WxBS: Wide Baseline Stereo Generalizations

Dmytro Mishkin[1], Jiri Matas[1], Michal Perdoch[2], Karel Lenc[3]

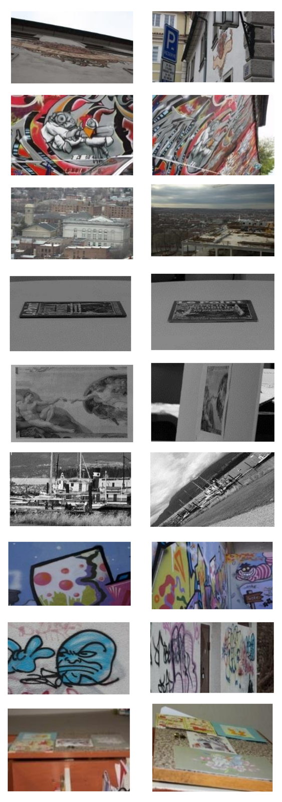
[1]Center for Machine Perception, Czech Technical University in Prague;
[2]Computer Vision Laboratory, ETH Zurich, Switzerland;
[3]Department of Engineering Science, University of Oxford, UK

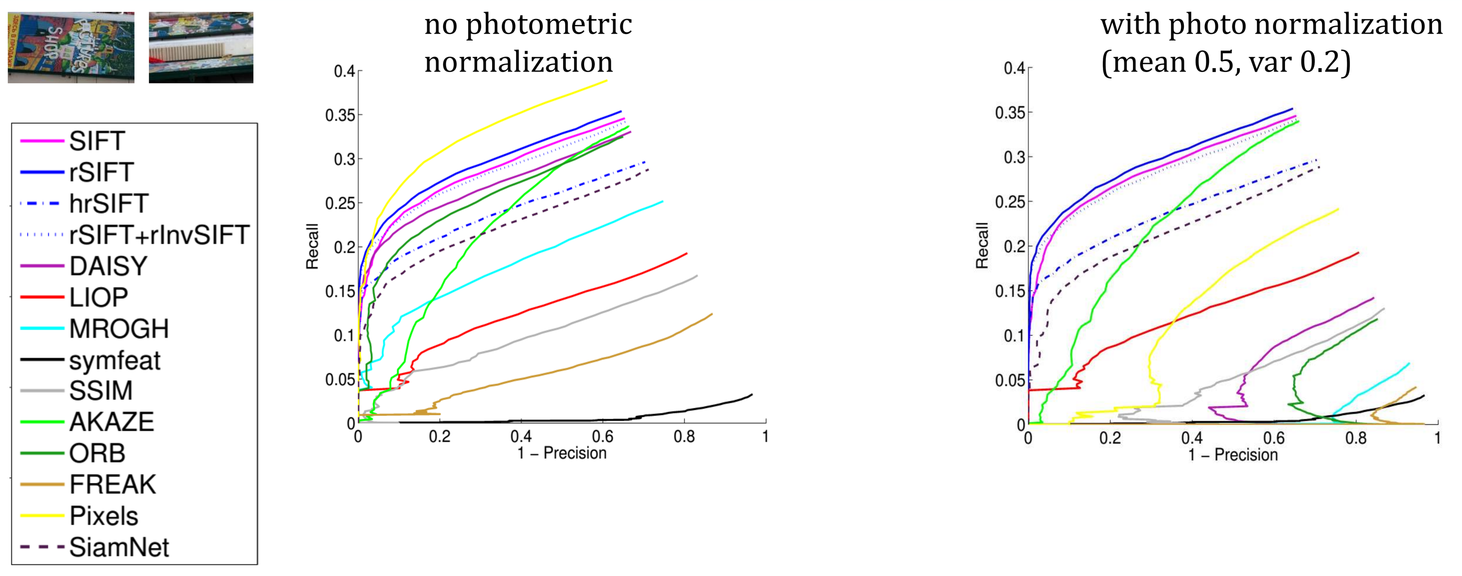
Download Poster

Abstract

- **Generalization of the wide baseline two-view matching problem - WxBS**
  x stands for different subsets of "wide baselines" in acquisition conditions.
- **Novel dataset** of ground-truthed image pairs which include multiple "wide baselines"
- **We show that state-of-the art matchers fail on almost all image pairs**.
- **WxBS-M - a novel matching algorithm for the WxBS problem is introduced.**
  We show experimentally that the WxBS-M matcher dominates the state-of-the-art methods both on the new and existing datasets

