

3DV 2020



THALES

# Learning to Guide Local Feature Matches



François Darmon<sup>1,2</sup>



Mathieu Aubry<sup>2</sup>



Pascal Monasse<sup>2</sup>

<sup>1</sup>Thales LAS France

<sup>2</sup>LIGM (UMR 8049), Ecole des Ponts, Univ. Gustave Eiffel, CNRS

Vendredi  
15 Mars 1902

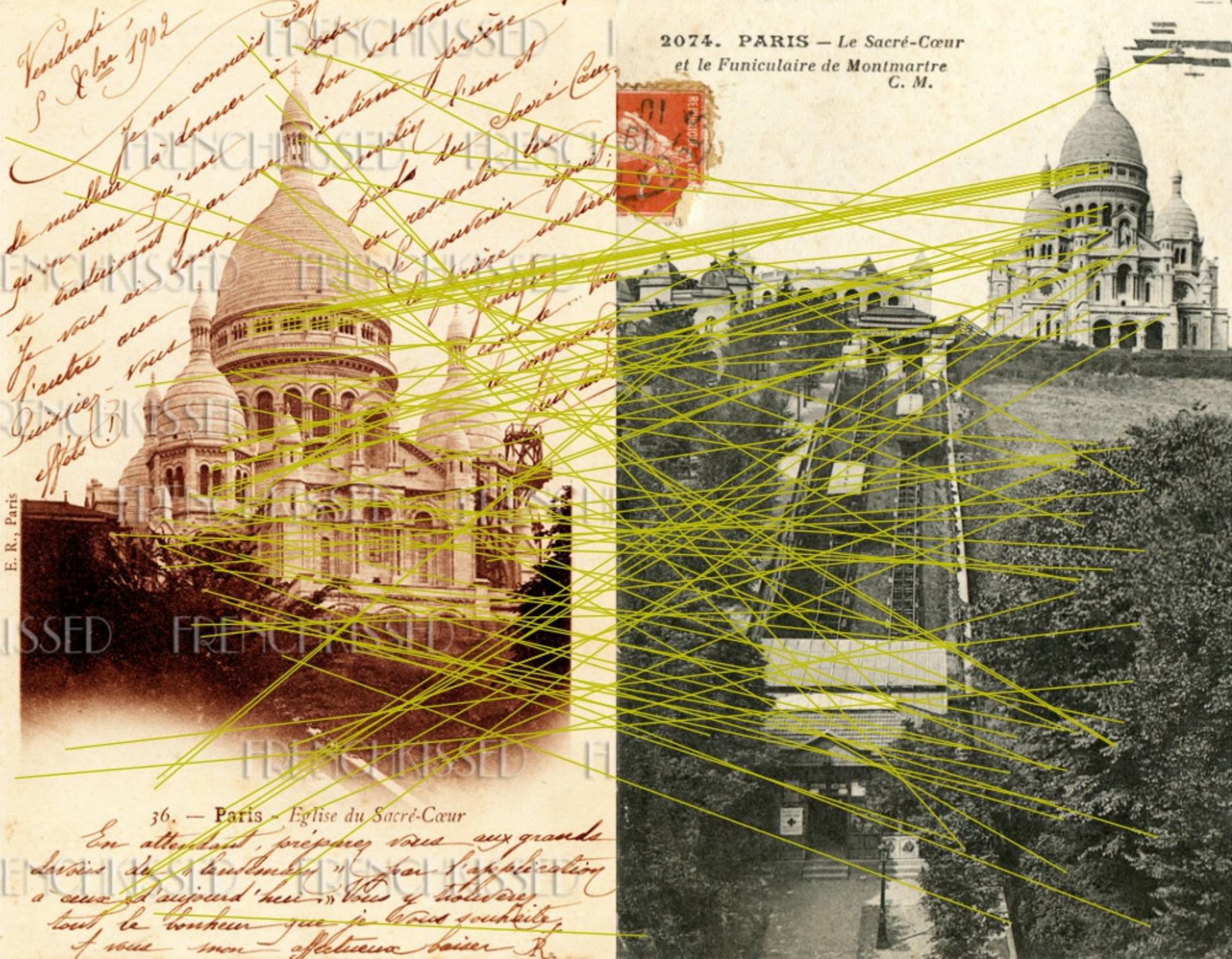
de mailler Je ne connais rien  
de meilleur à donner à un bon souvenir  
qu'un aime qui s'entend intime précieuse  
se traitant par une si jolie fleur et  
Je vous ai donné le poids de Sacré-Cœur  
l'autre avec en souvenir les  
Précieuses - vous le souvenir précieux  
effets !

2074. PARIS — Le Sacré-Cœur  
et le Funiculaire de Montmartre  
C. M.

E. R., Paris

36. — Paris — Eglise du Sacré-Cœur

En attendant, préparez vous aux grands  
services de « l'aujourd'hui » par l'application  
à ceux d'aujourd'hui « vous y trouverez  
tout le bonheur que je vous souhaite,  
à vous mon affectueux baiser R.



Vendredi  
 3 Mars 1902

de la meilleure Je ne connais rien  
 que son aïme qui s'en donne à son  
 se trahissant par une si bon souvenir  
 Je vous ai donné ce petit livre de  
 l'autre avec le pied de Sacré-Cœur  
 Puisse-t-elle vous en rappeler les  
 efforts. Vous la prière rapait;  
 Justifiez  
 console... Vous  
 le compagne  
 plus tard!



E. R., Paris

2074. PARIS — Le Sacré-Cœur  
 et le Funiculaire de Montmartre  
 C. M.



36. — Paris - Eglise du Sacré-Cœur

En attendant, préparez vous aux grands  
 services du « dimanche » par l'application  
 à ceux d'aujourd'hui « vous y trouverez  
 tout le bonheur que je vous souhaite,  
 et vous mon affectueux baiser R.

