(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111839551 A
(43)申请公布日 2020.10.30

(21)申请号 201910353028.6

(22)申请日 2019.04.29

(71)申请人 北京入思技术有限公司
地址 100083 北京市海淀区学清路甲18号中关村东升科技园学院园东配楼2层224室(东升地区)

(72)发明人 王春雷 尉迟学彪 毛鹏轩

(51)Int.Cl.
A61B 5/16(2006.01)
A61B 5/04(2006.01)
A61B 5/00(2006.01)

权利要求书2页 说明书5页 附图2页

(54)发明名称

一种基于视频和生理信号的多模态情感识别方法及系统

(57)摘要

本发明提供一种基于视频和生理信号的多模态情感识别方法及系统，所述方法包括：接收用户视频图像和皮肤电生理信号数据；分别对所述视频图像和皮肤电生理信号数据进行特征提取；利用深度信念网络对所述视频图像特征和皮肤电生理信号特征进行特征选择和融合，得到多模态融合特征；以及利用支持向量机对所述多模态融合特征进行分类，从而得到最终的情绪识别结果。本发明针对多模态情感识别，采用深度信念网络实现了特征选择过程的自动化，减少了特征选择对人工经验和实验次数的依赖性，对多模态情感识别提供了新思路。

CN 111839551 A

1.一种基于视频和生理信号的多模态情感识别方法,其特征在于,所述方法包括:

步骤S101:接收用户视频图像和皮肤电生理信号数据;

步骤S102:分别对所述视频图像和皮肤电生理信号数据进行特征提取;

步骤S103:利用深度信念网络对所述视频图像特征和皮肤电生理信号特征进行特征选择和融合,得到多模态融合特征;以及

步骤S104:利用分类器对所述多模态融合特征进行分类,从而得到最终的情绪识别结果。

2.如权利要求1所述的基于视频和生理信号的多模态情感识别方法,其特征在于,所述步骤S102包括:

针对所述视频图像,检测人脸并标记面部区域关键点,计算所述面部区域各关键点的位置与面部区域关键点平均位置的距离,并针对面部区域各关键点提取尺度不变特征变换(SIFT)特征,以得到视频图像特征向量;以及

针对所述皮肤电生理信号,通过低通滤波进行去噪预处理,然后分别计算原始信号及其一阶微分、二阶微分的均值、中值、标准差、最大值、最小值共计五类特征,以得到皮肤电生理信号特征向量。

3.如权利要求1所述的基于视频和生理信号的多模态情感识别方法,其特征在于,所述步骤S103包括:

针对所述视频图像特征向量,利用深度信念网络进行特征选择和融合,得到所述视频图像的高层特征;

针对所述皮肤电生理信号特征向量,利用深度信念网络进行特征选择和融合,得到所述皮肤电生理信号的高层特征;以及

针对所述视频图像的高层特征和所述皮肤电生理信号的高层特征,利用深度信念网络进行特征选择和融合,得到所述用户的多模态融合特征。

4.如权利要求1所述的基于视频和生理信号的多模态情感识别方法,其特征在于,所述步骤S104中的分类器为基于径向基函数的支持向量机。

5.一种基于视频和生理信号的多模态情感识别系统,其特征在于,所述系统包括:

数据接收模块,用于接收用户视频图像和皮肤电生理信号数据;

特征提取模块,用于提取用户视频图像特征和皮肤电生理信号特征;

特征融合模块,用于对所述视频图像特征和皮肤电生理信号特征进行特征选择和融合,以得到多模态融合特征;以及

情绪识别模块,用于对所述多模态融合特征进行分类,从而得到最终的情绪识别结果。

6.如权利要求5所述的基于视频和生理信号的多模态情感识别系统,其特征在于,所述特征提取模块通过如下方法实现:

针对用户视频图像,检测人脸并标记面部区域关键点,计算所述面部区域各关键点的位置与面部区域关键点平均位置的距离,并针对面部区域各关键点提取尺度不变特征变换(SIFT)特征,以得到视频图像特征向量;以及

针对用户皮肤电生理信号,通过低通滤波进行去噪预处理,然后分别计算原始信号及其一阶微分、二阶微分的均值、中值、标准差、最大值、最小值共计五类特征,以得到皮肤电生理信号特征向量。

7.如权利要求5所述基于视频和生理信号的多模态情感识别系统，其特征在于，所述特征融合模块通过如下方法实现：

针对所述视频图像特征向量，利用深度信念网络进行特征选择和融合，得到所述视频图像的高层特征；

针对所述皮肤电生理信号特征向量，利用深度信念网络进行特征选择和融合，得到所述皮肤电生理信号的高层特征；以及

针对所述视频图像的高层特征和所述皮肤电生理信号的高层特征，利用深度信念网络进行特征选择和融合，得到所述用户的多模态融合特征。

8.如权利要求5所述基于视频和生理信号的多模态情感识别系统，其特征在于，所述情绪识别模块通过使用基于径向基函数的支持向量机实现对所述多模态融合特征的分类过程。

一种基于视频和生理信号的多模态情感识别方法及系统

技术领域

[0001] 本发明涉及信号处理、情感识别技术领域，具体而言涉及一种基于视频和生理信号的多模态情感识别方法及系统。

背景技术

[0002] 情感识别的目的在于通过计算机对用户的生理信号进行分析和处理，得出用户的情感状态。目前针对语音或者生理信号的单模式情感识别技术已经相对成熟，但存在信息单一识别的结果不够可靠、准确的缺点。因此，利用不同性质的多模态特征的多模态情感识别技术值得进一步研究。

[0003] 多模态情感识别的主要步骤包括信息特征提取和分类器设计。分类器主要有支持向量机、神经网络、K近邻算法、贝叶斯方法等等。国内外研究人员在解决多模态情感识别问题时，大部采用这些分类算法。这类多模态情感识别方法极大地依赖于对情感特征的抽取，而目前采用的特征抽取方法大都是人工设计的，再通过特征选择算法剔除冗余或者不相关的特征，得出最优或者次优特征子集，这一步骤的目的是为了提高识别准确率和降低特征维度。这一过程极大地依赖人工专家的经验和反复实验，既需要大量的人力与计算资源，又很难得到最优的情感特征表达，从而影响了情感识别的最终效果。

[0004] 本发明针对现有多模态情感识别技术中特征提取方法的不足，利用深度信念网络在自动提取特征方面的优势，结合多模态情感识别技术，实现一种基于视频和生理信号的多模态情感识别方法。既利用了多模态特征的相关性和互补性，实现更加可靠稳定的情感识别，又能通过深度信念网络的非线性结构更好地学习复杂数据的结构和分布，自动提取更高级的特征然后分类，减少了情感特征提取对人的依赖性。

发明内容

[0005] 针对现有技术的不足，本发明提出一种基于视频和生理信号的多模态情感识别方法，所述方法包括：步骤S101：接收用户视频图像和皮肤电生理信号数据；步骤S102：分别对所述视频图像和皮肤电生理信号数据进行特征提取；步骤S103：利用深度信念网络对所述视频图像特征和皮肤电生理信号特征进行特征选择和融合，得到多模态融合特征；以及步骤S104：利用分类器对所述多模态融合特征进行分类，从而得到最终的情绪识别结果。

[0006] 示例性地，所述步骤S102包括：针对所述视频图像，检测人脸并标记面部区域关键点，计算所述面部区域各关键点的位置与面部区域关键点平均位置的距离，并针对面部区域各关键点提取尺度不变特征变换(SIFT)特征，以得到视频图像特征向量；以及针对所述皮肤电生理信号，通过低通滤波进行去噪预处理，然后分别计算原始信号及其一阶微分、二阶微分的均值、中值、标准差、最大值、最小值共计五类特征，以得到皮肤电生理信号特征向量。

[0007] 示例性地，所述步骤S103包括：针对所述视频图像特征向量，利用深度信念网络进行特征选择和融合，得到所述视频图像的高层特征；针对所述皮肤电生理信号特征向量，利

用深度信念网络进行特征选择和融合，得到所述皮肤电生理信号的高层特征；以及针对所述视频图像的高层特征和所述皮肤电生理信号的高层特征，利用深度信念网络进行特征选择和融合，得到所述用户的多模态融合特征。

[0008] 根据本发明的一个实施例，所述步骤S104中的分类器为基于径向基函数的支持向量机。

[0009] 另一方面，本发明还提供一种基于视频和生理信号的多模态情感识别系统，所述系统包括：数据接收模块，用于接收用户视频图像和皮肤电生理信号数据；特征提取模块，用于提取用户视频图像特征和皮肤电生理信号特征；特征融合模块，用于对所述视频图像特征和皮肤电生理信号特征进行特征选择和融合，以得到多模态融合特征；以及情绪识别模块，用于对所述多模态融合特征进行分类，从而得到最终的情绪识别结果。

[0010] 示例性地，所述特征提取模块通过如下方法实现：针对用户视频图像，检测人脸并标记面部区域关键点，计算所述面部区域各关键点的位置与面部区域关键点平均位置的距离，并针对面部区域各关键点提取尺度不变特征变换(SIFT)特征，以得到视频图像特征向量；以及针对用户皮肤电生理信号，通过低通滤波进行去噪预处理，然后分别计算原始信号及其一阶微分、二阶微分的均值、中值、标准差、最大值、最小值共计五类特征，以得到皮肤电生理信号特征向量。

[0011] 示例性地，所述特征融合模块通过如下方法实现：针对所述视频图像特征向量，利用深度信念网络进行特征选择和融合，得到所述视频图像的高层特征；针对所述皮肤电生理信号特征向量，利用深度信念网络进行特征选择和融合，得到所述皮肤电生理信号的高层特征；以及针对所述视频图像的高层特征和所述皮肤电生理信号的高层特征，利用深度信念网络进行特征选择和融合，得到所述用户的多模态融合特征。

[0012] 根据本发明的实施例，所述情绪识别模块中的分类器为基于径向基函数的支持向量机。

[0013] 本发明提供的基于视频和生理信号的多模态情感识别方法及系统，采用深度信念网络实现了特征选择过程的自动化，减少了特征选择对人工经验和实验次数的依赖性，对多模态情感识别提供了新思路。

附图说明

[0014] 本发明的下列附图在此作为本发明的一部分用于理解本发明。附图中示出了本发明的实施例及其描述，用来解释本发明的原理。

[0015] 附图中：

[0016] 图1示出了根据本发明的实施例的一种基于视频和生理信号的多模态情感识别方法100的流程图；以及

[0017] 图2示出了根据本发明的实施例的一种基于视频和生理信号的多模态情感识别系统200的结构框图。

具体实施方式

[0018] 在下文的描述中，给出了大量具体的细节以便提供对本发明更为彻底的理解。然而，对于本领域技术人员而言显而易见的是，本发明可以无需一个或多个这些细节而得以

实施。在其他的例子中，为了避免与本发明发生混淆，对于本领域公知的一些技术特征未进行描述。

[0019] 应当理解的是，本发明能够以不同形式实施，而不应当解释为局限于这里提出的实施例。相反地，提供这些实施例将使公开彻底和完全，并且将本发明的范围完全地传递给本领域技术人员。

[0020] 在此使用的术语的目的仅在于描述具体实施例并且不作为本发明的限制。在此使用时，单数形式的“一”、“一个”和“所述/该”也意图包括复数形式，除非上下文清楚指出另外的方式。还应明白术语“组成”和/或“包括”，当在该说明书中使用时，确定所述特征、整数、步骤、操作、元件和/或部件的存在，但不排除一个或更多其它的特征、整数、步骤、操作、元件、部件和/或组的存在或添加。在此使用时，术语“和/或”包括相关所列项目的任何及所有组合。

[0021] 为了彻底理解本发明，将在下列的描述中提出详细的步骤以及详细的结构，以便阐释本发明的技术方案。本发明的较佳实施例详细描述如下，然而除了这些详细描述外，本发明还可以具有其他实施方式。