WGBS – Wide Geometry Baseline Stereo



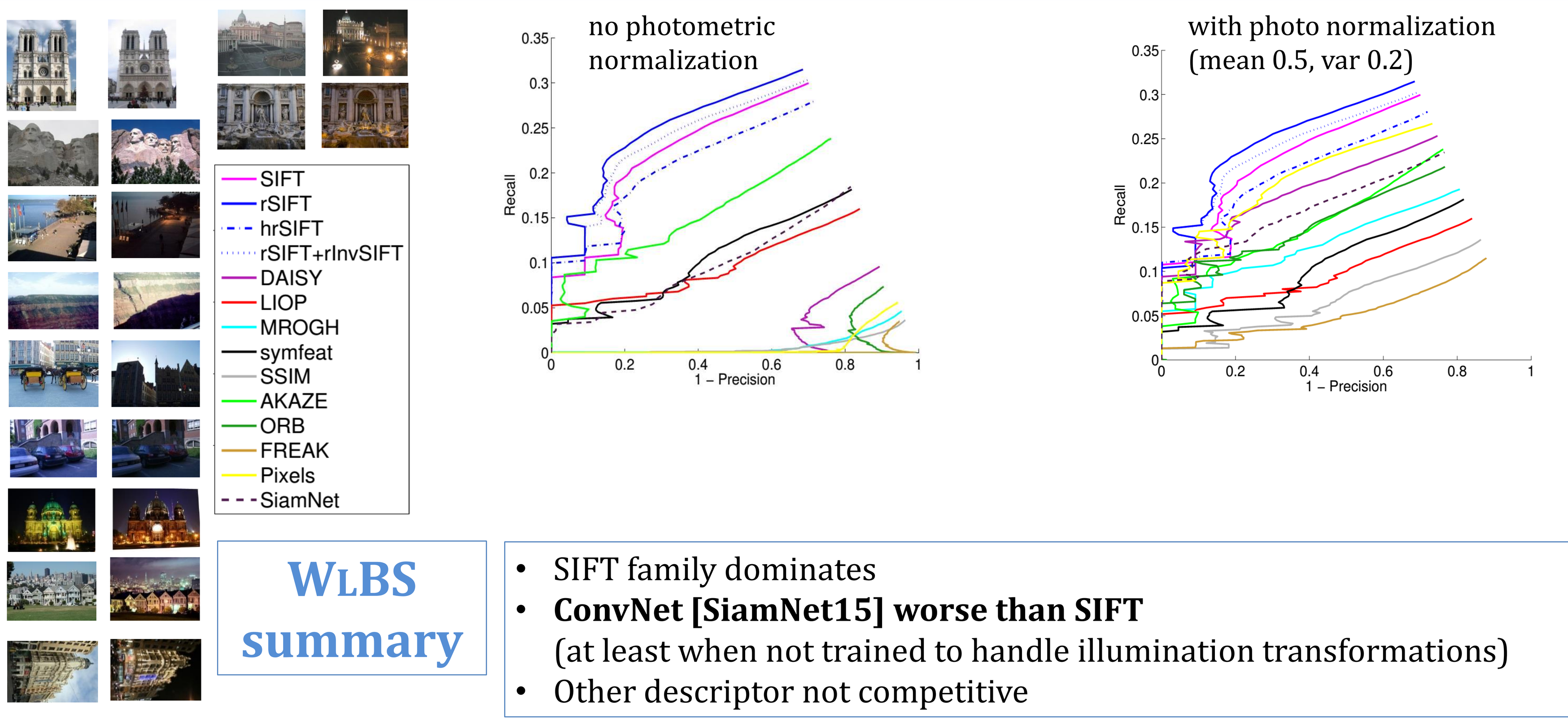


WGBS summary

- SIFT family dominates
- Photo-L2 normalized pixel intensities is a strong descriptor
- **ConvNet [SiamNet15] worse than SIFT**
  (at least when not trained to handle large transformations)
- Other descriptor not competitive

