An Improved Learning Framework for Covariant Local Feature Detection

Nehal Doiphode[1] | Rahul Mitra[2] | Shuaib Ahmed[3] | Arjun Jain[2]
University of Pennsylvania[1], IIT Bombay[2], Mercedes-Benz Research and Development India Pvt. Ltd.[3]

ACCV 2018
2 – 6 December 2018 Perth Western Australia

Motivation & Challenges

- Learn a CNN feature extractor which is covariant to geometric variations as well as discriminative from its neighbourhood.
- Existing methods [DDet, Cov Det] either uses pre-computed features as supervision or fails to be discriminative for subsequently used in matching. Using pre-computed features limits the performance of the learned extractor.

Contributions

- Introduce a novel framework, extending existing framework by incorporating additional geometric constraints increasing stability while maintaining discriminativeness.
- Model trained with our proposed framework achieves state-of-the-art performance in *repeatability* score on publicly available datasets.

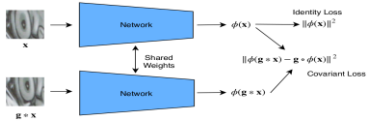
Covariant Feature Detection

- Covariant Constraint,

$$\phi(\mathbf{g} * \mathbf{x}) = \mathbf{g} \circ \phi(\mathbf{x}); \ \forall \mathbf{x} \in \mathcal{X}, \forall \mathbf{g} \in G$$

here $\mathbf{x}$ represents a patch from set of all patches $\mathcal{X}$. ϕ is the CNN feature extractor, selecting the position of feature within the patch. $\mathbf{g}$ parameterizes an affine transform. $\circ$ and $*$ are transformation composition and image warping functions respectively.

The training framework used in [Cov Det] is shown below,



The patches $\mathbf{x}$ are *standard* patches containing a *good* feature at its center. The covariant loss ensures detection of the same feature point in both the patches. Without the identity loss (marked in red), the learned is unstable and chooses different features at different training runs.

Proposed Framework

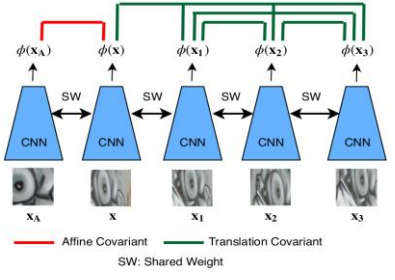
Our proposed framework utilizes stricter translational covariance constraints by considering *triplets* of patches. Furthermore, as an extension of [Cov Det] we consider patch pairs connected by generalized Affine transforms.

$$\ell_{cov-tran} = \|a\phi(\mathbf{x}_1) - \beta\phi(\mathbf{x}_2) - (\alpha - \beta)\phi(\mathbf{x}) - \mathbf{t}_{12}\|^2 + \|a\phi(\mathbf{x}_2) - \beta\phi(\mathbf{x}_3) - (\alpha - \beta)\phi(\mathbf{x}) - \mathbf{t}_{23}\|^2 + \|a\phi(\mathbf{x}_3) - \beta\phi(\mathbf{x}_1) - (\alpha - \beta)\phi(\mathbf{x}) - \mathbf{t}_{31}\|^2$$

$$\ell_{cov-aff} = \|\phi(\mathbf{x}_A) - A * \phi(\mathbf{x})\|^2$$

Here $\mathbf{x}$ is the reference patch, $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are translations of $\mathbf{x}$ while $\mathbf{x}_A$ is an affine warp. $\mathbf{t}_{12}$ is the relative translation between patches , $\mathbf{x}_1$and $\mathbf{x}_2$. α and β are taken as 2 and 1 respectively.

Corresponding network architecture,



Evaluation Metrics

- **Repeatability** - Measures the consistency in feature detection across images sharing the same contents. We use overlap ratio of detected feature regions in two images for this purpose.

Results

Ablation on performance on different variants of our proposed model along with std. dev based on 5 training runs is shown below,

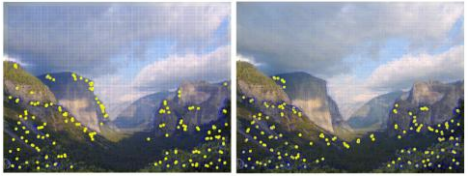
Method	Vgg-Aff	EF	Webcam
COVDET++	58.14 ± 0.12	36.41 ± 0.05	51.39 ± 0.05
Trip	54.65 ± 1.34	32.10 ± 0.30	48.45 ± 0.30
Cov + Aff	59.72 ± 1.15	34.14 ± 0.20	50.23 ± 0.30
Trip + Aff (proposed)	$\mathbf{63.12} \pm 0.20$	$\mathbf{36.50} \pm 0.05$	$\mathbf{51.40} \pm 0.05$

Our proposed approach records state-of-the-art performance while selecting stable features indicated by low standard deviation.

Comparison of repeatability scores on VGG-Affine, EF and Webcam datasets.

Method	Vgg-Affine		EF		Webcam	
	200	1000	200	1000	200	1000
TILDEP24 [18]	57.57	64.35	32.3	45.37	45.1	61.7
DDET [6]	50.9	65.41	24.54	43.31	34.24	50.67
COVDET [21]	59.14	68.15	35.1	46.10	50.65	**67.12**
Proposed	**63.12**	**69.79**	**36.5**	**46.49**	**51.4**	65.0259

Qualitative comparison of our proposed model against [Cov Det] on *Yosemite* scene of EF dataset.



(Proposed) (Cov Det)

References

[DDet] – Lenc et al. ECCV Workshop 2016.
[Cov Det] - Zhang et al. CVPR 2017.