GLAMpoints: Greedily Learned Accurate Match points

retinai ETH zürich

Prune Truong[1,2], Stefanos Apostolopoulos[1], Agata Mosinska[1], Samuel Stucky[1], Carlos Ciller[1], Sandro De Zanet[1]

[1] RetinAI Medical AG, Switzerland, [2] ETH Zurich, Switzerland

1. Motivation

- **Retinal image registration** helpful for ***diagnosis***, disease monitoring and ***treatment planning***
- In eye laser treatment, enables **real time tracking** of diabetic retinopathy laser treatment

Steps of mosaicking:

- Extract image interest points
- Compute descriptors
- Match descriptors
- Estimate transformation

➔ **Keypoint detection** is the **most important step** as it **conditions all others**

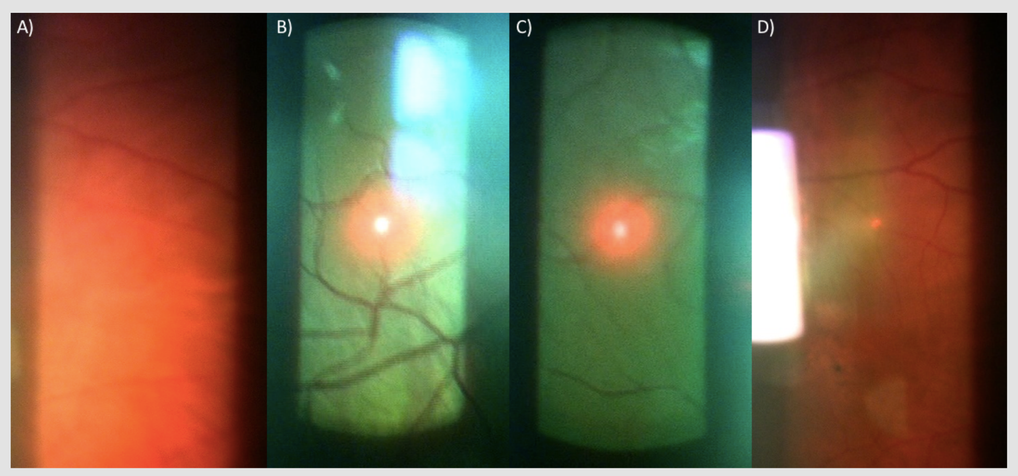
➔ **Classical detectors fail on medical data**

2. Contribution

- **Trained directly for the final matching accuracy** instead of optimized on indirect metrics such as repeatability
- **Semi supervised learning**
- Can be applied to different image domains

