

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06K 9/00 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200610152329.5

[43] 公开日 2007 年 4 月 4 日

[11] 公开号 CN 1940965A

[22] 申请日 2006.9.26

[21] 申请号 200610152329.5

[30] 优先权

[32] 2005.9.26 [33] JP [31] 2005-278782

[32] 2005.9.26 [33] JP [31] 2005-278783

[71] 申请人 佳能株式会社

地址 日本东京都大田区下丸子 3-30-2

[72] 发明人 金田雄司 真继优和 森克彦

[74] 专利代理机构 北京林达刘知识产权代理事务所
代理人 刘新宇 权鲜枝

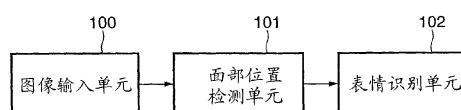
权利要求书 3 页 说明书 37 页 附图 43 页

[54] 发明名称

信息处理设备及其控制方法

[57] 摘要

本发明提供一种信息处理设备及其控制方法。
该信息处理设备包括：输入包含面部的图像数据的图像输入单元(100)、从图像数据中检测面部的特定部分的位置的面部位置检测单元(101)、及基于检测到的特定部分的位置从图像数据中检测面部的特征点并基于检测到的特征点判断面部表情的表情识别单元(102)。 以比检测特定部分的位置高的检测精度检测特征点。 特定部分的位置的检测对检测目标的变化具有鲁棒性。



1. 一种信息处理设备，包括：

输入单元，用于输入包含面部的图像数据；

第一检测单元，用于从所述图像数据中检测所述面部的特定部分的位置；

第二检测单元，用于基于检测到的所述特定部分的位置从所述图像数据中检测所述面部的特征点；以及

判断单元，用于基于检测到的所述特征点判断所述面部的表情，

其中，所述第二检测单元具有比所述第一检测单元的检测精度高的检测精度，所述第一检测单元对检测目标的变化具有鲁棒性。

2. 根据权利要求1所述的信息处理设备，其特征在于，所述第一检测单元基于分辨率比要由所述第二检测单元处理的图像数据低的图像数据检测所述面部的特定部分的位置。

3. 根据权利要求1所述的信息处理设备，其特征在于，还包括用于存储具有预定表情的面部的图像数据的存储单元，

其中，所述判断单元基于使用所述输入图像数据中的所述特征点计算的第一特征量与使用存储在所述存储单元中的所述图像数据中的所述特征点计算的所述第二特征量之间的差和比值中的一个判断第一表情。

4. 根据权利要求3所述的信息处理设备，其特征在于，所述判断单元在用预设的权重值乘以所述特征量之后执行所述判断。

5. 根据权利要求1所述的信息处理设备，其特征在于，

所述输入图像数据包括含有多个帧的运动图像，

所述第一检测单元检测每个所述帧中所述面部的特定部分的位置，

所述第二检测单元检测每个所述帧中所述面部的特征点，以

及

所述判断单元基于相互相邻的帧之间所述特征点的变化判断所述面部的表情的改变。

6. 根据权利要求1所述的信息处理设备, 其特征在于, 还包括用于校正所述输入图像数据的亮度以增强低亮度区域的色调的校正单元。

7. 根据权利要求1所述的信息处理设备, 其特征在于, 所述特定部分是面部区域的重心、眼睛区域的重心、嘴区域的重心、鼻孔重心位置、鼻边缘、及肤色区域的重心中的至少一个。

8. 根据权利要求1所述的信息处理设备, 其特征在于, 所述特征点是鼻孔重心、鼻边缘的重心、眉的端点、眉的上下特征点、眉的中点、眼睛的端点、眼睛的上下特征点、嘴的端点、及嘴的上下特征点中的至少一个。

9. 根据权利要求1所述的信息处理设备, 其特征在于, 所述特征点包括下列中的至少一个: 眉的形状、左眉和右眉之间的距离、眉和眼睛之间的距离、眼睛端点和嘴端点之间的距离、鼻孔重心和嘴端点之间的距离、鼻边缘和嘴端点之间的距离、面部区域的重心和嘴端点之间的距离、眼睛端点之间的距离、眼睛区域的上下边缘之间的距离、嘴端点之间的距离、嘴区域的上下边缘之间的距离、前额和眉间区域中的皱纹、及左右脸颊区域中的皱纹。

10. 根据权利要求1所述的信息处理设备, 其特征在于, 由相对于所述特定部分的相对坐标表示所述特征点。

11. 根据权利要求1所述的信息处理设备, 其特征在于, 所述判断单元基于所述输入图像数据中的所述特征点和所述特定部分之一的布局校正所述特征点的位置, 并基于具有校正过的位置的所述特征点来执行所述判断。

12. 根据权利要求1所述的信息处理设备，其特征在于，还包括用于当所述判断单元判断出预定表情时保持所述输入图像数据的保持单元。

13. 根据权利要求1所述的信息处理设备，其特征在于，还包括用于当所述判断单元判断出预定表情时显示所述输入图像数据的显示单元。

14. 根据权利要求1所述的信息处理设备，其特征在于，还包括用于基于由所述第一检测单元检测到的所述特定部分的位置在所述图像数据中设置区域的设置单元，

其中，所述第二检测单元从所设置的区域中检测所述面部的所述特征点。

15. 一种信息处理设备的控制方法，所述信息处理设备用于处理包含面部的图像数据，所述控制方法包括以下步骤：

输入步骤，用于输入包含面部的图像数据；

第一检测步骤，用于从所述图像数据中检测所述面部的特定部分的位置；

第二检测步骤，用于基于检测到的所述特定部分的位置从所述图像数据中检测所述面部的特征点；以及

判断步骤，用于基于检测到的所述特征点判断所述面部的表情，

其中，所述第二检测步骤具有比所述第一检测步骤的检测精度高的检测精度，所述第一检测步骤对检测目标的变化具有鲁棒性。

信息处理设备及其控制方法

技术领域

本发明涉及一种信息处理设备及其控制方法，尤其涉及一种图像识别技术。

背景技术

以往，已知使摄像装置拍摄被摄体来获取图像数据并通过分析图像数据计算被摄体的位置和方向的被摄体识别(图像识别)技术。

日本特开平09-282454号公报公开了下面的被摄体识别技术。首先，执行低分辨率的被摄体识别处理来粗略地得到整个识别目标被摄体的位置和方向(第一阶段的识别处理)。基于该识别结果在被摄体上的特征部分周围设置局部识别范围。只对该局部识别范围部分地执行高分辨率被摄体识别处理(第二阶段的识别处理)。被摄体上的特征部分包括，例如，螺钉或杆的孔、组件的凸起、及被摄体表面上的标记。基于在局部识别范围中的被摄体识别结果计算整个目标被摄体的位置和方向。

然而，在日本特开平09-282454号公报中公开的方案在第一阶段的识别处理和第二阶段的识别处理之间需要一预定时间。因为这个原因，在由于例如照明条件变化、识别目标被摄体的尺寸和形状变化、及识别目标被摄体的转动导致的摄像条件动态改变的环境中，很难精确识别被摄体。

因此，如果识别目标被摄体是人脸，且要识别在给定时间点的表情，则不能使用上面提到的传统技术。

另一方面，有另一项分析被摄面部图像的图像数据并基于分析结果识别被摄图像中识别目标的眼睛区域的传统技术。

日本专利第3452685号公开了面部图像处理技术。在该技术中，使用提取低亮度值的滤波器从面部图像中只提取低亮度值并二值化。计算二值图像的重心。将重心位置设置为面部的重心位置。基于重心位置设置眼睛存在区域。在所述存在区域中设置至少一个眼睛存在候选区域。该候选区域允许判断眼睛区域。

在日本专利第3452685号中公开的面部图像处理技术实现了对只含有面部的图像的处理。因此，如果图像中出现背景，则可能将面部重心位置识别为远不同于真实位置的位置。在这种情况下，不能正确设置眼睛区域。当通过日本专利第3452685号中公开的技术设置区域时，预先测量照相机及被摄体之间的距离，并基于测得的距离设置眼睛区域，而与被摄体的面部尺寸无关。因为这个原因，可能无法对任意的面部尺寸进行正确的区域设置。当发生例如转动等变化时，也可能无法进行正确的区域设置。

发明内容

考虑上述问题做出本发明，其目的是提供一种即使在拍摄条件动态改变的环境下也可精确识别被摄体的技术。本发明的另一个目的是提供一种在变化的拍摄条件下精确识别面部的技术。

为了达到以上目的，根据本发明的信息处理设备具有以下配置。所述信息处理设备包括：

输入单元，用于输入包含面部的图像数据；

第一检测单元，用于从所述图像数据中检测所述面部的特定部分的位置；

第二检测单元，用于基于检测到的所述特定部分的位置从所述图像数据中检测所述面部的特征点；以及

判断单元，用于基于检测到的所述特征点判断所述面部的表情，

其中，所述第二检测单元具有比所述第一检测单元的检测精度高的检测精度，所述第一检测单元对检测目标的变化具有鲁棒性。

为了达到上面的目的，根据本发明的信息处理设备的控制方法具有下面的配置。用于处理含有面部的图像数据的信息处理设备的控制方法包括以下步骤：

输入步骤，用于输入包含面部的图像数据；

第一检测步骤，用于从所述图像数据中检测所述面部的特定部分的位置；

第二检测步骤，用于基于检测到的所述特定部分的位置从所述图像数据中检测所述面部的特征点；以及

判断步骤，用于基于检测到的所述特征点判断所述面部的表情，

其中，所述第二检测步骤具有比所述第一检测步骤的检测精度高的检测精度，所述第一检测步骤对检测目标的变化具有鲁棒性。

通过下面参考附图对典型实施例的说明，本发明的进一步特征将变得明显。

附图说明

包含在说明书中并构成说明书一部分的附图示出本发明的实施例，并与本说明书一起用于解释本发明的原理。

图1是示出根据第一实施例的信息处理设备的功能配置的框图；

图2是示出神经网络的示意图；

图3是示意性示出直方图校正处理的图；

图4是示出给定层特征的神经元和前一层特征的多个神经元

之间的连接关系的图；

图5是示出与计算给定层特征的邻接神经元所需的前一层神经元的连接关系的图；

图6是示出表情识别单元的详细功能配置的框图；

图7是示出提取特征点的CNN的配置的图；

图8是示出将提取的特征点的示意图；

图9是示出从检测面部位置的CNN得到的面部、左/右眼、及嘴重心位置的示意图；

图10是示出得到鼻孔位置的鼻孔重心位置计算范围、得到重心计算范围所需的积-和运算(product-sum operation)范围、及得到重心计算范围所需的输入图像范围的示意图；

图11是示出左鼻孔和右鼻孔位置及鼻下边缘的示意图；

图12A、图12B、及图12C是示出计算左和右鼻孔位置的重心、右鼻孔位置的重心、及鼻下边缘的重心所需的各区域的示意图；

图13是示出得到左眉和右眉端点特征点的重心计算范围的示意图；

图14是示出得到左眉和右眉中央特征点的重心计算范围的示意图；

图15是示出得到左眼和右眼端点特征点的重心计算范围的示意图；

图16是示出得到左眼和右眼的上部和下部边缘的特征点的重心计算范围的示意图；

图17是示出得到嘴角特征点的重心计算范围的示意图；

图18是示出得到嘴的上部及下部边缘特征点的重心计算范围的示意图；

图19是示出前额、眉间、及脸颊区域的图；

图20是示出得到全部特征点所需的最小输入图像区域的图；

图21是示出用来检测尺寸变化和转动变化的左眼和右眼区域及面部区域的重心位置的图;

图22是示出当已发生了尺寸变化时左眼和右眼区域及面部区域的重心位置的图;

图23是示出当已发生了水平转动变化时左眼和右眼区域及面部区域的重心位置的图;

图24是示出当已发生了垂直转动变化时左眼和右眼区域及面部区域的重心位置的示意图;

图25是示出判断表情的CNN的配置的示意图;

图26是示出在根据特征量变化计算分数以判断表情“快乐”中特征量变化的权重的表;

图27是示出根据特征变量变化算出的分数的分布的曲线图;

图28是示出预先为表情“快乐”准备的分数分布模板的曲线图;

图29是示出根据第一实施例的整体处理过程的流程图;

图30是示出根据第二实施例的信息处理设备的功能配置的框图;

图31是示出表情识别单元的功能配置的框图;

图32是示出在 t [帧]及 $t+1$ [帧]图像中具有在面部检测位置的起点及在右外眼角特征点的终点的向量的示意图;

图33是示出运动向量的计算的示意图;

图34是示出内眼角间距离及具有在面部检测位置的起点和在右外眼角特征点的终点的向量的水平和垂直分量的图;

图35是示出当已发生了尺寸变化时内眼角间距离及具有在面部检测位置的起点和在右外眼角特征点的终点的向量的水平和垂直分量的图;

图36是示出根据第二实施例的整体处理过程的流程图;

图37是示出根据第三实施例的信息处理设备的功能配置的框图；

图38是示出根据第三实施例的整体处理过程的流程图；

图39是示意性示出根据第一到第三实施例的信息处理设备的硬件配置的框图；

图40是示出表113的内容的图；以及

图41是示出表313的内容的图。

具体实施方式

现在根据附图详细说明本发明的实施例。注意：下面实施例中的每个元件不意图限制本发明的范围，而仅仅是例子。

第一实施例

信息处理设备的硬件配置

首先参考图39说明根据本实施例的信息处理设备的硬件配置。图39是示意性示出本实施例的信息处理设备的硬件配置的框图。根据本实施例的信息处理设备是通过例如个人计算机(PC)、工作站(WS)、或个人数字助手(PDA)来实现的。

参考图39，CPU 390执行存储在硬盘(下文中称为HD)395中的应用程序、操作系统(OS)、及控制程序(稍后说明)。CPU 390还进行控制以将执行程序所需的信息和文件暂时存储在RAM 392中。

ROM 391存储包括基本I/O程序的程序及例如在文档处理中使用的字体数据和模板数据等各种数据。RAM 392暂时存储各种数据，并作为CPU 390的主存储器和工作区。

实现对记录介质的访问的外部存储驱动器393能够例如将程序从介质(记录介质)394加载到计算机系统。介质394可以是任意介质，例如：软盘(FD)、CD-ROM、CD-R、CD-RW、PC卡、

DVD、IC存储卡、MO、或记忆棒。

在本实施例中，外部存储装置395包括作为大容量存储装置的HD。HD 395存储应用程序、OS、控制程序、及相关程序。

指令输入装置396通过例如键盘、指示装置(例如鼠标)、及触摸式面板等装置来实现。用户通过使用指令输入装置396向本实施例的信息处理设备输入例如控制该信息处理设备的命令。

显示器397显示从指令输入装置396输入的命令或所述信息处理设备对该命令的响应输出。

系统总线399管理信息处理设备中的数据流。

摄像装置398拍摄被摄体并取得图像数据。摄像装置398包括例如成像光学系统、固态图像传感元件、及执行A/D转换的视频信号处理电路等组件。摄像装置398通过对从用作固态图像传感元件的CCD或CMOS传感器得到的电子信号进行A/D转换获取数字图像数据。在CPU 390的控制下对摄像装置398获取的图像数据进行缓冲处理，并通过DMA将其传输到例如RAM 392等存储器中。

代替硬件，可使用实现与上述硬件装置相同功能的软件。

在本实施例的例子中，将根据本实施例的程序及相关数据从介质394直接加载到RAM 392并执行。本实施例的程序可以预先安装在HD 395中并在每次运行本实施例的程序时从HD 395加载到RAM 392。可选择地，本实施例的程序可以作为部分存储器映射记录在ROM 391中并由CPU 390直接执行。

为了说明方便，本实施例的信息处理设备通过单一设备实现。然而，可以将资源分布到多个设备。例如，可以将存储及运算资源分布到多个设备。可以将资源分布到信息处理设备上的虚拟构成元件以执行并行处理。

信息处理设备的功能配置

下面参考图1说明用于上述信息处理设备进行被摄体识别的

功能配置。图1是示出根据本实施例的信息处理设备的功能配置的框图。

当前面参考图39说明的信息处理设备的CPU 390执行加载到RAM 392中的程序并与图1所示的每个硬件协作时，实现图1所示的功能框。所述功能框的一些或全部可以通过专用硬件来实现。

参考图1，图像输入单元100拍摄被摄体并获取图像数据。所述图像输入单元100与图39中的摄像装置398相对应。图像输入单元100获取图像数据并将其缓存到例如RAM 392等存储器中。

在本实施例中，由图像输入单元100输入的图像数据是面部图像的数据。在本实施例中，图像数据是包括多个帧的运动图像的数据。

面部位置检测单元101指定面部的位置，即，作为位置和方向计算目标的被摄体。所述面部位置检测单元101通过使用在图2中示意性示出的多层神经网络(第一CNN)指定面部位置。图2是神经网络的示意图。

在本实施例中，通过使用作为神经网络的卷积神经网络(convolutional neural network，下文中称为CNN)具体指定数字图像中的面部位置。所述CNN是已知技术，公开在例如M. Matsugu、K. Mori、M. Ishii、及Y. Mitarai, “Convolutional Spiking Neural Network Model for Robust Face Detection”, 9th International Conference on Neural Information Processing, pp.660-664, 2002年11月。所述CNN是通过本实施例的信息处理设备中的硬件和程序的协作来实现的。稍后详细说明面部位置检测单元101的操作。

表情识别单元102具有如图6所示的配置。图6是示出表情识别单元102的详细功能配置的框图。如图6所示，表情识别单元102包括：预定特征量提取单元110、特征量变化计算单元111、及表

情判断单元112。表情判断单元112通过查询包含特征量和表情之间对应关系的表113使神经元学习表情判断。

本实施例的配置使用两个网络：使面部位置检测单元101基于图像检测面部位置的CNN(第一CNN)，及使表情识别单元102得到识别表情所需的特征点的CNN(第二CNN)。

预定特征量提取单元110基于由面部位置检测单元101检测到的摄像目标的面部位置提取识别表情所需的预定特征量。特征量变化计算单元111根据由预定特征量提取单元110提取的特征量中的变化将特征量变化标准化(normalization)。在该标准化中，基于特征点在图像数据中的布局校正特征点的位置。表情判断单元112基于由特征量变化计算单元111标准化后的特征量变化判断表情。稍后详细说明包括在表情识别单元102中的预定特征量提取单元110、特征量变化计算单元111、及表情判断单元112。

整体处理

下面参考图29说明由本实施例的配置执行的整体处理。图29是示出根据本实施例的整体处理过程的流程图。

在步骤S270中，面部位置检测单元101对图像输入单元100获取的图像数据执行稀疏(decimation)及直方图校正。稀疏后的图像分辨率是例如 360×240 [像素]。

在步骤S271中，面部位置检测单元101使用CNN判断图像中的面部位置。通过稀疏将输入到判断面部位置的CNN的输入图像的分辨率进一步减少到例如 180×120 [像素]。

在步骤S272中，表情识别单元102判断是否检测到面部。如果检测到面部(步骤S272中“是”)，则处理前进到步骤S273。如果没有检测到面部(步骤S272中“否”)，则处理回到步骤S270以对下一帧的图像数据执行同样的处理。

在步骤S273中，预定特征量提取单元110通过使用由用于面

部位置检测的第一CNN提取的面部和眼睛位置设置鼻孔特征点提取范围。

在步骤S274中，预定特征量提取单元110基于在步骤S273中设置的提取范围提取鼻孔特征点。

在步骤S275中，预定特征量提取单元110通过使用判断面部位置的CNN获取的眼睛和嘴的位置及在步骤S274中提取的鼻孔特征点位置，来设置除鼻孔特征点之外的特征点提取范围。

在步骤S276中，预定特征量提取单元110基于在步骤S275中设置的提取范围使用第二CNN提取特征点。输入到提取特征点的第二CNN的输入图像的分辨率是例如 360×240 [像素]。

在步骤S277中，预定特征量提取单元110判断是否通过步骤S273到步骤S276中的处理提取了全部特征点。如果提取了全部特征点(步骤S277中“是”)，则处理前进到步骤S278。如果没有提取全部特征点(步骤S277中“否”)，则处理回到步骤S270以对下一帧执行相同的处理。

在步骤S278中，特征量变化计算单元111通过与预先准备的无表情的参考面部进行比较来计算特征量变化并根据变化将它们标准化。即，基于在图像数据中特征点的布局校正特征点的位置。将无表情参考面部的数据预先存储在例如HD 395等存储装置中。

在步骤S279中，表情判断单元112通过使用用于表情判断的NN判断表情。注意：NN表示神经网络(neural network)。

下面通过解释在每个功能配置中的处理来详细说明每个步骤中的处理。

面部位置检测单元101

下面详细说明面部位置检测单元101的功能。面部位置检测单元101基于面部的轮廓检测图像数据中面部的特定部分的位置(面部位置)。

面部位置检测单元101通过图像输入单元100获取存储在缓冲器中的图像数据，并作为预处理通过稀疏和直方图校正执行分辨率改变以减小照明条件的影响。面部位置检测单元101将校正后的图像数据输入到CNN。

如上所述，由图像输入单元100获取的图像数据暂时存储在缓冲器中。面部位置检测单元101通过稀疏从缓冲器中每隔一个像素读出所述图像数据。例如，如果缓存的图像数据的分辨率是 720×480 [像素]，则面部位置检测单元101通过稀疏获取分辨率为 360×240 [像素]的图像数据。

接下来，执行下面说明的直方图校正。如图3所示，生成输入图像的亮度值直方图130。图3是示意性示出直方图校正的图。亮度值直方图130表示输入图像(图像数据)的像素的亮度值的分布。横坐标代表亮度值，纵坐标代表像素数量(程度(degree))。

从亮度值直方图中提取在曲线端部的亮度值X 131和Y 132(最大和最小亮度值)。使用非线性函数133转换亮度值使得在曲线端部提取的亮度值131和132分别为例如255和0。选择用于减少例如阴影等照明条件的影响，即，增强低亮度区域的色调(tone)的函数，并将该函数作为所述非线性函数预先设置在所述信息处理设备中。

当以上述方式校正亮度来增强低亮度区域的色调时，可以独立于摄像条件精确地进行图像识别。

可以通过任何其它方法进行直方图校正。例如，预先设置上限和下限亮度值。将具有比下限值小的亮度值的像素转换为亮度值“0”。将具有等于或大于上限值的亮度值的像素转换为亮度值“255”。基于具有等于或小于下限值的亮度值的像素或具有等于或大于上限值的亮度值的像素适当地转换具有在下限值和上限值之间的亮度值的像素。也可以应用这个转换方法。

CNN的每层特征包括许多神经元。在本实施例中，一个神经元输出代表图像数据的一个像素的特征检测结果。例如，考虑只有一个前一层特征连接到给定层特征(CNN的子取样层或特征池(feature pooling)层)的情况。在这种情况下，如图4所示，某层的一个神经元121的内部状态值可以通过前一层特征的多个神经元120与对应于多个神经元120的权重因数数据的积和运算得到。图4是示出给定层特征的神经元和前一层特征的多个神经元之间的连接关系的图。

连接到一个神经元的前一层的神经元的数目依赖于被确定来提取特定特征的每个特征的感受野(receptive field)大小而改变。例如，如图5所示，如果得到某个特征所需的感受野大小是 3×5 ，则通过前一层中的 3×5 个神经元值122与 3×5 个权重因数的积和运算计算一个神经元的内部状态值124。图5是示出与计算给定层特征的邻接神经元所需的前一层神经元的连接关系的图。

可以通过前一层中从多个神经元122移位一个像素的区域的多个神经元123与权重因数的积和运算来计算紧邻神经元内部状态值124的神经元值125。即，通过将前一层中称为感受野的区域垂直和水平移位一个像素，并重复所设置的权重因数数据与位于每个感受野中的多个神经元值的积和运算来执行卷积运算。用这个处理，可以得到当前层的全部神经元的内部状态值。如果如图2所示多个前一层特征连接到给定层特征(CNN的特征检测层)，则在所连接的前一层特征中得到的内部状态值的和等于一个神经元的内部状态值。

通过使用预先给定的教授数据(supervisory data)进行学习得到所述权重因数数据。可以根据教授数据生成具有各种特征的CNN(层特征)。例如，当通过向检测面部位置的CNN的教授数据组给出例如照明变化、尺寸变化、及转动变化等各种变化来进行

学习时，与通过只给出特定变化例如只给出照明变化来进行学习的情况相比，位置检测精度降低。替代地，可以生成对这些变化具有鲁棒性的面部检测CNN(层特征)。可选择地，例如，可以通过只给出V型眼部端点的数据作为教授数据组来生成例如只能够精确检测V型眼部端点位置的层特征。

下面说明根据本实施例的CNN的每层。通过稀疏将图2中所示的输入到输入层的输入图像数据，即输入到指定图像数据中面部位置的CNN的图像数据的分辨率降低到 180×120 [像素]，以减少处理负担。

如图2所示，本实施例的CNN具有三层。第一层级(第一层201)提取总共四个特征：识别面部轮廓的倾斜(右上倾斜和右下倾斜)边缘、水平边缘、及垂直边缘。第二层级(第二层202)提取眼睛和嘴位置特征。

第三层级(第三层203)提取面部位置。如图9所示，面部位置包括在面部图像中预先定义的特定部分，即，眼睛区域重心位置160和161、嘴区域重心位置163、面部区域重心位置162、及鼻孔位置(稍后说明)。图9是示出从检测面部位置的CNN得到的面部、左/右眼、及嘴的重心位置的示意图。

即，根据本实施例的CNN的网络配置通过结合多个低阶特征(边缘级)检测结果提取中阶特征(眼睛和嘴)位置，然后从中阶特征(眼睛和嘴)检测结果提取高阶特征(面部位置)位置。

如上所述，因为使用了通过预先使用教授数据学习的权重因数，所以检测到这些特征。基于例如尺寸变化、转动变化、照明变化、及形状变化等各种变化的图像数据产生用于在检测面部的CNN中学习的教授数据。因此，建立起即使在多个变化的情况下也能够检测面部、眼睛、及嘴的位置的具有鲁棒性的网络。

可以基于在下面的条件下变化的环境中得到的图像进行例如

单个被摄体(面部)的图像数据学习。

- (1)尺寸变大三倍。
- (2)在垂直、水平及深度方向 45° 内发生转动变化。
- (3)在水平方向 45° 内发生平面转动变化。
- (4)在户内及户外照明环境下的摄像中照明条件改变。
- (5)眼睛和嘴的形状在垂直和水平方向上改变。

可以设计所述网络以进行学习,使得将眼睛、嘴、及面部的重心的外围区域认作它们正解(correct solution)位置。即,可以通过执行眼睛、嘴、及面部检测位置的积和运算结果的阈处理并计算等于或多于阈值的局部区域的重心位置来得到眼睛、嘴、及面部的正解位置。只有当确定了面部位置时才能确定眼睛和嘴的位置。即,在检测眼睛和嘴的位置的积和运算及阈处理中,检测到眼睛和嘴位置的候选位置。只用当通过确定面部位置的积和运算及阈处理确定了面部位置时,眼睛和嘴的位置才被确定。

CNN的层的数目、特征的数目、以及特征之间的连接关系可以被改变。例如使用最大神经元值而不是阈处理和重心计算的另一个方法可以基于眼睛、嘴、及面部的特征的神经元值计算位置信息。从图像输入单元得到的图像数据的分辨率不局限于 720×480 [像素]。输入到检测面部位置的CNN的输入图像的分辨率也不局限于 180×120 [像素]。

预定特征量提取单元110

下面说明包括在表情识别单元102中的预定特征量提取单元110。预定特征量提取单元110基于由面部位置检测单元101检测到的面部位置在图像数据中设置区域,稍后将详细说明。预定特征量提取单元110在所设置的区域中搜索面部特征点,然后基于找到的特征点判断表情。

如上所述,本实施例的配置使用两个网络:使面部位置检测

单元101基于图像检测面部位置的CNN(第一CNN)及使表情识别单元102得到识别表情所需的特征点的CNN(第二CNN)。预定特征量提取单元110使第二CNN基于输入图像及由面部位置检测单元101得到的眼睛、嘴、和面部检测位置提取识别表情所需的特征点。提取识别表情所需的特征点的第二CNN具有如图7所示的配置。图7是示出提取特征点的CNN的配置的图。

输入到提取特征点的第二CNN的输入图像是通过指定面部位置的第一CNN的预处理得到的直方图校正图像。图像分辨率是 360×240 [像素]。与检测面部位置的第一CNN不同,提取特征点的第二CNN以未经稀疏的 360×240 [像素]的高分辨率处理输入图像。这是因为存在于图像区域中的小区域中的特征点一定要精确地提取。提取特征点的第二CNN的输入图像分辨率不局限于 360×240 [像素]。

如图7所示,提取特征点的第二CNN具有两个层级(701和702)。第一层级701提取总共四个特征:倾斜(右上倾斜和右下倾斜)边缘、水平边缘、及垂直边缘。为了提取识别表情所需的特征点(左右眉特征点140到145、左右眼特征点146到153、鼻孔特征点154、及嘴特征点155到158),如图8所示,第二层级702对应于每个特征点准备CNN的一个特征。图8是示出要提取的特征点的示意图。

像检测面部的第一CNN一样,甚至提取特征点的第二CNN也可以通过使用由基于教授数据学习得到的权重因数精确地获取特征点。与检测面部位置的第一CNN不同,提取特征点的第二CNN使用仅特定变化的学习数据。因此,尽管不具有检测面部的第一CNN的高检测鲁棒性,但是提取特征点的第二CNN的特征位置检测精度非常高。

在本实施例中,使用只具有特定变化,即,眼睛和嘴形状变

化及照明变化的图像执行学习。然而，本发明不局限于此。例如，可以使用通过不降低特征点提取精度来改变照明变化宽度而获得的图像，即在多种照明环境下的图像，进行基于只有照明变化的学习。可以使用只具有例如照明变化和尺寸变化等其它特定变化的图像执行学习。可以与尺寸变化、转动变化、及照明变化中的每个相对应地准备单个特征点的特征。像检测面部的第一CNN一样，提取特征点的第二CNN的层的数目、特征的数目、及特征之间的连接关系可以改变。提取特征点的CNN不总是需要从一个特征提取一个特征点。例如右眼外眼角(V型)和左眼内眼角(V型)等相似特征的特征点可以从CNN的相同特征提取。

预定特征量提取单元110限定每层的每个特征的处理区域并通过使用提取特征点的第二CNN执行操作。具体来说，预定特征量提取单元110基于由检测面部位置的第一CNN(面部位置检测单元101)计算的面部位置确定提取每个特征点的处理区域限制范围。如图9所示，面部位置包括例如，眼睛区域重心位置160和161、嘴区域重心位置163、面部区域重心位置162、及鼻孔位置(稍后说明)。

区域限定处理

下面参考图10详细说明由预定特征量提取单元110执行的提取鼻孔重心位置的区域限定处理。图10是示出用于得到鼻孔位置的鼻孔重心位置计算范围(重心计算范围)，得到该重心计算范围所需的积和运算范围，及得到该重心计算范围所需的输入图像范围的示意图。

参考图10，区域173表示重心计算范围。如图10所示，重心计算范围173是具有基于右眼检测位置170和左眼检测位置171确定的水平范围的矩形区域。重心计算范围173的垂直范围基于右眼检测位置170或左眼检测位置171和嘴检测位置172来确定。

重心计算范围173被用以从得到的神经元值计算重心位置。为了在重心计算范围173中计算重心，重心计算范围173中必须存在神经元值。可以通过使用检测鼻孔的感受野尺寸及第一层的每个特征的感受野尺寸计算确保在重心计算范围173中存在神经元所需的输入图像数据的最小区域。

更具体来说，为了在鼻孔位置重心计算范围173中得到神经元值，检测鼻孔的延伸了感受野尺寸的 $1/2$ 的区域174的第一层的特征神经元值是必要的。因此，第一层级的每个特征需要区域174的神经元值。为了得到第一层中区域174的神经元值，检测第一层的每个特征的延伸了感受野尺寸的 $1/2$ 的区域175的输入图像数据是必要的。以这个方式，可以计算鼻孔位置重心计算范围所需的最小输入图像数据区域。在这些限定范围中，如上所述，可以通过执行前一层的神元值与权重因数的积和运算，然后执行阈处理及重心位置检测来计算鼻孔位置。

如图11所示的右鼻孔重心位置176、左鼻孔重心位置177、左和右鼻孔的重心位置、及鼻下边缘178中的任何一个可作为鼻孔重心位置计算。图11是示出左和右鼻孔位置及鼻下边缘的示意图。

为了学习，将包括要设置为鼻孔位置的部分的区域设置为感受野。通过将学习正解点设置到包括要设置为鼻孔位置的部分的区域的重心位置进行学习。图12A、图12B、及图12C是示出计算左右鼻孔位置的重心、右鼻孔位置的重心、及鼻下边缘的重心所需的感受野的示意图。

例如，如图12A所示，为了将左右鼻孔的重心位置作为鼻孔位置计算，将包括左右鼻孔的区域设置为感受野。通过将学习正解点设置到左右鼻孔的重心位置执行学习。如图12B所示，为了将右鼻孔重心位置176作为鼻孔位置计算，将包括右鼻孔的区域设置为感受野。通过将学习正解点设置到右鼻孔重心位置执行学习。

如图12C所示, 为了将鼻下边缘178作为鼻孔位置计算, 将包括鼻下边缘的区域设置为感受野。通过将学习正解点设置到鼻下边缘重心位置执行学习。在本实施例中, 将左右鼻孔的重心位置作为鼻孔位置计算。下面说明的剩余特征点由相对于鼻孔位置的相对位置表示。

特征点重心计算范围设置处理

下面参考图13到图18及图20说明设置提取除鼻孔特征点之外的特征点的特征点重心计算范围的处理。图13到图18是示出重心计算范围的图, 更具体来说, 所述重心计算范围分别是得到左右眉端点特征点、左右眉中间特征点、左右眼睛端点特征点、左右眼睛的上下边缘的特征点、嘴端点特征点、及嘴的上下边缘的特征点的重心计算范围。图20是示出得到全部特征点所需的最小输入图像区域的图。在下面的说明中, 右眼检测位置181和左眼检测位置182之间的距离定义为 L 。水平位置被定义为 X 轴位置, 垂直位置被定义为 Y 轴位置。

下面说明提取左右眉的每个特征点的重心计算范围。参考图13, 定义提取图8中特征点140的区域183以包括具有从“右眼检测位置181的 x 坐标 $-L/2$ ”到“右眼检测位置181的 x 坐标”的水平长度的 X 轴区域以及具有从“右眼检测位置181的 y 坐标 $-L/2$ ”到“右眼检测位置181的 y 坐标”的垂直长度的 Y 轴区域。定义提取图8中特征点142的区域184以包括具有从“右眼检测位置181的 x 坐标”到“鼻孔位置180的 x 坐标”的水平长度的 X 轴区域以及具有从“右眼检测位置181的 y 坐标 $-L/2$ ”到“右眼检测位置181的 y 坐标”的垂直长度的 Y 轴区域。

参考图14, 定义提取图8中特征点141的区域187以包括具有从“右眼检测位置181的 x 坐标 $-L/4$ ”到“右眼检测位置181的 x 坐标 $+L/4$ ”的水平长度的 X 轴区域以及具有从“右眼检测位置181的

y坐标 $-L/2$ ”到“右眼检测位置181的y坐标”的垂直长度的Y轴区域。左眉特征点提取区域185、186、及188像右眉特征点提取区域183、184、及187一样设置。

下面说明提取左右眼的每个特征点的重心计算范围。参考图15，定义提取图8中特征点146的区域189以包括具有从“右眼检测位置181的x坐标 $-L/2$ ”到“右眼检测位置181的x坐标”的水平长度的X轴区域以及从“右眼检测位置181的y坐标 $-L/2$ ”到“右眼检测位置181的y坐标 $+L/2$ ”的垂直长度的Y轴区域。定义提取图8中特征点149的区域190以包括具有从“右眼检测位置181的x坐标”到“鼻孔位置180的x坐标”的水平长度的X轴区域以及具有从“右眼检测位置181的y坐标 $-L/2$ ”到“右眼检测位置181的y坐标 $+L/2$ ”的垂直长度的Y轴区域。

参考图16，定义提取特征点147的区域193以包括具有从“右眼检测位置181的x坐标 $-L/8$ ”到“右眼检测位置181的x坐标 $+L/8$ ”的水平长度的X轴区域以及具有从“右眼检测位置181的y坐标 $-L/4$ ”到“右眼检测位置181的y坐标”的垂直长度的Y轴区域。定义提取图8中特征点148的区域194以包括具有从“右眼检测位置181的x坐标 $-L/8$ ”到“右眼检测位置181的x坐标 $+L/8$ ”的水平长度的X轴区域以及从“右眼检测位置181的y坐标”到“右眼检测位置181的y坐标 $+L/4$ ”的垂直长度的Y轴区域。左眼特征点提取区域191、192、195、及196像右眼特征点提取区域189、190、193、及194一样设置。

下面说明提取嘴的每个特征点的重心计算范围。将图17中鼻孔位置180和嘴检测位置197之间的距离定义为 L_1 。如上所述，将水平位置定义为X轴位置，将垂直位置定义为Y轴位置。

参考图17，定义提取图8中特征点155的区域198以具有从“嘴检测位置197的x坐标 $-2L/3$ ”到“嘴检测位置197的x坐标”的水

平长度以及从“嘴检测位置197的y坐标 $-L_1$ ”到“嘴检测位置197的y坐标 $+L$ ”的垂直长度。定义提取图8中特征点158的区域199以具有从“嘴检测位置197的x坐标”到“嘴检测位置197的x坐标 $+2L/3$ ”的水平长度以及从“嘴检测位置197的y坐标 $-L_1$ ”到“嘴检测位置197的y坐标 $+L_1$ ”的垂直长度。

参考图18, 定义提取图8中特征点156的区域200以具有从“嘴检测位置197的x坐标 $-L/4$ ”到“嘴检测位置197的x坐标 $+L/4$ ”的水平长度以及从“鼻孔位置180的y坐标”到“嘴检测位置197的y坐标”的垂直长度。定义提取图8中特征点157的区域201以具有从“嘴检测位置197的x坐标 $-L/4$ ”到“嘴检测位置197的x坐标 $+L/4$ ”的水平长度以及从“嘴检测位置197的y坐标”到“嘴检测位置197的y坐标 $+L_1$ ”的垂直长度。

如上所述, 预定特征量提取单元110基于由面部位置检测单元101检测的摄像目标的面部位置确定提取特征点的每个重心计算范围。以上述方式通过使用得到每个特征点的感受野尺寸及第一层的每个特征的感受野尺寸计算在输入图像中如图20中的阴影区域210一样的最小且必要的输入图像数据区域。因为限定了所述区域, 所以可以减少在特征点提取中CNN上的处理负担。

上述配置基于在当前帧中由用于面部检测的第一CNN得到的面部检测位置、左右眼检测位置、嘴检测位置、及鼻孔位置设置提取特征点的区域。然而, 本发明不局限于此。例如, 可以基于在前一帧中提取的上述内容(例如, 在前一帧中提取的鼻孔位置及特征点)来提取特征点。可选择地, 可以基于特征点之间的多个位置设置所述区域。本发明不局限于上述区域设置范围。

如上所述, 特征点坐标被表示为相对于鼻孔检测位置的相对位置(图8中特征点154)。然而, 本发明不局限于此。例如, 特征点坐标可以表示为相对于面部检测位置或内眼角特征点(图8中的

特征点149或150)的相对位置。

特征量

下面参考图8和图19说明从所得到的特征点识别表情所需的特征量。图19是示出前额、眉间、及脸颊区域的图。

在本实施例中，下面的特征量被提取并用于表情识别。下面列出的特征量只是例子，根据应用和目的可以使用任何其它值作为特征量。

眉的形状(例如，图8中由连接特征点140和141的线段及连接特征点141和142的线段形成的角(倾斜)和/或由连接特征点143和144的线段及连接特征点144和145的线段形成的角(倾斜))。

左眉和右眉之间的距离(图8中特征点142和143之间的距离)。

眉和眼睛之间的距离(图8中特征点140和146之间的距离、特征点141和147之间的距离、特征点142和149之间的距离、特征点143和150之间的距离、特征点144和151之间的距离、及特征点145和153之间的距离)。

眼睛端点和嘴端点之间的距离(图8中特征点146和155之间的距离及特征点153和158之间的距离)。

眼睛端点之间的距离(图8中特征点146和149之间的距离及特征点150和153之间的距离)。

眼睛区域的上下边缘之间的距离(图8中特征点147和148之间的距离及特征点151和152之间的距离)。

嘴端点之间的距离(图8中特征点155和158之间的距离)。

嘴区域的上下边缘之间的距离(图8中特征点156和157之间的距离)。

前额和眉间区域中的皱纹(图19中区域220和221的边缘密度)。

左和右脸颊区域中的皱纹(图19中区域222和223的边缘密

度)。

图19中的前额和眉间区域220是例如包括具有从“右眼检测位置181的x坐标”到“鼻孔位置180的x坐标”的水平长度的x轴区域及具有从“右眼检测位置181的y坐标- $2L/3$ ”到“右眼检测位置181的y坐标”的垂直长度y轴区域的矩形区域。右眼检测位置181和左眼检测位置182之间的距离是 L 。脸颊区域222是例如包括具有从“鼻孔位置180的x坐标- L ”到“鼻孔位置180的x坐标”的水平长度的x轴区域及具有从“鼻孔位置180的y坐标- $L/4$ ”到“嘴检测位置197的y坐标”的垂直长度的y轴区域的矩形区域。

可以通过基于CNN的第一层的边缘特征提取的结果对包含在区域中的边缘中的像素数目进行计数并由区域的面积去除像素的数目来计算边缘密度。

特征量变化计算单元111

下面说明特征量变化计算单元111。特征量变化计算单元111通过计算预先准备的无表情面部图像和当前帧的面部图像之间的每个特征量的比值计算每个特征量的变化。特征量变化计算单元111还根据图像中面部的尺寸及转动变化将特征量变化标准化。如上所述，标准化是基于图像数据中特征点的布局校正特征点的位置。

如图21所示，基于右内眼角特征点230的检测位置和内眼角中点233之间的距离 $a1$ 、左内眼角特征点231的检测位置和内眼角中点233之间的距离 $b1$ 、鼻孔位置232的检测位置和内眼角中点233之间的距离 $c1$ 检测变化。将在预先设置(准备)的无表情面部图像中的右内眼角特征点和内眼角中点233之间的距离、左内眼角特征点和内眼角中点233之间的距离、及鼻孔位置和内眼角中点233之间的距离分别表示为 a 、 b 、及 c 。

通过计算从图22中所示的当前帧得到的检测位置之间的距离

a_1 (图22中240)、 b_1 (图22中241)、及 c_1 (图22中242)与从预设的无表情面部图像得到的检测位置之间的距离 a 、 b 、及 c 的比值判断面部的尺寸变化。图22是示出当已发生尺寸变化时左右眼区域及面部区域的重心位置的图。例如,当 $a:b:c=a_1:b_1:c_1$,且 $a:a_1=1:2$ 时,面部尺寸变化是两倍。在这种情况下,通过用 $1/2$ 乘以每个算出的特征量变化来进行标准化。

可以通过例如将图23中所示的当前帧图像中的 $a_2:b_2$ (图23中250和251)与预先准备的无表情的正面的面部图像中的 $a:b$ 相比较来计算面部的水平转动变化。图23是示出当已发生水平转动变化时左右眼区域和面部区域的重心位置的图。

例如,考虑图23中所示的转到左侧的面部的识别。假设在预先准备的无表情的正面面部图像中 $a:b=5:5$,且在当前帧图像中 $a_2:b_2=5:3$ (图23中的250和251)。在这种情况下,可以通过用 $(a_2/b_2)/(a/b)$ 乘以对向左转动有影响的水平特征量,即,左眉端点之间的距离、左眼端点之间的距离、及嘴端点之间的距离来进行标准化。左眉端点之间的距离是例如图8中特征点143和145之间的距离。左眼端点之间的距离是例如图8中特征点150和153之间的距离。嘴端点之间的距离是例如图8中特征点155和158之间的距离。

眉形状可以通过用 $(a_2/b_2)/(a/b)$ 乘以从特征点143和144计算的水平区域及从特征点144和145计算的水平区域来标准化。

可以基于在当前帧的面部图像中的距离 c_3 (图24中的262)与在预先准备的无表情的正面面部图像中的距离 c 的比值判断面部的垂直转动变化。图24是示出当已发生垂直转动变化时左右眼区域及面部区域的重心位置的示意图。例如,当 $a/a_3=b/b_3=1$,且 $c:c_3=2:1$ 时,面部只在垂直方向发生了变化。在这种情况下,可以通过使用由用 c_3/c 乘以垂直特征量,即,眼睛端点和嘴端点之间的

距离、眉和眼睛之间的距离、眼睛区域的上下边缘之间的距离、及嘴区域的上下边缘之间的距离得到的值作为变化来执行标准化。

眼睛端点和嘴端点之间的距离包括,例如,图8中的特征点146和155之间的距离及特征点153和158之间的距离。眉和眼睛之间的距离包括,例如,图8中特征点140和146之间的距离、特征点141和147之间的距离、特征点142和149之间的距离、特征点143和150之间的距离、特征点144和151之间的距离、以及特征点145和153之间的距离。眼睛区域的上下边缘之间的距离包括,例如,图8中特征点147和148之间的距离及特征点151和152之间的距离。嘴区域的上下边缘之间的距离包括,例如,图8中特征点156和157之间的距离。

以上述配置,可以通过使用右内眼角特征点、左内眼角特征点、及鼻孔位置来检测变化。即使当转动和尺寸变化都已发生时,也可以由如上所述的同样处理(使用右内眼角特征点、左内眼角特征点、及鼻孔位置)来标准化特征量。上述标准化处理只是例子,本发明不局限于此。例如,可以通过使用面部部分例如右眼检测位置、左眼检测位置、及面部检测位置或其它特征点等检测变化,并且可以对特征量变化进行标准化。

表情判断单元112

下面参考图25说明表情判断单元112。图25是示出判断表情的CNN的配置的示意图。

表情判断单元112通过使用如图25所示的三层神经网络执行判断,该三层神经网络包括:接收由特征量变化计算单元111标准化了的特征量变化的输入层2501,中间层2502,及输出表情判断结果的输出层2503。在本实施例的配置中,给输入层的特征量变化和输出层的表情判断结果中的每一个分配一个神经元。

输入层2501接收标准化后的特征量变化。在本实施例中，输入层2501接收例如22个特征。

“眉的形状”特征量变化(4)

- “左右眉之间的距离”特征量变化(1)
- “眉和眼睛之间的距离”特征量变化(6)
- “眼睛端点和嘴端点之间的距离”特征量变化(2)
- “眼睛端点之间的距离”特征量变化(2)
- “眼睛区域的上下边缘之间的距离”特征量变化(2)
- “嘴端点之间的距离”特征量变化(1)
- “嘴区域的上下边缘之间的距离”特征量变化(1)
- “前额和眉间区域中的皱纹(边缘密度)”特征量变化(1)
- “左右脸颊区域中的皱纹(边缘密度)”特征量变化(2)

中间层(隐藏层)2502执行表情判断所需的中间处理。在本实施例中，中间层2502包括10个神经元(特征)。

输出层2503基于来自中间层2502的输入判断表情。在本实施例中，输出层2503包括输出表情“快乐”、“愤怒”、“伤心”、“可怜”、“无表情”、“担心”及“惊讶”8个特征(神经元)。

当识别目标面部形成某种表情时，特定特征量变化增加/减少。例如，在表情“快乐”中，特征量变化与无表情状态相比较以下面的方式增加/减少。眼睛端点和嘴端点之间(图8中146和155之间及153和158之间)的距离的变化减少。嘴端点之间(图8中155和158之间)距离的变化、面颊区域的边缘密度(图19中区域222和223的边缘密度)的变化、及外眼角和内眼角之间(图8中146和149之间及150和153之间)距离的变化增加。

可以基于增加或减少的特征量变化的种类及它们增加/减少的量判断识别目标面部的表情。在本实施例中，对应于每个表情为每个特征量变化设置阈值。基于阈值和检测到的特征量变化之

间的比较使NN学习表情。进行学习使得与基于特征量变化和阈值之间的数量关系判断的表情相对应的神经元输出“1”。输出层2503的输出值范围是0到1。

例如，以下面的方式设置与表情“快乐”相对应的特征量变化的阈值。无表情状态的特征量变化是“1”。

眼睛端点和嘴端点之间(图8中146和155之间及153和158之间)距离的变化：0.7。

· 嘴端点之间(图8中155和158之间)距离的变化(特征量变化2)：1.2

· 脸颊区域的边缘密度(图19中区域222和223的边缘密度)的变化(特征量变化4)：1.2

· 外眼角和内眼角之间(图8中146和149之间及150和153之间)距离的变化(特征量变化5)：1.1

· 其余特征量变化：1.0

当眼睛端点和嘴端点之间的距离的变化等于或小于阈值(0.7)，且嘴端点之间距离的变化、脸颊区域的边缘密度的变化、及外眼角和内眼角之间距离的变化等于或大于阈值(1.2、1.2、及1.1)时，NN学习“快乐”。即，NN学习使得对应于“快乐”的神经元输出“1”或几乎为“1”的值。阈值存储在表113中。图40是示出表113的内容的图。表情判断单元112通过查询表113控制神经元的学习。将表113预先定义在例如HD 395等存储装置中。

通过与到输入层的输入相对应地向NN的输出层2503提供教授数据来进行学习。因此，表情判断单元112可以通过参考在输出层接收特征量变化并判断表情的神经元来判断表情。

输入层2501、中间层2502、及输出层2503的配置不局限于上述配置。例如，可以预先为到输入层2501的输入及来自输出层2503的输出设置阈值。将等于或大于阈值的值定义为“1”，并将

等于或小于阈值的值定义为“0”，使得输入或输出“0”和“1”的值。要判断的表情不局限于“快乐”。例如，可以判断“愤怒”、“伤心”、“可怜”、“无表情”、“担心”及“惊讶”。

用于表情判断的NN的输出层可以输出多个具有强值(即，接近上限值的值)的特征。在这种情况下，基于输出强值的神经元组判断表情。例如，当得到表情“快乐”和“伤心”，即，对应于“快乐”的神经元和对应于“伤心”的神经元二者都输出强值时，将表情判断为几乎“喜极而泣”。当包括在输出层2503中的多个神经元输出强值时，可以以例如下面的方式进行表情判断。在例如HD 395等存储装置中准备存储输出强值的神经元组和表情之间对应关系的表。可以通过查询该表来判断表情。

在上述配置中，可以在例如用预设权重值乘以特征量之后进行所述判断。用于表情判断的配置不局限于基于上述方法的配置。下面参考图26到28说明基于不同方法的表情判断处理。图26是示出在根据判断表情“快乐”的特征量变化(variation)计算分数(score)中特征量变化的权重(权重值)的表。图27是示出从特征量变化计算的分数的分布的曲线图。图28是示出为表情“快乐”预先准备的分数分布模板的曲线图。

首先，如图26中所示，根据每个表情加权特征量变化。从所计算的权重值与特征量变化的乘积计算对应于特征量的分数。基于算出的分数生成表情分数分布。将生成的表情分数分布与为每个表情预设的分数分布模板相比较。将对应于具有相似分数分布的模板的表情判断为识别目标被摄体的面部所显示的表情。

例如，假设所算出的判断表情“快乐”的分数分布为图27中所示的分数分布。假设与图27中的分数分布相似的预设的分数分布模板为对应于图28中表情“快乐”的分数分布模板。在这种情况下，表情被判断为“快乐”。

如上所述，在根据本实施例的摄像装置中，基于面部轮廓检测图像数据中面部的特定部分的位置(面部位置)。基于检测到的面部位置设置搜索特征点的区域。不在图像数据的整个区域中而只在所设置的区域中搜索特征点。因此，可以有效地进行搜索操作。

在根据本实施例的摄像装置中，通过使用低分辨率图像数据检测面部位置。通过使用高分辨率图像数据执行特征点搜索。因为可以有效地搜索并精确地提取特征点，所以使得可以精确地判断表情。

在根据本实施例的摄像装置中，两个网络(神经网络)的使用使得即使当发生了各种变化时也允许精确地提取特征点。另外，通过精确地提取特征点，甚至可以识别面部特征改变非常小的表情改变。

第二实施例

在第一实施例中，将预先注册的无表情参考面部的特征量与识别目标面部的特征量相比较。基于所算出的特征量变化判断表情。然而，表情判断方法不局限于此。在第二实施例中，说明以下配置，在该配置中分析被测量图像的每个帧，并基于获取的运动向量判断表情中的变化。本实施例的信息处理设备具有与在第一实施例中相同的硬件配置。

信息处理设备的功能配置

首先参考图30说明根据本实施例的用于被摄体识别的功能配置。图30是示出根据本实施例的信息处理设备的功能配置的框图。

如图30所示，本实施例的信息处理设备的功能配置包括图像输入单元300、面部位置检测单元301、及表情识别单元302。在图像输入单元300和面部位置检测单元301中的处理与在第一实施例中的相同，所以省略其说明。

图31是示出表情识别单元302的功能配置的框图。在本实施例

中，如图31所示，表情识别单元302包括：预定特征点提取单元310、运动向量计算单元311、及表情判断单元312。表情判断单元312通过查询存储运动向量和表情改变之间的对应关系的表313使神经元学习表情改变判断。在预定特征点提取单元310中的处理与第一实施例中的相同，所以省略其说明。在本实施例中，基于面部检测位置表示特征点坐标。然而，本发明不局限于此。

运动向量计算单元311基于由面部位置检测单元301检测到的面部位置计算每个以面部位置为起点且以特征点为终点的运动向量。与在第一实施例中一样，表情检测单元112通过使用NN判断表情。

整体处理

下面参考图36说明由本实施例的配置执行的整体处理。图36是示出根据本实施例的整体处理过程的流程图。

在步骤S320中，面部位置检测单元301对图像输入单元300获取的图像数据执行稀疏和直方图校正。稀疏后的图像分辨率是例如 360×240 [像素]。

在步骤S321中，面部位置检测单元301通过使用CNN判断图像中的面部位置。通过稀疏将输入到判断面部位置的CNN的输入图像的分辨率进一步减少到例如 180×120 [像素]。

在步骤S322中，表情识别单元302判断是否检测到面部。如果检测到面部(在步骤S322中“是”)，则处理前进到步骤S323。如果没有检测到面部(在步骤S322中“否”)，则处理回到步骤S320以对下一帧的图像数据执行相同的处理。

在步骤S323中，预定特征点提取单元310通过使用由用于面部位置检测的第一CNN提取的面部和眼睛位置设置鼻孔特征点提取范围。

在步骤S324中，预定特征点提取单元310基于在步骤S323中

设置的提取范围使用第二CNN提取特征点。输入到用于特征点提取的第二CNN的输入图像的分辨率是例如 360×240 [像素]。

在步骤S325中，预定特征点提取单元310判断步骤S323和S324中的处理是否提取了全部特征点。如果提取了全部特征点(在步骤S325中“是”)，则处理进入到步骤S326。如果没有提取全部特征点(步骤S325中“否”)，则处理回到步骤S320以对下一帧执行相同的处理。

在步骤S326中，运动向量计算单元311通过将在前一帧中计算的向量与在当前帧中计算的向量相比较来计算特征点的运动向量。

在步骤S327中，基于在步骤S326中计算的运动向量，通过使用用于表情判断的NN判断表情。处理结束。

下面通过解释在每个功能配置中的处理详细说明每个步骤中的处理。

运动向量计算单元311

下面详细说明运动向量计算单元311的功能。基于由面部位置检测单元301检测到的面部位置，运动向量计算单元311计算每个以面部位置为起点且以特征点为终点的运动向量。运动向量的数目等于如图8中所示除鼻孔特征点之外的特征点的数目。

参考图32说明运动向量计算。图32是示出在 t [帧]和 $t+1$ [帧]图像中以面部检测位置为起点且以右外眼角特征点为终点的向量的示意图。

参考图32，附图标记3201表示作为参考点的面部检测位置；3202表示在 t [帧]中的外眼角特征点；3203表示在 $t+1$ [帧]中的外眼角特征点。如图32所示，在 t [帧]和 $t+1$ [帧]中，通过设置面部检测位置3201作为起点并设置外眼角特征点3202和3203作为终点来定义向量 c 和 b 。运动向量 a 定义为 $a=b-c$ 。

图33是示出运动向量的计算的示意图。对其余特征点类似地计算运动向量。计算除鼻孔特征点之外全部18个运动向量。根据计算运动向量的帧速率可以使用 t [帧]和 $t+2$ [帧]或 $t+3$ [帧]来代替使用 t [帧]和 $t+1$ [帧]。

变化改变了所计算的运动向量的方向和尺寸。执行标准化来处理尺寸改变。例如，基于内眼角间距离 $|f|$ 表示每个向量的尺寸。

例如，参考图34，当将向量 f 定义为标准化的参考时，标准化后的向量 d 可以用 $d/|f|$ 表示。如图35所示，如果尺寸变化并且内眼角距离改变到 $|g|$ ，则标准化后的图35中的向量 e 可以用 $e/|g|$ 表示。通过该标准化，如果只有尺寸变化而不发生例如眼睛及嘴等面部特征的改变，则向量 d 等于向量 e 。这可以抑制由摄像角度引起的识别错误。

如果已发生了水平转动变化，则只有图34中的向量的水平分量改变。根据转动变化对图34中向量 d 的水平分量 d_2 的大小进行标准化。在标准化中，通过使用面部检测位置和左右眼检测位置检测转动，并基于检测到的转动校正特征点的布局，如在第一实施例中说明的一样。

例如，在图23中，用 a_2/b_2 乘以从转动方向区域中的特征点得到的每个向量的水平分量。如图33所示，从 $b-c=a$ 计算运动向量 a 。在转动方向区域中的特征点是例如图8中的特征点143、144、145、150、151、152、153、及158。

即使在垂直转动变化中，用 c/c_3 乘以从除鼻孔特征点之外的全部特征点得到的每个向量的垂直分量 d_1 的大小。之后，如图33所示，从 $b-c=a$ 计算运动向量 a 。

从特征点计算出的向量的起点不局限于上述面部检测位置。可选择地，可以使用鼻孔特征点位置(图8中的特征点154)、内眼角特征点(图8中的特征点149和150)、由面部检测CNN得到的眼睛

检测位置(图9中的右眼检测位置160和左眼检测位置161)、及嘴检测位置(图9中的163)。

表情判断单元312

下面说明表情判断单元312。与在第一实施例中相同，表情判断单元312通过使用NN判断表情。在第一实施例中，输入由与预先准备在例如HD395等存储装置中的无表情面部相比较得到的22个标准化后的特征量变化。在第二实施例中，例如，将18个运动向量的水平和垂直分量，即总共36个向量的尺寸和方向输入到NN。例如，运动向量(4, -3)可以被分解为水平分量+4和垂直分量-3。输入向量的分量的尺寸和方向。

另一方面，所述输出包括输出从“0”到“1”的值的八个表情判断神经元。输出系统的神经元与第一实施例的相同。下面说明表情的学习。如在第一实施例中所述，当作为识别目标被摄体的面部呈现某种表情时，特定特征量变化增加/减少。当作为识别目标被摄体的面部呈现某种表情时，运动向量还具有特定方向和尺寸。因为这个原因，当将代表某种表情的运动向量的特定方向和尺寸输入到输入层的特征时，使得输出层中代表该表情的神经元输出接近“1”的值。由此执行学习。

表313存储运动向量的参数(例如，代表方向和尺寸的值)和表情之间的对应关系。图41是示出表313的内容的图。表情判断单元312通过查询表313控制神经元的学习。例如，如果预先定义的运动向量1的参数1趋向增加同时参数2趋向减少，则控制学习以增加“快乐”的输出级。表113预先定义在例如HD 395等存储装置中。

如上所述，在根据本实施例的配置中，基于根据相邻帧中的特征点计算的运动向量判断表情。因此，可以有效地检测到表情中的改变。

第三实施例

在第一和第二实施例中，将信息处理设备假设为PC、WS、或PDA。然而，本发明不局限于此。例如，上述配置可以由例如数字照相机等摄像装置实现。

本实施例的配置将面部检测和表情识别功能合并例如数字照相机(照相机)等摄像装置中以使它能够自动检测出被摄体呈现出预设的所希望的表情(例如，“快乐”)并将其自动记录。另外，显示所记录的图像。

图37是示出根据第三实施例的信息处理设备的功能配置的框图。本实施例的信息处理设备包括：图像输入单元400、面部位置检测单元401、表情识别单元402、图像显示单元403、及图像存储单元404，如图37所示。

图像输入单元400、面部位置检测单元401、及表情识别单元402执行与在第一及第二实施例中相同的处理。

图像显示单元403在显示器397上显示由表情识别单元402判断的具有预设表情的图像。即，将暂时存储在例如RAM 392等缓冲存储器中的图像数据显示在显示器397上。这时，可以每几个像素隔行扫描并显示图像数据。在这种情况下，可高速显示。

图像存储单元404将显示在显示器397上的图像数据存储例如RAM或存储器(例如，闪速存储器)394等存储装置中。

整体处理

下面参考图38说明本实施例的配置执行的整体处理。图38是示出根据本实施例的整体处理过程的流程图。

在步骤S410中，面部位置检测单元401对图像输入单元400获取的图像数据执行稀疏和直方图校正。稀疏后的图像分辨率是例如 360×240 [像素]。

在步骤S411中，面部位置检测单元401通过使用CNN判断图像中的面部位置。通过稀疏将输入到判断面部位置的CNN的输入

图像的分辨率进一步减少到例如 180×120 [像素]。

在步骤S412中，表情识别单元402判断是否检测到面部。如果检测到面部(步骤S412中“是”)，则处理进入到步骤S413。如果没检测到面部(步骤S412中“否”)，则处理回到步骤S410以对下一帧的图像数据执行相同的处理。

在步骤S413中，表情识别单元402通过使用由用于面部位置检测的第一CNN提取的面部和眼睛位置设置鼻孔特征点提取范围。

在步骤S414中，表情识别单元402基于在步骤S413中设置的提取范围通过使用第二CNN提取特征点。输入到用于特征点提取的第二CNN的输入图像的分辨率是例如 360×240 [像素]。

在步骤S415中，表情识别单元402判断步骤S413和S414中的处理是否提取了全部特征点。如果提取了全部特征点(步骤S415中“是”)，则处理进入到步骤S416。如果没有提取全部特征点(步骤S415中“否”)，则处理回到步骤S410以对下一帧执行同样的处理。

在步骤S416中，表情识别单元402通过将在前一帧中计算的向量与在当前帧中计算的向量相比较来计算特征点的运动向量。

在步骤S417中，基于在步骤S416中计算的运动向量通过使用用于表情判断的NN判断表情。

在步骤S418中，判断在步骤S417中是否识别出表情。如果识别出表情(步骤S418中“是”)，则处理进入到步骤S419。如果没有识别出表情(步骤S418中“否”)，则处理回到步骤S410以继续该处理。

在步骤S419中，将具有识别出的表情的图像数据显示在显示器397上。根据需要以低分辨率进行该显示。除图像数据之外，在显示器397上还显示允许用户选择是否将所显示的图像数据存储

在例如介质394等存储装置中的用户界面。

如果用户在步骤S420中选择存储图像数据(步骤S420中的“是”),则处理进入到步骤S421。如果未选择存储(步骤S420中的“否”),则处理回到步骤S410以继续该处理。

在步骤S421中,将图像数据以高分辨率存储在介质394(例如,闪速存储器)。该处理结束。

可以根据例如下面的方式执行步骤S418到步骤S421中的处理。预先设置将显示在显示器397上和/或存储在例如介质394等存储装置中的表情。在步骤S418中,判断识别目标图像是否被识别为具有预设表情。如果识别到该表情(在步骤S418中“是”),则处理进入到步骤S419。如果没有识别到该表情(在步骤S418中“否”),则处理回到步骤S410。

在步骤S419中,将图像数据显示在显示器397上。处理跳过步骤S420进入到步骤S421。

在步骤S421中,将图像数据存储存储在例如介质394等存储装置中。

如上所述,根据本实施例的摄像装置自动识别图像数据的表情并只显示及存储对应于预设的表情的图像数据。因此,用户不会因为失去按快门的机会而得不到希望的图像。

其它实施例

上面详细地说明了本发明的实施例。本发明可以采用例如系统、设备、方法、程序、或存储介质等形式。更具体来说,本发明适用于包括多个装置的系统或包括单个装置的设备。

甚至通过直接或从远程站点向系统或设备提供实现上述实施例的功能的程序并使该系统或设备的计算机读出并执行所提供的程序代码也可以实现本发明。

因此,所述安装在计算机中以由计算机实现本发明的功能处

理的程序代码本身也包括在本发明的技术范围中。即，本发明还包括实现本发明的功能处理的计算机程序。

在这种情况下，该程序可以采取任何方式，例如目标代码、由编译器执行的程序、或要提供到OS的脚本数据等，只要可以得到程序的功能即可。

所述提供程序的记录介质包括：例如软盘®、硬盘、光盘、磁光盘、MO、CD-ROM、CD-R、CD-RW、磁带、非易失性存储卡、ROM、或DVD(DVD-ROM或DVD-R)。

作为另一个程序提供方法，可以使用客户计算机中的浏览器将客户计算机连接到因特网上的主页，并且可以从主页将本发明的计算机程序本身或含有自动安装功能的压缩文件下载到例如硬盘等的记录介质。可以将包含在本发明的程序中的程序代码分为多个文件，且可从不同的主页下载所述文件。即，使多个用户下载使计算机实现本发明的功能处理的程序文件的WWW服务器也包括在本发明的权利要求中。

下面的提供形式也是有效的。可以将本发明的程序加密，存储在例如CD-ROM等存储介质中，并分发给用户。任何满足预定条件的用户都可以被允许通过因特网从主页下载用于解密的密钥信息，通过使用该密钥信息执行加密的程序，并将程序安装在计算机中。上述提供方式也是有效的。

当由计算机执行所读出的程序时，以及当例如在计算机上运行的OS基于该程序的指令执行部分或全部实际处理时都实现了上述实施例的功能。

当从记录介质读出的程序被写进插入到计算机的功能扩展板或连接到计算机的功能扩展单元的存储器中，且该功能扩展板或功能扩展单元的CPU基于该程序的指令执行部分或全部实际处理时，上述实施例的功能也被实现。

如上所述，根据所述实施例，可以提供在各种摄像条件下以高精度识别面部的技术。

虽然参考典型实施例说明了本发明，但应理解，本发明不局限于所公开的典型实施例。所附权利要求的范围符合最宽的解释以包括全部这种改变及等同结构和功能。

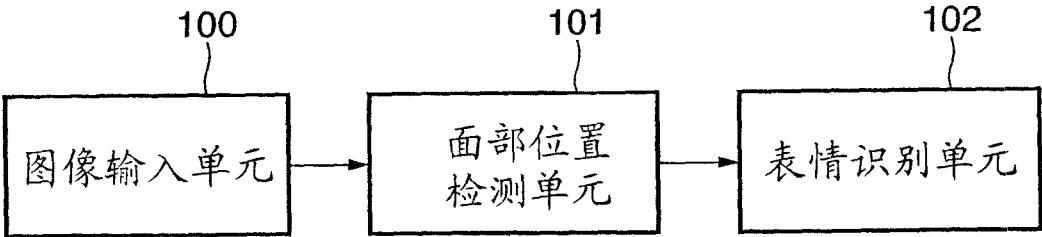


图 1

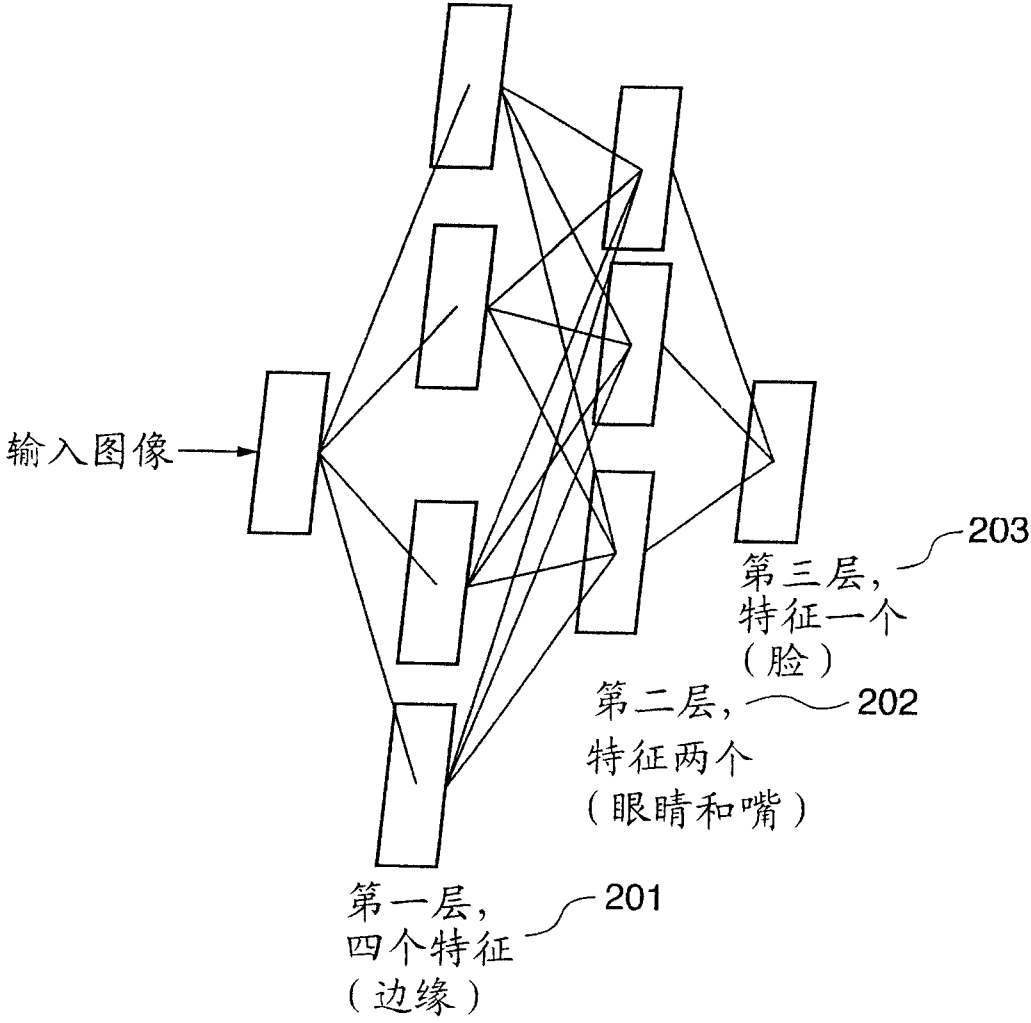


图 2

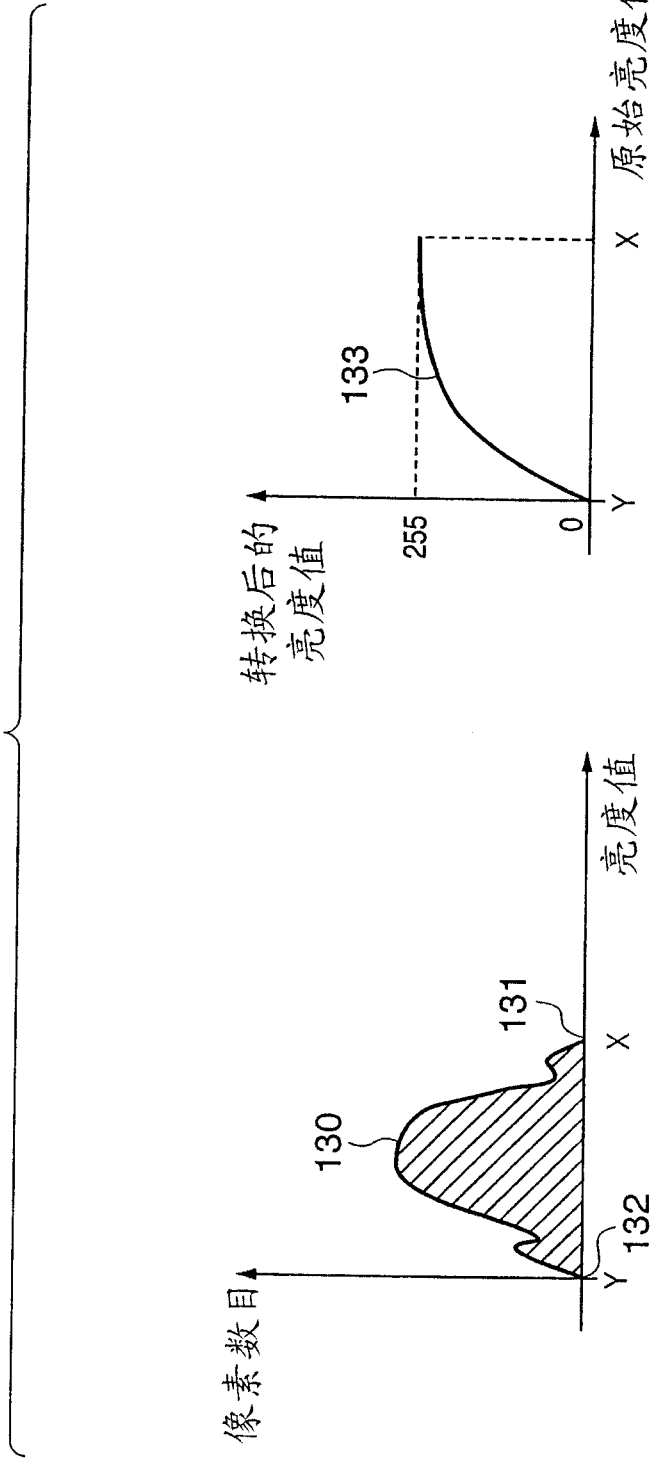


图 3

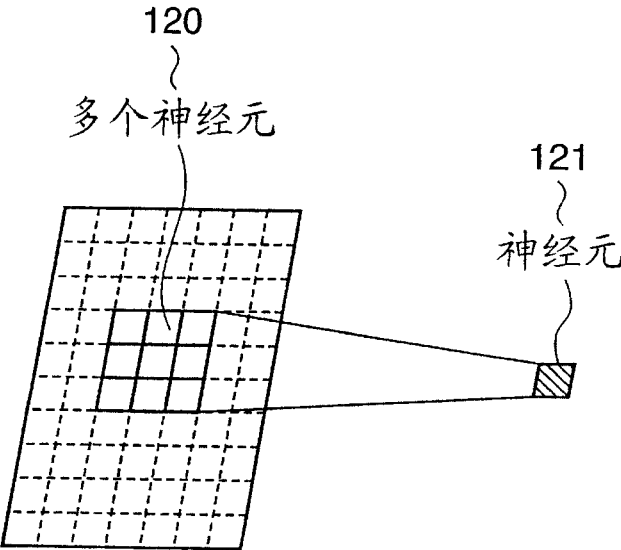


图 4

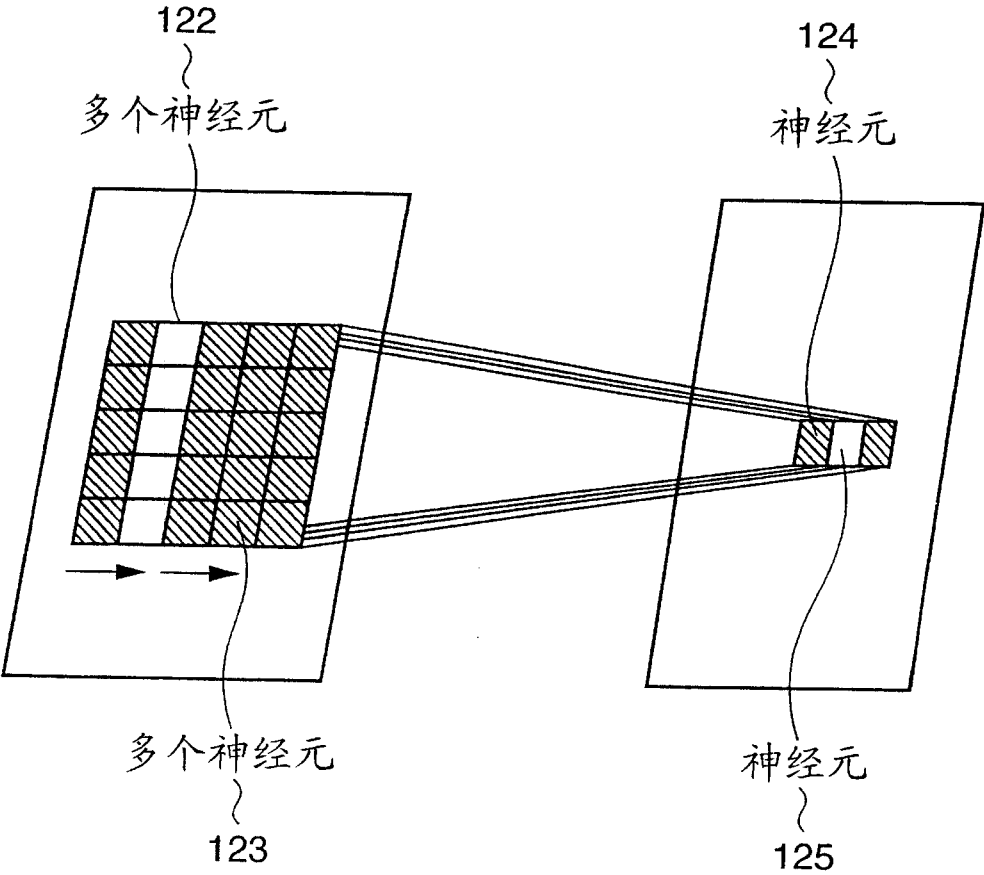


图 5

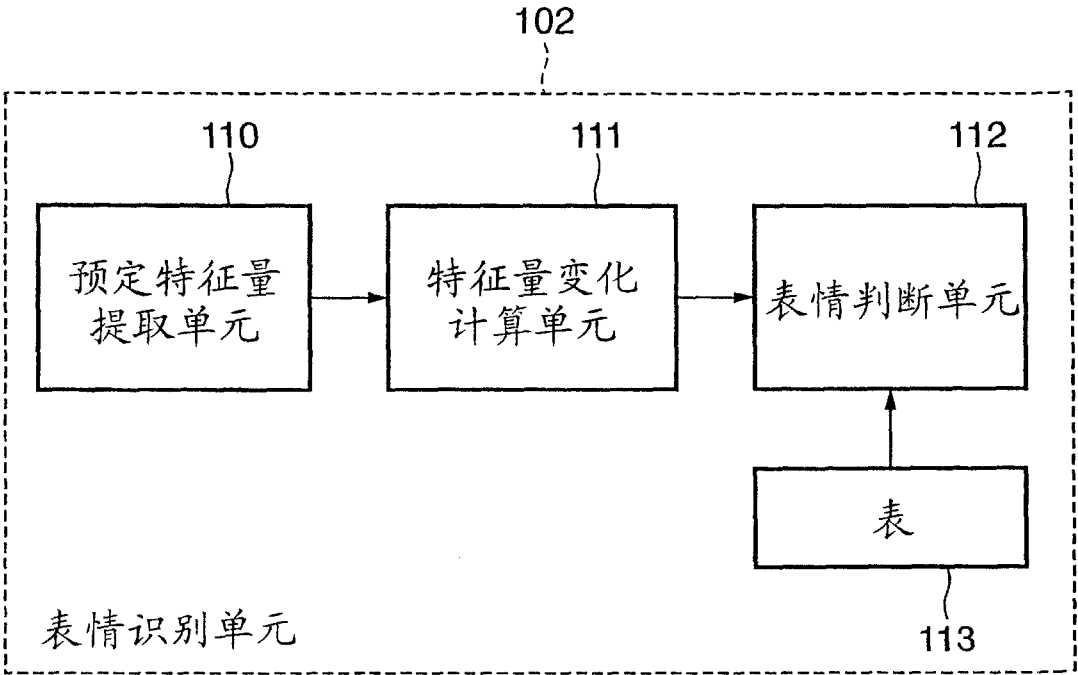


图 6

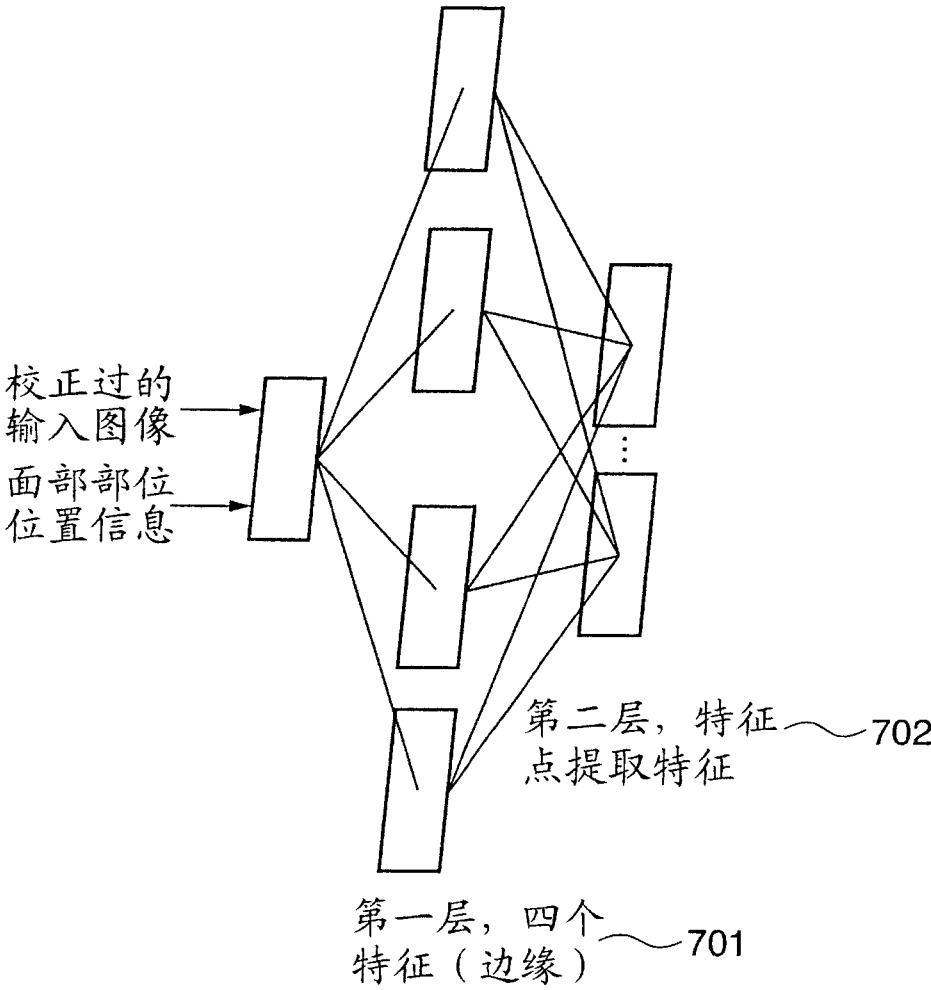


图 7

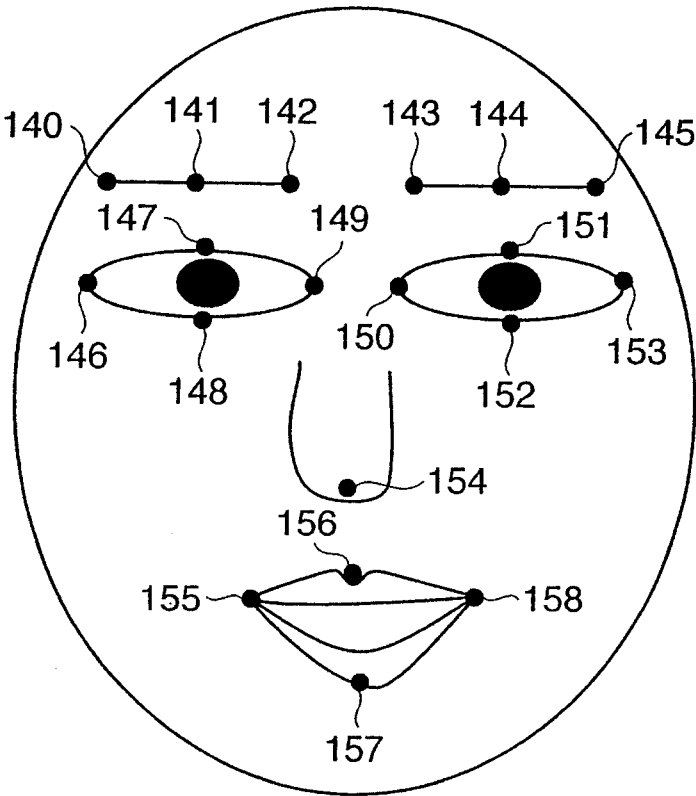


图 8

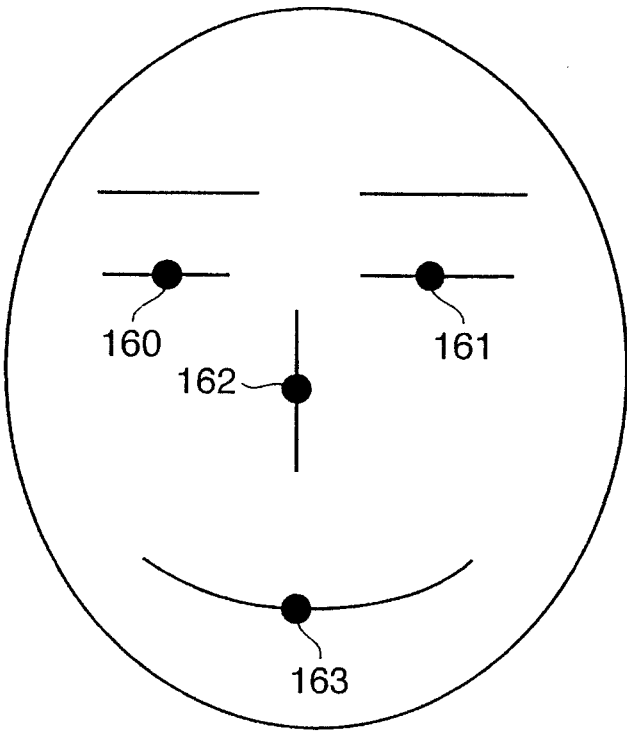


图 9

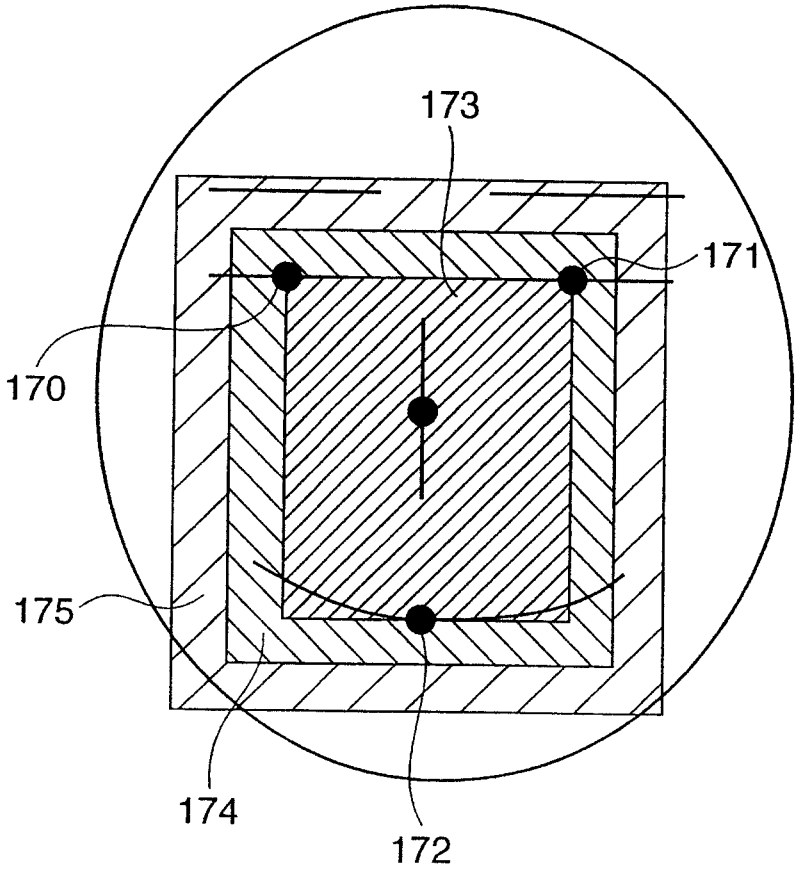


图 10

