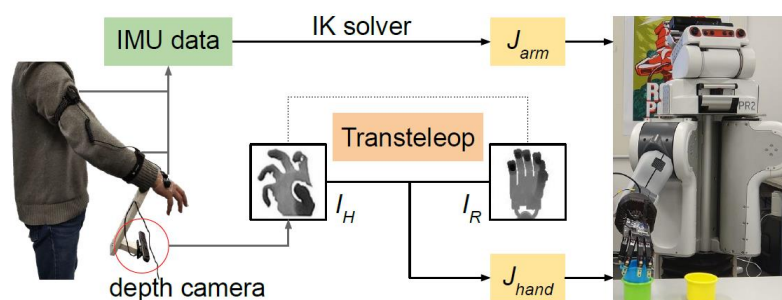


基于视觉和惯性传感器的移动机器人手遥操作系统



论文：A Mobile Robot Hand-Arm Teleoperation System by Vision and IMU

该研究提出了一种多模式移动遥操作系统,该系统由一个新颖的基于视觉的手势回归网络 (Transteleop) 和一个基于 IMU 的手臂跟踪方法组成。Transteleop 通过低成本的深度相机观察人的手,并通过图像到图像的转换过程,不仅生成关节角度,而且还生成配对的机器人手姿势的深度图像。基于关键点重建损失探索了人类和机器人手在外观和解剖结构上的相似之处,并丰富了重建图像的局部特征。穿戴式摄像机支架可实现同时的手臂控制,并促进整个远程操作系统的移动性。测试数据集上的网络评估结果以及除简单的取放操作之外的各种复杂操作任务,显示了多模式遥操作系统的效率和稳定性。

相关视频和代码: <https://Smilels.github.io/multimodal-translation-teleop>

1. 研究问题

拟人化机械手的遥操作以执行灵巧操作仍是挑战。无标记的基于视觉的遥操作具有成本低,侵入性小的强大优势。

问题 a: 由于机器人手和人手占据两个不同的领域,因此如何补偿它们之间的运动学差异在基于无标记视觉的遥操作中起着至关重要的作用。

问题 b: 远程操作员的手应停留在摄像机系统有限的视野范围内。这种限制阻碍了操作员完成需要广阔工作区域的操纵任务。

2. 解决方法

解决方法 a: 提出了一种称为 Transteleop 的基于视觉的新型遥操作方法,该方法基于图像到图像的翻译方法提取配对的人和机器人手之间的连贯姿势特征。Transteleop 将人手的深度图像作为输入,然后估计机器人手的关节角度,并生成机器人手重建图像。本着监督学习的精神,为了增强从图像翻译结构中提取的特征的丰富性,研究人员设计了一个基于关键点重建损失,以专注于手部关键点周围的局部重建质量。

解决方法 b: 为了实现真正的移动式手臂遥控操作系统,研究人员开发了一种摄像机支架,将摄像机安装在人的手臂上。

3. 研究目标

建立一个移动机器人手臂远程操作系统,其中远程操作员可以在无限的工作空间中进行自然的手部动作以完成一系列操作任务。为了建立这样的系统,研究人员制定了一种新颖的基于视觉的方法来对拟人化的手进行遥控操作,并利用基于 IMU 的设备来同时控制手臂。假设 I_H 是人类的图像,展示了通过深度相机观察到的操作任务的手部姿势。视觉部分旨在训练输入 I_H 并预测机器人关节角度 J_{hand} 的神经模型,而 IMU 部分则打算将人手臂的绝对运动映射到机器人手臂。

4. 具体研究方法

4.1 Transteleop 的提出和使用

根据机器人和人的共享姿态特征 Zpose 可以获取 Jhand(机器人关节角度)。文章采用了一种生成结构,该结构从图像 IH 映射到图像 IR,并从该结构的瓶颈层检索姿态作为 Zpose。网络结构如下图所示:

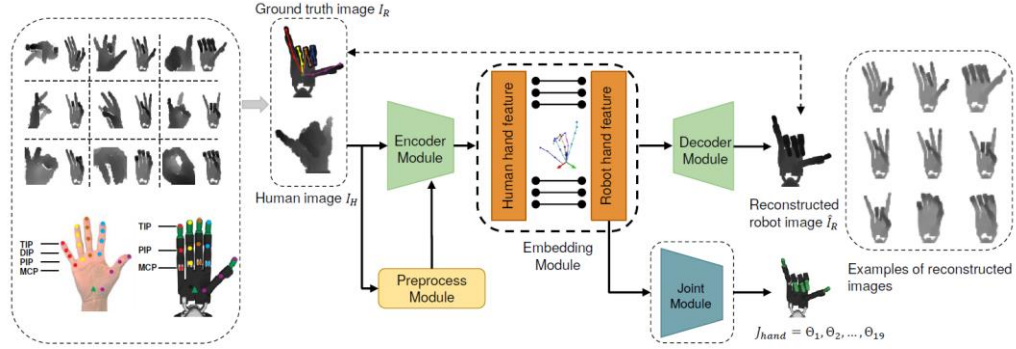


Fig. 2. Transteleop Architecture. Left: Examples of paired human-robot hand datasets used to train Transteleop. Center: The encoder-decoder structure is an image-to-image translation branch, which feeds depth images of a human hand I_H and produces reconstructed depth images of the robot hand I_R . The joint module takes the pose embedding from the encoder-decoder structure and predicts the robot's joint angles J_{hand} . The preprocess module is a spatial transformer network which explicitly permits the spatial manipulation of input images. Right: Examples of reconstructed robot images are generated from Transteleop, whose inputs are the dataset examples shown on the left.

编码器-解码器模块: 编码器从各种角度拍摄人手 IH 的深度图像,并发现人手和机器人手之间的潜在特征 Zpose。研究人员使用六个卷积层,其中包含四个下采样层和两个具有相同输出尺寸的残差块。因此,给定大小为 96×96 的输入图像,编码器将计算抽象 $6 \times 6 \times 512$ 维特征表示。考虑到数据集集中的 IH 和 IR 中的像素区域不匹配,文章中使用完全连接的层而不是卷积层连接编码器和解码器。

解码器旨在根据潜在姿势特征 Zpose 从固定视点重建机器人手的深度图像。一层全连接层将要素从 Zpose 连接到机器人要素向量 ZR。紧跟着的是具有学习滤波器的四个上卷积层和一个用于图像生成的卷积层。与其他图像翻译不同的是本文更加关注局部特征(例如指尖的位置)的准确性,而不是全局特征(例如图像样式)的准确性。研究人员设计了一个基于关键点重建损失,以捕获手的整体结构,并专注于手的 15 个关键点周围的像素。每个像素误差的比例因子取决于该像素与所有关键点的距离,并认为每个关键点的八个相邻像素与这些关键点本身一样重要。

4.2 手臂支架设计

手臂的移动比较大,会导致遥控操作员的手很容易从摄像机的视野中消失,研究中通过便宜的 3D 打印摄像机支架解决了这个问题,该摄像机支架可以安装在远程操作员的前臂上,因此,相机将随手臂一起移动。

由于摄像机位置的不确定性,研究中使用惯性动作捕捉设备控制机器人的手臂,将可穿戴设备的全局坐标系设置为与机器人基坐标系平行。根据获取的旋转数据和机器人手臂的链接长度,计算出机器人的手腕姿势。通过将该姿势输入到 BioIK 解算器中来计算机器人手臂的关节角度。此后,研究人员通过计算和缩放当前帧和前一帧的期望关节角度之间的前馈关节差异以及期望关节角度之间的反馈关节差异来设置每个关节的角速度。

$$\mathcal{V}_{n,t} = \delta_1 \cdot (J_{n,t}^{ik} - J_{n,t-1}^{ik}) + \delta_2 \cdot (J_{n,t}^{ik} - J_{n,t}^{robot})$$

5. 实验分析

输入的深度图像是从原始深度图像中提取出来的,大小调整为 96×96 。为优化网络,研

究中使用批训练随机梯度下降法，并应用 Adam 优化器，其学习率 0.002 和动量参数设置为 0.5 和 0.999，并在每个卷积层之后添加一个批处理归一化（BN）层和一个修正线性单元（ReLU）。

(1) 网络框架对比

为了评估 Transtelep 是否可以学习指示性的视觉表示，研究中将 Transtelep 与两种网络结构进行了比较：TeachNet 和 GANteleop。为了显示来自机器人自身域的回归结果，研究中训练了一个模型 Robotonly，该模型删除了 Transtelep 中的解码器模块，仅提供了机器人手的图像。为提供输入图像的空间变换功能，研究人员在 Transtelep 上增加了 STN 网络。

