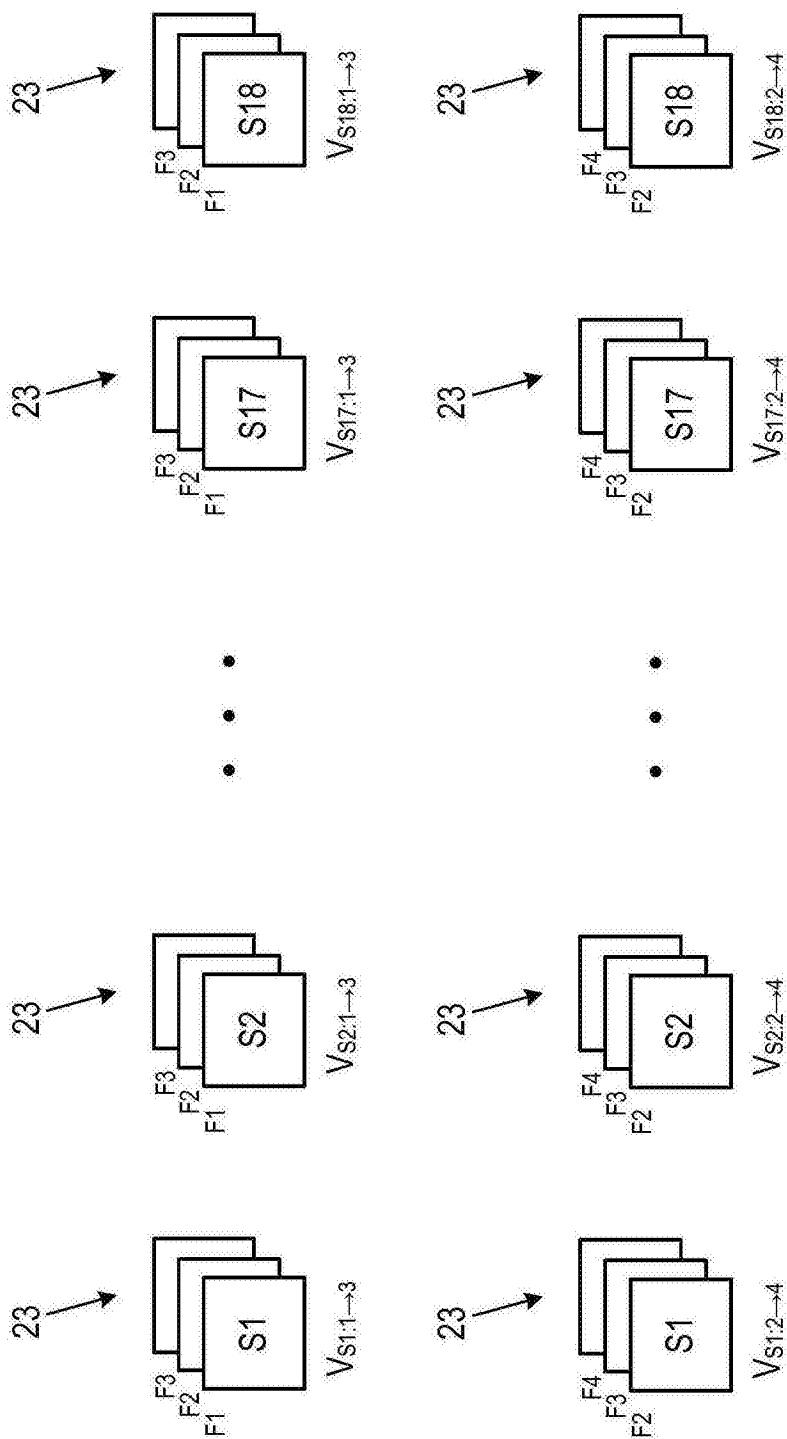


图4(c)

