

파운데이션 피쳐 기반 뉴럴 디스크립터 필드를 활용한 비강체 Mesh-to-RGBD 정합

Non-Rigid Mesh-to-RGBD Registration via Foundation Feature-Based Neural Descriptor Field

이 영 준¹ · 임 종 우[†]
Yeongjun Lee¹, Jongwoo Lim[†]

Abstract: In this paper, we present a framework for non-rigid alignment between a canonical 3D mesh of a non-rigid object and a single RGB-D image in a category-agnostic manner. Conventional optimization-based methods typically rely solely on geometric features, making them highly sensitive to ambiguities caused by occlusions and partial observations. To address these limitations, we propose the Neural Descriptor Field (NDF), which leverages semantic and geometric representations of pre-trained foundation models to construct a continuous 3D descriptor field via instance-specific optimization. During registration, point-wise features extracted from the input RGB-D image are compared with those from the NDF to establish robust correspondences, which are then used within a scoring-based mechanism to estimate a deformation-tolerant coarse pose. This coarse alignment is further refined through a test-time non-rigid optimization that minimizes features and geometric inconsistencies. The proposed two-stage framework ensures accurate registration even under substantial pose variations and shape deformations. To evaluate our approach under challenging conditions not covered by existing benchmarks, we construct a new synthetic dataset with diverse deformations and occlusions. We demonstrate that our method achieves superior accuracy and robustness compared to state-of-the-art techniques.

Keywords: Non-rigid Registration, Robot Vision, Neural Descriptor Field, Test-Time Optimization

1. 서 론

비강체 정합(non-rigid registration)은 물체에 외력이나 자중 등으로 인한 변형이 발생했을 때, 기준 모델과 실제 관측 데이터 간의 형태변화를 추정하여 두 형상을 정밀하게 일치시키는 과정이다. 이는 변형체의 추적, 로봇 조작, 3D복원 등 다양한 로봇틱스 및 컴퓨터 비전 응용 분야에서 형상 일치와 일관성을 보장하기 위한 필수적인 기술이다. 그러나 강체 정합과 달리 비강체 정합은 회전과 이동 뿐만 아니라 형상의 비선형적인 구조적

변형까지 함께 추정해야 하므로 자유도와 복잡도가 매우 높다. 특히 로봇 조작 환경과 같이 단일 시점 RGB-D 센서에 의존하는 경우, 깊이 영상의 구조적 부분 관측과 빈번한 가려짐으로 인해 안정적인 정합을 달성하기 어렵다.

기존 연구들은 이러한 문제를 해결하기 위해 기하적 거리 최소화에 집중해왔다. 초기 최적화 기반 방법론^[1,2]이나 최근의 신경 표현(Neural Representation)을 활용한 연구^[3,4]들은 변형을 국소 제약을 포함한 그래프 구조나 MLP를 활용한 연속적인 함수로 모델링하여 자연스러운 변형을 유도하고자 했다. 그러나 이러한 접근 방식들은 대다수 점군 간의 Chamfer Distance와 같은 유클리드 거리를 최소화하는 기하적 손실함수에 의존한다는 근본적인 한계가 있다. 기하적 정보에만 의존할 경우, 텍스처가 부족하거나 형상 특징이 모호한 영역에서는 대응점이 표면을 따라 미끄러지는 현상이 발생하기 쉽다. 또한 심한 가려짐이나 부분관측 상황에서는 최근접점이 실제 대응점과

Received : Oct. 23. 2025; Revised : Dec. 5. 2025; Accepted : Jan. 3. 2026

1. Master Student, Department of Mechanical Engineering, Seoul National University, Seoul, Korea; Senior Engineer, Global Technology Research, Samsung Electronics, Suwon, Korea (yeongjunlee@snu.ac.kr)

† Professor, Corresponding author: Department of Mechanical Engineering, Seoul National University, Seoul, Korea (jongwoo.lim@snu.ac.kr)

일치하지 않아 국소 최적해에 빠지며, 최종 형상은 목표 형상과 유사해 보일지라도 실제 대응점의 위치 정밀도가 크게 저하되는 문제를 야기한다.

이에 본 연구에서는 별도의 대규모 데이터셋을 활용한 사전 학습 없이, 범용적인 파운데이션 모델을 활용하여 인스턴스별 최적화만으로 수행가능한 Neural Descriptor Field (NDF) 기반의 비강체 정합 프레임워크를 제안한다. 제안 방법은 기준 메쉬(Canonical Mesh) 위에서 정의된 연속적인 디스크립터를 활용하여, 입력된 부분 관측 데이터를 기준 좌표계로 강건하게 투영하는 구조이다. 3D 좌표를 시각적 및 기하적 파운데이션 피쳐를 결합한 고차원 피쳐공간으로 매핑함으로써, 기하적으로 모호한 영역이나 가려진 영역에서도 안정적인 정합을 유도한다. 이를 바탕으로 변형에 강건한 초기 자세(Coarse Pose)를 추정 한 뒤, Test-time 최적화를 통해 정밀한 비강체 정합을 수행하는 Coarse-to-Fine 전략을 사용한다.

본 연구의 주요 기여는 다음과 같다.

- (1) 시각적-기하적 파운데이션 피쳐를 결합하여, 별도의 사전 학습 없이 인스턴스별 최적화만으로 가려짐이나 형상적 모호함에 강건한 비강체 Mesh-to-RGBD 정합 프레임워크를 제안하였다.
- (2) 피쳐 유사도 기반 Coarse-to-fine 전략을 통해, 큰 초기 정렬 오차가 존재하는 상황에서도 안정적인 정합 성능을 달성함을 보였다.
- (3) 물리적 변형을 반영한 Mesh-to-RGBD 합성 데이터셋을 구축하였으며, EPE 및 Accuracy 지표에서 기존 최신 기법 대비 정량적으로 우수한 성능을 입증하였다.

2. 관련 연구

2.1 기하 기반 비강체 정합 및 신경 표현 기법

초기 비강체 정합 연구는 국소적 변형의 연속성을 가정하여 변형을 모델링 하였다. Non-Rigid ICP (NICP)^[1]는 각 정점(vertex)에 로컬 아핀(affine) 변환을 부여하고 인접 변환 간 차이를 최소화하는 정규화(Regularization)를 통해, Embedded Deformation (ED)^[2]은 그래프 기반의 국소 강성(Local Stiffness)을 유지하는 방식으로 자연스러운 변형을 모델링 하였다. 최근에는 딥러닝 기반의 암시적 표현(Implicit Representation)을 활용하여 이러한 변형을 연속함수로 모델링하는 연구가 활발하다. NSFP^[3]는 네트워크의 Inductive Bias를 활용하여 부드러운 변형을 유도하였으며, NDP^[4]는 변형을 단단계 피라미드 구조로 분해하여 국소 최적해를 방지하고 수렴 속도를 개선하였다. 또한 DPF^[5]는 포인트 기반 표현과 암시적 필드를 결합하여 계산 효율성을 극대화 하였다.