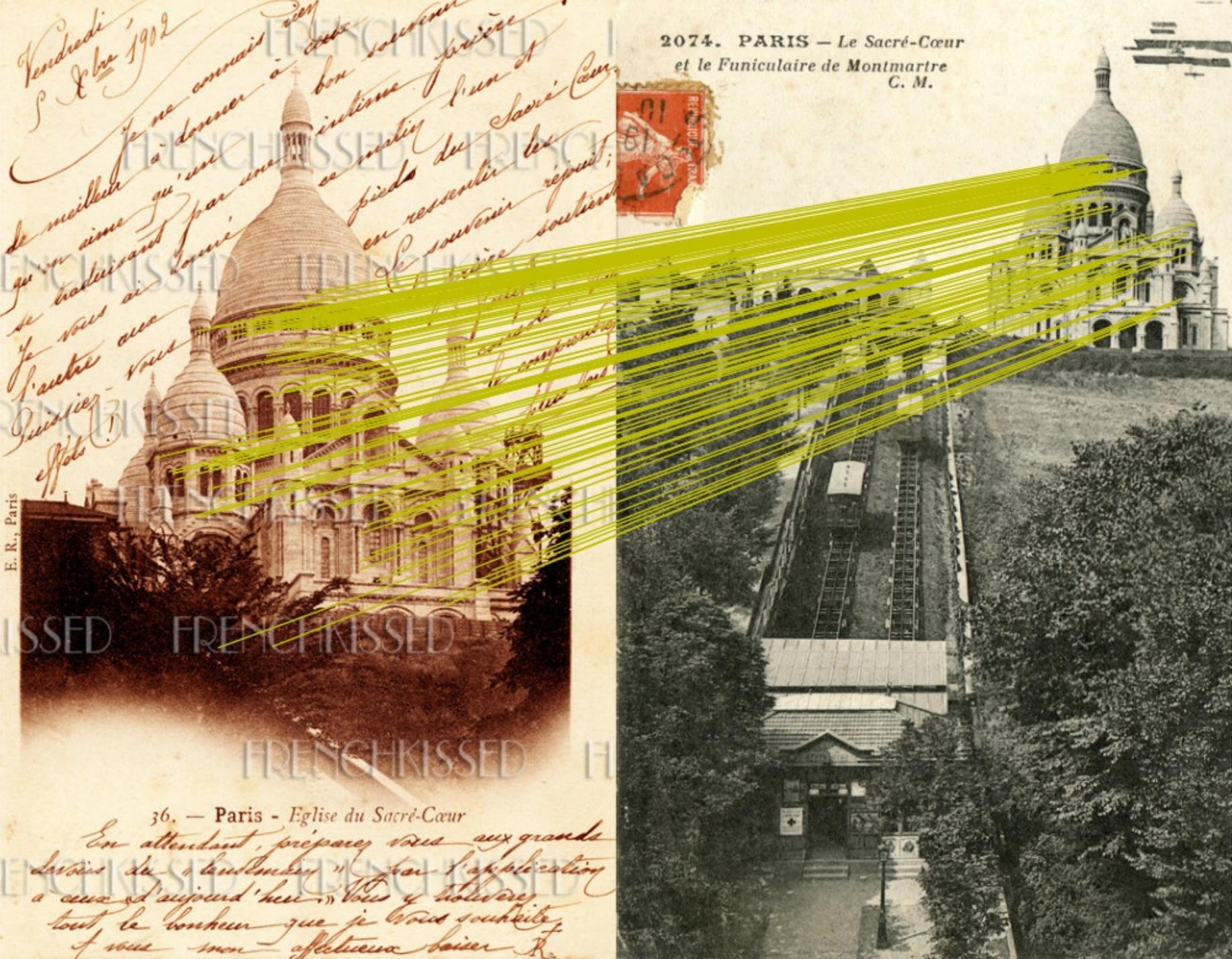
Vendredi  
 15 Mars 1902

de meilleur Je ne connais rien  
 que mon aïme qui' un bon souvenir  
 se traitant par une intime prière  
 Je vous ai donné le pied de Sacré-Cœur  
 l'autre avec en souvenir répit  
 Précisiez-vous le souvenir  
 Effets!

2074. PARIS — Le Sacré-Cœur  
 et le Funiculaire de Montmartre  
 C. M.



E. R., Paris



36. — Paris - Eglise du Sacré-Cœur

En attendant, préparez vous aux grands  
 services du « Landwey » par l'application  
 à ceux d'aujourd'hui « vous y trouverez  
 tout le bonheur que je vous souhaite,  
 et vous mon affectueux baiser R.

# Related work

- Local features:
  - Classical methods: SIFT and others
  - Deep learning based features: Superpoint, Contextdesc, D2Net, many others
- Guided feature matching
  - Very natural idea but not studied much
  - Guide local features with the best features

➔ Our approach is the first guided matching to improve results on challenging setups

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.

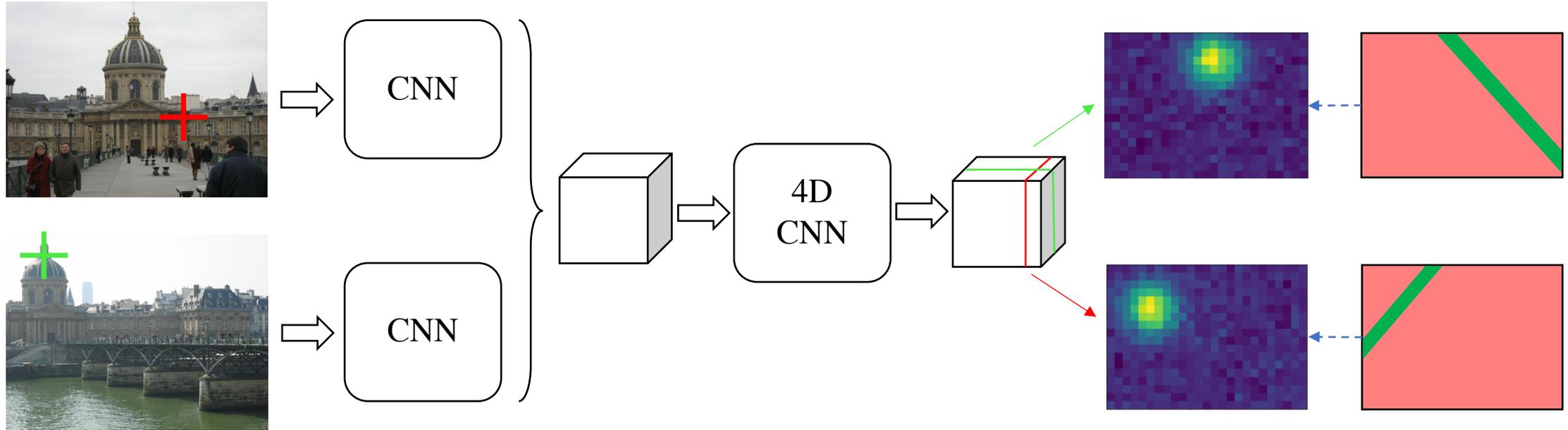
Shah, R., Srivastava, V., & Narayanan, P. J. (2015, January). Geometry-aware feature matching for structure from motion applications. In 2015 IEEE Winter Conference on Applications of Computer Vision (pp. 278-285). IEEE.

DeTone, D., Malisiewicz, T., & Rabinovich, A. (2018). Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 224-236).

Luo, Z., Shen, T., Zhou, L., Zhang, J., Yao, Y., Li, S., ... & Quan, L. (2019). Contextdesc: Local descriptor augmentation with cross-modality context. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2527-2536).

Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., & Sattler, T. (2019). D2-net: A trainable cnn for joint description and detection of local features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8092-8101).