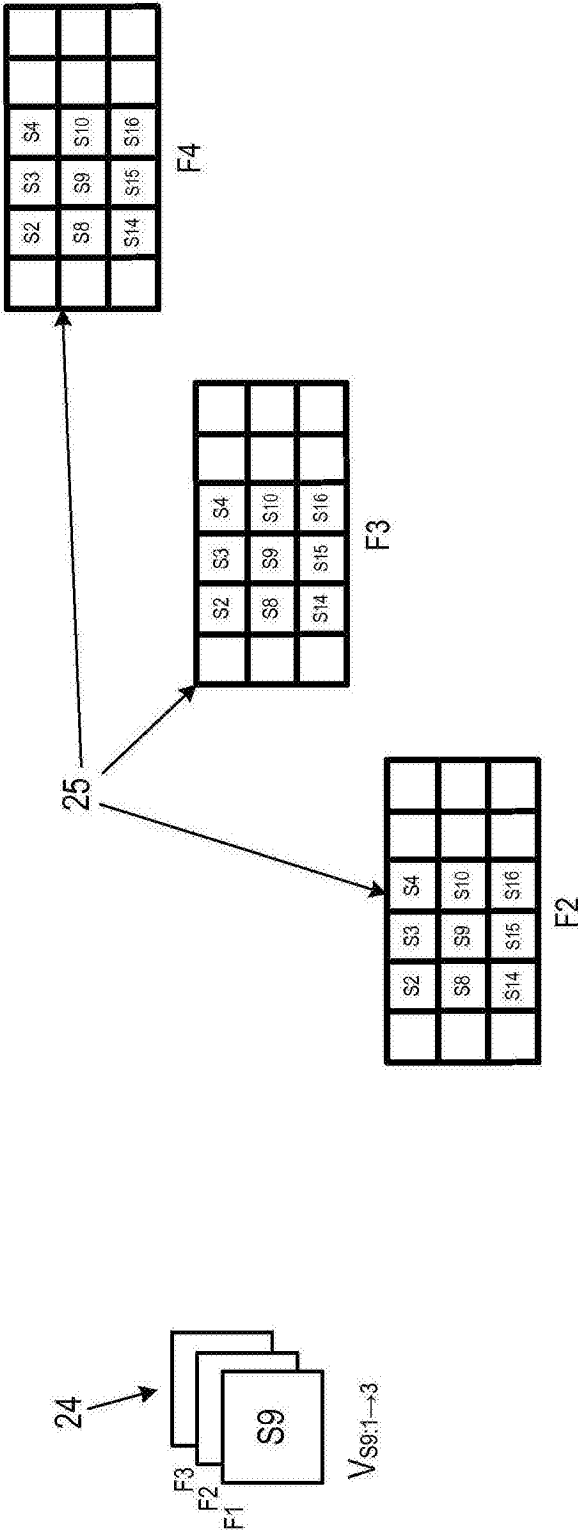


图4 (d)

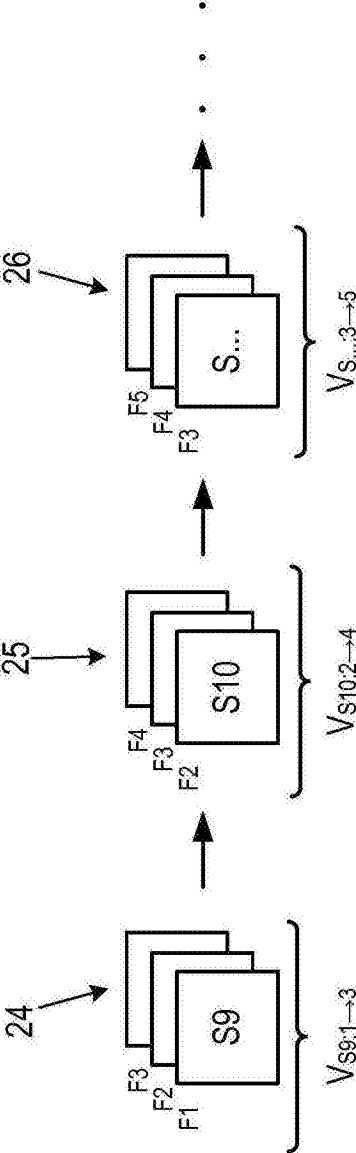


图4 (e)

