

第27届柔性自动化与智能制造国际会议, FAIM2017
2017年6月27日至30日, 意大利摩德纳

一种用于机器人制造业中的复杂视觉识别的机器学习方法

P. Aivaliotis^a, A. Zampetis^a, G. Michalos^a, S. Makris^{a*}

帕特雷大学制造系统与自动化实验室, Rion Patras, 26504, Greece

摘要

本研究提出了一种使用机器学习服务来对部件进行视觉识别的方法, 以实现复杂的部件的操作。由于零件定位的不确定性, 机器人操作复杂零件是一项具有挑战性的应用。这里抓手抓握的不稳定性, 是因为它的非对称和复杂的几何形状所引起的, 因此在抓握之后会导致部件的方向稍微发生变化, 这就超出了处理/组装过程所允许的误差。为了弥补这一点, 我们通过分类器来实现视觉识别方法。最后, 通过一个案例研究, 对消费品的操纵进行了论证和评估。

关键词: 视觉识别, 掌握不稳定性, 复杂零件

1. 引言

目前机器人操纵广泛应用于制造业, 以简化和加速生产处理。随着机器人精度的提高, 可以应用的领域也越来越多。例如现在的手术可以通过比人手更精确和稳定的机器人系统来执行操作[1]。特别是在制造过程中, 任何部件的操作都应该高度重视进行, 准确性这一点变的至关重要[2]。

* Corresponding author. Tel.: +30-2610-910-160; fax: +30-2610-997-774.
E-mail address: makris@lms.mech.upatras.gr

机械式夹具用于大多数抓取操作; 然而, 尽管他们在驱动和传感方面的技术进步, 他们仍然无法稳定地抓住物体复杂的几何形状。

过去, 大量的研究都与提高夹具的抓握能力有关 灵敏度为此评估感知能力对抓取的影响的方法 并且在抓手设计上, 已经开发出了掌握能力图[3]。极端学习机 (ELM) 和支持向量回归 (SVR) 方法也被用于更高的操作精度实现[4]。其他尝试集中在基于几何的抓握质量测量的发展上 可以直接从3D点云[5]确定。

与上述关注部分预抓阶段的研究相比, 所提出的方法更加突出。在掌握后, 识别被操纵部件的位置和方向误差。就这样克服了夹具不稳定性造成的不确定性, 机器人的行为可以调整到执行 操作任务成功。通过机器学习服务进行视觉识别的方法用于此 目的, 因为这种服务可以执行任务, 即快速准确地识别对象 三维空间中的方向和位置。过去开发了大量视觉识别方法。归因于关键步骤 识别是提取低级特征, 它对图像中的局部视觉特征进行编码 提供在属性预测步骤[6,7]中使用的表示。DeCAF是一个开源实现的 深度卷积激活功能, 以及所有相关的网络参数, 以实现实验 跨越一系列视觉概念学习范式的深层表征[8]。使用最佳代码条目 (BCE) 和中级分类权重

(MLCW) 允许改善表示的质量, 同时 保持实时表现。BCE完全专注于稀疏性并提高稳定性和计算性 在编码阶段的效率, MLCW增加了视觉表示的可辨别性, 因此, 系统的整体识别准确性, 通过利用数据监督[9]。

基于上述方法, 本文着重于视觉识别的集成和实现 在工业机器人操纵中的服务, 如下所述。在第2节中, 描述了该方法详情。在第3节中, 该方法的软件和硬件实现是在真实的基础上提出的 工业试点案例。通过更改参数, 多次实现了相同的方案 每次视觉识别服务。因此, 该方法的评估已经在进行, 实验结果 在第4节中有介绍。

2.方法

这项研究的目的是把视觉识别服务作为一个复杂零件的机器人来操纵外部工具, 用于在抓取零件后估算零件的位置和方向。抓握阶段对于成功操作非常关键, 因为夹具引起的部件可能的姿势的偏差而造成其不稳定