그러나 이러한 최적화 기반 기법들은 여전히 Chamfer Distance와 같은 기하적 거리 척도에 의존한다는 근본적 한계를 공유한다. OAREg^[6]가 Correntropy를 도입하여 이상치에 대한 강건성을 높였으나, 피쳐 정보의 부재로 인해 텍스처가 부족하거나 형상이 반복되는 모호한 영역에서는 대응점이 표면을 따라 미끄러지는 오정합이 빈번하게 발생한다. 이는 본 연구가 의미론적 정보를 포함한 파운데이션 피쳐를 도입하게 된 주된 배경이다.

2.2 파운데이션 모델 기반 3D 비전

대규모 데이터로 학습된 파운데이션 모델(Foundation Model)의 강건한 의미론적 특징을 3D 비전으로 확장하려는 시도가 활발하다. 특히 자기지도학습 기반의 DINOv2^[7]는 다양한 조명 변화나, 가려짐, 시점 차이에도 강건한 의미론적 일관성을 유지하는 피쳐를 제공하여, 텍스처가 없거나, 기하적 특징이 부족한 물체의 대응점 추정 문제를 해결할 단서를 제공한다.

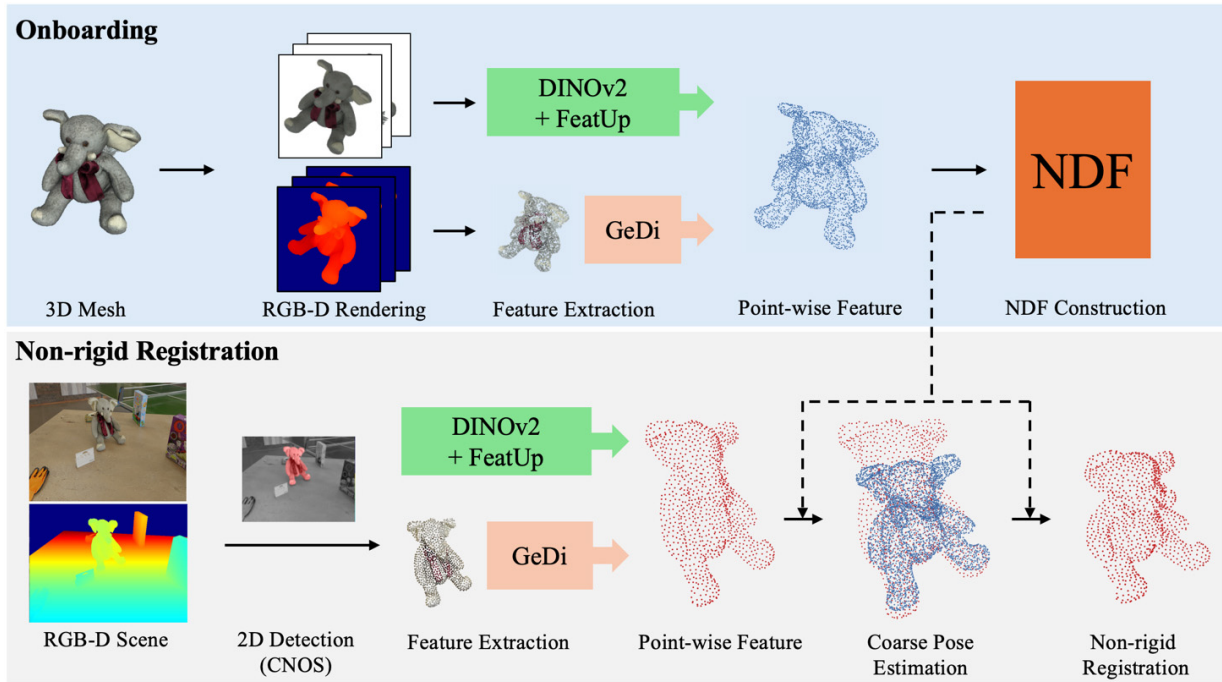
그러나 ViT(Vision Transformer)^[8] 기반 모델들은 패치 단위 연산으로 인해 해상도가 낮아, 픽셀 단위의 정합에 활용하기에는 한계가 있었다. 이를 극복하기 위해 제안된 FeatUp^[9]은 저해상도 피쳐를 고해상도로 업샘플링하여 정밀한 처리를 가능하게 하였다. 이와 더불어 GeDi^[10]와 같은 기하 파운데이션 모델은 3D 점군의 국소적 곡률이나 위상 정보를 강건한 임베딩으로 표현함으로써, 시각 정보만으로는 구분이 모호한 형상의 기하적 특징을 효과적으로 나타낸다.

최근 FreeZe^[11] 등은 이러한 시각적, 기하적 파운데이션 피쳐를 결합하여 사전 학습 없이도 강체 물체(Rigid Body)의 자세 추정 분야에서 우수한 성과를 보였다. 하지만 이는 템플릿 형상이 고정된 강체만을 대상으로 하며, 이를 복잡한 자유도를 가진 비강체 정합의 변형 최적화 문제로 확장한 연구는 아직 미비하다. 본 연구는 시각적, 기하적 파운데이션 피쳐를 통합한 NDF 프레임워크를 제안하여 기존 기하 기반 방법론들이 겪는 형상적 모호함과 초기 정렬의 한계를 극복하여 강건한 비강체 정합을 달성하고자 한다

3. 제안 방법

3.1 개요

본 연구에서는 시각적 및 기하적 파운데이션 피쳐를 통합한 Neural Descriptor Field(NDF) 기반의 비강체 Mesh-to-RGBD Scene 정합 프레임워크를 제안한다. [Fig. 1]과 같이, 전체 파이프라인은 다음 세 단계로 구성된다. (1) 기준 메쉬(Canonical Mesh)로부터 NDF를 구축하는 온보딩(Onboarding) 단계, (2) 피쳐 유사도를 통해 초기 정렬을 수행하는 초기 자세 추정



[Fig. 1] Overview of the proposed pipeline for NDF-based non-rigid registration. The pipeline consists of an onboarding stage for NDF construction from rendered RGB-D views of 3D meshes, and a non-rigid registration stage that aligns RGB-D scenes to the canonical mesh using coarse pose estimation and NDF-guided deformation

(Coarse Pose Estimation) 단계 (3) Test-time 최적화를 이용하여 정밀한 변형을 수행하는 비강체 정합(Non-rigid Registration) 단계이다.