# Coarse matching network



ResNet Feature  
extraction

Correlation  
volume

Output  
volume

Heatmaps

Supervision  
signal

# Image supervision

Positive label



Negative label



$$l_{image} = \pm \sum_{i,j} \max_{k,l} s_{ijkl}$$

# Epipolar supervision



$$l_{epipolar} = \frac{1}{|\mathcal{N}|} \sum_{(i,j) \in \mathcal{N}} \max_{kl} s_{ijkl} - \frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} \max_{kl} s_{ijkl}$$

# Point supervision



$$l_{points} = - \sum_{ij} \max_{(k,l) \in \mathcal{M}(i,j)} s_{ijkl}.$$

# Evaluation

- Coarse matching evaluation on MegaDepth
- Visual localization on Aachen Daynight
- Two-views geometry estimation on YFCC and sun3D
- Structure from Motion (SfM) on LTLL

# Coarse matching stage

- MegaDepth SfM points as ground truth

- Metrics: proportion of points within 8/16/32 pixels of a coarse match



Threshold	Frozen features			Finetuned features		
	8	16	32	8	16	32
Image	34.5	55.0	65.36	36.3	57.8	68.7
Epipolar	43.1	62.4	70.7	<b>47.7</b>	<b>67.6</b>	<b>75.8</b>
Point	40.3	58.5	67.8	45.0	63.5	72.5

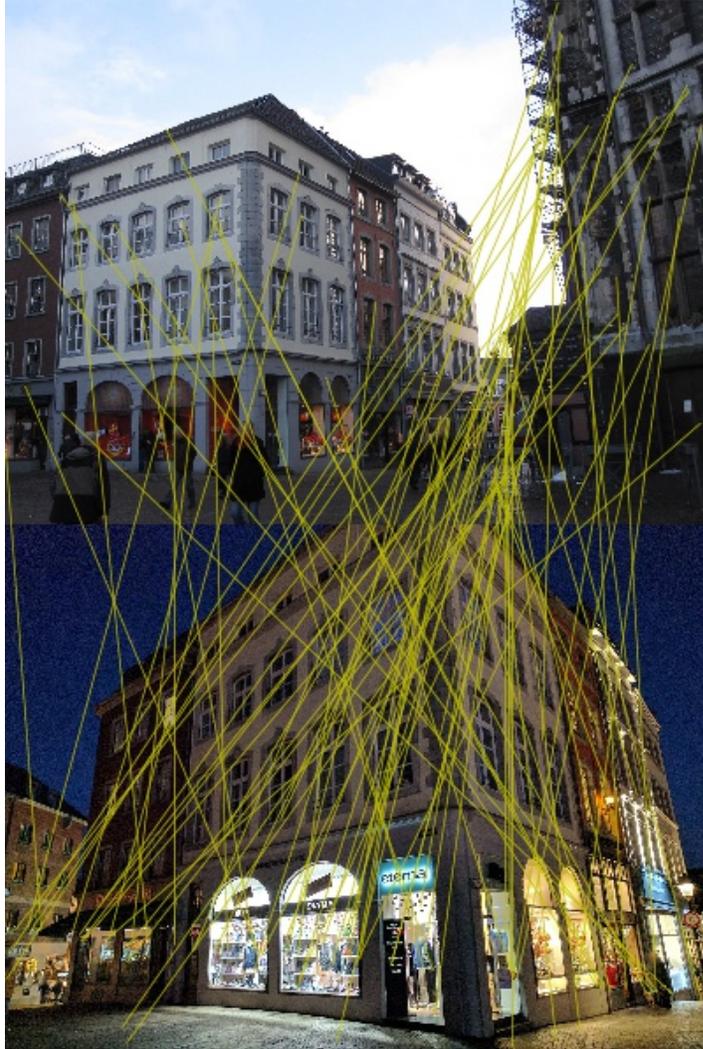
# Day night visual localization

- Aachen daynight dataset: local feature benchmark
- Input:
  - Set of calibrated day-time images
  - Set of query night-time images
  - List of overlapping image pairs to match
- Metrics: localization accuracy of query images at several thresholds

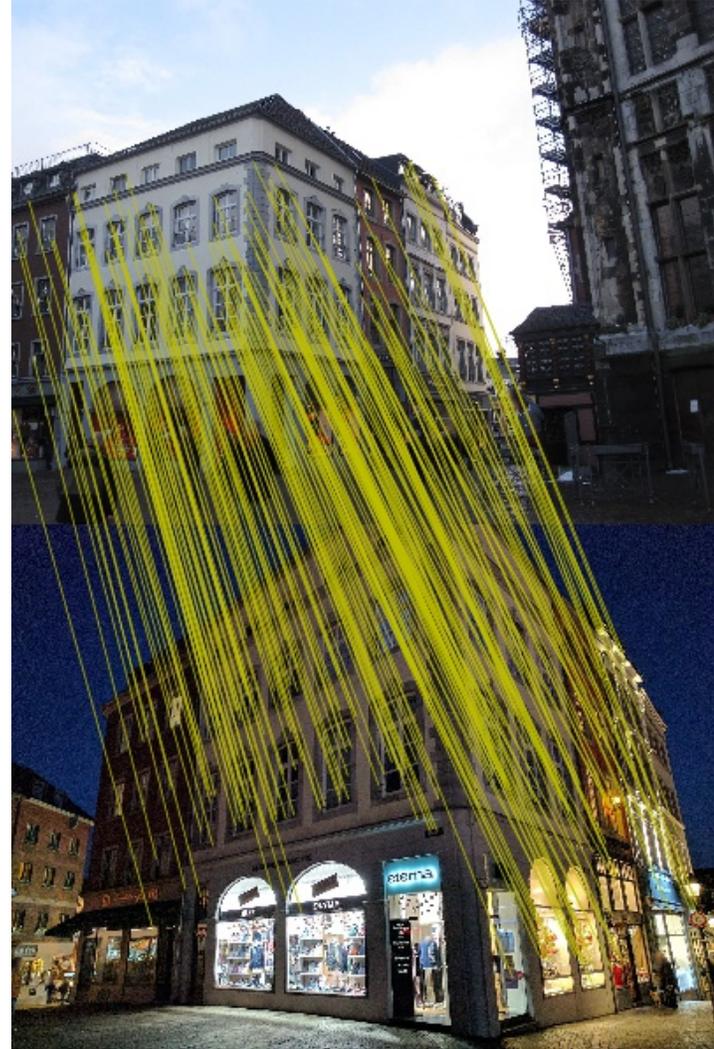


Features	Matching	Localization thresholds		
		(0.25m, 2°)	(0.5m, 5°)	(5m, 10°)
SIFT	Raw	38.8	51.0	58.2
	Ours	<b>66.3</b>	<b>84.7</b>	<b>96.9</b>
ContextDesc	Raw	60.2	74.5	87.8
	Ours	<b>75.5</b>	<b>85.7</b>	<b>98.0</b>
Superpoint	Raw	70.4	77.6	85.7
	Ours	<b>75.5</b>	<b>89.8</b>	<b>99.0</b>
D2-Net	Raw	<b>78.6</b>	85.7	<b>100</b>
	Ours	76.5	<b>87.8</b>	99.0