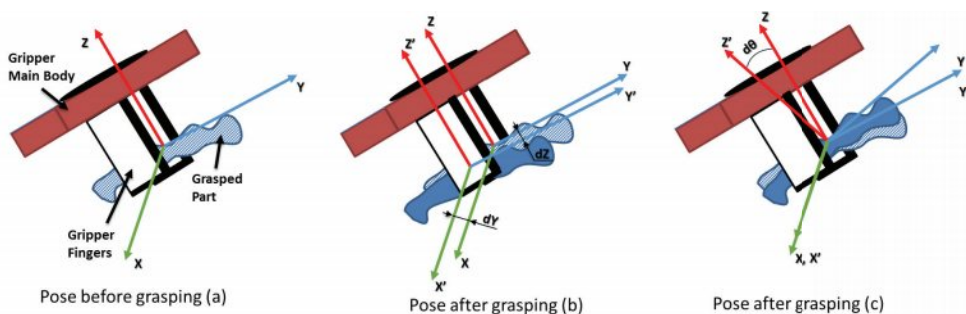


图1.抓握后姿势的偏差。

如图1 (a) 所示，复杂部分在被抓住之前的姿势已经知道并且是固定的， 无论是因为它是由生产过程定向还

是使用其他视觉技术（例如立体相机或 其他）。所有部件的尺寸也被认为是已知的，并且夹具的抓取是理想的， 机器人应该知道该部件的准确姿势。但实际上，如图1 (b) 和 (c) 所示， 夹具的不稳定性与其复杂的几何形状相结合，导致零件在它被抓住之后的方向略微变化。因此，机器人不知道抓握部分的确切实际姿势，而事实并不能达到所要求的准确度，以便执行任务。为了克服这个问题，一个 视觉系统与视觉识别服务相结合，被整合到机器人操作中 处理。使用这些技术执行任务的过程如图3所示。

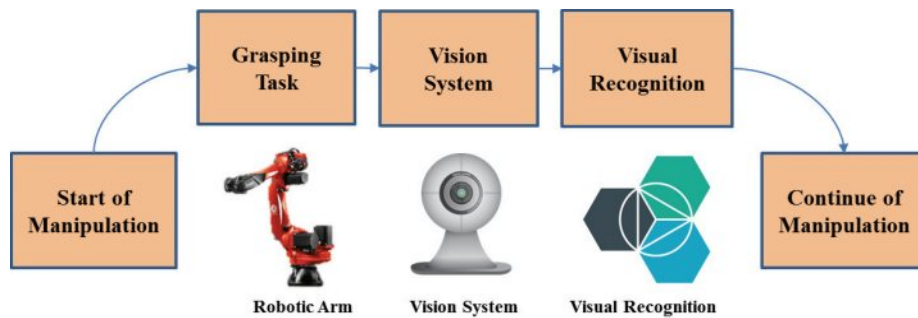


图2.使用视觉识别服务的机器人操作。

大多数的机器学习服务，用于实现视觉识别，是卷积神经网络（CNN / ConvNets）。与普通神经网络相比，ConvNets包含多层感受神经元。最初，这些神经元通过标记数据进行训练，被人类所赋予的权利。在训练阶段之后，每个神经元都准备好执行它的任务，即处理一部分未标记数据，旨在提供有关它们的信息作为输出。最终，所有的输出基于它们的输入区域重叠，以便更好地表示未标记的数据。对每个神经元重复相同的过程，最后，多个标签用于表示输入未标记的数据。这些标签的特征还在于得分，确定答案的置信度。

ConvNets结合了多个神经元簇的输出，使它们比正常情况更快，更准确。神经网络做[10,11,12]。基于ConvNets的视觉识别程序与上面提到的类似。首先，在训练阶段，具有相同特征的一组图像用于创建由用户标记的类，根据其共同特点。多个类用于训练，从而产生分类器。其次，使用此分类器，可以分析未标记的图像。根据ConvNets的信心，得分描绘了已经计算出未标记图像与每个类的匹配率。计算分数，每个类独立，而未标记的图像可以匹配多个类。