3.2 Onboarding

3.2.1 Mesh 렌더링 및 피쳐 추출

기준 메쉬(Canonical Mesh)는 42개의 균등 뷰포인트에서 RGB-D로 렌더링된다. 각 RGB이미지로부터 DINOv2^[7] 및 FeatUp^[9]을 통해 고해상도 시각 피쳐를 추출하고 PCA를 통해 64차원으로 압축한다. 깊이(Depth) 정보를 이용해 복원한 포인트 클라우드에서 GeDi^[10]를 통해 기하적 피쳐를 추출하고, 시각 피쳐 또한 해당 포인트 클라우드에 투영한다. 여기서 GeDi 피쳐는 복수의 반경(0.15D, 0.3D)에서 정의된 피쳐를 결합하여 사용하며 여기서 D는 기준 메쉬의 지름을 의미한다. 결과적으로 포인트 클라우드의 각 포인트는 시각 피쳐와 기하 피쳐를 가지게 된다. 이 피쳐 추출 과정은 FreeZe^[11]와 유사한 피쳐 추출 절차를 따르며, 본 연구에서는 FeatUp^[9]을 통한 고해상도 피쳐를 활용하여 이를 비강체 정합 문제로 확장하였다. 포인트별 피쳐를 포함한 42개의 포인트 클라우드는 렌더링 시의 Extrinsic 파라미터를 이용해서 동일 좌표계로 정렬되며, 최종적으로 NDF 구축을 위한 입력 데이터로 사용된다.

3.2.2 Neural Descriptor Field 구축

Neural Descriptor Field(NDF)는 3D 좌표 $x \in \mathbb{R}^3$ 를 입력으로 받아 시각 피쳐 \hat{f}^{vis} , 기하 피쳐 \hat{f}^{geo} , 법선 \hat{n} 을 출력하는 식 (1)과 같은 연속함수로 정의된다.

$$NDF(x) = (\hat{f}^{vis}, \hat{f}^{geo}, \hat{n}) \quad (1)$$

입력좌표 x 는 HashGrid 인코딩을 통해 다중 해상도 표현으로 변환된 후^[12], 경량 Fully-Fused MLP를 거쳐 잠재 임베딩 $h \in \mathbb{R}^{128}$ 으로 투영된다. 이 잠재 벡터는 세 개의 선형 헤드를 통해 시각, 기하 피쳐 및 법선 정보를 각각 예측한다. NDF는 최적화 시에 파운데이션 피쳐 및 법선 (f^{vis*}, f^{geo*}, n^*)을 회귀하도록 설계되며, 법선 방향을 기준으로 한 비등방 가우시안 증강을 통해 접평면 방향의 표면 평활성을 유지함과 동시에 메쉬 표면 근처 영역에서도 안정적인 피쳐를 예측한다. 전체 손실은 식 (2)와 같이 시각 손실, 기하 손실, 자코비안 근사 평활 손실, 법선 정합 손실 네 항으로 구성된다.

$$L_{NDF} = \lambda_v L_{vis} + \lambda_g L_{geo} + \lambda_j L_{jac} + \lambda_n L_{normal} \quad (2)$$

시각 피쳐는 비교적 안정적인 분포를 가지므로 식 (3)의 평균제곱오차(MSE) 기반으로 최적화한다.

$$L_{vis} = \frac{1}{N} \sum_{i=1}^N w_i \|\hat{f}^{vis} - f^{vis*}\|_2^2 \quad (3)$$

기하 피쳐는 깊이 이미지의 복원 과정에서 Self-occlusion이 나, 경계 불연속에 의한 피쳐 불일치를 보상하기 위해 식 (4)와 같이 로그 불확실도 \hat{b}_i 를 함께 예측하는 Heteroscedastic 회귀 손실을 사용한다.

$$L_{geo} = \frac{1}{N} \sum_{i=1}^N w_i \left[e^{-\hat{b}_i} \|\hat{f}^{geo} - f^{geo*}\|_2^2 + \hat{b}_i \right] \quad (4)$$

식 (5)의 자코비안 근사 평활 손실은 접평면 방향의 미소 변위 $\delta_i = \epsilon(t_{1,i}u + t_{2,i}v)$ 에 대해 기하 피쳐의 변화가 최소화되도록 하여 표면에서 안정적이며 노이즈에 민감하지 않도록 한다. 여기서 $u, v \sim N(0, 1)$, $\epsilon \ll 1$ 이며 $t_{1,i}, t_{2,i}$ 는 법선과 직교하는 두 접선 벡터이다.

$$L_{jac} = \frac{1}{N} \sum_{i=1}^N \|\hat{f}^{geo}(x_i) - \hat{f}^{geo}(x_i + \delta_i)\|_2^2 \quad (5)$$

법선 정합 손실은 식 (6)의 로그 코사인 형태를 사용하여, 법선 불일치 시 gradient가 커지도록 함으로써 초기 수렴속도를 향상시킨다.

$$L_{normal} = \frac{1}{N} \sum_{i=1}^N -\log\left(\frac{\hat{n}_i^\top n_i^* + 1}{2} + \epsilon\right) \quad (6)$$

모든 온보딩은 기준 메쉬(Canonical Mesh) 단위로 수행된다. NDF에서 출력한 디스크립터는 메쉬의 표면 근처에서만 의미를 가지므로 추론 시에는 SDF 기반 가우시안 가중과 함께 활용한다.

3.3 Coarse Pose Estimation

비강체 정합을 진행하기 전 초기 정렬(Coarse Pose Alignment)을 진행한다. 장면 RGB-D에서 대상체를 CNOS^[13]를 통해 탐지하고, 탐지된 마스크 영역으로부터 포인트 클라우드를 복원한 후 피쳐를 추출한다. 이후, 해당 피쳐를 NDF에서 출력된 디스크립터와 매칭하여 강체 초기자세 (R, t) 를 추정한다.

3.3.1 피쳐 대응점 생성

장면에서 추출한 포인트 $\{x_i\}_{i=1}^N$ 별 장면 피쳐 f_i^{scene} 와, 메쉬 표면에서 샘플링한 포인트를 NDF에 쿼리하여 얻은 모델 피쳐 f_j^{model} 간 유사도 행렬 $S_{ij}^{scene} = \langle f_i^{scene}, f_j^{model} \rangle$ 을 계산한다. 각 장면 포인트는 가장 유사도가 높은 모델 피쳐를 대응점

으로 선택하여 대응쌍 집합 C_i 을 형성한다. 여기서 피쳐 $f = [f^{vis}, f^{geo}]$ 이다.

3.3.2 NDF 기반 자세 스코어링

대응쌍 집합에서 3쌍을 무작위로 반복 샘플링하여, 각 샘플에 대해 SVD를 이용하여 강체 변환 가설 (R, t) 을 추정한다. 각 가설 (R, t) 에 대해 장면 포인트 및 법선을 변환하여 $x'_i = Rx_i + t, n'_i = Rn_i$ 를 계산한 뒤 변환된 포인트 $\{x'_i\}_{i=1}^N$ 를 NDF에 쿼리하여 예측 디스크립터, 법선, SDF 가중 $(\hat{f}_i, \hat{n}_i, \omega_i^{SDF})$ 을 얻는다. 이를 식 (7)을 통해 피쳐 유사도 점수를 계산하여 가설의 적합도를 평가한다. 이때 식 (8)의 법선 일치도 척도를 이용하여 올바르게 정렬되지 않은 표면의 영향을 완화한다.

$$score(R, t) = \sum_{i=1}^N \gamma_i \omega_i^{SDF} \left(\frac{\hat{f}_i^\top f_i^{scene}}{\|\hat{f}_i\| \|f_i^{scene}\|} \right) \quad (7)$$

$$\gamma_i = 1 \left(\frac{\hat{n}_i^\top n'_i}{\|\hat{n}_i\| \|n'_i\|} \geq \cos \theta_{thre} \right) \quad (8)$$

모든 가설 중 $score(R, t)$ 최대인 것을 Coarse Pose (R_0, t_0) 로 선택한다. 이 과정은 배치단위로 병렬 수행된다.

