DispVoxNets: Non-Rigid Point Set Alignment with Supervised Learning Proxies

Soshi Shimada\textsuperscript{1,2} Vladislav Golyanik\textsuperscript{3} Edgar Tretschk\textsuperscript{3}
Didier Stricker\textsuperscript{1,2} Christian Theobalt\textsuperscript{3}

\textsuperscript{1}University of Kaiserslautern \quad \textsuperscript{2}DFKI \quad \textsuperscript{3}MPI for Informatics, SIC
Non-rigid Point Set Registration (NRPSR)

Objective: given two point sets, find displacements (or correspondences) between the point sets.

2D point set registration
[Myronenko and Song 2010]

3D face registration
[Taetz et al. 2016]

3D pose registration
[Golyanik et al. 2017]
Related Works, General-Purpose Methods
Related Works, General-Purpose Methods

Iterative Closest Point (ICP)
[Besl and McKay 1992]
the image is taken from [Smistad et al. 2015]
Related Works, General-Purpose Methods

Iterative Closest Point (ICP)
[Besl and McKay 1992]
the image is taken from [Smistad et al. 2015]

Gaussian Mixture Model Registration (GMR)
[Jian et al. 2005]
Related Works, General-Purpose Methods

Iterative Closest Point (ICP)  
[Besl and McKay 1992]  
the image is taken from [Smistad et al. 2015]

Gaussian Mixture Model Registration (GMR)  
[Jian et al. 2005]

Coherent Point Drift (CPD)  
[Myronenko and Song 2010]
Related Works, General-Purpose Methods

Gravitational Approach for NRPSR
Often fails with large deformations and articulated motions between the point sets.
Related Works, General-Purpose Methods

Target

CPD

GLTP

[Ge et al. 2014]
Related Works, General-Purpose Methods

Target

CPD

GLTP
[Ge et al. 2014 ]

Relatively accurate however sensitive to noises
Related Works, Class-Specific Methods

[Ge and Fan 2015]
Perform well with large deformations and articulated motions between the point sets. However, the generalisability is limited.
Related Works, Neural Network Based Approaches (Other Fields)

3D-PhysNet
[Wang et al. 2018]

DEMEA
[Tretschk et al. 2019]
Pipeline
Pipeline

Y

\(\text{(Mx3)}\)

X

\(\text{(Nx3)}\)

- \(Y\): template point set, \(X\): reference point set
Pipeline

- $Y$: template point set, $X$: reference point set
- Assume $M$ is not equal to $N$ in general
Pipeline

- **Y**: template point set, **X**: reference point set
- Assume $M$ is not equal to $N$ in general

Displacement Estimation (DE) → Refinement

$Y + v(Y, X)$ (Mx3)
Pipeline

- Y: template point set, X: reference point set
- Assume M is not equal to N in general
- DE stage regresses global displacements between Y and X
Pipeline

Y: template point set, X: reference point set
Assume M is not equal to N in general
DE stage regresses global displacements between Y and X
Refinement stage improves the initial displacements
Pipeline

- Y and X are firstly converted into voxel representation (P2V)
● Y and X are firstly converted into voxel representation (P2V)
● During the conversion, point-voxel correspondence information is stored in an affinity table
Y and X are firstly converted into voxel representation (P2V)

During the conversion, point-voxel correspondence information is stored in an **affinity table**
Y and X are firstly converted into voxel representation (P2V).
During the conversion, point-voxel correspondence information is stored in an affinity table.
DispVoxNet accepts two voxel grids and returns voxel displacements.
• Y and X are firstly converted into voxel representation (P2V)
• During the conversion, point-voxel correspondence information is stored in an affinity table
• DispVoxNet accepts two voxel grids and returns voxel displacements
• The displacements are applied using the affinity table at the end of DE stage
The outputs from the DE stage are further sent to the Refinement stage after P2V.
The outputs from the DE stage are further sent to the Refinement stage after P2V.

The new instance of DispVoxNet returns small displacements for refinement.
The outputs from the DE stage are further sent to the Refinement stage after P2V
The new instance of DispVoxNet returns small displacements for refinement
The inferred displacements are added to the template points
Pipeline

\[ \mathcal{L}_{\text{Disp.}} \rightarrow \text{GT displacement} \]

Displacement Estimation (DE)

\[ (3 \times Q^3) \]

Refinement

\[ (3 \times Q^3) \]

Template (after DE)

\[ (Q^3) \]

\[ Y + v(Y, X) \]

(\(M \times 3\))

\( (N \times 3) \)
The network in the DE stage is trained in a supervised manner (displacement loss)
The network in the DE stage is trained in a supervised manner (displacement loss)

The network in the Refinement stage is trained in an unsupervised manner (point projection loss)
Loss Functions - Displacement Loss

\[ \mathcal{L}_{\text{Disp.}} = \left\| \text{Network output} - \text{GT displacement} \right\|_2^2 \]
Loss Functions - Point Projection Loss

(I) After DE Stage

PP Loss Computation

(II) After Refinement

Template Point

Reference Point
Problem 1: Discretisation effect due to the nature of voxel grids

Problem 2: Indifferentiability problem
Affinity Table: Inferred Displacement

I. Compute trilinear weights for each template point using its 8 nearest inferred displacements
II. Record the weights and indices of the 8 nearest displacements in the affinity table
III. Compute the point projection loss
IV. Distribute gradients following the IDs and weights information recorded in the affinity table in II.
Datasets
Datasets

thin plate
[Golyanik et al. 2018]

FLAME
[Li et al. 2017]

DFAUST
[Bogo et al. 2017]

cloth
[Bednařík et al. 2018]
Evaluation
Quantitative Results - Baseline and Outliers
Quantitative Results - Baseline and Outliers

