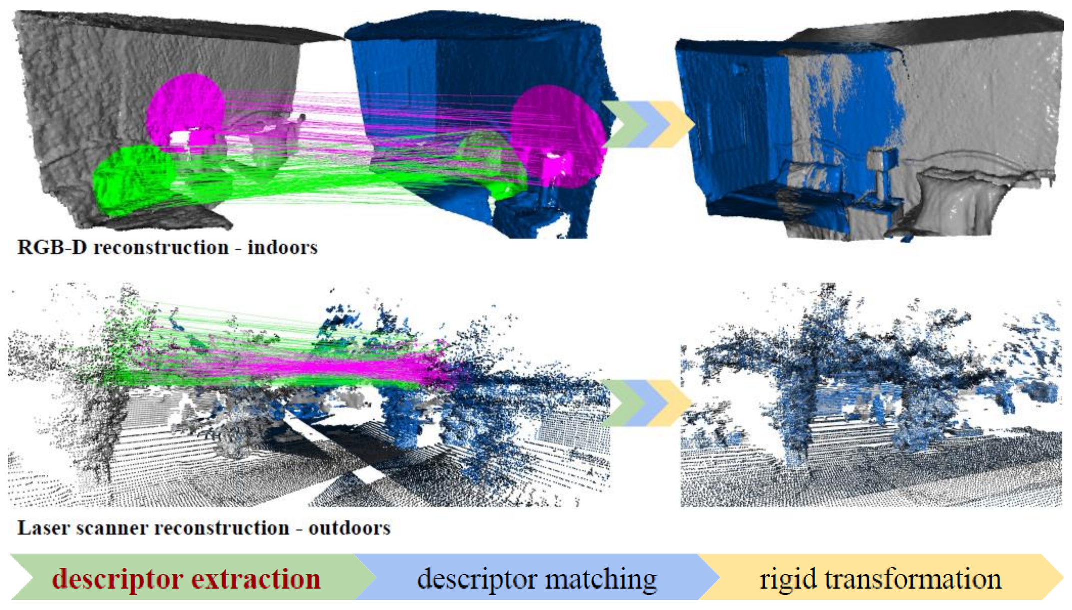
Distinctive 3D local deep descriptors

Fabio Poiesi, Davide Boscaini

<poiesi, dboscaini>@fbk.eu

1. Introduction

Goal: build robust descriptors to register point clouds without requiring an initial alignment



