可伸缩 RGB-D 人脸识别的局部表示学习

摘要

本文提出了一种新的基于 RGB-D 学习的人脸识别局部表示方法。该方法的主要贡献是基于数据驱动描述子的图像局部特征提取参考点。我们探讨了深度学习和统计图像之间的互补性作为数据驱动描述符的特性。此外，我们还提出了一种基于利用图像与深度模式互补性的稀疏表示算法-以及使用的数据驱动特性。我们的方法在四个著名的在人脸识别的情况下，证明其对已知挑战的稳健性的基准。产科医生所获得的实验结果与最先进的方法相比具有竞争力，同时提供了可扩展的以及自适应 RGB-D 人脸识别方法。

1. 介绍

多年来，用于人身自动识别的人脸识别技术提供了用户最友好且无创的方式，因此备受关注。对基于标准二维（2-D）图像的人脸识别进行了广泛研究，但仍然存在与成像条件和人脸姿势变化有关的问题. 多亏了三维（3-D）的进步技术，最近的研究已经从 2-D 转变为 3-D(Abbad,Abbad, & Tairi, 2018)。实际上，3-D 人脸表示可确保可靠的表面形状描述，并在面部特征中添加几何形状信息。最近，一些研究人员提议使用从具有成本效益的 RGB-D 传感器（如 MS Kinect 或 Intel RealSense）中捕获的图像和深度数据，而不是笨重且昂贵的 3-D 扫描仪。除了颜色图像，RGB-D 传感器提供描述场景的深度图通过主动视觉或替代技术获得 3D 形状。通过驱动这类传感器的出现和最新的进展深度学习技术，RGB-D 人脸识别现在正在成为是最近几项研究的核心。确实，如今非常清楚的是，例如，使用卷积神经网络(CNN)进行数据驱动的特征提取在许多计算机视觉任务(例如对象检测）（Szegedy，Toshev 和＆Erhan，2013 年），图像方面都优于传统的手工特征。 分类（Krizhevsky，Sutskever 和 Hinton，2012 年）等。当涉及到 RGB-D 人脸识别时，观察到的挑战基本上涉及人脸姿势变化，部分遮挡，成像条件和判别特征提取。

在本文中，我们以可扩展的方式为 RGB-D 人脸分类提出了多模式数据驱动的表示形式，以处理典型问题，例如照明变化，头部姿势变化和受控环境中的伪装。我们的贡献是多方面的：首先，拟议的管道不需要任何关于面部姿势的先验知识，

也不需要在进行识别之前依赖于面部的语义分析。通常，可以检测图像兴趣点（例如 SURF 或 SIFT），从而在不同的观看条件下提供可重复，高效和稳定的结果。从面部图像中提取这些兴趣点，然后简单地提取 RGB-D 色块以获得局部的面部区域而不是整个面部。其次，与手工制作的特征相反，我们建议基于深度学习技术和统计二进制特征来学习判别局部数据驱动的特征，以获得最佳的面部补丁表示。证明了在图像和深度模态上组合这些表示的有效性。 第三，我们提出了一种基于可稀疏表示分类（SRC）方法的补丁匹配算法。 初步，SRC 算法具有一个字典，该字典中充满了画廊中的所有样本。但这会减慢每个补丁与图库中所有补丁的匹配过程。本文其余部分的结构如下。首先，第 2 节概述了相关工作。 然后我们在第 3 节中详细介绍了建议的 RGB-D 人脸识别方法。第 4 节总结了进行的实验和观察到的内容。取得了验证我们的方法的结果。 最后，我们得出结论在第 5 节中进行研究，并提供一些关于真正的工作。

2. 相关工作

在本节中，我们提供与我们的工作密切相关的 RGB-D 人脸识别方法的概述. 可以根据以下三个类别直观地进行讨论. 首先包括为开发姿势不变式解决方案所做的初步努力，其中主要贡献通常集中在预处理部分。这可以用第一个低成本 RGB-D 传感器获取的深度数据质量差来解释。在第二类中，其他解决方案探索了标准手工图像描述符的改编，以表征 RGB-D 人脸数据。 最后，作为最后的趋势，RGB-D 人脸识别技术最近从使用手工功能转变为基于深度学习技术应用学习的功能。

