RANSAC-Flow: generic two-stage image alignment

Shen H., Darmon F., Efros A. and Aubry M.

Presented by: Joshua Heipel
Advisor: Philipp Schroeppel
Examiner: Prof. Dr. Thomas Brox

Outline

- Introduction
- Approach
  - Coarse Alignment (RANSAC)
  - Fine Alignment (Optical Flow Estimation)
  - Training Procedure
  - Inference
- Experiments
- Conclusion

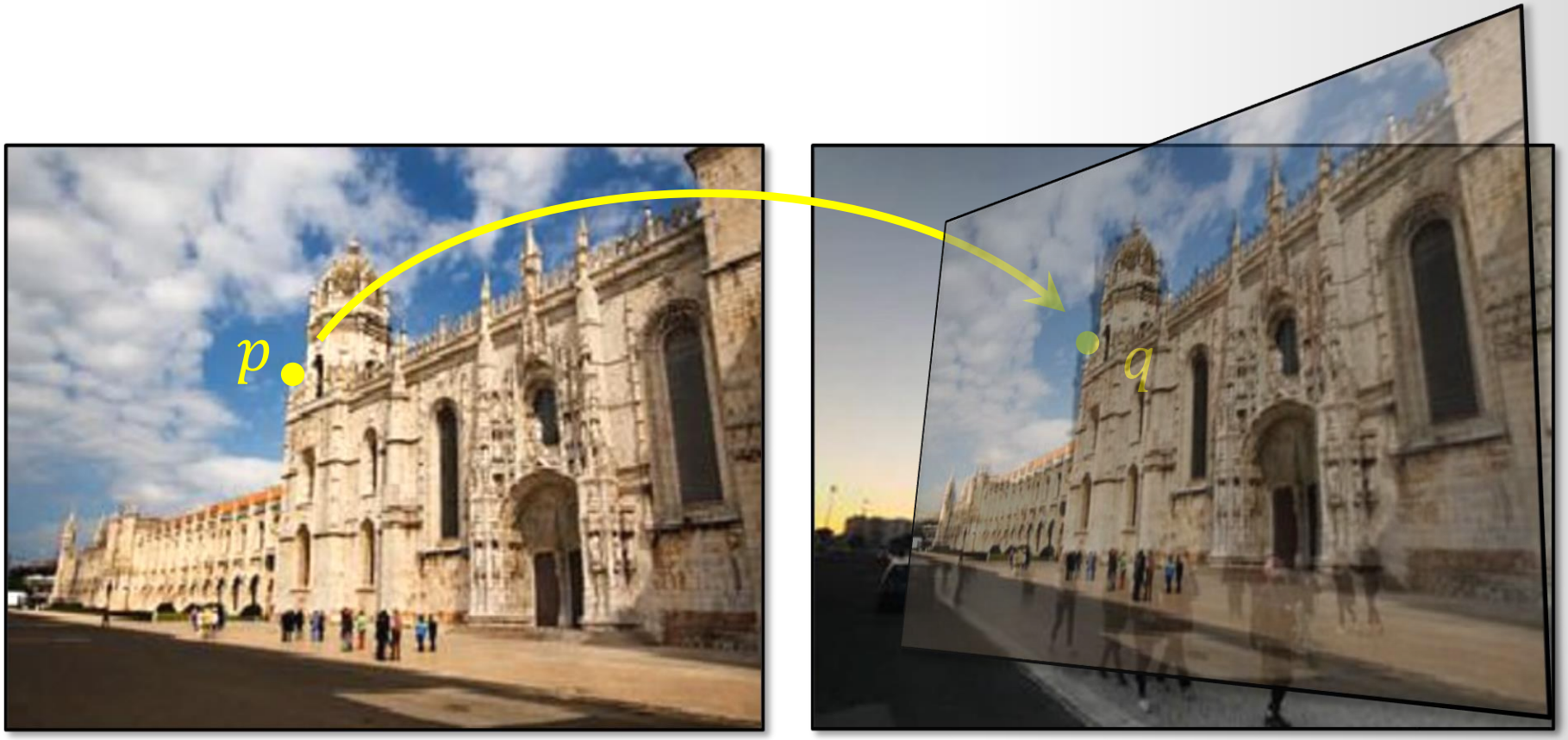
Introduction

Motivation

source

target

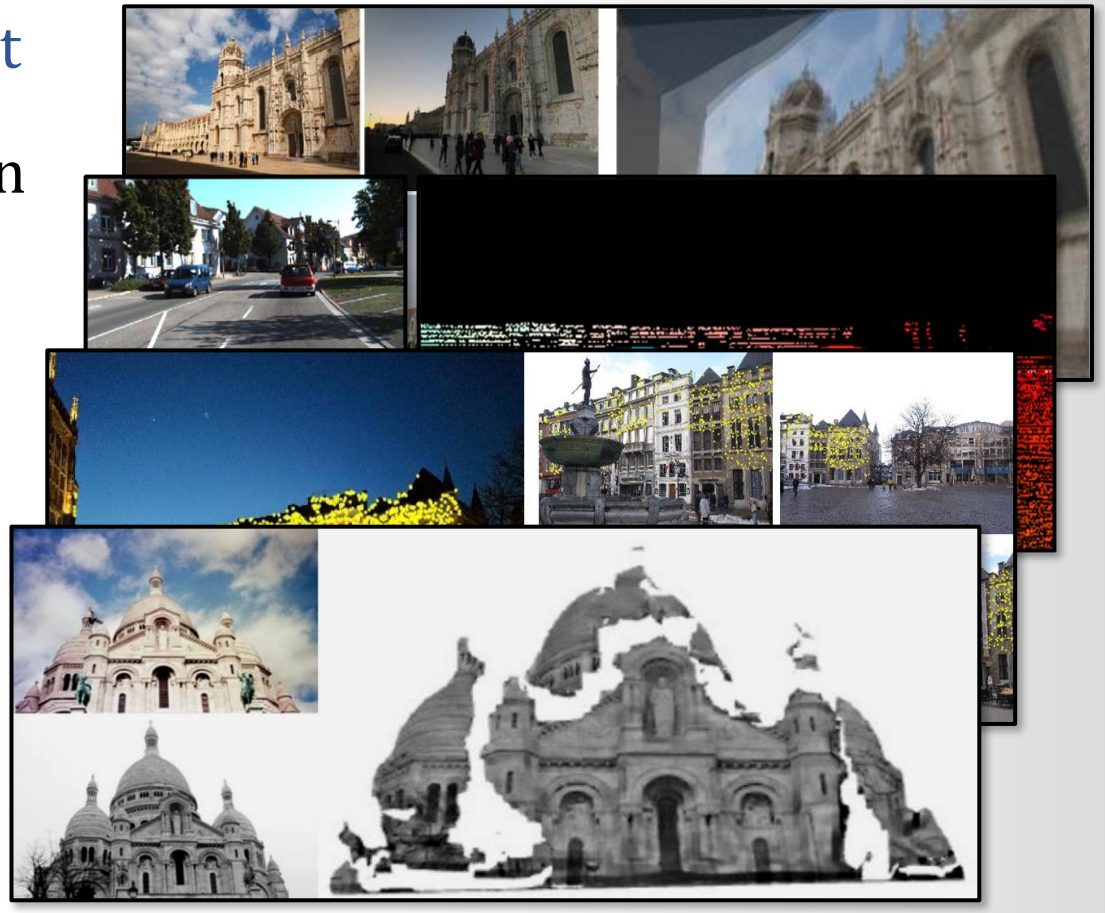
Motivation



source target

Motivation

- Dense Image Alignment
- Optical Flow Estimation
- Visual Localization
- 3D Reconstruction
- Artwork Alignment
- Texture Transfer
- …



Background

Parametric Methods

- Compute a global transformation $H(p_i) = q_i$ (affine, homographic, etc.)
- Based on sparse local image features (e.g. SIFT)
- Can deal with large displacements