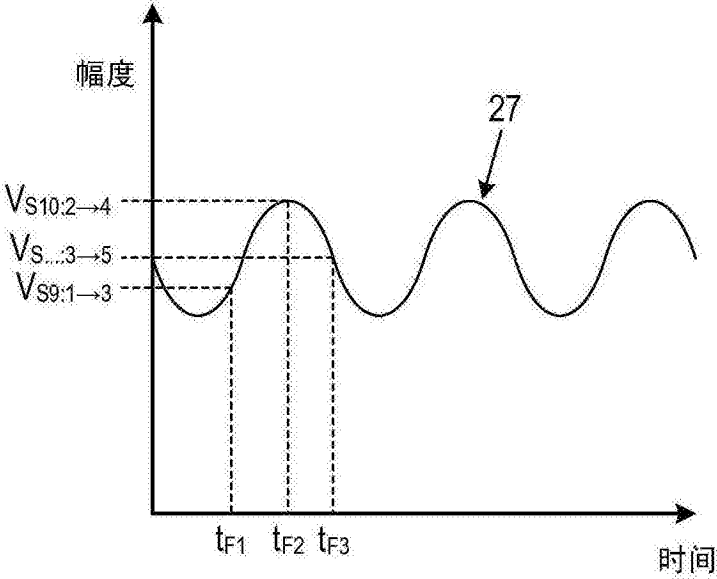


图4 (f)