3.4 Non-Rigid Registration

3.3에서 얻은 Coarse Pose (R_0, t_0) 를 통해 장면의 포인트 클라우드를 기준 메쉬(Canonical Mesh) 형태에 비강체 정합한다. 이를 위해 공간 좌표 기반 MLP 변형필드 $D: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ 를 Test-time 최적화한다. 각 반복 수행 과정에서 장면에서의 포인트 x_i 는 Coarse Pose를 이용해 $x'_i = R_0 x_i + t_0$ 로 강체 변환한 뒤, 변형필드를 적용하여 $\tilde{x}_i = x'_i + D(x'_i)$ 로 업데이트 된다. 이 변형필드는 식 (9)의 손실을 최소화하도록 최적화한다. 이때 각 항의 가중치 λ 는 초기 단계에서 각 손실항의 크기가 유사하도록 실험적으로 결정되었다.

$$L_D = \lambda_{feat} L_{feat} + \lambda_{cham} L_{cham} + \lambda_{corr} L_{corr} \quad (9)$$

3.4.1 NDF 피쳐 유사도 손실

변형된 포인트 \tilde{x}_i 를 NDF에 쿼리하여 예측 디스크립터 \hat{f}_i 와 SDF기반 표면 가중치 $\omega_i^{SDF} \in [0, 1]$ 를 얻는다. 장면 포인트의 피쳐 f_i 와 코사인 유사도를 최대화하는 손실은 식 (10)과 같이 정의한다. 이때 코사인 유사도에 ReLU를 적용한 뒤 1에서 감산

하는 형태를 취하여, 손실 값이 항상 0 이상의 양수가 되도록 설계하였다.

$$L_{feat} = \frac{1}{N} \sum_i \left[\left(1 - \max \left(0, \frac{\hat{f}_i^\top f_i}{\|\hat{f}_i\| \|f_i\|} \right) \right) \omega_i^{SDF} \right] \quad (10)$$

ω_i^{SDF} 는 SDF의 가우시안 가중치로, 표면에 근접할수록 높은 가중치를 부여하여, 표면부근의 피쳐의 유사성을 강하게 유도한다.

3.4.2 피쳐 일관성 기반 Soft Chamfer Distance

단순 Chamfer Distance는 기하적 근접성에만 의존하여 최근접점을 찾기 때문에, 가려짐이나 기하적 모호성이 존재하는 영역에서 잘못된 표면으로 수렴하는 오정합이 발생하기 쉽다. 이를 보완하기 위해 식 (11)의 피쳐 일관성 기반 Soft Chamfer Distance를 제안한다. Chamfer Distance와 같은 기하 거리항에 식 (12)의 가우시안 거리 가중(ω^{dist})과, 디스크립터 유사도 기반의 피쳐 일관성 가중(ω^{feat})을 결합하여 나타낸다.

$$L_{cham} = \frac{1}{|\tilde{X}|} \sum_i \omega_i^{dist} \omega_i^{feat} \min_j \|\tilde{x}_i - y_j\|^2 + \frac{1}{|\tilde{Y}|} \sum_j \omega_j^{dist} \omega_j^{feat} \min_i \|y_j - \tilde{x}_i\|^2 \quad (11)$$

$$\omega^{dist} = e^{-d^2/2\sigma^2}, \quad \omega^{feat} = 1 + \max \left(0, \frac{f_x^\top f_y}{\|f_x\| \|f_y\|} \right) \quad (12)$$

ω^{dist} 는 거리가 먼 이상치를 효과적으로 억제하여 Chamfer 항을 안정화하고, ω^{feat} 은 피쳐 일관성이 높은 대응쌍을 강화한다. ω^{feat} 의 하한을 1로 둔 것은 일반적으로 디스크립터 유사도가 낮은 대응쌍일수록 기하적 거리 또한 큰 경향이 있어 Chamfer Distance가 이미 낮은 유사도에 대한 패널티를 충분히 제공한다. 이때 추가적인 down-weighting을 적용할 경우 gradient가 약화되어 비강체 정합 과정의 수렴 안정성을 저하시킬 수 있다.

3.4.3 대응쌍 손실

대응쌍 손실의 경우 §3.3.1에서 구성한 대응쌍 집합 C_i 내의 포인트 쌍의 유클리드 거리오차를 최소화하도록 정의한다. 즉, 대응쌍 간의 MSE 손실을 다음 식 (13)과 같이 계산한다.

$$L_{corr} = \frac{1}{|C_i|} \sum_{(i, j) \in C_i} \|\tilde{x}_i - y_j\|_2^2 \quad (13)$$

기하적 최근접점에 의존하는 L_{cham} 과 달리, 대응쌍 손실은 피쳐 유사도에 기반한 명시적 대응관계를 이용한다. 기하적 거

리가 먼 대변형 상황에서도 전역적인 정합 방향을 제시하여 국소 최적해(Local Minima)를 방지하고 전체적인 정합의 안정성을 보장하는 역할을 수행한다.

4. 실험 및 결과

4.1 데이터셋

기존 비강체 정합 벤치마크로 널리 사용되는 4DMatch^[14] 등은 주로 두 점군 간의 직접 정합을 목표로 하기 때문에 주 관측 간의 시점 차이가 크지 않고 중심점 정렬이나 PCA를 통해 비교적 쉽게 초기화가 가능하다. 그러나 본 연구는 변형되지 않은 기준 메쉬(Canonical Mesh)를 변형된 물체의 단일 RGB-D 관측에 정렬하는 Mesh-to-Scene 문제를 다루며, 이 경우 임의의 뷰포인트에서 관측된 장면과 기준 메쉬 사이에는 예측하기 어려운 큰 회전·변환 차이가 존재한다. 따라서 기존 벤치마크는 본 연구의 초기 정렬 난이도를 충분히 반영하지 못한다.

이러한 점을 고려하여, 본 연구에서는 실제 응용 환경의 조건을 반영하면서도 데이터셋의 공정성을 확보하기 위해 새로운 합성 Mesh-to-Scene 데이터셋을 구축하였다. 우선, 특정 모델에 대한 편향을 배제하기 위해 학계에서 널리 사용되는 Google Scanned Objects (GSO)^[15] 데이터베이스에서 인형, 가방, 장난감 등 10종의 일상 물체를 선정하였다. 각 객체에는 대해 굽힘, 비틀기, 전단, 스케일 변형 등 8종의 물리적 변형함수를 무작위로 조합하여 적용하였으며, 변형 전후 정점 인덱스(Index)를 유지하여 정확한 일대일 Ground Truth 대응 관계를 생성하였다.

