Structured Epipolar Matcher for Local Feature Matching

Jiahao Chang*, Jiahuan Yu*, Tianzhu Zhang†

University of Science and Technology of China

* Equal contribution † Corresponding author

Project Homepage:
https://sem2023.github.io

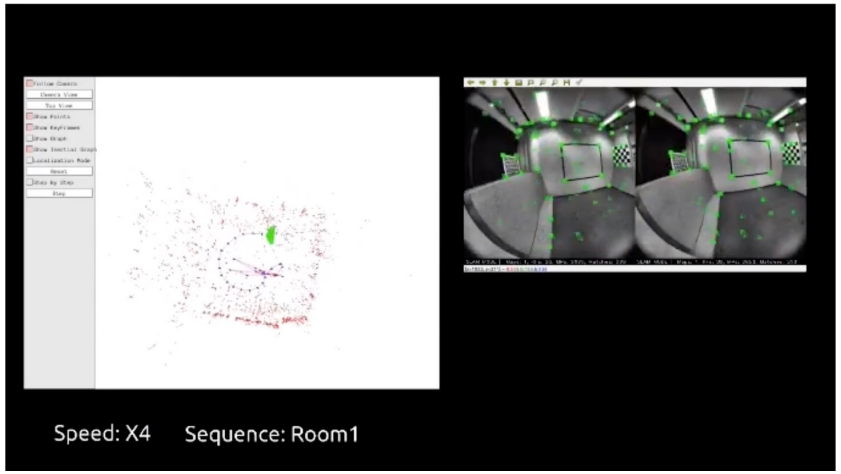
Contents

- **Introduction**
- Motivation
- Novelty
- Evaluation
- Visualization
- Conclusion