[0022] 本发明提出一种基于视频和生理信号的多模态情感识别方法及系统，其通过捕捉说话人的视频图像和皮肤电生理信号数据来检测说话人的情绪状态。本发明提供的基于视频和生理信号的多模态情感识别方法及系统仅需要普通的摄像装置、皮肤电采集装置以及相应的软件系统即可实现。

[0023] 图1示出了根据本发明实施例的一种基于视频和生理信号的多模态情感识别方法100的流程图。下面参照图1来具体描述根据本发明实施例的一种基于视频和生理信号的多模态情感识别方法100。

[0024] 根据本发明的实施例的，基于视频和生理信号的多模态情感识别方法100包括如下步骤：

[0025] 步骤S101：接收用户视频图像和皮肤电生理信号数据。例如，用户回答提问者的问题，这时摄像装置录制其视频图像，皮肤电采集装置采集其皮肤电生理信号数据。示例性地，本步骤中用户的视频图像可以通过普通的基于可见光的彩色或灰度摄像装置进行采集，所述摄像装置例如普通摄像头、网络摄像头、手机的前置摄像头等。采集得到视频图像序列由根据本发明实施例的基于视频和生理信号的多模态情感识别系统逐帧进行接收。

[0026] 步骤S102：分别对所述视频图像和皮肤电生理信号数据进行特征提取。当然，本步骤也可以不提取用户视频图像的每一帧的图像特征，而是选择性地提取用户视频图像中的某一帧或某几帧，以减少运算量。也就是说，在本步骤中，提取所述用户视频图像中至少一帧的图像特征。

[0027] 在本实施例中，针对所述视频图像，检测人脸并标记面部区域关键点，计算所述面部区域各关键点的位置与面部区域关键点平均位置的距离，并针对面部区域各关键点提取尺度不变特征变换(SIFT)特征，以得到视频图像特征向量；针对所述皮肤电生理信号，通过低通滤波进行去噪预处理，然后分别计算原始信号及其一阶微分、二阶微分的均值、中值、标准差、最大值、最小值共计五类特征，以得到皮肤电生理信号特征向量。。

[0028] 步骤S103：利用深度信念网络对所述视频图像特征和皮肤电生理信号特征进行特征选择和融合，得到多模态融合特征。在本实施例中，针对所述视频图像特征向量，利用深

度信念网络进行特征选择和融合，得到所述视频图像的高层特征；针对所述皮肤电生理信号特征向量，利用深度信念网络进行特征选择和融合，得到所述皮肤电生理信号的高层特征；以及针对所述视频图像的高层特征和所述皮肤电生理信号的高层特征，利用深度信念网络进行特征选择和融合，得到所述用户的多模态融合特征。

[0029] 步骤S104：利用分类器对所述多模态融合特征进行分类，从而得到最终的情绪识别结果。在本实施例中，所述分类器为基于径向基函数的支持向量机。

[0030] 根据本发明的另一方面，还提供了一种基于视频和生理信号的多模态情感识别系统。图2示出了根据本发明实施例的基于视频和生理信号的多模态情感识别系统200的结构框图。

[0031] 如图2所示，基于视频和生理信号的多模态情感识别系统200包括：数据接收模块201、特征提取模块202、特征融合模块203和情绪识别模块204。其中，数据接收模块201用于接收用户视频图像和皮肤电生理信号数据；特征提取模块202用于提取用户视频图像特征和皮肤电生理信号特征；特征融合模块203用于对所述视频图像特征和皮肤电生理信号特征进行特征选择和融合，以得到多模态融合特征；情绪识别模块204用于对所述多模态融合特征进行分类，从而得到最终的情绪识别结果。

[0032] 根据本发明一个实施例，特征提取模块202可以包括：针对用户视频图像，检测人脸并标记面部区域关键点，计算所述面部区域各关键点的位置与面部区域关键点平均位置的距离，并针对面部区域各关键点提取尺度不变特征变换(SIFT)特征，以得到视频图像特征向量；以及针对用户皮肤电生理信号，通过低通滤波进行去噪预处理，然后分别计算原始信号及其一阶微分、二阶微分的均值、中值、标准差、最大值、最小值共计五类特征，以得到皮肤电生理信号特征向量。

