



(12) 发明专利

(10) 授权公告号 CN 113033398 B

(45) 授权公告日 2022.02.11

(21) 申请号 202110319068.6

G06N 3/04 (2006.01)

(22) 申请日 2021.03.25

G06N 3/08 (2006.01)

(65) 同一申请的已公布的文献号

G06V 10/774 (2022.01)

申请公布号 CN 113033398 A

G06V 10/764 (2022.01)

(43) 申请公布日 2021.06.25

(56) 对比文件

(73) 专利权人 深圳市康冠商用科技有限公司

CN 112486317 A, 2021.03.12

地址 518000 广东省深圳市龙岗区坂田街

CN 108960163 A, 2018.12.07

道五和大道北4023号1#楼第一层B区

CN 109978838 A, 2019.07.05

和第三、四层

CN 110399847 A, 2019.11.01

US 2015138078 A1, 2015.05.21

(72) 发明人 严海兵 肖伟华 李振乐

EP 3702961 A1, 2020.09.02

(74) 专利代理机构 深圳市精英专利事务所

彭玉青 等. 复杂背景下基于深度学习的手势识别.《机器人》.2019,第41卷(第4期),

44242

Bing Shuai 等.Improving Fully

代理人 武志峰

Convolution Network for Semantic

(51) Int. Cl.

Segmentation.《Computer Science》.2016,

G06V 40/10 (2022.01)

审查员 陈琪蒙

G06F 3/01 (2006.01)

G06K 9/62 (2022.01)

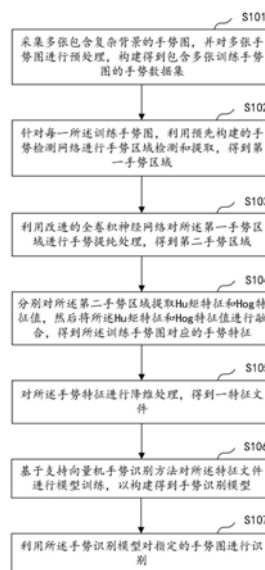
权利要求书3页 说明书15页 附图4页

(54) 发明名称

一种手势识别方法、装置、计算机设备及存储介质

(57) 摘要

本发明公开了一种手势识别方法、装置、计算机设备及存储介质,该方法包括采集构建包含多张训练手势图的手势数据集;利用预先构建的手势检测网络进行手势区域检测和提取,得到第一手势区域;利用改进的全卷积神经网络对第一手势区域进行手势提纯处理,得到第二手势区域;分别对第二手势区域提取Hu矩特征和Hog特征值,然后将Hu矩特征和Hog特征值进行融合,得到训练手势图对应的手势特征;对手势特征进行降维处理,得到一特征文件;基于SVM手势识别方法对特征文件进行模型训练,以构建得到手势识别模型;利用手势识别模型对指定的手势图进行识别。本发明通过融合手势图的Hu矩特征和Hog特征值,可提高对于手势识别的准确度。



1. 一种手势识别方法,其特征在于,包括:

采集多张包含复杂背景的手势图,并对多张手势图进行预处理,构建得到包含多张训练手势图的手势数据集;

针对每一所述训练手势图,利用预先构建的手势检测网络进行手势区域检测和提取,得到第一手势区域;

所述手势检测网络依次包括数据层、第一中间卷积层、第一中间池化层、第二中间卷积层、第二中间池化层、第三中间卷积层、第三中间池化层、第四中间卷积层、第四中间池化层、第五中间卷积层、第五中间池化层、第六中间卷积层、第六中间池化层、第七中间卷积层、第八中间卷积层、第九中间卷积层、第十中间卷积层、输出卷积层,其中,所述数据层的输出为 $320 \times 224 \times 3$,所述输出卷积层的卷积核个数为30个、卷积核大小为 1×1 、步长为1、输出为 $10 \times 7 \times 30$;所述第七中间卷积层、第八中间卷积层、第九中间卷积层和第十中间卷积层的卷积核个数依次为1024、256、512和128,卷积核尺寸依次为 3×3 、 1×1 、 3×3 和 1×1 ;

利用改进的全卷积神经网络对所述第一手势区域进行手势提纯处理,得到第二手势区域;

所述利用改进的全卷积神经网络对所述第一手势区域进行手势提纯处理,得到第二手势区域,包括:

将所述第一手势区域连续三次依次输入改进的全卷积神经网络的第一卷积层、激活层、最大池化层和局部归一化层,并输出第一特征,其中,所述第一卷积层的卷积核大小为 3×3 ,步长为1;

将所述第一特征依次输入至改进的全卷积神经网络的第二卷积层和第三卷积层,并输出第二特征,其中,所述第二卷积层和第三卷积层的卷积核大小为 1×1 ,所述第三卷积层的卷积核个数为2个;

对所述第二特征进行高低维特征融合,然后采用反卷积的方法对所述第二特征进行上采样,得到所述第二手势区域;

分别对所述第二手势区域提取Hu矩特征和Hog特征值,然后将所述Hu矩特征和Hog特征值进行融合,得到所述训练手势图对应的手势特征;

对所述手势特征进行降维处理,得到一特征文件;

基于SVM手势识别方法对所述特征文件进行模型训练,以构建得到手势识别模型;

利用所述手势识别模型对指定的手势图进行识别。

2. 根据权利要求1所述的手势识别方法,其特征在于,所述针对每一所述训练手势图,利用预先构建的手势检测网络对所述训练手势图进行手势区域检测和提取,得到第一手势区域,包括:

将所述训练手势图划分为 10×7 的网格单元,其中每一网格单元的步长为32;

将包含所述训练手势图中手势区域中心位置的网格单元作为检测框;

基于所述检测框,按照下式对所述训练手势图的置信度进行计算:

$$C_b = P(\text{Object}) \cdot \text{IOU}_{\text{pred}}^{\text{truth}}$$

式中, C_b 表示置信度, $P(\text{Object})$ 表示边界框包括手势的概率, $\text{IOU}_{\text{pred}}^{\text{truth}}$ 表示预测框与真实框的交并比,其值介于0到1之间;

基于所述置信度确定所述训练手势图的第一手势区域。

3. 根据权利要求1所述的手势识别方法,其特征在于,所述分别对所述第二手势区域提取Hu矩特征和Hog特征值,然后将所述Hu矩特征和Hog特征值进行融合,得到所述训练手势图对应的手势特征,包括:

按照下式计算所述第二手势区域的Hu矩特征:

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-x_0)^p (y-y_0)^q f(x,y) dx dy$$

式中, $f(x,y)$ 为所述第一手势区域像素点 (x,y) 处的灰度值, $(x-x_0)^p$ 和 $(y-y_0)^q$ 为变核;

对所述第二手势区域进行颜色空间归一化;

按照下式计算归一化后的第二手势区域的梯度值;

$$G_x(x,y) = I(x+1,y) - I(x-1,y)$$

$$G_y(x,y) = I(x,y+1) - I(x,y-1)$$

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2}$$

$$\theta(x,y) = \arctan\left[\frac{G_y(x,y)}{G_x(x,y)}\right]$$

式中, G_x 和 G_y 分别表示所述第二手势区域水平方向和垂直方向的梯度值, $G(x,y)$ 为所述第二手势区域像素点 (x,y) 的梯度幅度, $\theta(x,y)$ 代表所述第二手势区域梯度方向,取值范围为 $[0,\pi)$;

对所述第二手势区域设置梯度方向直方图,然后对所述梯度方向直方图进行归一化处理,并将归一化处理后的梯度方向直方图作为所述Hog特征值;

对所述Hu矩特征和Hog特征值按照1:1的比例进行串联融合,得到融合的手势特征。

4. 根据权利要求1所述的手势识别方法,其特征在于,所述对所述手势特征进行降维处理,得到一特征文件,包括:

按照下式计算所述手势特征的协方差 χ :

$$\chi = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T, \text{其中 } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i;$$

式中, x_i 为第*i*个手势特征的特征向量, N 为手势特征的总个数;

计算所述手势特征对应的协方差矩阵的特征值,并将特征值按照从大到小进行排序,然后选取前*d*个特征值对应的特征向量作为主成分特征向量;

基于投影变换矩阵对所述主成分特征向量进行降维处理,并根据降维处理后的主成分特征向量构建所述特征文件。

5. 根据权利要求4所述的手势识别方法,其特征在于,所述基于SVM手势识别方法对所述特征文件进行模型训练,以构建得到手势识别模型,包括:

通过一对多方法构造SVM多分类器;

对所述特征文件中的主成分特征向量寻找最优分类超平面;

基于所述最优分类超平面获取所述主成分特征向量对应的线性分类器,并通过所述线性分类器输出所述主成分特征向量对应的输出值;

在所述特征文件中,选择输出值最大的主成分特征向量作为所述特征文件的分类结

果,以此构建所述手势识别模型。

6. 一种手势识别装置,其特征在于,包括:

采集单元,用于采集多张包含复杂背景的手势图,并对多张手势图进行预处理,构建得到包含多张训练手势图的手势数据集;