Li，Mian，Liu 和 Krishna（2013）的方法是最早提出的用于 RGB-D 人脸识别的方法之一。预处理包括通过将球体居中于鼻子尖端（从人工选择的最接近传感器的点）对中，从 3-D 扫描中裁剪面部数据。然后，使用迭代最近点（ICP）算法将所有裁剪的面部扫描与通用面部模型对齐，以生成图像和深度数据的规范正面视图。然后将对称填充过程应用于由于非正面姿势中的自闭塞而导致的深度数据丢失。 对于图像数据，判别色空间（DCS）运算符用作特征提取器。然后，在执行后期融合以获得给定探针的最终身份之前，通过应用 SRC 算法分别对预处理后的深度图和 2-D DCS 特征进行分类。

Hsu，Liu，Peng 和 Wu（2014）将 3-D 人脸模型拟合到人脸数据，以为画廊中的每

个人构建 3-D 纹理人脸模型。对于一个新的探针，面部姿势是根据面部标志检测（Zhu＆Ramanan，2012）进行估计的，以便能够将其应用于画廊中的所有 3D 纹理模型。 这允许通过平面投影生成与探头面部姿势相对应的二维图像。 然后，将本地二进制模式（LBP）描述符应用于所有投影的二维图像，以使用 SRC 算法执行分类。类似地，Sang，Li 和 Zhao（2016）使用 ICP 算法将模板面部模型与深度数据对齐，从而从探针样本中估计面部姿势，然后可以将图库中的图像数据渲染为与探针相同的视图 。对于特征提取，将众所周知的“定向梯度直方图”（HOG）运算符应用于图像和深度数据，然后使用“联合贝叶斯分类器”，并根据从图像获得的相似性分数的加权总和得出最终决策和深度分类。

可以清楚地观察到，这些先前的方法集中于预处理，特别是在通过将探针与画廊样本对齐来处理姿势变化时。虽然这种顺序处理可能会导致从姿势估计到分类的错误传播，但仍产生了可喜的结果（Hsu et al., 2014）。另外，为了应对姿势变化，Ciaccio，Wen 和 Guo（2013）建议通过从单个 RGB-D 数据生成一组新图像来完成画廊，这些新图像对应于很大范围的预定义面部朝向。这可以通过以下方式实现。首先，所有面部都被裁剪，然后基于面部标志检测器进行对齐（Zhu＆Ramanan，2012）。然后，每张脸每 5 度围绕 Y 轴旋转一次以渲染新图像。 现在，对于所有画廊样本，包括原始样本和生成的样本，每个人脸图像都由一组 10×10 像素的密集采样小块表示，步长为 5 像素。识别与自遮挡部分相对应的面片，然后根据估计的姿势将其丢弃。其余色块由 LBP 和从像素位置，强度导数和边缘方向计算出的协方差描述符描述。然后，基于特征空间中的欧几里得距离，仅使用每个画廊面部图像的滤波后的补丁执行匹配部分。将所选补丁上的相似性度量整合在一起并进行归一化，以获取探针与图库集之间的最终相似性得分。最后，将结果分数与概率积分进行组合，并执行贝叶斯决策。

在第二类方法中，本概述重点介绍了主要对从 RGB-D 人脸数据中提取特征的部分感兴趣的那些方法。在 Dai，Yin，Ouyang，Liu 和 Wei（2015）中，增强型局部混合导数模式描述符分别应用于从图像和深度数据中提取的二维 Gabor 特征。 该描述符是不同阶的本地导数模式和本地二进制模式的混合特征描述符。为了归因于给定探针的身份，对每个模态分别应用最近邻居搜索算法，并通过组合针对图像和深度模态计算的得分来产生最终相似度得分。在 Goswami，Vatsa 和 Singh（2014）中，人脸由一组根据图像和深度数据计算出的纹理特征和几何属性表示。 对于纹理特征，将 HOG

运算符应用于从图像和深度数据获得的显着性和熵贴图。基于位于深度图上的面部界标之间的欧几里得距离来计算几何属性集。 最后，随机森林分类器用于分类部分。Boutellaa, Hadid, Bengherabi, and Ait-Aoudia (2015)探索了更多的特征组合，将一堆手工制作的特征（例如 LBP，局部相位量化（LPQ）和 HOG）分别应用于 RGB 和深度面部作物，最后使用支持向量机（SVM）分类器 分类。Kaashki 和 Safabakhsh（2018）也探索了类似特征提取器（如 HOG，LBP 和 3DLBP）的用法。 然而，这些描述符被局部地应用在定位的面部标志周围的补丁上。 SVM 分类器也用于分类。在 Hayat，Bennamoun 和 El-Sallam（2016）中，提出了一种图像集分类用于 RGB-D 人脸识别。 对于给定的图像集，首先使用 Fanelli，Gall 和 Van Gool（2011）检测面部区域和头部姿势，然后根据估计的姿势将其聚类为多个子集。 来自 LBP 特征的基于块的协方差矩阵表示可用于对黎曼流形空间上的每个子集建模。 作为分类器，SVM 用于两种模式的每个子集，并做出具有多数表决规则的最终决策。