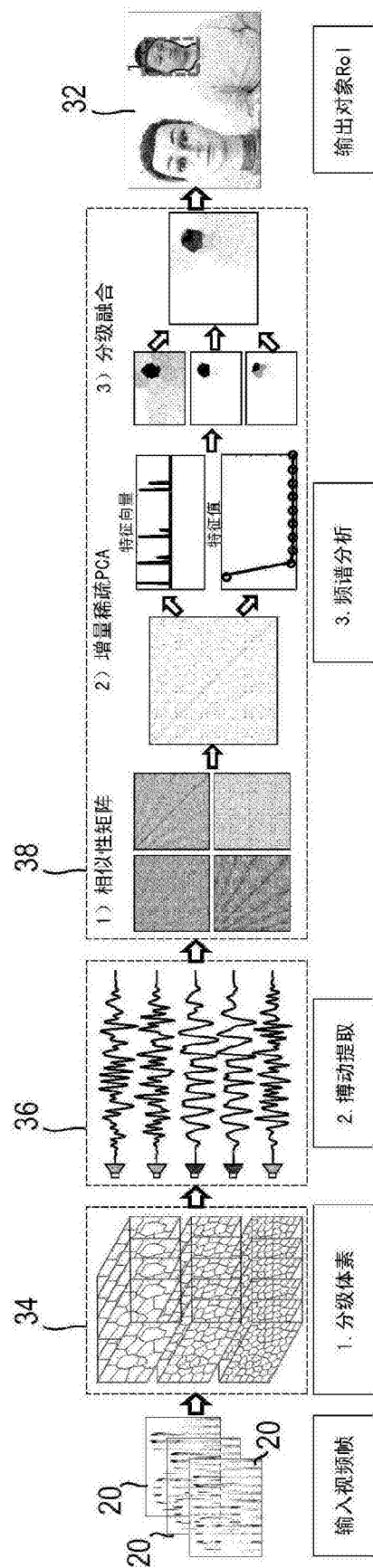


图5

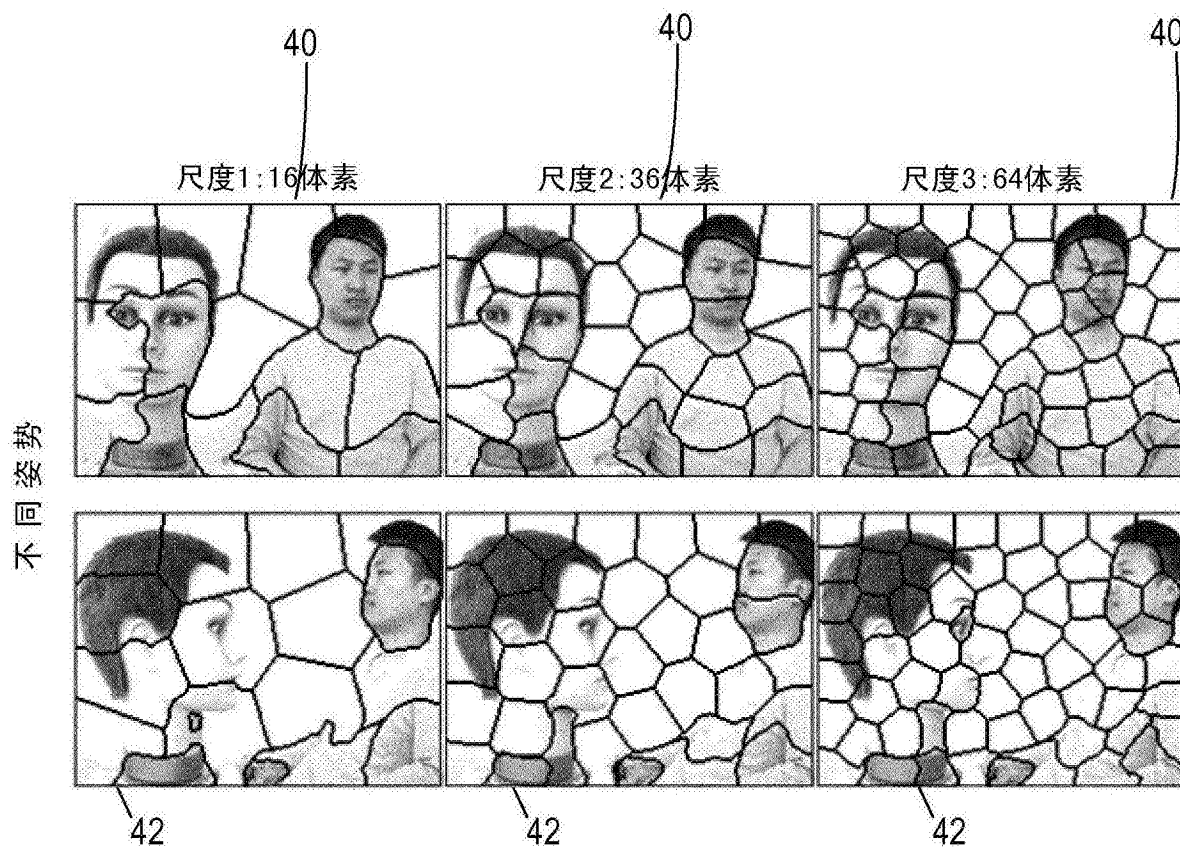


