Hao Huang

Personal Information

Status:	**PH.D. STUDENT**
Program:	**Computer Science and Engineering**
School:	**Tandon School of Engineering, New York University**
Website:	**https://haohuang40.github.io/**
Period:	**From 2019-09 to Present**

Biography

I am a Ph.D. student at New York University and advised by Professor Yi Fang. During my Ph.D. period, I work as a research assistant in NYU Multimedia and Visual Computing (MMVC) Lab. I am broadly interested in 3D Computer Vision and Deep Learning.

Research Project: Multi-image Matching Based on Model-free Cycle Consistency

Description

We present a novel model-free multi-images matching paradigm that 1) can estimate highly complex transformations without any prior knowledge or assumption about the transformations; 2) can predict correspondences among a group of images in the same category efficiently; 3) can be trained in an end-to-end fashion. Furthermore, we adopt cycle consistency as a *bridge* to match multiple images, which distinguishes our method from previous work that just utilized cycle consistency to improve the pair-wise image matching performance.

Method

To solve the problem mentioned above, we propose a model-free transformation estimation module to match images without any assumption about the types of transformations. The key of our method lays in regarding transformation estimation as motion prediction. Instead of regressing parameters of a certain type of transformations, we estimate a continuous motion field between a pair of images. The model-free module takes the best advantage of the powerful expressiveness of neural networks to model arbitrary transformation patterns. Our proposed method 1) can estimate transformations with high degree-of-freedom without any pre-defined assumption; 2) can preserve spatial continuity in estimated motion fields without applying any interpolation operation or adding any penalization term. Beyond pair-wise image matching, we further devise a multi-image matching framework based on the proposed model-free transformation estimation module. The most important constraint for joint matching is *cycle consistency* whose key idea is that the composition matches along a loop of images should be identity. We instantiate this high-level definition of cycle consistency into a concrete form under our model-free transformation setting. We develop an efficient framework to match multiple images simultaneously using cycle consistency as a "bridge" to connect different images. Both the model-free module and the multi-image matching framework can be trained jointly in

an end-to-end manner.

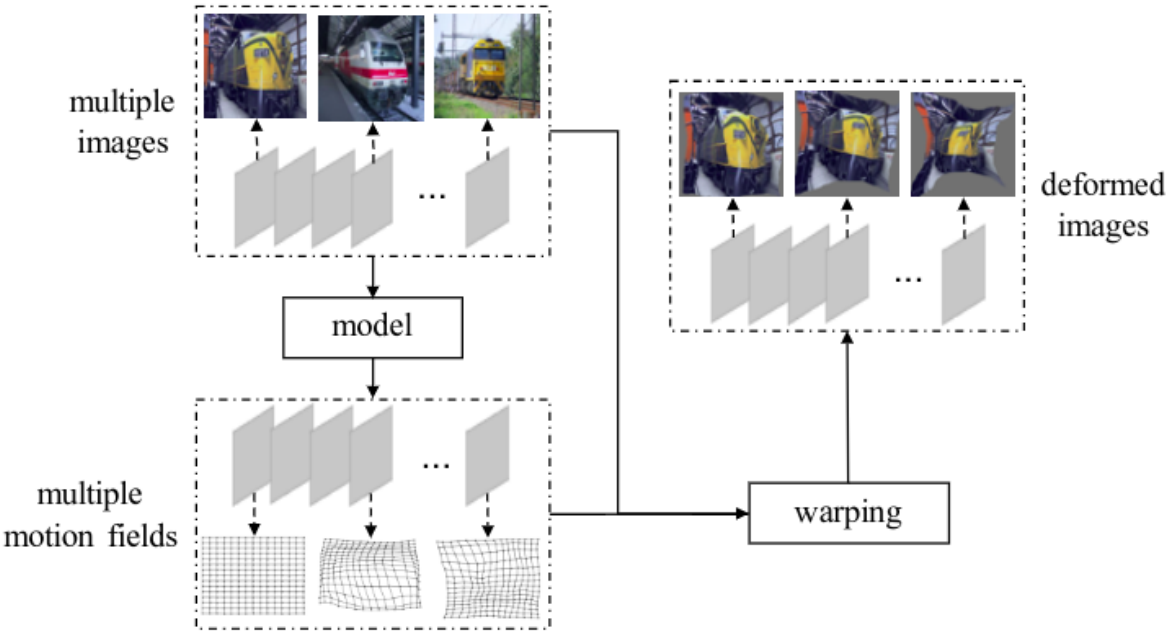
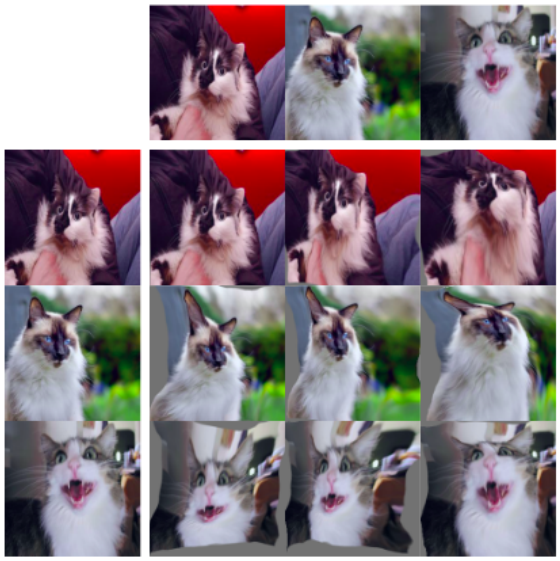