提取单元,用于针对每一所述训练手势图,利用预先构建的手势检测网络进行手势区域检测和提取,得到第一手势区域;

所述手势检测网络依次包括数据层、第一中间卷积层、第一中间池化层、第二中间卷积层、第二中间池化层、第三中间卷积层、第三中间池化层、第四中间卷积层、第四中间池化层、第五中间卷积层、第五中间池化层、第六中间卷积层、第六中间池化层、第七中间卷积层、第八中间卷积层、第九中间卷积层、第十中间卷积层、输出卷积层,其中,所述数据层的输出为 $320 \times 224 \times 3$,所述输出卷积层的卷积核个数为30个、卷积核大小为 1×1 、步长为1、输出为 $10 \times 7 \times 30$;所述第七中间卷积层、第八中间卷积层、第九中间卷积层和第十中间卷积层的卷积核个数依次为1024、256、512和128,卷积核尺寸依次为 3×3 、 1×1 、 3×3 和 1×1 ;

提纯单元,用于利用改进的全卷积神经网络对所述第一手势区域进行手势提纯处理,得到第二手势区域;

所述提纯单元包括:

第一输入单元,用于将所述第一手势区域连续三次依次输入改进的全卷积神经网络的第一卷积层、激活层、最大池化层和局部归一化层,并输出第一特征,其中,所述第一卷积层的卷积核大小为 3×3 ,步长为1;

第二输入单元,用于将所述第一特征依次输入至改进的全卷积神经网络的第二卷积层和第三卷积层,并输出第二特征,其中,所述第二卷积层和第三卷积层的卷积核大小为 1×1 ,所述第三卷积层的卷积核个数为2个;

上采样单元,用于对所述第二特征进行高低维特征融合,然后采用反卷积的方法对所述第二特征进行上采样,得到所述第二手势区域;

第一融合单元,用于分别对所述第二手势区域提取Hu矩特征和Hog特征值,然后将所述Hu矩特征和Hog特征值进行融合,得到所述训练手势图对应的手势特征;

降维单元,用于对所述手势特征进行降维处理,得到一特征文件;

模型训练单元,用于基于SVM手势识别方法对所述特征文件进行模型训练,以构建得到手势识别模型;

识别单元,用于利用所述手势识别模型对指定的手势图进行识别。

7. 一种计算机设备,其特征在于,包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如权利要求1至5任一项所述的手势识别方法。

8. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有计算机程序,所述计算机程序被处理器执行时实现如权利要求1至5任一项所述的手势识别方法。

一种手势识别方法、装置、计算机设备及存储介质

技术领域

[0001] 本发明涉及人机交互技术领域,特别涉及一种手势识别方法、装置、计算机设备及存储介质。

背景技术

[0002] 手势识别作为一种人机交互方式,相比于使用鼠标、键盘,触摸屏等方式有着更加简便、自然、通俗易懂等优势。在PC端,在简单背景和较复杂背景条件下,手势识别的研究已趋于成熟,但在人手与人脸重叠、手势周围存在大面积类肤色、光照不均匀(光线太亮或太暗)等复杂背景下,其检测效果一般,识别率较低。手势识别在安卓系统平台(电子白板、智能电视等)的应用是一个趋势,但由于硬件平台受限,安卓系统平台的运算能力与内存容量相比PC端还是有较大的差距,因此实际硬件的处理能力也是一个需要考虑的重要因素。复杂算法通常有较好的识别率,但复杂算法通常具有复杂度高、运算量大、不满足实时性等缺点,降低用户体验;而简单算法虽然复杂度低、运算量小、可以满足实时性,但简单算法的识别率低,所以需要平衡好算法的复杂性和系统的实时性。

[0003] 手势识别系统主要分为两个部分,第一部分是手势检测与分割,第二部分是手势特征提取及手势识别,其中第一部分是关键,直接影响到第二部分手势识别结果的好坏,特别是在复杂背景条件下。对于手势检测与分割技术,当前主要包括肤色检测与分割、区域对比度、肤色-区域对比度、贝叶斯推理视觉注意力、YOLO算法(一种对象检测算法)、Viola-Jones+Haar(一种人脸检测方法)、Viola-Jones+LBP(一种人脸检测方法)等,而对于手势特征提取及手势识别技术,主要有Hu不变矩+模板匹配、Hog特征+SVM、CNN等。

[0004] 但是上述手势检测与分割技术通常具有一定的缺陷。例如肤色在简单背景中识别率较高,而在复杂环境条件下识别率较差;区域对比度技术在当前手势区域与相邻手势区域的颜色对比度不显著,或者背景中存在与周围区域对比度较大区域时,容易检测不到手势区域,误将背景中的显著区域分割出来;肤色-区域对比度方法是区域对比度方法的改进,该方法虽然引入肤色特征以进一步降低复杂背景中非类肤色区域的干扰,但并没有解决当手势在图中不显著或手势周围存在大面积类肤色区域而无法准确检测到手势并进行正确分割的问题;贝叶斯推理视觉注意力方法通过模仿人类视觉机制提取出特征,然后输入到基于贝叶斯推理的注意力模型中,这样虽然识别率有了较大的提升,但计算时间较长、实时性较差,且由于负责特征提取的注意力模型层数较少、简单,在光线照度不均匀、颜色不一、形状多样、类肤色等复杂环境下所提取的特征较低级,不够全面、丰富,无法有效表达手势信息,最终识别率达不到要求;YOLO算法虽然有优于其它所有目标检测算法的速度,能够满足实时性,但由于其结构复杂,需要借助GPU(图形处理器)才能满足实时性要求,因而导致Android平台下的硬件平台达不到要求;Viola-Jones+Haar和Viola-Jones+LBP虽然计算简单、运算速度快,可以满足实时性要求,并且在目标进行检测时具有较高准确性与稳定性,但在复杂背景下特别是当人脸与手势重合时检测效果差,极易出现手势检测不完整的现象。

[0005] 同样的,当前的手势特征提取及手势识别技术也具有一定的缺陷。例如,Hu不变矩具有平移、旋转、尺度不变性等特性,计算量较小、计算速度快,能够满足实时性要求,对图像中完整的物体外部形状有很好的描述,但对纹理丰富的图像识别较差,不能够很好地描述图像中的细节信息;模板匹配技术在建立标准模板库时需考虑不同背景下的各种形状、距离、角度等因素,导致建模困难且耗时较长;Hog特征广泛应用于手势特征提取中,但只采用Hog特征并不能完全描述好不同的手势,特别是当手势检测不完全、分割不准确的时候;另外,直接使用CNN对整张图或手势检测图进行识别,在复杂背景特别是受类肤色区域干扰条件下效果较差,其主要是因为使用CNN所提取的不同手势的特征差别不大,远没有使用Hu矩特征或者Hog特征所提取的特征可以更好地描述每个不同的手势信息。

发明内容

[0006] 本发明实施例提供了一种手势识别方法、装置、计算机设备及存储介质,旨在提高在复杂背景下手势识别的精度。

[0007] 第一方面,本发明实施例提供了一种手势识别方法,包括:

[0008] 采集多张包含复杂背景的手势图,并对多张手势图进行预处理,构建得到包含多张训练手势图的手势数据集;

[0009] 针对每一所述训练手势图,利用预先构建的手势检测网络进行手势区域检测和提取,得到第一手势区域;

[0010] 利用改进的全卷积神经网络对所述第一手势区域进行手势提纯处理,得到第二手势区域;

[0011] 分别对所述第二手势区域提取Hu矩特征和Hog特征值,然后将所述Hu矩特征和Hog特征值进行融合,得到所述训练手势图对应的手势特征;

[0012] 对所述手势特征进行降维处理,得到一特征文件;

[0013] 基于SVM手势识别方法对所述特征文件进行模型训练,以构建得到手势识别模型;

[0014] 利用所述手势识别模型对指定的手势图进行识别。

[0015] 第二方面,本发明实施例提供了一种手势识别装置,包括:

[0016] 采集单元,用于采集多张包含复杂背景的手势图,并对多张手势图进行预处理,构建得到包含多张训练手势图的手势数据集;

[0017] 提取单元,用于针对每一所述训练手势图,利用预先构建的手势检测网络进行手势区域检测和提取,得到第一手势区域;

[0018] 提纯单元,用于利用改进的全卷积神经网络对所述第一手势区域进行手势提纯处理,得到第二手势区域;

[0019] 第一融合单元,用于分别对所述第二手势区域提取Hu矩特征和Hog特征值,然后将所述Hu矩特征和Hog特征值进行融合,得到所述训练手势图对应的手势特征;

[0020] 降维单元,用于对所述手势特征进行降维处理,得到一特征文件;

[0021] 模型训练单元,用于基于SVM手势识别方法对所述特征文件进行模型训练,以构建得到手势识别模型;

[0022] 识别单元,用于利用所述手势识别模型对指定的手势图进行识别。

[0023] 第三方面,本发明实施例提供了一种计算机设备,包括存储器、处理器及存储在所

述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如第一方面所述的手势识别方法。

[0024] 第四方面,本发明实施例提供了一种计算机可读存储介质,所述计算机可读存储介质上存储有计算机程序,所述计算机程序被处理器执行时实现如第一方面所述的手势识别方法。