# Qualitative examples



SIFT matching

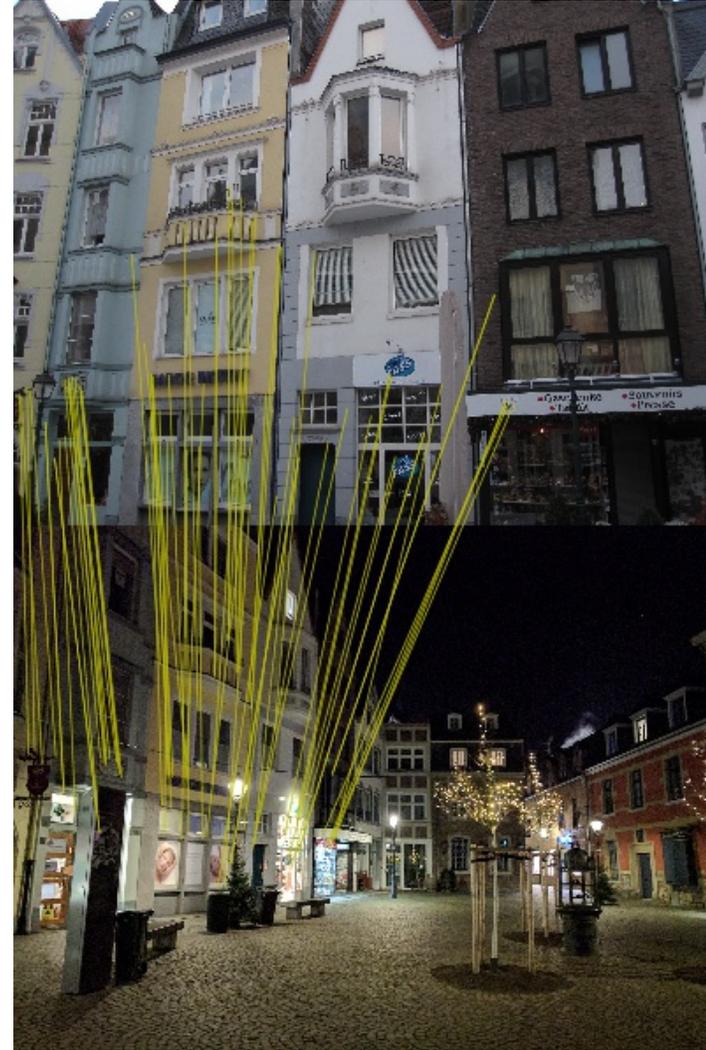


Our Guided SIFT matching

# Qualitative examples

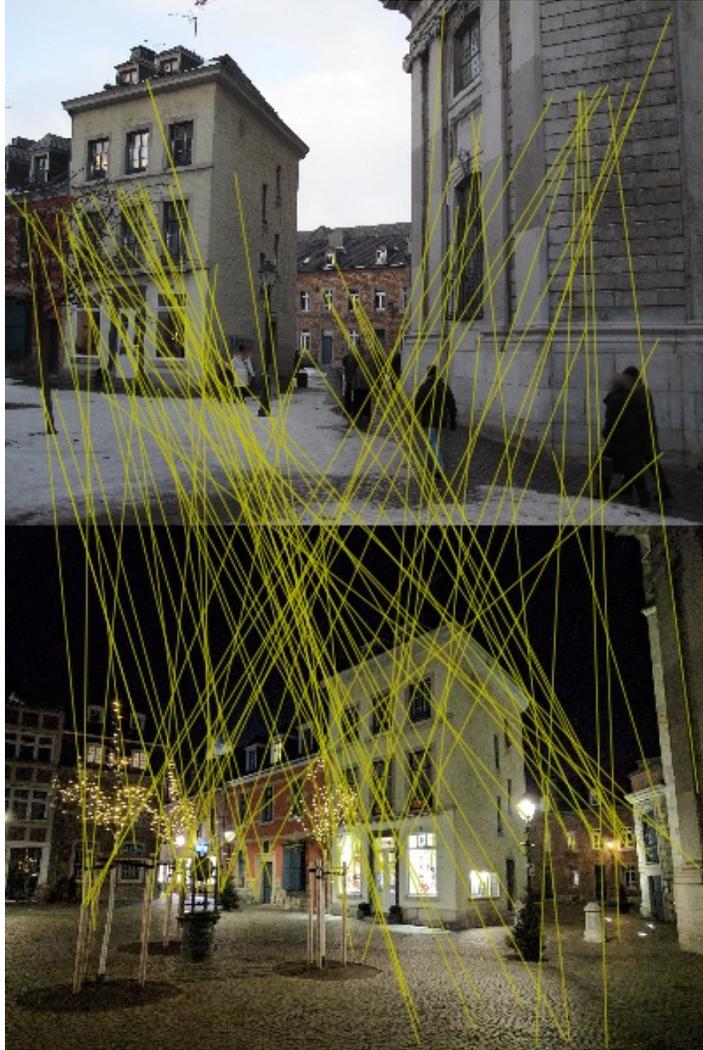


SIFT matching



Our Guided SIFT matching

# Qualitative examples

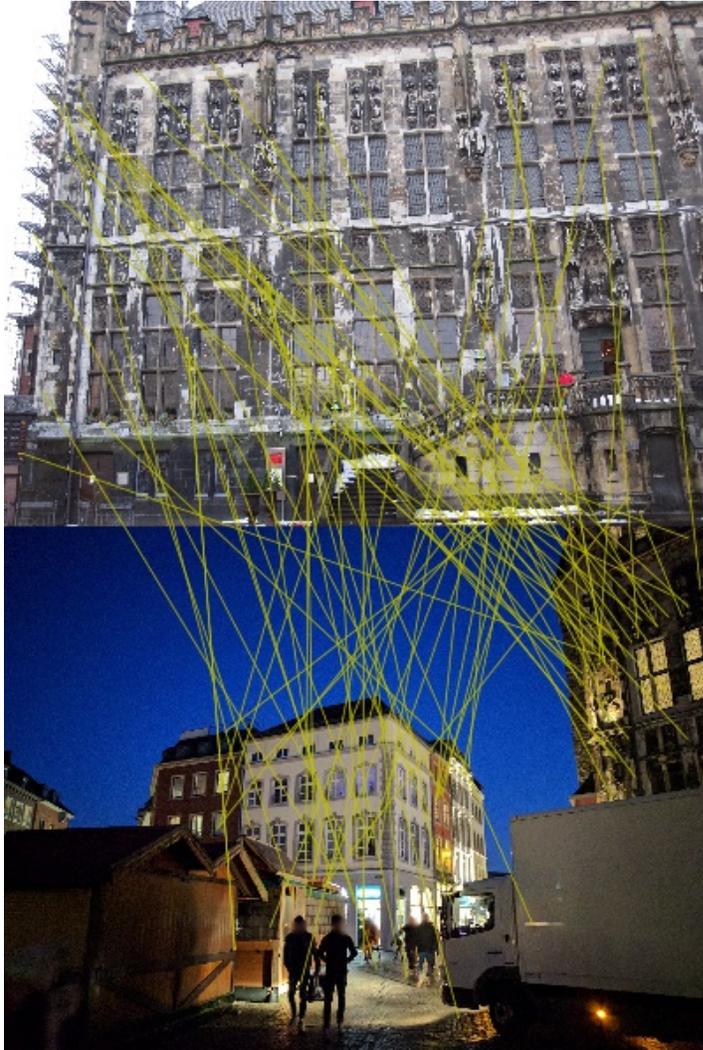


SIFT matching

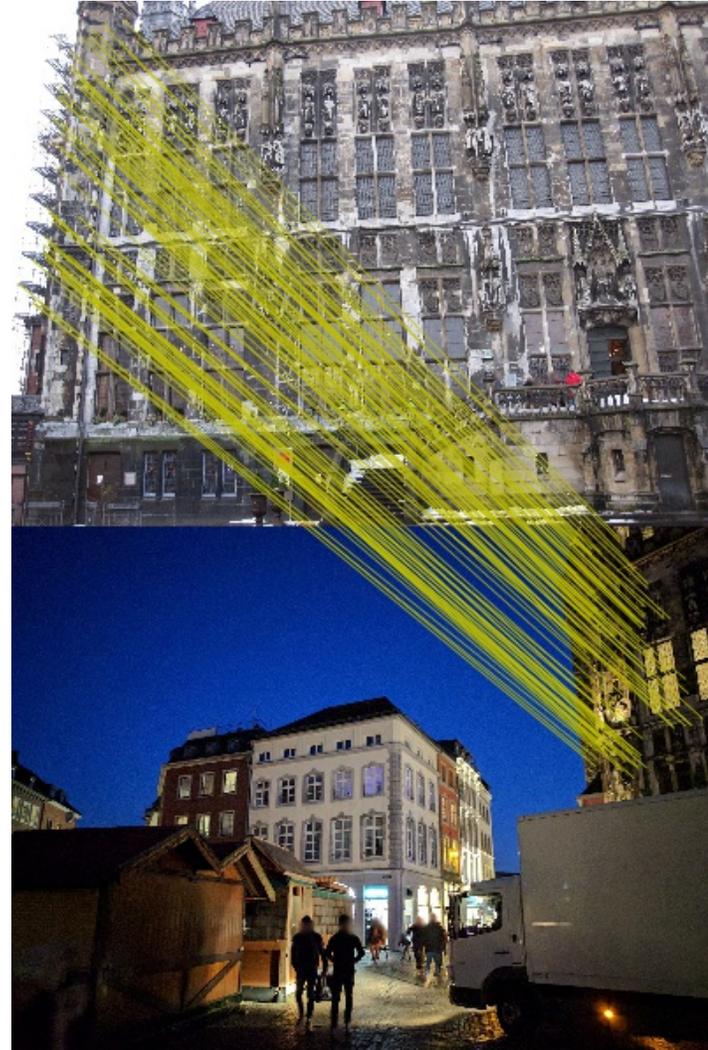


Our Guided SIFT matching

# Qualitative examples



SIFT matching



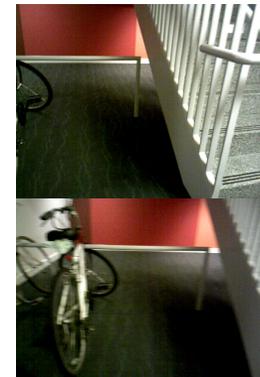
Our Guided SIFT matching

# Two-view geometry estimation

- Dataset:
  - 4 scenes of YFCC outdoor internet images
  - 15 scenes of sun3D indoor images
- Matches used for essential matrix estimation then relative pose estimation
- Metrics: AUC for localization accuracy



YFCC pairs



Sun3D pairs

# Quantitative results

Matches	Pre-filtering	YFCC (internet)			Sun3D (indoor)		
		5°	10°	20°	5°	10°	20°
Raw	None	8.45	13.80	22.4	2.34	4.70	9.61
	Bidirectional check	27.70	36.43	47.73	6.96	11.72	19.89