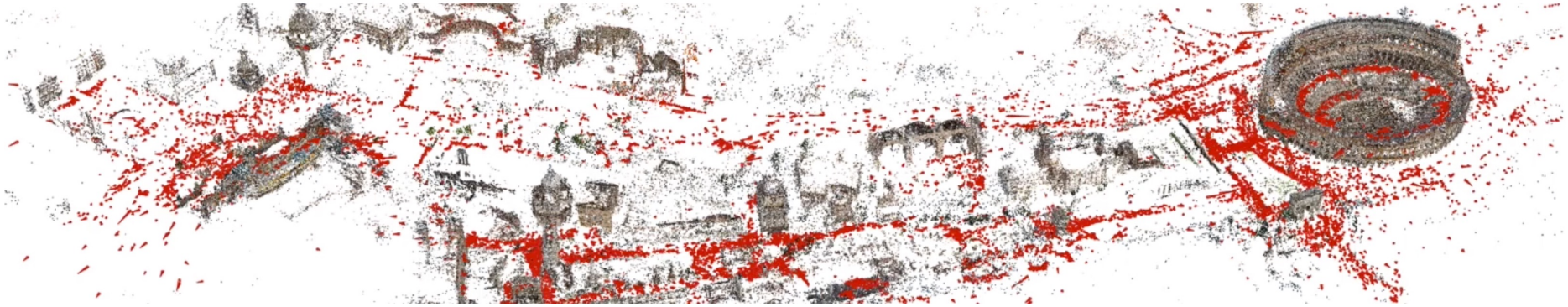
Introduction

- Local feature matching serves as a fundamental task in many 3D vision tasks

Visual Localization



SLAM

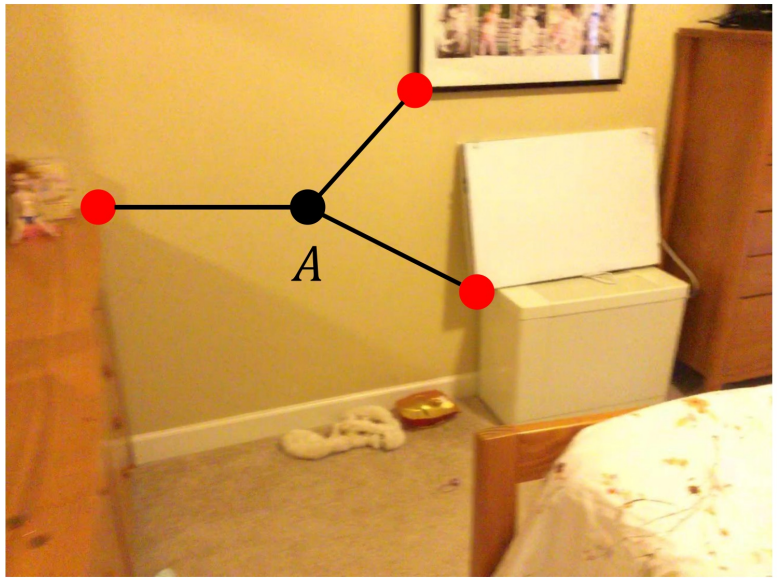


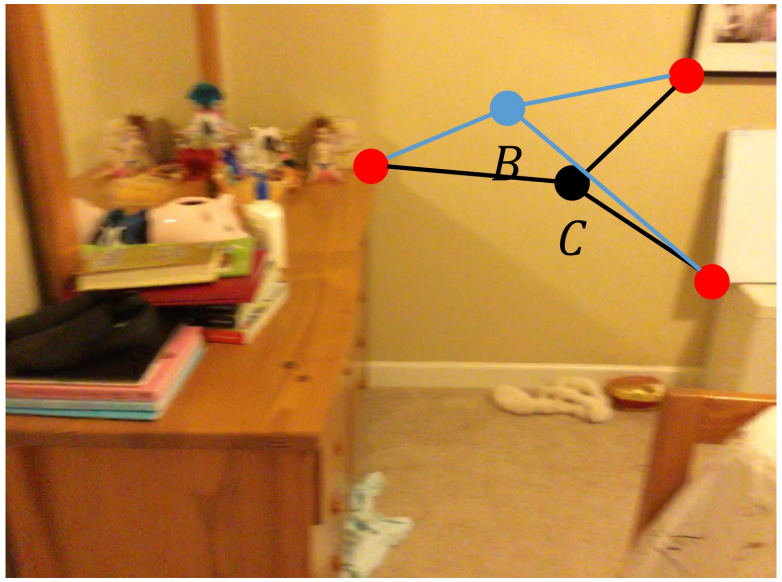
Structure from Motion (SfM)

Contents

Motivation

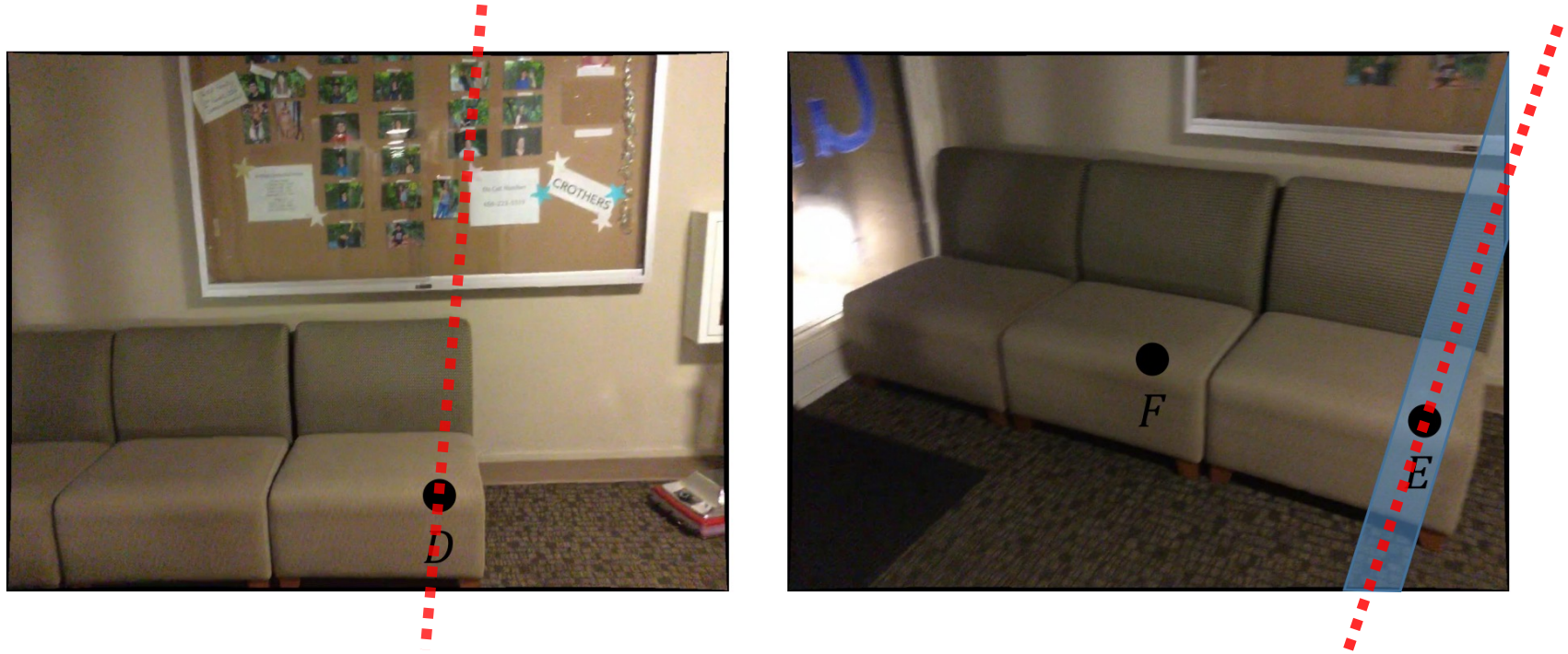
- **Appearance feature** is not distinguishable enough
  - Anchor points with rich texture can be easily matched with appearance feature
  - But points with **poor texture** (**A**, **B**, **C**) are similar in appearance feature
- **Structured feature** is ignored
  - **Relative position** to anchor points can help to find correct matching (**A, C**) instead of **B**





