GIFT: Learning Transformation-Invariant Dense Visual Descriptors via Group CNNs

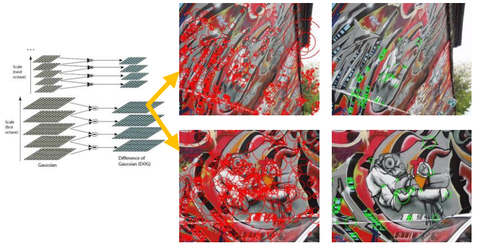
Yuan Liu, Zehong Shen, Zhixuan Lin, Sida Peng, Hujun Bao, Xiaowei Zhou
State Key Lab of CAD&CG, ZJU-Sensetime Joint Lab of 3D Vision, Zhejiang University

NeurIPS 2019
Vancouver
Dec. 8-14, 2019

1. Background

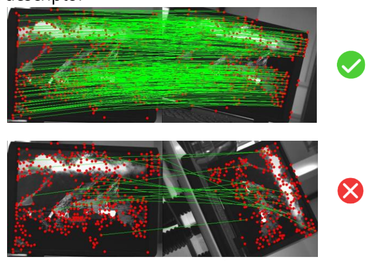
Problem: How to make image feature descriptors invariant to scaling and rotation?

1) Traditional DoG detector



Only few covariant regions can be reliably detected

2) CNN-based descriptor

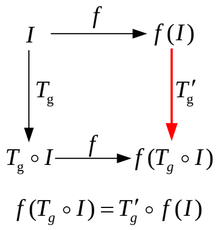


Convolutions are not invariant to geometric transformation naturally.

Can we learn an invariant descriptor with theoretical guarantees?

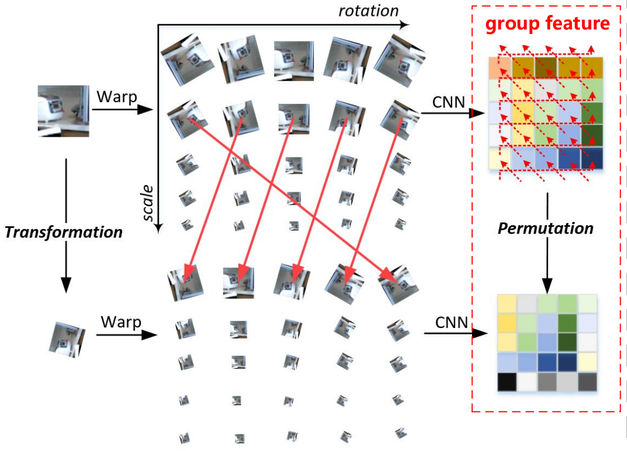
2. Group Feature & Equivariance Property

1) Equivariance definition



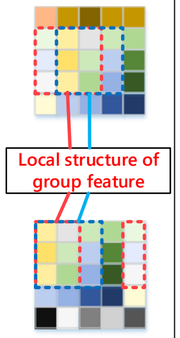
I input data.
f feature extractor
T_g transformation of input
T'_g transformation of feature

2) Equivariance property of group feature extraction



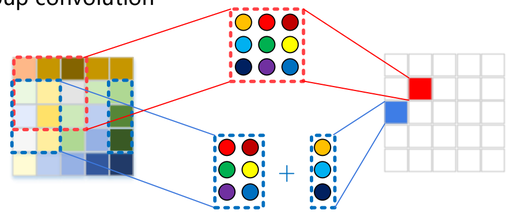
Transformation results in permutation.

3) Invariance of local structure



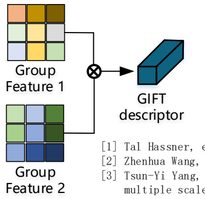
3. Group Convolution & Bilinear Pooling

1) Group convolution



i) encodes local structures of group feature ii) preserves equivariance property

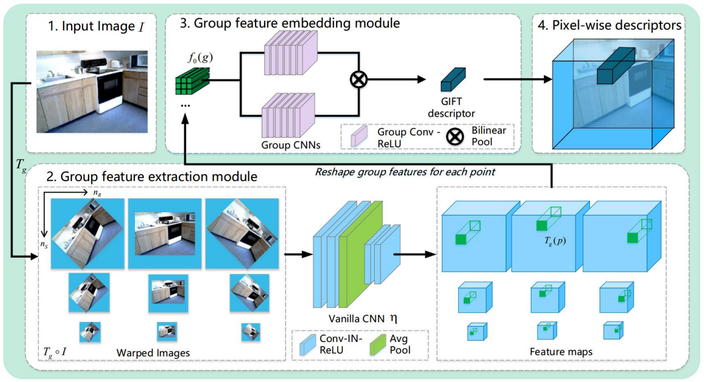
2) Bilinear Pooling



Advantages over Max/Average Pooling:
i) Second order statistics are more informative.
ii) Generalized form of previous descriptors [1, 2, 3].

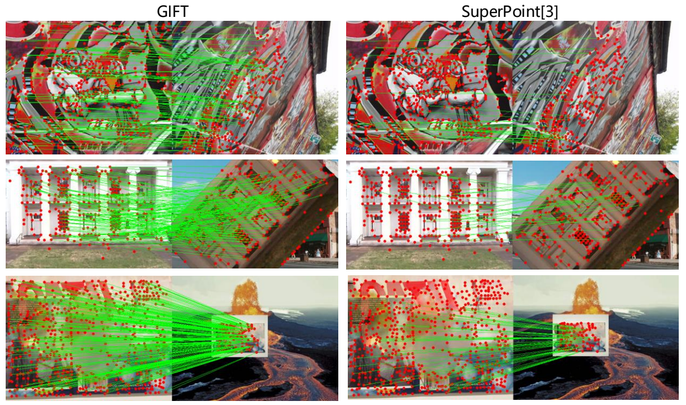
[1] Tal Hassner, et al. On sifts and their scales. In CVPR, 2012.
[2] Zhenhua Wang, et al. Affine subspace representation for feature description. In ECCV, 2014.
[3] Tsun-Yi Yang, et al. Accumulated stability voting: A robust descriptor fromdescriptors of multiple scales. In CVPR, 2016.

4. Pipeline

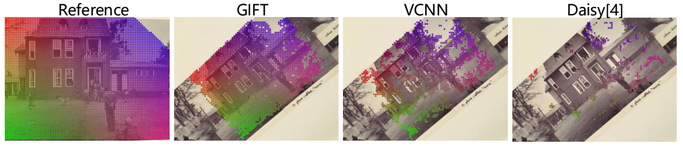


5. Results

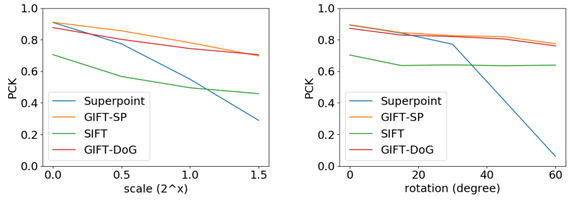
1) Sparse correspondence estimation



2) Dense correspondence estimation



3) Robustness of GIFT



[4] Daniel DeTone, et al. SuperPoint: Self-supervised interest pointdetection and description. In CVPR Workshops, 2018.
[5] James Philbin, et al. Descriptor learning for efficient retrieval. In ECCV, 2010.