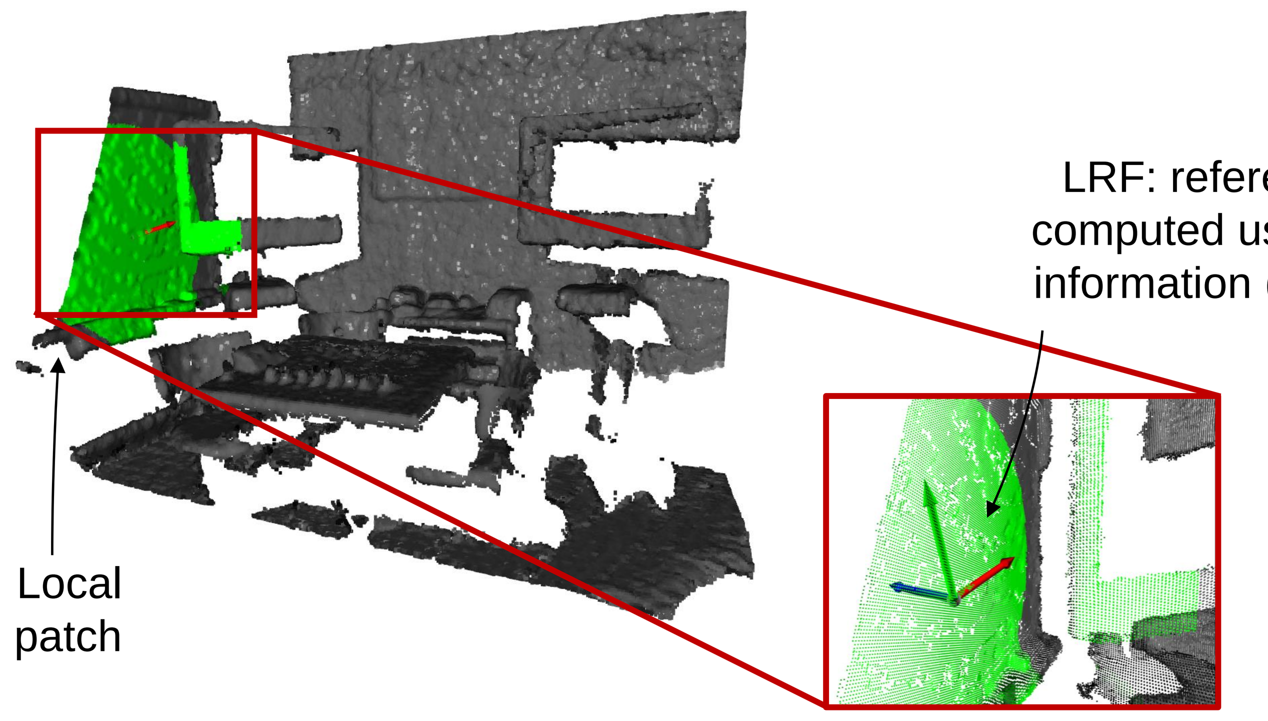
Desired descriptor properties

- Compact
- Efficient to compute
- Generalise across sensor modalities
- Learnable end-to-end

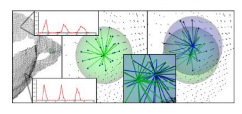
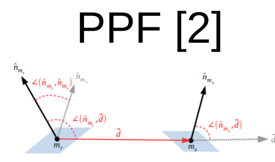
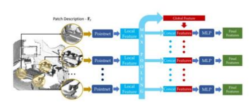
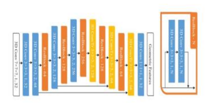
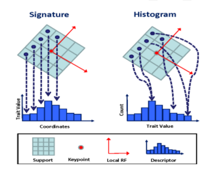
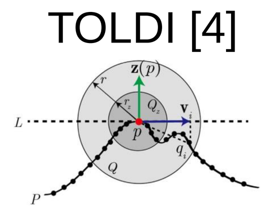
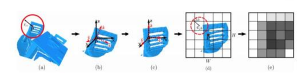
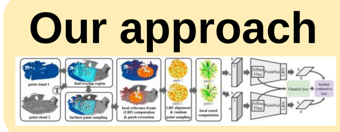
Our method highlights

- Achieves high scalability because it computes descriptors over local patches
- Produces rotation-invariant descriptors thanks to a Local Reference Frame transformation
- Learns an attention mechanism to quantify the quality of each descriptor