	Ratio test	41.75	51.63	62.23	13.48	20.93	31.48
	Ratio test + bid. check	46.80	57.41	67.80	14.52	22.74	34.22
	Ratio test + GMS [1]	30.43	38.30	48.16	11.49	17.89	27.46
Raw	CNNNet [30, 53]	47.98	58.13	68.67	15.98	-	-
	N <sup>3</sup> Net [32, 53]	49.13	-	-	15.38	-	-
	DFE [33, 53]	49.45	-	-	16.45	-	-
	OANet [53]	<b>52.08</b>	<b>62.38</b>	<b>72.66</b>	<b>17.25</b>	<b>26.60</b>	<b>39.50</b>
Guided epipolar [42]	Ratio test + bid. check	45.88	55.59	65.20	<u>15.86</u>	<u>24.52</u>	36.31
Guided homography	Ratio test + bid. check	46.00	55.65	65.46	15.15	23.55	35.36
Guided VGG4 [46, 49]	Ratio test + bid. check	31.23	40.49	51.51	3.97	7.23	13.16
Ours image guided	Ratio test + bid. check	43.50	52.99	63.24	15.45	23.84	35.81
Ours point guided	Ratio test + bid. check	47.43	57.71	68.59	15.61	24.24	36.37
Ours epipolar guided	Ratio test + bid. check	<u>49.60</u>	<u>60.36</u>	<u>71.37</u>	15.72	24.35	<u>36.40</u>

