频谱域中的三维直接视觉伺服

Maxime Adjigble[1] Brahim Tamadazte[2] Cristiana de Farias[1] Rustam Stolkin[1] Naresh Marturi[1]

[1] 极限机器人实验室，伯明翰大学，埃德巴斯顿，B15 2TT，英国。[2] 索邦大学，法国国家科学研究中心 UMR 7222，法国国家健康与医学研究院 U1150，智能系统与机器人研究所，F-75005，巴黎，法国。

m.k.j.adjigble@bham.ac.uk

arXiv:2304.00892v1 [cs.RO] 3 Apr 2023

摘要

本文提出了一种三维直接视觉伺服方案，利用频谱域视觉信息对点云（或物体）进行自动对齐。具体而言，我们提出了一种适用于三维模型/点云的对齐方法，通过谐波域数据分析，估计参考点云与目标点云之间的全局变换。本文采用 $\mathbb{R}^3$ 中的三维离散傅里叶变换（*TFD*）估计平移，并采用 $\boldsymbol{SO(3)}$ 中的实球谐函数估计旋转。该方法使我们能够推导出一个六自由度解耦视觉伺服控制器。随后，我们展示了如何将该方法用作机器人手臂的控制器，以执行定位任务。与现有的三维视觉伺服方法不同，我们的方法能够处理部分点云，也适用于初始位置与期望位置之间存在较大初始变换的情况。此外，与使用空间数据不同，使用频谱数据估计变换使我们的方法能够抵抗传感器引入的噪声和部分遮挡。我们已在实验中成功验证了该方法，所用点云由安装在机器人手臂上的深度相机采集。

关键词

点云、视觉伺服、位姿计算、三维配准、傅里叶变换

摘要

本文提出一种直接的 *3D* 视觉伺服方案，利用频域中的视觉信息自动对齐点云（或物体）。具体而言，我们提出一种 *3D* 模型/点云对齐方法，通过谐波域数据分析估计参考点云与目标点云之间的全局变换。本文在 $\mathbb{R}^3$ 中使用三维离散傅里叶变换（*DFT*）进行变换

估计，并在 $\boldsymbol{SO}$（**3**）中使用实球谐函数进行旋转估计。该方法使我们能够推导出一种具有 *6* 个自由度的解耦视觉伺服控制器。随后，我们展示如何将该方法用作机械臂控制器，以执行定位任务。与现有的 *3D* 视觉伺服方法不同，我们的方法在点云不完整以及初始位置与期望位置之间存在较大初始变换的情况下仍能良好工作。此外，在变换估计中使用频域数据（而非空间域数据），使我们的方法能够抵抗传感器引入的噪声和部分遮挡。我们已通过实验，成功验证了该方法在由安装于机械臂上的深度相机获取的点云上的有效性。

关键词

点云、视觉伺服、位姿估计、三维配准、傅里叶变换

1 引言

在过去三十年中，人们日益重视视觉伺服方法，以在工业、国防、自动驾驶车辆、医学等多个领域执行机器人任务。视觉伺服是指利用连续视觉反馈对系统进行动态控制。因此，经典视觉伺服控制器的主要组成部分包括视觉特征提取、特征匹配以及随时间跟踪。然而，经典视觉伺服方法的可行性和效率与视觉信息的检测、匹配和跟踪能力密切相关，而这些能力在某些情况下可能受到限制（纹理稀少的图像、缺少显著几何形状、遮挡等）。近年来，出现了不依赖视觉跟踪的视觉伺服方法。它们直接在控制回路中利用图像的全局信息。这些方法称为直接视觉伺服 [1]方法。文献中研究了不同类型的全局信息，例如强度

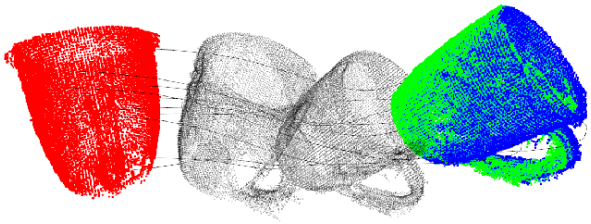

图1 – 使用所提出的方法进行模型配准的示意图。红色点云为参考模型，灰色点云是在收敛过程中产生的中间候选点云，绿色点云为目标场景的点云，蓝色模型则是与目标配准后的最终模型。灰色轨迹曲线表示模型的收敛过程。

像素 [2, 3], 时空梯度 [4], 直方图 [5], 互信息 [6], 高斯混合 [7], 等。最近，一些作者提出利用小波 [8] 和剪切波 [9]获取时频视觉信息。然而，与传统方法相比，直接方法的收敛范围显然更窄。为了解决这一问题，一些研究提出使用以频谱域表示的视觉特征。这些特征已被证明具有较强的抗噪声能力，并用于计算机视觉和机器人学中的多种应用，例如图像相关 [10], 、深度传感器生成模型的配准 [11],、机器人抓取 [12], 等。在 [13], 中，离散余弦变换（DCT）系数被用于直接视觉伺服；在 [14, 15],中，傅里叶变换的时移性质被用于一种视觉伺服解耦控制方案。