[0025] 本发明实施例提供了一种手势识别方法、装置、计算机设备及存储介质,该方法包括采集多张包含复杂背景的手势图,并对多张手势图进行预处理,构建得到包含多张训练手势图的手势数据集;针对每一所述训练手势图,利用预先构建的手势检测网络进行手势区域检测和提取,得到第一手势区域;利用改进的全卷积神经网络对所述第一手势区域进行手势提纯处理,得到第二手势区域;分别对所述第二手势区域提取Hu矩特征和Hog特征值,然后将所述Hu矩特征和Hog特征值进行融合,得到所述训练手势图对应的手势特征;对所述手势特征进行降维处理,得到一特征文件;基于SVM手势识别方法对所述特征文件进行模型训练,以构建得到手势识别模型;利用所述手势识别模型对指定的手势图进行识别。本发明实施例通过预先构建手势检测网络和改进的全卷积神经网络对具有复杂背景的手势图进行特征提取和特征提纯,并通过融合手势图的Hu矩特征和Hog特征值,有效解决单个特征描述手势的缺陷,从而提高对于手势识别的准确度。

附图说明

[0026] 为了更清楚地说明本发明实施例技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0027] 图1为本发明实施例提供了一种手势识别方法的流程示意图;

[0028] 图2为本发明实施例提供了一种手势识别方法的子流程示意图;

[0029] 图3为本发明实施例提供了一种手势识别方法中改进的全卷积神经网络的网络结构示意图;

[0030] 图4为本发明实施例提供了一种手势识别装置的示意性框图;

[0031] 图5为本发明实施例提供了一种手势识别装置的子示意性框图。

具体实施方式

[0032] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0033] 应当理解,当在本说明书和所附权利要求书中使用时,术语“包括”和“包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在,但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0034] 还应当理解,在此本发明说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本发明。如在本发明说明书和所附权利要求书中所使用的那样,除非上下文清楚地指明其它情况,否则单数形式的“一”、“一个”及“该”意在包括复数形式。

[0035] 还应当进一步理解,在本发明说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合,并且包括这些组合。

[0036] 下面请参见图1,图1为本发明实施例提供的一种手势识别方法的流程示意图,具体包括:步骤S101~S107。

[0037] S101、采集多张包含复杂背景的手势图,并对多张手势图进行预处理,构建得到包含多张训练手势图的手势数据集;

[0038] S102、针对每一所述训练手势图,利用预先构建的手势检测网络进行手势区域检测和提取,得到第一手势区域;

[0039] S103、利用改进的全卷积神经网络对所述第一手势区域进行手势提纯处理,得到第二手势区域;

[0040] S104、分别对所述第二手势区域提取Hu矩特征和Hog特征值,然后将所述Hu矩特征和Hog特征值进行融合,得到所述训练手势图对应的手势特征;

[0041] S105、对所述手势特征进行降维处理,得到一特征文件;

[0042] S106、基于SVM手势识别方法对所述特征文件进行模型训练,以构建得到手势识别模型;

[0043] S107、利用所述手势识别模型对指定的手势图进行识别。

[0044] 本实施例中,在对具有复杂背景的手势图进行识别时,首先对每一手势图进行预处理,例如随机裁剪、旋转、翻转等数据增强方式,从而得到所述手势数据集。然后对手势数据集中的训练手势图(即用于模型训练的手势图)进行手势区域的检测和提取,以及手势提纯,以得到具有较少干扰的手势图(即所述第二手势区域)。再分别对所述第二手势区域进行Hu矩特征和Hog特征值提取,并将二者融合得到所述手势特征。接着对所述手势特征进行降维处理,并通过SVM手势识别方法对经过降维处理后的手势特征进行模型训练,以构建所述手势识别模型,从而可以利用所述手势识别模型对指定的手势图进行识别。

[0045] 进一步的,在构建所述手势识别模型后,利用所述手势数据集中的测试手势图(即用于模型测试的手势图)对所述手势识别模型进行测试优化,以提高所述手势识别模型的精度。另外,本实施例所述的复杂背景包括模糊图像、光线强度、类肤色区域干扰等背景,当然,还可以包括其他对于手势识别具有较大影响的背景。

[0046] 本实施例通过预先构建的手势检测网络对训练手势图进行特征提取,可以缩小感受野以及减小卷积步长,进而可以提取到更加丰富的特征,以提高对小目标的检测准确度。同时,通过所述改进的全卷积神经网络将手势从干扰背景(即所述第一手势区域)中提纯出来,使得提纯得到的第二手势区域中具有更少的干扰背景。另外,本实施例将Hu矩特征与Hog特征值相融合,可以有效解决单个特征描述手势的缺陷,把两者的优势结合起来,从而极大的提高对于手势识别的准确度。

[0047] 在一实施例中,所述预先构建的手势数据集为从NUS-II数据集中选取出的部分数据集。NUS-II数据集采集于室内外各种嘈杂环境,其背景复杂多变,手势采集的对象是40个不同种族、不同性别、年龄在22~56岁之间的主体,该数据集考虑了不同的手势尺寸大小(年龄不同或距离不同),并且因考虑的情况比较全面、齐全,例如手势在图像中的位置、光照强度的不同,所以非常适合用来对复杂背景条件下的手势识别进行研究。NUS-II数据集分为两个子集,即subset-A和subset-B,每种数据集有10个不同的手势,其中subset-A是一

般复杂背景数据集,包含各种不同的自然背景,共2000张GRB图片,而subset-B是较复杂背景数据集,包含许多复杂的人脸、手臂、身体等人体背景,共750张GRB图片。为了体现本实施例在复杂背景特别是受类肤色区域干扰下手势识别的优越性,构建了所述手势数据集,所述手势数据集包括从subset-A中选取的650张图像和subset-B的全部图像(750张),其中subset-A中所选取的数据集基本是受类肤色区域干扰的图像,目的是使训练的模型能够更好地检测分割出受人脸、手臂或其它类肤色物体干扰的手势区域,然后对这1400张图像进行随机裁剪至 320×224 ,再经过旋转、翻转等数据增强方式把数据集从1400变成14000。在一具体实施例中,将所述手势数据集划分为训练集和测试集,其中训练集为11200张,测试集为2800张。

[0048] 在一实施例中,所述手势检测网络依次包括数据层、第一中间卷积层、第一中间池化层、第二中间卷积层、第二中间池化层、第三中间卷积层、第三中间池化层、第四中间卷积层、第四中间池化层、第五中间卷积层、第五中间池化层、第六中间卷积层、第六中间池化层、第七中间卷积层、第八中间卷积层、第九中间卷积层、第十中间卷积层、输出卷积层,其中,所述数据层的输出为 $320 \times 224 \times 3$,所述输出卷积层的卷积核个数为30个、卷积核大小为 1×1 、步长为1、输出为 $10 \times 7 \times 30$ 。

[0049] 本实施例中,所述手势检测网络针对现有的YOLO网络结构冗余、单目标(手势)检测效率不高等问题,去除了YOLO网络中一部分卷积层和池化层,并且减少了卷积核个数,以满足检测的实时性要求,进而提高检测效率;同时,针对YOLO网络在检测小目标(例如手势)任务上效果较差的问题,本实施例所述的手势检测网络将第一层卷积核的大小由 7×7 改为 3×3 ,同时将卷积的步长由2改为1,如此,可以通过缩小感受野以及减小卷积步长提取到更加丰富的特征,以提高对小目标的检测准确度。

[0050] 结合表1,所述手势检测网络的输入层为data(数据)层,最后一层输出层添加了 1×1 的卷积层。进一步的,对所述数据层进行去均值处理,使得每个样本(即所述训练手势图)的每一个维度具有零均值和单位方差,其目的有两个,一是消除公共部分的影响,以凸显样本间不同的特征以及差异性;二是避免梯度消失,减少过拟合情况的发生,加快训练/测试迭代速度。所述输出层网络为 1×1 卷积层,如此可以使得 3×3 与 1×1 的卷积核进行交替卷积,在保证特征图大小不变的情况下,可以更加有效地提取特征,增强对小目标(手势)的检测效果。所述手势检测网络的中间层主要由卷积层和池化层组成,与YOLO网络相比,所述手势检测网络减少了卷积核的个数,可以解决YOLO网络的结构冗余、检测效率不高的问题。同时,由于手势是小目标,为了解决YOLO在检测小目标(手势)任务效果较差,容易出现漏检与误检的问题,所述手势检测网络的第一层(即所述第一中间卷积层)的卷积核大小为 3×3 ,然后将整个网络中的卷积步长由2改为1。由表1可以看到,所述手势检测网络的前半部分包括6组相似的部分组成,即每一组包含一个卷积层和一个池化层,其中卷积层主要负责特征提取,其并不改变特征图大小,Padding设置为1主要是防止图像失真,池化层使用的最大池化,经过池化层后的特征图大小减半;所述手势检测网络的后半部分,先是 3×3 与 1×1 进行两轮的交替卷积,在这个过程中保持特征图大小不变,而最后输出特征图大小为 $S1 \times S2 \times X$,其中 $S1 = 10$, $S2 = 7$, X 的计算公式为 $X = (5+N) \times B$,其中 N 代表待检测的目标个数, B 代表要检测的目标边框数,由于本实施例中需要检测的只有人手一个类别,所以 $N = 1$,所述手势检测网络中取 $B = 5$,所以 $X = 30$ 。