# Quantitative results

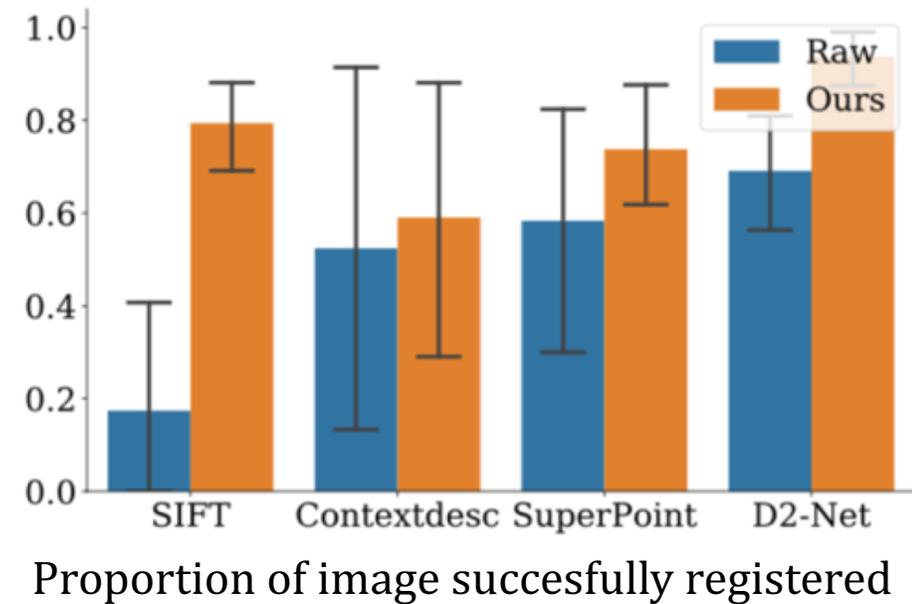
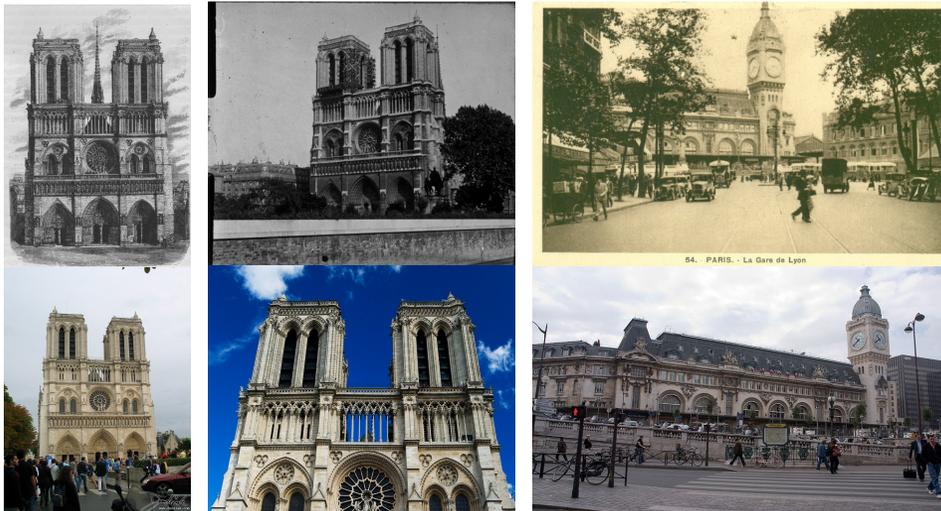
Matches	Pre-filtering	YFCC (internet)			Sun3D (indoor)		
		5°	10°	20°	5°	10°	20°
Raw	None	8.45	13.80	22.4	2.34	4.70	9.61
	Bidirectional check	27.70	36.43	47.73	6.96	11.72	19.89
	Ratio test	41.75	51.63	62.23	13.48	20.93	31.48
	Ratio test + bid. check	46.80	57.41	67.80	14.52	22.74	34.22
	Ratio test + GMS [1]	30.43	38.30	48.16	11.49	17.89	27.46
Raw	CNNNet [30, 53]	47.98	58.13	68.67	15.98	-	-
	N <sup>3</sup> Net [32, 53]	49.13	-	-	15.38	-	-
	DFE [33, 53]	49.45	-	-	16.45	-	-
	OANet [53]	<b>52.08</b>	<b>62.38</b>	<b>72.66</b>	<b>17.25</b>	<b>26.60</b>	<b>39.50</b>
Guided epipolar [42]	Ratio test + bid. check	45.88	55.59	65.20	<u>15.86</u>	<u>24.52</u>	36.31
Guided homography	Ratio test + bid. check	46.00	55.65	65.46	15.15	23.55	35.36
Guided VGG4 [46, 49]	Ratio test + bid. check	31.23	40.49	51.51	3.97	7.23	13.16
Ours image guided	Ratio test + bid. check	43.50	52.99	63.24	15.45	23.84	35.81
Ours point guided	Ratio test + bid. check	47.43	57.71	68.59	15.61	24.24	36.37
Ours epipolar guided	Ratio test + bid. check	<u>49.60</u>	<u>60.36</u>	<u>71.37</u>	15.72	24.35	<u>36.40</u>