以上介绍的大多数视觉伺服方法都基于二维视觉信息。然而，直接在控制回路中使用三维信息的视觉伺服方法却很少见 [16, 17]。使用三维视觉信息的一个优势是，无需经过位姿估计或标定流程，即可直接表达控制器。尽管这些方法取得了令人鼓舞的结果，但它们需要密集的深度数据，而且收敛能力有限。文献 [18] 介绍了一种虚拟视觉伺服方法，该方法使用由点云离线生成的多边形网格。

本文介绍了一种直接的三维视觉伺服方法，利用以频谱域表示的视觉信息。据我们所知，尽管已有极少数研究在二维视觉伺服控制律中使用频谱信息 [13–15, 19],，但文献中尚无研究使用源自点云的三维频谱信息。我们的方法的核心思想是估计参考点云与另一个目标之间的空间变换（6 个自由度）。两个点云均被变换至频谱域；随后通过傅里叶分析估计平移，通过球面相关估计旋转。接着采用优化过程，迭代地最小化平移和旋转的代价函数。值得注意的是，物体的平移估计与旋转估计彼此独立，因此我们能够构建完全解耦的控制器。所提出的方法在笛卡尔空间中使用三维快速傅里叶变换 $\mathbb{R}^3$，并在单位球面 $\boldsymbol{S}^2$ 和旋转群 $\boldsymbol{SO(3)}$上使用实球谐函数，分别计算平移代价和旋转代价的梯度。

图1展示了在简单场景中配准物体的一个示例，其中一个参考模型与包含同一物体的场景点云进行配准。当前场景或目标场景的点云由静态或安装在机器人手臂上的深度传感器在线采集。

本文的主要贡献如下：

— 我们提出一种基于频谱域的新方法，用于对3D物体模型进行配准，即估计两个点云之间的全局平移和旋转。— 我们提出一种新的6自由度直接3D视觉伺服方法，直接使用以频谱域表示的点云（完整或部分）。

所提出的方法具有多项优势。与需要密集且完整深度数据的现有方法不同，我们的方法能够利用部分物体点云高效运行。与使用深度数据的直接视觉伺服方法相比，我们的方法具有更大的收敛域。使用频谱数据使我们的方法天然具备抗噪能力，尤其是在使用3D距离传感器采集的点云时。由于无需颜色信息或强度信息，所提出的方法适用于无纹理物体或弱光照环境。最后，所提出的方法既可用于将物体模型配准到包含多个不同物体的场景中，也可用于在任务空间中定位机器人操作器。

2 方法论

在本节中，我们介绍频谱域中的直接 3D 视觉伺服方法。如前所述，我们方法的主要基础是

通过频谱域中的变换后点云进行物体模型对齐策略。基于这一思路，我们首先介绍该方法所采用的表示方式，接着阐述笛卡尔空间和单位球面上的相位相关概念。最后，介绍模型的迭代对齐控制律。

2.1 点云表示

我们的视觉伺服流水线的第一步，是分别将点云中的点和表面法线表示为体素网格和扩展高斯图像（IGE，即 Extended Gaussian Image）。

以体素网格形式表示的点。点云离散化是一个简单的过程。给定一个由 N 个点组成的点云，可以构建一个分辨率为 $r \in \mathbb{R}^+$ 的 3D 体素网格。因此，对于点云中的每个点 $p = (x,\ y,\ z)$，该点 $p_{ijk} = (i,\ j,\ k)$ 所在体素的索引计算如下：

$$i = [x/r] \qquad j = [y/r] \qquad k = [z/r] \tag{1}$$

运算 $[./.]$ 表示整数除法，即只保留除法结果的整数部分。定义 $v_t : \mathbb{R}^3 \to \mathbb{N}^3$ 为笛卡尔坐标与体素索引之间的映射函数。点云的体素网格函数 [1] $f_t : \mathbb{R}^3 \to \mathbb{N}$ 定义如下：

$$f_t(p) = f_t(x, y, z) = v_{ijk} \tag{2}$$

其中 $v_{i,j,k} \in [0, 1]$。这里，$v_{i,j,k} = 1$ 当点云中至少有一个点的索引等于 $v_T(p) = (i, j,$ {v24) 时，以及 $v_{ijk} = 0$ 在其他情况下。我们的方法使用实值体素网格；不过，也可以用同样的方式使用二值体素网格。[21] 中介绍的 LoCoMo（局部接触矩）评分，可能是提升体素所含信息量的良好候选方案。

作为扩展高斯图像（IGE）的表面法线。IGE 是一种常用表示，适用于定义在单位球面上的函数。在相关文献中，它被广泛用作物体表面法线的形状描述符 [12, 22–24]。利用式（3），将表面法线 $n = (n_x,\ n_y,\ n_z) \in \mathbb{R}^3$ 从笛卡尔坐标转换为球坐标 $n = (r,\ \theta,\ \phi)$，即可将表面法线表示在单位球面上。

$$r = \sqrt{n_x^2 + n_y^2 + n_z^2} \qquad \theta = \arctan \frac{\sqrt{n_x^2 + n_y^2}}{n_z} \qquad \phi = \arctan(\frac{n_y}{n_x}) \tag{3}$$