[0051] 进一步的,所述手势检测网络的网格单元对应有B(B=5)个检测框,每个检测框对应有5个基本参数和N个(N=1,即手势)类别概率参数。每个网格单元对应一个向量,该向量的维数是X(X=30),1维~6维分别代表第一个检测框的x坐标、y坐标、宽度w、高度h、置信度、检测框所包含物体属于待检测手势的概率,而7维~12维分别代表第二个检测框的x坐标、y坐标、宽度w、高度h、检测框所包含物体属于待检测手势的概率,以此类推。

类型	卷积核	尺寸/步长/Pad	输出	类型	卷积核	尺寸/步长/Pad	输出
data	—	—	320×224×3	conv pool	256	3×3 /1 /1 2×2 /2 /0	20×14×256 10×7×256
conv pool	16	3×3 /1 /1 2×2 /2 /0	320×224×16 160×112×16	conv pool	512	3×3 /1 /1 2×2 /1 /0	10×7×256 10×7×256
conv pool	32	3×3 /1 /1 2×2 /2 /0	160×112×32 80×56×32	conv	1024	3×3 /1 /1	10×7×1024
conv pool	64	3×3 /1 /1 2×2 /2 /0	80×56×64 40×28×64	conv	256	1×1 /1 /0	10×7×256
conv pool	128	3×3 /1 /1 2×2 /2 /0	40×28×128 20×14×128	conv	512	3×3 /1 /1	10×7×512
				conv	128	1×1 /1 /0	10×7×128
				conv	30	1×1 /1 /0	10×7×30

[0052] 表1

[0053] 在一实施例中,所述步骤S102包括:

[0054] 将所述训练手势图划分为10×7的网格单元,其中每一网格单元的步长为32;

[0055] 将包含所述训练手势图中手势区域中心位置的网格单元作为检测框;

[0056] 基于所述检测框,按照下式对所述训练手势图的置信度进行计算:

$$[0057] \quad C_b = P(\text{Object}) \cdot \text{IOU}_{\text{pred}}^{\text{truth}}$$

[0058] 式中, C_b 表示置信度, $P(\text{Object})$ 表示边界框包括手势的概率, $\text{IOU}_{\text{pred}}^{\text{truth}}$ 表示预测框与真实框的交并比,其值介于0到1之间;

[0059] 基于所述置信度确定所述训练手势图的第一手势区域。

[0060] 本实施例中,由表1可知,所述手势检测网络的最后输出为 $S1 \times S2 \times X$,其中 $S1=10$ 、 $S2=7$ 、 $X=30$,且 $X=(5+N) \times B$,其中 $N=1$ 、 $B=5$ 。也就是说,所述手势检测网络最后输出的特征图(即所述第一手势区域)大小为10×7,特征图中的每个单元格对应有一个手势对应的最佳边界框(5个边界框中最佳的一个),但只有一个边界框是用于检测手势的。为了找出用于检测手势区域的网络单元,首先将训练手势图,即大小为320×224的图像分成10×7的网格单元,其中每个网格单元的步长是32。根据算法规则,若手势区域的中心位置坐标落在某个网格单元,那么这个网格单元就负责检测手势,检测框一般包括三个内容:预测的边界框(boundingbox)、边界框物体的类别、属于这个类别的置信度。本实施例由于只对手势进行检测,所以只有手势这个类别。

[0061] 在计算所述训练手势图的置信度时,其中的交并比计算公式如下所示:

$$[0062] \quad \text{IOU}_{\text{pred}}^{\text{truth}} = \frac{B_{\text{truth}} \cap B_{\text{pred}}}{B_{\text{truth}} \cup B_{\text{pred}}}$$

[0063] 另外,所述手势检测网络的网格单元还需为该单元所对应的每个检测框进行类别概率的预测。具体的,令 $P(s/\text{Object})$ 表示检测框包含手势的前提下,手势属于类别s的概

率,则当类别概率与对应检测框相乘,便可得每个检测框属于某个类别的类别置信度。并按照下式计算类别置信度:

$$[0065] \quad C_{s_i} = P(s_i | Object) \cdot P(Object) \cdot IOU_{pred}^{truth} = P(s_i) \cdot IOU_{pred}^{truth}$$

[0066] 式中, s_i 表示检测框的第 i 个类别, $i = 1, 2, \dots, N$ 。由此可知, 类别置信度同时包含了检测框中手势所属类别概率和检测框定位的准确度信息。

[0067] 进一步的, 利用均方误差作为损失函数对所述手势检测网络进行优化调参, 误差计算公式如下所示:

$$[0068] \quad loss = E_c + E_L + E_s$$

[0069] 式中, $loss, E_c, E_L, E_s$ 分别代表总体误差、预测框与真实框之间的坐标误差、预测框与真实框之间的交并比误差、分类误差。更进一步的, E_c, E_L, E_s 的计算公式分别如下所示:

$$[0070] \quad E_c = \lambda_c \sum_{i=0}^{g_w \times g_h} \sum_{j=0}^B \Pi_{ij}^o [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_c \sum_{i=0}^{g_w \times g_h} \sum_{j=0}^B \Pi_{ij}^o [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2]$$

$$[0071] \quad E_L = \sum_{i=0}^{g_w \times g_h} \sum_{j=0}^B \Pi_{ij}^o (C_i - \hat{C}_i)^2 + \lambda_n \sum_{i=0}^{g_w \times g_h} \sum_{j=0}^B \Pi_{ij}^n (C_i - \hat{C}_i)^2$$

$$[0072] \quad E_s = \sum_{i=0}^{g_w \times g_h} \Pi_i^o \sum_{c \in S} (P_i(c) - \hat{P}_i(c))^2$$

[0073] 其中 x, y, w, h, P, C 分别表示预测框中所预测物体的横坐标、纵坐标、宽度、高度、所属类别概率以及对应该置信度, 而 $\hat{x}, \hat{y}, \hat{w}, \hat{h}, \hat{P}, \hat{C}$ 则为真实值, c 代表预测框中待检测手势, s 代表类别, g_w, g_h 分别表示输出特征图的宽和高。 λ_c 表示坐标预测的惩罚系数, λ_n 表示检测到物体边界时的置信度惩罚系数, Π_i^o 表示单元格 i 中是否存在物体, 存在取值为 1, 否则取值为 0, Π_{ij}^o, Π_{ij}^n 分别表示所要检测的手势是否落在或未落入单元格 i 的第 j 个检测框内, 若框内存在所要检测的手势, 则 Π_{ij}^o 为 1, 检测有效, 否则为 0, 检测无效。 Π_{ij}^n 取值则相反。

[0074] 在一实施例中, 结合图 3, 所述步骤 S103 包括:

[0075] 将所述第一手势区域连续三次依次输入改进的全卷积神经网络的第一卷积层、激活层、最大池化层和局部归一化层, 并输出第一特征, 其中, 所述第一卷积层的卷积核大小为 3×3 , 步长为 1;

[0076] 将所述第一特征依次输入至改进的全卷积神经网络的第二卷积层和第三卷积层, 并输出第二特征, 其中, 所述第二卷积层和第三卷积层的卷积核大小为 1×1 , 所述第三卷积层的卷积核个数为 2 个;

[0077] 对所述第二特征进行高低维特征融合, 然后采用反卷积的方法对所述第二特征进行上采样, 得到所述第二手势区域。

[0078] 本实施例中, 通过 IFCN 网络 (即所述的改进的全卷积神经网络, Improved Fully Convolutional Networks) 分割算法对检测后的手势区域图 (即所述第一手势区域) 进行手势提纯。经过所述手势检测网络提取到的第一手势区域, 在复杂背景条件特别是手势周围存在大面积类肤色区域的情况下, 第一手势区域仍很可能存在诸如人脸、脖子、衣服等类肤色背景的干扰。因此本实施例对所述第一手势区域进一步进行提纯, 如此, 与直接将训练手势图直接输入 IFCN 网络中进行手势分割相比, 有两大好处: 一是效率更高, 能满足实时性;

二是准确度更容易得到保证,泛化能力更强。具体来说,由于所述手势检测网络的检测速度快,且可以满足硬件平台的实时性要求(即不需要GPU支持),同时相比于原图来说,经过检测后的第一手势区域的区域面积大幅度减小,这使得IFCN网络的结构不需要太复杂便可以达到比较好的效果,可在一定程度地避免过拟合现象,提高泛化能力,并且所需时间更少,效果更高。另外,相比于输入原图,因所述第一手势区域通过检测阶段排除了很多干扰的背景因素,进行手势提纯将更加容易,准确性也相应会更高。