# Quantitative results

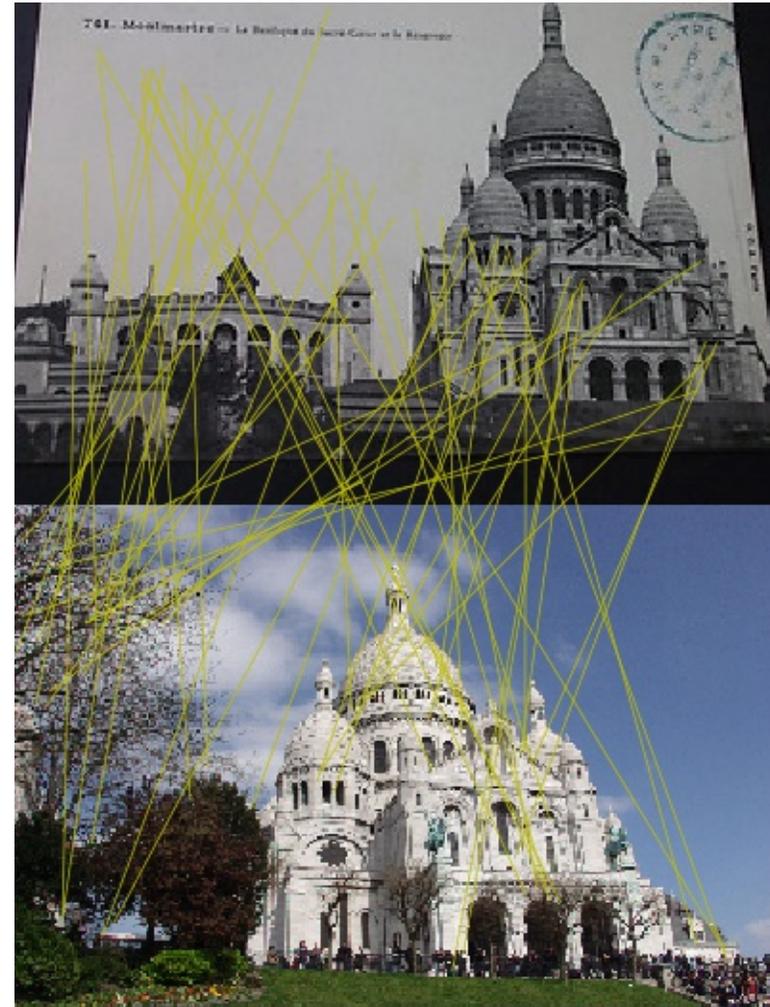
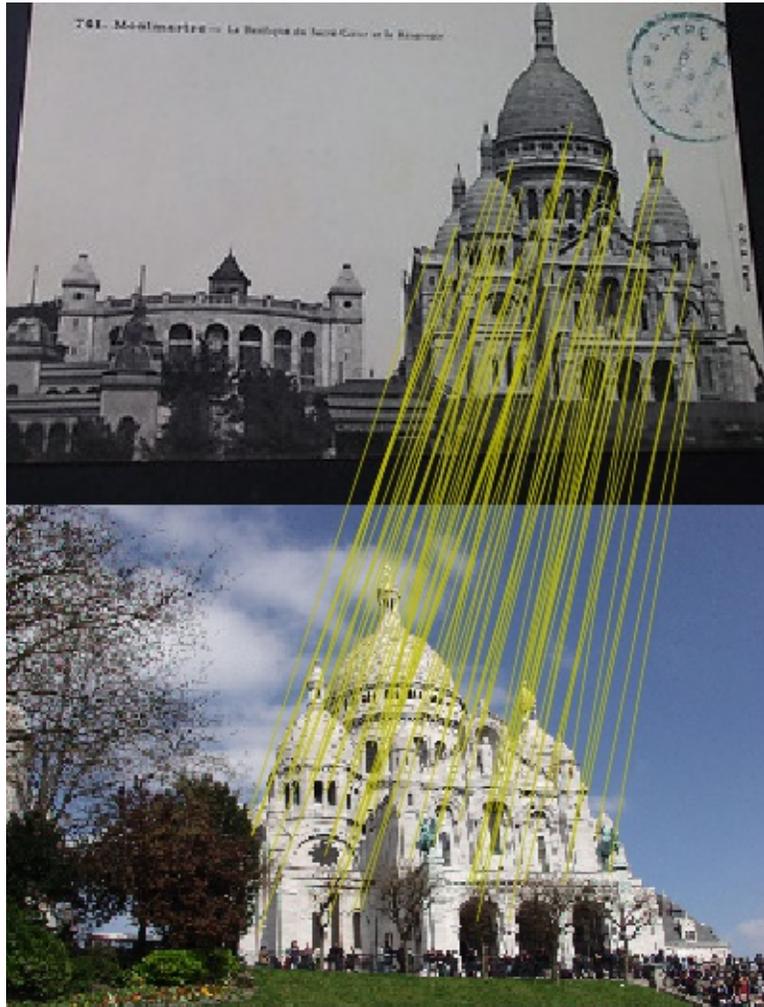
Matches	Pre-filtering	YFCC (internet)			Sun3D (indoor)		
		5°	10°	20°	5°	10°	20°
Raw	None	8.45	13.80	22.4	2.34	4.70	9.61
	Bidirectional check	27.70	36.43	47.73	6.96	11.72	19.89
	Ratio test	41.75	51.63	62.23	13.48	20.93	31.48
	Ratio test + bid. check	46.80	57.41	67.80	14.52	22.74	34.22
	Ratio test + GMS [1]	30.43	38.30	48.16	11.49	17.89	27.46
Raw	CNNNet [30, 53]	47.98	58.13	68.67	15.98	-	-
	N <sup>3</sup> Net [32, 53]	49.13	-	-	15.38	-	-
	DFE [33, 53]	49.45	-	-	16.45	-	-
	OANet [53]	<b>52.08</b>	<b>62.38</b>	<b>72.66</b>	<b>17.25</b>	<b>26.60</b>	<b>39.50</b>
Guided epipolar [42]	Ratio test + bid. check	45.88	55.59	65.20	<u>15.86</u>	<u>24.52</u>	36.31
Guided homography	Ratio test + bid. check	46.00	55.65	65.46	15.15	23.55	35.36
Guided VGG4 [46, 49]	Ratio test + bid. check	31.23	40.49	51.51	3.97	7.23	13.16
Ours image guided	Ratio test + bid. check	43.50	52.99	63.24	15.45	23.84	35.81
Ours point guided	Ratio test + bid. check	47.43	57.71	68.59	15.61	24.24	36.37
Ours epipolar guided	Ratio test + bid. check	<u>49.60</u>	<u>60.36</u>	<u>71.37</u>	15.72	24.35	<u>36.40</u>

# 3D reconstruction from historical data

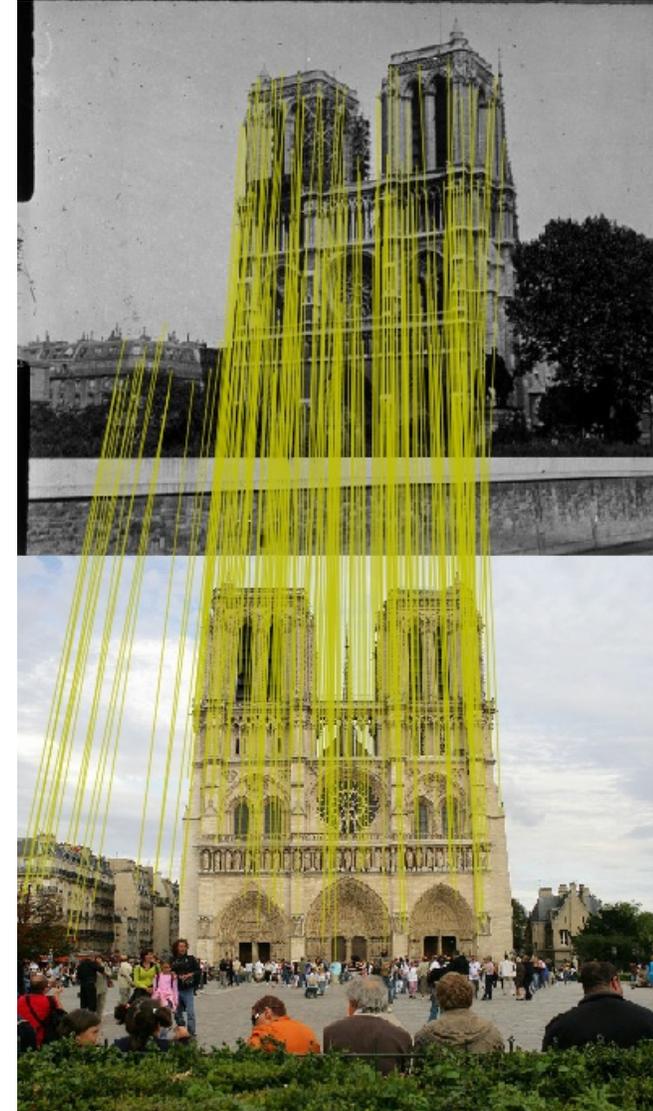
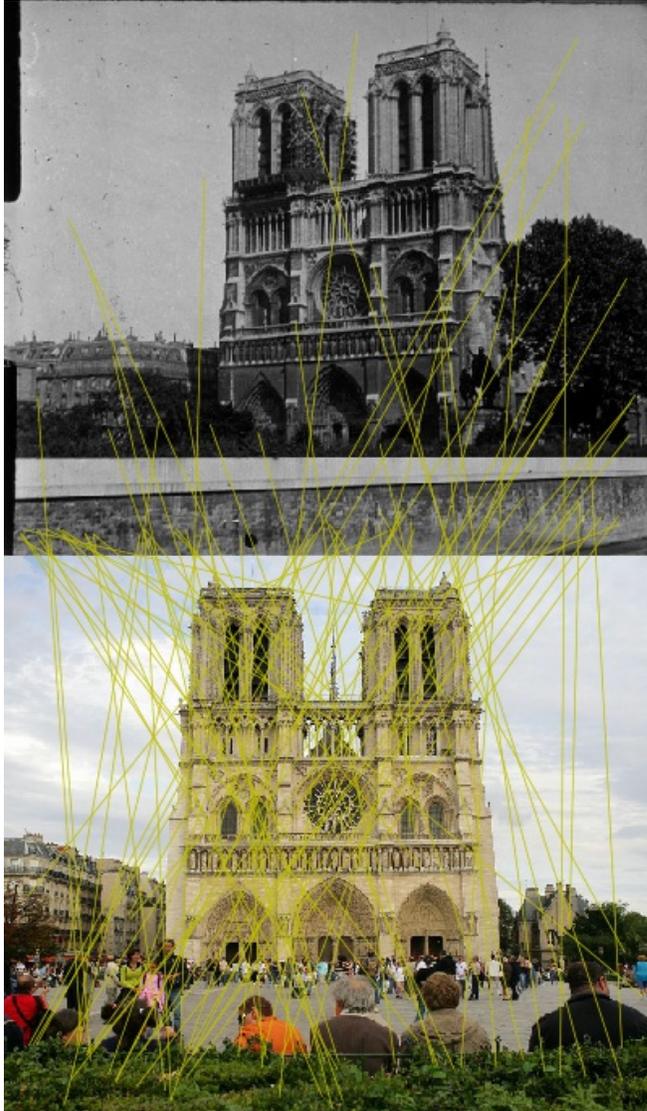
- LTL dataset: set of images with historic pictures and recent ones
- SfM on sets of historical pictures
- Metrics: average proportion of image successfully registered