由于所有表面法线都是单位向量，因此径向距离 $r = 1$ 对所有表面法线都成立。因此，集合 (θ, ϕ) 足以描述单位球面上表面法线的分布。为进行数值计算，需要对球面进行离散表示。经纬度方向采用以下离散化方式：$\theta_j = {}^{\pi}{}_{(2j+1)\ 4B}$ 和 $\phi_k = \frac{\pi k}{B}$，$(j, k) \in \mathbb{N}$，并满足约束条件 $0 \leq j, k < 2B$ 和 $B \in \mathbb{N}$，其~~中 B~~ 为带宽。~~带~~宽通常取 2 的幂次，这意味着 $B = 2^n$，$n \in \mathbb{N}^+$。因此，点云表面法线的 IGE 可表示为函数 $f_r : \boldsymbol{S}^2 \to \mathbb{N}$：

$$f_r(\theta, \phi) = c_r(\theta_j, \phi_k) \tag{4}$$

其中，$c_r \in \mathbb{N}$ 表示点云中经离散化后经度和纬度均为（θ_j，ϕ_k）的表面法线数量。在这种情况下，使用整数值而非二进制值。其优点在于，单位球面上的表面法线分布比简单的二进制分布能提供更多关于物体几何形状的信息。图2展示了一个物体以及包含多个物体的场景的 IGE 示例，这些示例均以点云的形式表示。

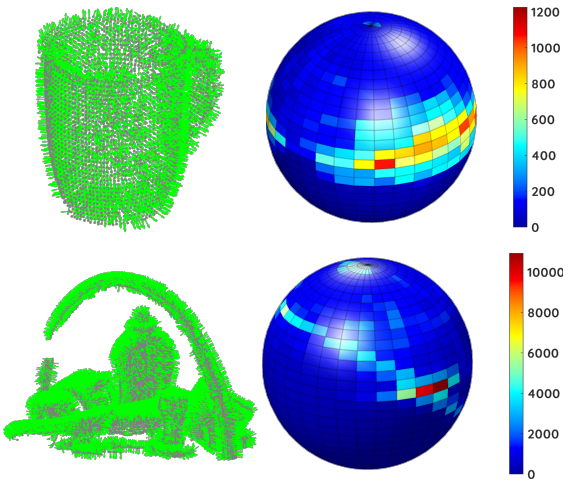

图2 – 多个物体的杯子（上）和场景（下）的 IGE，它们以带有表面法线（小绿色箭头）的点云形式表示。

1. 索引 t 表示该函数用于平移估计；同样，索引 r 将用于与旋转估计相关的函数。

2.2 通过傅里叶分析在 $\mathbb{R}^3$中的平移估计

利用频谱域中的三维相位相关和傅里叶分析，估算两个点云（目标点云和参考点云）之间的平移。基于傅里叶分析的方法的主要优点是能够抵抗不同类型的噪声 [11, 15]。相位相关法基于傅里叶时间平移性质，可将笛卡尔空间中的平移转换为频谱域中的相位位移。

定义 f_t： $\mathbb{R}^3 \to \mathbb{N}$ 为物体或场景点云的体素表示。f_t 的傅里叶系数计算如下:

$$F_t(u,v,w) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\sum_{z=0}^{L-1} f_t(x,y,z)e^{-i2\pi(\frac{u}{M}x+\frac{v}{N}y+\frac{w}{L}z)} \tag{5}$$

其中， M， N， $L \in N^+$ 分别是X、 Y和 Z 上傅里叶系数分解的最大阶数，而（u， v， w）是频谱域中的对应坐标。假设物体或场景平移了 $T = (\tau_x, \tau_y, \tau_z) \in \mathbb{R}^3$，且$g_t$： $\mathbb{R}^3 \to \mathbb{N}$ 是平移后点云的新体素表示。根据时移性质， g_t的傅里叶系数 G_t 可通过以下公式计算:

$$G_t(u,v,w) = F_t(u,v,w)e^{-i2\pi(\frac{u}{M}\tau_x+\frac{v}{N}\tau_y+\frac{w}{L}\tau_z)} \tag{6}$$

平移估计的目标是在已知 f_t 和 g_t的情况下求出 T。具体而言，可以先计算 F_t 和 G_t 的归一化互功率谱 $\mathcal{C}_t$，然后利用式（7）应用傅里叶逆变换来实现。

$$\begin{aligned} \mathcal{C}_t(u,v,w) &= \frac{F_t(u,v,w)\overline{G_t(u,v,w)}}{|F_t(u,v,w)\overline{G_t(u,v,w)}|} \\ \delta(\tau_x,\tau_y,\tau_z) &= \mathcal{F}^{-1}(\mathcal{C}_t(u,v,w)) \end{aligned} \tag{7}$$