Figure 1: Instead of matching only one pair of images, our model takes in a collection of images andestimate transformations betweeneachpair simultaneously.

Results

We evaluate our proposed model on two datasets: PF-Pascal and PF-WILLOW. The PF-Pascal dataset consists of 1,351 semantically related image pairs. Annotated correspondences for each image pair are provided. We follow the split from to split all image pairs into training, validation and test sets. The PF-WILLOW dataset comprises 100 images which are grouped into 900 image pairs. All pairs are divided into four semantically related subsets. For each image, 10 keypoint annotations are provided. Note that all manual annotations are only used for evaluation. We adopt the percentage of correct keypoints (PCK) metric to evaluate our model. PCK measures the percentage of keypoints whose transformation errors are below a given threshold. The threshold is defined as $\alpha \max(h,w)$ where h and w are height and width of the object bounding box. For both datasets, we set the threshold $\alpha = 0.1$. Experiments on two real image datasets setup a benchmark for the following similar research.



(a) A 3-cycle example of PF-Pascal dataset.

(b) A 2 cycle example of PF-WILLOW dataset.

Figure 2: In each sub-figure, the first row and the first column are the original images. The diagonal images represent the results of identity transformations. (a) An example of qualitative results from PF-Pascal dataset. (b) An example of qualitative results from PF-WILLOW dataset.

Method	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	d.table	dog	horse	moto	person	plant	sheep	sofa	train	tv	mean
LOM [10]	73.3	74.4	54.4	50.9	49.6	73.8	72.9	63.6	46.1	79.8	42.5	48.0	68.3	66.3	42.1	62.1	65.2	57.1	64.4	58.0	62.5
Ours 2-cycle	79.2	73.0	71.0	47.2	61.0	75.2	86.4	71.5	53.5	81.3	51.6	61.0	54.1	66.6	50.0	65.7	60.0	46.2	62.8	42.2	64.5
Ours 3-cycle	79.7	72.7	72.3	38.9	64.1	73.7	84.2	69.3	54.8	81.3	48.4	58.5	55.0	67.4	50.0	55.7	60.0	49.2	61.8	45.6	65.6
Ours 4-cycle	77.0	72.2	65.6	45.8	64.1	75.3	85.7	72.1	60.4	79.2	51.0	60.9	56.2	65.4	51.8	68.6	60.0	46.2	65.0	45.6	65.7
Ours 5-cycle	75.1	74.9	67.7	47.2	64.1	74.6	87.1	70.8	52.2	72.9	56.3	59.2	56.2	68.0	52.3	65.7	60.0	49.3	62.8	47.2	65.8