생성된 변형 객체는 Blender를 사용해 사실적인 배경, 조명 조건에서 객체당 30개의 RGB-D 장면으로 렌더링하였다. 추가적으로 심각한 가려짐(Occlusion)의 영향 분석을 위해 20~50%의 가려짐을 포함하는 추가 세트를 객체당 10장씩 별도로 렌더링하여 4.5절의 평가에 사용하였다. 본 연구에서는 사전학습 과정 없이 기준 메쉬(Canonical Mesh)를 직접 입력으로 사용하므로, 생성된 데이터셋 전부를 테스트 데이터셋으로 활용하였으며, 데이터셋의 구성 형식은 강체의 자세추정의 분야에서 널리 사용되는 BOP (Benchmark for 6D Object Pose Estimation)^[16] 형식을 준용하였다. 각 샘플은 RGB 이미지, 깊이 맵, 카메라 Intrinsic·Extrinsic 파라미터, 그리고 GT 대응점 정보를 포함한다

4.2 구현 상세

4.2.1 NDF 은보딩

본 연구에서 사용한 NDF는 Instant-NGP 스타일의 Hash Grid-MLP 백본을 기반으로 구현하였다. 입력 좌표 $x \in \mathbb{R}^3$ 는

8-레벨 헤시 그리드 인코딩을 거치며, 각 레벨당 피쳐 수는 2, base resolution은 8, per-level scale은 1.5로 설정하였다. 인코딩 결과는 2-layer MLP (hidden dimension=64)에 입력되어 포인트 별 디스크립터를 출력한다. 식 (2)의 가중치는 각 손실항의 초기 크기 균형을 맞추기 위해 $\lambda_{vis} = 10$, $\lambda_{geo} = 0.05$, $\lambda_{jac} = 50$, $\lambda_{norm} = 0.1$ 으로 설정하였다. 최적화는 Adam 옵티마이저 ($lr = 1 \times 10^{-3}$)로 수행하였으며 각 기준 객체에 대해 독립적으로 NDF를 구축하였다. RTX 4080 GPU 환경에서 렌더링, 피쳐 추출 및 NDF 구축을 포함한 온보딩 과정은 객체당 약 300초 미만의 시간 내에 완료되었다. 이 과정은 추론과 별도로 테스트 전 1회만 수행된다.

4.2.2 변형장 최적화

비강체 변형장은 입력 좌표에 대한 변위를 출력하는 128차원 3개 은닉층 구조의 MLP를 파라미터화 하였다. 최적화는 Adam($lr = 5 \times 10^{-5}$)을 사용하였으며, 학습률은 Exponential-LR(감쇠율 $\gamma = 0.999$)을 통해 점진적으로 감소시켰다. 식 (9)의 가중치는 각 손실항의 초기 크기 균형을 맞추기 위해 $\lambda_{feat} = 2$, $\lambda_{cham} = 10$, $\lambda_{corr} = 20$ 으로 설정하였다. 기준 메쉬에서는 점 5K개, 장면의 탐지 마스크에서는 점 1K개를 샘플링하여 비교하였으며, 제안하는 알고리즘의 주요 연산 비용(실행 시간 및 메모리 사용량 등) 분석 결과는 §4.7에 서술되어 있다.

4.3 평가 지표

정량적 평가는 일반적으로 비강체 정합의 성능평가에 널리 사용되는 End-Point-Error (EPE), Accuracy Strict (AccS), Accuracy Relaxed (AccR), Outlier Ratio를 사용하였다. EPE는 모든 점에 대한 변형 오차의 크기의 평균을 나타내며, AccS와 AccR은 각각 대응점의 오차가 물체 지름 대비 특정 비율 이하인 점의 비율로, 실험에서는 각각 1% 이하, 2.5% 이하인 점의 비율로 설정하였다. Outlier Ratio는 오차가 30%를 초과한 점의 비율로, 수렴 불안정성을 측정한다.

4.4 정량평가

4.4.1 초기 정렬 성능

§3.3에서 제안한 유사도 스코어링 기반 초기 정렬의 유효성을 검증하기 위해 중심점 정렬, PCA, FoundationPose^[17], FreeZe^[11] 및 제안 방법의 초기화에 따른 최종 정합 성능을 비교하였다.

[Table 1]과 같이, 제안 방법으로 초기화 한 경우, 타 방법 대비 가장 낮은 EPE를 기록하여 우수한 성능을 입증하였다. 중심

점 및 PCA 정렬은 Mesh-to-Scene 환경의 특성인 부분관측과 큰 시점 차이로 인해 초기화에 실패하였으며, 최신 6D 자세추정 연구인 FoundationPose나 FreeZe는 표준적인 강체 정합 벤치마크에서는 SOTA 성능을 보이지만, 본 실험과 같이 물체의 변형이 존재하는 경우 성능이 저하되었다. 이는 비교방법들이 강체 가정에 기반하고 있어, 기하적 불일치를 이상치로 처리하거나 Render-and-Compare 과정에서 템플릿과 관측 사이의 형상 차이로 추정에 실패하기 때문이다. 반면, 제안하는 NDF 기반 스코어링은 전체적인 피쳐 유사도를 기반으로 물체에 변형이 존재하는 상황에서도 강건한 초기 자세를 제공하여 안정적인 정합을 유도함을 확인하였다.

4.4.2 대응쌍 손실을 사용하지 않은 비강체 정합 성능

명시적 대응쌍(Correspondence Pair) 정보를 활용하지 않는 기존 연구들과 공정한 비교를 위해, 제안 방법에서도 대응쌍 손실(L_{corr})을 제외한 조건 하에 비강체 정합 성능을 검증하였다. 모든 비교 실험은 동일한 NDF기반 초기정렬 결과를 사용하여 수행되었으며, 정합 과정에서는 명시적 대응점 정보없이 특징 및 기하 정보만으로 정합을 수행하였다.

[Table 2]에서 확인할 수 있듯이, 제안 방법은 모든 평가지표에서 비교 방법들을 상회하는 성능을 기록하였다. 이는 본 연구

[Table 1] Quantitative comparison of non-rigid registration results under different coarse alignments.

Coarse pose estimation	EPE ↓ (mm)	AccS ↑ (%)	AccR ↑ (%)	Outlier ↓ (%)
Centroid	77.06	5.46	23.82	27.85
PCA	83.23	13.65	29.97	33.26
FoundationPose ^[17]	22.56	<u>54.08</u>	82.07	5.48
FreeZe ^[11]	<u>19.37</u>	53.04	<u>83.16</u>	<u>4.07</u>
NDF(Ours)	15.63	57.82	87.11	2.70

[Table 2] Quantitative comparison of non-rigid registration results without correspondence loss

Non-rigid Registration	EPE ↓ (mm)	AccS ↑ (%)	AccR ↑ (%)	Outlier ↓ (%)
NICP ^[1]	27.78	13.23	62.66	<u>3.37</u>
Nerfies ^[18]	58.58	13.52	35.72	18.82
NSFP ^[3]	30.81	19.46	59.15	5.31
DPF ^[5]	36.93	19.20	48.97	6.28
NDP ^[4]	<u>23.63</u>	39.71	<u>73.13</u>	4.20
OAREg ^[6]	25.24	<u>46.76</u>	69.58	5.16
NDF(Ours) w/o L_{corr}	17.70	49.71	83.94	3.00

[Table 3] Quantitative comparison of non-rigid registration results with correspondence loss. For all methods, point correspondences were established using NDF feature descriptors for a fair comparison.

