1. Background

Problem: How to make image feature descriptors invariant to scaling and rotation?
1) Traditional DoG detector
2) CNN-based descriptor

Only few covariant regions can be reliably detected
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
2) Equivariance property of group feature extraction
3) Invariance of local structure

\[ T \] is a transformation of input
\[ T^{\prime} \] is a transformation of feature
\[ f(T) \] is the feature extractor
\[ f \] is the feature function
\[ T_f \] is the transformation of input
\[ T_g \] is the transformation of group feature

Transformation results in permutation.

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].

4. Pipeline

1) Input image
2) Group feature extraction module
3) Group feature embedding module
4) Pixel-wise description
5) Sparse correspondence estimation

5. Results

1) Sparse correspondence estimation
   - GIFT
   - SuperPoint

2) Dense correspondence estimation
   - Reference
   - GIFT
   - VCNN
   - Daisy

3) Robustness of GIFT

References: