



(12) 发明专利

(10) 授权公告号 CN 109033978 B

(45) 授权公告日 2023. 04. 18

(21) 申请号 201810684333.9

G06V 10/774 (2022.01)

(22) 申请日 2018.06.28

G06V 10/776 (2022.01)

G06V 10/82 (2022.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 109033978 A

(43) 申请公布日 2018.12.18

(73) 专利权人 济南大学

地址 250022 山东省济南市南辛庄路336号

济南大学信息学院

(72) 发明人 冯志全 李健

(74) 专利代理机构 北京中索知识产权代理有限公司

11640

专利代理师 商金婷

(56) 对比文件

CN 106227341 A, 2016.12.14

WO 9942920 A1, 1999.08.26

CN 106529470 A, 2017.03.22

CN 106529470 A, 2017.03.22

CN 107766842 A, 2018.03.06

US 2014079297 A1, 2014.03.20

CN 108027873 A, 2018.05.11

CN 107704072 A, 2018.02.16

审查员 高文滔

(51) Int. Cl.

G06V 40/10 (2022.01)

G06V 10/764 (2022.01)

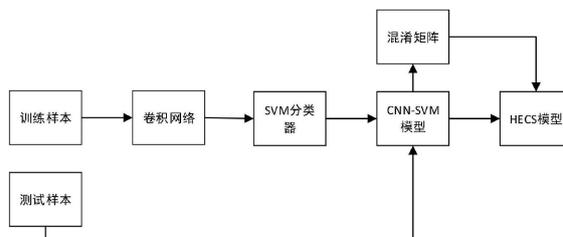
权利要求书2页 说明书9页 附图4页

(54) 发明名称

一种基于纠错策略的CNN-SVM混合模型手势识别方法

(57) 摘要

本发明提供了一种基于纠错策略的CNN-SVM混合模型手势识别方法,属于人机交互领域。所述基于纠错策略的CNN-SVM混合模型手势识别方法首先对采集到的手势数据进行预处理,然后自动提取特征并进行预测分类得到分类结果,最后利用纠错策略对所述分类结果进行纠正。利用本发明方法降低了易混淆手势之间的误识率,提高了静态手势的识别率。



1. 一种基于纠错策略的CNN-SVM混合模型手势识别方法,其特征在于:所述方法首先对采集到的手势数据进行预处理,然后自动提取特征并进行预测分类得到分类结果,最后利用纠错策略对所述分类结果进行纠正;

所述方法包括:

第一步:对采集到的数据进行预处理得到训练样本和测试样本;

第二步:获得CNN-SVM混合模型;

第三步:将测试样本输入到第二步得到的CNN-SVM混合模型中进行训练,得到分类结果以及分类结果的概率估计以及混淆矩阵;

第四步:基于第三步得到的概率估计以及混淆矩阵得到纠错策略,然后利用纠错策略对分类结果进行纠正;

所述第一步的操作包括:

(11) 采集静态手势,分别获取手部的深度图像和彩色图像;

(12) 对所述深度图像进行处理获得掩模图像;

(13) 对彩色图像和掩模图像进行与运算得到粗糙的手势区域图像;

(14) 利用贝叶斯肤色模型对所述粗糙的手势区域图像进行肤色分割得到分割后的图像,将分割后的图像分为两部分,一部分作为训练样本,另一部分作为测试样本;

所述第四步的操作包括:

在N分类问题中,设 M_i 为对分类结果为i的所有测试样本进行纠错的一个阈值,对于 M_i 的描述如下:

$$M_{i,j} = \frac{1}{S_{i,j}} \sum_{r=0}^{r=S_{i,j}} (P_n(i) - P_n(j)) \quad (1)$$

其中, $M_{i,j}$ 表示预测结果为i,但真实值为j的样本所计算出来的均值, M_i 是一个j维向量; $S_{i,j}$ 表示预测结果为i,但真实值为j的所有样本的数量, S_i 表示预测为i类的所有测试样本的数量, $P_n(i)$ 代表在所有预测为i类的所有测试样本中第n个测试样本的概率估计的最大值, $P_n(j)$ 代表次大值;i表示分类估计中最大值属于的类,j表示分类估计中次大值属于的类;

当概率估计满足以下条件时,将概率估计的最大值对应的类修改为次大值所对应的类:

$$\text{s.t.} \begin{cases} w_n(i,j) < M_{i,j} \\ \frac{1}{1+e^{-\frac{w_n(i,j)}{r_{ij}}}} > \text{rand}(0, 1) \end{cases} \quad (2)$$

其中 $w_n(i)$ 表示预测结果为i类的概率估计最大值与概率估计次大值的距离,即在数值上等于 $P_n(i) - P_n(j)$, p_{ij} 表示在混淆矩阵中分类结果为i但真实值为j的概率。

2. 根据权利要求1所述的基于纠错策略的CNN-SVM混合模型手势识别方法,其特征在于:所述步骤(11)中是采用Kinect采集静态手势。

3. 根据权利要求1所述的基于纠错策略的CNN-SVM混合模型手势识别方法,其特征在于:所述第二步是这样实现的:用SVM分类器代替CNN分类器的最后的输出层。

4. 根据权利要求1所述的基于纠错策略的CNN-SVM混合模型手势识别方法,其特征在

于:所述第二步的操作包括:

(21) 将所述训练样本输入到CNN分类器的输入层,经过CNN分类器的训练直到训练过程收敛或者达到最大的迭代次数,得到训练好的CNN模型;

(22):将所述训练样本输入到所述训练好的CNN模型中进行自动特征提取获得训练样本的特征向量;

(23):将所述训练样本的特征向量输入到SVM分类器中进行二次训练,训练完成后得到CNN-SVM混合模型。

5. 根据权利要求2所述的基于纠错策略的CNN-SVM混合模型手势识别方法,其特征在于:所述纠错策略是指:规定一个阈值,根据该阈值将错误的分类结果筛选出来,然后依据实验得出的统计数据,对最终的分类结果进行纠正。

一种基于纠错策略的CNN-SVM混合模型手势识别方法

技术领域

[0001] 本发明属于人机交互领域,具体涉及一种基于纠错策略的CNN-SVM混合模型手势识别方法。

背景技术

[0002] 随着计算机在当今社会越来越普及,一种便捷自然的人机交互(HCI)方式对使用者来说尤为重要。在众多的人机交互方式中,手势作为一种自然、简洁、直观的人机交互方式受到了越来越多人的关注,而且其在各种现实场景中都能发挥重要的作用,如体感游戏、手语识别、智能穿戴设备和智能教学等方面。手势识别的目的在于设计一种算法使计算机能够识别出图片或者人体的手势,理解手势的含义,从而实现人与计算机的交互。在手势识别过程中,手势通常是处于复杂的环境下,为了精准的进行人机交互,所设计的手势识别算法应该在各种光线、角度、背景以及其他复杂环境下都有良好的识别能力。

[0003] 传统的手势识别算法主要基于隐马尔可夫模型(HMM)和模板匹配。其中,基于隐马尔可夫模型的手势识别方法,该模型可以用于表达一个隐含未知参数的马尔科夫过程,而手势识别的过程可以看做是一个含有时间序列的马尔科夫链,因此该模型可以应用于手势识别。基于模板匹配的手势识别方法将手势的轮廓、边缘、空间分布等信息作为特征建立手势模板,应用模板匹配算法实现手势识别。这两种方法需要人工提取特征,而人工提取的手势特征需要大量的经验基础,并且人工提取的特征具有一定的主观性和局限性,使得其容易忽视一些显著性的特征,因此传统方法往往识别能力有限且效率不高。

[0004] 卷积神经网络(Convolutional Neural Network,CNN)是目前机器视觉和图像处理领域应用最广泛的模型之一,卷积神经网络可以通过训练学习得到输入图像的局部和全局特征,解决了人工提取特征带来的特征提取不充分的问题。近几年来,卷积神经网络已经成功的应用于图像检索、人脸识别、表情识别和目标检测。已经有学者将CNN应用于手势识别领域,Jawad Nagi等人将最大池化层与卷积神经网络相结合(MPCNN)用于手势识别取得了不错的效果,Takayoushi等人提出了一种端到端的深度卷积网络实现了手势识别,同时提高了手势识别的准确率。在手势识别应用上,一般采用的都是比较浅的网络,在传统的静态手势识别方法中,基于人工特征提取的手势识别方法耗时长,识别率低。

发明内容

[0005] 本发明的目的在于解决上述现有技术中存在的难题,提供一种基于纠错策略的CNN-SVM混合模型手势识别方法,采用的网络更深,能够学习到更加深层次的特征,降低模型对易混淆手势的误识率,最终实现静态手势的识别。

[0006] 本发明是通过以下技术方案实现的:

[0007] 一种基于纠错策略的CNN-SVM混合模型手势识别方法,首先对采集到的手势数据进行预处理,然后自动提取特征并进行预测分类得到分类结果,最后利用纠错策略对所述分类结果进行纠正。

[0008] 所述方法包括：

[0009] 第一步：对采集到的数据进行预处理得到训练样本和测试样本；

[0010] 第二步：获得CNN-SVM混合模型；

[0011] 第三步：将测试样本输入到第二步得到的CNN-SVM混合模型中，得到分类结果以及分类结果的概率估计以及混淆矩阵；

[0012] 第四步：基于第三步得到的概率估计以及混淆矩阵得到纠错策略，然后利用纠错策略对分类结果进行纠正。

[0013] 所述第一步的操作包括：

[0014] (11) 采集静态手势，分别获取手部的深度图像和彩色图像；

[0015] (12) 对所述深度图像进行处理获得掩模图像；

[0016] (13) 对彩色图像和掩模图像进行与运算得到粗糙的手势区域图像；

[0017] (14) 利用贝叶斯肤色模型对所述粗糙的手势区域图像进行肤色分割得到分割后的图像，将分割后的图像分为两部分，一部分作为训练样本，另一部分作为测试样本。

[0018] 所述步骤(11)中是采用Kinect采集静态手势。

[0019] 所述第二步是这样实现的：用SVM分类器代替CNN分类器的最后的输出层

[0020] 所述第二步的操作包括：

[0021] (21) 将所述训练样本输入到CNN分类器的输入层，经过CNN分类器的训练直到训练过程收敛或者达到最大的迭代次数，得到训练好的CNN模型；

[0022] (22)：将所述训练样本输入到所述训练好的CNN模型中进行自动特征提取获得训练样本的特征向量；

[0023] (23)：将所述训练样本的特征向量输入到SVM分类器中进行二次训练，训练完成后得到CNN-SVM混合模型。

[0024] 所述纠错策略是指：规定一个阈值，根据该阈值将错误的分类结果筛选出来，然后依据实验得出的统计数据，对最终的分分类结果进行纠正。

[0025] 所述第四步的操作包括：

[0026] 在N分类问题中，设 M_i 为对分类结果为i的所有测试样本进行纠错的一个阈值，对于 M_i 的描述如下：

$$[0027] \quad M_{ij} = \frac{1}{S_{ij}} \sum_{n=0}^{N-1} S_{ij} (P_n(i) - P_n(j)) \quad (1)$$

[0028] 其中， $M_{i,j}$ 表示预测结果为i，但真实值为j的样本所计算出来的均值， M_i 是一个j维向量； $S_{i,j}$ 表示预测结果为i，但真实值为j的所有样本的数量， S_i 表示预测为i类的所有测试样本的数量， $P_n(i)$ 代表在所有预测为i类的所有测试样本中第n个测试样本的概率估计的最大值， $P_n(j)$ 代表次大值；i表示分类估计中最大值属于的类，j表示分类估计中次大值属于的类；

[0029] 当概率估计满足以下条件时，将概率估计的最大值对应的类修改为次大值所对应的类：

$$[0030] \quad \text{s.t.} \quad \begin{cases} w_n(i,j) < M_{i,j} \\ \frac{1}{1+e^{\frac{p_{ij}}{w_n(i,j)}}} > \text{rand}(0, 1) \end{cases} \quad (2)$$

[0031] 其中 $w_n(i)$ 表示预测结果为i类的概率估计最大值与概率估计次大值的距离,即在数值上等于 $P_n(i) - P_n(j)$, p_{ij} 表示在混淆矩阵中分类结果为i但真实值为j的概率。

[0032] 与现有技术相比,本发明的有益效果是:利用本发明方法降低了易混淆手势之间的误识率,提高了静态手势的识别率。

附图说明

[0033] 图1-1.九种不同的手势的照片

[0034] 图1-2 对应图1-1中九种不同的手势的深度图像

[0035] 图2 本发明方法中的图像预处理不足框图

[0036] 图3 预处理过程中的图片

[0037] 图4 本发明方法所采用的CNN网络结构图

[0038] 图5 在不同数据集上测试准确率的曲线

[0039] 图6 本发明方法的步骤框图。

具体实施方式

[0040] 下面结合附图对本发明作进一步详细描述:

[0041] 本发明结合卷积神经网络和支持向量机的优点,提出了一种混合模型来自动提取特征并提高模型的泛化能力,并且用一种基于概率估计的纠错策略来降低易混淆手势的误识率。

[0042] 如图6所示,本发明方法包括:首先,对Kinect采集的手势数据进行分段预处理,以减少复杂背景和人体其他部分的干扰。然后,混合模型自动提取特征并进行预测分类。最后,利用纠错策略对分类决策进行调整。在所建立的数据库上进行实验,最终得到没有使用纠错策略的识别率为95.81%,使用纠错策略后得到平均准确率为97.32%。

[0043] 本发明方法中的数据收集如下:

[0044] 本发明系统采用Kinect2.0采集静态手势,分别获取手部的深度图像与彩色图像,然后建立相应的手势数据库。所建立的手势库共包含17类手势,由300个在校大学生在不同光照背景下采集的静态图像组成。在本发明中挑选了人类常用的9种手势,每种手势包含3300张图片。图1-1、图1-2分别为操作者完成的9种手势的照片和深度图像。

[0045] 数据预处理如下:

[0046] 从采集到的手势图像不难看出,虽然彩色图像中人手势的图像清晰可辨,但要完成精确的识别还是十分困难,这是因为采集到的手势受视角、外观、形状、人体其他部位和复杂背景的影响。而在采集的深度图像中,一方面深度信息不受人手自身的颜色、纹理特征以及光照的影响,鲁棒性好,精度高;另一方面,深度图像中的深度信息反映的是人手到采集设备之间的距离,因此在手势区域的深度差别并不是很大。因为深度图像在采集过程中已经进行了分割,因此,利用该特点能够帮助分割出彩色图像感兴趣的手势区域,从而减少了彩色图像中人体其他部位以及复杂背景的干扰。分割预处理的步骤如图2所示。

[0047] 在预处理过程中,本发明将所采集的深度图像二值化,由于在采集过程中将深度图像已经转换为灰度深度图像,即将深度值到的取值范围调整到灰度值 [0-255]之间。由于在采集过程中已经对深度图进行了手势区域的分割,可以利用灰度值的大小来得到手势区域的二值图像。直接将掩模图像(对灰度图像设定一个阈值本发明中为128,大于128的像素点赋值为1,小于128的像素点赋值为0)与彩色图像进行逻辑与运算只能得到粗糙的手势区域图像,由于在 Kinect采集过程中,深度图像与彩色图像存在分辨率不同的问题,周围还会有非手势像素点的干扰。对得到的粗糙手势区域进行肤色分割,利用贝叶斯肤色模型(请参见文献“M.J.Jones,et al.Statistical color models with application to skin detection[J].International Journal of Computer Vision(IJCV),2002,46(1):81-96”)得到精确的手势区域图像。

[0048] 在本发明随机的选择了一张图像以检验分割预处理的有效性,其中彩色图像、深度图像、掩模图像、粗糙手势区域以及分割后图像分别如图3-1到图3-5所示。

[0049] 可以明显的看到本发明方法中的分割预处理可以有效的去处复杂背景以及人体其他部分的影响,而且最后利用贝叶斯肤色模型也能够准确的保留手势区域的有效信息,为后期的训练工作提供了良好的数据保障。

[0050] 混合CNN-SVM模型如下:

[0051] SVM分类器:支持向量机通过选择不同的核函数将低维输入空间线性不可分的样本转换成高维特征空间使其线性可分,以机构风险最小化原则为理论基础在特征空间中构建最优超平面,得到了对数据分布的结构化描述,因此降低了对数据规模和数据分布的要求,有效降低了独立测试集误差,被认为是最常用、效果最好的分类器之一。

[0052] 在实验中使用LIBSVM(请参见文献“Chih-Chung Chang,Chih-Jen Lin.LIBSVM:A library for support vector machines[J].ACM Transactions on Intelligent Systems and Technology(TIST),2011,2(3):1-27”)来构建SVMs,LIBSVM是一种快速有效的用于分类和回归的软件包,使用一对一策略来解决多分类问题。LIBSVM不仅能够预测分类结果而且可以给每一个测试样本提供分类的概率信息。对于一个k分类问题,其目的就是去估计样本 \vec{x} 属于每一类的概率:

$$[0053] \quad p_i = P(y = i | \vec{x}), i = 1, \dots, k \quad (1)$$

[0054] 对于一对一策略, p_i 通过以下解决下列优化问题来得到:

$$[0055] \quad \min: W(\vec{p}) = \frac{1}{2} \sum_{i=1}^k \sum_{j \neq i} (r_{ji} p_i - r_{ij} p_j)^2 \quad (2)$$

$$[0056] \quad \text{s.t.} \sum_{i=1}^k p_i = 1, p_i \geq 0, \forall i \quad (3)$$

[0057] 其中 r_{ij} 是一个成对的概率定义为:

$$[0058] \quad r_{ij} \approx P(y = i | y = i, \text{ or } j, \vec{x}) \quad (4)$$

[0059] 在本发明的实验中,SVMs被训练用于预测带概率的分类结果,这些分类结果的概率值将被应用于对于易混淆手势的纠错,去决定分类结果是否直接应用,还是通过本发明所采用的一种策略来进行重新分类。

[0060] CNN分类器:卷积神经网络是一种深度前馈的神经网络,将图像直接作为网络的输

入,不需要人工定义和特征选择,避免了在传统识别算法中的特征选择和特征提取的环节,同时还具有良好的容错能力、并行处理能力和自学习能力。

[0061] 代替使用在MPCNN,(请参见文献“Chih-Chung Chang,Chih-Jen Lin.LIBSVM:A library for support vector machines[J].ACM Transactions on Intelligent Systems and Technology (TIST),2011,2(3):1-27”)本发明采用了一种在文献“A.Krizhevsky,S.Ilya,and G.E.Hinton.Imagent classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems 2 (NIPS),2012:1106-1114”中所提到的更为复杂的CNN进行了训练,其中网络结构如图4所示。这个网络总共有8层,包括5个卷积层和3个全连接层,最后一个全连接层输出一个9维的softmax 来表达对于9个类别的预测。第一层卷积层将 $224 \times 224 \times 3$ 的输入图像与96个 $11 \times 11 \times 3$ 的卷积核做卷积运算,步长为4。第二层卷积层将第一层经过响应归一化和池化的输出与256个 $5 \times 5 \times 48$ 的卷积核做卷积运算。第三层卷积层用384个 $3 \times 3 \times 256$ 的卷积核与进过归一化核池化的第二层输出做卷积运算。第四层卷积层的卷积核个数为384大小为 $3 \times 3 \times 192$,第五层卷积层有256个 $3 \times 3 \times 192$ 大小的卷积核。每一个全连接层都有4096个神经元。由于网络结构复杂,本发明采用放大数据集的方式来应对过拟合。通过从 256×256 的图片中随机抽取 224×224 的区块以及水平镜像来实现这种方法,并在这些收取得到的区块上来训练神经网络。如果不采用这种方法,网络会出现严重的过拟合,迫使采用更小的网络,从而导致在SVM训练中无法使用深层次的特征。

[0062] CNN-SVM混合模型:本发明所采用的混合CNN-SVM模型是将CNN中最后的输出层用SVM代替。首先,将经过处理的图像传入输入层,经过原始CNN多次的训练直到训练过程收敛或者达到最大的迭代次数。然后将训练样本输入训练好的 CNN模型,得到训练样本的特征向量,将其输入SVM分类器进行二次训练,训练完成后得到CNN-SVM模型,将测试样本输入模型得到分类结果。

[0063] 基于概率估计的纠错策略(HECS):LIBSVM在最后的预测结果中给出了每一个样本被分到各类的一个概率估计,最终选择的分类结果是其中概率值最大的一个,表1列出了一些预测分类结果错误的测试样本最终的概率分布,表中的第一列表示测试样本的真实类别编号,表中的第二列代表的该测试样本的预测分类编号,剩余的其他列分别代表该样本属于某列的概率,从中可以观察到在预测错误的测试样本的概率估计中,估计概率的最大值为预测值而次大值就为其真实值。