Non-Parametric Methods

- Compute an individual displacement vector $\vec{w}_i$ for each pixel p_i
- Based on pixel similarities (e.g. brightness constancy)
- Are flexible towards the underlying transformation

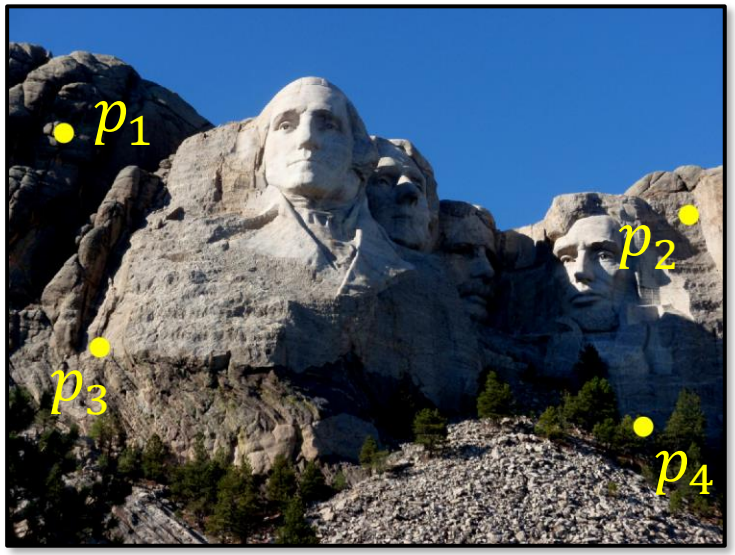
$\Rightarrow$ Hybrid two-stage image alignment

Main Idea (Illustration)

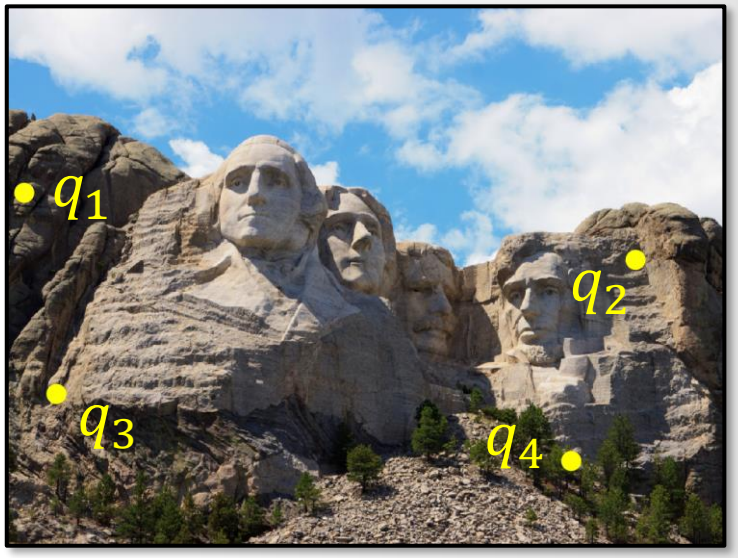
source

target

Main Idea (Illustration)

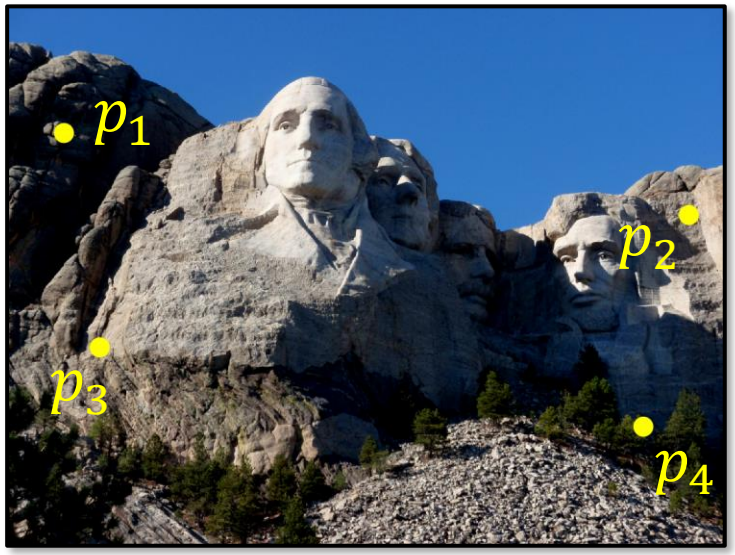


source

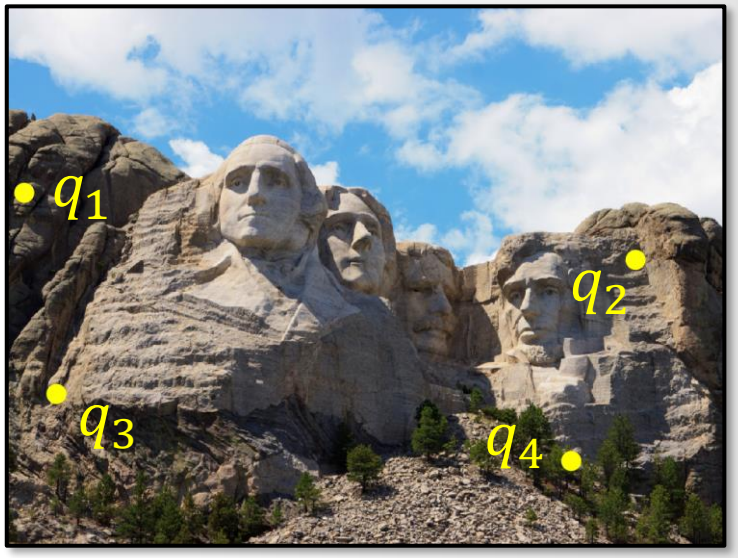


target

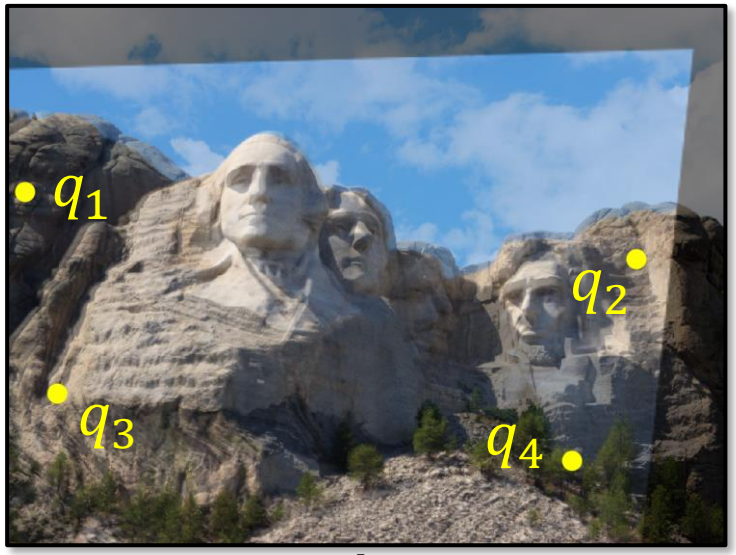
Main Idea (Illustration)



source



target



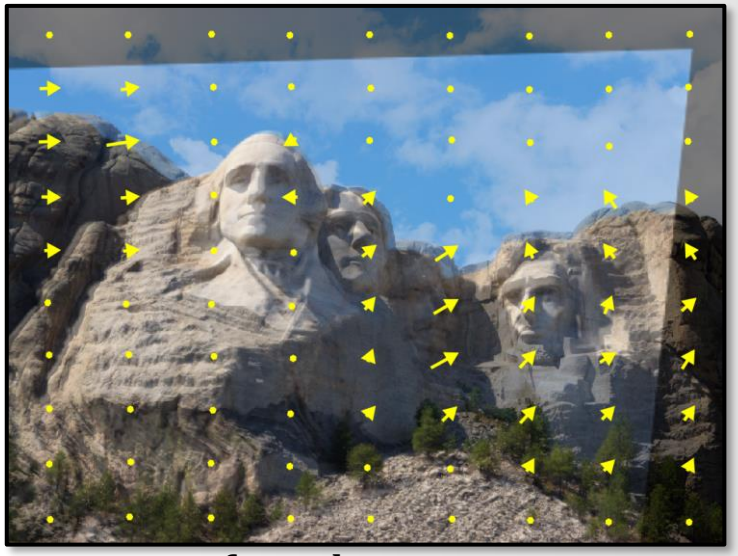
coarse alignment

Main Idea (Illustration)