3. Training Dataset

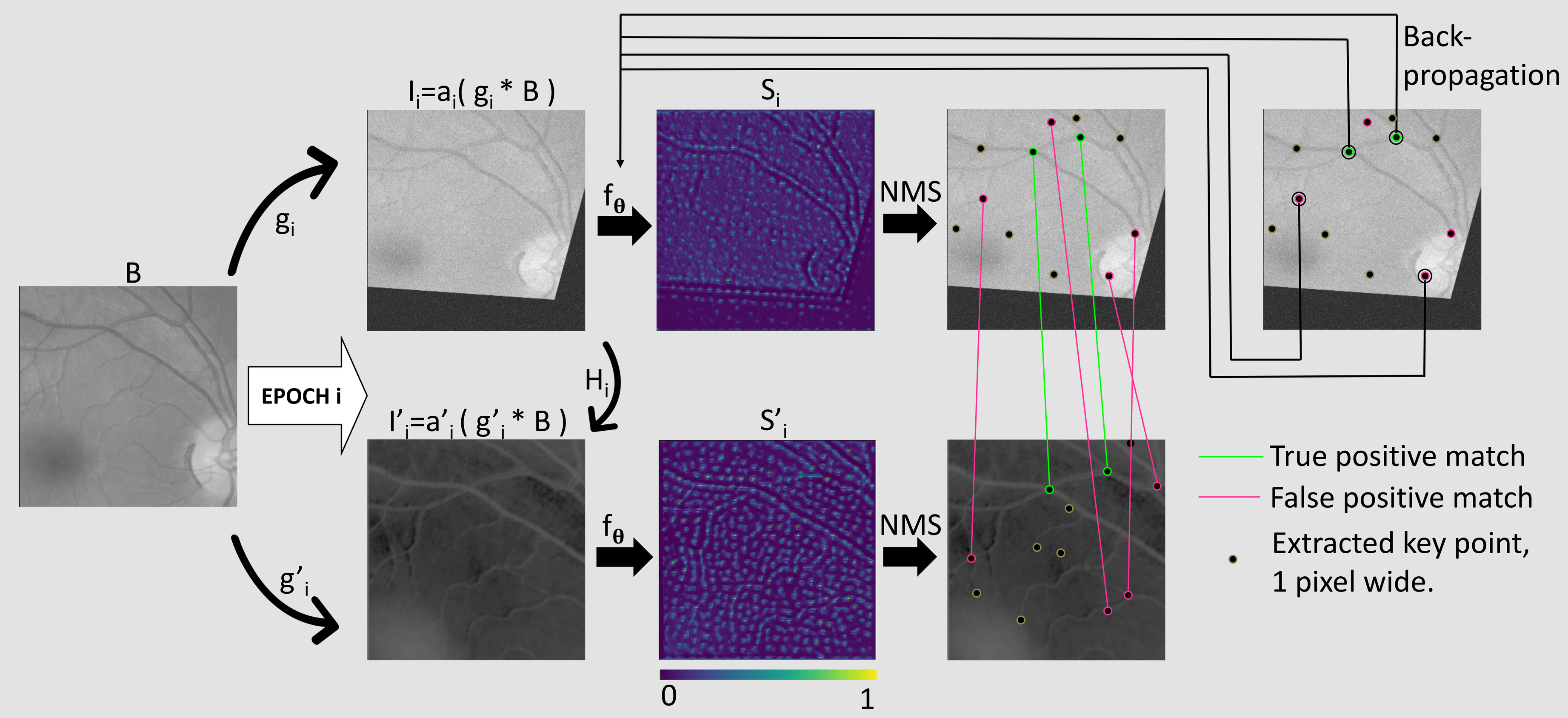
- Dataset based on **slit lamp fundus** videos used in laser treatment. 1336 base images
- **Challenging appearance**: low vascularization and over-exposure (A), motion blur (B), focus blur (C), acquisition artifacts and reflections (D)



- For training, pairs of images synthetically generated from base image by random homographies

References

[1] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. IJCV, Nov 2004.
[2] Pablo Fernandez Alcantarilla et al. KAZE Features. ECCV, 2012.
[3] Daniel DeTone et al. SuperPoint: Self-Supervised Interest Point Detection and Description. CVPR Workshops, 2018.
[4] Kwang Moo Yi et al. LIFT: Learned Invariant Feature Transform. ECCV, 2016.
[5] Yuki Ono et al. LF-Net: Learning Local Features from Images. NIPS, 2018.



4. Method

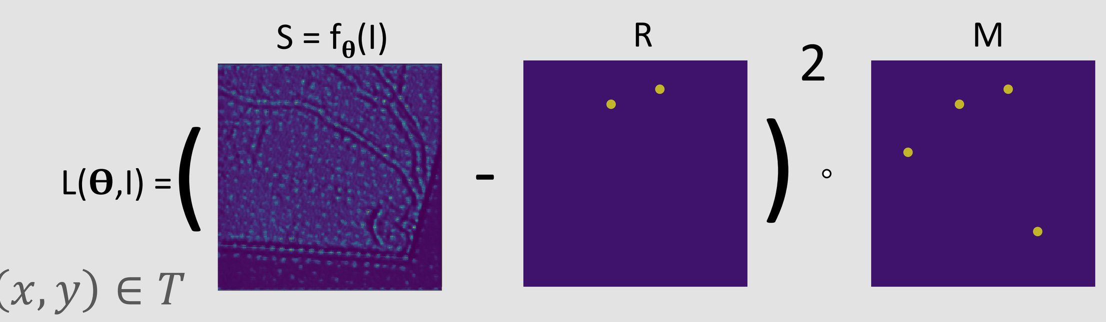
Our network predicts the **location** of stable **interest points**, called **GLAMpoints**, on a gray-scale image.
The training steps:

1. For a pair of images I_i and I'_i our model provides a score map for each image $S_i = f_\theta(I_i)$ and $S'_i = f_\theta(I'_i)$
2. Interest point locations are extracted from both score maps using NMS
3. A 128 root-SIFT feature descriptor is computed for each detected keypoint
4. Keypoint descriptors from image I_i are matched to those of image I'_i and checked according to ground truth H_i.

The set of true positive is denoted as T.

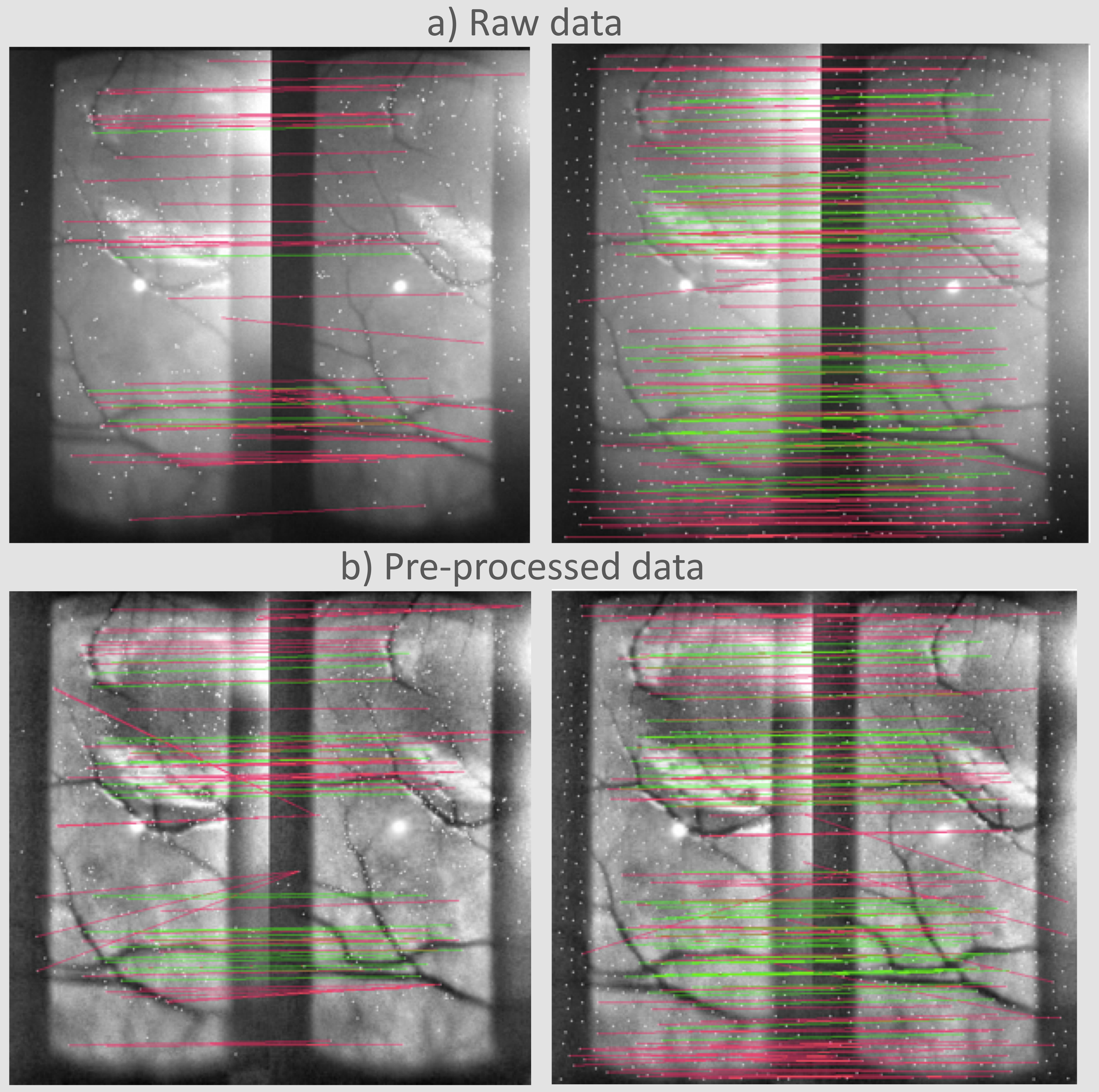
$$L(\theta, I) = \frac{\sum (f_\theta(I) - R)^2 . M}{\sum M}$$