Non-rigid Registration	EPE ↓ (mm)	AccS ↑ (%)	AccR ↑ (%)	Outlier ↓ (%)
NICP+corr	24.11	34.49	74.66	4.85
DPF+corr	33.43	23.12	56.78	5.54
NDP+corr	20.58	47.41	76.92	3.46
OAReg+corr	<u>17.20</u>	<u>55.42</u>	<u>81.34</u>	2.16
NDF(Ours)	15.63	57.82	87.11	<u>2.70</u>

가 단순히 기하적 거리를 최소화하는 것에 그치지 않고, 피쳐 일관성을 유지하면서 변형을 유도하도록 최적화했기 때문이다. 반면, Chamfer Distance와 같은 기하 거리 척도에 의존하는 기존 방법들은 외형은 정합이 된 것처럼 보일지라도, 실제로는 의미적으로 다른 부위 간 대응이 형성되는 오정합이 발생함을 확인하였다.

4.4.3 대응쌍 손실을 포함한 비강체 정합 성능

§3.3.1의 방법으로 구성된 대응쌍 집합 C_i 을 모든 비교군에 동일하게 적용하여 대응쌍 손실을 포함했을 때의 비강체 정합 성능을 비교하였다. 대변형이 존재하는 경우, 명시적 대응 정보가 정합 안정성에 중요한 역할을 하며 [Table 3]의 결과에서도 대응쌍 손실을 도입할 경우 전반적인 성능이 뚜렷하게 향상됨을 확인할 수 있다.

특히 제안 방법은 대응쌍 손실을 함께 사용할 때도 여전히 가장 낮은 EPE와 가장 높은 AccS/AccR을 기록하며, 기존 방법론 대비 우수한 성능을 보였다. 이는 제안된 프레임워크가 피쳐 기반 정합과 대응점 제약을 효과적으로 결합하여, 대규모 변형 상황에서도 강건하게 비강체 정합을 수행함을 의미한다. 나아가 [Fig. 2]의 정성적 결과에서 볼 수 있듯이, 제안 방법은 국소적인 변형이 심한 영역에서도 오정합 없이 형상을 복원하여, 정량적 지표의 향상이 실제 시각적 품질의 향상으로 직결됨을 입증하였다.

4.5 가려짐 강건성 평가

§4.1에서 추가로 생성한 심각한 가려짐 상황의 세트에 대해, 제안 방법의 가려짐 강건성을 정량적으로 평가하였다. 객체별 10장면씩, 총 100장의 이미지를 가려짐 수준별 20~30%, 30~40%, 40~50%로 구분하여 평가하였다. [Table 4]에서 확인할 수 있듯이, 제안 방법은 NDF에 인코딩된 파운데이션 피쳐와의 의미론적 일관성을 활용하여, 국소적인 가려짐이 발생하더라도

[Table 4] Quantitative evaluation of occlusion robustness at each occlusion level.

EPE (mm)	Nerfies	NSFP	DPF	NDP	OAReg	NDF (Ours)
20~30%	77.45	32.05	46.33	<u>29.60</u>	33.72	13.40
30~40%	86.08	49.02	63.26	<u>48.31</u>	53.61	32.01
40~50%	81.32	45.89	67.09	48.95	<u>48.11</u>	25.28

도 보이는 영역의 피쳐 정보만으로 기존 메쉬와 정합을 강건하게 추론할 수 있어 비교적 안정적인 정합 성능을 유지하였다.

4.6 정성평가

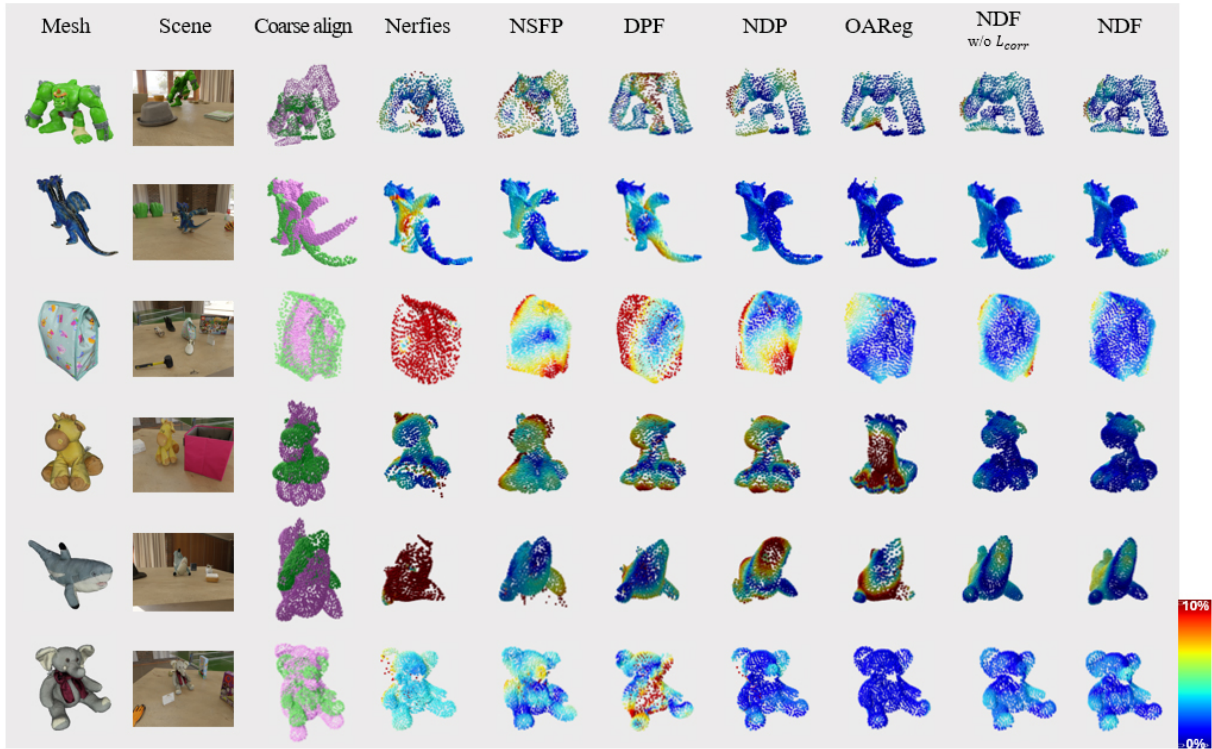
§4.1의 방법으로 구성된 합성 데이터셋에 대하여 각 방법에 따른 정합 결과에 시각적으로 평가하였다. 모든 방법은 동일한 NDF 기반 초기 정렬 결과를 입력으로 사용하였다. [Fig. 2]는 정합 결과에 대한 포인트별 변형 오차를 색상으로 시각화한 것이다.