2. What is the Local Reference Frame (LRF)?



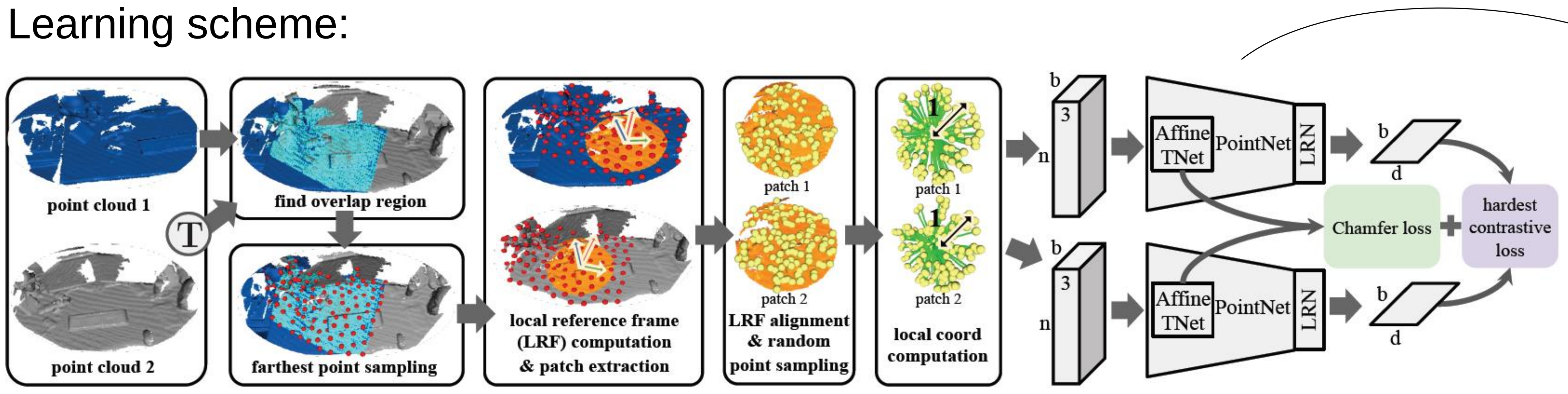
3. Related work

	Hand crafted	Data driven
One-stage (without LRF)	FPFH [1]  PPF [2] 	PPFNet [5]  FCGF [6] 
Two-stage (with LRF)	SHOT [3]  TOLDI [4] 	3DSmoothNet [7]  **Our approach** 

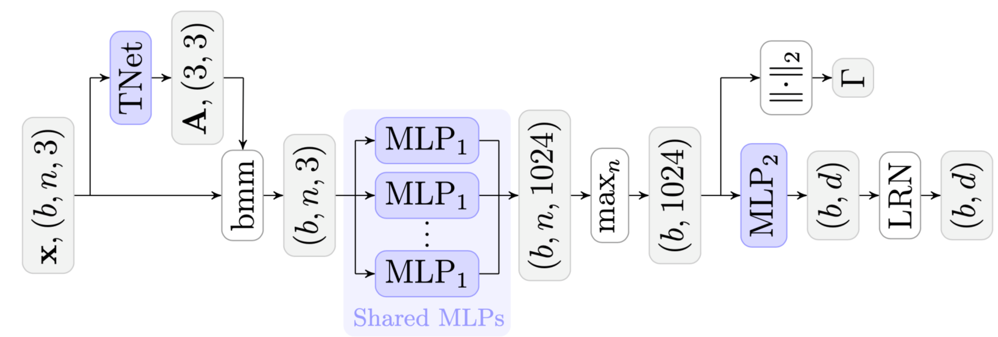
4. Our approach

Tnet: Transformation Network
MLP: Multi Layer Perceptron
LRN: Local Response Normalisation

Learning scheme:



PointNet backbone



Loss functions:

Global loss

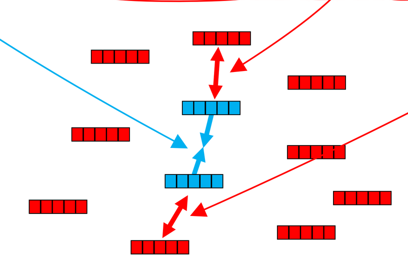
$$\ell = \ell_h + \frac{1}{b} \sum_{\mathcal{X} \in \mathcal{P}} \ell_c(\mathcal{X})$$

hardest contrastive loss — Chamfer loss

Hardest contrastive loss [6]

$$\ell_h = \frac{1}{b} \sum_{(\mathbf{f},\mathbf{f}') \in \mathcal{C}_+} \left(\frac{1}{|\mathcal{C}_+|} [d(\mathbf{f},\mathbf{f}') - m_+]_+^2 + \frac{1}{2|\mathcal{C}_-|} [m_- - \underbrace{\min_{\tilde{\mathbf{f}} \in \mathcal{C}_-} d(\mathbf{f},\tilde{\mathbf{f}})}_{d(\mathbf{f},\mathbf{f}_-)}]_+^2 + \frac{1}{2|\mathcal{C}_-|} [m_- - \underbrace{\min_{\tilde{\mathbf{f}} \in \mathcal{C}_-} d(\mathbf{f}',\tilde{\mathbf{f}})}_{d(\mathbf{f}',\mathbf{f}'_-)}]_+^2 \right)$$

$\mathbf{f}$ descriptor
$(\mathbf{f}, \mathbf{f}')$ pair of anchors
$(\mathbf{f}_-, \mathbf{f}'_-)$ hardest negatives
m_- m_+ margins



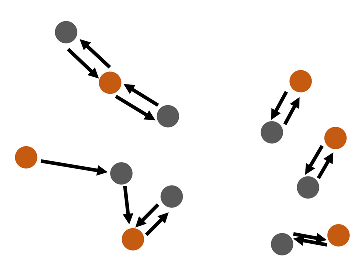
Descriptors (illustration)
positive
negative

Chamfer loss [8]

$$\ell_c(\mathcal{X}) = \frac{1}{2n} \left(\sum_{\mathbf{x} \in \mathcal{X}} \min_{\mathbf{x}' \in \mathcal{X}'} \|\mathbf{A}\mathbf{x} - \mathbf{A}'\mathbf{x}'\|_2 + \sum_{\mathbf{x}' \in \mathcal{X}'} \min_{\mathbf{x} \in \mathcal{X}} \|\mathbf{A}\mathbf{x} - \mathbf{A}'\mathbf{x}'\|_2 \right)$$

$\mathbf{A} \in \mathbb{R}^{3\times3}$
unconstrained during training

● $\mathbf{x}'$
● $\mathbf{x}$

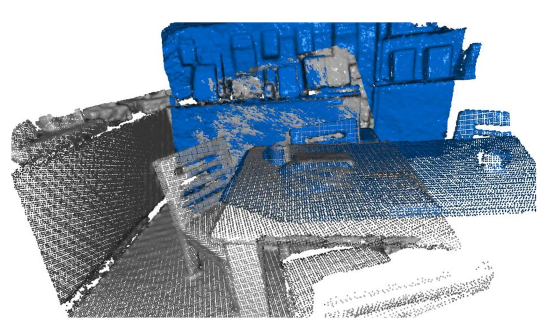


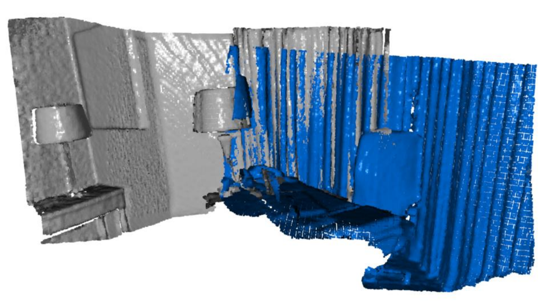
5. Results

Training
- 3DMatch dataset [9]
- about 16K point-cloud pairs
- each pair is 256 descriptors
- 40 epochs

Testing
- 3DMatch and ETH dataset [7]