Motivation

- **Irrelevant regions** is not properly filtered:
  - Mainstream methods: **all-pixel-to-all-pixel** attention and matching are applied, accuracy is influenced by irrelevant regions
  - **Geometric prior** is ignored: according to **epipolar constraint** (epipolar lines marked as red), correct correspondence of **D** must rely in blue region, another similar point **F** can be filtered

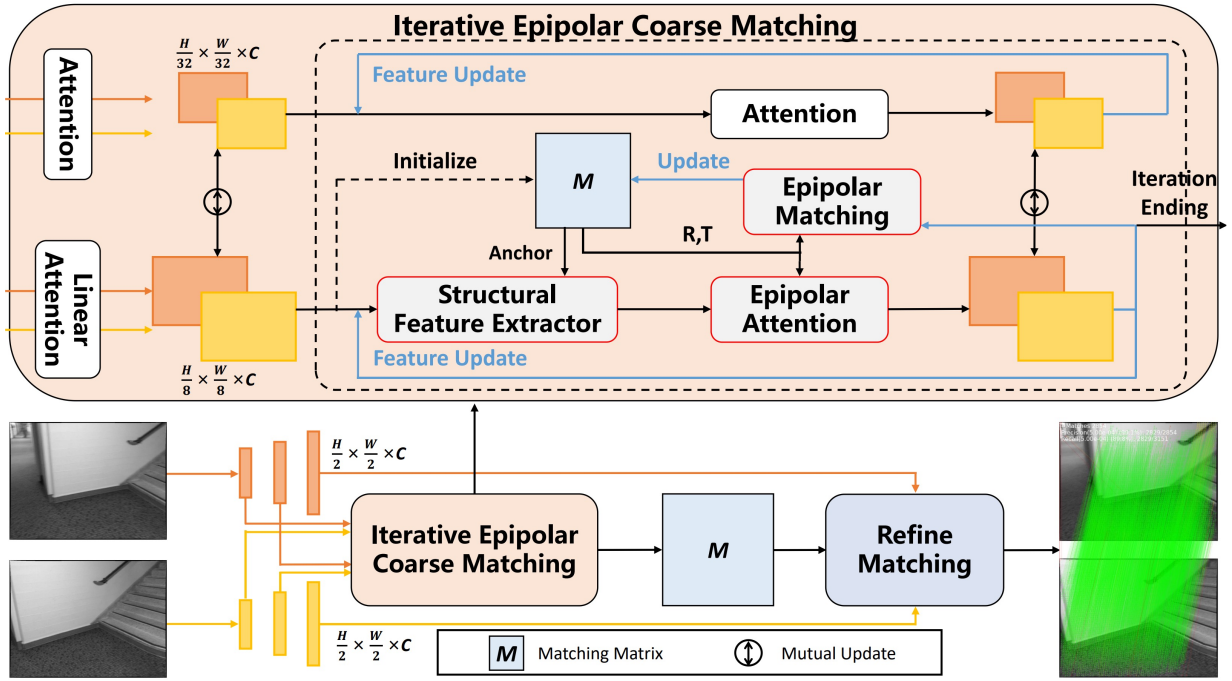


Contents

- Introduction
- Motivation
- **Novelty**
- Evaluation
- Visualization
- Conclusion