결과에서 보듯 제안한 방법은 시각적으로도 다른 방법들에 비해 국소적 왜곡이나 오정합 없이 의미적으로 일관된 비강체 정합 결과를 달성하였다. 특히 세번째 행(가방)과 같이 표면이 매끄러워 기하적 특징 모호한 경우, 기존 방법들은 기하적 거리를 최소화하는 과정에서 최종 형상의 윤곽은 목표와 유사하게 복원하는 것처럼 보이나, 실제로는 대응점이 표면을 따라 미끄러지는 현상이 발생하여 텍스처나 패턴이 불일치하는 오정합이 관찰된다. 반면, 제안하는 방법은 시각 및 기하 피쳐를 활용하여 형상 뿐만 아니라 의미론적 대응관계까지 고려한 오정합 없는 일관된 결과를 달성하였다. 또한 여섯 번째 행(상어 인형)과 같이 얇고 긴 부위가 변형된 경우에도, 기존 방법들은 지느러미나 꼬리를 인접한 몸통 표면에 잘못 정합 시키는 국소 최적해 문제가 발생한 반면, 제안한 NDF는 강건한 정합을 수행함을 시각적으로 보여준다.

4.7 실행 시간 및 메모리 사용량

제안하는 방법의 실용성을 평가하기 위해, 기존 비강체 정합 방법들과 연산비용(추론 시간 및 GPU 메모리 사용량)을 비교하였다. 모든 실험은 동일한 RTX 4080 GPU 환경에서 수행되었으며, 결과는 [Table 5]와 같다.

본 방법은 GPU 메모리 사용량 측면에서 비교군 대비 높은 수치를 보였다. 이는 제안 방법의 핵심 요소인 파운데이션 기반 고투상도 피쳐가 메모리 집약적 구조를 가지기 때문이며, 단순 기하적 근접성만으로 해결할 수 없는 구조적 모호성을 해소하



[Fig. 2] Visual comparison of non-rigid registration results on the synthetic datasets. The figure shows qualitative results of various methods including Nerfies, NSFP, DPF, NDP, OAReg, compared with the proposed NDF framework. Each row represents a different scene, and color indicate the point-wise deformation error (blue = 0%, red = 10% of object diameter). While geometry-only methods exhibit local distortions or mismatched regions, our NDF produces semantically consistent and geometrically accurate alignments across all examples. In the coarse align column, the pink point cloud represents the reconstructed scene point cloud, and the green point cloud represents the points sampled from the mesh surface. Each result visualizes the scene point cloud after being deformed to the mesh shape by the corresponding method

[Table 5] Computational cost comparison of non-rigid registration (peak GPU memory usage and runtime, excluding coarse alignment)

	Nerfies	NSFP	DPF	NDP	OAReg	NDF (Ours)
Mem. (MB)	30.6	27.1	26.2	23.3	23.9	5616.0
Time (s)	4.87	0.21	2.42	0.45	6.04	3.62

고, 강건한 정합을 확보하기 위해 발생하는 트레이드 오프이다. 그럼에도 불구하고 전체 메모리 사용량은 보급형 GPU 환경에서 운용 가능한 수준이며, 추론 시간 또한 비교군 대비 유사한 수준을 유지하였다. 공정한 비교를 위해, 모든 방법에는 Coarse Alignment 단계(0.83s)를 제외한 순수 비강체 정합 시간만을 보고하였다. 이러한 결과는 제안 방법이 실시간성 요구가 없는 정합·스캐닝·로보틱스 응용에서 실질적으로 사용 가능한 효율성을 보유함을 의미한다.

4.8 Ablation Study

제안한 방법의 각 구성요소가 정합 성능에 미치는 영향을 분

[Table 6] Ablation Study on loss components

	EPE ↓ (mm)	AccS ↑ (%)	AccR ↑ (%)	Outlier ↓ (%)
[Only L_{feat}]	21.11	38.84	79.25	3.46
[w/o L_{feat}]	17.09	49.84	84.65	2.66
[Chamfer Dist.]	15.91	56.02	86.88	2.71
NDF (Ours)	15.63	57.82	87.11	2.70

석하기 위해 Ablation Study를 수행하였다. 손실 함수에서 특정 항을 제거하거나, 기존의 일반적인 손실 함수로 대체한 후 §4.1의 합성 데이터셋에서의 성능 변화를 측정하였다. [Table 6]은 각 실험 설정에 따른 정량적 결과를 요약한 것이며, 각 비교 설정에 대한 분석은 다음과 같다.

- [Only L_{feat}]: 기하적 제약 없이 NDF 피쳐 유사도 손실만을 사용하여 최적화를 수행한 경우이다. 이 경우 피쳐가 유사한 영역으로 점들이 이동하려는 경향은 보이지만, 표면 간의 정밀한 밀착을 유도하는 기하적 인력이 부족하여 EPE와 정확도(AccS)가 크게 하락하였다. 이는 정밀한 정합을 위해서는 기하적 제약 역시 필수적임을 의미한다.

- [w/o L_{feat}]: 피쳐 일관성 기반 Soft Chamfer Distance L_{cham} 와 대응쌍 손실 L_{corr} 만을 사용한 경우이다. 성능이 소폭 하락한 것은 NDF 피쳐 유사도 손실 L_{feat} 이 변형된 모든 포인트에 대해 NDF상의 올바른 위치로 지속적인 그래디언트를 제공하여, 대응쌍이 희소하거나, 적절한 최근접점이 없어 기하학적 손실이 커버하지 못하는 영역에서도 정합을 가이드하는 역할을 수행하기 때문이다.
- [Chamfer Dist.]: 제안한 피쳐 일관성 기반 Soft Chamfer Distance 대신 단순 Truncated Chamfer Distance를 사용한 경우이다. 기존 Chamfer Distance는 단순히 가장 가까운 점을 끌어당기기 때문에 구조적으로 모호한 형상에서 오정합이 빈번하게 발생하였다. 반면, 제안 방법은 피쳐 일관성 가중치를 통해 기하적으로 인접하면서도 의미적으로 일치하는 점들에 집중함으로써 더 높은 EPE와 정확도를 달성하였다.

5. 결론

본 연구에서는 별도의 대규모 사전 학습 없이, 인스턴스별 최적화만으로 수행 가능한 Neural Descriptor Field (NDF)기반의 비강체 Mesh-to-RGBD 정합 프레임워크를 제안하였다. 제안한 방법은 사전 학습된 파운데이션 모델의 시각 및 기하 피쳐를 결합하여, 3D 공간상의 각 점을 고차원 피쳐 공간으로 매핑하는 연속적 디스크립터 필드를 구성함으로써, 기하적 정보만으로는 해결하기 어려운 구조적 모호성을 해소하고 가림 상황에도 강건한 대응을 형성한다.