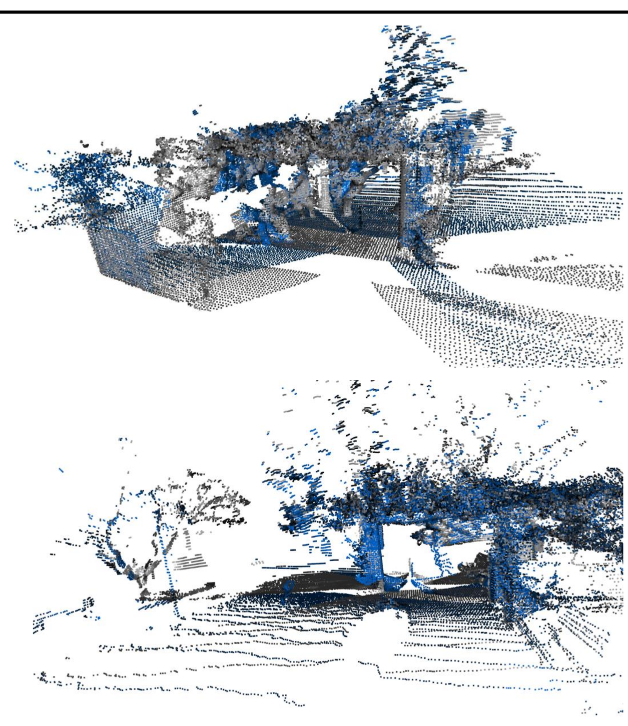
Evaluation
- Feature Matching Recall [7]





3DMatch dataset (train RGBD → test RGBD)

Method	3DMatch Ξ	3DMatch std	3DMatchRotated Ξ	3DMatchRotated std	Feat. dim.	Time [ms]
Spin	.227	.114	.227	.121	153	.133
SHOT	.238	.109	.234	.095	352	.279
FPFH	.359	.134	.364	.136	33	.032
USC	.400	.125	-	-	1980	3.712
CGF	.582	.142	.585	.140	32	1.463
3DMatch	.596	.088	.011	.012	512	3.210
Folding	.613	.087	.023	.010	512	.352
PPFNet	.623	.108	.003	.005	64	2.257
PPF-FoldNet	.718	.105	.731	.104	512	.794
DirectReg	.746	.094	-	-	512	.794
CapsuleNet	.807	.062	.807	.062	512	1.208
PerfectMatch	.947	.027	.949	.024	32	5.515
FCGF	.952	.029	.953	.033	32	.009
D3Feat	**.958**	.029	**.955**	.035	32	-
DIP	.948	.046	.946	.046	32	4.870



ETH dataset (train RGBD → test LIDAR)

Method	Gazebo Summer	Gazebo Winter	Wood Autumn	Wood Summer	Average
FPFH	.386	.142	.148	.208	.221
SHOT	.739	.457	.609	.640	.611
3DMatch	.228	.083	.139	.224	.169
CGF	.375	.138	.104	.192	.202
PerfectMatch	**.913**	.841	.678	.728	.790
FCGF	.228	.100	.148	.168	.161
D3Feat	.859	.630	.496	.480	.563
DIP	.908	**.886**	**.965**	**.952**	**.928**

6. Conclusions

- Compact descriptor
- Efficient to compute
- Generalise across sensor modalities
- Learnable end-to-end

CODE AVAILABLE
https://github.com/fabiopoiesi/dip

acknowledgment: FONDAZIONE CARITRO

References

[1] R.B. Rusu et al., "Fast Point Feature Histograms (FPFH) for 3D registration," ICRA 2009
[2] B. Drost, et al., "Model globally, match locally: Efficient and robust 3d object recognition," CVPR 2010
[3] F. Tombari et al., "Unique Signatures of Histograms for Local Surface Description," ECCV 2010
[4] Yang et al., "TOLDI: An effective and robust approach for 3D local shape description," Patt. Rec. 2017
[5] Deng et al., "PPFNet: Global Context Aware Local Features for Robust 3D Point Matching," CVPR 2018
[6] Choy et al. "Fully Convolutional Geometric Features," ICCV 2019
[7] Gojcic et al., "The Perfect Match: 3D Point Cloud Matching with Smoothed Densities," CVPR 2019
[8] Y. Zhao et al., "3D point capsule networks," CVPR 2019
[9] A. Zeng et al., "3DMatch: Learning the matching of local 3D geometry in range scans," CVPR 2017