[0033] 根据本发明一个实施例，特征融合模块203可以包括：针对所述视频图像特征向量，利用深度信念网络进行特征选择和融合，得到所述视频图像的高层特征；针对所述皮肤电生理信号特征向量，利用深度信念网络进行特征选择和融合，得到所述皮肤电生理信号的高层特征；针对所述视频图像的高层特征和所述皮肤电生理信号的高层特征，利用深度信念网络进行特征选择和融合，得到所述用户的多模态融合特征。

[0034] 根据本发明一个实施例，情绪识别模块204中的分类器为基于径向基函数的支持向量机。

[0035] 本发明提供的基于视频和生理信号的多模态情感识别系统，采用深度信念网络实现了特征选择过程的自动化，减少了特征选择对人工经验和实验次数的依赖性，对多模态情感识别提供了新思路。因此，该系统易于实现并且在使用上非常灵活和方便。

[0036] 进一步地，根据本发明实施例的上述基于视频和生理信号的多模态情感识别系统所需要的外界输入仅有普通的视频图像序列和皮肤电生理信号数据，并且只需要在屏幕上与用户进行交互，其可以部署在普通个人计算机、智能手机、平板电脑等常见终端上运行，无需特殊硬件，因此对硬件要求较低。

[0037] 本领域的技术人员可以理解，本发明实施例的基于视频和生理信号的多模态情感识别系统200还可以包括上述各种类型的摄像装置和皮肤电采集装置，用于用户视频图像和皮肤电生理信号数据的采集，在此并不进行限定。

[0038] 本发明实施例的各个模块可以以硬件实现，或者以在一个或者多个处理器上运行

的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器(DSP)来实现根据本发明实施例的基于视频和生理信号的多模态情感识别系统中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序(例如，计算机程序和计算机程序产品)。这样的实现本发明的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在存储载体上提供，或者以任何其他形式提供。

[0039] 本发明已经通过上述实施例进行了说明，但应当理解的是，上述实施例只是用于举例和说明的目的，而非意在将本发明限制于所描述的实施例范围内。此外本领域技术人员可以理解的是，本发明并不局限于上述实施例，根据本发明的教导还可以做出更多种的变型和修改，这些变型和修改均落在本发明所要求保护的范围以内。本发明的保护范围由附属的权利要求书及其等效范围所界定。