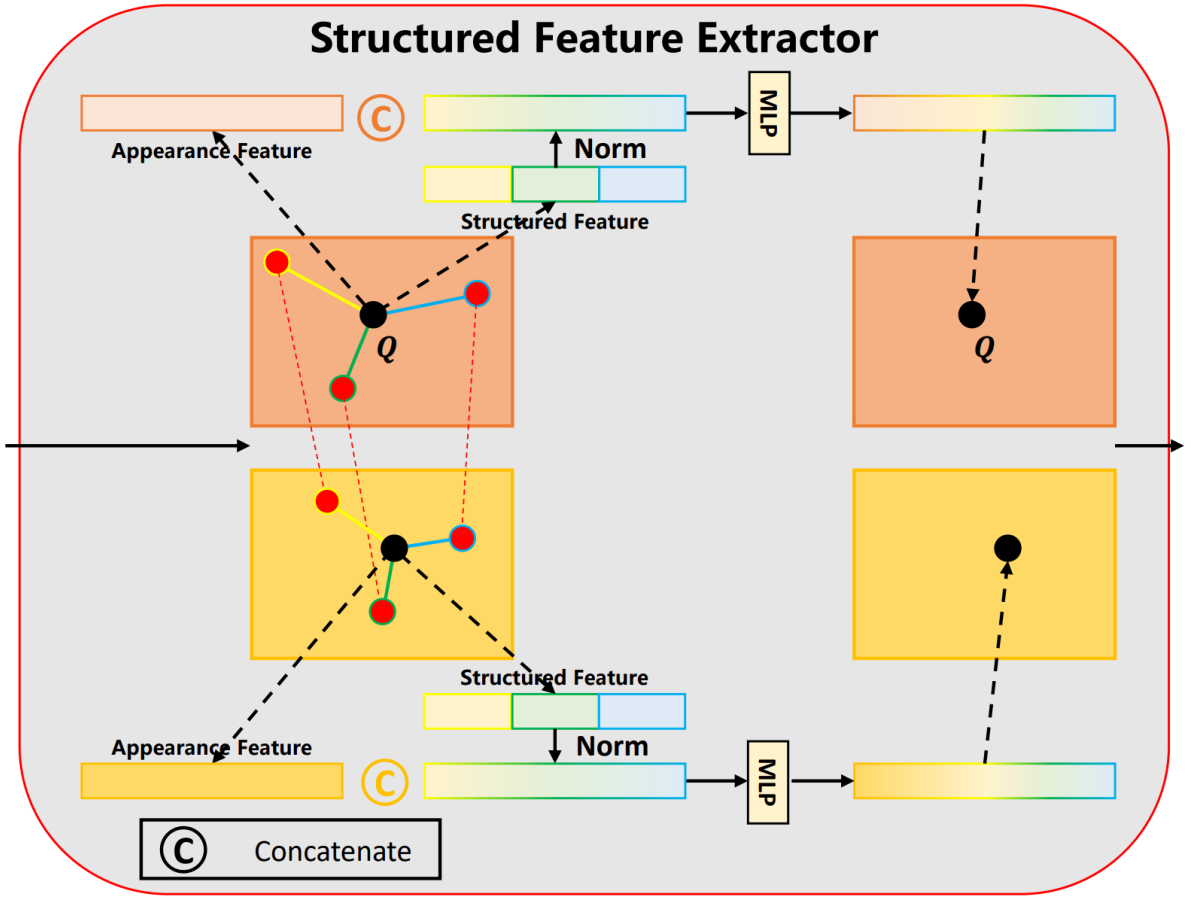
Novelty

- A unified coarse-to-fine **iterative** architecture named Structured Epipolar Matcher (**SEM**) taking structured feature and geometric prior into consideration
  - Iterative Epipolar Coarse Matching: iteratively update coarse-stage feature with **Structured Feature Extractor** and **Epipolar Attention/Matching**
  - **Structured Feature Extractor**: extract structured feature and fuse into appearance feature
  - **Epipolar Attention/Matching**: apply attention/matching with epipolar constraint