其中，$\overline{\ }$ G_t 是 G_t 的复共轭， $\mathcal{F}^{-1}$ 是傅里叶逆变换。结果 δ 是狄拉克 ***delta*** 函数，其最大值在横坐标轴上的位置对应于平移 T。因此，可以通过最大化函数δ来求得平移 T。

$$T = \nabla_{glob}T = \text{argmax}\{\delta(\tau_x,\tau_y,\tau_z)\} \tag{8}$$

尽管可以直接求得平移的全局解 $\nabla_{glob}T$，但在 3D 视觉伺服的语境下，每次迭代只会使用向量的一小部分，该部分等于 $\nabla T = \lambda_t \nabla_{glob}T$，其中 $\lambda_t \in \mathbb{R}^+$ 和 $\lambda_t < 1$。这样既能同时估计平移和旋转，也能控制控制器的收敛速率。代价函数 $J_t(T)$ 可以构造如下函数，用于评估平移估计算法在 $\mathbb{R}^3$ 上的性能:

$$J_t(T) = \frac{1}{2}||g_t(x) - f_t(x+T)||^2 \tag{9}$$

2.3 利用 S^2上的傅里叶分析估计旋转

与平移一样，也可以使用谱分析来估计两个点云（目标和参考）之间的旋转。这里，利用数据在单位球面上的酉表示来编码物体的表面法线信息。在这种情况下，我们通过 IGE相关性估计全局旋转。通过直接寻找使相关性达到最大值的旋转，可以求得“精确”的旋转，但这需要计算一个计算量很大的二重积分。作为解决方案，我们采用梯度优化，迭代计算使两个点云之间的相关性最大的旋转。

在 $\boldsymbol{S}^2$ 和 $\boldsymbol{SO}$(3)中。设 $f_r : \boldsymbol{S}^2 \to \mathbb{N}$ 为物体表面法线的 IGE。由于函数 f_r 的取值属于集合$\mathbb{N} \subset \mathbb{R}$，因此可使用 [25], 中引入的 $\boldsymbol{SO}(\boldsymbol{3})$实调和分析来计算傅里叶系数。给定带宽 B， f_r 在 $\boldsymbol{S}^2$ 中的傅里叶变换定义为:

$$f_r(\theta,\phi) = \sum_{l=0}^{B-1} (F_r^l)^T S^l(\theta,\phi) \tag{10}$$

其中， $F_r^l \in \mathbb{R}^{(2l+1)\times 1}$ 是傅里叶变换的系数，$S^l \in \mathbb{R}^{2l+1}$ 是在 $\boldsymbol{S}^2$ 上定义的实值函数的正交轴。向量S^l 由实球谐函数$Y^l(\theta,\ \phi)$和一个系数为复数的矩阵 $T^l \in \mathbb{C}^{(2l+1)\times(2l+1)}$ 构造而成，具体如下:

$$S^l(\theta,\phi) = T^l Y^l(\theta,\phi) \tag{11}$$

有关球谐函数的更多详情，请参阅 [25, 26] 。

假设点云绕其重心发生旋转 $R \in SO(3)$。 R采用 ZYZ 欧拉角约定进行参数化， α， $\gamma \in [0, 2\pi[$ 和 $\beta \in [0, \pi]$，其中 g_r： $\boldsymbol{S}^2 \to \mathbb{N}$是旋转后点云的 IGE。因此，旋转矩阵 R 可表示为:

$$R = R(\alpha,\beta,\gamma) = \exp(\alpha\hat{e}_z)\exp(\beta\hat{e}_y)\exp(\gamma\hat{e}_z) \tag{12}$$

其中， e_y 和 e_z 分别为向量 (0，1，0) 和 (0，0，1)。算子 $\hat{\cdot}$: $\mathbb{R}^3 \to \mathfrak{so}(3)$ 将一个三维向量变换为其维数为 3×3 的对称矩阵，形式为李代数： $\mathfrak{so}(3) = \{S \in R^{3\times3} | S + S^T = 0\}$。尽管式 (12) 的表示存在固有奇异性

但它非常适合计算 $SO(3)$中的傅里叶变换。与式 (10) 一样， g_r 的傅里叶变换由下式给出：

$$g_r(\theta,\phi)=\sum_{l=0}^{B-1}(G_r^l)^T S^l(\theta,\phi) \tag{13}$$

其中 $G_r^l\in\mathbb{R}^{(2l+1)\times 1}$ 是傅里叶系数。
考虑到 g_r 是 f_r经过旋转变换后的版本，由此得到关系式 (14)， g_r 的傅里叶变换可利用 f_r 的傅里叶系数通过 (15) 计算。

$$g_r(\theta,\phi)=f_r(R^T(\theta,\phi)) \tag{14}$$

其中， $R^T(\theta,\ \phi)$ 是表达式 $M_{s2c}^{-1}(R^T M_{s2c}(\theta,\ \phi))$ 的简写，其中 M_{s2c}：$\boldsymbol{S}^2\to\mathbb{R}^3$ 是将球坐标转换为笛卡尔坐标的函数，而 M_{s2c}^{-1}：$\mathbb{R}^3\to\boldsymbol{S}^2$是其逆函数，可通过 (3) 求得。将 f_r 替换为其数值并改写 (14)，可得：

$$\begin{aligned} g_r(\theta,\phi)&=\sum_{l=0}^{B-1}(F_r^l)^T S^l(R^T(\theta,\phi))\\ &=\sum_{l=0}^{B-1}(U^l(R)F_r^l)^T S^l(\theta,\phi)\end{aligned} \tag{15}$$