source

target

coarse alignment



fine alignment

Main Idea (Illustration)

source

target

coarse alignment

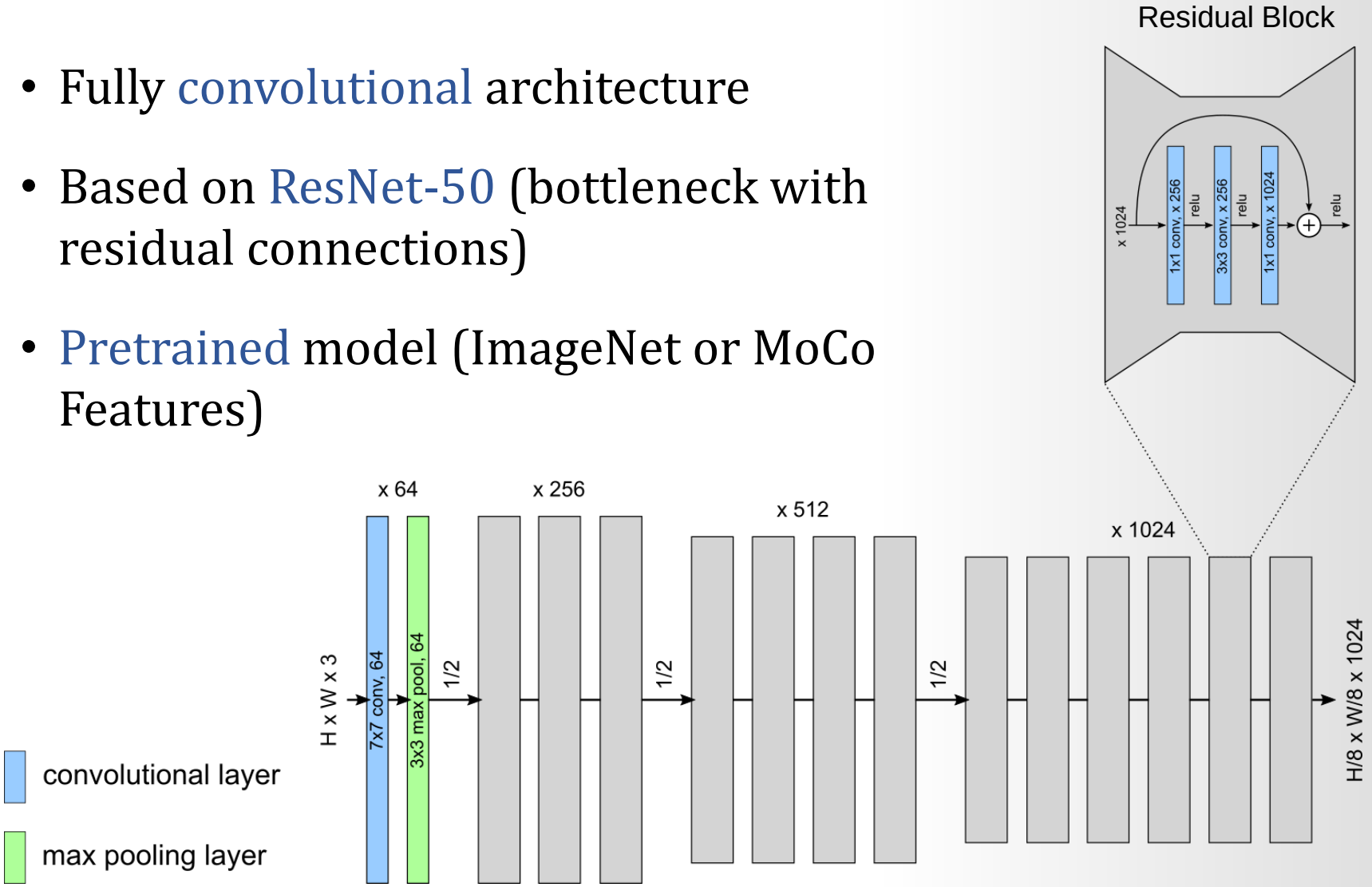
fine alignment

Approach

1. Coarse Alignment (RANSAC)

Deep Feature Extraction

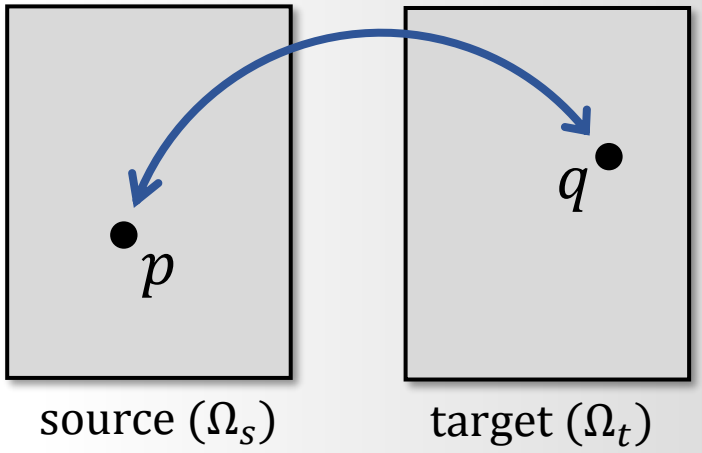
- Fully convolutional architecture
- Based on ResNet-50 (bottleneck with residual connections)
- Pretrained model (ImageNet or MoCo Features)



Deep Feature Extraction

- Source and target images (I_s and I_t) are processed independently
- Calculate dot product similarities between extracted feature maps ($\vec{f}_s$ and $\vec{f}_t$)
- A pair of positions (p, q) is a mutual match if:

$$\vec{f}_s(p)^{\mathrm{T}} \vec{f}_t(q) = \max_{q' \in \Omega_t} \{\vec{f}_s(p)^{\mathrm{T}} \vec{f}_t(q')\}$$
$$\vec{f}_s(p)^{\mathrm{T}} \vec{f}_t(q) = \max_{p' \in \Omega_s} \{\vec{f}_s(p')^{\mathrm{T}} \vec{f}_t(q)\}$$