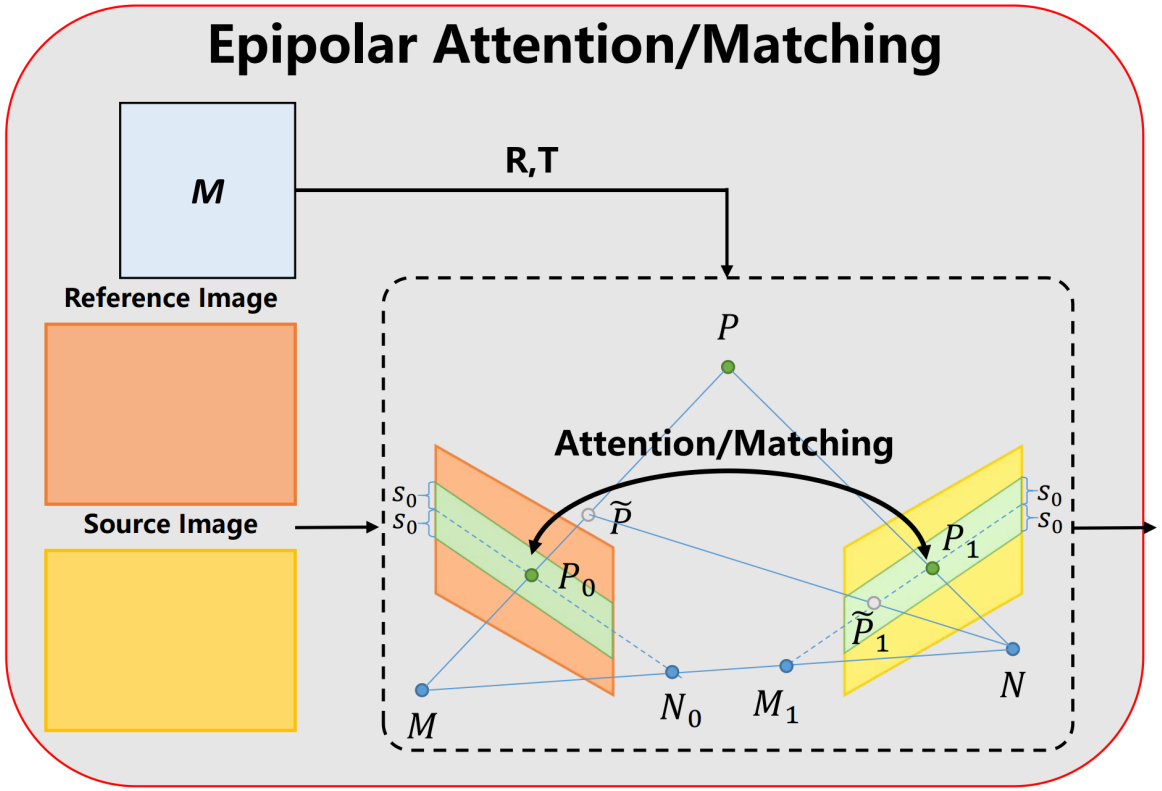
Novelty -- Structured Feature Extractor

- Anchor points (x_i, y_i): high confidence correspondences
- For each point **Q** (x, y), calculate structured feature:
  - Coordinate difference:
    - $\boldsymbol{\Delta X} = (x - x_1, x - x_2, \dots, x - x_N)$
    - $\boldsymbol{\Delta Y} = (y - y_1, y - y_2, \dots, y - y_N)$
  - Euclidean distances:
    - $\boldsymbol{D} = \sqrt{(\boldsymbol{\Delta X})^2 + (\boldsymbol{\Delta Y})^2}$
- Structured feature:
  - $\boldsymbol{F}^{sf} = \mathrm{norm}(\boldsymbol{\Delta X}) \parallel \mathrm{norm}(\boldsymbol{\Delta Y}) \parallel \mathrm{norm}(\boldsymbol{D})$
  - norm for **scaling invariance**
  - $\boldsymbol{D}$ for **rotational invariance**
- Fusing with appearance feature $\boldsymbol{F}^{af}$:
  - $\boldsymbol{F} = \mathrm{MLP}(\boldsymbol{F}^{sf} \parallel \boldsymbol{F}^{af})$



Novelty -- Epipolar Attention/Matching

- Take **epipolar constraint** into attention and matching in an **iterative** manner
- **Relative position** R, T obtained from previous matching matrix M by RANSAC
- Calculate epipolar lines P_0N_0 and P_1N_1 from R, T by epipolar geometry
- Broaden lines to $2s_0$ width regions for **error tolerance**
- Attention and matching are applied between P_0 and corresponding region, features and matching matrix are **updated**
- **Irrelevant areas are filtered**



Contents

- Introduction
- Motivation
- Novelty
- **Evaluation**
- Visualization
- Conclusion

Evaluation

- Homography Estimation (HPatches)

Category	Method	Homography est. AUC @3px	Homography est. AUC @5px	Homography est. AUC @10px	matches
Detector-based	D2Net [9]+NN	23.2	35.9	53.6	0.2K
	R2D2 [28]+NN	50.6	63.9	76.8	0.5K
	DISK [45]+NN	52.3	64.9	78.9	1.1K
	SP [8]+SuperGlue [32]	53.9	68.3	81.7	0.6K
	Patch2Pix [50]	46.4	59.2	73.1	1.0K
Detector-free	Sparse-NCNet [29]	48.9	54.2	67.1	1.0K
	COTR [16]	41.9	57.7	74.0	1.0K
	DRC-Net [19]	50.6	56.2	68.3	1.0K
	LoFTR [35]	65.9	75.6	84.6	1.0K
	PDC-Net+ [44]	66.7	76.8	85.8	1.0k
	SEM(ours)	**69.6**	**79.0**	**87.1**	1.0K

- Visual Localization (InLoc)

Method	DUC1 $(0.25m, 10^\circ)$ / $(0.5m, 10^\circ)$ / $(1m, 10^\circ)$	DUC2 $(0.25m, 10^\circ)$ / $(0.5m, 10^\circ)$ / $(1m, 10^\circ)$
LoFTR [35]	47.5 / 72.2 / 84.8	54.2 / 74.8 / **85.5**
MatchFormer [46]	46.5 / 73.2 / 85.9	**55.7** / 71.8 / 81.7
ASpanFormer [5]	51.5 / 73.7 / 86.4	55.0 / 74.0 / 81.7
SEM(ours)	**52.0 / 74.2 / 87.4**	50.4 / **76.3** / 83.2

- Relative Pose Estimation (MegaDepth & ScanNet)

MegaDepth

Category	Method	Pose estimation AUC @5°	Pose estimation AUC @10°	Pose estimation AUC @20°
Detector-based	SP [8]+SuperGlue [32]	42.2	59.0	73.6
	SP [8]+SGMNet [4]	40.5	59.0	73.6
Detector-free	DRC-Net [19]	27.0	42.9	58.3
	PDC-Net+(H) [44]	43.1	61.9	76.1
	LoFTR [35]	52.8	69.2	81.2
	MatchFormer [46]	53.3	69.7	81.8
	QuadTree [39]	54.6	70.5	82.2
	ASpanFormer [5]	55.3	71.5	83.1
	SEM(ours)	**58.0**	**72.9**	**83.7**