with M a binary mask for sample mining, and $R_{x,y} = \begin{cases} 1, for\ (x,y) \in T \\ 0, otherwise \end{cases}$



5. Result on slit lamp dataset

- **Success rate** for failed (lack of keypoints or matches), acceptable (MEE < 10 and MAE < 30) or inaccurate (otherwise) registrations
- Raw and pre-processed data



SIFT GLAMpoints

White dots are detected keypoints; true positive matches are green; false positive are in red.

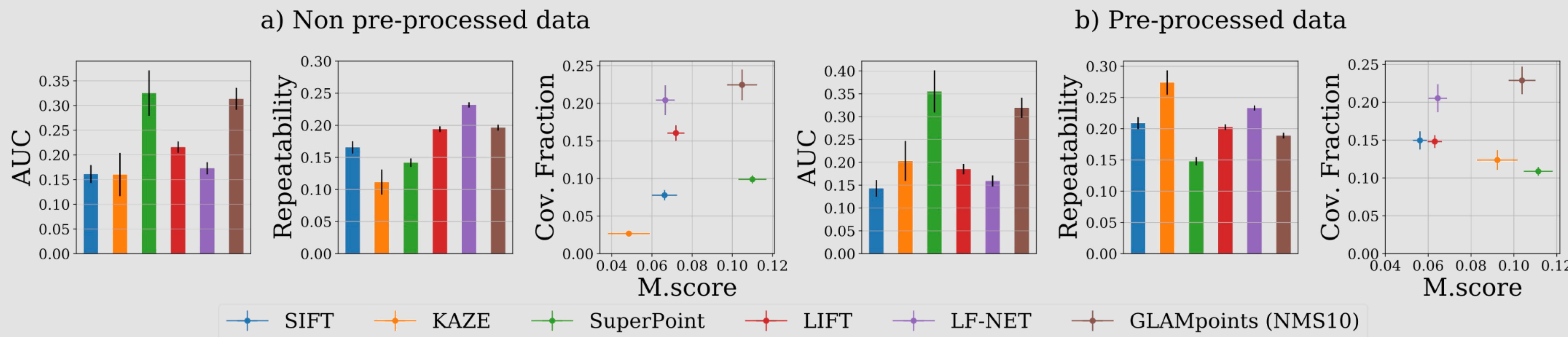
a) Raw data

Detector	Failed [%]	Inaccurate [%]	Acceptable [%]
SIFT [1]	14.56	63.11	22.33
KAZE [2]	24.27	61.65	14.08
SuperPoint [3]	17.48	48.54	33.98
LIFT [4]	0.00	43.69	56.31
LF-Net [5]	0.00	39.81	60.19
GLAMpoints (OURS)	0.00	36.41	**63.59**