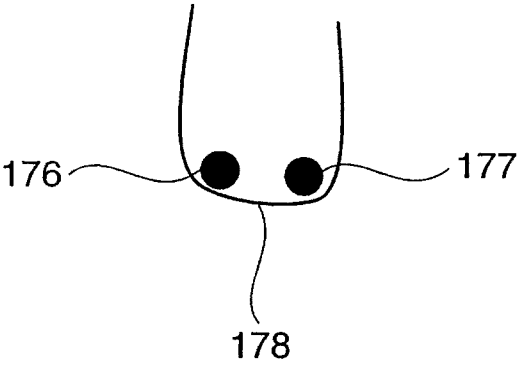


图 11

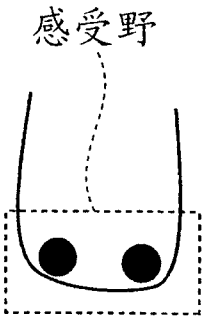


图 12A

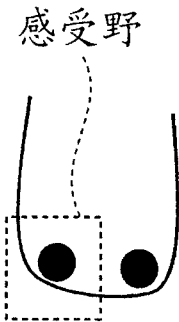


图 12B

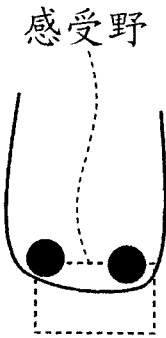


图 12C

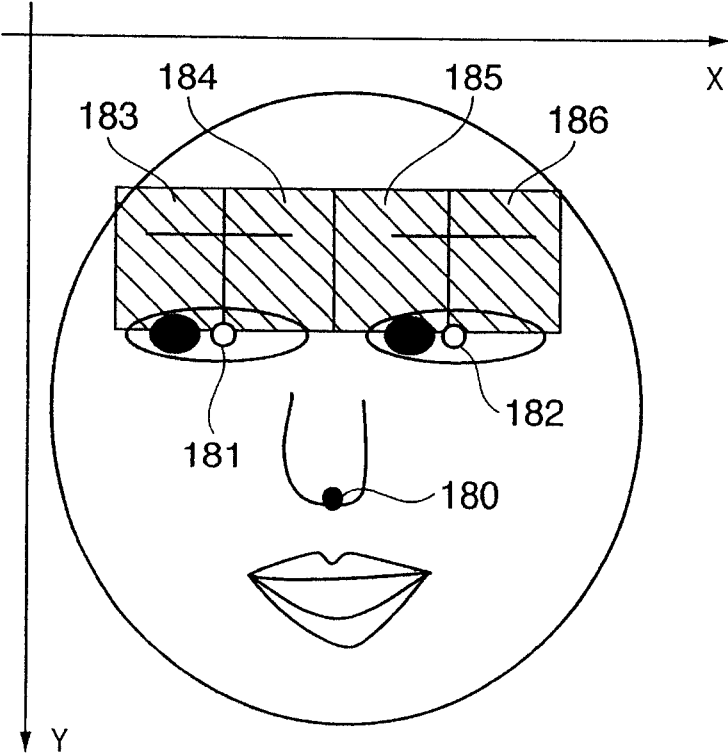


图 13

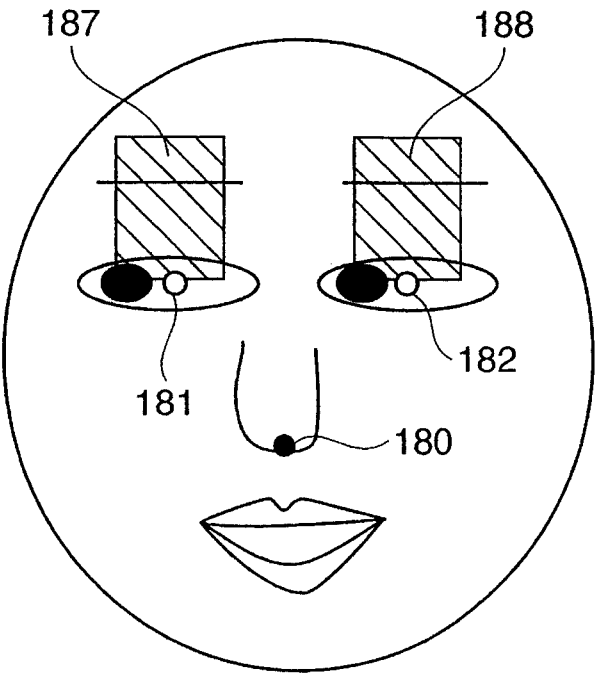


图 14

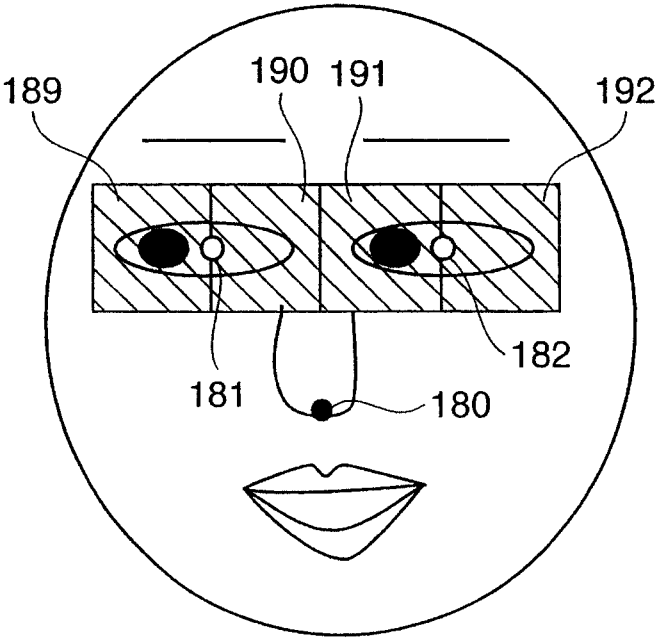


图 15

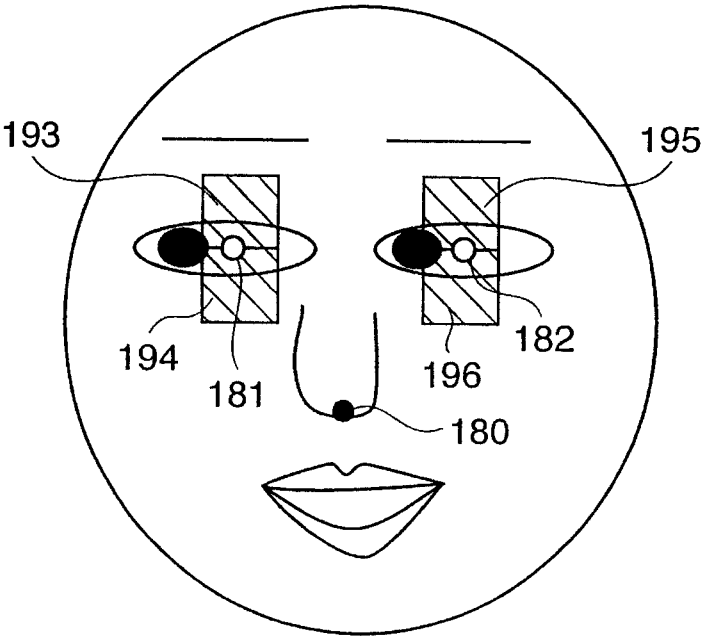


图 16

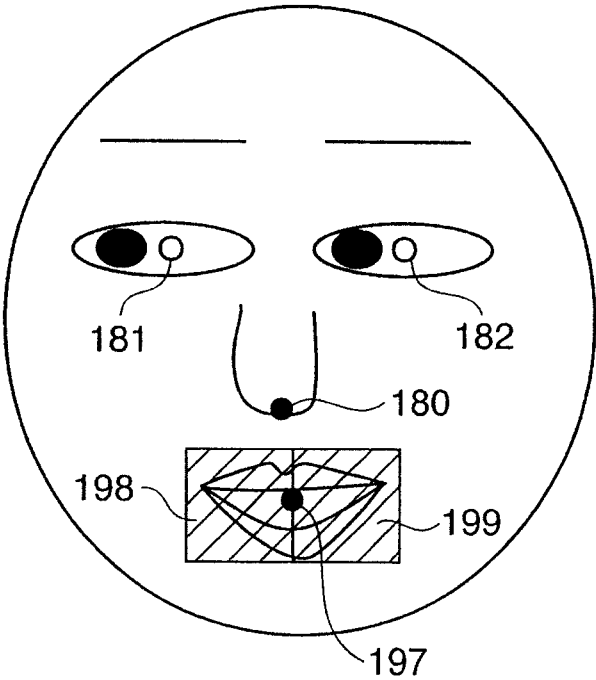


图 17

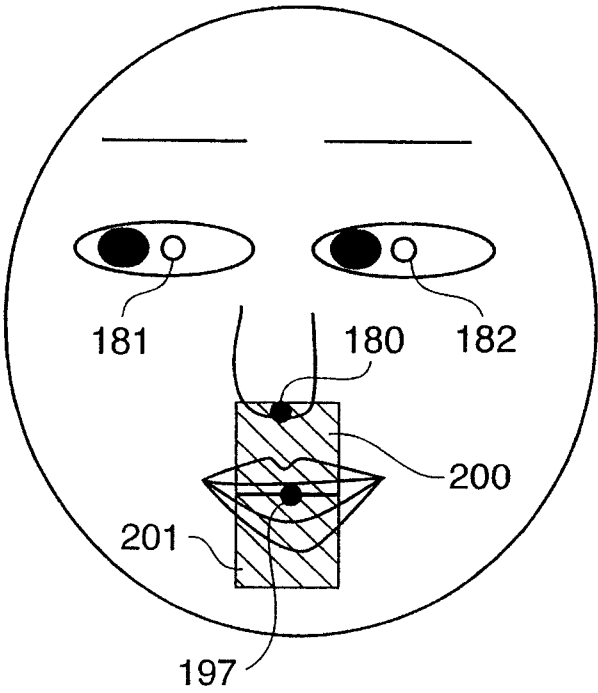


图 18

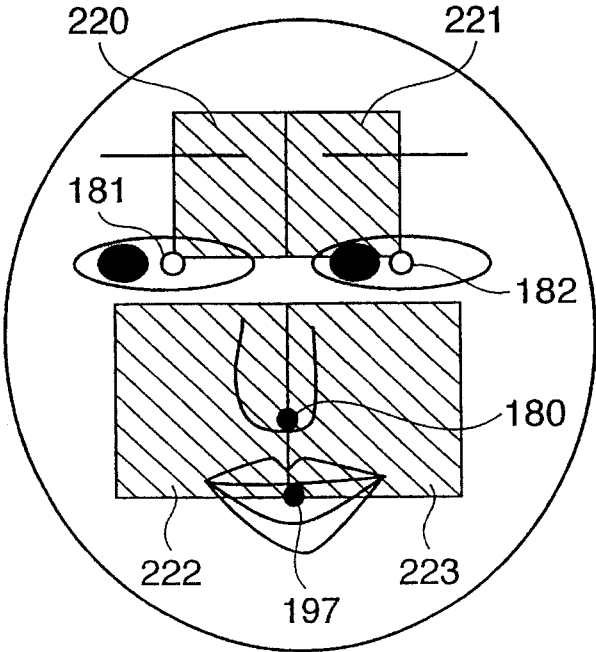


图 19

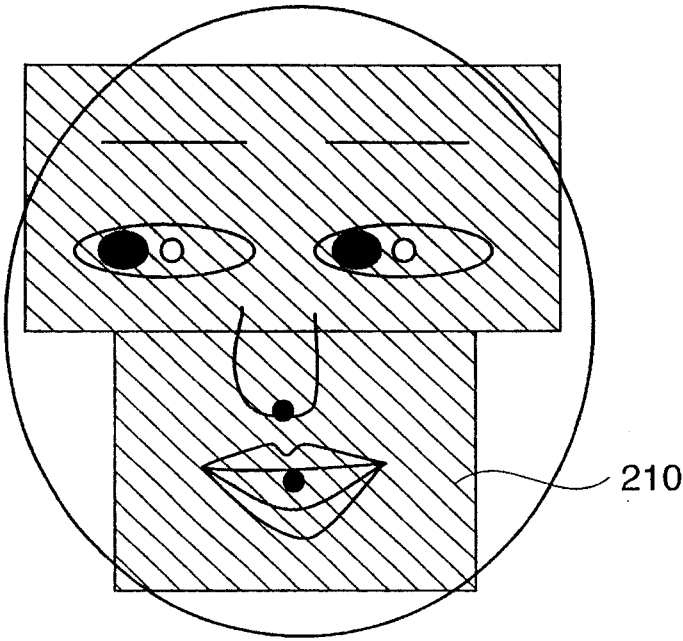


图 20

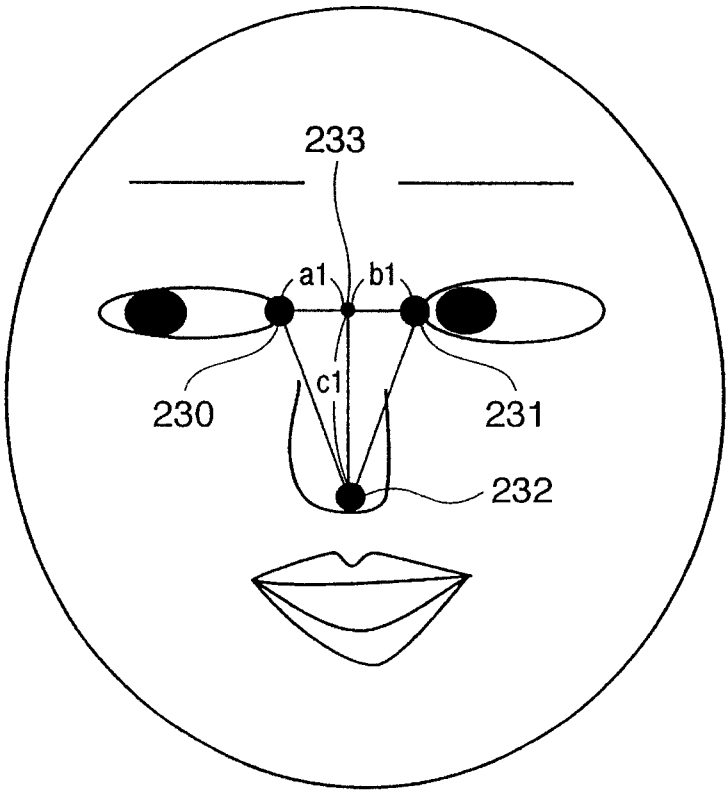