其中， $U^l(R)\ \overline{=}\ T^l D^l(R)(T^l)^T$。 T^l 是 T^l的复共轭，而 $\boldsymbol{D}^l$ 是 Wigner $\boldsymbol{D}$ 矩阵。式 (15) 可以展开，因为旋转在频谱域中表示为 Wigner $\boldsymbol{D}$ 矩阵，并且对基函数 S^l 应用旋转，等价于通过与该旋转对应的 Wigner $\boldsymbol{D}$ 矩阵对基函数进行线性变换。根据 (13) 和 (15)，我们发现 $G_r^l=U^l(R)F_r^l$。因此，对 f_r的傅里叶系数应用变换 $U^l(R)$即可得到 G_r^l。关于 Wigner 矩阵常用性质的更多细节，请参阅 [25–27]。旋转估计的目标是在已知 f_r 和 g_r的情况下求出 R。

相关性（ $SO(3)$ 及其导数。 f_r 与 g_r 之间的相关性按如下方式计算：

$$\mathcal{C}_r(R)=corr(f_r,g_r)=\frac{1}{4\pi}\sum_{l=0}^{B-1}(G_r^l)^T U^l(R)F_r^l \tag{16}$$

该结果经过化简后得到：将f_r 和 g_r 替换为它们的傅里叶表示式 (10) 和 (15)，利用傅里叶变换的卷积定理 [2] 以及基底的正交原理 S^l。关系式 $\langle S^l(\theta,\phi),(S^l(R^T(\theta,\phi))^T)\rangle=\frac{1}{4\pi}U^l(R)$ 直接由基向量 S^l的正交性得出，其中，算子 $\langle.\rangle$ 是 $\mathcal{L}^2(SO(3))$ 中的内积。在 (16) 中，只有 U^l 取决于旋转 R，因此，可以通过计算 U^l的导数来求得 $\mathcal{C}_r$ 的导数。相对于基本旋转 $R_\epsilon=\exp(\hat{\epsilon\eta})$ ($\epsilon\approx 0$ 和 $\eta\in\mathbb{R}^3$)，在 R处求得的 U^l的导数计算如下：

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0}U^l(R\exp(\epsilon\hat{\eta}))=U^l(R)\left.\frac{d}{d\epsilon}\right|_{\epsilon=0}U^l(\exp(\epsilon\hat{\eta})) \tag{17}$$

在前一个方程中，使用了 U^l的同态性质，即，对于 $R_1,\ \ U^l(R_1R_2)=U^l(R_1)U^l(R_2)\,R_2\in\boldsymbol{SO}(3)$。随后按如下方式计算 $\mathcal{C}_r$的导数：

$$\begin{aligned}\left.\frac{d}{d\epsilon}\right|_{\epsilon=0}\mathcal{C}_r(\exp(\epsilon\hat{\eta}))&=\frac{1}{4\pi}\sum_{l=0}^{B-1}(G_r^l)^T U^l(R)u^l(\eta)F_r^l\cdot\eta\\&=\nabla\mathcal{C}_r(R,\eta)\cdot\eta\end{aligned} \tag{18}$$

其中， $\nabla\mathcal{C}_r(R,\ \eta)\in\mathbb{R}^3$ 是 $\mathcal{C}_r(R)$绕 η轴的梯度，而 $u^l(\eta)=\left.\frac{d}{d\epsilon}\right|_{\epsilon=0}U^l(\exp(\hat{\epsilon\eta}))$。在 $\eta=e_x,\ e_y,\ e_z$处计算梯度$\nabla\mathcal{C}_r(R,\ \eta)$，即可求得基本旋转；该旋转与 R 组合后，会提高相关性 $\mathcal{C}_r$的值。更形式化地说：

$$\nabla\mathcal{C}_r(R,e_k)|_{k\in\{x,y,z\}}=\frac{1}{4\pi}\sum_{l=0}^{B-1}(G_r^l)^T U^l(R)u^l(e_k)F_r^l \tag{19}$$

计算 $u^l(e_k)$很简单，因为这只是对维格纳矩阵 $\boldsymbol{D}$ 的输入直接求导，而该矩阵在 [25]中有解析导数。

现在可以使用梯度下降方法迭代搜索理想旋转。可以构造代价函数 J_r ，用于评估在$SO(3)$中旋转估计算法的性能：

$$J_r(R)=\frac{1}{2}||g_r(\theta,\phi)-f_r(R^T(\theta,\phi))||^2 \tag{20}$$

2.4 控制律

为了估计当前点云与参考点云之间的变换 $H=(R,\ T)\in\boldsymbol{SO}(3)\times\mathbb{R}^3$，采用了式 (21) 给出的控制律。

$$\begin{aligned}T&=T+\lambda_t\nabla_{glob}T\\R&=R\exp(\lambda_r\widehat{\nabla\mathcal{C}_r})\end{aligned} \tag{21}$$