*Images from Extreme View (EVD) and Oxford-Affine(OxAff) Datasets

WLBS – Wide iLlumination Baseline Stereo

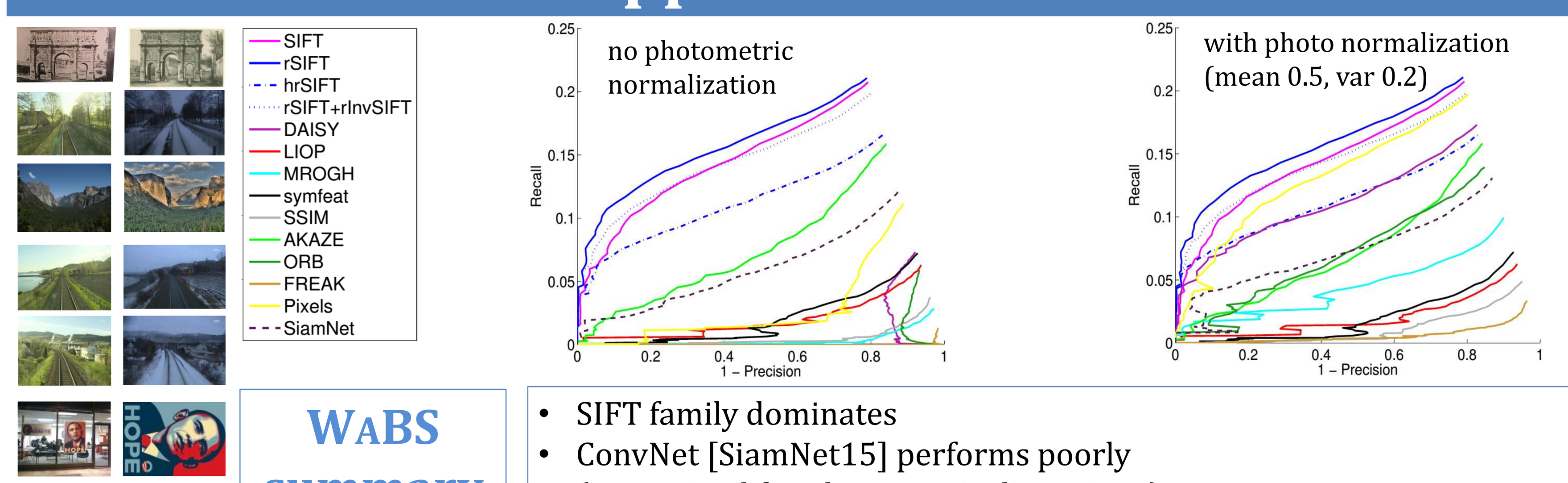


WLBS summary

- SIFT family dominates
- **ConvNet [SiamNet15] worse than SIFT**
  (at least when not trained to handle illumination transformations)
- Other descriptor not competitive