最后，假设服务的准确性，得分最高的类是表示输入的正确标签的类图片。整个系统的方法如图3所示。对于类的创建，应该由相机捕获图像集。相机的配置是在此阶段定义并在完整操作过程中保持不变。另外，相机应该是固定，而物体的背景应尽可能清晰，以实现高精度和匹配分数。关于相机的规格，拍摄率，高对比度，饱和度和清晰度都是可以主要提高视觉识别性能的参数。每节课至少需要10节课图像，而代表性标签应该表征它们。

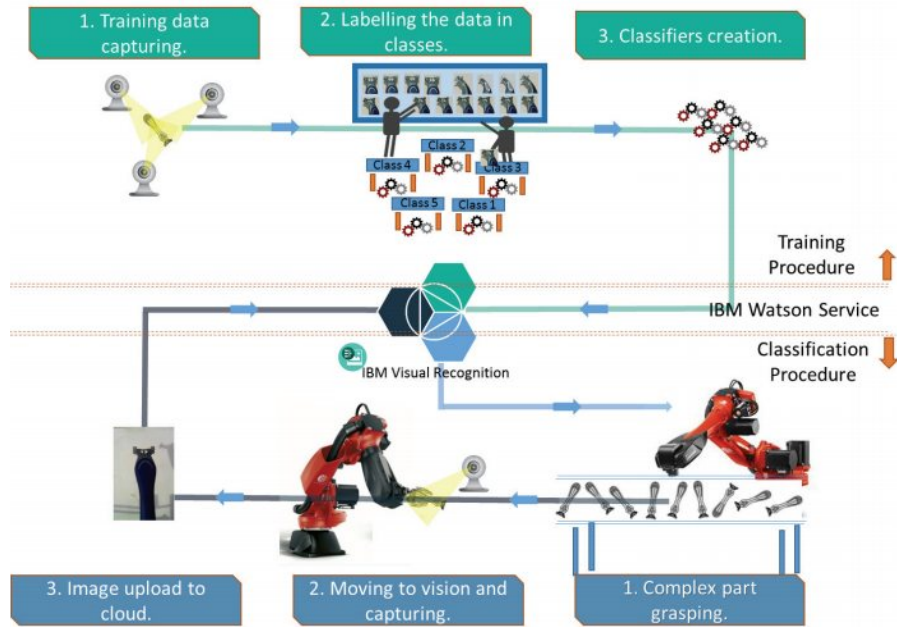


图3.系统方法。

3.系统实施 3.1。软件实施在这个机器学习和视觉识别的特定实现中，它是IBM的Watson Visual 利用认可服务。IBM的Watson是一个基于ConvNets的回答计算系统的问题 体系结构，适用于许多机器学习技术。它包含了大量的处理能力 记忆，使其能够80 TeraFLOPs。结合其视觉识别服务，它可以提供一个 对所提供图像进行非常准确和快速的分类。此外，它还可以选择创建自定义类。因此，它可以很容易地集成到工业机器人操纵过程中。该 服务可通过IBM的Bluemix平台访问，该平台提供许多其他机器学习服务云 服务和虚拟服务器也是如此。沃森视觉识别服务的培训和分类阶段与此类似 该方法中描述的相应程序。最初，创建了所需的图像集。利用一个 简单的代码，在下面的实验中，它是使用Python 2.7.12；图像集分为类和 上传到IBM的Bluemix平台，作为培训阶段的输入。作为训练阶段的输出，IBM的Watson Visual Recognition服务创建了一个具有唯一ID的分类器。那么，一个简单的应用，它可以在本地执行，也可以在Bluemix虚拟服务器上执行，允许对未标记的进行分类 image，使用创建的分类器。Watson视觉识别服务返回了匹配的分数的 每个类的未标记图像输入。得分最高的班级的标签是该班的最终结果 分类。

3.2。硬件实现 为了演示具体方法，建立了机器人操纵的工业试点案例。一个 机械臂，配备平行夹具，用于抓住剃须刀的弯曲手柄。申请是 旨在实现这些复杂零件的喂养。手柄的抓握导致其轻微

