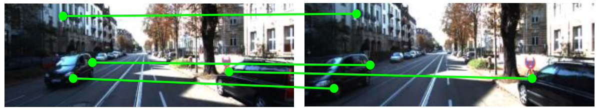
1073: Feature Point Matching in Cross-Spectral Images with Cycle Consistency Learning

Ryosuke Furuta, Naoaki Noguchi, Xueting Wang, and Toshihiko Yamasaki

The University of Tokyo

Background:

- **Feature Point Matching**



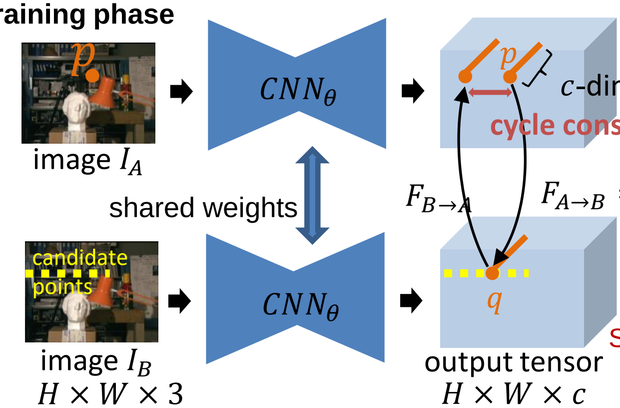
- Applications:
  - Image stitching
  - 3D reconstruction
  - etc.
- Deep learning-based approaches:
  - HardNet [Mishchuk, NeurIPS17]
  - AffNet [Mishkin, ECCV18]
  - D2Net [Dusmanu, CVPR19]

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

- E.g., between RGB and Near-Infrared (NIR) images
  Not easy to obtain ground truth correspondences to train CNNs
- We propose a self-supervised learning method
  by utilizing the cycle consistency of the corresponding points

Proposed Method:

- **Training phase**

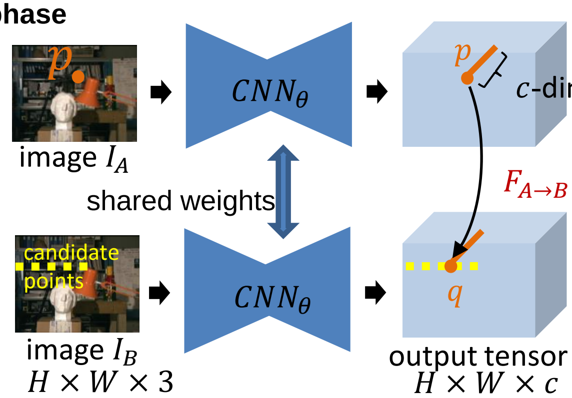


$$F_{A\to B} = \sum_q w_{pq} q$$

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

Soft nearest neighbor [Dwibedi, CVPR19]

- **Test phase**



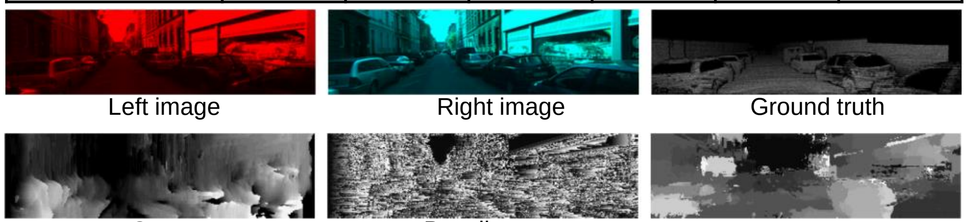
$$F_{A\to B} = \underset{q}{\operatorname{argmin}} ||f_A(p) - f_B(q)||^2$$

Simple nearest neighbor

Experimental Results:

- Stereo Matching on KITTI 2012 dataset (Three simulated cross-spectral settings)

	Error rate [%] ↓			Mean error [pix] ↓		
	RGB stereo	RGB2gray	anaglyph	RGB stereo	RGB2gray	anaglyph
Ours	**39.0**	**35.4**	**27.6**	**5.29**	**4.93**	**4.20**
Baseline	52.8	49.9	57.9	7.31	6.92	8.20
CVF [Hosni, TPAMI12]	43.9	43.7	34.0	5.65	5.59	4.71



Left image | Right image | Ground truth

Ours (Learned feature + NN matching) | Baseline (Handcrafted feature + NN matching) | CVF (Handcrafted feature + Filtering + Post-processing)

- Stereo Matching on PittsStereo dataset (RGB2NIR)

	Mean error [pix]	Inference time [s]	Note
Ours	8.20	0.00619	Python + GPU
Ours + guided filter	6.27	0.00776	Python + GPU
Baseline	11.2	0.0111	Python + CPU
CVF [Hosni, TPAMI12]	2.29	14.4	Python + CPU
CMA [Chiu, BMVC11]	4.00	227	-
ANCC [Heo, TPAMI11]	2.63	119	-
DASC [Kim, CVPR15]	1.28	44.7	-
DMC [Zhi, CVPR18]	0.80	0.02	Tailored for cross-spectral stereo
F-cycle GAN [Liang, AAAI19]	0.84	0.04	Tailored for cross-spectral stereo

- Better accuracy than hand-crafted methods on KITTI dataset
- Not as accurate as the state-of-the-art methods but much faster on PittsStereo dataset