Truth	Forecast	0	1	2	3	4	5	6	7	8
0	8	0.052	0.022	0.004	0.001	0.001	0.000	0.000	0.000	0.895
6	4	0.000	0.000	0.002	0.000	0.963	0.001	0.016	0.000	0.000
3	2	0.000	0.000	0.841	0.155	0.001	0.000	0.002	0.000	0.000
2	3	0.000	0.002	0.005	0.989	0.002	0.000	0.000	0.000	0.001
5	6	0.000	0.000	0.000	0.000	0.000	0.006	0.990	0.002	0.000

[0065] 表1

[0066] 根据LIBSVM最终的决策特点以及从最后的实验结果可以得知,在预测分类结果错误的样本中,预测分类与真实分类之间的概率估计差距很小,而在预测结果正确的样本中,预测分类与其他各个分类结果之间的概率估计差距都比较大。本发明依据这种特点,提出了一种基于概率估计的纠错策略以达到减少此种情况下所产生的分类错误。在N分类问题中,本发明采用 M_i 作为预测结果为i的所有测试样本进行纠错的一个阈值,对于 M_i 的描述如下:

$$[0067] \quad M_{ij} = \frac{1}{S_{ij}} \sum_{n=0}^{S_{ij}} (P_n(i) - P_n(j)) \quad (5)$$

[0068] 其中 S_i 表示预测为i类的所有测试样本的数量, $P_n(i)$ 代表在所有预测为i类的所有图片中第n个测试样本概率估计的最大值, $P_n(j)$ 代表次大值。i表示概率估计中最大值属于的那一类,j表示概率估计中次大值属于的那一类。

[0069] 当概率估计满足以下条件时,将最大值的对应的类修改为次大值所对应的类。

$$[0070] \quad \text{s.t.} \begin{cases} w_n(i, j) < M_{i, j} \\ \frac{1}{1 + e^{-\frac{p_{ij}}{w_n(i, j)}}} > \text{rand}(0, 1) \end{cases} \quad (6)$$

[0071] 其中 $w_n(i)$ 表示预测结果为i类的概率估计最大值与次大值的距离,即在数值上等于 $P_n(i) - P_n(j)$, p_{ij} 表示在混淆矩阵中,预测结果为i但真实值j的概率。

[0072] 本发明模型的优点如下:

[0073] 本发明构建CNN-SVM模型以期达到弥补CNN和SVM分类器的限制,结合两种分类器的优点。卷积神经网络的理论学习方法与多层感知机(MLP)(可参考文献“E.A.Zanaty.Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification[J].Egyptian Informatics Journal,2012,13(3):177-183”)的学习方法是相同的,因此实质上是MLP的一个扩展。MLP理论基于经验风险最小化的,在训练过程中将训练的误差不断的减小。当进行反向传播计算时发现一个极小值,无论其是否是全局的最小值都会使得训练结果收敛在这一点,而不再继续改进算法的解。SVM是在训练样本集分布固定的情况下,利用结构风险化最小原则去寻找一个最优的超平面,最小化数据上的泛化误差,因此SVM的泛化能力要优于MLP。

[0074] CNN的优点在于能够自动地提取输入图像深层次的特征,而且输入图像在一定程度上移动和扭曲时特征仍然具有不变性。然而,人工进行提取特征需要进行精心的设计,在手势识别方法上传统的人工提取特征的方法(如文献“Jiang Y.An HMM based approach for video action recognition using motion trajectories[C]//IEEE International Conference on Intelligent Control and Information Porcessing,2010:359-464.”、“Liu Jie,Huang Jin,Han Dongqi,Tian Feng,et al.Template Matching Algorithm for 3D Gesture Recognition[J].Journal of Computer-Aided Design&Computer Graphics,2016,28(8):1365-1372”提供的方法)忽视了手的局部视觉特征,只注意到了手势的轮廓和颜色信息,诸如手指的弯曲,手指之间的距离这在手势识别方面都是非常重要的特征。人工设计的特征提取容易忽视和丢失一些特征。因此,利用CNN来进行特征的提取能够比传统方法收集到更多的具有代表性和相关性的信息。

[0075] 纠错策略实际上就是规定了一个阈值,将有可能出错的预测分类结果筛选出来,然后依据实验得出的统计数据,以一定概率对最终的分类决策进行纠正。通过CNN-SVM模型对样本进行分类已经可以得到很好的效果,但对于一些由于遮挡或者采集图像质量问题而导致的两个样本难以分别的情况并不能进行准确的判断,本发明所提出的纠错策略能够在最终的决策中将易混淆样本的分类结果进行一定的纠正以提高最后整体的准确率。

[0076] 对本发明方法进行实验与分析如下:

[0077] 实验环境:在本实验中,手势识别模型运行在Windows操作系统上,硬件配置为: Intel (R) Core (TM) i5-6500处理器,NVIDIA GeForceGT730,内存为 8G,显存为2G.CNN网络是由Caffe搭建,本发明采用径向核 (Gaussian RBF),利用LIBSVM软件包实现SVM分类器。实验中所有算法在Matlab2014a平台运行。

[0078] 实验结果与分析如下:

[0079] 在本发明的实验中,首先将彩色图像与深度图像进行分割预处理,将得到的分割出的手势图像共计29700张作为本发明的数据集,其中27000张图片用于模型的训练,2700张图片用于测试。在CNN训练过程采用30000次为最大的迭代次数,可以从图5中看出,迭代约10000次时系统已经达到收敛,最后利用迭代30000次的模型进行测试,在测试集上的准确率为88.35%。然后再建立 CNN-SVM模型,将最后的全连接层代替为SVM分类器,将4096维的特征向量放入SVM中进行训练和测试。在本发明实验中,SVM采用RBF核函数,为了去寻找最优的乘法系数C和最优的核参数g在训练集上采用5折交叉验证法得到最优的结果。这两个参数寻找的范围分别是: $g = [2^3, 2^1, \dots, 2^{-15}]$ 和 $C = [2^{15}, 2^{13}, \dots, 2^{-5}]$ 。总共尝试了 $11 \times 10 = 110$ 种不同的组合,最终确定 $C = 64, g = 0.00024414$ 。然后用得到的这两个参数用于混合模型的训练,最终的到训练的准确率为99.94%,在2700张测试图片上的准确率达到95.81%。表2列出了在本发明所准备的数据集上,使用CNN和使用CNN-SVM的训练准确率与测试准确率。

[0080] 由图5可以看出,在最大迭代次数同为30000次时,彩色图像的准确率最低,最多只能达到37.92%的准确率,深度图像与彩色图像相比有了明显的改善能够达到79.07%,经过预处理后的图像准确最高能够达到88.35%。这是由于直接用未处理的彩色图像进行训练时,训练样本本身存在大量的噪声信息(复杂的背景信息以及人体其他部分的信息),用分割出的深度图像虽然没了背景和人体其他部分的干扰,但是因为所采集的深度图像是将深度信息投影到了 $[0, 255]$ 的灰度信息中保存起来的,因此深度图像也会有一部分信息的缺失,而进过本发明预处理分割后的手势不仅能够有效的去处复杂背景和人体其他部分大干扰同时能够保留手势区域完整的彩色信息,使得在进行CNN网络训练的时候能够提取出更为丰富的特征用于分类。通过将测试样本放入混合模型进行分类预测,可以统计出一个混淆矩阵如表2所示:

	Truth	Prediction								
		0	1	2	3	4	5	6	7	8
[0081]	0									24
	1	1		2	2	1				7
	2		15		12		1			1
	3		2	8				2		
	4			5	3		1	1		
	5					3		3	3	1

	6		7	2	
[0082]	7			2	2
	8	1	1		

[0083] 表2

[0084] 在100次试验中,纠错率主要集中在[3%,5%]之间,准确率在[97%,98%]最为集中,平均纠错率为4.12%,平均准确率为97.32%。

[0085] 表3给出了本发明方法和其他方法在所提供的数据集下手势识别的准确率。与本发明方法不同,文献“Yamashita T,Watasue T.Hand posture recognition based on bottom-up structured deep convolutional neural network with curriculum learning[C]//Image Processing(ICIP),2014 IEEE International Conference on.IEEE,2014:853-857”是用了一种比较简单的卷积神经网络,将最大池化层与卷积神经网络构成MPCNN,在测试集上得到了 68.89%的识别准确率。文献“Shao-Zi Li,Bin Yu,Wei Wu,Song-Zhi Su,Rong-Rong Ji.Feature learning based on SAE-PCA network for human gesture recognition in RGBD images[J].Neurocomputing,2015,151(2):565-573”使用的是一种端到端的卷积神经网络,得到手势识别准确率为85.43%。文献“Xiao-Xiao Niu,Ching Y.Suen.A novel hybrid CNN-SVM classifier for recognizing handwritten digits[J].Pattern Recognition,2012,45(4):1318-1325”先利用深度信息与肤色信息进行手势分割,然后同过基于特征学习的SAE-PCA模型提取特征,最后采用SVM分类器进行分类,最终的到手势识别的准确率为93.32%,不同手势识别方法在本发明数据集上的准确率如表3所示:

	Method	Test Accuracy (%)
	MPCNN [9]	68.89
[0086]	Bottom-Up Deep CNN [10]	85.43
	SAE-PCA [11]	93.32
		88.35
	CaffeNet [14]	95.81
[0087]	CNN-SVM	97.32
	HECS	

[0088] 表3

[0089] 可以看出,本发明方法在识别准确方面相比其他方法都有明显的提升。

[0090] 本发明方法首先对手势的深度数据与彩色数据进行了分割预处理,消除了彩色数据人体以及复杂背景的影响;然后利用卷积神经网络提取手势的特征,避免了根据手势的轮廓和几何特性人为设计特征的复杂过程;再通过支持向量机进行手势的概率估计;最后,基于所得到的概率估计结合实验得出的混淆矩阵提出一种纠错策略对模型的分类结果进行纠错。大量的实验结果表明本方法能够有效的识别静态手势,且能够在一定程度上优化CNN-SVM模型分类易混淆手势的能力,在整体上能够提高最终识别的准确率。

[0091] 上述技术方案只是本发明的一种实施方式,对于本领域内的技术人员而言,在本发明公开了应用方法和原理的基础上,很容易做出各种类型的改进或变形,而不仅限于本