Homography Estimation

source

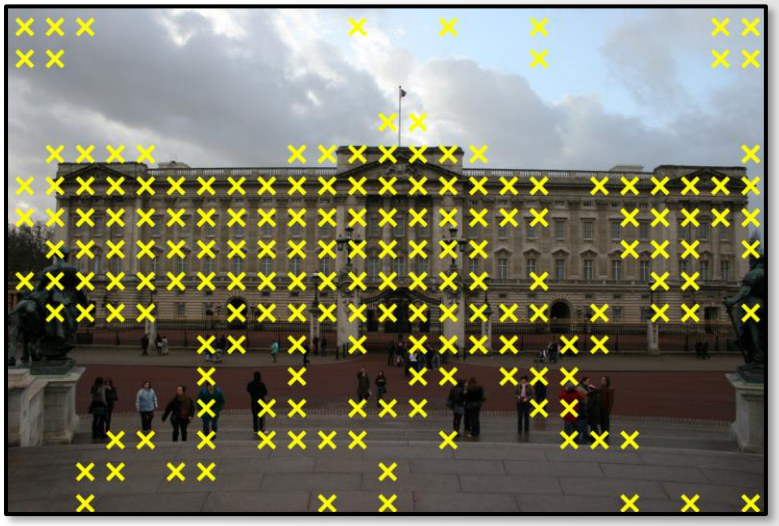
target

- **Goal**: Calculate a homographic transformation H such that

$$H(p_i) = q_i \qquad 1 \leq i \leq n$$

Homography Estimation

source

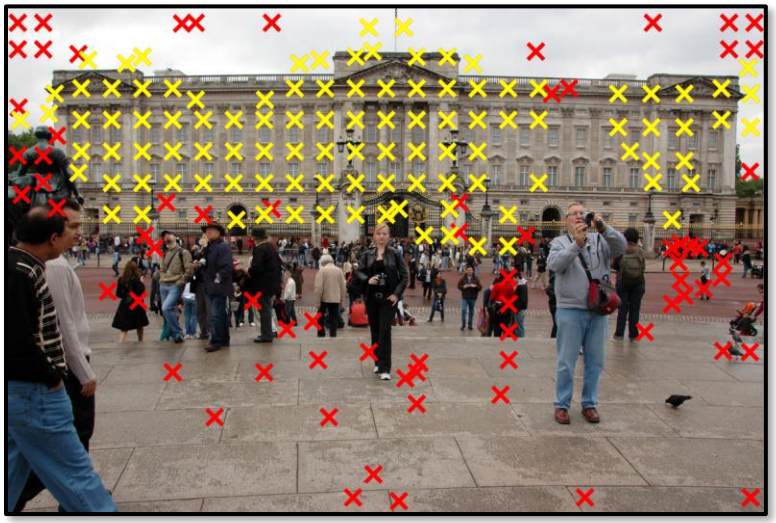


target

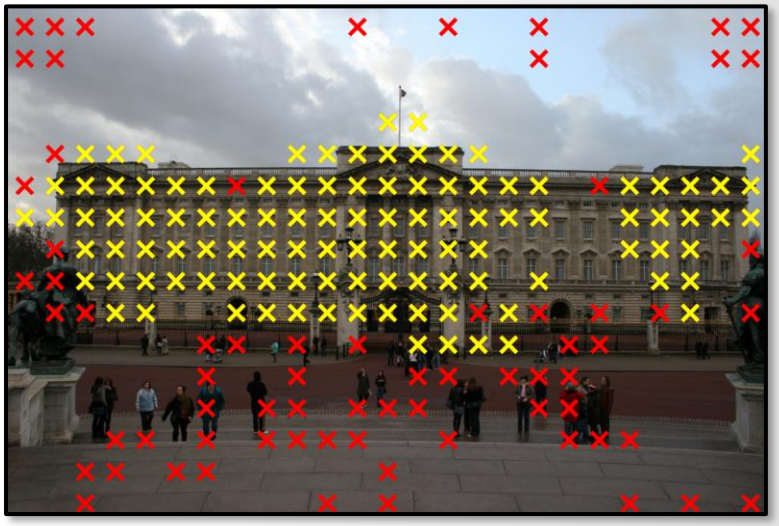
- **Goal**: Calculate a homographic transformation H such that

$$H(p_i) = q_i \qquad 1 \leq i \leq n$$

Homography Estimation



source



target

- **Goal**: Calculate a homographic transformation H such that

$$H(p_i) = q_i \qquad 1 \leq i \leq n$$

- **Problem**: set of correspondences $M = \{(p_i, q_i) \mid 1 \leq i \leq n\}$ contains false matches

$$\Rightarrow \text{RANSAC algorithm}$$

RANSAC (**RAN**dom **SA**mple **C**onsensus)

Algorithm:

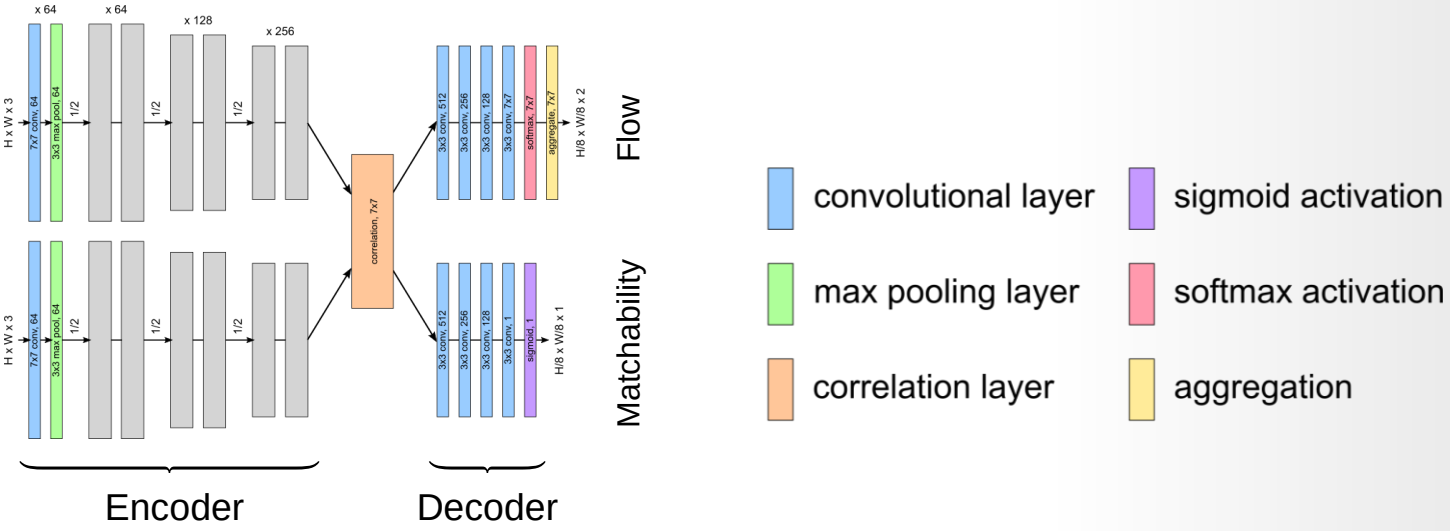
1. Draw 4 random samples (p_j, q_j) from the set of matches $M = \{(p_i, q_i) \mid 1 \leq i \leq n\}$
2. Estimate parameters θ of homography matrix H_θ such that $H_\theta(p_j) = q_j$ for all $1 \leq j \leq 4$
3. Compute the number of inliers (consensus set) $|C| := |\{ (p_i, q_i) \in M \mid \|H_\theta(p_i) - q_i\|_2 < \epsilon \}|$

➢ Repeat steps 1. – 3. for k iterations and return H_θ with maximum $|C|$

Approach

2. Fine Alignment
   (Optical Flow Estimation)

Encoder-Decoder Architecture



- Siamese Encoder based on ResNet-18
- Correlation Layer computing cosine similarities between encoded feature maps
- Two separate Decoder Streams for:
  - Optical Flow: $\overrightarrow{w}_{s\to t},\ \overrightarrow{w}_{t\to s}$
  - Matchability (Confidence): $m_{s\to t},\ m_{t\to s} \in [0,1]$

Approach

3. Training Procedure

Training Procedure

- Extract deep features and precompute homographic transformations (coarse alignment)
- Learn optical flow from unlabeled training data using a combined loss function (fine alignment):

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda \mathcal{L}_m + \mu \mathcal{L}_c$$