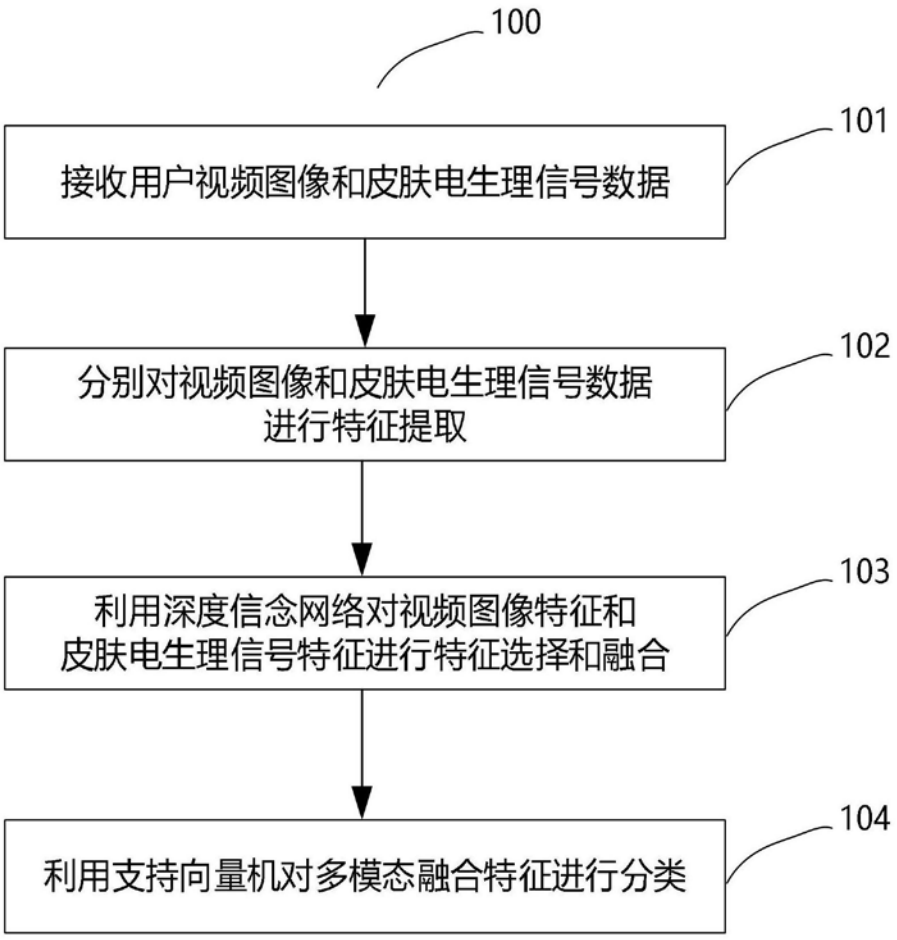
100
101
接收用户视频图像和皮肤电生理信号数据
102
分别对视频图像和皮肤电生理信号数据
进行特征提取
103
利用深度信念网络对视频图像特征和
皮肤电生理信号特征进行特征选择和融合
104
利用支持向量机对多模态融合特征进行分类

图1

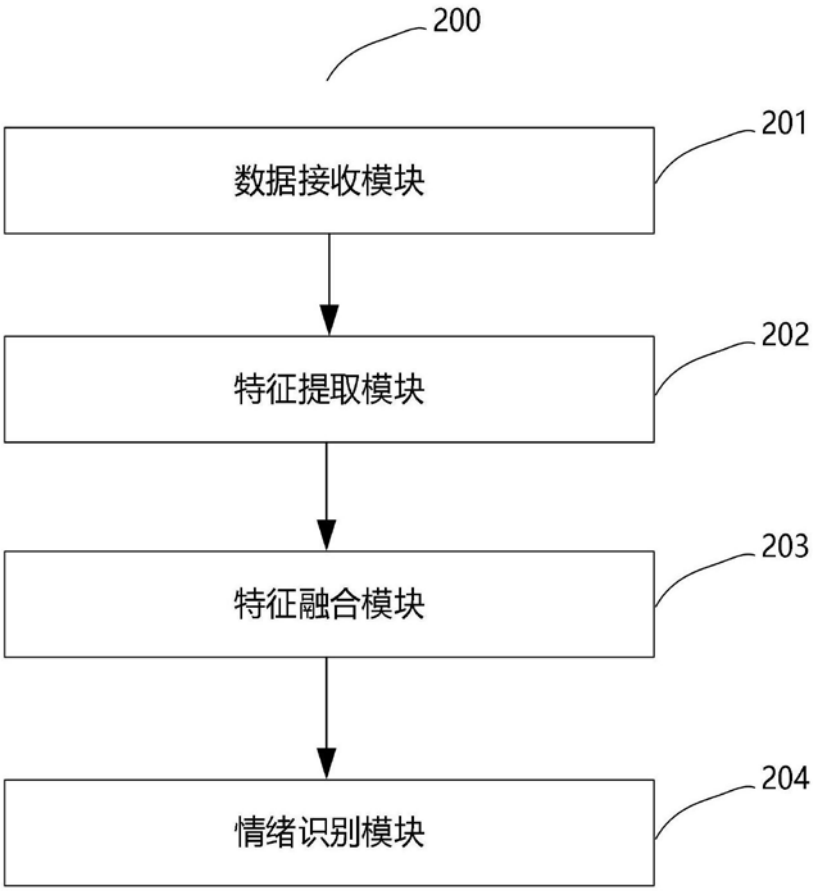

图2