b) Pre-processed data

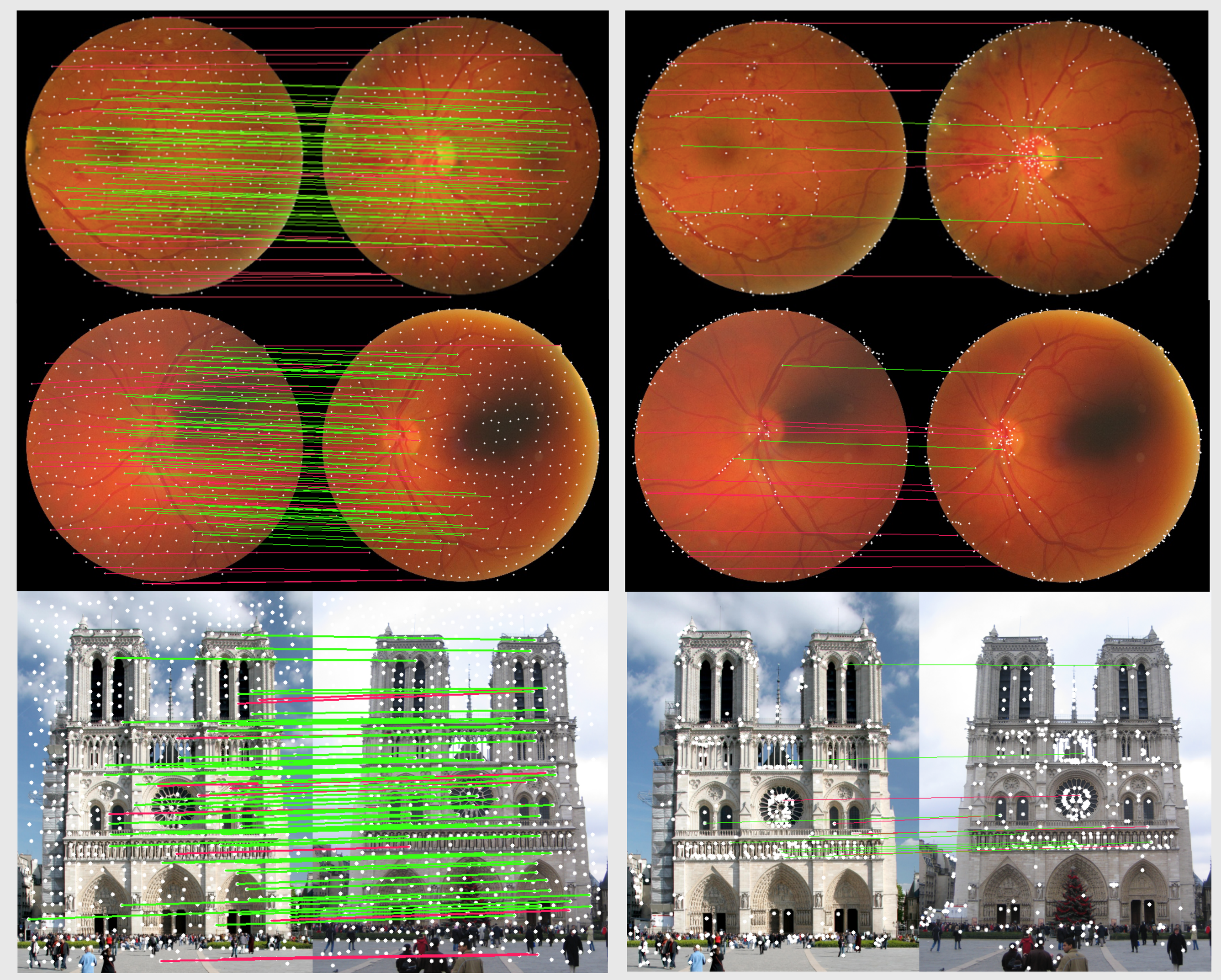
Detector	Failed [%]	Inaccurate [%]	Acceptable [%]
ORB	9.71	83.01	7.28
GLAMpoints (ORB)	0.00	88.35	**11.65**
BRISK	16.99	66.02	16.99
GLAMpoints (BRISK)	1.94	75.73	**22.33**
Random grid (SIFT)	0.00	62.62	37.38
SIFT	1.94	47.75	50.49
KAZE	1.46	54.85	43.69
KAZE (SIFT)	4.37	57.28	39.35
SuperPoint	7.77	51.46	40.78
SuperPoint (SIFT)	6.80	54.37	38.83
LIFT	0.00	39.81	60.19
LF-NET	0.00	36.89	63.11
LF-NET (SIFT)	0.00	40.29	59.71
GLAMpoints (SIFT)	0.00	31.55	**68.45**