*Images from SymBench, GDBootstrap, EgdeFoci (EF) datasets

WABS – Wide Appearance Baseline Stereo

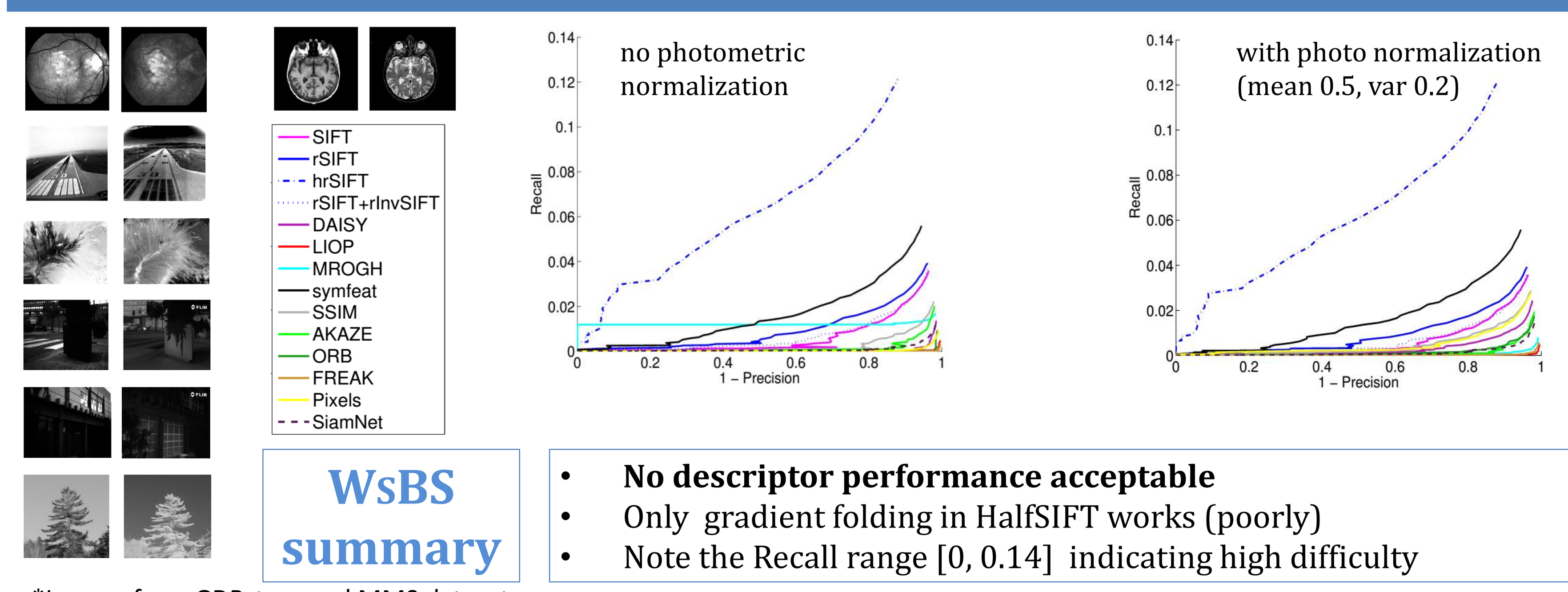


WABS summary

- SIFT family dominates
- ConvNet [SiamNet15] performs poorly
  (not trained for photometric distortions)
- Other descriptor not competitive