旋转。复杂几何零件的操纵是具有挑战性的机器人应用程序由于过多的位移和旋转，即使是几毫米或几度的距离在抓取阶段可能出现的名义位置。因此，手柄的位置和方向被更改，机器人控制器无法获取3D中物体姿势的正确信息空间。为了计算手柄的新姿势，执行了以下程序。首先是机器人如上所述，将把手移动到精确的点，进入固定摄像机的视觉区域。一个led光源被安装在相机旁边的倾斜位置，以便为复杂的部件提供更强大的功能，均匀照明。其次，捕获图像并上传到Bluemix平台以进行处理通过视觉识别服务。最后，将分类结果反馈给机器人告知手柄的新姿势。

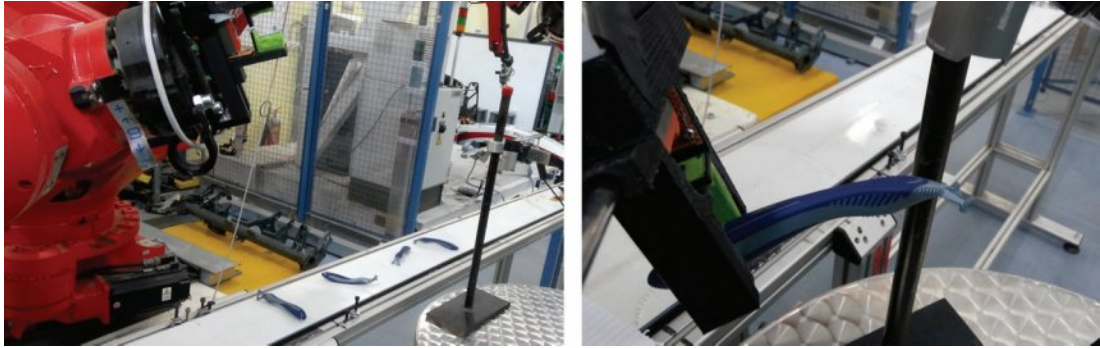


图4.所示方法的演示。

复杂部件操作的整个过程由主PC控制，主PC具有a的作用 设备/服务之间的通信节点。TCP/IP是用于的通信协议 实现这种方法。整个演示的软件序列图如下所示 接下来是图5。

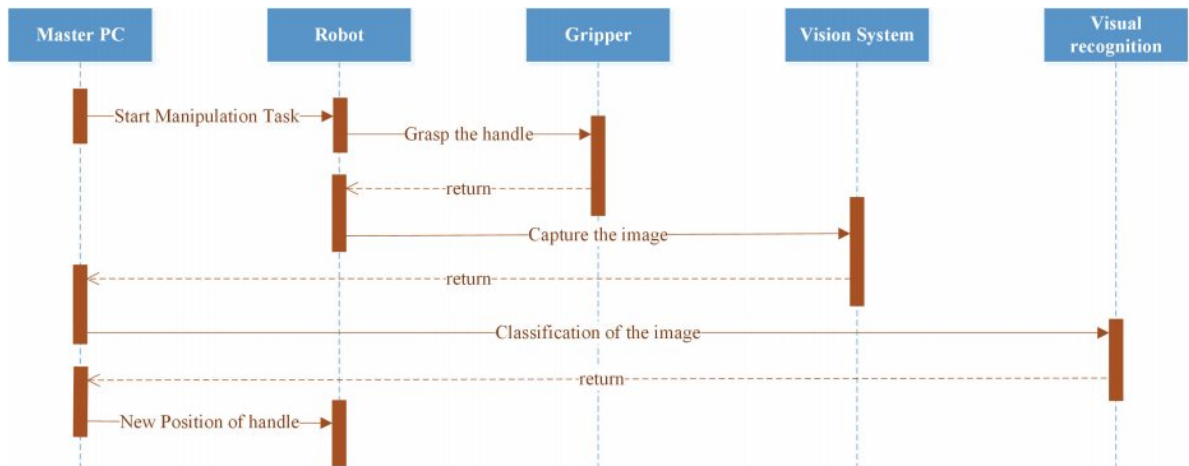


图5.工业场景的实施程序。

4.实验

为了评估该方法，针对视觉的不同参数重复上述实施 识别服务，例如图像分辨率和每类培训中的照片数量。答案得分和答案 每个实验的服务响应时间计算并在本节中介绍。在第一组实验中，类的数量以及每个类的图像数量是稳定的，而视觉系统的分辨率正在发生变化。捕获后的手柄照片被捕获并且 上传进行分类。实验进行了60个等级（围绕x轴旋转15个等级，15个等级 围绕y轴旋转的类，用于Z轴平移的15个类和用于x-中的平移的15个类 轴）和每班30张图像。由于课程数量众多，因此无法为所有课程提供分数 他们。然而，关于其旋转，围绕X轴的手柄分类的代表性示例 抓取后的（ R_x ）如表1所示。最初，手柄处于其标称位置（ $R_x = R_y = R_z = 0$ O，在抓住之前 $D_x = D_y = D_z = 0$ ）。



图6.用于分类的捕获图像。

表1.具有不同图像分辨率的实验

Image Resolution	Answer Score for Class 1: $R_x = -1$	Answer Score for Class 2: $R_x = 0$	Answer Score for Class 3: $R_x = 1$	Answer Score for Class 4: $R_x = 2$	Service Response Time
100%	0.14	0.21	0.35	0.98	4 sec
85%	0.14	0.22	0.38	0.96	3.7 sec
63%	0.19	0.24	0.4	0.89	3.1 sec
41%	0.26	0.3	0.42	0.81	2.4 sec
15%	0.35	0.38	0.47	0.72	1.6 sec
10%	0.4	0.42	0.48	0.67	1.3 sec
5%	0.47	0.49	0.51	0.55	1.5 sec

假设只考虑了这些课程，答案分数最高的课程 在所有情况下都是4级。答案得分取决于图像分辨率。最终，轮换 抓握后的手柄为2度（ $R_x = 2$ ）。如上所述，答案得分随着时间的推移而下降 图像分辨率降低。即使我们考虑服务响应时间的度量，也不意味着 最高分辨率是图像分类的最佳选择。如果没有任何时间限制 操作任务，可以使用最高的图像分辨率。另外，对于图像的最低百分比

£~Æä·ÖÀàÊÇ³É¹µÄ«ÊÇËüµÄËjÖµÖ»ÄÜÄãÄãÇ¿ÇÒÏjËÓÐÄË¿ÉÄÜµÄ·çlÖ

±í 2. Experiments with different number of images in each class.

Number of images in each class	Answer Score for Class 1: Rx = -1	Answer Score for Class 2: Rx = 0	Answer Score for Class 3: Rx = 1	Answer Score for Class 4: Rx = 2	Service Response Time
30	0.14	0.21	0.35	0.98	1.4 sec
25	0.14	0.22	0.37	0.95	1.2 sec
20	0.16	0.24	0.37	0.93	1.12 sec
15	0.16	0.25	0.39	0.92	0.95 sec
10	0.17	0.25	0.39	0.92	1.5 sec

分辨率，分类成功，但所有类别的答案分数值非常接近，从而产生可能出错的风险。实际上，图像分辨率不是唯一的参数，它会对答案得分和服务产生影响 响应时间。在第二组实验中，对每个实验重复上述过程。数字类和图像分辨率都是稳定的，而每个类的图像数量都在变化。该实验发生了60个班级和90%的图像分辨率，同时影响了不同数量的图像 每个班级的表格如表2所示。表2.每个类别中具有不同数量的图像的实验。每个中的图像数量 类 答案分数 第1类：Rx = -1 答案分数 第2类：Rx = 0 答案分数 第3类：Rx = 1 答案分数 第4类：Rx = 2 服务响应时间 三十 0.14 0.21 0.35 0.98 1.4秒 25 0.14 0.22 0.37 0.95 1.2秒 20 0.16 0.24 0.37 0.93 1.12秒 15 0.16 0.25 0.39 0.92 0.95秒 10 0.17 0.25 0.39 0.92 1.5秒 以同样的方式，假设只考虑这些类，似乎数量 每个类中的图像主要影响服务响应时间，而答案分数非常接近所有 实验。