Figure 3: Per-class PCK on the PF-Pascal dataset with $\alpha = 0.1$

Method	car(S)	car(G)	car(M)	duc(S)	mot(S)	mot(G)	mot(M)	win(w/o C)	win(w/C)	win(M)	Avg.
LOM [10]	0.86	0.58	0.52	0.65	0.48	0.28	0.28	0.91	0.37	0.65	0.56
CGC [33]	0.89	0.62	0.56	0.70	0.49	0.31	0.28	0.91	0.52	0.72	0.60
Ours 2-cycle	0.73	0.50	0.62	0.62	0.53	0.33	0.38	0.92	0.38	0.49	0.55
Ours 3-cycle	0.75	0.50	0.63	0.62	0.52	0.34	0.38	0.94	0.38	0.50	0.56
Ours 4-cycle	0.78	0.53	0.66	0.63	0.53	0.35	0.36	0.94	0.39	0.51	0.57
Ours 5-cycle	0.79	0.52	0.66	0.63	0.55	0.36	0.36	0.94	0.39	0.52	0.57

Figure 4: Per-class PCK on the PF-WILLOW dataset with $\alpha = 0.1$

Research Project: Robust Image Matching By Dynamic Feature Selection

Description

Estimating dense correspondences between images is a long-standing image understanding task. Recent works introduce convolutional neural networks (CNNs) to extract high-level feature maps and find correspondences through feature matching. However, high-level feature maps are in low spatial resolution and therefore insufficient to provide accurate and fine-grained features to distinguish intra-class variations for correspondence matching. To address this problem, we generate robust features by dynamically selecting features at different scales. To resolve two critical issues in feature selection, i.e., how many and which scales of features to be selected, we frame the feature selection process as a sequential Markov decision-making process (MDP) and introduce an optimal selection strategy using reinforcement learning (RL). Experimental results show that our method achieves comparable/superior performance with state-of-the-art methods on three benchmarks, demonstrating the effectiveness of our feature selection strategy.

Method

We frame the feature selection problem as a sequential Markov decision-making process (MDP) and tackle it using reinforcement learning. Specifically, based on the selected features, each individual action either requires new features or terminates the selection episode by referring a matching score. The learning process is driven by reward functions. Without manually imposed prior knowledge about image pairs, the proposed method can select the optimal collection of features that are suitable for image matching. Compared with beam search, there is no strict selection order in our proposed method, i.e., from low to high levels, leading to a larger search space and a higher possibility to find the optimal solution. We test the proposed method on three public datasets to demonstrate the effectiveness of our proposed feature selection strategy for robust image matching.

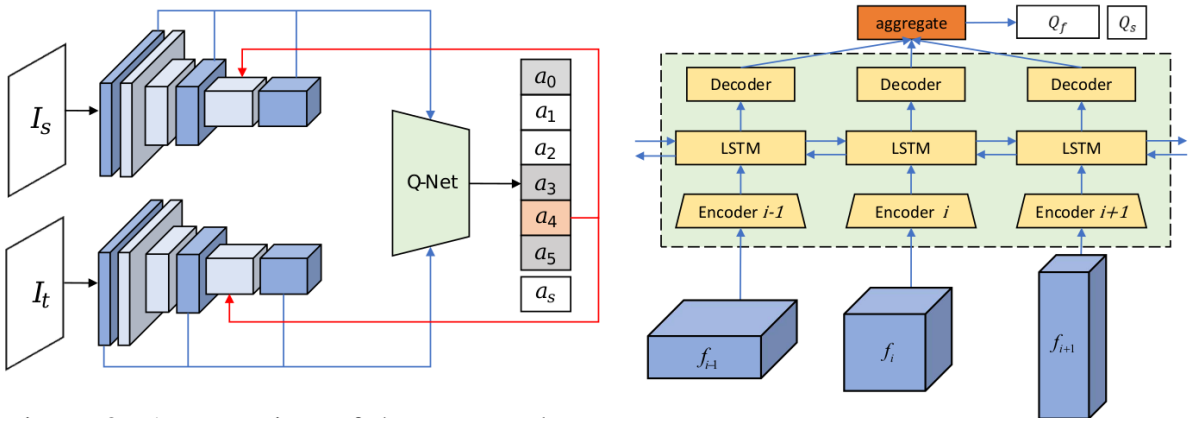


Figure 5: (Left) An overview of the proposed approach. Actions in grey color are invalid actions at the current step, as the corresponding features have been selected previously. (Right) Q-network. It adopts the encoder- LSTM-decoder structure. The input are the current state, i.e., selected features, and the output is the predicted Q-value.