[0079] 由图3可知,IFCN网络主要分为三部分,第一部分是由3组conv+relu+MaxPool+LRN(局部响应归一化)的结构组成,由于输入为所述第一手势区域,所以为了适应小尺寸输入的特征提取,所述IFCN网络的第一部分的卷积核的大小设置为 3×3 ,步长设置为1;relu为激活函数,用于增加网络模型的非线性能力以完成复杂任务;池化层使用最大池化(即MaxPool)以保留显著性特征,从而避免平均池化所带来的模糊化效果,设置池化步长小于池化核大小,可以使池化输出之间有重叠,进而增加了所提取特征的丰富性,同时减少了信息的丢失;LRN主要是对图像进行局部归一化。

[0080] 所述IFCN网络的第二部分包括两个 1×1 的卷积层(即所述第二卷积层和第三卷积层),分别用于升维和降维,即输出特征图大小不变,通道数增加或减少。第二部分通过将全连接层转化成卷积层,以及将输出由一维向量变成一个热图(高维特征图)类型的图像,可以避免空间信息的丢失,为图像的分割提供必要前提。因为第三部分进行手势分割是二值分割,即分类时只分手势区域与非手势区域,故将所述第三卷积层的卷积核的个数设置为2。

[0081] 在所述IFCN网络的第三部分中,由于经过多次conv+pool操作之后,图像分辨率越来越低,尺寸也越来越小,特征也越来越抽象。虽然抽象特征有利于图像分类,但因丢失轮廓、纹理等细节严重而不利于图像分割,为了对输入图像中的手势进行有效的分割提纯,IFCN网络采用deconvolution(反卷积)的方法进行upsampling(上采样),使上采样后的特征图大小与网络输入图像大小相同,最后对上采样后的特征图进行逐像素分类。本实施例能够克服传统CNN图像分割算法存储开销大、计算效率低的缺点,从图像级的分类进一步拓展到像素级别的分类。

[0082] 进一步的,若只对所述IFCN网络的最后一层输出进行反卷积(例如8倍放大)至输入原图大小,得到的分割结果可能并不是特别准确,有些细节可能没有恢复出来,因此,为了得到更加准确的分割效果,本实施例在采用反卷积方法之前,先采用高低维特征相融合的方法对所述第二特征进行融合处理。具体来说,对最后一层(第8层)输出进行步长为2的上采样,得到特征图A,对第4层输出使用 1×1 的卷积核进行卷积实现降维,得到特征图B,接着分别对特征图A和特征图B上的每个像素使用softmax分类函数进行二分类(手势or非手势),选取概率大的那个为所属类别,得到对应的手势分割图A和分割图B,由于特征图A与特征图B大小相同,所以二者对应的分割结果图大小也相同,然后这两个分类图求和相加得到分割结果图C,最后利用反卷积技术实现步长为4的上采样,得到和输入图像尺寸大小相同的手势分割结果,相当于对分割图C直接放大4倍。

[0083] 在一实施例中,如图2所示,所述步骤S104包括:步骤S201~S205。

[0084] S201、按照下式计算所述第二手势区域的Hu矩特征:

$$[0085] \quad \mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-x_0)^p (y-y_0)^q f(x,y) dx dy$$

[0086] 式中, $f(x,y)$ 为所述第一手势区域像素点 (x,y) 处的灰度值, $(x-x_0)^p$ 和 $(y-y_0)^q$ 为变核;

[0087] S202、对所述第二手势区域进行颜色空间归一化;

[0088] S203、按照下式计算归一化后的第二手势区域的梯度值;

$$[0089] \quad G_x(x,y) = I(x+1,y) - I(x-1,y)$$

$$[0090] \quad G_y(x,y) = I(x,y+1) - I(x,y-1)$$

$$[0091] \quad G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2}$$

$$[0092] \quad \theta(x,y) = \arctan\left[\frac{G_y(x,y)}{G_x(x,y)}\right]$$

[0093] 式中, G_x 和 G_y 分别表示所述第二手势区域水平方向和垂直方向的梯度值, $G(x,y)$ 为所述第二手势区域像素点 (x,y) 的梯度幅度, $\theta(x,y)$ 代表所述第二手势区域梯度方向, 取值范围为 $[0, \pi)$;

[0094] S204、对所述第二手势区域设置梯度方向直方图, 然后对所述梯度方向直方图进行归一化处理, 并将归一化处理后的梯度方向直方图作为所述Hog特征值;

[0095] S205、对所述Hu矩特征和Hog特征值按照1:1的比例进行串联融合, 得到融合的手势特征。

[0096] 本实施例中, 在得到提纯后的第二手势区域后, 需要对第二手势区域提取手势特征, 若手势所提取的特征太少, 则不能很好地识别手势, 若手势所提取的特征太多, 则会造成识别速度太慢, 不能满足实时性要求。因此, 本实施例在提取手势特征的过程中遵循以下四个原则:

[0097] ①对于不同样本, 具有很好的区分性。

[0098] ②对于同一样本, 特征值应该近似相等, 即具有很好的可靠性。

[0099] ③在保证识别率的情况下, 特征数量应尽可能少, 以保证系统实时性。

[0100] ④特征间相对独立, 关联性小。

[0101] 手势轮廓是描述手势形状最有效的方式之一, 计算简单高效且关联性不大, 能够很好地描述手势, 可作为手势识别的有效特征。其中Hu不变矩 (即所述Hu矩特征) 作为描述轮廓特征的一种经典特征提取方法, 因其计算量小、速度快, 能够满足实时性要求, 而且具有平移、旋转和尺度不变性的图像特征, 有较好的识别率, 但其对小的和纹理较丰富的手势很难识别, 因为Hu矩特征只用到低阶矩, 不能有效全面地描述手势的特征, 同时, 虽然Hu矩特征对 (手势) 形状有很好的描述, 但前提是轮廓 (边缘) 信息要较为完整和准确, 但在复杂背景特别是类肤色干扰下并不能得到很好的保证。Hog (方向梯度直方图) 作为图像的三个特征之一, 使用梯度或边缘方向的分布对局部的物体外观或形状进行描述, 能够很好地描述局部形状信息, 本质上是计算梯度的统计信息, 而梯度主要存在于物体边缘, 由于不同手势轮廓差别较大, 所以采用Hog特征识别手势可以达到不错的效果, 但Hog也存在弊端, 首先, 因Hog没有选择主方向, 因而没有旋转不变性, 其次, Hog不具备尺度不变性, 其尺度不变性只能通过缩放检测窗口大小实现, 最后, Hog对噪声很敏感。

[0102] 本实施例为了能够提取更加丰富的手势特征以提高手势识别率,对Hu矩特征与Hog特征值进行融合,将两者的优点相结合。具体的,首先对待提取手势的第二手势区域进行归一化处理,接着分别提取Hu矩特征与Hog特征值,然后进行融合。在一具体实施中,Hu矩特征是7维向量,Hog特征是2268维向量,对两种特征进行1:1的比例进行串联融合,那么便得到2275维的融合特征。

[0103] 在提取Hu矩特征过程中,首先按照下式计算矩:

$$[0104] \quad m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy$$

[0105] 式中, $f(x, y)$ 为概率密度分布函数,在灰度图像中指的是像素点 (x, y) 处的灰度值, x^p 和 y^q 是变核。

[0106] 通过上式可以得到零阶矩 m_{00} , 一阶矩 m_{01} 和 m_{10} , 进而得到质心 (x_0, y_0) 的坐标为:

$$[0107] \quad x_0 = \frac{m_{10}}{m_{00}}, \quad y_0 = \frac{m_{01}}{m_{00}}$$

[0108] 接着,按照下式计算中心矩:

$$[0109] \quad \mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-x_0)^p (y-y_0)^q f(x, y) dx dy$$

[0110] 于是,可以所述中心矩得到Hu矩,即:

$$[0111] \quad h_1 = \mu_{20} + \mu_{02}$$

$$[0112] \quad h_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2$$

$$[0113] \quad h_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2$$

$$[0114] \quad h_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2$$

$$[0115] \quad h_5 = (\mu_{03} - 3\mu_{12}) (\mu_{30} + \mu_{12}) [(\mu_{30} + 3\mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03}) (\mu_{21} + \mu_{03}) [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$$

$$[0116] \quad h_6 = (\mu_{20} - \mu_{02}) [(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11} (\mu_{30} + \mu_{12}) (\mu_{21} + \mu_{03})$$

$$[0117] \quad h_7 = (3\mu_{21} - \mu_{03}) (\mu_{03} + \mu_{12}) [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{12} - \mu_{30}) (\mu_{21} + \mu_{03}) [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]。$$