图6

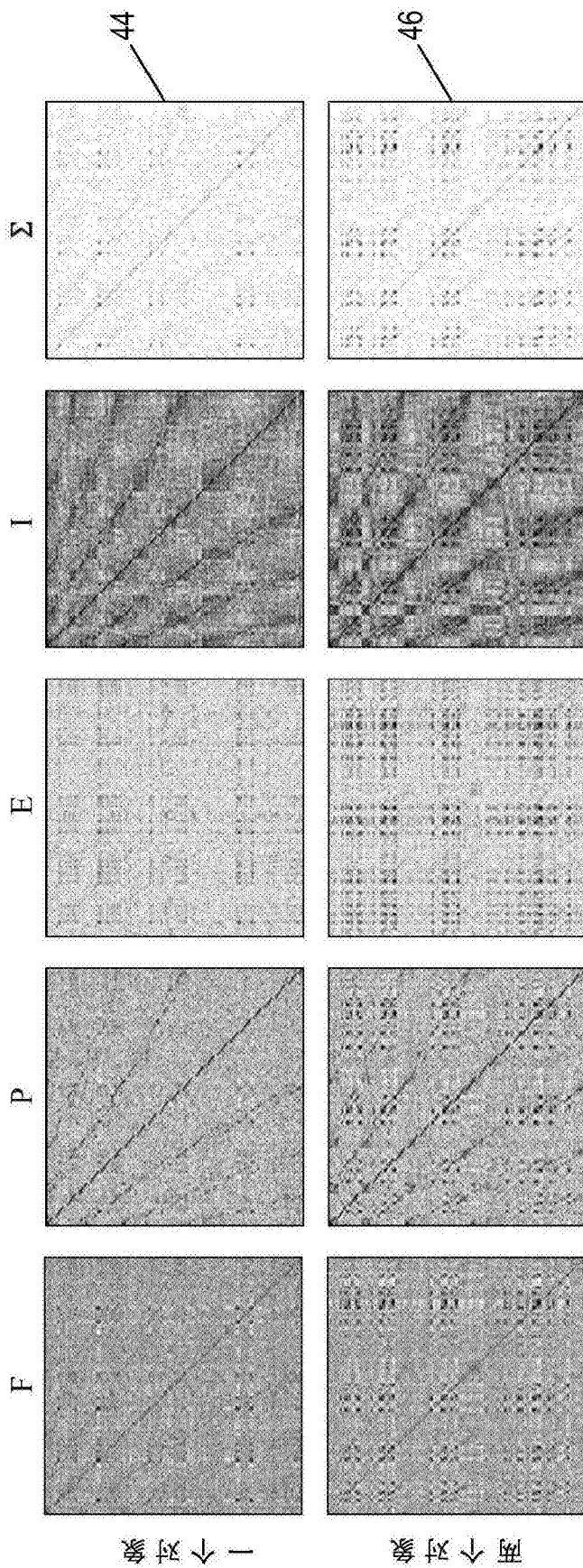


图7