*Images from SymBench, VPRiCE 2015, EgdeFoci (EF) datasets

WSBS – Wide Sensor Baseline Stereo

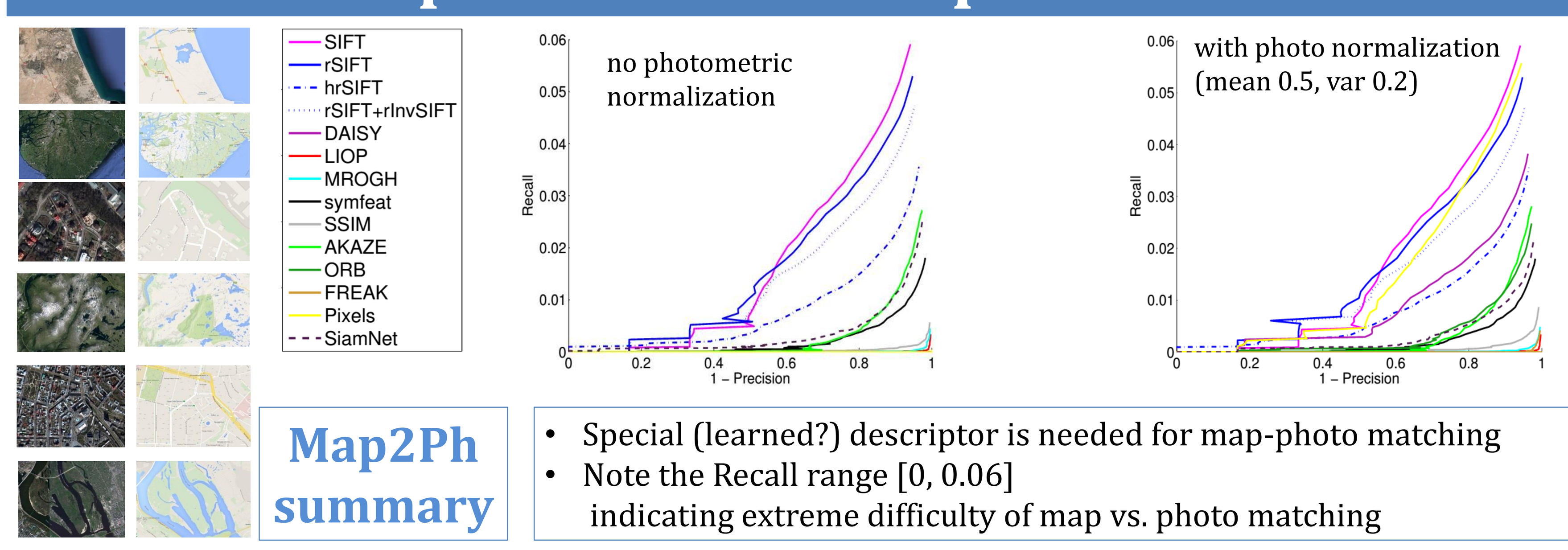


WSBS summary

- **No descriptor performance acceptable**
- Only gradient folding in HalfSIFT works (poorly)
- Note the Recall range [0, 0.14] indicating high difficulty

*Images from GDBstrap and MMS datasets

Map2Photo: WABS special case



Map2Ph summary

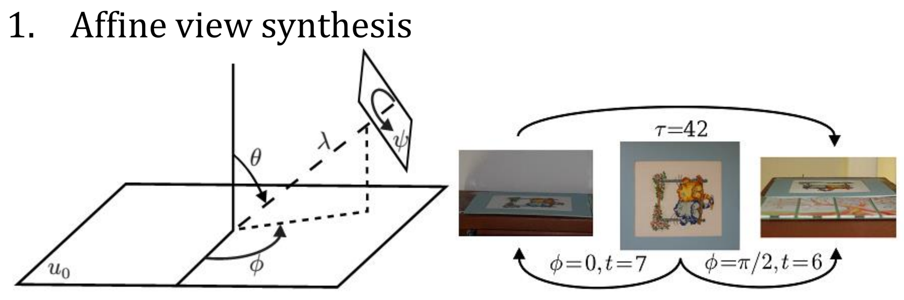
- Special (learned?) descriptor is needed for map-photo matching
- Note the Recall range [0, 0.06]
  indicating extreme difficulty of map vs. photo matching

*map2ph dataset with this paper

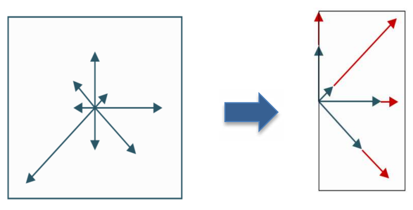
WxBS-Matcher

Input: I_1, I_2 - two images,
$\boldsymbol{\Theta_m}$ - minimum required number of matches,
$\boldsymbol{S_{max}}$ - maximum number of iterations
Output: Fundamental or homography matrix **F** or **H**;
a list of corresponding local features
while $(N_{matches} < \Theta_m)$ **and** $(Iter < S_{max})$ **do**
for I_1 and I_2 **separately do**
1 Generate synthetic views according to the scale-tilt-rotation-detector setup for Iter
2 Detect local features using adaptive thresholding
3 Extract rotation invariant descriptors with:
3a RootSIFT and **3b HalfRootSIFT**
4 Reproject local features to I_1, I_2
end for
5 Generate tentative correspondences based on 1st geom. Inconsistent rule for RootSIFT and HalfRootSIFT separately using kD-tree
6 Filter duplicates
7 Geometric verification of all TC with modified DEGENSAC estimating F or H
8 Check geometric consistency of the local affine features with est. F or H
end while

1. Affine view synthesis



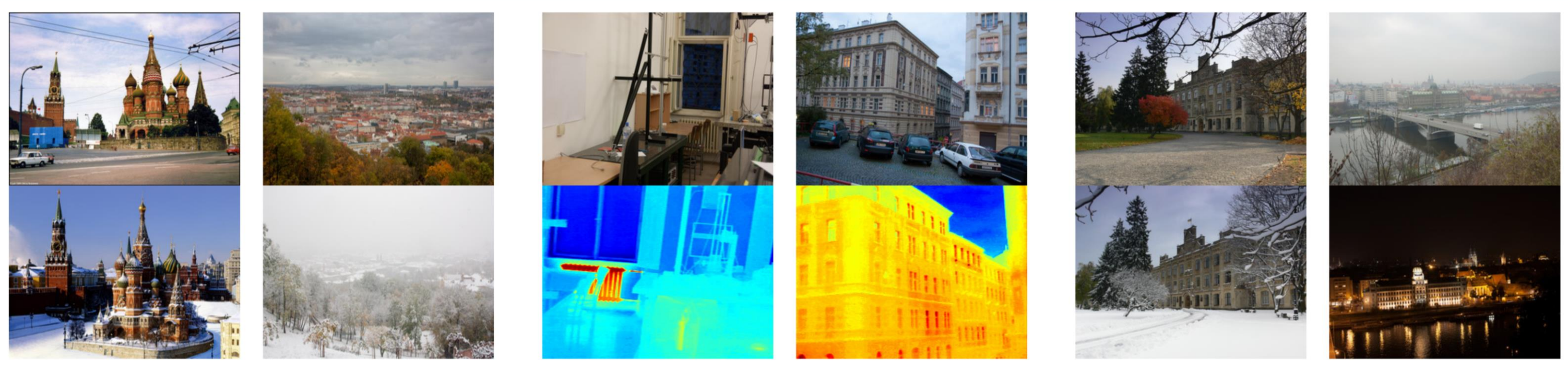
2. Adaptive thresholding:
   if #HesAffs < θ_{HesAff}, lower the detection threshold
3. HalfRootSIFT:
   SIFT bin HalfSIFT bin



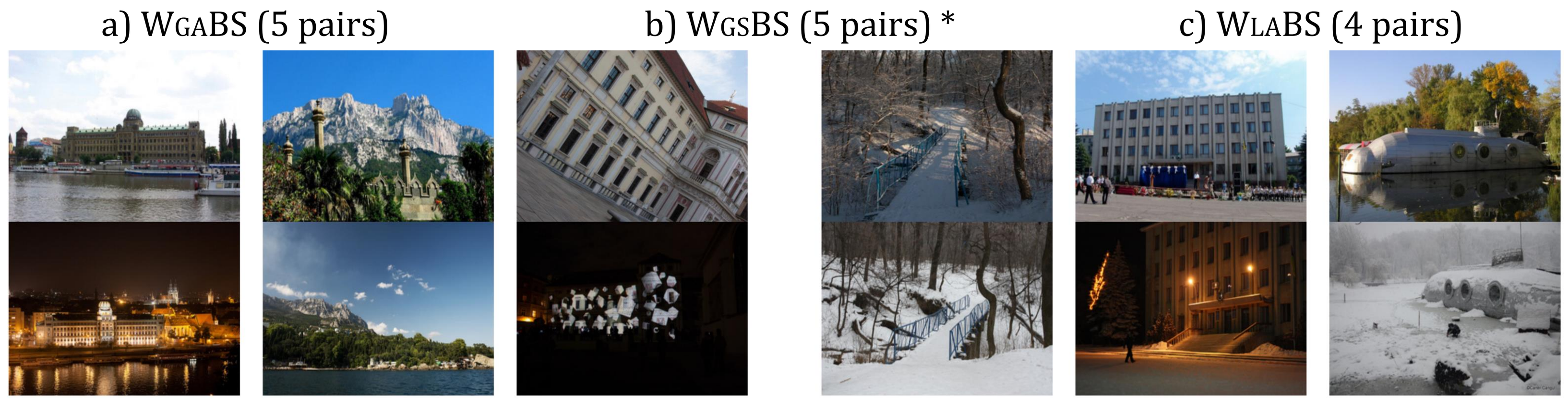
5. 1st geom. Inconsistent rule:
   use for second nearest distance ratio only patches, which are inconsistent with closest one (yellow, not red)

6. Filter duplicates: discard redetections (red patches)

WxBS: Multiple Wide Baselines



a) WGABS (5 pairs) b) WGSBS (5 pairs) * c) WLABS (4 pairs)



d) WGLBS (9 pairs) e) WGALBS (8 pairs)