研究中使用手姿势估计中的标准指标评估了测试数据集上 Transtelep 和四个基线的回归性能：a.最大关节角度误差低于阈值的帧比例；b.最大关节距离误差低于阈值的帧比例；c.在所有角度上的平均角度误差。结果如下图所示：

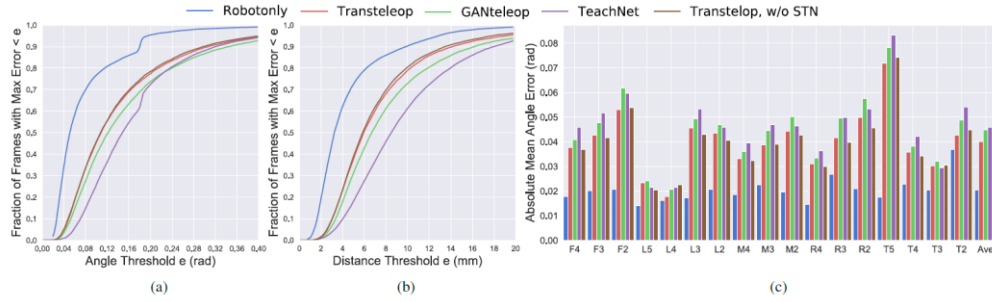


Fig. 5. (a) and (b): The fraction of frames whose absolute maximum joint angle/distance error is below a threshold between the Transtelep approach and different baselines on our test dataset. (c): Comparison of the absolute average angle error on the individual joint between the Transtelep approach and different baselines on our test dataset. F means the first finger, L means the little finger, M means the middle finger, R means the ring finger, T means the thumb. 2, 3, 4, 5 mean the n-th joint of the finger.

实验结果分析：

由于具有匹配的域和相同的视角，因此，Robotonly 模型在所有评估指标上均明显优于其他基准。Transtelep 和 GANteleop 均在最大关节角度以下显示了平均 10.63% 的精度的提升，该精度高于 TeachNet。研究人员推断，这两种图像到图像的转换方法都比 TeachNet 抓住了机器人更多的姿势特征。此外，GANteleop 之所以比 Transtelep 差，是因为 GANteleop 中的判别网络专注于追求逼真的图像，并削弱了对关节的监督。比较 Transtelep 和不使用 STN 的 Transtelep，即使使用了 STN 网络，也没有明显的改进。这表明附加的空间变换仅为该任务带来了一点外观标准化效果，但并未显著促进手部姿势变换为规范姿势。如上图(c)所示，所有方法联合回归的绝对平均误差均低于 0.05 rad，最高误差发生在拇指关节 5 上，因为人的拇指与阴影拇指之间存在很大差异。

(2) 实验操作设置：

系统地评估了四种类型的物理任务，对多模型远程操作方法进行了分析，涉及精度和功率控制的分析，有力和无力操纵以及双臂交接任务的分析。

手臂的速度控制频率为 20Hz，人手臂的起始姿势始终与机器人手臂的起始姿势一致。同时，机器人的手臂在每个任务上总是以几乎相似的姿势开始和结束，手部轨迹控制的频率设置为 10Hz。一名女性和两名男性测试人员参加了以下机器人实验，并且其中每一项任务都是随机执行的。

1)取放。研究人员准备了两个测试方案：挑选一个薯条罐并将其放在同一张桌子上的红色碗中；在桌子上选择一个立方体，并将其放在方块的顶部。第一种情况需要机械手的动力掌握技巧，第二种情况需要机械手的精确掌握技巧，并为远程操作员提供足够的工作空间。

2)插入杯子。三个同心杯应相互插入。此任务检查精确掌握和释放的能力。

3)推动物体。研究人员设置方块的随机初始姿势，然后将方块推入指定的位置。此任务

包含推动，滑动和精确抓握的挑战。

4)双臂交接。左臂将一卷纸交给右手。操作员还利用惯性可穿戴设备控制 PR2 的左臂和左夹具。此任务测试整个远程操作系统的协调能力。由于系统的可移动性，人们可以面对面坐着而不是与机器人平行，从而获得清晰的视野。

