Diffusion 3D features (Diff3F): Decorating Untextured Shapes with Distilled Semantic Features

Niladri Shekhar Dutt[1,2] Sanjeev Muralikrishnan[1] Niloy J. Mitra[1,3]

[1] University College London [2] Ready Player Me [3] Adobe Research

CVPR JUNE 17-21, 2024 SEATTLE, WA

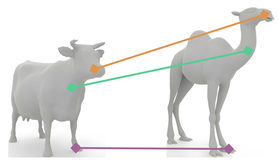
Motivation

- Meshes lack semantic features; existing methods only capture geometry
- Geometric features → curvature, normals, etc
- Semantic features → understands eyes, nose, neck, etc
- 2D Image Foundation models have semantic features but untextured meshes are not suitable for image descriptors
- Diff3F → generate textures using diffusion & extract intermediate features

$S \in \mathbb{R}^{V\times 3}$ Input Shape → **Diff3F** → $S \in \mathbb{R}^{V\times 1280}$ Decorated Shape

3D Shape Correspondence

- For every point in the source shape, find a corresponding point in the target



Point-to-Point Correspondence

Dense Correspondence

Source

DPC [1]

SE-ORNet [2]

Diff3F (Ours)

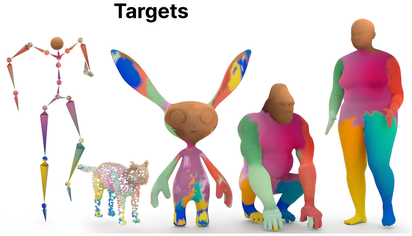
	3D-CODED [3]	DPC [1]	SE-ORNet [2]	FM+WKS [4+5]	Ours
No 3D training data?	✗	✗	✗	✓	✓
Unsupervised?	✗	✓	✓	✓	✓
Class agnostic?	✗	✗	✗	✓	✓
Handles meshes?	✓	✓	✓	✓	✓
Handles point cloud?	✓	✓	✓	✗	✓

No training

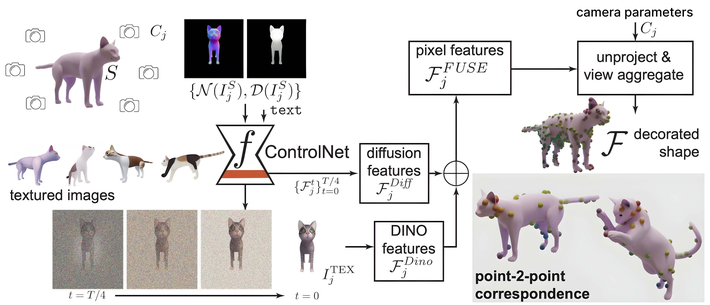
No 3D data required

Source

Targets



Method



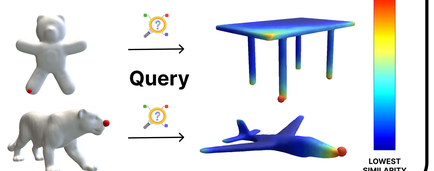
TL;DR: multi-view render → texture w/ ControlNet → extract diffusion features → unproject 2D features to 3D → aggregate

Method ↦ / ↓ Dataset	DPC [1] acc ↑	DPC [1] err ↓	SE-ORNet [2] acc ↑	SE-ORNet [2] err ↓	3DCODED [3] acc ↑	3DCODED [3] err ↓	FM [4] acc ↑	+WKS [5] err ↓	DIFF3F (ours) acc ↑	DIFF3F (ours) err ↓	DIFF3F (ours)+FM [4] acc ↑	DIFF3F (ours)+FM [4] err ↓
TOSCA	*30.79*	**3.74**	**33.25**	*4.32*	0.5*	19.2*	✗	✗	20.27	5.69	✗	✗
SHREC'19	17.40	6.26	21.41	4.56	2.10	8.10	4.37	3.26	**26.41**	*1.69*	*21.55*	**1.49**
SHREC'20	31.08	2.13	31.70	1.00	✗	✗	4.13	7.29	**72.60**	*0.93*	*62.34*	**0.71**

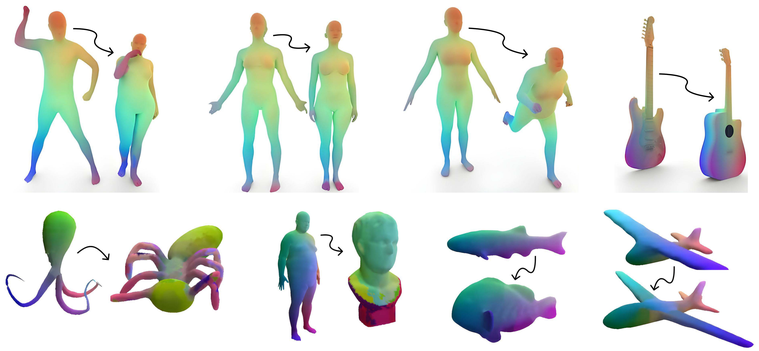
Ablation

	SHREC'19 acc ↑	SHREC'19 err ↓	SHREC'20 acc ↑	SHREC'20 err ↓
w/o ControlNet (untextured)	17.20	2.04	65.48	**0.69**
TEXTure [6] +DINO	17.20	2.04	65.48	**0.69**
w/o Fusion with DINO	**26.53**	2.06	64.89	1.60
w/o Normal Maps	25.68	**1.67**	69.71	1.17
w/o Time Aggregation	25.73	1.71	68.95	*0.87*
w/o Ball query	25.72	1.73	**74.10**	0.99
DIFF3F (full method)	*26.41*	*1.69*	*72.60*	0.93

Semantic Understanding

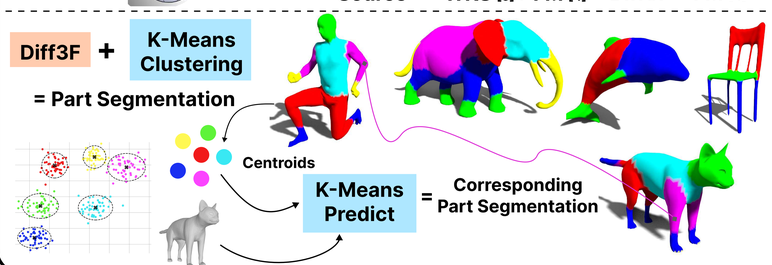


Results



Diff3F + **Functional Maps [4]** = Continuous Surface Maps

Diff3F + **K-Means Clustering** = Part Segmentation



SCAN ME

References

1. Lang, Itai, et al. DPC: Unsupervised deep point correspondence via cross and self construction. 3DV'21.
2. Deng, Jiacheng, et al. SE-ORNet: Self-ensembling orientation-aware network for unsupervised point cloud shape correspondence. CVPR'23.
3. Groueix, Thibault, et al. 3D-Coded: 3D correspondences by deep deformation. ECCV'18.
4. Ovsjanikov, Maks, et al. Functional maps: a flexible representation of maps between shapes. ACM ToG '12.
5. Aubry, Mathieu, et al. The wave kernel signature: A quantum mechanical approach to shape analysis. ICCVW'11.
6. Richardson, Elad, et al. Texture: Text-guided texturing of 3d shapes. SIGGRAPH'23.

ARC®

CVPR
JUNE 17-21, 2024
SEATTLE, WA

PRESENTER NAME:

Niladri Shekhar Dutt

PRESENTATION DATE:

June 18, 2024

AM

Poster Number /Paper ID

29710

TAG / NOTES

Doing both presentation & workshop, need poster on June 18 AM