发明上述具体实施方式所描述的方法,因此前面描述的方式只是优选的,而并不具有限制性的意义。

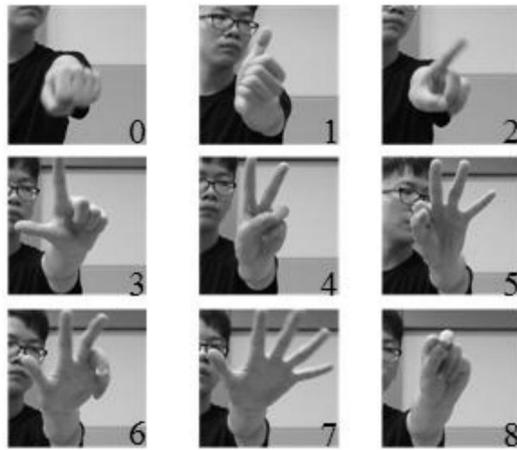


图1-1

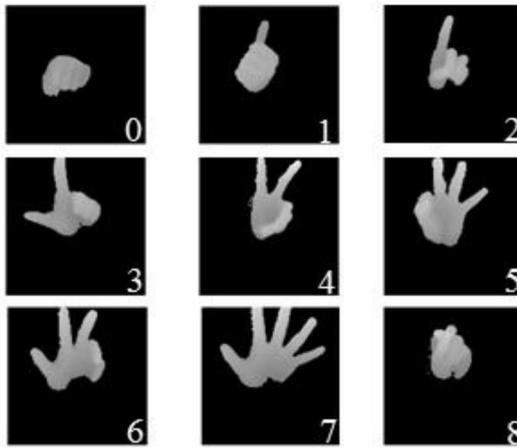


图1-2

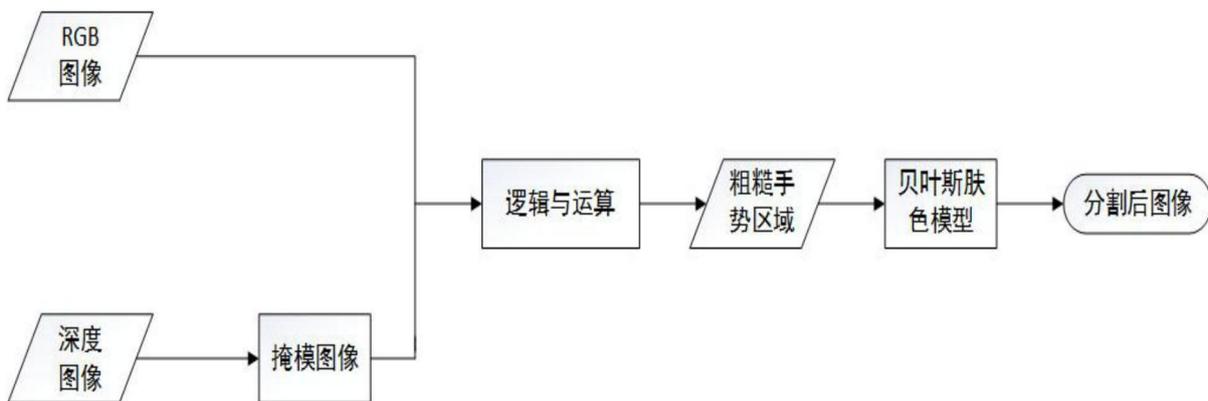