<table>
<thead>
<tr>
<th>Method</th>
<th>( e )</th>
<th>( \sigma )</th>
<th>( e )</th>
<th>( \sigma )</th>
<th>( e )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{thin plate}\cite{17}</td>
<td>0.0103</td>
<td>0.0402</td>
<td>0.0083</td>
<td>0.0192</td>
<td>0.2189</td>
<td></td>
</tr>
<tr>
<td>\textit{FLAME}\cite{33}</td>
<td>0.0059</td>
<td>0.0273</td>
<td>0.0102</td>
<td>0.0083</td>
<td>1.0121</td>
<td></td>
</tr>
<tr>
<td>\textit{DFAUST}\cite{5}</td>
<td>0.0063</td>
<td>0.0588</td>
<td>0.0043</td>
<td>0.0094</td>
<td>0.0056</td>
<td></td>
</tr>
<tr>
<td>\textit{cloth}\cite{2}</td>
<td>0.0009</td>
<td>0.0454</td>
<td>0.0008</td>
<td>0.0005</td>
<td>0.0007</td>
<td></td>
</tr>
</tbody>
</table>

Baseline Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>( e )</th>
<th>( \sigma )</th>
<th>( e )</th>
<th>( \sigma )</th>
<th>( e )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{thin plate}\cite{17}</td>
<td>0.0107</td>
<td>0.0668</td>
<td>0.0218</td>
<td>0.0386</td>
<td>0.4415</td>
<td></td>
</tr>
<tr>
<td>\textit{FLAME}\cite{33}</td>
<td>0.0061</td>
<td>0.0352</td>
<td>0.0148</td>
<td>0.0067</td>
<td>1.4632</td>
<td></td>
</tr>
<tr>
<td>\textit{DFAUST}\cite{5}</td>
<td>0.0108</td>
<td>0.0334</td>
<td>0.0479</td>
<td>0.0471</td>
<td>0.4287</td>
<td></td>
</tr>
<tr>
<td>\textit{cloth}\cite{2}</td>
<td>0.0062</td>
<td>0.0281</td>
<td>0.0101</td>
<td>0.0038</td>
<td>1.3832</td>
<td></td>
</tr>
</tbody>
</table>

Outlier
Quantitative Results - Uniform Noises
Quantitative Results - Uniform Noises

**thin plate**

**FLAME**

**DFAUST**

**cloth**

<table>
<thead>
<tr>
<th>Error vs template noise ratio (%)</th>
<th>GMR</th>
<th>CPD</th>
<th>CPD (FGT)</th>
<th>NR-ICP</th>
<th>DispVoxNets (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 25 50 75 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Template  
Reference  
Template  
Reference  
Template  
Reference  
Template  
Reference
Quantitative Results - Runtime
Quantitative Results - Runtime

- With 10K points, our approach requires only a second per registration whereas others require around 2 hours - 15 seconds.
Qualitative Results
Baseline Comparison
# Baseline Comparison

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Template</th>
<th>Reference</th>
<th>DispVoxNets (Ours)</th>
<th>NR-ICP</th>
<th>CPD</th>
<th>CPD (FGT)</th>
<th>GMR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DispVoxNets (Ours)</strong></td>
<td><img src="image1" alt="Template" /></td>
<td><img src="image2" alt="Reference" /></td>
<td><img src="image3" alt="DispVoxNets" /></td>
<td><img src="image4" alt="NR-ICP" /></td>
<td><img src="image5" alt="CPD" /></td>
<td><img src="image6" alt="CPD (FGT)" /></td>
<td><img src="image7" alt="GMR" /></td>
</tr>
</tbody>
</table>
Outliers
Outliers

Inputs

<table>
<thead>
<tr>
<th>Template</th>
<th>Reference</th>
<th>DispVoxNets (Ours)</th>
<th>NR-ICP</th>
<th>CPD</th>
<th>CPD (FGT)</th>
<th>GMR</th>
</tr>
</thead>
</table>

48
Uniform Noises
Uniform Noises

Inputs

<table>
<thead>
<tr>
<th>Template</th>
<th>Reference</th>
<th>DispVoxNets (Ours)</th>
<th>NR-ICP</th>
<th>CPD</th>
<th>CPD (FGT)</th>
<th>GMR</th>
</tr>
</thead>
</table>

50
Real Face Dataset
Real Face Dataset

Datasets: [Dai et al. 2017], [Li et al. 2017]
Summary
Summary
Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.
Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

- Our approach outperforms other existing general-purpose methods in the presence of large deformations, articulated motion, noise, outliers and missing data.
Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

- Our approach outperforms other existing general-purpose methods in the presence of large deformations, articulated motion, noise, outliers and missing data.

- Runs orders of magnitude faster than previous techniques.
Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

- Our approach outperforms other existing general-purpose methods in the presence of large deformations, articulated motion, noise, outliers and missing data.

- Runs orders of magnitude faster than previous techniques.

- Limitation: topology preserving is not yet fully satisfying.
Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

- Our approach outperforms other existing general-purpose methods in the presence of large deformations, articulated motion, noise, outliers and missing data.

- Runs orders of magnitude faster than previous techniques.

- Limitation: topology preserving is not yet fully satisfying.

References

Questions?
DispVoxNets: Non-Rigid Point Set Alignment with Supervised Learning Proxies

Soshi Shimada$^{1,2}$  Vladislav Golyanik$^3$  Edgar Tretschk$^3$

Didier Stricker$^{1,2}$  Christian Theobalt$^3$

$^1$University of Kaiserslautern  $^2$DFKI  $^3$MPI for Informatics, SIC
Thank you