其中， $\lambda_t,\ \lambda_r\in\mathbb{R}^+$ 和 $\lambda_t,\ \lambda_r<1$。 $\nabla_{glob}T$ 和 $\nabla\mathcal{C}_r$ 分别根据（8）和（19）计算。在第一次迭代时，旋转矩阵 R和向量 T可以分别随机初始化，或分别设为单位矩阵和零向量。控制器在满足以下条件时收敛：

$$||\nabla_{glob}T||+||\nabla\mathcal{C}_r||<\epsilon_g \tag{22}$$

2. 频谱域中的傅里叶系数相乘，等价于空间域中的卷积

算法1： 频谱域中的三维直接视觉伺服

频谱域

用单位矩阵初始化 R

用零向量初始化 T

初始化增益 λ_t, λ_r 和阈值 ϵ_g

计算目标点云的 f_t (2)、 f_r (4)、 F_t (5)、 F_r^l

</style}时 $||\nabla_{glob}T|| + ||\nabla \mathcal{C}_r|| >= \epsilon_g$ **do**

  获取场景点云

  计算 g_t (2)、 g_r (4)、 G_t (5)、 G_r^l 参考点云点

  计算 $\nabla_{glob}T$ (8)和 $\nabla \mathcal{C}_r$ (19)

  应用更新规则 (21)

  计算代价函数 $J = J_t(T) + J_r(R)$

  使用 (23) 控制机器人

结束

获取最终变换 $H = (R, T)$

其中， $\epsilon_g \in \mathbb{R}^+$ 为容差阈值。为控制机器人，采用以下控制律:

$$\dot{\mathbf{q}} = \mathcal{J}_c^+ \dot{X}_c \tag{23}$$

其中， $\mathcal{J}_c^+$为机器人雅可比矩阵的伪逆，在相机坐标系中表示； $\dot{\mathbf{q}}$ 是机器人关节速度的向量，而 X_c 是相机的笛卡尔速度，根据 (21) 计算得出。相应的控制律算法见算法 1。

3 实验验证

3.1 实验平台描述

实验验证使用深度相机采集的点云进行。本文介绍了两项不同的实验。首先，我们验证模型配准过程，即将物体的完整参考模型与场景点云对齐。随后，我们展示使用一台配备深度相机（Ensenso N35）的 7 个自由度机器人（KUKA iiwa）进行的直接视觉伺服实验，相机安装在机器人末端执行器上。在此情况下，使用整个点云将机器人定位到目标物体位置。点云库（PCL） [28]用于处理点云，FFTSO3 [25]和 FFTW [29]库则用于谱分析。

如第 1 节所述，用于模型配准实验的参考物体点云是在离线状态下，将从不同视角获取的多个点云组合而成，如 [20] 所示。使用 Ensenso N35 传感器时，每个点的表面法线可在采集时直接获得。点云以分辨率为 8 mm 的网格进行体素化，而表面法线则使用带宽 $B = 16$ 在单位球面上进行离散化。球面上谐波系数展开的最高阶数为 $l_{max} = 32$。这些值通过经验估算，在计算速度和配准精度方面效果良好。影响我们方法收敛速度的三个主要因素是体素网格的分辨率、IGE 的带宽以及参数 λ_t 和 λ_r。更精细的网格需要计算更多傅里叶系数，这会降低算法的收敛速度。采用前文提到的参数时，我们的方法目前平均处理速度为每次迭代 8.7 ms/。

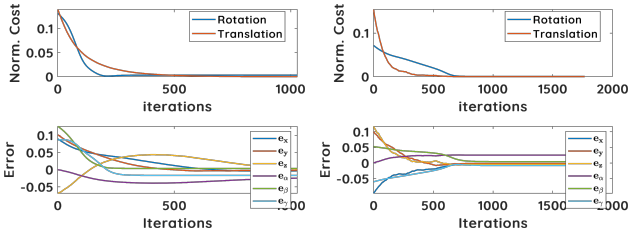

图 3 – 图 1 所示物体的收敛曲线：杯子（左）和油门把手（右）。

3.2 配准分析

我们进行了以下三项实验，以验证所提方法的模型配准能力：（C— 1）将物体的完整点云配准到经过任意旋转和平移变换后的版本（图 1 和图 3）；（C—2）将物体的完整点云配准到同一物体的目标局部点云（图 4）；以及（C—3）将物体的完整点云配准到包含多个物体的场景（图 5）。在这些测试中，参考点云与目标点云共享同一全局坐标系。实验采用了多种日常物体；如图 5 所示，多物体场景通过随机放置多个物体构建。图 1、3、4 和 5 展示了配准过程中收敛情况的示例，以及配准过程中误差（平移和旋转）的变化。

根据所得结果可以看出，在所有配置下，参考模型都能与目标点云精确配准。表 1 列出了 C–1 和 C–2 条件下最终平均收敛代价函数，以及 C–3 条件下最终梯度平均值。结果清楚地证明了我们所用方法的精度。此外，我们的方法在复杂条件下也表现出相当出色的性能，例如将完整模型与部分观测的模型配准，以及处理完全无结构的场景。

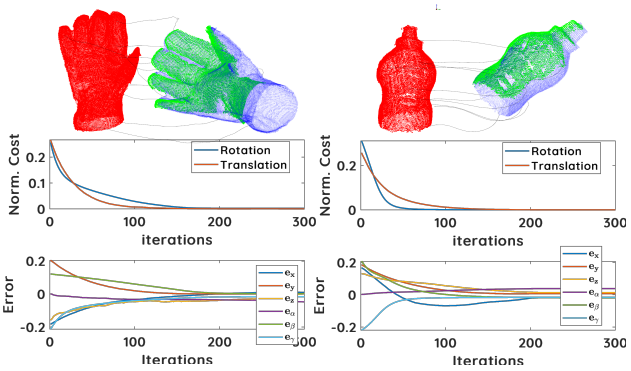

图4 – 完整参考模型配准到部分观测场景时，对齐过程示例。红色、蓝色和绿色点云分别表示参考模型、对齐模型和目标。展示了两个物体的结果：手套（左）和一个火柴盒（右）。