[0118] 在提取Hog特征值过程中,首先对所述第二手势区域进行颜色空间归一化,可有效减少光线(如光线太暗或太亮)的干扰。颜色空间归一化具体包括图像的灰度化与Gamma矫正(伽马矫正,一种图像处理方法)。其中,按照下式对所述第二手势区域进行图像灰度化:

$$[0119] \quad \text{Gray} = 0.3 \times R + 0.59 \times G + 0.11 \times B$$

[0120] Gamma矫正定义如下式所示,其中 γ 取0.5。

$$[0121] \quad Y(x, y) = I(x, y)^\gamma$$

[0122] 考虑到输入原图大小为 320×224 , 以及所述第二手势区域的实际大小,本实施例将手势提取区域大小统一归一化为 64×80 , 并且每 8×8 个像素组成一个 cell, 每 2×2 个 cell 组成一个 block, 以8个像素为步长,那么水平方向有7个扫描窗口,垂直方向有9个扫描窗口,取每个 cell 对应的直方图组数为9,那么对于 64×80 的手势图片可得到 $7 \times 9 \times 4 \times 9 = 2268$ 维的特征向量。

[0123] 由于一个 block 由 2×2 个 cell 组成,每个 cell 是 8×8 像素的,且步长也为8个像素,所以 block 之间会有重叠部分。在实际应用中,光线等因素可能会导致某些区域像素的梯度变化较为激烈,从而使得不同 block 之间特征差距较大不利于最终的分类,因此,本实施例

按照下式对重叠块内的梯度直方图进行归一化以减少光照、边缘对比度所带来的影响：

$$[0124] \quad L2 - Norm : v \rightarrow \frac{v}{\sqrt{\|v\|_2^2 + \varepsilon^2}}$$

[0125] 式中, v 为手势特征的特征向量, ε 为归一化系数, $\varepsilon = 0.25$ 。

[0126] 在一实施例中, 所述步骤S105包括：

[0127] 按照下式计算所述手势特征的协方差 χ ：

$$[0128] \quad \chi = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T, \text{ 其中 } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i ;$$

[0129] 式中, x_i 为第 i 个手势特征的特征向量, N 为手势特征的总个数；

[0130] 计算所述手势特征对应的协方差矩阵的特征值, 并将特征值按照从大到小进行排序, 然后选取前 d 个特征值对应的特征向量作为主成分特征向量；

[0131] 基于投影变换矩阵对所述主成分特征向量进行降维处理, 并根据降维处理后的主成分特征向量构建所述特征文件。

[0132] 在融合Hu矩特征与Hog特征值的过程中, 可能会存在两个问题: 一是特征维数过高导致计算时长增加, 进而导致系统实时性得不到保证, 即不满足特征提取的第三个原则 (③在保证识别率的情况下, 特征数量应尽可能少, 以保证系统实时性); 二是融合的手势特征可能会存在较多的冗余信息, 而冗余信息会使识别结果变差。

[0133] 针对上述问题, 本实施例采用PCA算法 (主成分分析算法) 对融合的收拾特征进行特征降维。具体来说, 首先对手势特征进行归一化处理, 使手势特征的特征分量落在 $[0, 1]$ 之间, 然后采用PCA算法对2275维的手势特征的特征向量进行降维处理, 把降维后的特征向量最后处理成一个特征文件, 作为下一步骤中分类器的输入。本实施例在保证信息完整性的前提下, 尽可能缩短计算时长, 综合权衡实时性与信息完整性后, 将2275维的融合特征 (即所述手势特征) 降为75维。

[0134] 在基于投影变换矩阵对所述主成分特征向量进行降维处理时, 按照下式计算得到主成分特征向量：

$$[0135] \quad U = (u_1, u_2, \dots, u_i, \dots, u_d) \text{ 其中 } u_i = (u_{i1}, u_{i2}, \dots, u_{im})$$

$$[0136] \quad y_i = U^T x_i$$

[0137] 式中, U 为投影变换矩阵, y_i 为主成分特征向量。

[0138] 在一实施例中, 所述步骤S106包括：

[0139] 通过一对多方法构造SVM多分类器；

[0140] 对所述特征文件中的主成分特征向量寻找最优分类超平面；

[0141] 基于所述最优分类超平面获取所述主成分特征向量对应的线性分类器, 并通过所述线性分类器输出所述主成分特征向量对应的输出值；

[0142] 在所述特征文件中, 选择输出值最大的主成分特征向量作为所述特征文件的分类结果, 以此构建所述手势识别模型。

[0143] 由于SVM (支持向量机) 手势识别方法在解决小样本、非线性以及高维特征数据中具有比较好的优势, 并且SVM手势识别方法泛化能力强, 在优化目标过程中可以使结构风险最小, 使其在手势识别技术领域取得不错的效果。因此本实施例采用SVM手势识别方法对所

述特征文件进行模型训练。具体过程如下所示：

[0144] 使用一对多 (one-versus-rest, 简称OVR SVMs) 的方法构造出SVM多分类器；

[0145] OVR方法本质上是通过组合多个二分类器进而实现多分类器的功能。若手势有C个类,那么训练集对应的特征向量需分成C组,每组选择其中一个类对应的特性向量作为正集,剩余C-1个类对应的特征向量作为负集,每组相当于训练一个二分类的SVM。

[0146] 对于第c ($c=1, 2, \dots, C$) 组训练集对应的特征向量,寻找最优分类超平面；

[0147] SVM手势识别方法实现二分类的基本原理是能够找到使分类间隔最大的最优超平面。当线性可分时,选择线性核函数linear直接构造线性分类器;当线性不可分时,选择高斯核函数rbf (径向基函数),将样本映射到高维空间,然后在高维空间中构造线性分类器。考虑到使用OVR模式训练数据的不平衡性,在训练过程中会选择距离分类面较近的少量样本作为训练样本,而这些样本包含大部分分类所需的信息,且数量较少,对它们所对应的特征子集的分类等同于对所有样本的分类,靠近分类面的特征子集称为支持向量 (Support SV),此外,解决样本不均衡的另外一种方式是对样本点较少的正集采用较大的惩罚因子C。

[0148] 若支持向量集对应的训练样本为 $x = \{x_1, x_2, \dots, x_N\}$,对应的样本标签为 $\{y_1, y_2, \dots, y_N\}$, $y_k \in \{C_{-1}, C_1\}$,每个样本 $x_k = (x_k^1, x_k^2, \dots, x_k^d)$, $k=1, 2, \dots, N$.输入的特征向量就是上述经过PCA降维后的主成分特征向量,维数为d。当 $x_k \in C_{-1}$ 时, $y_k = -1$;当 $x_k \in C_1$ 时, $y_k = 1$ 。于是,样本集 $x = \{x_1, x_2, \dots, x_N\}$ 对应的超平面可表示为:

$$[0149] \quad w^T x + w_0 = 0$$

[0150] 其中 $w = (w_1, w_2, \dots, w_d)^T$,此处目的是找到上述超平面公式中所表示的最优超平面,即两类最靠近的样本距离达到最大时的超平面。实际上,上式问题可转化为求解拉格朗日对偶问题,实践表明,通过朗格朗日对偶,总可以找到这个最优超平面,进而得到最优解 w^* 和 w_0 ,记所有支持向量的集合为S,且支持向量对应的样本数记为 N_S ,那么,最优解为:

$$[0151] \quad w^* = \sum_{x_k \in S} \lambda_k y_k x_k$$

$$[0152] \quad w_0 = \frac{1}{N_S} \sum_{x_k \in S} y_k - \frac{1}{N_S} (w^*)^T \sum_{x_k \in S} x_k$$

[0153] 式中, λ_k 为拉格朗日乘子,也叫对偶变量,且 $\lambda_k \geq 0$,可以通过SMO算法 (Sequential Minimal Optimization,用于针对求解SVM问题的Lagrange对偶问题) 求解得到。求解出 λ_k 之后,通过上述最优解公式便可求出 w^* 和 w_0 ,于是得到最终的最优超平面为:

$$[0154] \quad (w^*)^T x + w_0 = 0.$$

[0155] 通过上述得到的最优超平面对每组特征向量对应的线性分类器进行求解训练。训练完毕之后,得到C个训练结果文件。进一步的,使用预先设置的验证集进行结果验证,即把验证集对应的特征向量输入这C个结果文件中,得到对应的C个输出值,取C个值中最大的作为该验证样本的分类结果。最后使用交叉验证法 (Cross-Validation) 以及网格搜索法 (GridSearchCV) 进行参数调节,得到最优的所述手势识别模型。

[0156] 图4为本发明实施例提供的一种手势识别装置400的示意性框图,该装置400包括:

[0157] 采集单元401,用于采集多张包含复杂背景的手势图,并对多张手势图进行预处理,构建得到包含多张训练手势图的手势数据集;

[0158] 提取单元402,用于针对每一所述训练手势图,利用预先构建的手势检测网络进行

手势区域检测和提取,得到第一手势区域;

[0159] 提纯单元403,用于利用改进的全卷积神经网络对所述第一手势区域进行手势提纯处理,得到第二手势区域;

[0160] 第一融合单元404,用于分别对所述第二手势区域提取Hu矩特征和Hog特征值,然后将所述Hu矩特征和Hog特征值进行融合,得到所述训练手势图对应的手势特征;

[0161] 降维单元405,用于对所述手势特征进行降维处理,得到一特征文件;

[0162] 模型训练单元406,用于基于SVM手势识别方法对所述特征文件进行模型训练,以构建得到手势识别模型;