Recon-struction Loss

Matchability Loss

Cycle Consistency Loss

- (Optional) fine-tuning on the test dataset

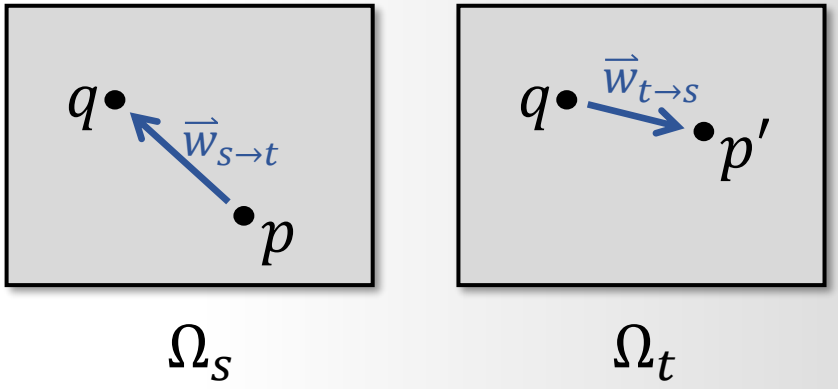
Combined Loss Function

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda \mathcal{L}_m + \mu \mathcal{L}_c$$

$$\vec{w}_{s\to t}(p) = q$$
$$\vec{w}_{t\to s}(q) = p'$$

- Cycle Consistency Loss:

$$\mathcal{L}_c = \sum_{q \in \Omega_t} m(q) \cdot \| p - p' \|_2$$



Ω_s Ω_t

- Reconstruction Loss:

$$\mathcal{L}_{rec} = \sum_{q \in \Omega_t} m(q) \cdot (1 - SSIM(p, q)) \qquad \text{with } SSIM(p, q) \in [-1, 1]$$

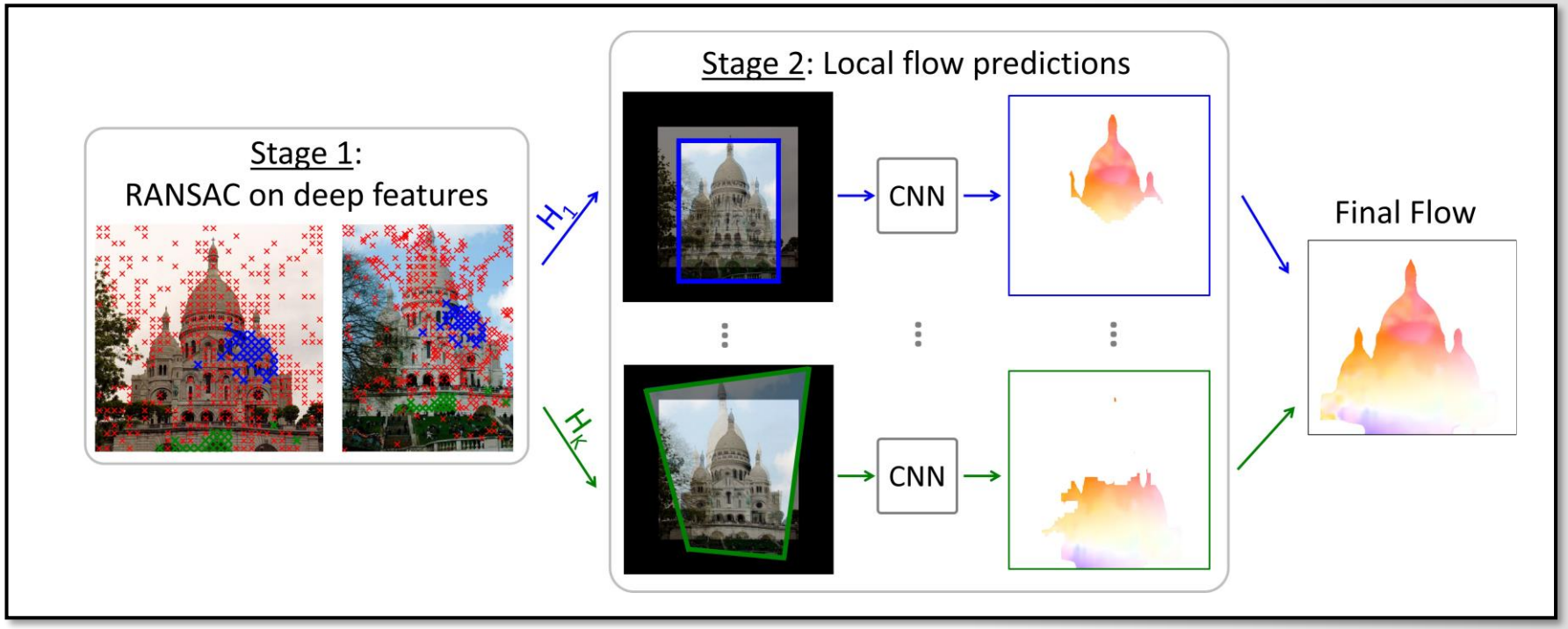
- Matchability Loss:

$$\mathcal{L}_m = \sum_{q \in \Omega_t} | m(q) - 1 | \qquad \text{with } m(q) = m_{t\to s}(q) \cdot m_{s\to t}(p)$$
$$m(q) \in [-1, 1]$$

Approach

4. Inference

Iterative Inference Process



1. Fit homography and estimate optical flow
2. Repeat procedure for pixels q_i with low confidence $m(q_i)$
3. Aggregate final flow predictions

Experiments

1. Quantitative & Qualitative Results

Optical Flow Estimation

KITTI 2015

source & target

ground truth

prediction

HPatches

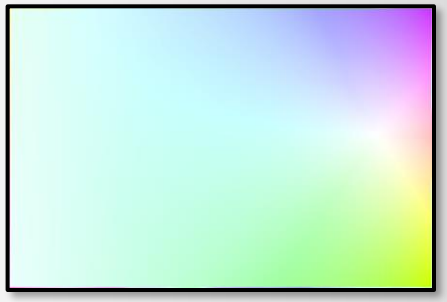
source

target

coarse alignment

fine alignment

coarse flow



fine flow

Optical Flow Estimation

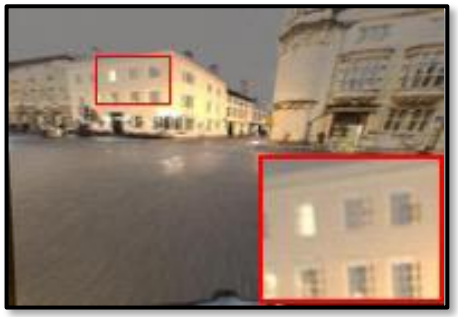
Method	KITTI 2015 (AAE ↓)		HPatches Viewpoint (AAE ↓)				
	noc	all	1	2	3	4	5
FlowNet2	4.93	10.06	5.99	15.55	17.09	22.13	30.68
PWC-Net	-	10.35	4.43	11.44	15.47	20.17	28.30
ImageNet + H	13.49	17.26	1.33	3.34	3.71	6.04	10.07
MoCo + H	13.86	17.60	1.47	2.96	3.43	7.73	10.53
DSTFlow	6.96	16.79	-	-	-	-	-
EpicFlow	4.45	**9.57**	-	-	-	-	-
RANSAC-Flow (MoCo)	4.15	12.63	0.52	**2.13**	4.83	5.13	6.36
RANSAC-Flow (ImageNet)	**3.87**	12.48	**0.51**	2.36	**2.91**	**4.41**	**5.12**

Sparse Correspondences

RobotCar

source

target



coarse alignment

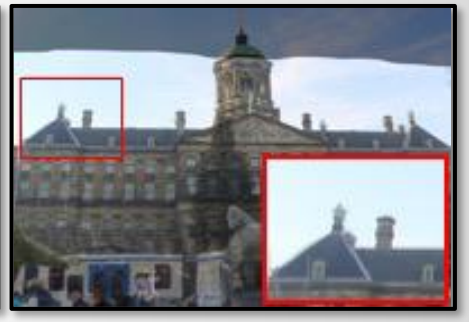
fine alignment

MegaDepth

source

target

coarse alignment



fine alignment

Sparse Correspondences

Method	RobotCar Acc(≤ d pixels ↑)			MegaDepth Acc(≤ d pixels ↑)		
	1	3	5	1	2	3
ImageNet + H	1.03	8.12	19.21	3.49	23.48	43.94
MoCo + H	1.08	8.77	20.05	3.70	25.12	45.45
SIFT-Flow	1.12	8.13	16.45	8.70	12.19	13.30
DGC-Net	1.19	9.35	20.17	3.55	20.33	34.28
Glu-Net	**2.16**	**16.77**	**33.38**	25.20	51.00	56.80
RANSAC-Flow (MoCo)	2.10	16.07	31.66	**53.47**	**83.45**	**86.81**
RANSAC-Flow (ImageNet)	2.10	16.09	31.80	53.15	83.34	86.74