*WGSBS contains image pairs of thermal camera vs visible

Detector and matcher comparison

Alg.	EF # 33	EF time [s]	EVD # 15	EVD time [s]	MMS # 100	MMS time [s]	WGABS # 5	WGABS time [s]	WGALBS # 8	WGALBS time [s]	WGLBS # 9	WGLBS time [s]	WGSBS # 5	WGSBS time [s]	WLABS # 4	WLABS time [s]	Past # 172	Past time [s]	OxAff # 40	OxAff time [s]	SymB # 46	SymB time [s]	GDB # 22	GDB time [s]
Threshold adaptation																								
MSER	16	1.4	3	1.4	1	0.3	0	2.0	0	1.3	0	1.3	0	0.8	1	1.2	8	1.3	40	3.5	23	2.4	9	2.4
AdMSER	25	3.4	8	4.0	6	1.0	0	4.0	0	3.2	0	3.3	0	1.4	1	2.6	11	2.9	40	5.7	26	4.6	13	6.9
DoG	29	2.3	0	2.8	10	0.8	0	2.7	0	2.3	0	2.1	0	1.0	1	2.4	13	2.0	38	4.8	29	2.7	12	4.7
iiDoG	29	3.1	0	3.0	11	1.2	0	3.2	0	2.9	0	2.8	0	1.2	1	2.5	13	2.2	38	8.0	29	2.9	12	6.1
AdDoG	29	2.6	0	3.4	11	1.2	0	3.3	0	3.0	0	3.0	0	1.5	1	2.7	13	2.7	38	4.1	30	3.0	12	4.8
HesAf	32	4.6	1	5.2	15	1.2	0	5.5	0	3.8	0	4.2	0	2.0	1	3.6	24	4.0	40	11	35	5.8	17	9.1
AdHesAf	33	5.7	2	7.6	35	2.9	0	7.2	1	6.5	0	6.0	0	3.2	1	4.9	25	5.4	40	10	35	7.2	18	13
Other detectors																								
WαSH	0	1.8	0	5.4	0	0.6	0	2.8	0	2.5	0	1.4	0	1.8	0	1.2	0	1.9	24	4.1	3	2.8	3	6.9
ORB	3	4.1	0	3.6	1	0.8	0	2.8	0	2.7	0	3.6	0	1.6	0	2.8	1	2.3	28	8.7	5	3.0	3	6.1
SURF	27	2.3	0	2.4	7	1.0	0	2.5	0	1.9	0	2.1	0	0.9	1	1.4	10	1.9	38	5.8	31	2.9	15	4.0
AKAZE	28	4.3	0	3.6	10	0.8	1	4.7	0	3.4	0	4.0	0	1.3	1	2.7	25	3.6	38	13	35	5.6	17	6.4
FOCI	29	12	0	39	14	11	1	32	0	29	0	29	0	20	1	29	21	13	38	35	35	27	17	45
SFOP	25	11	0	16	12	4.7	0	12	0	10	0	10	0	9.2	0	7.5	11	12	36	15	24	11	8	17
WADE	16	14	0	20	0	3.4	0	58	0	11	0	14	0	7.9	1	8.3	20	23	34	60	34	46	13	77
TILDE-StL-ns	22	3.7	0	6.6	20	2.8	0	5.0	0	4.5	0	5.0	0	4.6	1	4.2	-	-	29	5.5	28	4.6	8	8.4
TILDE-StL	27	18.	0	32.	31	13.	0	22.	0	20.	0	21.	0	17.	1	21.	-	-	35	24.	29	22.	9	35.
TILDE-Cha	26	16	0	30	5	11	0	21	0	21	0	20	0	16	1	21	13	19	38	25	30	22	8	31
TILDE-Cou	28	18	0	30	42	13	0	23	0	22	0	24	0	17	1	21	18	20	37	26	31	22	8	37
TILDE-Fra	23	18	1	32	33	13	0	22	0	21	0	23	0	17	1	23	14	20	37	25	31	22	9	34
TILDE-Mex	24	17	0	29	5	12	0	23	0	23	0	23	0	18	1	21	13	20	36	24	26	22	8	33
TILDE-Pan	29	18	0	30	42	13	0	26	1	24	0	23	0	18	1	23	15	20	36	26	32	21	11	36
State-of-art matchers																								
ASIFT	23	27	5	12	18	3.2	0	52	0	32	0	35	0	12	1	30	62	32	40	102	27	14	15	41
MODS	33	4.8	15	11	27	11	2	41	2	31	1	46	0	17	1	26	94	27	40	3.4	42	18	18	11
DBstrap	31	26	0	18	79	9.3	0	11	0	13	0	13	0	4.7	0	15	16	28	36	24	38	21	16	17
Proposed matcher																								
WXBS-M	33	4.7	15	14	82	12	3	40	3	63	3	61	0	26	3	28	107	42	40	5.1	43	18	22	12

TILDE detector results are post-CR deadline

Best results among single detectors (AdHesAf) and view-synth based matchers (WxBS-M)

Take away

- **SIFT family is still the best local descriptor,**
  outperforms novel CNN [SiamNet2015] approaches.
- (adaptive) Hessian-Affine is the best detector with broad applicability
- Affine view synthesis greatly helps for non-geometrical problems.
- Datasets and WxBS-Matcher available http://cmp.felk.cvut.cz/wbs/
- We need more diverse datasets for learning local descriptors than Yosemite and Liberty

Download WxBS-Matcher

References

- [SiamNet15] S. Zagoruyko, N. Komodakis. Learning to Compare Image Patches via Convolutional Neural Networks. In CVPR 2015
- [HalfSIFT10] J. Chen, J. Tian, N. Lee, J. Zheng, R. Smith, and A. Laine. A partial intensity invariant feature descriptor for multimodal retinal image registration. Biomedical Engineering, IEEE Transactions on, 2010.
- [MODS15] D. Mishkin and J. Matas and M. Perdoch. MODS: Fast and Robust Method for Two-View Matching. Accepted to CVIU, 2015.
- [DEGENSAC05] O.Chum, T. Werner, J. Matas. Two-view Geometry Estimation Unaffected by a Dominant Plane. In CVPR 2005