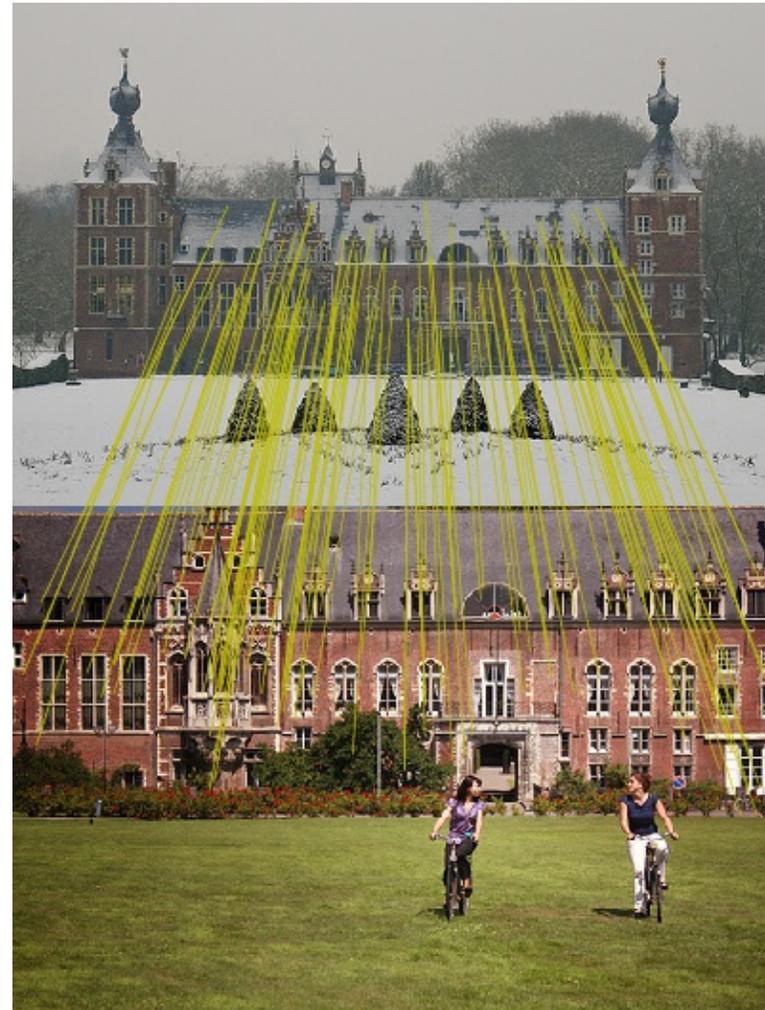
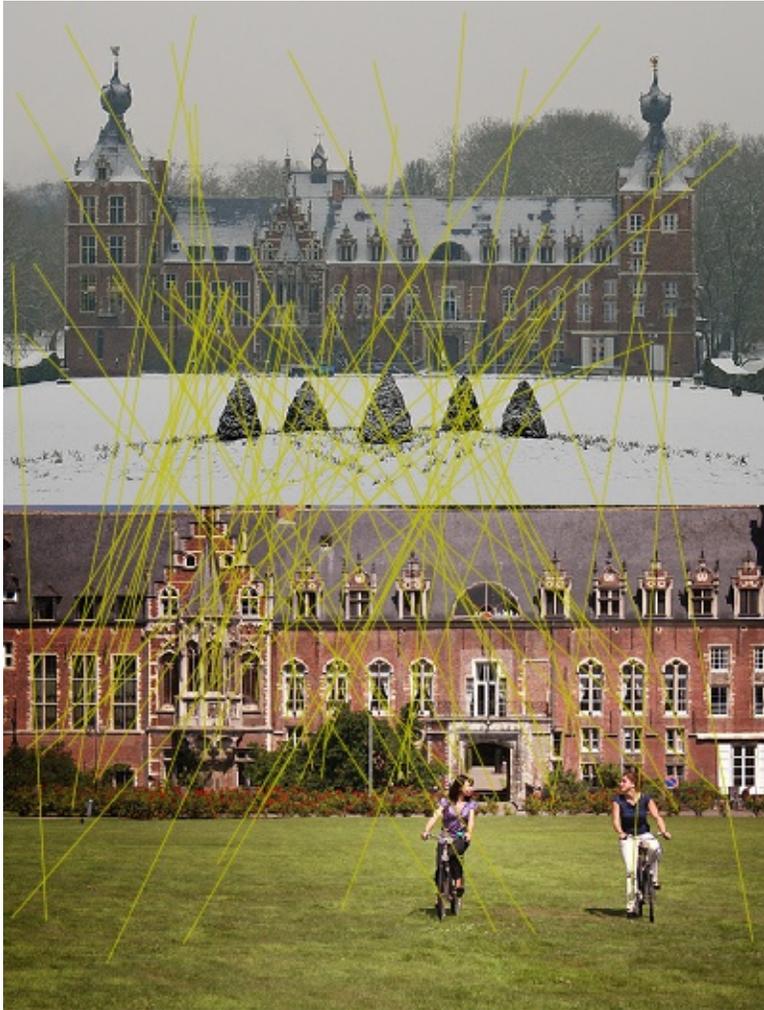
# Qualitative examples



# Qualitative examples



# Qualitative examples



# Conclusion

- We propose a new guided matching technique
- SIFT features are improved to the level of current Deep Features
- On some dataset, features can still be improved with our matching