与手工制作的功能不同，功能学习开始在基于二维图像的方法中以及目前在基于 RGB-D 数据的方法中引起了越来越多的关注。 最近，对于图像集分类，Hayat，Bennamoun 和 An（2014，2015）提出了一种基于自动编码器（AE）的深度学习方法，以为每组图像学习一种称为“深度重构模型”（DRM）的特定于类的模型。 。 在离线阶段，首先使用高斯受限玻尔兹曼机（GRBM）初始化模板深度重构模型（TDRM）权重，然后针对训练图像集的每个类别进行微调。

在测试阶段，给定一个新的探针，使用所有学习到的类别特定模型分别对面部图像和深度数据进行编码和解码。 基本思想是基于对所有学习模型的原始面部数据和重建的面部数据之间的残留误差（即解码器网络的输出）的评估来执行分类。 这种方法的主要缺点是缺乏可扩展性，因为它需要学习一个特定的模型以供任何新人添加到图库中，并且在测试阶段，所有模型都应在输入数据集上运行以评估模型。 重建错误。 这意味着运行时间线性地取决于画廊中的人数。

Lee，Chen，Tseng 和 Lai（2016）建议从图像和深度数据中学习深度特征。 首先，对 CNN 模型进行彩色和灰度面部图像训练。 然后，在深度面部数据上微调获得的模型以进行转移学习。通过将深度像素投影到 3-D 空间上并在一系列处理步骤（例如降噪，深度融合，孔填充，姿势估计，正面化）之后再次执行深度增强步骤，以恢复面部深度图像 对于分类，将支持向量机应用于概率估计，同时考虑深度表示相似

性，头部姿势和数据库相似性标准差，以估计置信度得分并做出最终决策。

Zhang，Han，Cui，Shan 和 Chen（2018）引入了一种新颖的方法，该方法可以处理多模态和跨模态匹配，从而可以测量图像和深度数据之间的相似性。 更详细地，从图像和深度数据中学习了一组互补和共同的特征。 一方面，作者从学习两个基于 Inception-v2 的特定于模态的特征网络开始（Ioffe＆Szegedy，2015 年），然后他们引入了一种联合损失架构，从两个网络中激活以实施补充性的特征学习。另一方面，为了从图像和深度数据中学习异构特征，再次使用特定于模态的特征来获得 RGB 到 RGB 和 RGB 到深度的匹配分数。 最后，将得到的相似性分数与加权和规则相结合。

内图（Neto），马拉纳（Marana），法拉利（Ferrari），贝拉蒂（Berretti）和宾博（Bimbo）（2019）通过从 3D-LBP 图像中学习提出了一种基于深度的面部识别方法。首先，从深度图像中计算出两个 3D-LBP 变体。 然后，设计一个浅层的 CNN 进行分类。 每个描述符图像在最后一个 softmax 层中得到的分数与加权和规则相结合以获得最终决策。

先前的概述清楚地表明，所有这些方法都需要一个预处理步骤，以精确定位面部，估计其姿势，甚至准确检测在连续处理中可能易于错误传播的面部标志，并进一步增加对面部的依赖性。 方法。 例如，Hsu 等。 （2014），Li 等。 （2013），Ciaccio 等。 （2013）和 Sang 等。 （2016 年）的主要目的是通过在不同的视图中生成新图像来通过姿势校正或画廊完成来克服姿势变化。对于数据表示，上述工作（Boutellaa 等，2015； Dai 等，2015； Goswami 等，2014； Kaashki＆Safabakhsh，2018）适应了经典手工描述符（即 HOG）的改编。 ，LBP 等），而可以通过数据驱动的学习技术来提取更合适的特征。后来，随着深度学习技术新时代的到来，诸如 Lee 等人的替代方法也应运而生。（2016），Hayat，Bennamoun 和 An（2015）和 Zhang 等。（2018）开始从学习更多合适的功能中受益，并提高了 RGB-D 人脸识别性能。 与 Hayat 等人相反。 Zhang 等人（2015 年）着重于类内部的紧密性，并没有考虑最大化类间可分离性的类之间的关系。 （2018），Lee 等。 （2016）提出学习识别特征以将整个面部作为输入的多模式识别。虽然只有少数技术将本地学习的特征应用于 RGB-D 人脸识别，但我们的方法着重介绍了如何学习人脸数据中的局部区域的判别式表示，并且在这种情况下显示了与标准手工特征竞争的不可否认的能力 RGB-D 人脸识别。 与全脸图像的全局描述相比，已证明局部特征对许多变化尤其是遮挡具有鲁棒性（Tan，Chen，

Zhou 和 Zhang，2006 年）。通常对于局部特征提取，应用特征检测器以从局部区域提取面部区别信息。通过以固定的步幅在网格中或在检测到的地标周围对输入图像进行采样来裁剪这些区域。在我们的方法中，我们只是考虑在面部上定位一组图像兴趣点，以摆脱面部标志检测或进一步的面部分析。给定的脸部由显着检测到的图像兴趣点周围的一组补丁表示。在提供分类部分之前，使用学习到的描述符（即 CNN 和二值化统计图像特征（BSIF））对这些补丁中的每一个进行变换（Kannala＆Rahtu，2012）。对于 CNN，我们提出了一种有效的训练算法，该算法导致了针对面部补丁表示的可区分空间。此外，尽管大多数深度学习方法通常依赖于深度架构和大量数据的可用性以实现最新的性能，但我们相信，使用较少的参数构建深度深度模型可以有效地学习判别式从小数据中获取本地特征并实现并发性能。 BSIF 是一种流行的统计描述符，用于几种计算机视觉任务。 Boutellaa 等。 （2015 年）证明了低分辨率深度数据在不同的面部分析任务中的有用性，与手工制作的特征相比，具有很高的分类率。在这里，我们展示了结合统计和 CNN 功能来正确描述面部局部斑块的有效性。最后，基于 SRC 算法和动态补丁字典选择，执行补丁之间的对应关系。最终的分类决定通过多数表决规则获得。

3. 建议的 RGB-D 人脸识别方法

图 1 概述了提出的方法的总体流程。它涉及在线和离线阶段共享一些处理块，例如原始数据预处理（即中值和双边过滤），人脸定位，补丁提取和特征向量计算。离线阶段主要致力于训练或更新数据驱动的描述符并构建图库。在线阶段专用于给定面部查询的身份识别。此在线阶段遵循以下步骤。首先，将脸部定位在图像中。然后，通过在面部提取的图像兴趣点周围裁剪的一组补丁来表示。我们考虑了两个数据驱动的描述符，即 CNN 和 BSIF，它们同时应用于输入和库补丁。根据输入特征码的特征向量并使用稀疏表示算法，将输入特征码与图库的特征码进行匹配。此算法在每个补丁上的应用分别产生一组投票。稍后将它们组合以获得输入面的最终标识。本节的其余部分详细介绍了我们建议的方法所涉及的主要模块

3.1 人脸预处理和补丁提取

我们系统的离线和在线阶段之间共享的人脸预处理包括深度图和人脸定位的中值和双边过滤（Zhu＆Ramanan，2012）。 1 对纹理图像执行面部检测，然后将获得的边

界框映射到深度图像。裁剪后的面部区域被调整为 96×96 像素，以确保标准化的面部空间分辨率。为了摆脱人脸地标的定位，我们仅考虑图像兴趣点而无需任何进一步的语义分析，并且不会失去一般性。换句话说，我们不会尝试捕捉特定的面部标志。尽管使用面部标志似乎是直观的，但我们认为，总体而言，在我们的上下文中，图像兴趣点的使用更适合且健壮，可以应对与面部识别相关的所有变化和挑战。从根本上讲，我们认为在头部姿势变化，面部表情下，精确的面部标志定位仍然具有挑战性，并且在面部遮挡下仍不够鲁棒。此外，面部标志检测是一个独立的研究领域，我们希望摆脱管道中的其他依赖。不精确的界标检测错误可能传播到管道的其余部分并影响识别性能这一事实可以证明这一点。相反，图像兴趣点检测不需要对面部区域进行任何高级分析。它相当稳定，直接且快速。

我们使用 SURF 检测器（Bay，Tuytelaars 和 Van Gool，2006 年）在裁剪和调整大小的面部图像上提取图像兴趣点。 SURF 提出了一种有效的比例尺和旋转不变检测器和描述符，使用积分图像和 Hessian 逼近的概念在重复性，独特性，鲁棒性和速度方面优于所有其他特征检测技术（Bay，Ess 和 Tuytelaars，2008； Bay 等） 。，2006）。使用传感器校准将检测到的 SURF 兴趣点的坐标从图像映射到深度数据（示例请参见图 2）。 在每个兴趣点周围，我们从图像和深度数据中提取两个 21×21 像素的色块同样，通过 RGB-D 传感器校准可以确保图像和深度数据之间的映射。BSIF 描述符及其如何适应局部面部表情的重新表示。