图 21

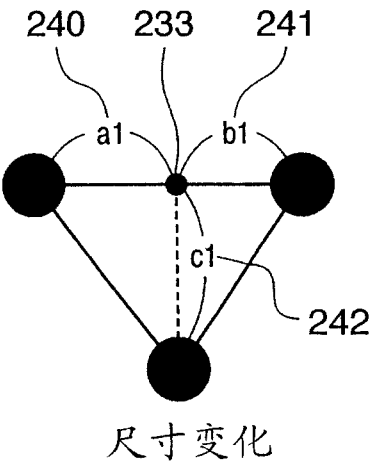


图 22

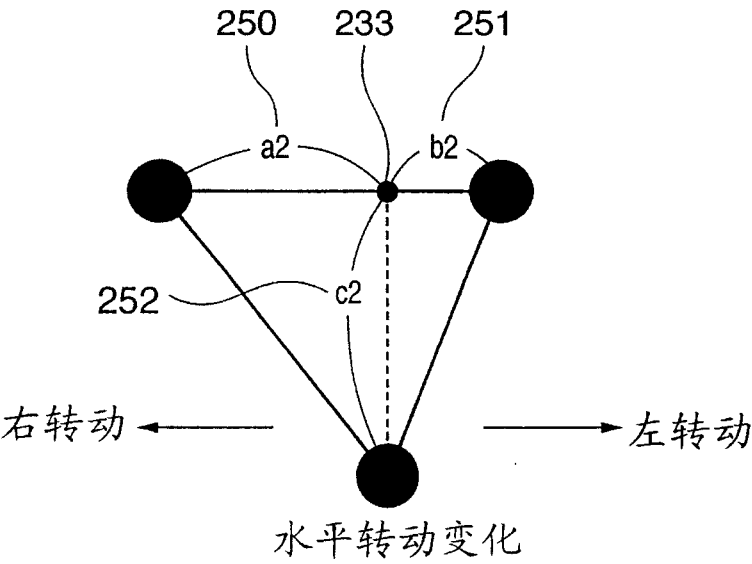


图 23

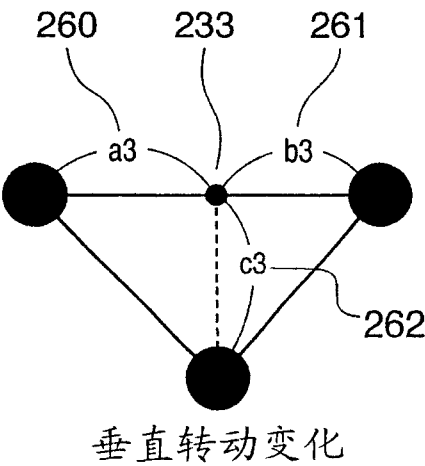


图 24

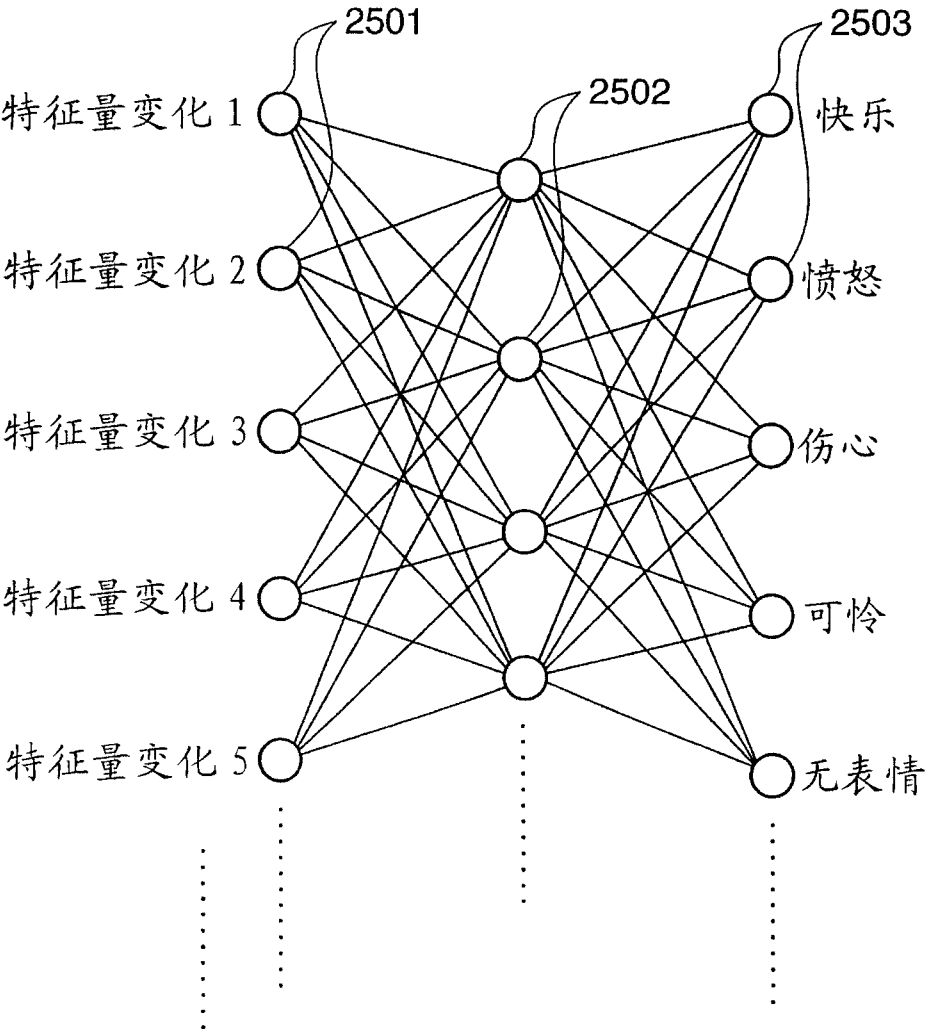


图 25

特征量	权重
眼睛端点和嘴端点之间的距离	大
嘴端点之间的距离	大
嘴区域上下边缘之间的距离	中
脸颊区域的边缘密度	大
前额和眉间区域的边缘密度	小

图 26

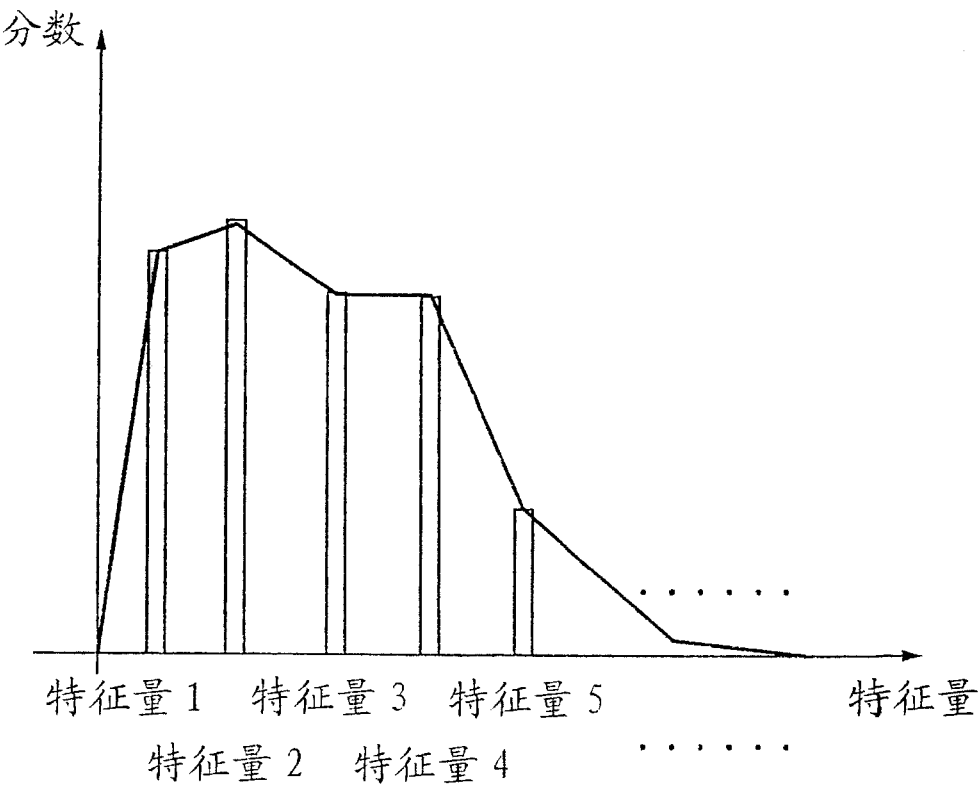


图 27

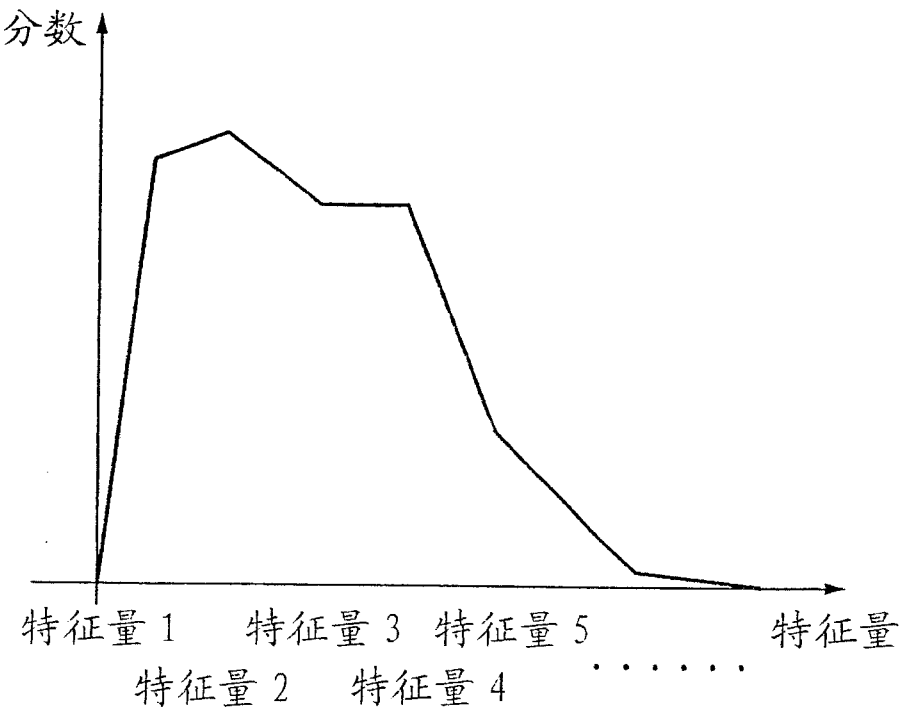


图 28

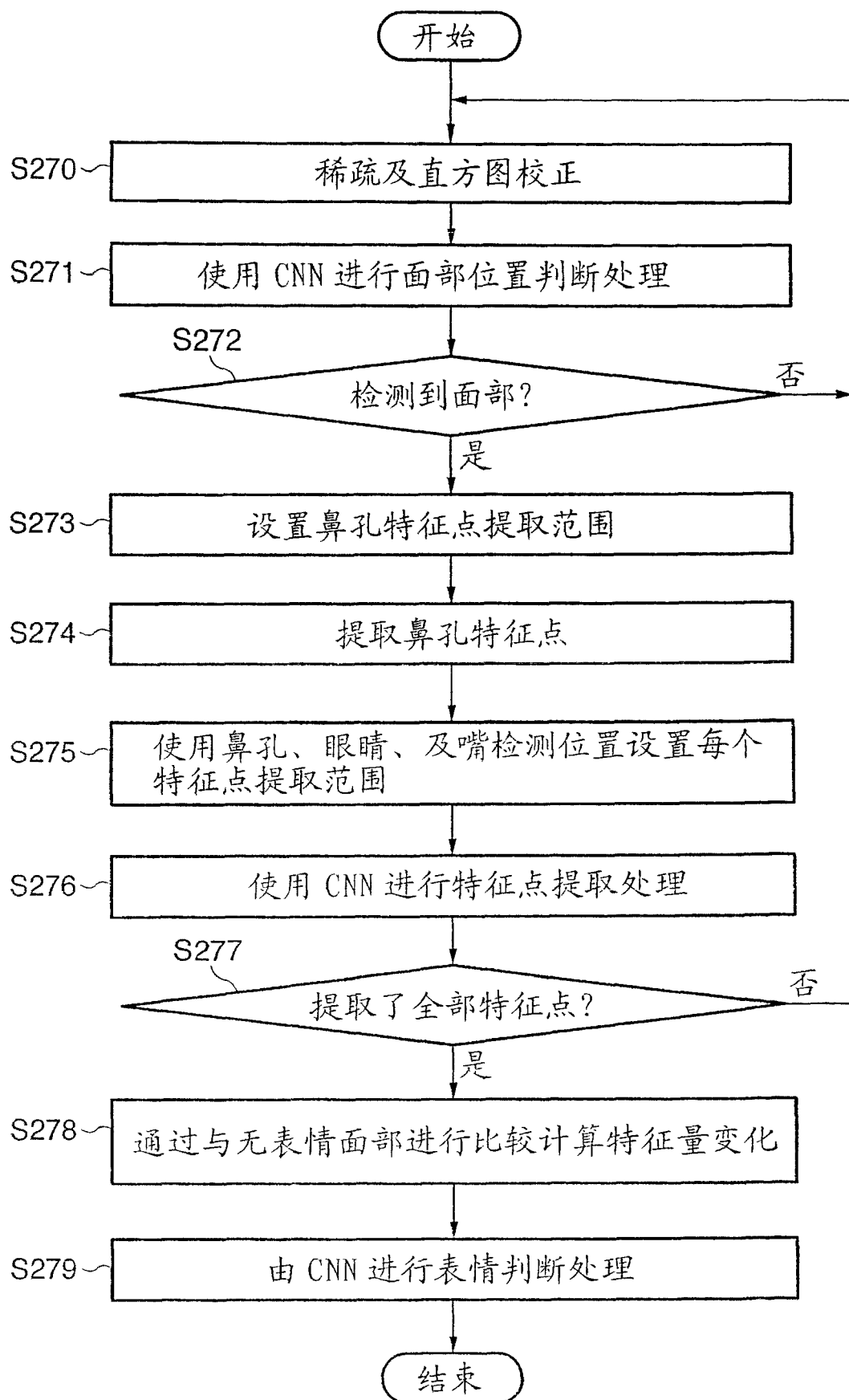


图 29

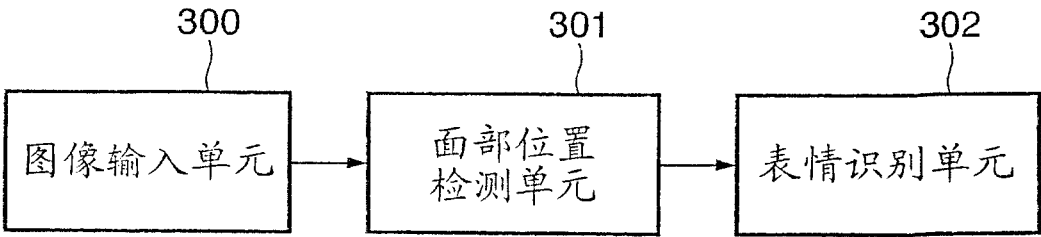


图 30

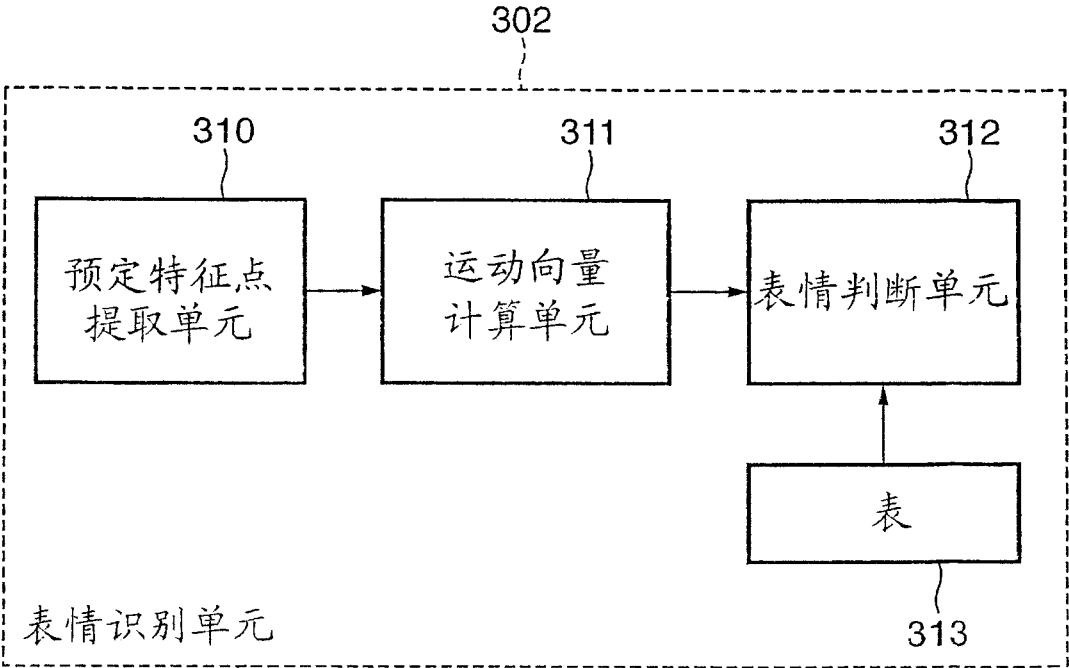


图 31

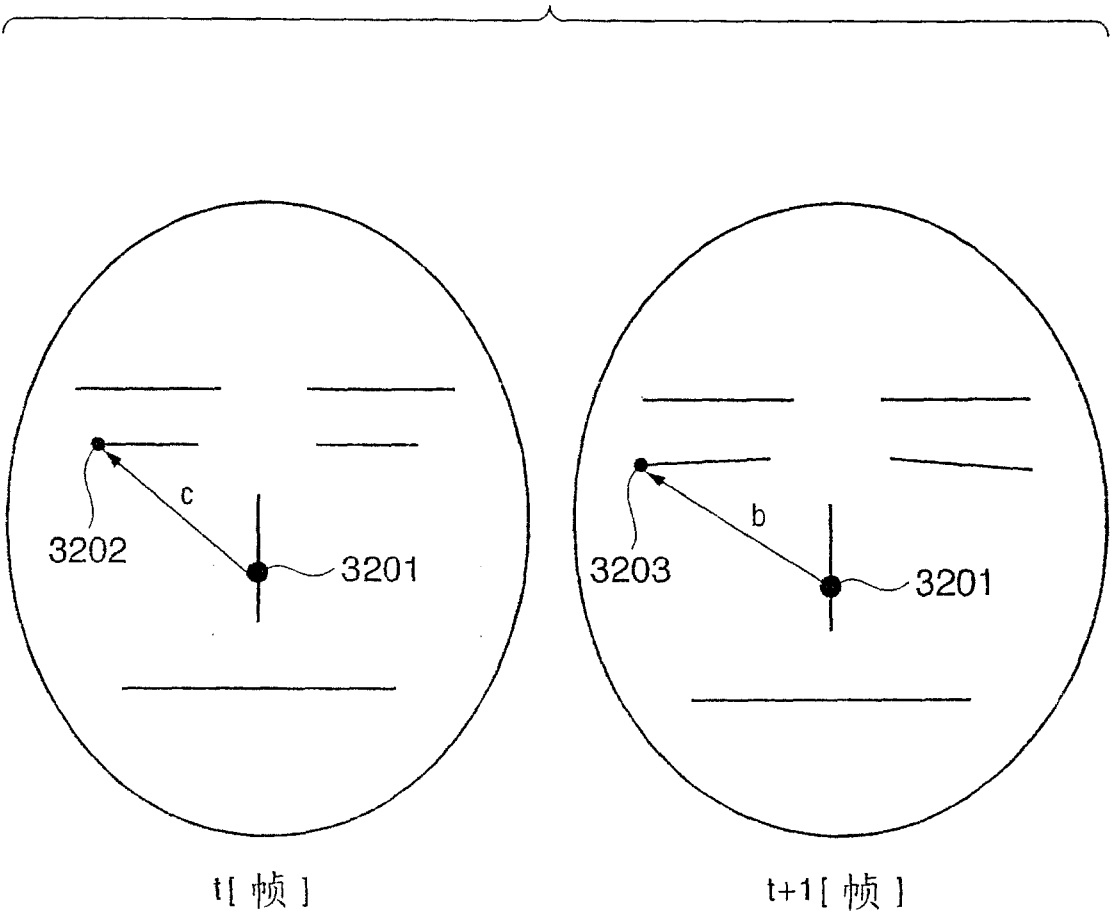


图 32

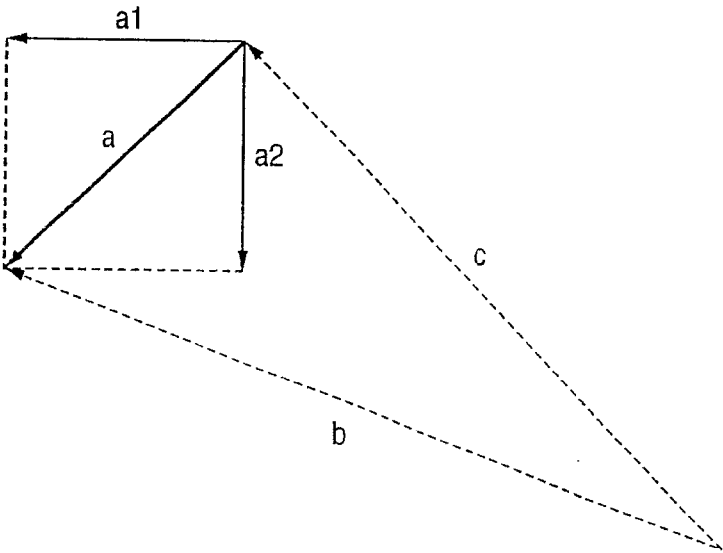