更具体地说，保持相同的图像分辨率，响应时间和正确的分数 答案取决于每个班级的图像数量。随着每个班级使用的图像越多，图像越高 服务的响应时间和正确答案的得分。但是，每班15张图片是一个下限 服务的响应计时器。如果图像的数量，pes类，小于15，则响应时间为 服务增加。考虑到上述实验，最好选择服务 参数可以放置。在所有情况下，图像的分类是成功的，而不同的实验 介绍视觉识别服务参数对答案得分和服务响应时间的影响。5。结论 所提出的方法可以作为克服发生的可能不确定性的良好解决方案 在执行抓取任务期间，抓手的不稳定性。讨论的方法的直接好处 很快讨论了复杂零件的机器人操纵。首先，操作复杂的部分可以通过使用视觉识别服务作为补偿抓取位置误差的工具来管理。可能 由于其非对称和复杂的几何形状，可以管理复杂零件的轻微旋转。该 更改视觉识别服务参数的灵活性是另一个好处，因为用户可以调整它们，取决于可用时间和期望的匹配分数。未来的研究将侧重于通过减少服务响应来增加分类图像的答案分数 时间同时。预计应用程序将更准确并且能够识别图像 也有复杂的背景。最后，未来的工作将旨在增加分类器的数量以及 将它们组合起来，以便能够使用多视觉系统同时分类更复杂的部件。

引用

- [1] Huu M. L., Thanh N. D., Soo J. Ph. A survey on actuators-driven surgical robots. *Sensors and Actuators A: Physical* 2016;247:323-354.
- [2] Chen F., Cannella F., Canali C., Hauptman T., Sofia G., Caldwell D. In-Hand Precise Twisting and Positioning by a Novel Dexterous Robotic Gripper for Industrial High-Speed Assembly. *IEEE International Conference on Robotics and Automation* 2013:270-275.
- [3] Eizicovits D., Tuijl B., Berman S., Edan Y. Integration of perception capabilities in gripper design using graspability maps. *Biosystems Engineering* 2016;146:98-113.
- [4] Petković D., Danesh A. S., Dadkhah M., Misaghian N., Shamshirband S., Zalnezhad E., Pavlović N. D. Adaptive control algorithm of flexible robotic gripper by extreme learning machine. *Robotics and Computer-Integrated Manufacturing* 2016;37:170-178.
- [5] Eizicovits D., Berman S. Efficient sensory-grounded grasp pose quality mapping for gripper design and online grasp planning. *Robotics and Autonomous Systems* 2014;62:1208-1219.
- [6] Danaci E. G., Ikizler-Cinbis N. Low-level features for visual attribute recognition: An evaluation. *Pattern Recognition Letters* 2016;84:185-191.
- [7] Zhanga Ch., Xuea Z., Zhu X., Wangc H., Huang Q., Tiane Q. Boosted random contextual semantic space based representation for visual recognition. *Information Sciences* 2016;369:160-170.
- [8] Donahue J., Jia Y., Vinyals O., Hoffman J., Zhang N., Tzeng E., Darrell Tr. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. *International Conference in Machine Learning (ICML)* 2014;
- [9] Fanello S. R., Ciliberto C., Noceti N., Metta G., Odone F. Visual recognition for humanoid robots. *Robotics and Autonomous Systems* 2016.
- [10] Krizhevsky A., Sutskever I., Geoffrey E Hinton G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in neural information processing systems* 2012;25:1097-1105.
- [11] Zeiler M. D., Fergus R. Visualizing and Understanding Convolutional Networks. *Computer Vision – ECCV* 2014: 8689:818-833.
- [12] Shelhamer E., Long J., Darrell T. Fully Convolutional Networks for Semantic Segmentation. *Computer Vision and Pattern Recognition* 2016; 3431-3440.
- [13] EU VERSATILE Project. <http://versatile-project.eu/>