[0163] 识别单元407,用于利用所述手势识别模型对指定的手势图进行识别。

[0164] 在一实施例中,所述手势检测网络依次包括数据层、第一中间卷积层、第一中间池化层、第二中间卷积层、第二中间池化层、第三中间卷积层、第三中间池化层、第四中间卷积层、第四中间池化层、第五中间卷积层、第五中间池化层、第六中间卷积层、第六中间池化层、第七中间卷积层、第八中间卷积层、第九中间卷积层、第十中间卷积层、输出卷积层,其中,所述数据层的输出为 $320 \times 224 \times 3$,所述输出卷积层的卷积核个数为30个、卷积核大小为 1×1 、步长为1、输出为 $10 \times 7 \times 30$ 。

[0165] 在一实施例中,所述提取单元402包括:

[0166] 划分单元,用于将所述训练手势图划分为 10×7 的网格单元,其中每一网格单元的步长为32;

[0167] 检测框设置单元,用于将包含所述训练手势图中手势区域中心位置的网格单元作为检测框;

[0168] 置信度计算单元,用于基于所述检测框,按照下式对所述训练手势图的置信度进行计算:

$$[0169] \quad C_b = P(\text{Object}) \cdot \text{IOU}_{\text{pred}}^{\text{truth}}$$

[0170] 式中, C_b 表示置信度, $P(\text{Object})$ 表示边界框包括手势的概率, $\text{IOU}_{\text{pred}}^{\text{truth}}$ 表示预测框与真实框的交并比,其值介于0到1之间;

[0171] 第一确定单元,用于基于所述置信度确定所述训练手势图的第一手势区域。

[0172] 在一实施例中,所述提纯单元403包括:

[0173] 第一输入单元,用于将所述第一手势区域连续三次依次输入改进的全卷积神经网络的第一卷积层、激活层、最大池化层和局部归一化层,并输出第一特征,其中,所述第一卷积层的卷积核大小为 3×3 ,步长为1;

[0174] 第二输入单元,用于将所述第一特征依次输入至改进的全卷积神经网络的第二卷积层和第三卷积层,并输出第二特征,其中,所述第二卷积层和第三卷积层的卷积核大小为 1×1 ,所述第三卷积层的卷积核个数为2个;

[0175] 上采样单元,用于对所述第二特征进行高低维特征融合,然后采用反卷积的方法对所述第二特征进行上采样,得到所述第二手势区域。

[0176] 在一实施例中,如图5所示,所述第一融合单元404包括:

[0177] Hu矩特征计算单元501,用于按照下式计算所述第二手势区域的Hu矩特征:

$$[0178] \quad \mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-x_0)^p (y-y_0)^q f(x,y) dx dy$$

[0179] 式中, $f(x, y)$ 为所述第一手势区域像素点 (x, y) 处的灰度值, $(x-x_0)^p$ 和 $(y-y_0)^q$ 为变核;

[0180] 归一化单元502, 用于对所述第二手势区域进行颜色空间归一化;

[0181] 梯度值计算单元503, 用于按照下式计算归一化后的第二手势区域的梯度值;

$$[0182] \quad G_x(x, y) = I(x+1, y) - I(x-1, y)$$

$$[0183] \quad G_y(x, y) = I(x, y+1) - I(x, y-1)$$

$$[0184] \quad G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$$

$$[0185] \quad \theta(x, y) = \arctan\left[\frac{G_y(x, y)}{G_x(x, y)}\right]$$

[0186] 式中, G_x 和 G_y 分别表示所述第二手势区域水平方向和垂直方向的梯度值, $G(x, y)$ 为所述第二手势区域像素点 (x, y) 的梯度幅度, $\theta(x, y)$ 代表所述第二手势区域梯度方向, 取值范围为 $[0, \pi)$;

[0187] 直方图设置单元504, 用于对所述第二手势区域设置梯度方向直方图, 然后对所述梯度方向直方图进行归一化处理, 并将归一化处理后的梯度方向直方图作为所述Hog特征值;

[0188] 第二融合单元505, 用于对所述Hu矩特征和Hog特征值按照1:1的比例进行串联融合, 得到融合的手势特征。

[0189] 在一实施例中, 所述降维单元405包括:

[0190] 协方差计算单元, 用于按照下式计算所述手势特征的协方差 χ :

$$[0191] \quad \chi = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T, \text{ 其中 } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i;$$

[0192] 式中, x_i 为第 i 个手势特征的特征向量, N 为手势特征的总个数;

[0193] 特征值选择单元, 用于计算所述手势特征对应的协方差矩阵的特征值, 并将特征值按照从大到小进行排序, 然后选取前 d 个特征值对应的特征向量作为主成分特征向量;

[0194] 特征文件构建单元, 用于基于投影变换矩阵对所述主成分特征向量进行降维处理, 并根据降维处理后的主成分特征向量构建所述特征文件。

[0195] 在一实施例中, 所述模型训练单元406包括:

[0196] 构造单元, 用于通过一对多方法构造SVM多分类器;

[0197] 寻找单元, 用于对所述特征文件中的主成分特征向量寻找最优分类超平面;

[0198] 获取单元, 用于基于所述最优分类超平面获取所述主成分特征向量对应的线性分类器, 并通过所述线性分类器输出所述主成分特征向量对应的输出值;

[0199] 分类结果选择单元, 用于在所述特征文件中, 选择输出值最大的主成分特征向量作为所述特征文件的分类结果, 以此构建所述手势识别模型。

[0200] 由于装置部分的实施例与方法部分的实施例相互对应, 因此装置部分的实施例请参见方法部分的实施例的描述, 这里暂不赘述。

[0201] 本发明实施例还提供了一种计算机可读存储介质, 其上存有计算机程序, 该计算机程序被执行时可以实现上述实施例所提供的步骤。该存储介质可以包括: U盘、移动硬盘、只读存储器 (Read-Only Memory, ROM)、随机存取存储器 (Random Access Memory, RAM)、磁碟

或者光盘等各种可以存储程序代码的介质。

[0202] 本发明实施例还提供了一种计算机设备,可以包括存储器和处理器,存储器中存有计算机程序,处理器调用存储器中的计算机程序时,可以实现上述实施例所提供的步骤。当然计算机设备还可以包括各种网络接口,电源等组件。

[0203] 说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。应当指出,对于本技术领域的普通技术人员来说,在不脱离本申请原理的前提下,还可以对本申请进行若干改进和修饰,这些改进和修饰也落入本申请权利要求的保护范围内。

[0204] 还需要说明的是,在本说明书中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的状况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