Table: Registration success rates (%) on the slit lamp testing dataset. Descriptor in parentheses, where not original algorithm used.



6. Results on FIRE dataset and natural images

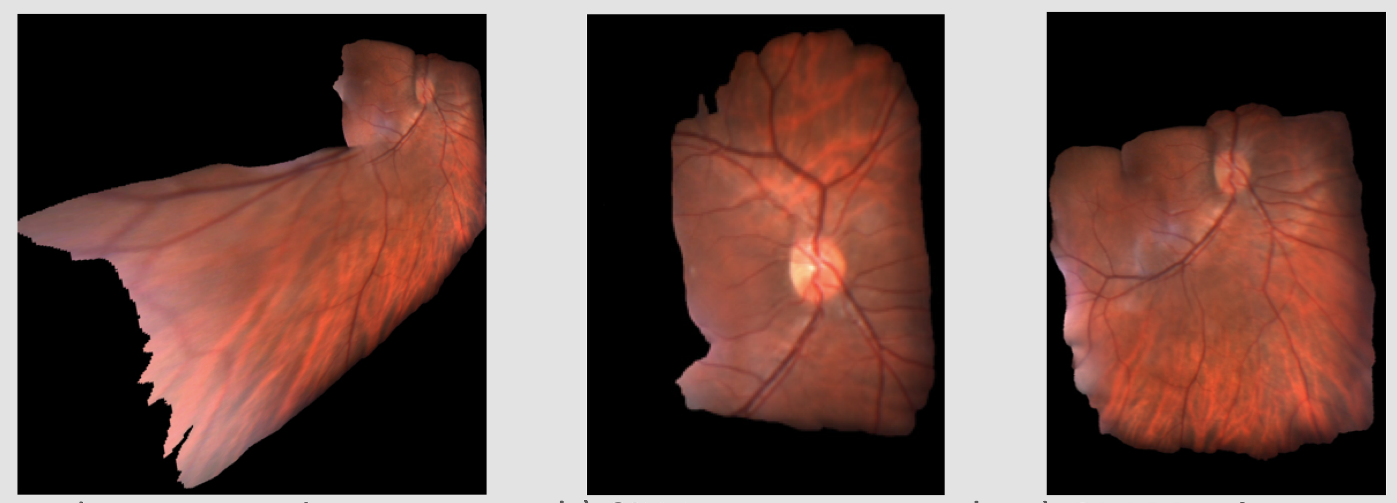
- FIRE dataset: **GLAMpoints 1st** in success rate of acceptable registration
- Natural images: GLAMpoints 3rd in success rate (75.38 % against 85.13 %)
  ➔ possible **extension of GLAMpoints** to **natural images**



a) GLAMpoints (OURS) b) SIFT

7. Image Mosaicking

- Consecutive frame registration, no bundle adjustment
- Average number of registered frames before failure :
  1.04 frame for SIFT against 9.98 for GLAMpoints



a) **SIFT**, raw data, **11 frames** b) **SIFT**, pre-processed data, **34 frames** c) **GLAMpoints**, raw data, **53 frames**