（存在遮挡）且包含多个物体。

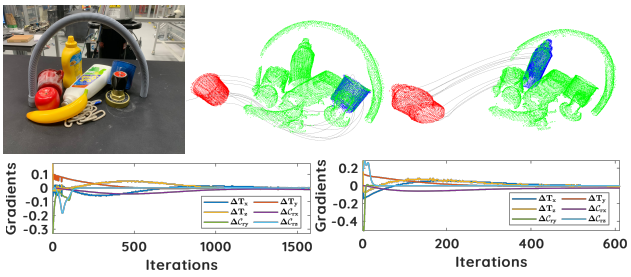

图5 – 多物体场景中的模型对齐示例。用于对齐两个不同物体的场景（上）；以及相应的曲线（下），显示对齐过程中平移和旋转梯度的变化。

表1 – 收敛时的平均值。

	C–1（代价）	C–2（代价）	C–3（梯度）
平移误差[1]	1.775e-5	8.113e-5	6.4e-05
旋转误差[1]	2.747e-2	3.2e-2	1.199e-12

[1] 计算值为（真实值与估计值）的平均值[2]。注意：此计算仅适用于 C–1 和 C–2。

3.3 机器人手臂定位任务：三维直接视觉伺服

在这项实验中，我们考虑一项定位任务，机器人位置由我们的直接视觉伺服方法控制。正如前文所述，整个点云都用于控制回路，即不使用分割方法。该测试使用由多个物体构成的场景进行，以检验我们的方法的通用性，尤其是在复杂场景下。机器人处于期望位置时，首先采集参考点云。随后将机器人移动到任务空间中的一个随机位置，并确保发生较大的空间变换。结果如图 6 所示。从这些结果可以清楚地看出，即使从只能看到点云部分内容的位置出发，该方法仍能很好地将整个点云对齐。收敛曲线显示，机器人能够平稳运动至目标位置。

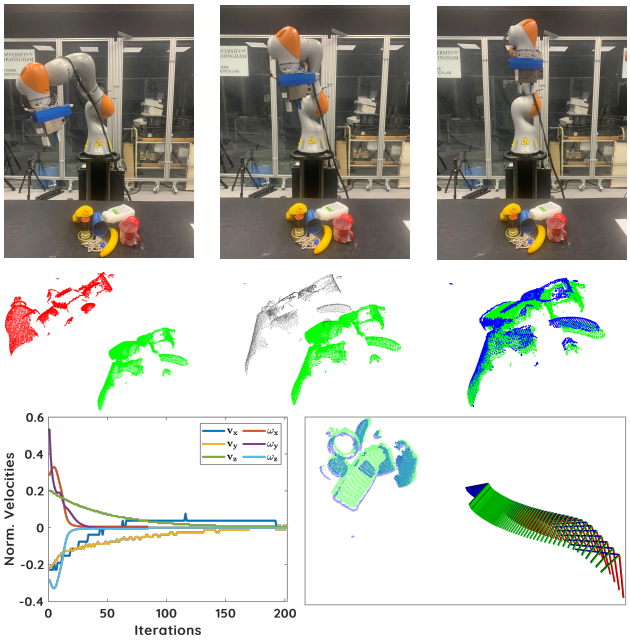

图 6 – 三维直接视觉伺服示意图。上排显示机器人在起始、中间和最终位置的状态。中排显示初始、中间和最终点云。下排显示收敛曲线以及机器人末端执行器的运动轨迹。本次测试使用完整点云，未进行任何局部模型匹配。

4 结论

本文介绍了一种直接三维视觉伺服方法，该方法基于使用频谱域中的点云。所提出的方法利用傅里叶变换的时间平移特性来估计平移，并利用实球谐函数在 $SO(3)$中估计旋转，从而逐步将参考点云与目标点云对齐。这种方法最初用于将三维模型与不同场景对齐，随后用于控制机器人手臂的位置，以便自动执

行定位任务。实验结果表明，我们的方法在精度和控制器行为方面均有效，包括在不利条件下（例如存在遮挡时）。未来的工作将侧重于将所提出的方法用于操纵静止或运动中的物体。