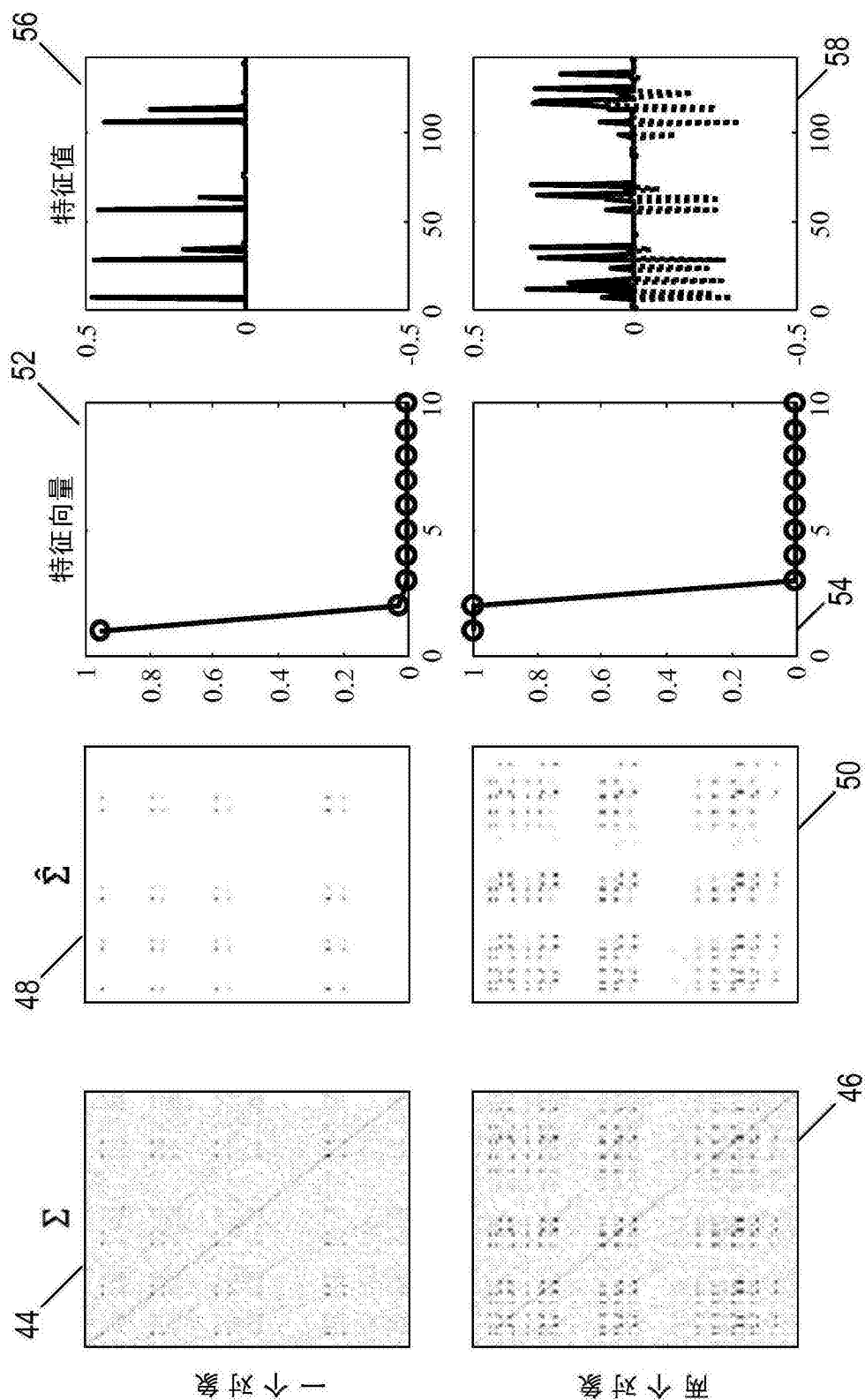


图8

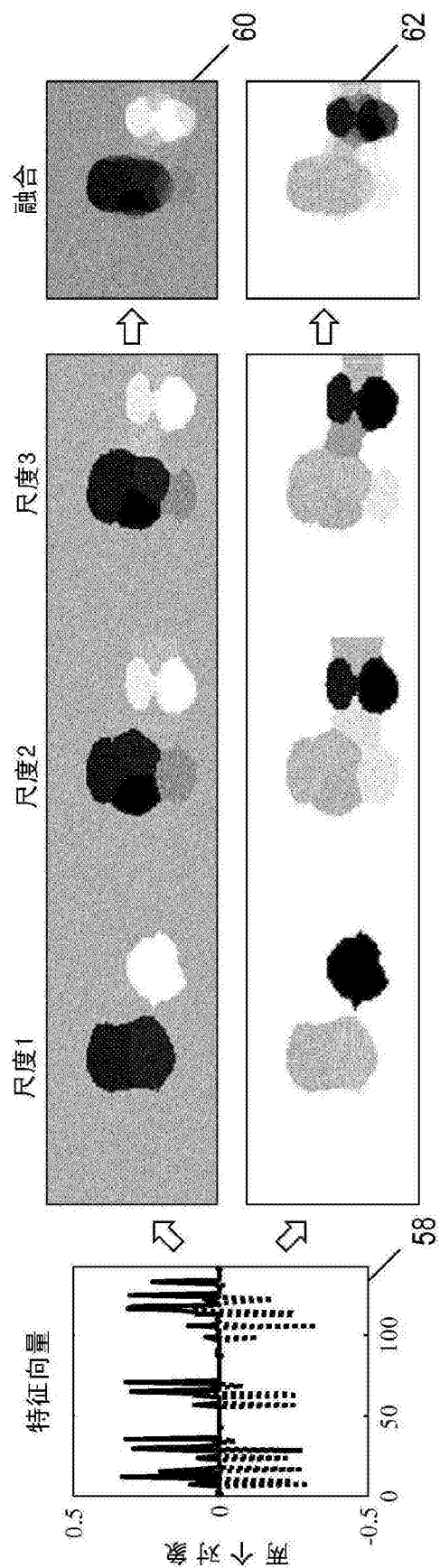


图9