ScanNet (* train on MegaDepth)

Category	Method	Pose estimation AUC @5°	Pose estimation AUC @10°	Pose estimation AUC @20°
Detector-based	D2-Net [9]+NN	5.3	14.5	28.0
	SP [8]+OANet [49]	11.8	26.9	43.9
	SP [8]+SuperGlue [32]	16.2	33.8	51.8
Detector-free	DRC-Net [19]*	7.7	17.9	30.5
	MatchFormer [46]*	15.8	32.0	48.0
	LoFTR-OT [35]*	16.9	33.6	50.6
	SEM(ours)*	**18.7**	**36.6**	**52.9**

Evaluation

- Ablation study on MegaDepth

Proposed Module

Index	Multi-Level	SF	EAM	Pose estimation AUC @5°	@10°	@20°
1				45.6	62.2	75.3
2	✓			46.7	63.1	76.3
3	✓	✓		47.3	64.3	76.8
4	✓	✓	✓	**48.1**	**64.7**	**77.4**

SF = Structured Feature
EAM = Epipolar Attention/Matching

Different Epipolar Region half-width s_0

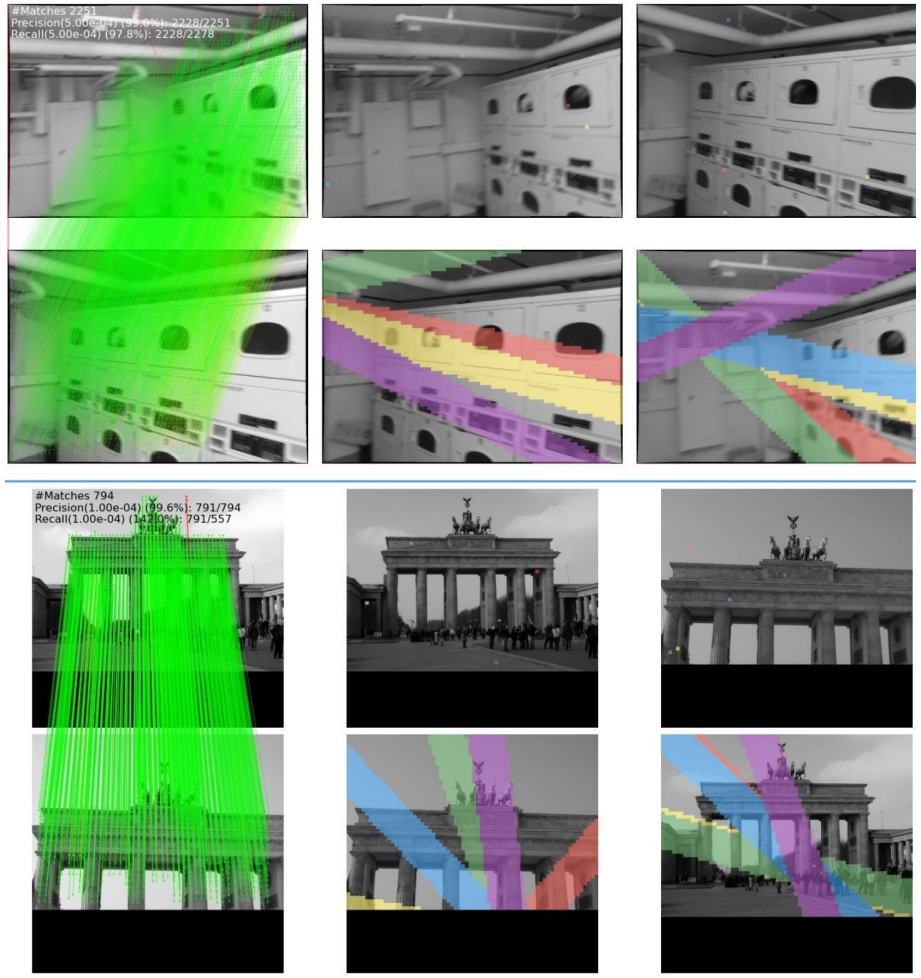
s_0	Pose estimation AUC @5°	@10°	@20°
5	45.6	62.7	76.2
10	**48.1**	**64.7**	**77.4**
15	47.5	64.3	77.2
20	46.7	62.4	76.4

