Phototourism Challenge

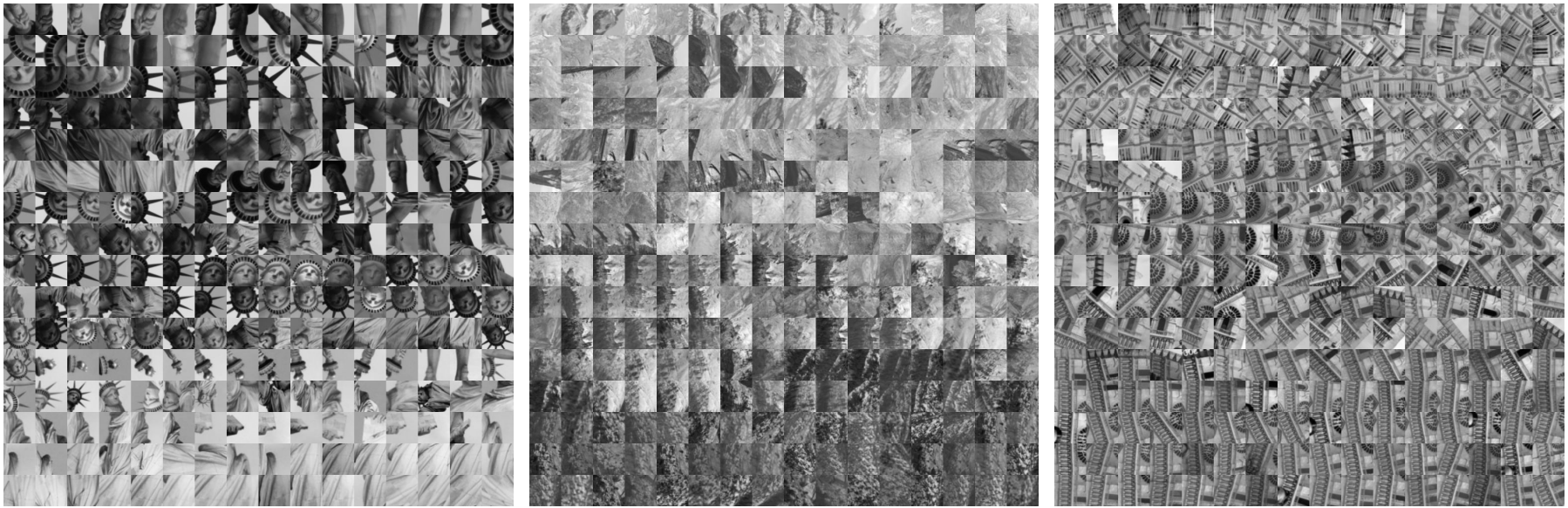
Eduard Trulls (Google)
Kwang Moo Yi (U. Victoria)
Sri Raghu Malireddi (U. Victoria)
Yuhe Jin (U. Victoria)

How good is

<insert-your-favorite-method-here>

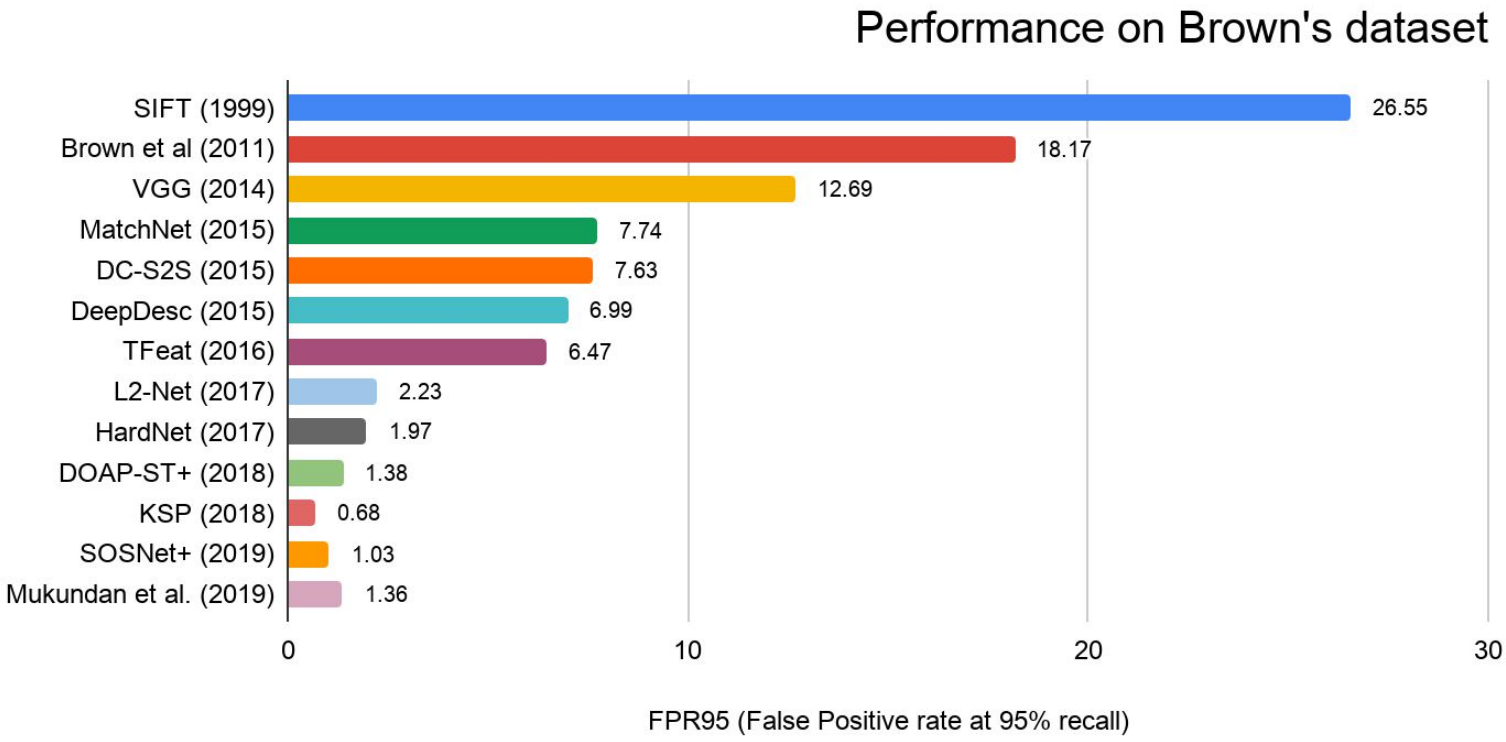
in practice?

Current benchmarks are **saturated**

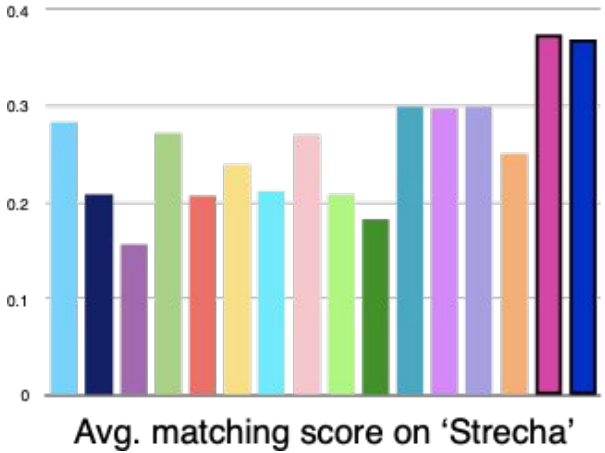


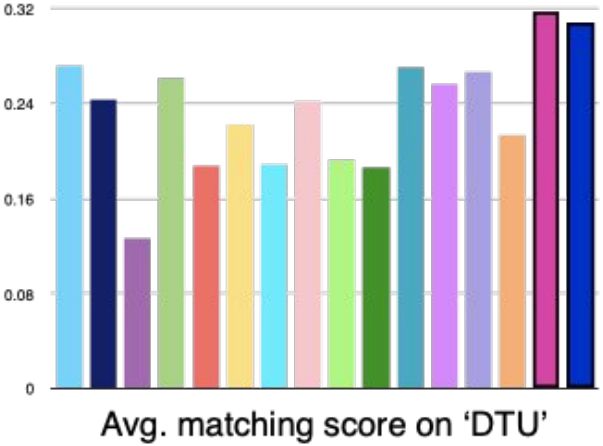
Discriminative Learning of Local Image Descriptors. Brown et al., PAMI'10

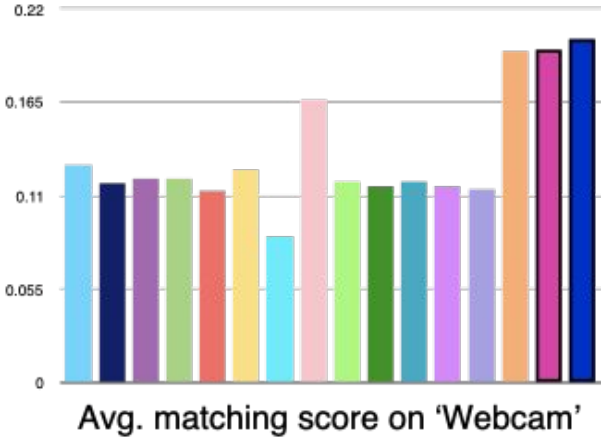
Current benchmarks are **saturated**



Current benchmarks are **not representative**



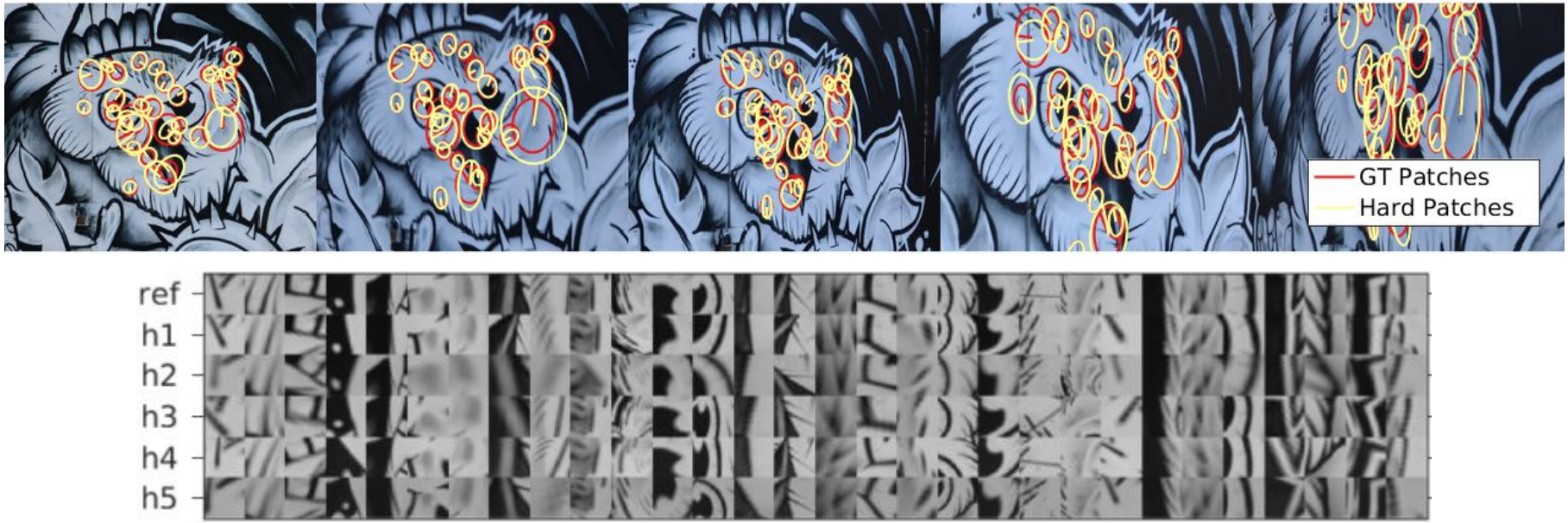




LIFT: Learned Invariant Feature Transform. Yi et al., ECCV'16

Towards proper benchmarking -- H(omography)Patches

Task: patch matching under affine transformation or illumination changes

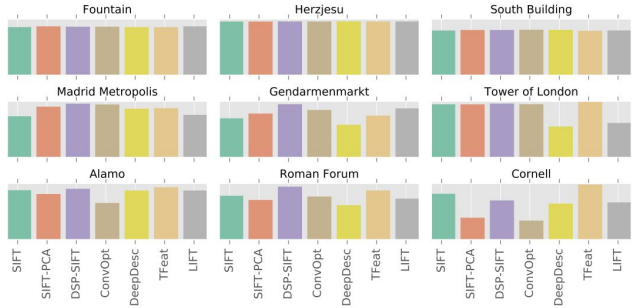


HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. V. Balntas et al., CVPR'17
Source: github.com/hpatches/hpatches-dataset

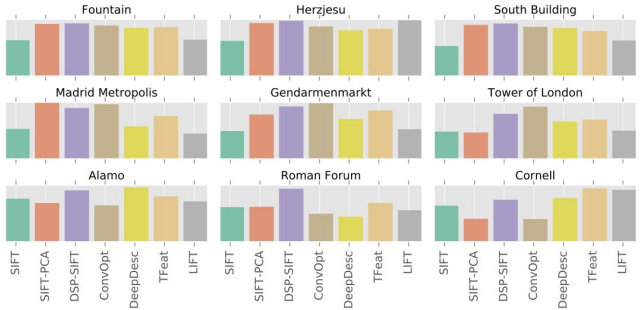
Towards proper benchmarking -- SfM (COLMAP)

Task: 3D reconstruction with local features

Number of registered images

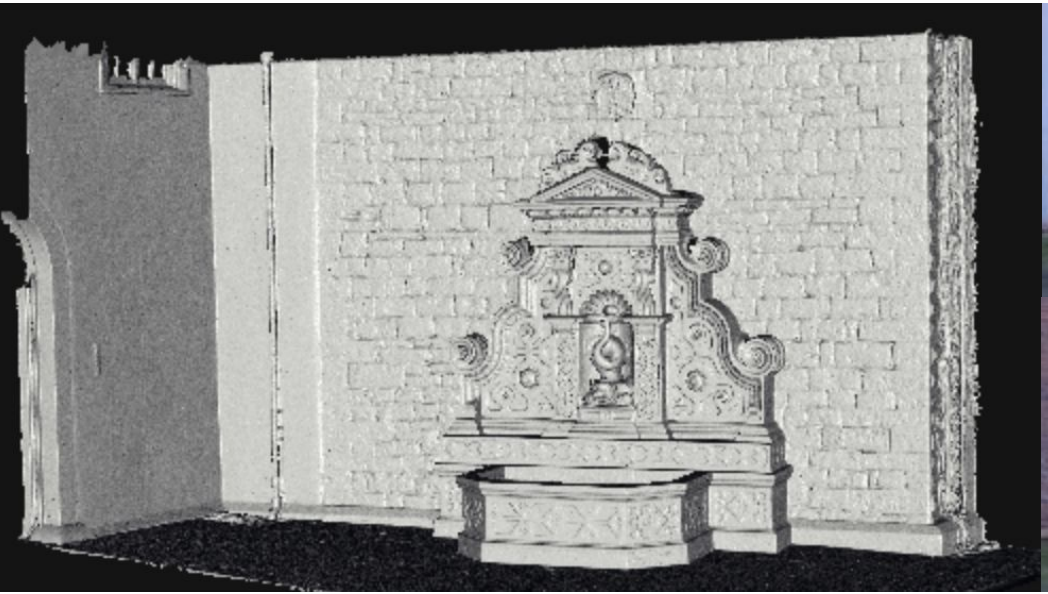


Number of registered 3D points



Comparative Evaluation of Hand-Crafted and Learned Local Features.
Schönberger et al., CVPR'17. Source: github.com/ahojnnes/local-feature-evaluation

Depth comes at a cost



On benchmarking camera calibration and multi-view stereo for high resolution imagery. Strecha et al., CVPR'08.

How good is

<insert-your-favorite-method-here>

in practice?

How good is

<insert-your-favorite-method-here>

in **practice**?

Towards practical evaluation

- Variation + Volume

Towards practical evaluation

- Variation + Volume
  - Phototourism data: viewpoint, sensors, illumination, motion blur, occlusions, etc
  - Large-scale: ~30k images
  - Images, poses & depth: suitable for multiple tasks

Towards practical evaluation

- Variation + Volume
  - Phototourism data: viewpoint, sensors, illumination, motion blur, occlusions, etc
  - Large-scale: ~30k images
  - Images, poses & depth: suitable for multiple tasks

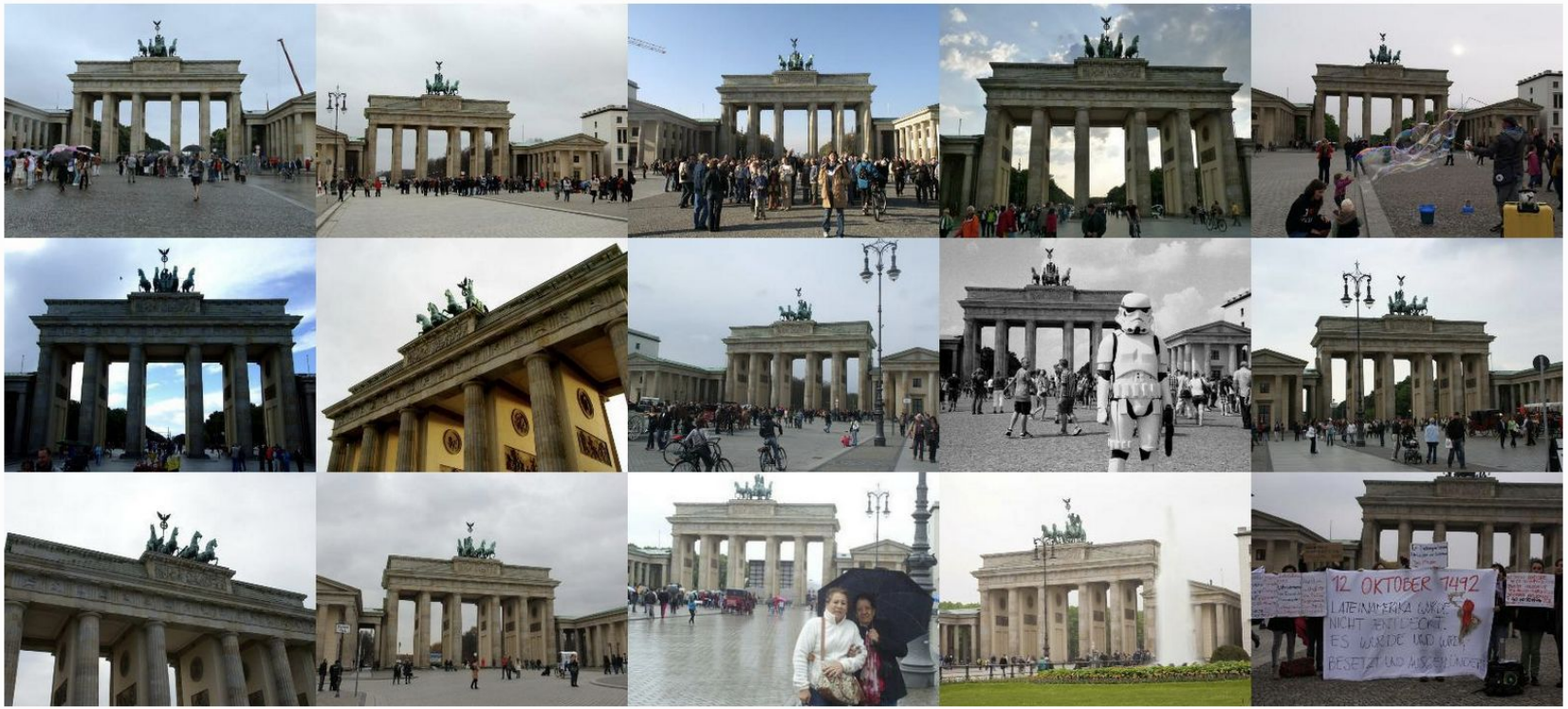
- Image-level evaluation
  - Matching scores

Towards practical evaluation

- Variation + Volume
  - Phototourism data: viewpoint, sensors, illumination, motion blur, occlusions, etc
  - Large-scale: ~30k images
  - Images, poses & depth: suitable for multiple tasks

- Image-level evaluation
  - Matching scores
  - Stereo: Camera pose accuracy
  - SfM: Camera pose accuracy + Metrics by Schönberger et al. CVPR'17

The phototourism challenge: Data



Examples from brandenburg_gate

The phototourism challenge: Data

The phototourism challenge: Data

Training sequences	Num. images	Num. 3D points
brandenburg_gate	1363	100040
buckingham_palace	1676	234052
colosseum_exterior	2063	259807
grand_place_brussels	1083	229788
hagia_sophia_interior	888	235541
notre_dame_front_facade	3765	488895
palace_of_westminster	983	115868
pantheon_exterior	1401	166923
prague_old_town_square	2316	558600
sacre_coeur	1179	140659
st_peters_square	2504	232329
taj_mahal	1312	94121
temple_nara_japan	904	92131
trevi_fountain	3191	580673
westminster_abbey	1061	198222
Total	25.6k	3.7M

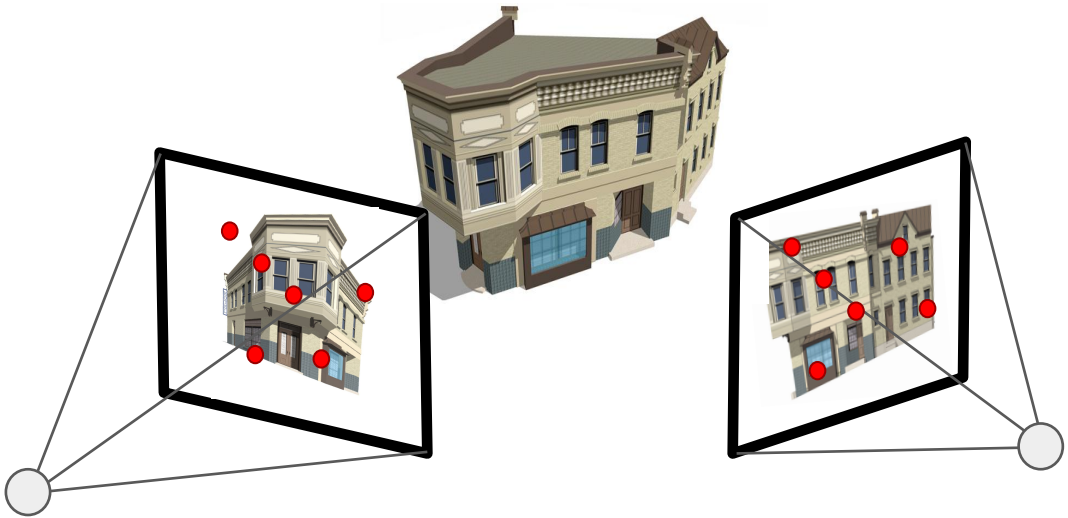
- 25k images in total for training.
- “Quasi” ground truth data is generated by performing SfM with COLMAP with all images.
  - Assumption: Images registered in COLMAP are accurate given enough images.
- Valid pairs are generated via simple visibility check.

The phototourism challenge: Data

Test sequences	Num. images	Num. 3D points
british_museum	660	73569
florence_cathedral_side	108	44143
lincoln_memorial_statue	850	58661
london_bridge	629	72235
milan_cathedral	124	33905
mount_rushmore	138	45350
piazza_san_marco	249	95895
reichstag	75	17823
sagrada_familia	401	120723
st_pauls_cathedral	615	98872
united_states_capitol	258	35095
Total	4107	696k

- 4k images in total for testing.
- Random bags of images are subsampled to form test subsets (size: 3, 5, 10, 25).

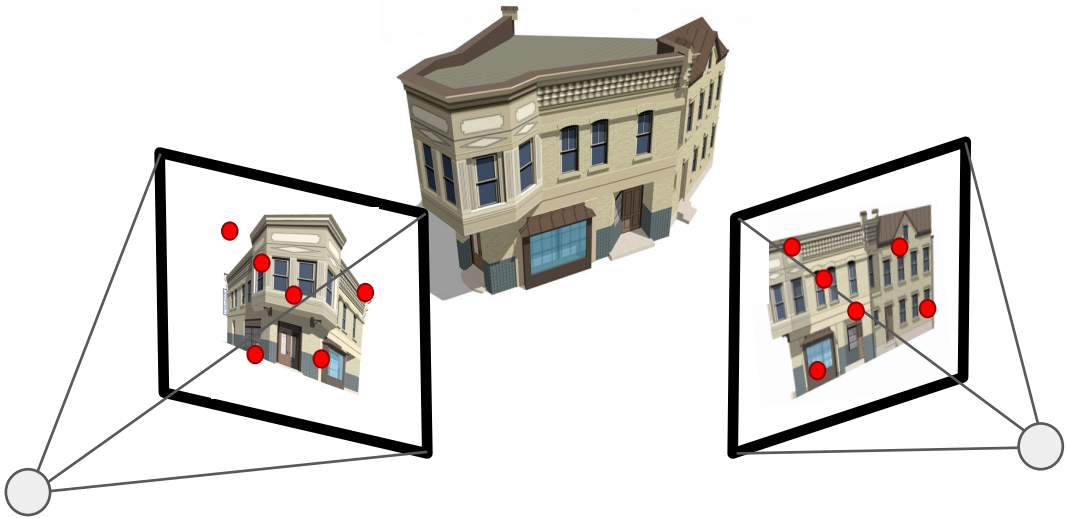
The phototourism challenge: local features



- Submission: Features
- IMW evaluates them via a typical stereo/SfM pipeline
  - Nearest neighbor matching
  - 1-to-1 matching
  - RANSAC
  - COLMAP

Hotel Images are in the public domain. Modified to simulate 3D rotation

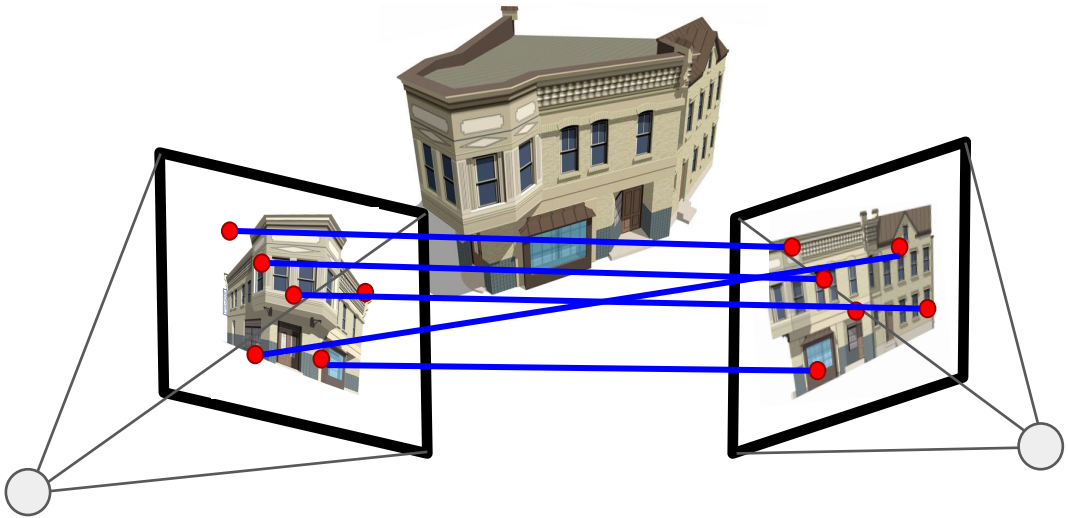
The phototourism challenge: local features



- Submission: Features
- IMW evaluates them via a typical stereo/SfM pipeline
  - Nearest neighbor matching
  - 1-to-1 matching
  - RANSAC_F
  - COLMAP

Hotel Images are in the public domain. Modified to simulate 3D rotation

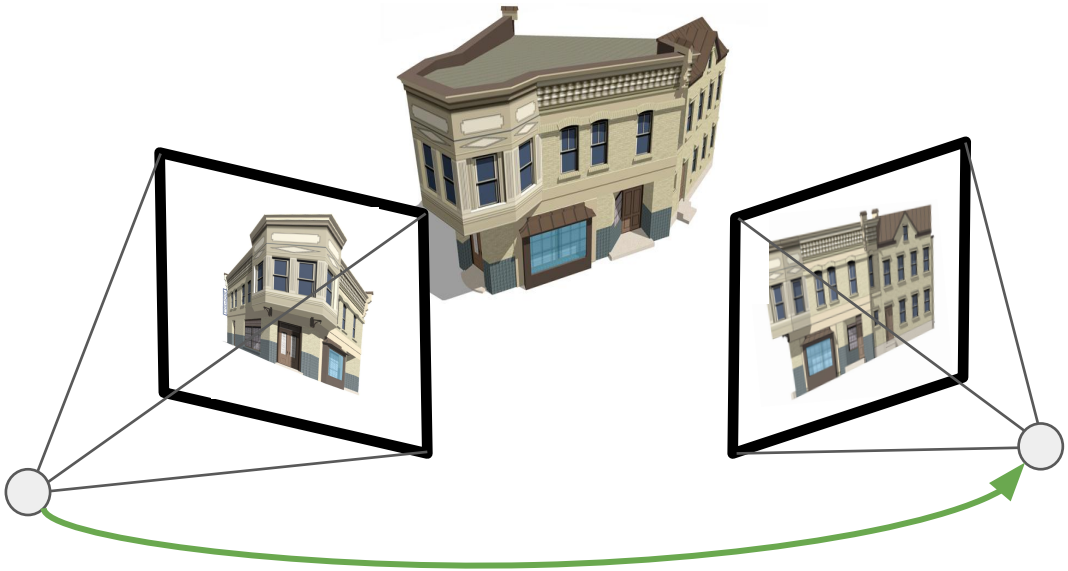
The phototourism challenge: matches



- Submission: Features + Matches
- IMW evaluates them via a typical stereo/SfM pipeline
  - ~~Nearest neighbor matching~~
  - ~~1-to-1 matching~~
  - RANSAC_F
  - COLMAP

Hotel Images are in the public domain. Modified to simulate 3D rotation

The phototourism challenge: poses



- Submission: Poses
- IMW evaluates them via a typical stereo/SfM pipeline
  - ~~Nearest neighbor matching~~
  - ~~1-to-1 matching~~
  - ~~RANSAC_F~~
  - ~~COLMAP~~

Hotel Images are in the public domain. Modified to simulate 3D rotation

The phototourism challenge: Stereo

Stereo — averaged over all sequences

Method	Date	Type	#kp	MS	mAP$^{5°}$	mAP$^{10°}$	mAP$^{15°}$	mAP$^{20°}$	mAP$^{25°}$
SIFT (OpenCV) kp:8000, match:nn	19-04-24	F	7884.4	0.2148	0.0004	0.0068	0.0277	0.0692	0.1303
SIFT (OpenCV) kp:2048, match:nn	19-05-17	F	2048.0	0.2215	0.0004	0.0055	0.0223	0.0570	0.1060

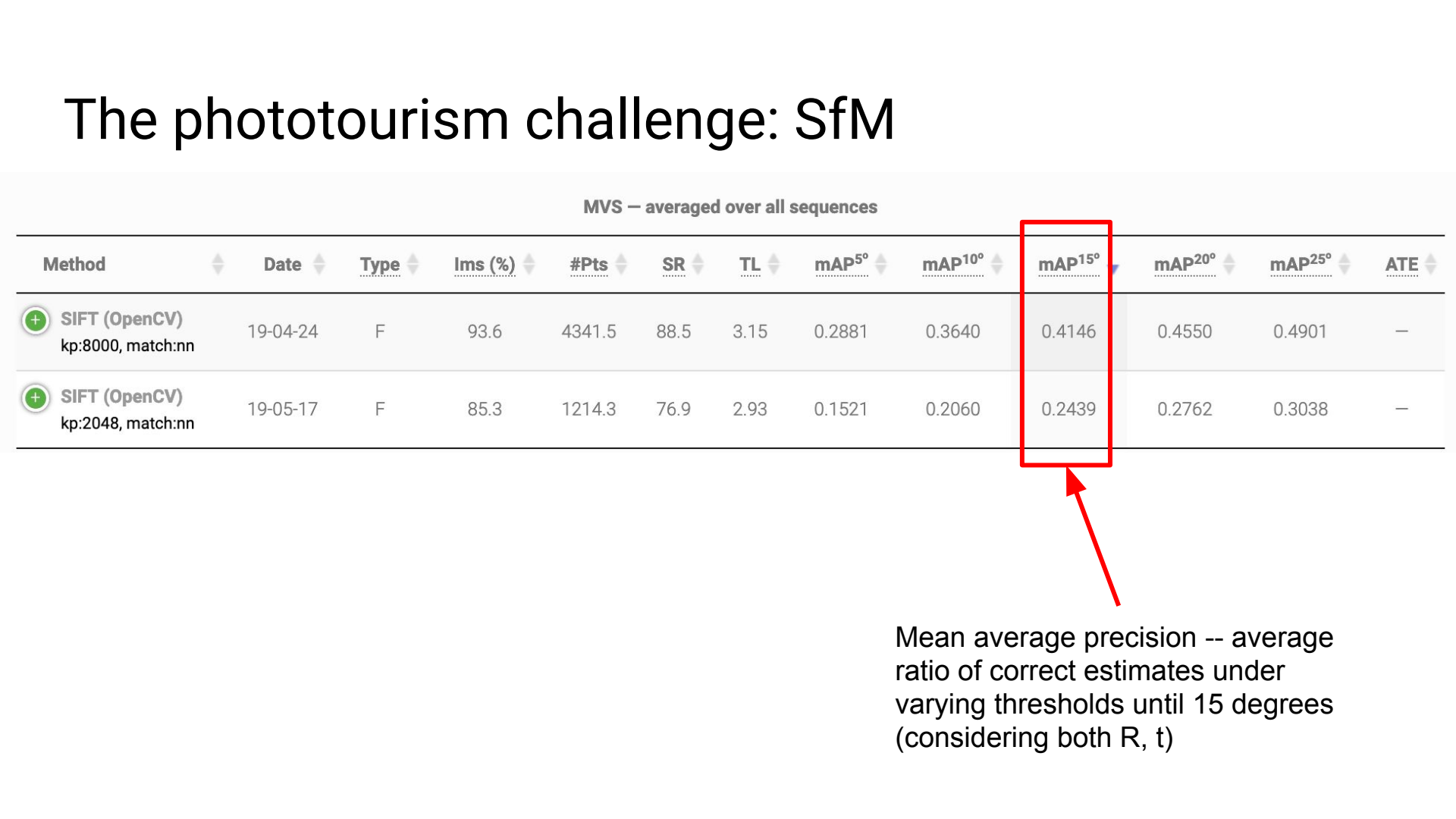
Matching score, but with symmetric epipolar distance for thresholding.

Mean average precision -- average ratio of correct estimates under varying thresholds until 15 degrees (considering both R, t)

The phototourism challenge: SfM

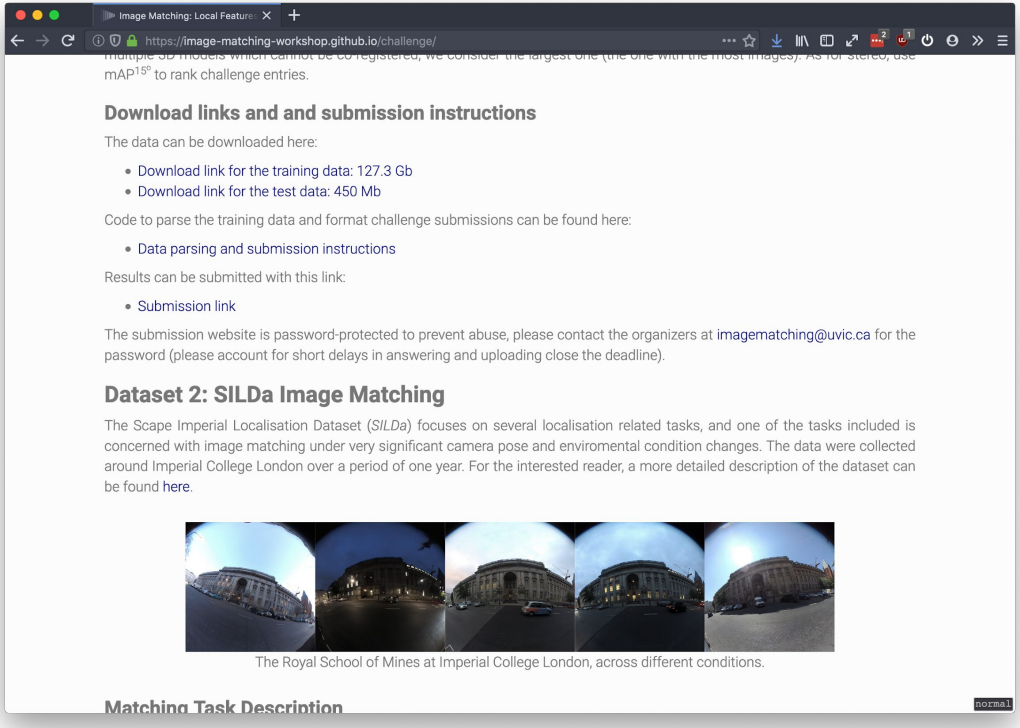
MVS — averaged over all sequences

Method	Date	Type	Ims (%)	#Pts	SR	TL	mAP$^{5°}$	mAP$^{10°}$	mAP$^{15°}$	mAP$^{20°}$	mAP$^{25°}$	ATE
SIFT (OpenCV) kp:8000, match:nn	19-04-24	F	93.6	4341.5	88.5	3.15	0.2881	0.3640	0.4146	0.4550	0.4901	–
SIFT (OpenCV) kp:2048, match:nn	19-05-17	F	85.3	1214.3	76.9	2.93	0.1521	0.2060	0.2439	0.2762	0.3038	–



Mean average precision -- average ratio of correct estimates under varying thresholds until 15 degrees (considering both R, t)

The phototourism challenge: Submission



- Upload server is password protected
  - Contact us for password
- Submission rules to be updated soon
  - We used roughly 55 core-years for this year challenge alone :-)

compute canada

- Code release soon
  - Welcoming contributions (and criticism!)

SILDa Challenge

Vassileios Balntas (Scape)

SILDa Challenge

Vassileios Balntas (Scape)

Axel Baroso (Imperial College London)
Krystian Mikolajczyk (Imperial College London)
Rigas Kouskouridas (Scape Technologies)
Duncan Frost (Scape Technologies)
Huub Heijnen (Scape Technologies)

SILDa: Key facts

- 14k images collected around Imperial College London over 1.5 year
- Rain, snow, sun, evening, night, morning
- Significant variations in the scenes

Commonwealth Coaches

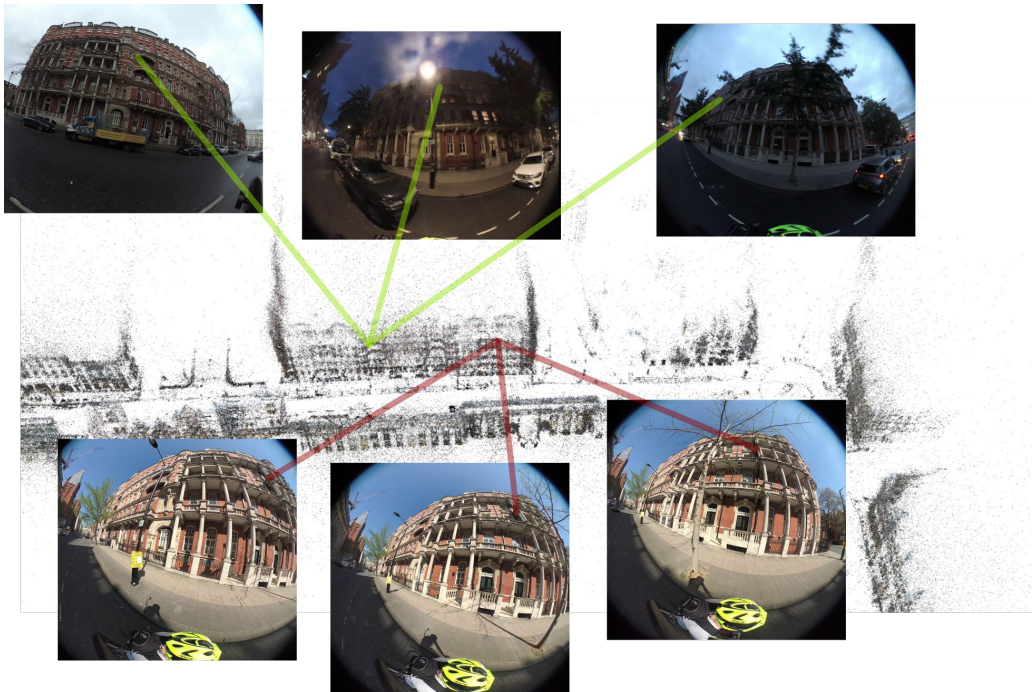
Imperial College
London

3D Reconstruction

- SfM with calibrated spherical cameras
- Chain SfM to help out matches: e.g. day -> evening & evening -> night.
- 1.4M points in the point cloud
- Covering almost 20 passes of 1.6km road

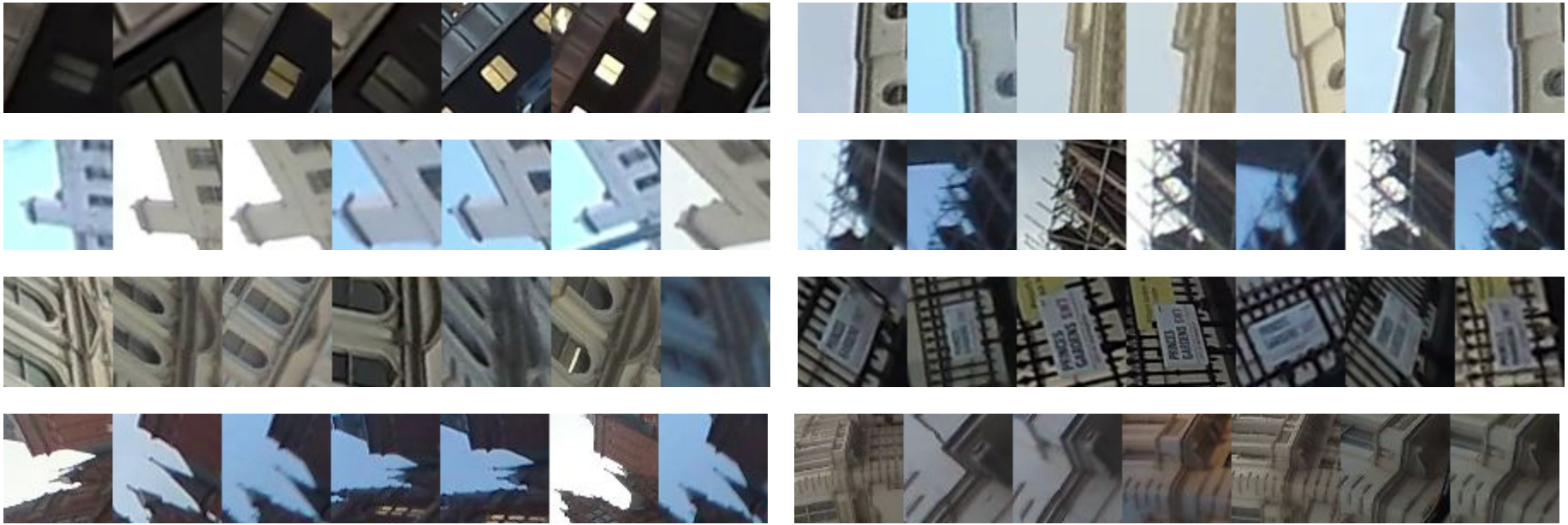
Local patches

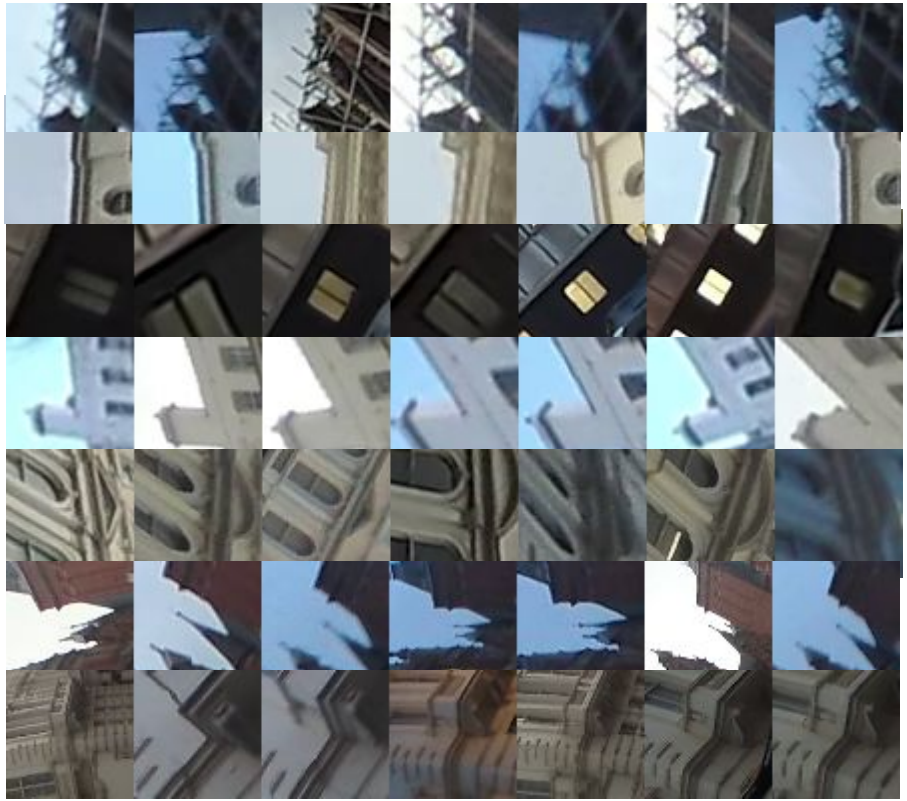
- Similarly to Brown and HPatches we extract a set of patches from the 3d points across different days, times and conditions

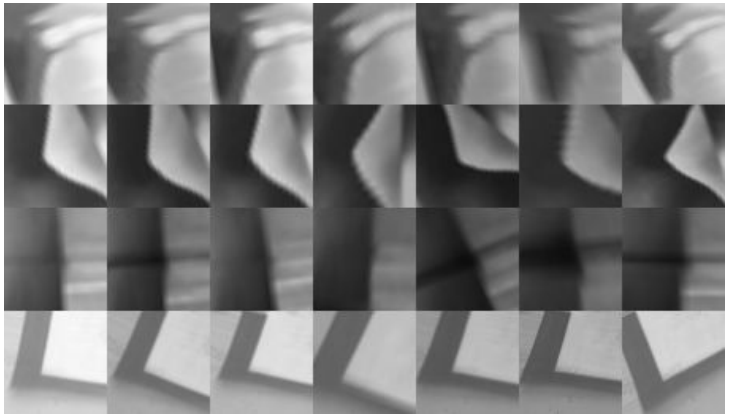


Local patches

- Similarly to Brown and HPatches we extract a set of patches from the 3d points across different days, times and conditions







Are patches still relevant?

- Is colour important for descriptors (CNN)?
- Is patch matching a good proxy for image matching?
- Is the separate evaluation of detector/descriptor the best strategy?

IMW Challenge: Image Pairs

- We generate 100k image pairs, which are deemed *difficult*
  - *difficult*: small number of inlier matches (<100) during the SfM process, but contain common point cloud points.
  - why focus on *difficult?*
    - *lots of SfM pairs are very incremental in terms of camera motion and end up having a big amount of inliers (>1000)*

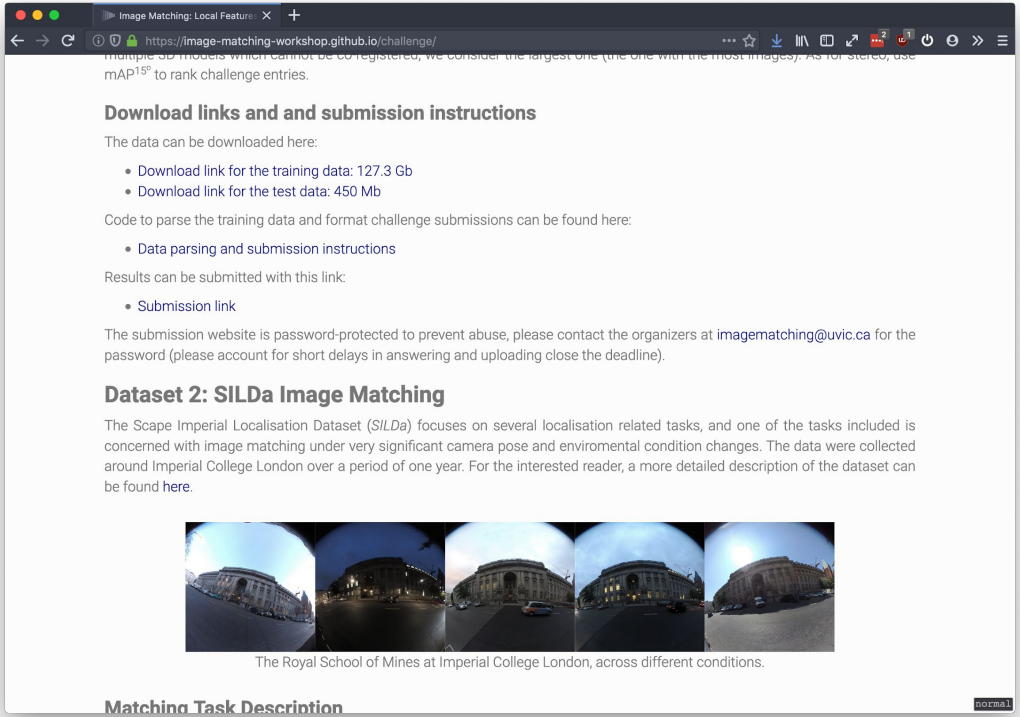
SOLO

MAERSK

Evaluation Protocol: Epipolar Arcs

Evaluation Protocol: Epipolar Arcs

SILDa challenge: Submission



- Online server will be available later on
- Hidden test set
- Future: more baselines D2Net, ContextDesc etc...

SILDa Matching Challenge: 3 Evaluation Metrics

- **Matching Scores**: Define a threshold on epipolar arc distance error, and use this to compute correct matches
- **Epipolar Arc Distance Statistics**: average/median epipolar arc distances between matches
- **Number of image pairs with more than 8 inliers**

https://image-matching-workshop.github.io/

Program

8:45 - 9:00	**Welcome**
9:00 - 9:30	**Amir Zamir (Stanford/UC Berkeley)** *Collection of Large-scale Densely-labeled 3D Data from the Real World Without a Single Click*
9:30 - 10:15	**Jiri Matas (CTU Prague)** *On the Art of Establishing Correspondence*
10:15 - 11:00	**Coffee Break + Poster Session**
11:15 - 12:00	**Torsten Sattler (Chalmers U. of Technology, Gothenburg)** *In Defense of Local Features for Visual Localization*
12:00 - 12:15	**IMW2019 Challenge**
12:15 - 12:30	**Zixin Luo (HKUST)** Winner of the Phototourism Challenge
12:30 - 12:45	**Challenge results and awards**