Feature Point Matching in Cross-Spectral Images with Cycle Consistency Learning

Ryosuke Furuta    Naoaki Noguchi    Xueiting Wang    Toshihiko Yamasaki
The University of Tokyo
Introduction

Feature Point Matching

Applications
• Image stitching
• 3D reconstruction
• Camera calibration
• etc.

Deep learning-based approaches have been proposed as local descriptors
• HardNet [Mishchuk, NeurIPS17]
• AffNet [Mishkin, ECCV18]
• D2Net [Dusmanu, CVPR19]

feature vector: $f$
Introduction

Feature Point Matching in **Cross-Spectral Images**

E.g., between RGB and Near-Infrared (NIR) images
Not easy to obtain ground-truth correspondences

We propose a **self-supervised learning method** to train feature extraction networks by utilizing the cycle consistency of the corresponding points.
Proposed Method

Training phase

- Image $I_A$
- Image $I_B$ of size $H \times W \times 3$
- Shared weights
- $CNN_\theta$
- Output tensor of size $H \times W \times c$
Proposed Method

Training phase

\[ \text{image } I_A \]

\[ \text{image } I_B \quad H \times W \times 3 \]

\[ \text{shared weights} \]

\[ \text{output tensor} \quad H \times W \times c \]

\[ \text{c-dim. feature } f_A(p) \]
Proposed Method

Training phase

image $I_A$

shared weights

image $I_B$

$H \times W \times 3$

$C_{NN_\theta}$

candidate points

$c$-dim. feature $f_A(p)$

output tensor

$H \times W \times c$

$F_{A \rightarrow B} = \arg\min_q ||f_A(p) - f_B(q)||^2$

$C_{NN_\theta}$
Proposed Method

Training phase

- Image $I_A$:
  - $H \times W \times 3$
  - Candidate points

- Image $I_B$:
  - $H \times W \times 3$

- CNN$_\theta$:
  - $c$-dim. feature $f_A(p)$
  - Cycle consistency loss
  - $F_{A\rightarrow B} = \arg\min_q ||f_A(p) - f_B(q)||^2$

- Output tensor:
  - $H \times W \times c$
Proposed Method

Training phase

image $I_A$

$c$-dim. feature $f_A(p)$

cycle consistency loss

shared weights

$F_{B\rightarrow A}$

$F_{A\rightarrow B} = \sum_q w_{pq} q$

$w_{pq} = \frac{\exp(-||f_A(p) - f_B(q)||^2)}{\sum_q \exp(-||f_A(p) - f_B(q)||^2)}$

Soft nearest neighbor [Dwibedi, CVPR19]

output tensor

$H \times W \times c$

image $I_B$

$H \times W \times 3$

$c$-dim. feature $f_A(p)$

candidate points

shared weights

candidate points
Proposed Method

Test phase

Image $I_A$

$H \times W \times 3$

Candidate points

CNN$_\theta$

$F_{A\rightarrow B} = \underset{q}{\text{argmin}} ||f_A(p) - f_B(q)||^2$

Image $I_B$

Output tensor

$H \times W \times c$

$c$-dim. feature $f_A(p)$
Experiments

Stereo matching on KITTI 2012 dataset
• 390 image pairs for training
• 194 image pairs for testing

Simulated three types of cross-spectral settings
• RGB stereo
• RGB2gray
• anaglyph

Compared methods
• Hand-crafted cost function + nearest-neighbor matching:
  ➢ Baseline
• Hand-crafted cost function + smoothness regularization (guided filter) + post-processing:
  ➢ Cost-volume filtering (CVF) [Hosni, TPAMI12]
### Experiments

<table>
<thead>
<tr>
<th>Method</th>
<th>Error rate [%] ↓</th>
<th>Mean error [pix] ↓</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RGB stereo</td>
<td>RGB2gray</td>
<td>anaglyph</td>
<td>RGB stereo</td>
<td>RGB2gray</td>
<td>anaglyph</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td>39.0</td>
<td>35.4</td>
<td>27.6</td>
<td>5.29</td>
<td>4.93</td>
<td>4.20</td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td>52.8</td>
<td>49.9</td>
<td>57.9</td>
<td>7.31</td>
<td>6.92</td>
<td>8.20</td>
</tr>
<tr>
<td><strong>CVF [Hosni, TPAMI12]</strong></td>
<td>43.9</td>
<td>43.7</td>
<td>34.0</td>
<td>5.65</td>
<td>5.59</td>
<td>4.71</td>
</tr>
</tbody>
</table>
Ours
(Learned feature + NN matching)

Baseline
(Handcrafted feature + NN matching)

CVF
(Handcrafted feature + Filtering + Post-processing)
Conclusions

• General feature point matching including cross-spectral settings

• Proposed method:
  – Self-supervised method with cycle consistency learning

• Experimental results on cross-spectral stereo matching:
  – Better accuracy than hand-crafted methods on KITTI dataset
  – Not as accurate as the compared methods but much faster on PittsStereo dataset

• Future works:
  – Deal with occlusions for better accuracy
  – Apply to other feature point matching problems such as image stitching and optical flow estimation