Contents

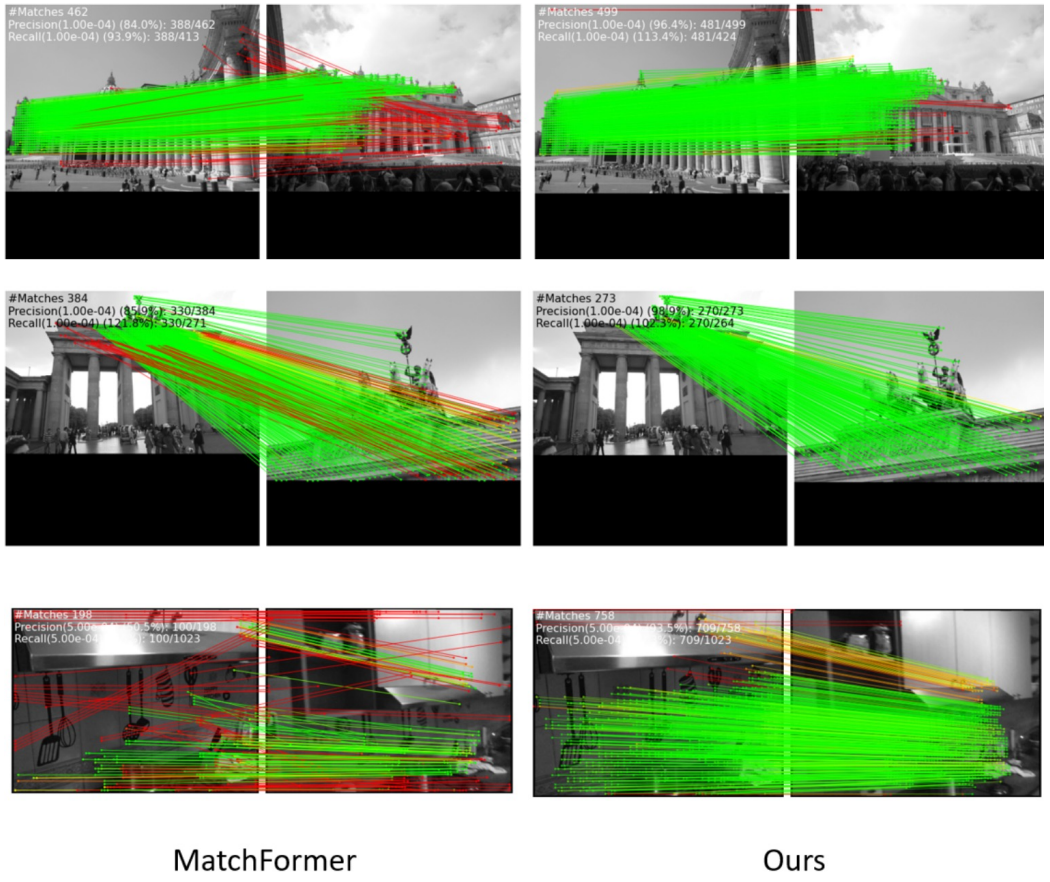
- Introduction
- Motivation
- Novelty
- Evaluation
- **Visualization**
- Conclusion

Visualization

Epipolar Banded Areas



Qualitative Comparison



MatchFormer

Ours

Contents

- Introduction
- Motivation
- Novelty
- Evaluation
- Visualization
- **Conclusion**

Conclusion

- A novel Structured Epipolar Matcher (**SEM**) for local feature matching
- Two novel module:
  - **Structured Feature Extractor:** generating and fusing structured feature to complement the appearance features
  - **Epipolar Attention/Matching:** utilizing epipolar constraint to filter out irrelevant matching regions as much as possible
- SOTA performance in extensive experimental

Thanks!

Epipolar Matcher for Local Feature Matching

Jiahuan Yu*, Jiahao Chang*, Tianzhu Zhang†

University of Science and Technology of China

* Equal contribution † Corresponding author

Project Homepage:
https://sem2023.github.io