Ablation Studies

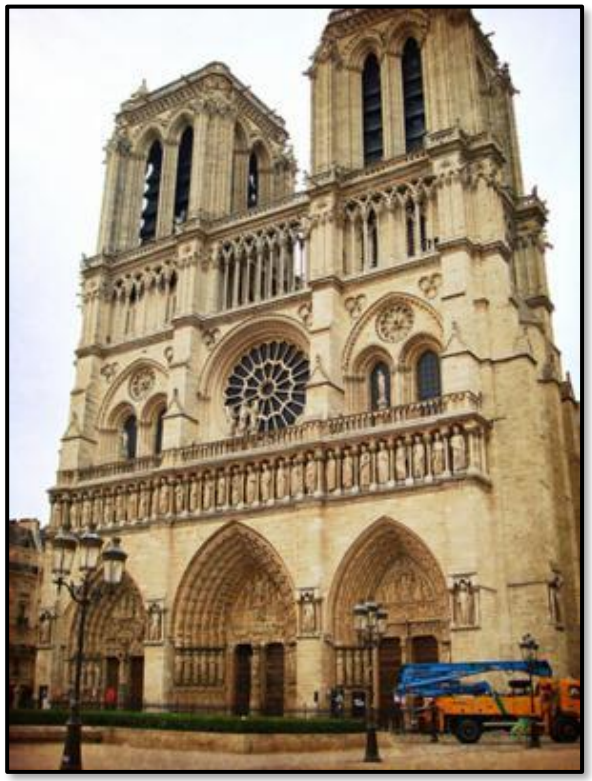
	KITTI 2015 (AAE ↓)		HPatches Viewpoint (AAE ↓)				
	noc	all	1	2	3	4	5
RANSAC-Flow (MoCo)	**4.15**	**12.63**	**0.52**	2.13	4.83	**5.13**	**6.36**
w/o fine-tuning	4.67	13.51	0.53	**2.04**	**2.32**	6.54	6.79
w/o Multi-H	7.04	14.02	-	-	-	-	-

	RobotCar Acc(≤ d pixels ↑)			MegaDepth Acc(≤ d pixels ↑)		
	1	3	5	1	2	3
RANSAC-Flow (MoCo)	**2.10**	**16.07**	**31.66**	**53.47**	83.45	**86.81**
w/o fine-tuning	2.09	15.94	31.61	52.60	**83.46**	86.80
w/o Multi-H	2.06	15.77	31.05	50.65	78.34	81.59