Results

We report the average PCK scores of our method and recent methods that are directly comparable on PF-PASCAL dataset. Our proposed method achieves state-of-the-art performance. For PF-WILLOW dataset, in order to verify the generalization ability of our proposed method, we directly apply the same layers selected for PF-PASCAL dataset and test on PF-WILLOW dataset without fine-tuning. Similar to PF-WILLOW dataset, we directly test our method on Caltech-101 using selected layers based on PF-PASCAL without fine-tuning. Our method achieves the same performance as state-of-the-art methods.

	Model	PCK ($\alpha = 0.1$) PF-WILLOW	PF-PASCAL
Hand-crafted	DeepFlow [35]	0.20	0.21
	GMK [10]	0.27	0.27
	DSP [19]	0.29	0.30
	SIFTFlow [26]	0.38	0.33
	ProposalFlow [12]	0.56	0.45
CNN-based	FCSS + PF-LOM [20]	0.58	0.46
	GeoCNN (SS) [36]	0.68	0.68
	A2Net [16]	0.69	0.67
	GeoCNN (WS) [37]	0.71	0.72
	SFNet [22]	<u>0.74</u>	0.79
	HPFlow [29]	<u>0.74</u>	<u>0.85</u>
	Ours	**0.75**	**0.86**

	Methods	LT-ACC	IoU
Hand-crafted	DeepFlow [35]	0.74	0.40
	SIFTFlow [26]	0.75	0.48
	GMK [10]	0.77	0.40
	DSP [19]	0.77	0.47
	ProposalFlow [12]	0.78	0.50
	OADSC [43]	0.81	0.55
CNN-based	SCNet-AG[14]	0.79	0.51
	A2Net [16]	0.80	0.57
	FCSS + PF-LOM [20]	0.83	0.52
	GeoCNN (SS) [36]	0.83	0.61
	GeoCNN (WS) [37]	0.85	<u>0.63</u>
	SFNet [22]	**0.88**	**0.67**
	HPFlow [29]	<u>0.87</u>	<u>0.63</u>
	Ours	<u>0.87</u>	<u>0.63</u>

Figure 6: (Left) The average PCK results on PF-WILLOW and the test split of PF-PASCAL dataset with $\alpha = 0.1$. Numbers of the top-1 performance are in bold and the top-2 performance are underlined. (Right) The average quantita- tive results on Caltech-101 dataset. Numbers of the top-1 performance are in bold and the top-2 performance are underlined.