특히, 피쳐 유사도 기반의 자세 스코어링 기법을 도입하여 변형이 있는 물체의 초기 자세를 안정적으로 추정하였으며, 이후 Test-time 최적화를 통해 정밀한 비강체 정합을 달성하였다. 합성 데이터셋으로 실험한 결과, 제안한 방법은 기존 최신 기법 (OAReg) 대비 EPE를 약 38% 감소시켰으며, 정합 정확도 (AccR) 또한 87.11%를 달성하여 구조적 변형에 대한 강건성을 정량적으로 확인하였다.

향후 연구에서는 본 프레임워크를 실제 환경으로 확장하여, 비강체 물체의 로봇 조작 및 핸들링 작업으로의 적합성을 검증할 계획이다.

References

- [1] B. Amberg, S. Romdhani, and T. Vetter, "Optimal Step Nonrigid ICP Algorithms for Surface Registration," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, USA, pp. 1-8, 2007, DOI: 10.1109/CVPR.2007.383165.
- [2] R. W. Sumner, J. Schmid, and M. Pauly, "Embedded Deformation for Shape Manipulation," *ACM Transactions on Graphics*, vol. 26, no. 3, p. 80, 2007, DOI: 10.1145/1276377.1276478.
- [3] X. Li, J. K. Pontes, and S. Lucey, "Neural Scene Flow Prior," in *35th Conf. on Neural Information Proc. Systems*, pp. 7838-7851, 2021, [Online], <https://proceedings.neurips.cc/paper/2021/file/41263b9a46f68f22668476661614478-Paper.pdf>.
- [4] Y. Li and T. Harada, "Non-rigid Point Cloud Registration with Neural Deformation Pyramid," in *36th Conf. on Neural Information Proc. Systems*, New Orleans, USA, pp. 27757-27768, 2022, [Online], https://proceedings.neurips.cc/paper_files/paper/2022/hash/b2077e6d66da612fcb701589efa9ce88-Abstract-Conference.html.
- [5] S. Prokudin, Q. Ma, M. Raafat, J. Valentin, and S. Tang, "Dynamic Point Fields," in *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, Paris, France, pp. 7930-7942, 2023, DOI: 10.1109/ICCV51070.2023.00732.
- [6] M. Zhao, G. Meng, and D.-M. Yan, "Occlusion-aware Non-Rigid Point Cloud Registration via Unsupervised Neural Deformation Correntropy," in *Int. Conf. on Learning Representations (ICLR)*, Singapore, 2025 [Online], <https://openreview.net/pdf?id=cjJqU40nYS>.
- [7] M. Oquab et al., "DINOv2: Learning Robust Visual Features without Supervision," *Transactions on Machine Learning Research (TMLR)*, [Online], <https://openreview.net/pdf?id=a68SUt6zFt>, Accessed: Jan. 23, 2026.
- [8] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Int. Conf. on Learning Representations (ICLR)*, 2021, [Online], <https://openreview.net/pdf?id=YicbFdNTTy>.
- [9] S. Fu, M. Hamilton, L. E. Brandt, A. Feldmann, Z. Zhang, and W. Freeman, "FeatUp: A Model-Agnostic Framework for Features at Any Resolution," in *Int. Conf. on Learning Representations (ICLR)*, Vienna, Austria, 2024, [Online], https://proceedings.iclr.cc/paper_files/paper/2024/file/c5601d99ed028448f29d1dae2e4a926d-Paper-Conference.pdf.
- [10] F. Poiesi and D. Boscaini, "Learning General and Distinctive 3D Local Deep Descriptors for Point Cloud Registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3979-3985, 2023, DOI: 10.1109/TPAMI.2022.3175371.
- [11] A. Caraffa, D. Boscaini, A. Hamza, and F. Poiesi, "FreeZe: Training-Free Zero-Shot 6D Pose Estimation with Geometric and Vision Foundation Models," in *Computer Vision - ECCV 2024: 18th European Conf.*, Milan, Italy, pp. 414-431, 2024, DOI: 10.1007/978-3-031-73226-3_24.
- [12] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1-15, 2022, DOI: 10.1145/3528223.3530127.
- [13] V. N. Nguyen, T. Groueix, G. Ponimatkin, V. Lepetit, and T. Hodan, "CNOS: A Strong Baseline for CAD-based Novel Object Segmentation," in *IEEE/CVF Int. Conf. on Computer Vision Workshops (ICCVW)*, Paris, France, pp. 2126-2132, 2023, DOI: 10.1109/ICCVW60793.2023.00227.

- [14] Y. Li and T. Harada, "Lepard: Learning partial point cloud matching in rigid and deformable scenes," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, USA, pp. 5544-5554, 2022, DOI: 10.1109/CVPR52688.2022.00547.
- [15] L. Downs, A. Francis, N. Koenig, B. Kinman, R. Hickman, K. Reymann, T. B. McHugh, and V. Vanhoucke, "Google Scanned Objects: A High-Quality Dataset of 3D Scanned Household Items," in *Int. Conf. on Robotics and Automation (ICRA)*, Philadelphia, USA, pp. 2553-2560, 2022, DOI: 10.1109/ICRA46639.2022.9811809.
- [16] T. Hodaň, F. Michel, E. Brachmann, W. Kehl, A. G. Buch, D. Kraft, B. Drost, J. Vidal, S. Ihrke, X. Zabulis, C. Sahin, F. Manhardt, F. Tombari, T.-K. Kim, J. Matas, and C. Rother, "BOP: Benchmark for 6D Object Pose Estimation," in *Computer Vision - ECCV 2018: 15th European Conf.*, Munich, Germany, pp. 19-35, 2018, DOI: 10.1007/978-3-030-01249-6_2.
- [17] B. Wen, W. Yang, J. Kautz, and S. Birchfield, "FoundationPose: Unified 6D Pose Estimation and Tracking of Novel Objects," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Seattle, USA, pp. 17868-17879, 2024, DOI: 10.1109/CVPR52733.2024.01692.
- [18] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, "Nerfies: Deformable Neural Radiance Fields," in *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, Montreal, Canada, pp. 5845-5854, 2021, DOI: 10.1109/ICCV48922.2021.00581.



이 영 준

2017 연세대학교 기계공학과(학사)
 2017~현재 삼성전자 책임연구원
 2024~현재 서울대학교 기계공학과 석사과정

관심분야: Robot Vision, Object Pose Estimation



임 종 우

1997 서울대학교 계산통계학과 전산과학 전공(학사)
 2003 UIUC 컴퓨터과학(석사)
 2005 UIUC 컴퓨터과학(박사)
 2023~현재 서울대학교 기계공학과 교수

관심분야: Robot Vision, Depth Estimation, Visual SLAM, 3D Modeling, Visual Object Tracking