Experiments

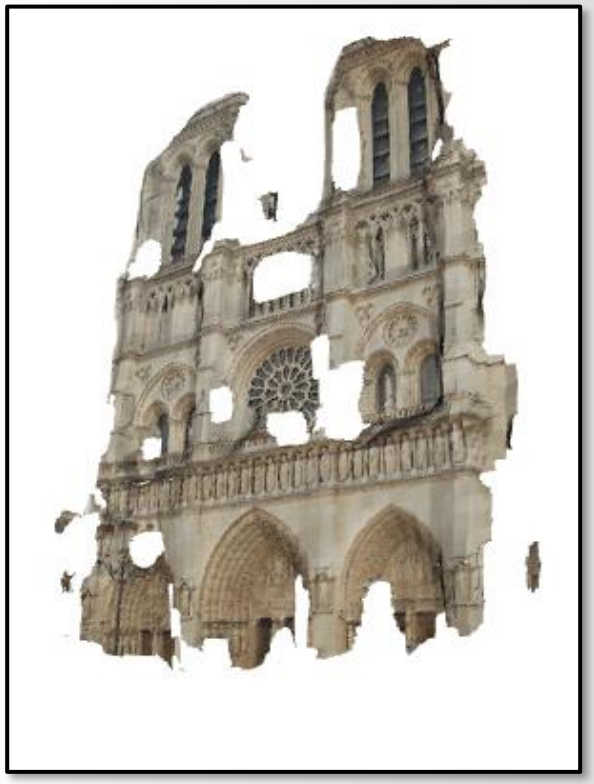
2. Downstream Applications

3D Reconstruction



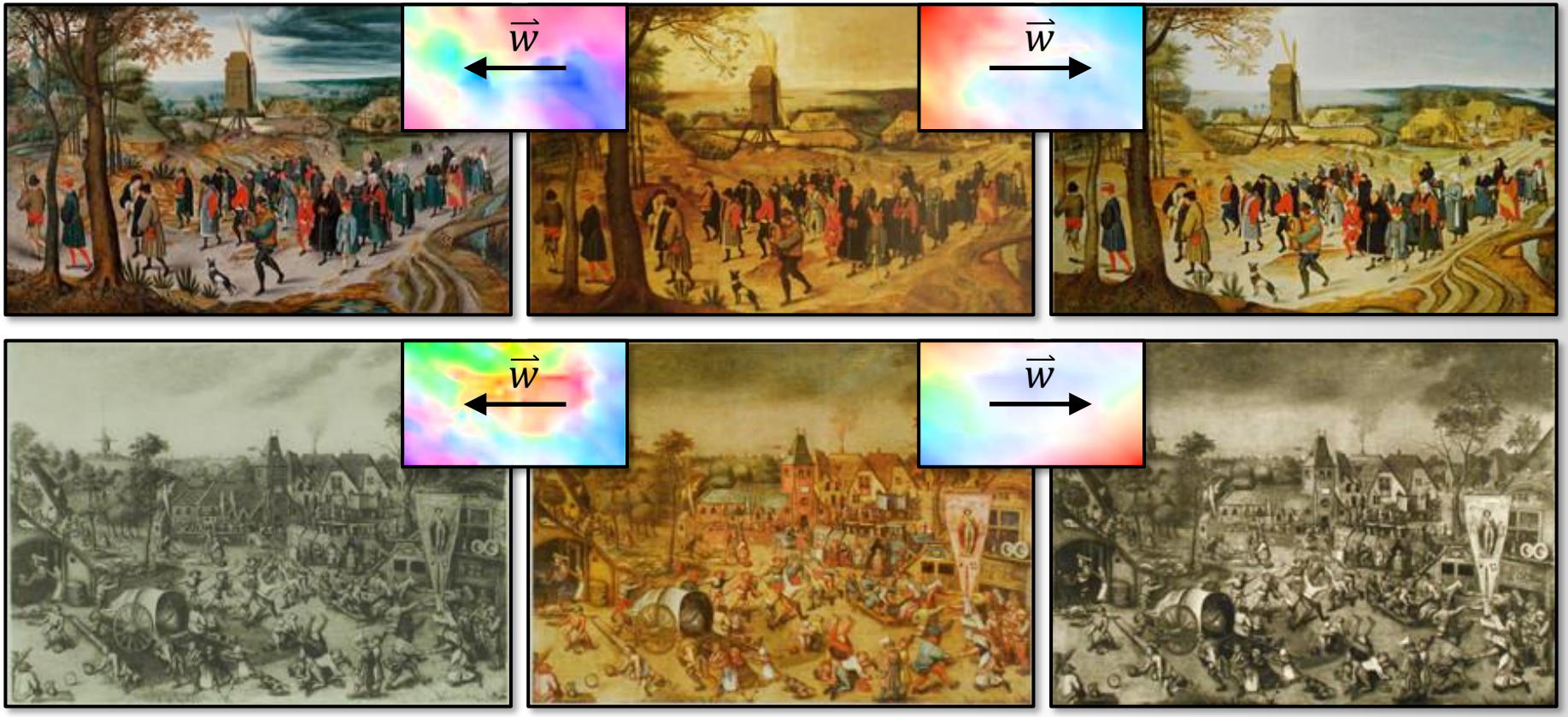
source

target



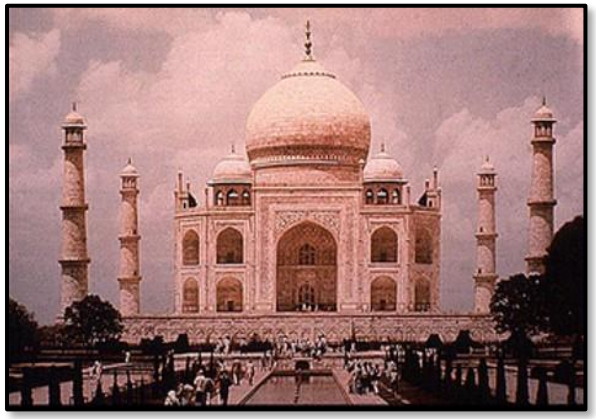
3D reconstruction

Artwork Alignment



target 1 source target 2

Texture Transfer



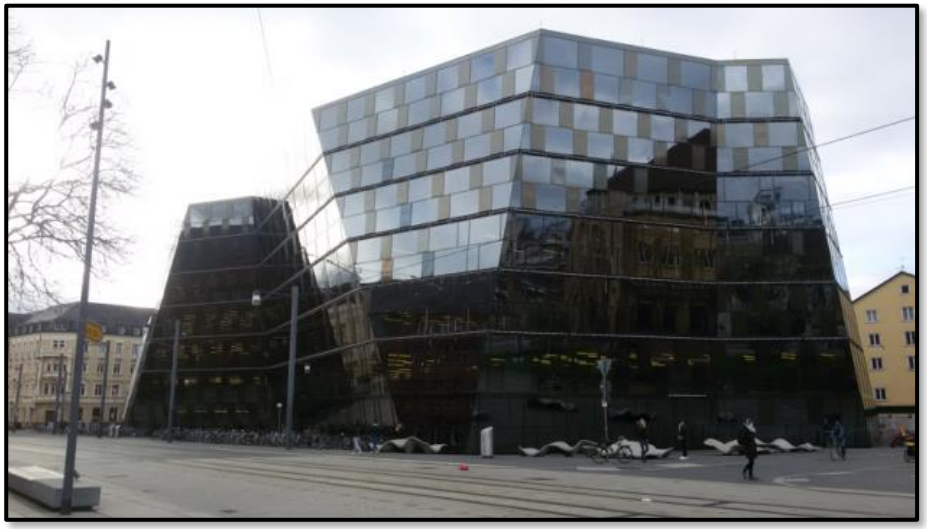



source | target | texture transfer

Experiments

3. Demo

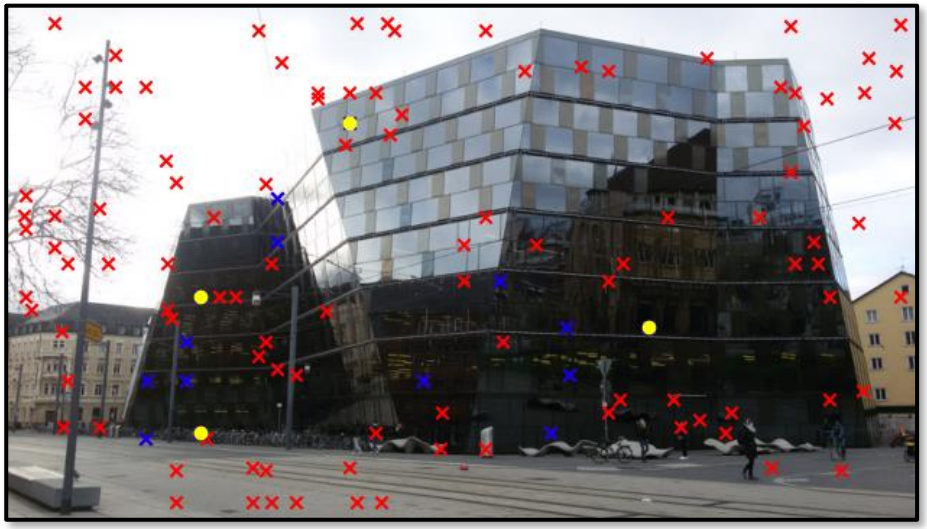
Demo



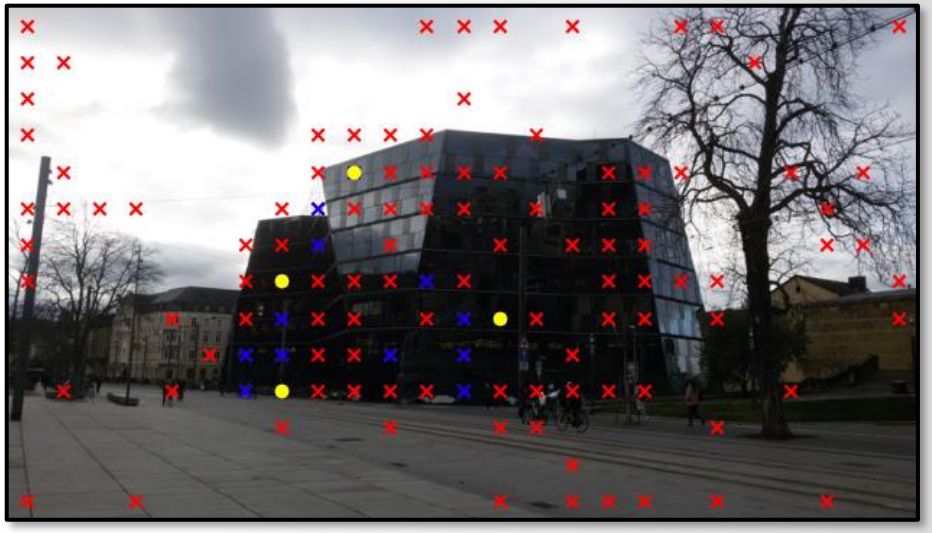
source

target

Demo

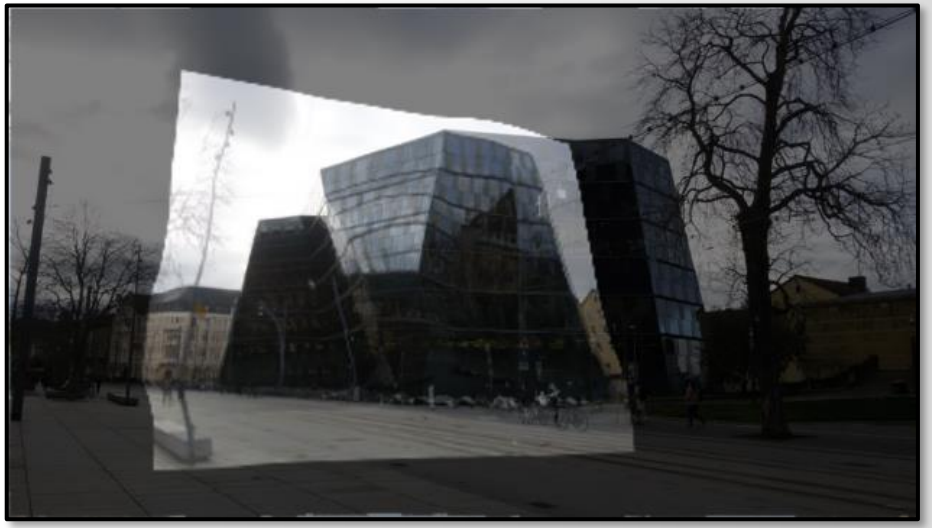


source



target

coarse alignment



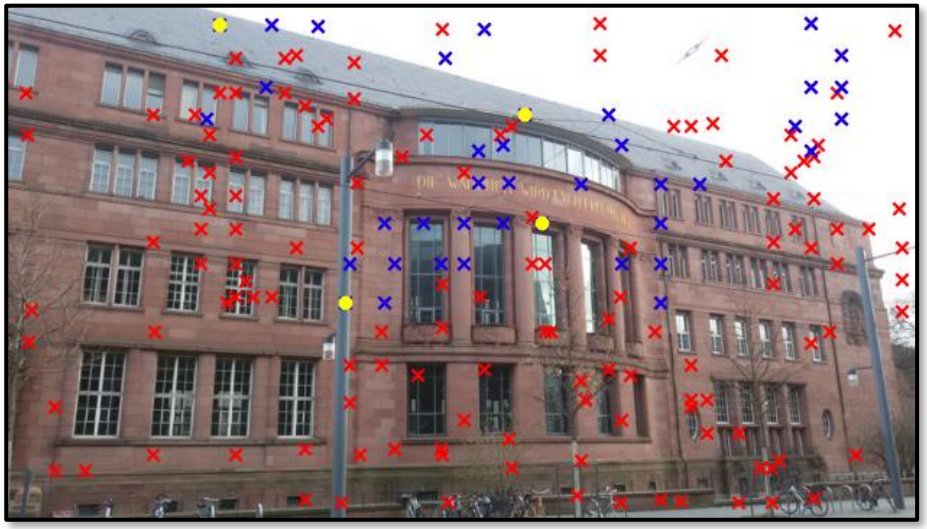
fine alignment

Demo

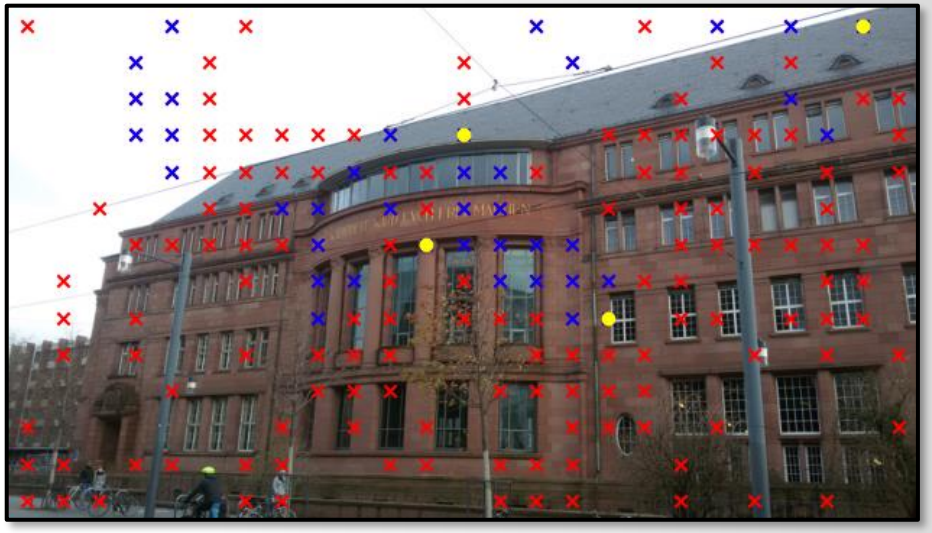
source

target

Demo



source



target

coarse alignment

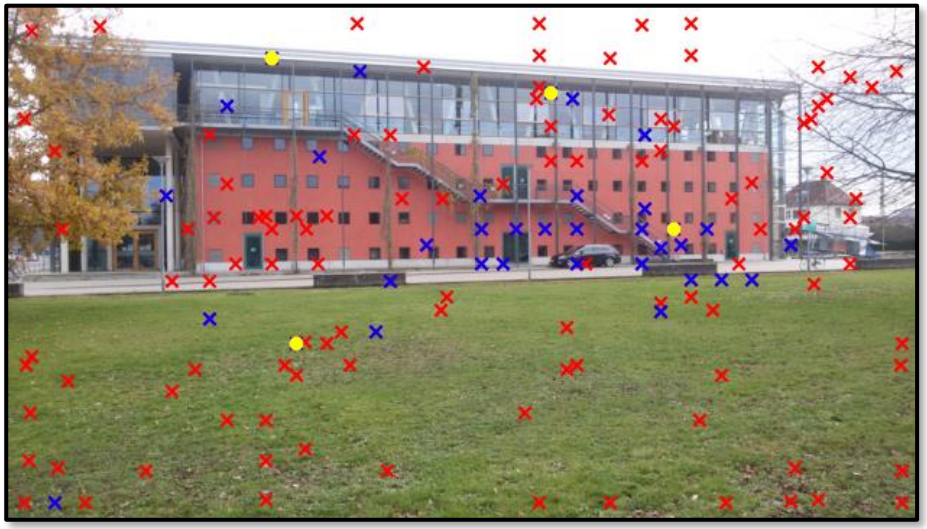
fine alignment

Demo

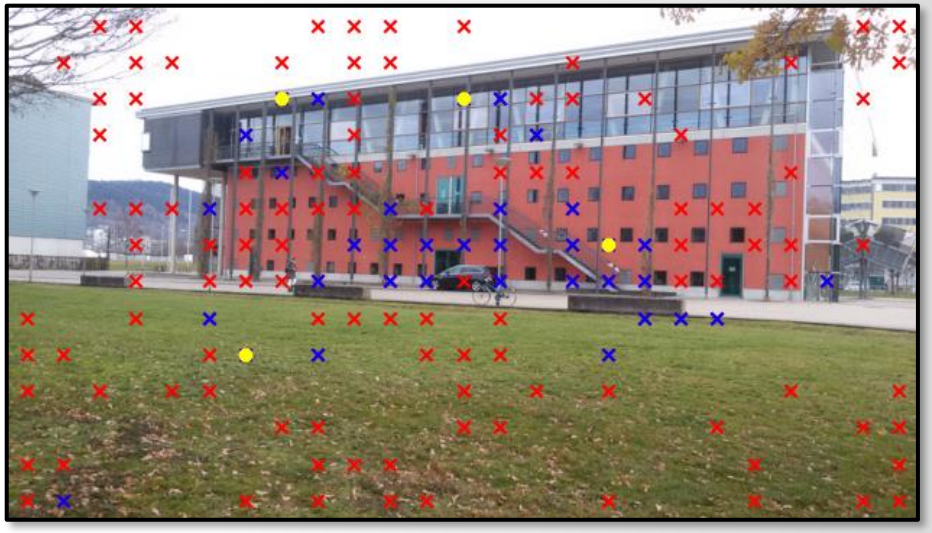
source

target

Demo



source



target

coarse alignment

fine alignment

Conclusion

Hybrid Two-Stage Image Alignment

- Combines the benefits of parametric and non-parametric methods
- Robust and precise correspondence estimation
- Unsupervised training procedure
- Various applications (3D reconstruction, artwork alignment, texture transfer, etc.)

References