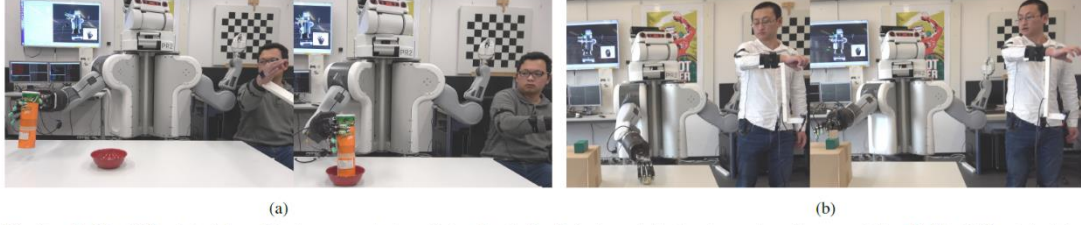


Fig. 6. (a) The PR2 robot picks a Pringles can and places it in a bowl. The Pringles and the bowl are set on the same table. (b) The PR2 robot picks a wooden cube on the table and places it on a rectangular brick on a box. The height of the box is 325 mm.



Fig. 7. (a) The PR2 robot inserts three cups into each other. (b) The PR2 robot hands a roll of paper from its left gripper over to its right hand.

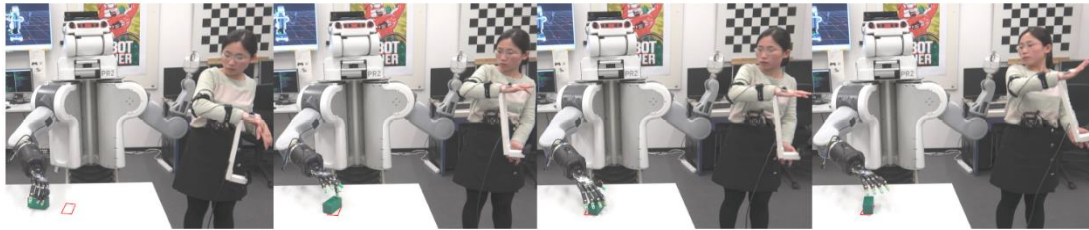


Fig. 8. The PR2 robot pushes a rectangular brick to a specified goal. The red rectangular on the table represents the target pose of the brick.

实验练习：操作员在实际测试之前对每个任务进行了五次非连续尝试的预热训练阶段。对于诸如拾取和放置之类的简单任务，经过三轮试验，操作员可以很好地完成任务。但是对于移交任务，遥控操作员进行了更多尝试，以适应手臂的相反操作方向。每个任务由一名演示者执行了五次。下表为实验结果：

TABLE I
AVERAGE COMPLETION TIME AND SUCCESS RATE OF EACH TASK

	pick1	pick2	cup	pushing	handover
Ave. time	18.5	37.2	25.5	62.0	36.33
Ave. success rate	100%	100%	100%	80%	60%

上表显示了远程操作员完成任务所花费的平均时间以及成功率。完成时间是在机器人开始移动直到返回初始姿势时计算得出的。两项拾取和放置任务的成功率高且完成时间短，并且杯子插入任务表明系统具有精确性和力量把握能力。与拾取和放置任务相比，方块比碗小得多，因此机器人需要更长的时间才能找到放置立方体的宝贵位置。在推入任务期间，机器人可以使用多个手指将方块快速推入目标位置。然而，为了使推入误差小于 5mm，操作者花费了很长时间来处理砖的方向。移交任务的成功率较低，主要是因为对左夹具的控制不精确，从而导致机器人意外丢失了物体。研究人员认为：基于视觉的方法比基于 IMU 的方法更适合于多指控制。

6 总结和未来工作

本文结合了基于视觉的联合估计方法，Transteleop 和基于 IMU 的手臂遥操作方法，提出了一种手臂远程传送系统。

研究不足和研究计划：首先，摄像机支架给操作员带来了额外的负担，这在长期遥控操作过程中不舒服。其次，研究中缺乏手势控制，因此一些高精度的任务（例如开瓶和拧紧螺丝）对于当前系统仍然是棘手的。因此，研究人员计划收集一个手部运动数据集，更多地集中在拇指、无名指和中指的微妙姿势上。最后，研究人员希望通过滑动检测和力估计以减轻用户的控制负担并避免机器人的意外碰撞。

发布链接：<https://mp.weixin.qq.com/s/mZFMBdDWp2HpJts2RVrsgA>