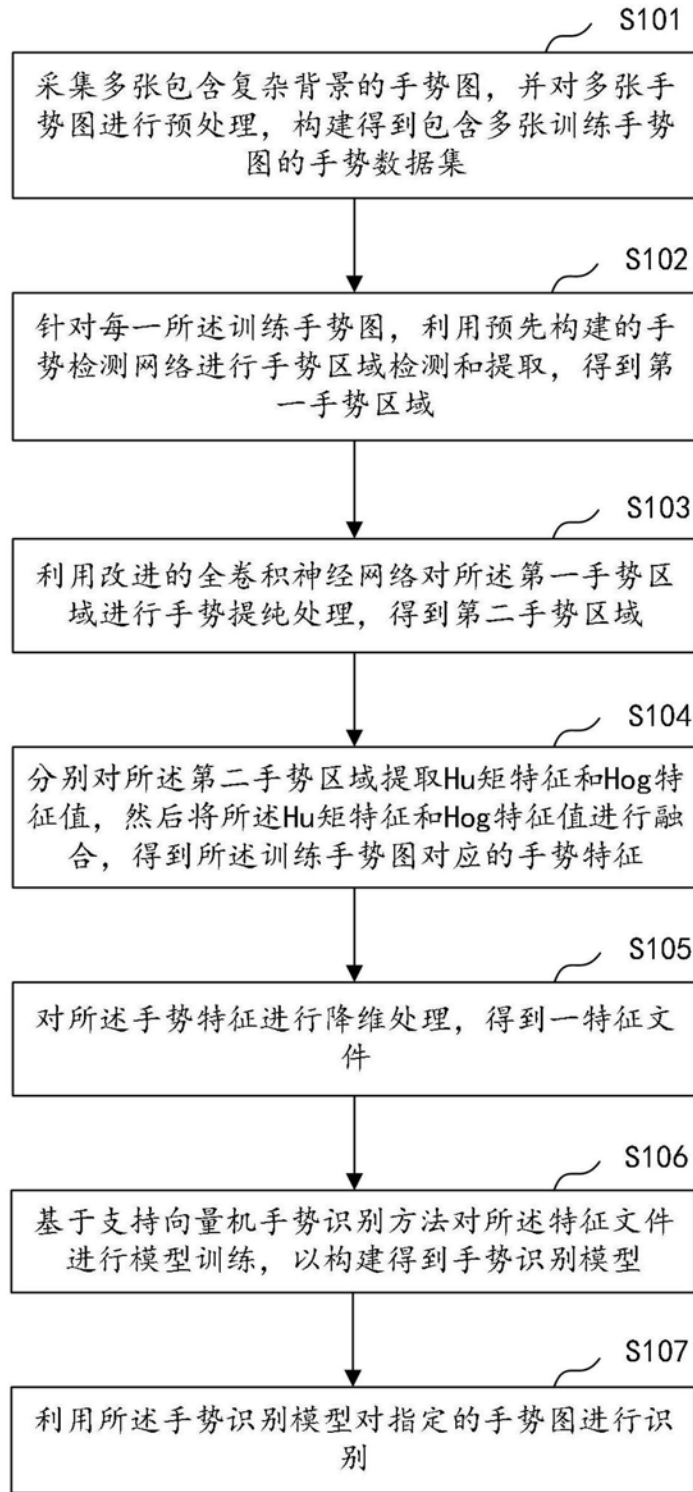


图1

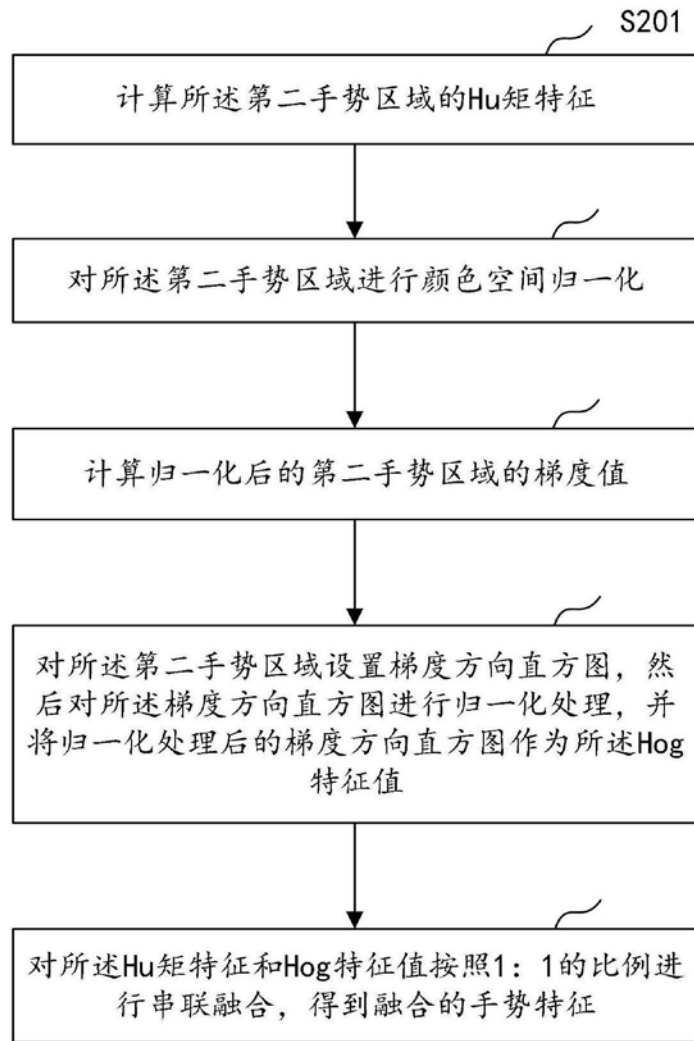


图2

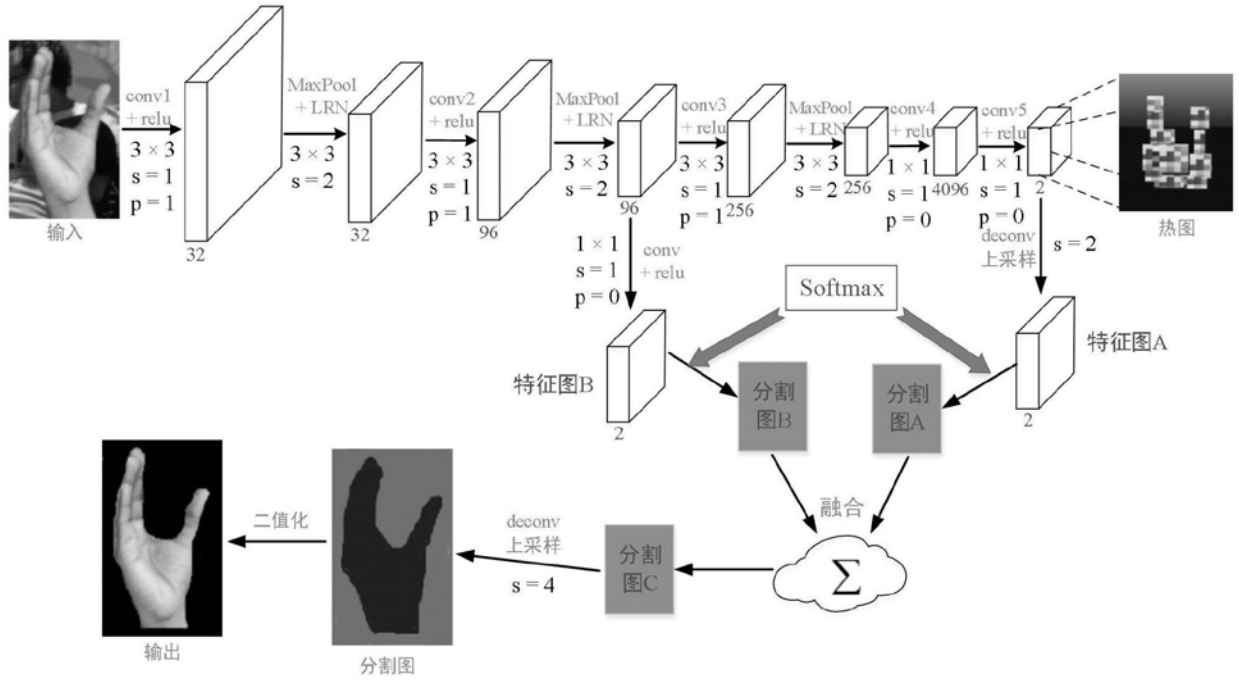


图3

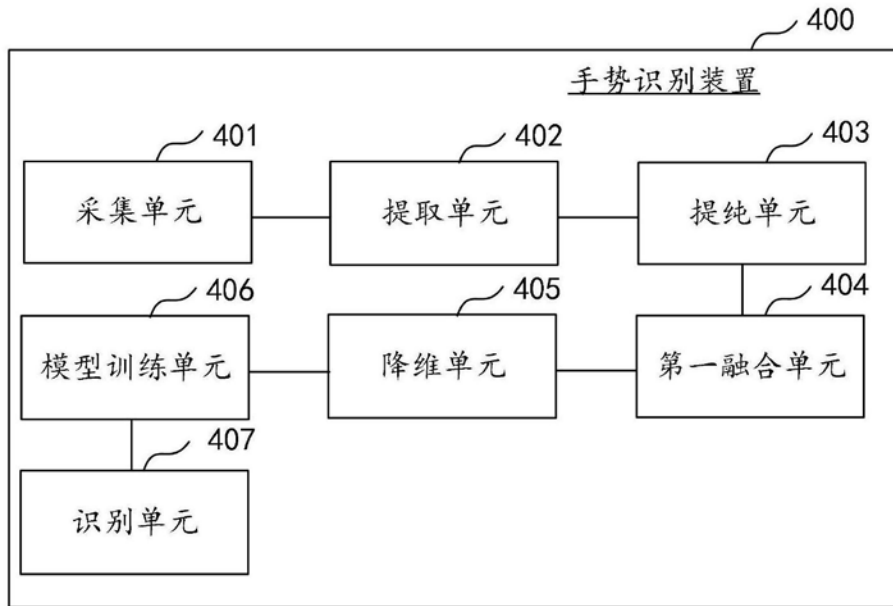


图4

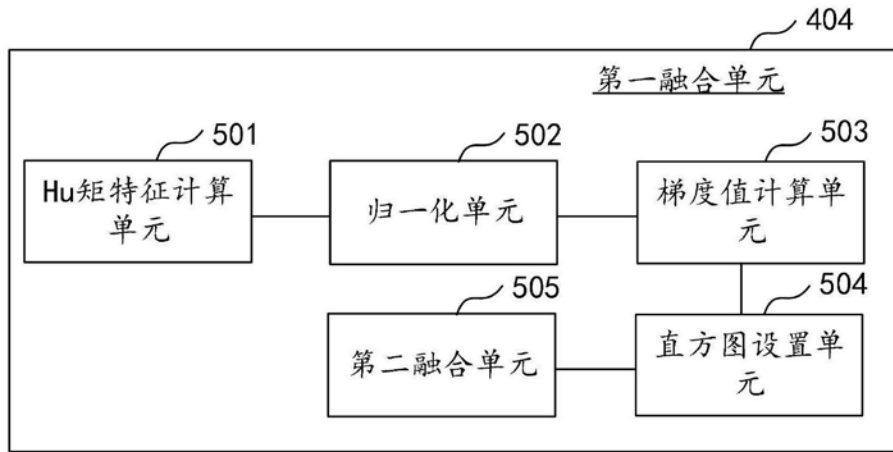


图5