参考

[1]K. Deguchi, “结合相机与物体运动对动态图像进行直接解释，以实现视觉引导的机器人控制”，国际计算机视觉杂志，第 37 卷，第 7–20 页，2000年。[2] B. Tamadazte、N. L.-F. Piat 和 E. Marchand，“一种用于自动纳米定位的直接视觉伺服方案”，*IEEE/ASME* 机电学报，第 728–736 页，2012年。[3] C. Collewet 和 E. Marchand，“光度视觉伺服”，*IEEE* 机器人学汇刊，第 27 卷，，第 828–834 页，2011年。[4] E. Marchand 和 C. Collewet，“将图像梯度用作视觉伺服的视觉特征”，载于*IEEE/RSJ* 智能机器人与系统国际会议，2010年，第 5687–5692 页。[5] Q. Bateux 和 E. Marchand，“基于直方图的视觉伺服”，*IEEE* 机器人与自动化快报，第 80–87 页，2016年。[6]A. Dame 和 E. Marchand，“基于互信息的视觉伺服”，*IEEE* 机器人学汇刊，第 27 卷，第 958–969 页，2011年。[7] N. Crombez、E. M. Mouaddib、G. Caron 等，“以光度高斯混合模型作为稠密特征的视觉伺服”，*IEEE* 机器人学汇刊，第 35 卷，第 49–63 页，2018年。[8] M. Ourak、B. Tamadazte、O. Lehmann 等，“使用小波系数的直接视觉伺服”，*IEEE/ASME* 机电学报，第 1129–1140 页，2019年。[9]L.-A. Duflot、R. Reisenhofer、B. Tamadazte 等，“用于视觉伺服的小波与剪切波图像表示”，国际机器人研究杂志，第 38 卷，第 422–450 页，2019年。[10] F. Larsson、M. Felsberg 和 P.-E. Forssen，“用于道路标志识别的局部图像块傅里叶描述符相关”，*IET* 计算机视觉，第 5 卷，第 4 期，第 244–254 页，2011年。[11] H. Bülow 和 A. Birk，“部分重叠的含噪三维距离数据的谱域六自由度配准”，*IEEE* 模式分析与机器智能汇刊，第 35 卷，第 954–969 页，2012 年。[12] M. Adjigble 、 C. de Farias、R. Stolkin 和 N. Marturi，“Spectgrasp：基于谱相关的机器人抓取”，载于*IEEE/RSJ* 智能机器人与系统国际会议，2021年，第 3987–3994 页。[13] E. Marchand，“频域中的直接视觉伺服”，*IEEE* 机器人与自动化快报，第 5 卷，第 620–627 页，2020年。[14] N. Marturi、B. Tamadazte、S. Dembélé 等，“用于扫描电子显微镜下自动纳米定位的视觉伺服方案

扫描电子显微镜”，载于*IEEE* 机器人与自动化国际会议，2014年，第 981–986 页。[15] N. Marturi、B. Tamadazte、S. Dembélé 等，“用于扫描电子显微镜的图像引导纳米定位方案”，*IEEE* 自动化科学与工程汇刊，第 15 卷，第 45–56 页，2016年。[16] C. Teuliere 和 E. Marchand，“使用深度图进行视觉伺服的稠密直接方法”，*IEEE* 机器人学汇刊，第 30 卷，第 1242–1249 页，2014年。[17] B. Dahroug、B. Tamadazte 和 N. Andreff，“基于 PCA 并使用光学相干断层成像的视觉伺服”，*IEEE* 机器人与自动化快报，第 5 卷，第 3430–3437 页，2020年。[18]C. Kingkan、S. Ito、S. Arai 等，“基于点云数据的模型式虚拟视觉伺服”，载于*IEEE/RSJ* 智能机器人与系统国际会议，2016年，第 5549–5555 页。[19]M. Ourak、B. Tamadazte、O. Lehmann、等，“基于小波的六自由度视觉伺服”，载于*IEEE* 机器人与自动化国际会议，2016年，第 3414–3419 页。[20] N. Marturi、M. Kopicki、A. Rastegarpanah 等，“移动物体的动态抓取与轨迹规划”，自动化与机器人，第 43 卷，第 1241–1256 页，2019年。[21] M. Adjigble、N. Marturi、V. Ortenzi 等，“通过局部接触力矩匹配实现无模型、无学习抓取”，载于 *IEEE/RSJ* 智能机器人与系统国际会议，2018年，第 2933–2940 页。[22] J. J. Little，“扩展高斯图像、混合体积与形状重建”，载于第一届计算几何年会，1985年，第 15–23 页。[23] S. K. Nayar、A. C. Sanderson、L. E. Weiss 等，“利用结构化高光和高斯图像进行镜面表面检测。”*IEEE* 机器人与自动化汇刊，第 6 卷，第 208–218 页，1990年。[24] B. Lowekamp、P. Rheingans 和 T. S. Yoo，“利用交互式高斯图像探索表面特性（案例研究）”，载于 *IEEE* 可视化会议，2002年，第 553–556 页。[25] T. Lee，“特殊正交群上的实调和分析”，*arXiv* 预印本 *arXiv :1809.10533*，2018年。[26] M. A. Blanco、M. Flórez 和 M. Bermejo，“实球谐基下旋转矩阵的计算”，分子结构杂志：理论化学，第 419 卷，第 19–27 页，1997年。[27] P. J. Kostelec 和 D. N. Rockmore，“旋转群上的快速傅里叶变换”，傅里叶分析与应用杂志，第 14 卷，第 145–179 页，2008年。[28] R. B. R. 和 S. C.，“3D 时代已至：点云库（PCL）”，载于*IEEE* 机器人与自动化国际会议，2011年。[29] M. Frigo 和 S. G. Johnson，“FFTW：一种自适应 FFT 软件架构”，载于*IEEE* 声学、语音与信号国际会议，第 3 卷，1998年，第 1381–1384 页。