图2



图3-1



图3-2



图3-3



图3-4



图3-5

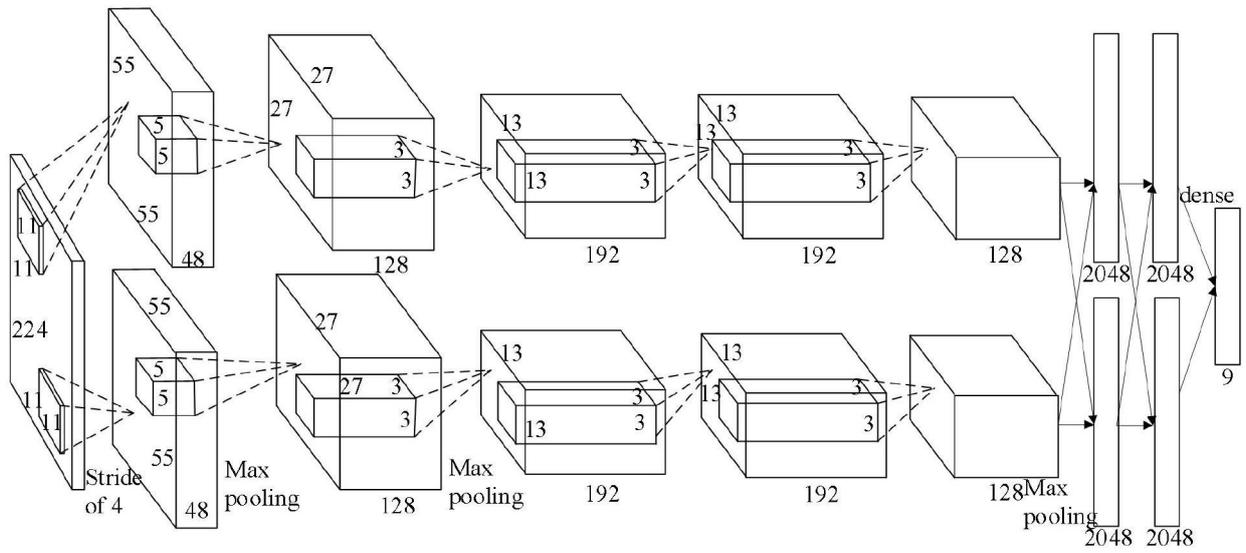


图4

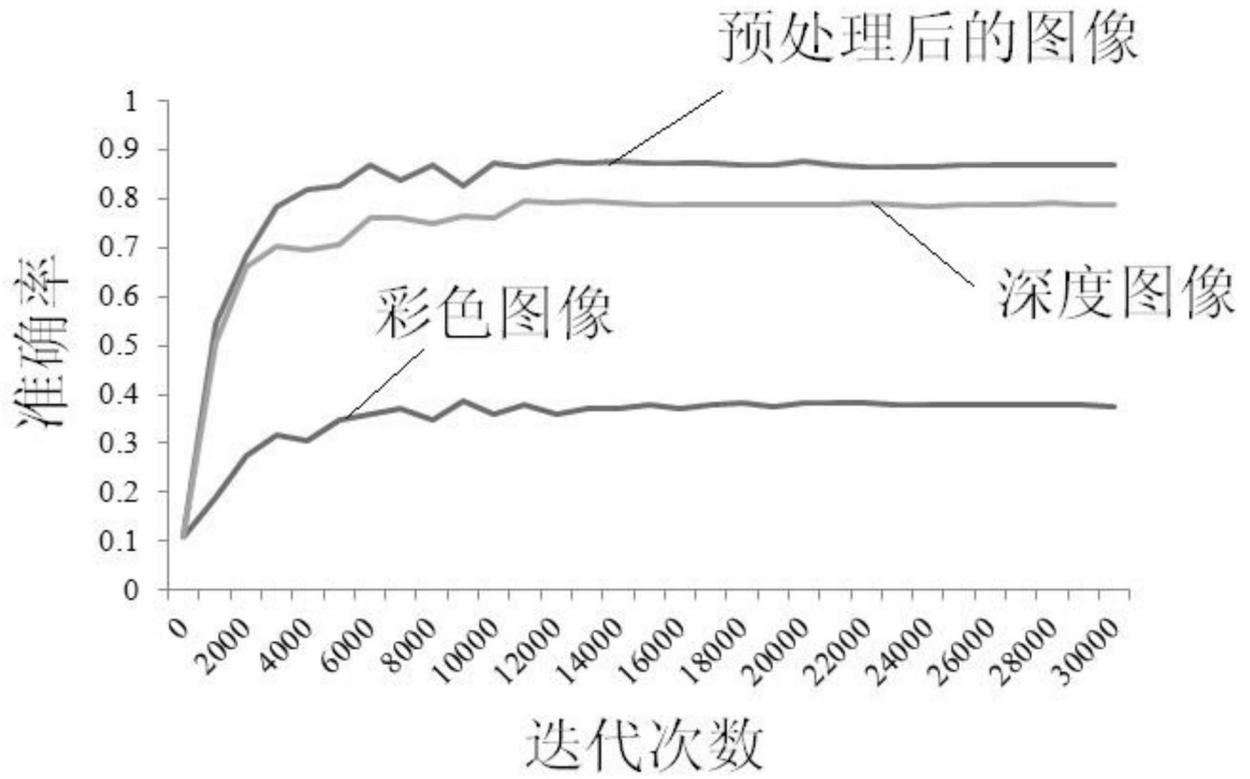


图5

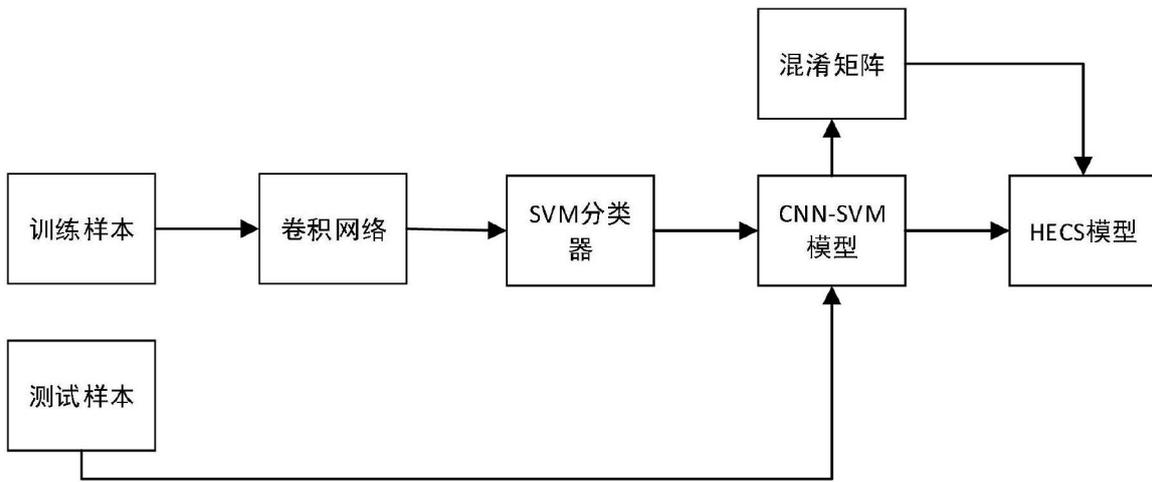


图6