图 33

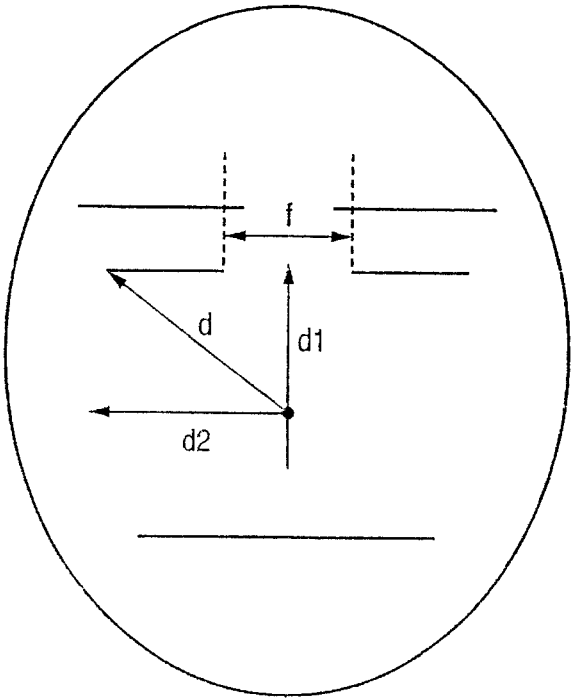


图 34

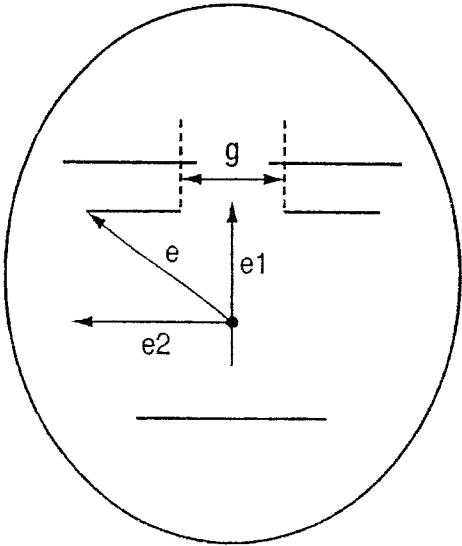


图 35

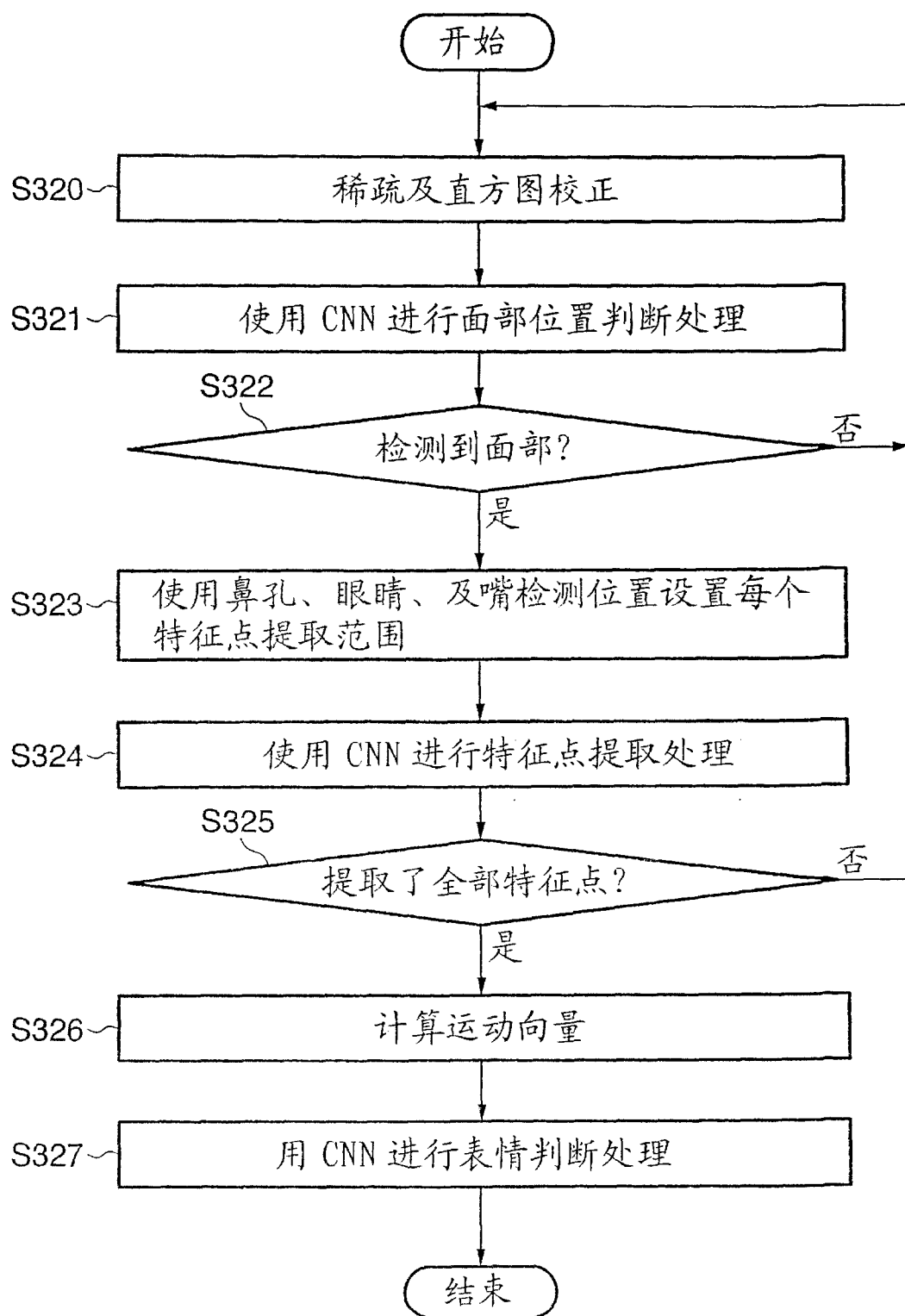


图 36

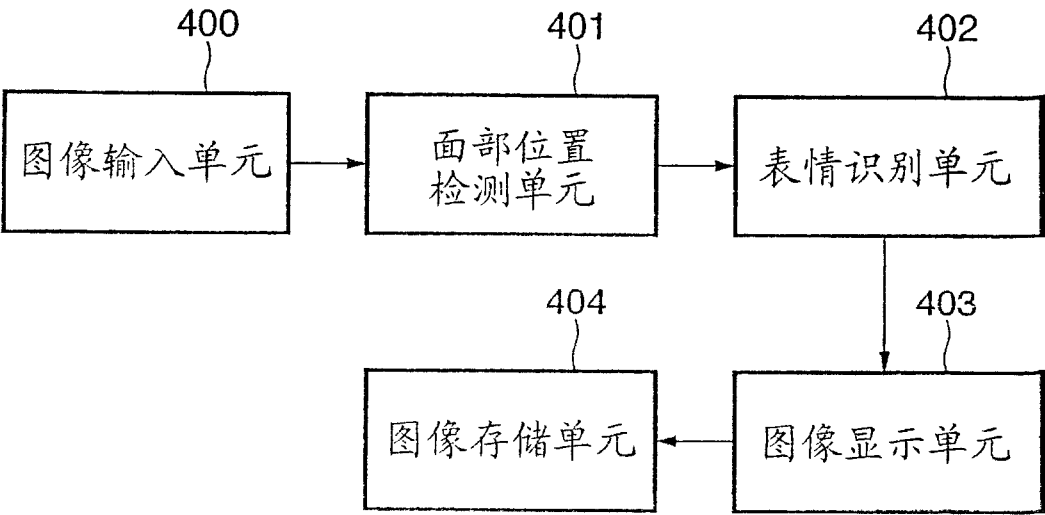


图 37

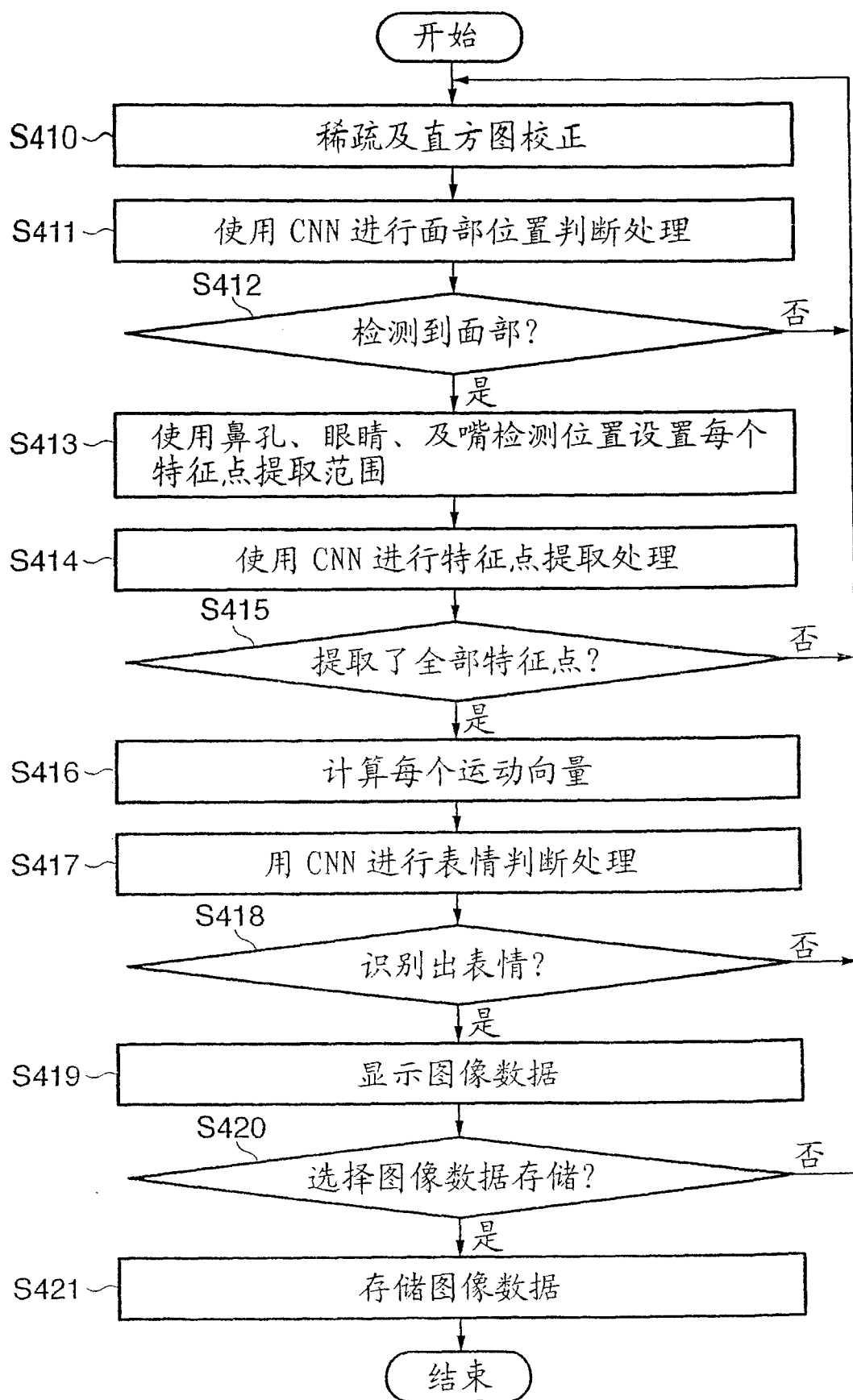


图 38

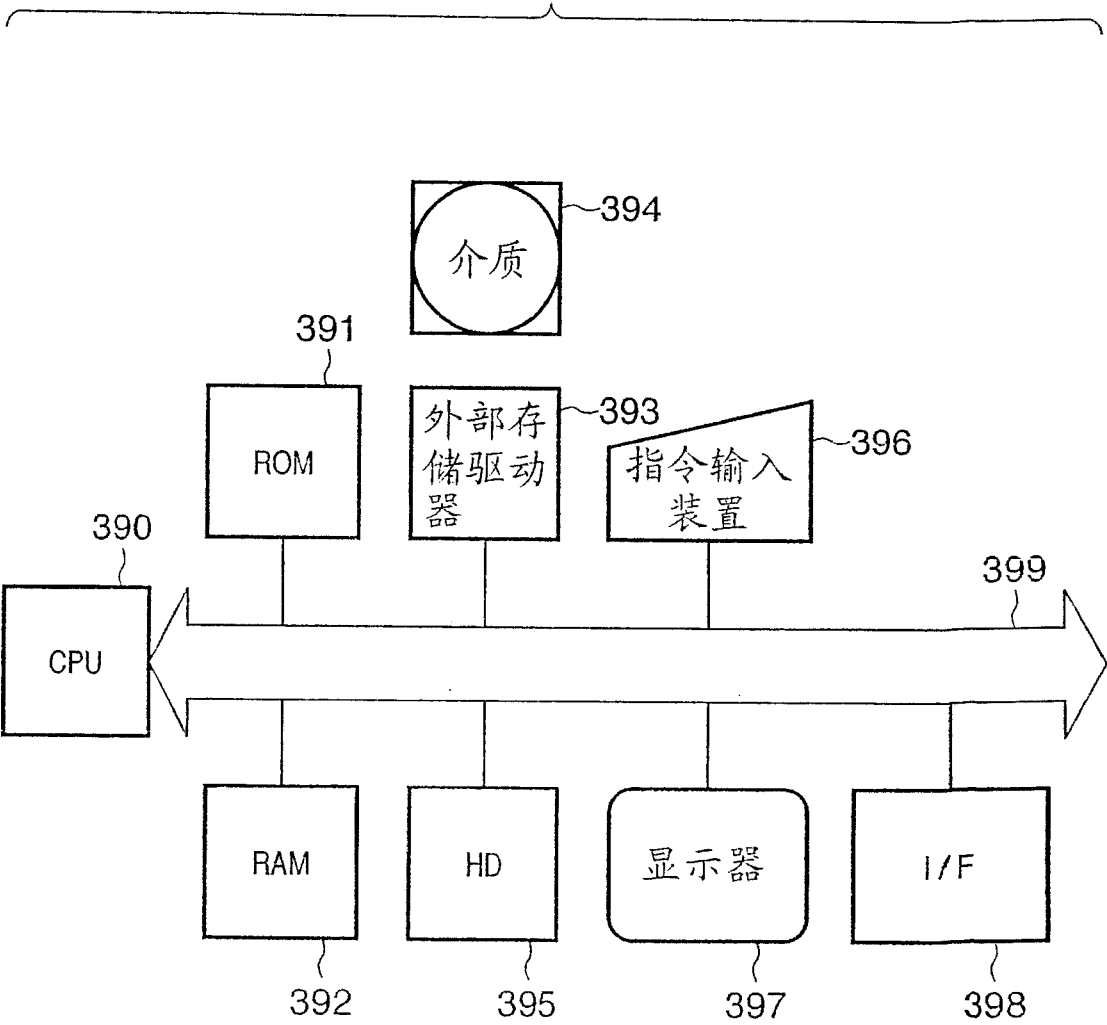


图 39

113

表情	眼睛端点 和嘴端点 之间的距 离	嘴端点之 间的距离	脸颊区域 的边缘密 度	外眼角和 内眼角之 间的距离	-----
快乐	0.7	1.2	1.2	1.1	-----
-----	-----	-----	-----	-----	-----

图 40

313

表情	运动向量 1		运动向量 2			-----
	参数 1	参数 2	-----	参数 1	参数 2	
快乐	增加	减少	-----	增加	增加	-----
愤怒	减少	减少	-----	增加	减少	-----
-----	-----	-----	-----	-----	-----	-----

图 41