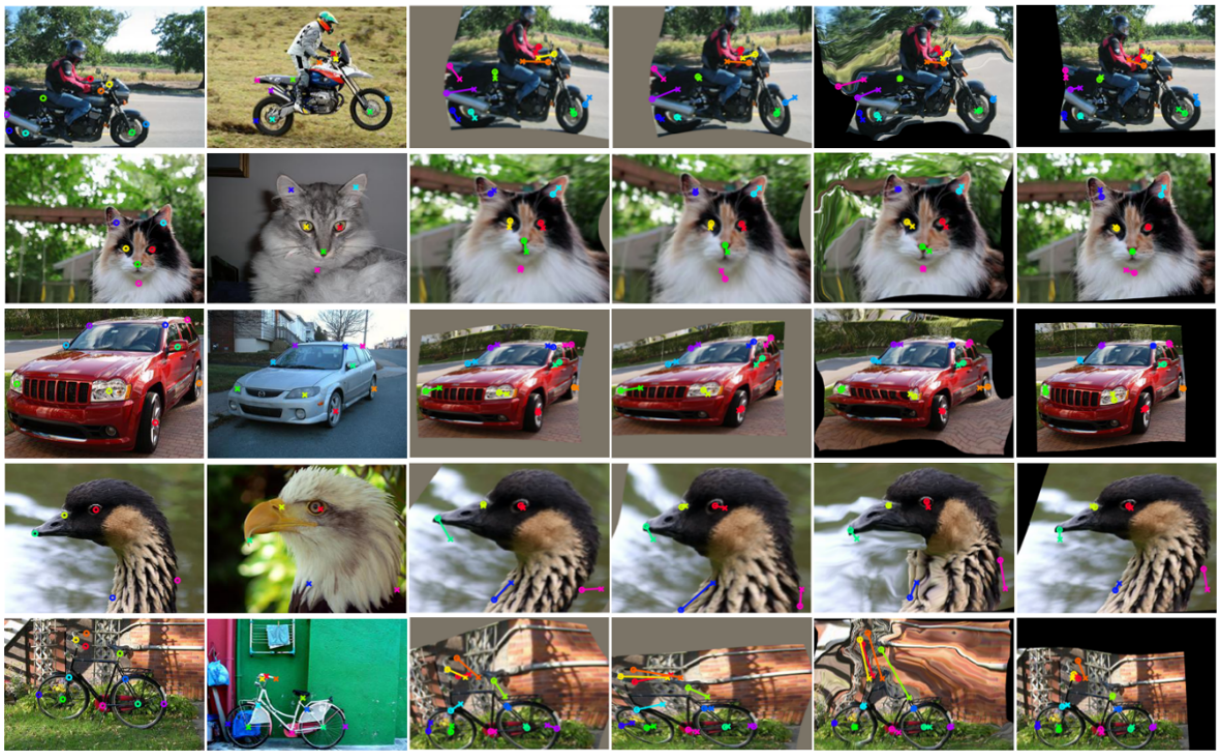


Figure 7: Examples of qualitative results from PF-PASCAL dataset. Keypoints of the source and target images are shown in circles and crosses, respectively. Compared to GeoCNN (WS), A2Net and SFNet, our method is more robust to intra-class variations.

Research Project: Learning to Learn 3D Point Signature for 3D Dense Shape Correspondence

Description

Point signature, a representation describing the structural neighborhood of a point in 3D shapes, can be applied to establish correspondences between points in 3D shapes. Conventional methods apply a weight-sharing network, e.g., graph neural network, across all neighborhoods to generate point signatures directly and gain the generalization ability by extensive training over a large number of training samples from scratch. However, these methods lack the flexibility in rapidly adapting to unseen neighborhood structures and thus generalizes poorly on new point sets. In this paper, we propose a novel meta-learning based 3D point signature model, named 3D meta point signature (MEPS) network, that is capable of learning robust point signatures in 3D shapes. We evaluate the MEPS model on a dataset for 3D shape correspondence. Experimental results demonstrate that our method gains significant improvements over the baseline model and achieves state-of-the-art results.

Method

We propose to develop a novel meta-learning based 3D point signature model, dubbed as 3D meta point signature (MEPS) network, that is capable of dynamically capture neighborhood features including unseen ones and thus generating robust point signatures. We treat the process of learning point signature in each neighborhood as a task. Instead of directly mapping each center point along with its neighboring points into a point signature by a network model, we introduce a meta- learner to generate a group of models, denoted as base-learners, that achieve optimal performance for each corresponding task. Compared with conventional methods, the meta-learner in our MEPS network learns the distribution of all tasks and generates base-learner models that are dynamically tailored for unseen tasks based on the task distribution. Each concrete point signature learning process is performed by different base-learners, while the meta-learner learns to "learn point signature".

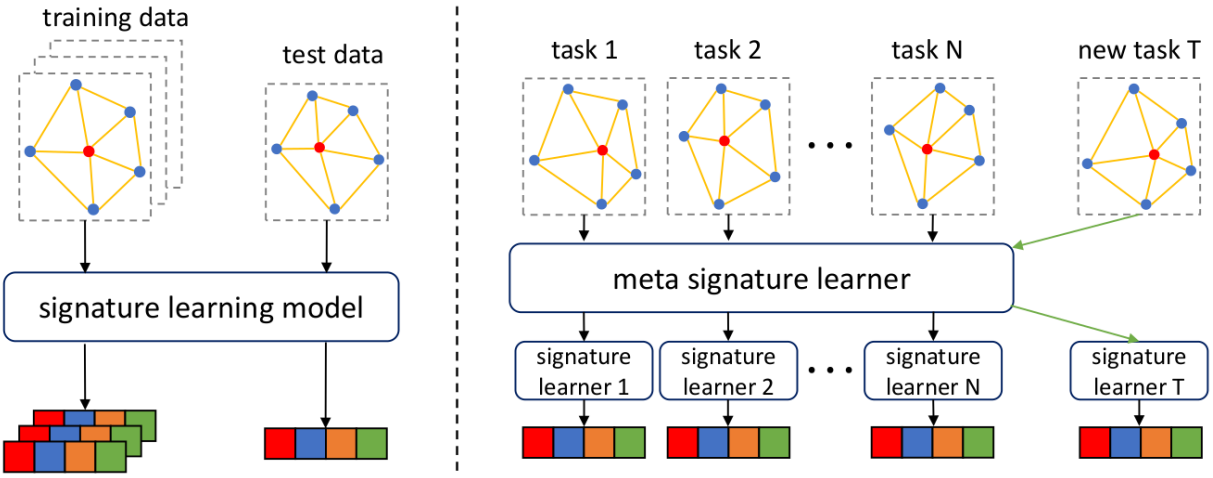


Figure 8: (Left) Conventional point signature learning process. (Right) Our 3D meta point signature (MEPS) network. Colored bar represents generated point signatures.

Results

We evaluate the proposed method on a 3D vision dataset, i.e., FAUST dataset for shape correspondence estimation, and achieve state-of-the-art performance by obtaining a significant improvement over the baseline model. Note that the proposed method can be further modified and adapted to 2D image correspondence by treating 2D pixels as 3D points.

Method	Input	Accuracy
Logistic Regression	SHOT	39.9%
PointNet [30]	SHOT	49.7%
GCNN [24], w/o refinement	SHOT	42.3%
GCNN [24], w/ refinement	SHOT	65.4%
ACNN [5], w/o refinement	SHOT	60.6%
ACNN [5], w/ refinement [29]	SHOT	62.4%
MoNet [26], w/o refinement	SHOT	73.8%
MoNet [26], w/ refinement [40]	SHOT	88.2%
FeaStNet-S [39], w/o refinement	XYZ	88.1%
FeaStNet-S [39], w/ refinement [40]	XYZ	92.2%
Ours, w/o refinement	XYZ	**94.0%**
Ours, w/ refinement [40]	XYZ	**98.7%**

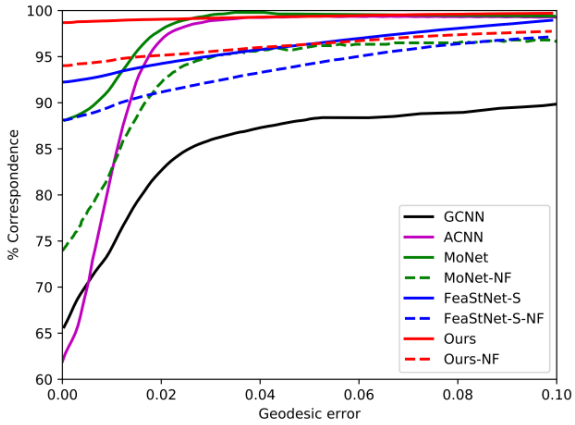


Figure 9: (Left) Comparison of correspondence accuracy on FAUST dataset of our model (single-scale) and the state-of-the-art approaches. Note that "-S" denotes single-scale architecture. (Right) Comparison of fraction of geodesic shape correspondence errors within a certain distance with state-of-the-art approaches. The "-NF" denotes results without refinement.

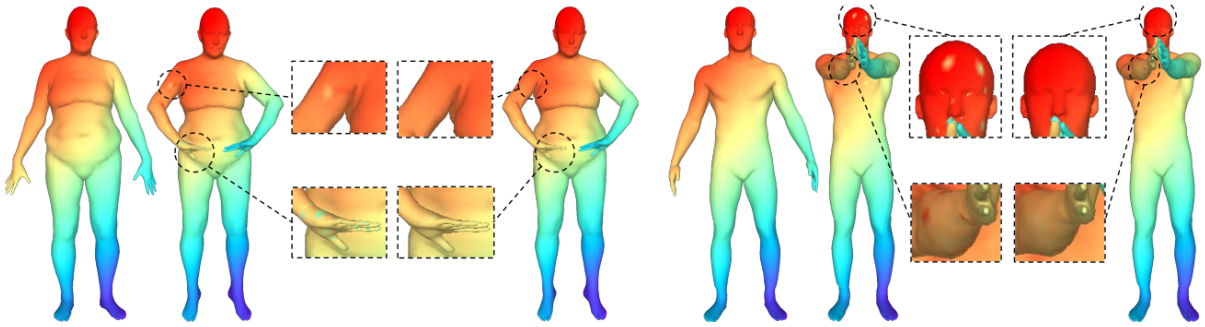


Figure 10: Two groups of correspondence results obtained by our approach on FAUST test set: reference shape (left), result without refinement (middle), result with refinement (right). Corresponding points are painted with same color.