[1] Fischler, M. A. & Bolles, R. C. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography" *Communications of the ACM* **24**, 381–395 (1981).

[2] He, K., Fan, H., Wu, Y., Xie, S. & Girshick, R. "Momentum Contrast for Unsupervised Visual Representation Learning" in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 9726–9735 (2020).

[3] He, K., Zhang, X., Ren, S. & Sun, J. "Deep Residual Learning for Image Recognition" in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (2016).

[4] Raguram, R., Chum, O., Pollefeys, M., Matas, J. & Frahm, J.-M. "USAC: A Universal Framework for Random Sample Consensus" *IEEE Transaction on Pattern Analysis and Machine Intelligence* **35**, 2022–2038 (2013).

[5] Shen, X., Darmon, F., Efros, A. A. & Aubry, M. "RANSAC-Flow: Generic Two-Stage Image Alignment" in *Computer Vision – ECCV 2020* (eds. Vedaldi, A., Bischof, H., Brox, T. & Frahm, J.-M.) 618–637 (2020).

[6] Shen, X., Efros, A. A. & Aubry, M. "Discovering Visual Patterns in Art Collections With Spatially-Consistent Feature Learning" in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 9270–9279 (2019).

[7] Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. "Image Quality Assessment: From Error Visibility to Structural Similarity" *IEEE Transactions on Image Process.* **13**, 600–612 (2004).