Supplementary Material

Keypoint Matching for Instrument-Free 3D Registration in Video-based Surgical Navigation

Tânia Baptista[1,2(✉)], Carolina Raposo[1,2], Miguel Marques[2], Michel Antunes[1,2], and Joao P. Barreto[1,2]

[1] Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal
tania.baptista@isr.uc.pt
[2] Perceive3D, Coimbra, Portugal

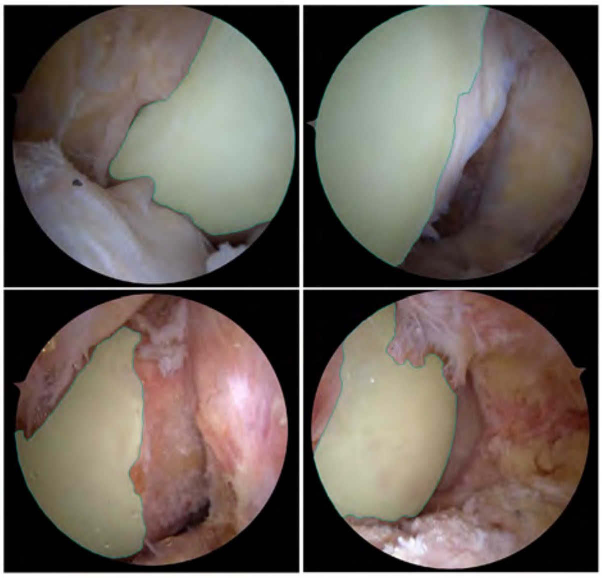

Fig. 1: Overlay of exemplary segmentation masks predicted by the automatic segmentation model on unseen Specimens 5 (top row) and 6 (bottom row).

Table 1: Description of the dataset used for segmenting arthroscopic images. Two approaches were used: an Overfitting model (OverfitM) tailored to each cadaver specimen or the General model (GenM). The latter was trained using predictions from the three highlighted Overfitting models.

		# Training Images	# Testing Images	Segment. Model
Specimen	1	97	1823	OverfitM
	2	163	**3297**	OverfitM
	3	728	**4146**	OverfitM
	4	271	**2944**	OverfitM
	7	85	2071	OverfitM
	5	10 387	1114	GenM
	6		4095	

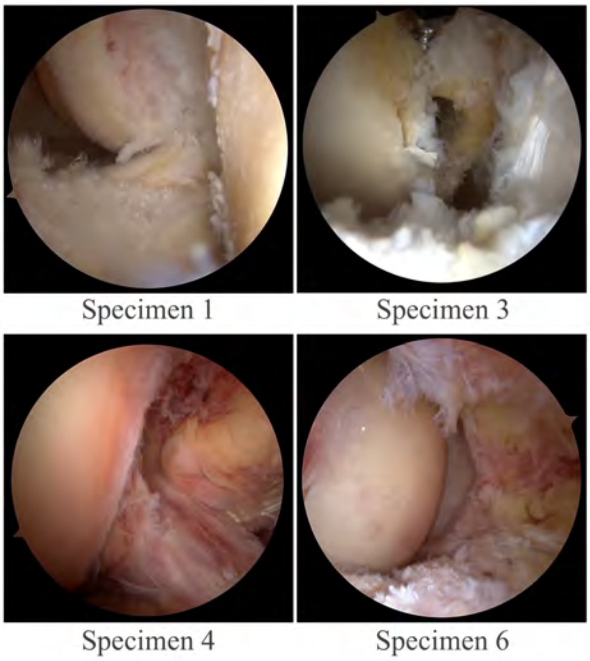

Fig. 2: Different knee specimens present significantly different levels of texture on bone and cartilage regions. While Specimen 1 presents considerable texture in the entire condyle and Specimen 3 also contains texture in the region of transition from bone to cartilage, Specimens 4 and 6 are mostly untextured. The higher the texture, the easier the matching process, and subsequently the more accurate the 3D registration.

Table 2: Number of failed registrations (# Failed) for each specimen, method and number of pairs (# Pairs) used to perform the registration.

		Methods	# Pairs	# Failed
Specimen	3	SIFT+NN	5	4
			10	1
		DISK+NN	5	3
		SP+LG	5	1
	4	SIFT+NN	5	2
			10	2
		LOFTR	5	1
	6	SIFT+NN	5	2
		DISK+NN	5	2
		RoMa	30	1

Table 3: Optimized parameters for each feature extraction and matching method tailored for arthroscopic images. Both SIFT and DISK use the NN matching method. Common parameters across all methods include: match_threshold: 0.0001; keypoint_threshold: 1e-6; and max_num_keypoints: 1e6.

<table>
<tr><th>SIFT</th><th>DISK</th><th>SP+LG</th></tr>
<tr><td>resize_max: 1600
peak_threshold: 1e-6
first_octave: 0
patch_size: 32
mr_size: 12</td><td>resize_max: 1600
nms_window_size: 5
detection_threshold: 0</td><td rowspan="3">resize_max: 1024
nms_dist: 2
width_confidence: 0.9999
depth_confidence: 0.80
filter_threshold: 0.05</td></tr>
<tr><th colspan="2">NN</th></tr>
<tr><td colspan="2">do_mutual_check: True
distance_threshold: None
ratio_threshold: None</td></tr>
</table>

LoFTR	RoMa	DKM
resize_max: 1024 weights: outdoor-ds	resize_max: 1024 weights: outdoor	resize_max: 1024 weights: outdoor

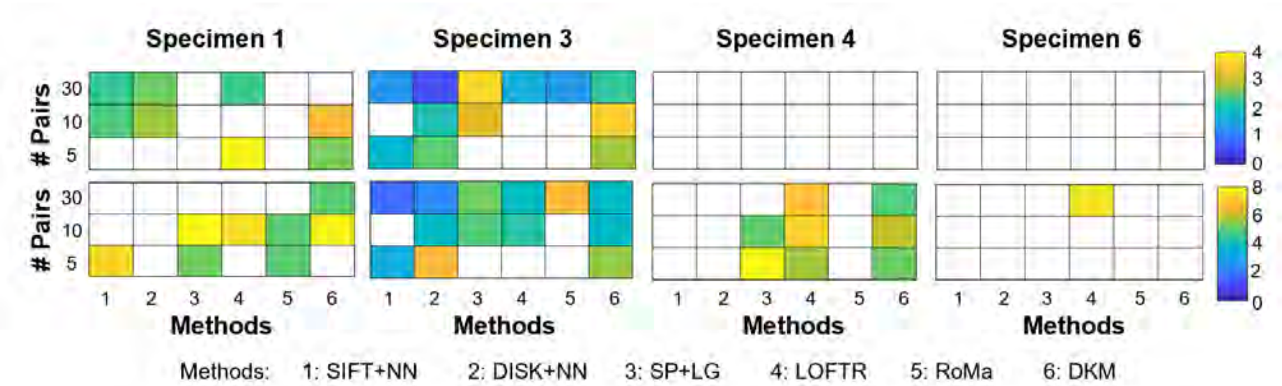

Fig. 3: Assessment of the registration accuracy in terms of tunnel placement for reconstructions obtained with the 6 different methods for 5, 10 and 30 pairs of images. The top row represents the median error in the entry point, in mm, and the bottom row represents the median error in tunnel direction, in degrees. White cells correspond to failure cases or errors larger than 4 mm or 8º. No semantic segmentation was considered for generating these results.

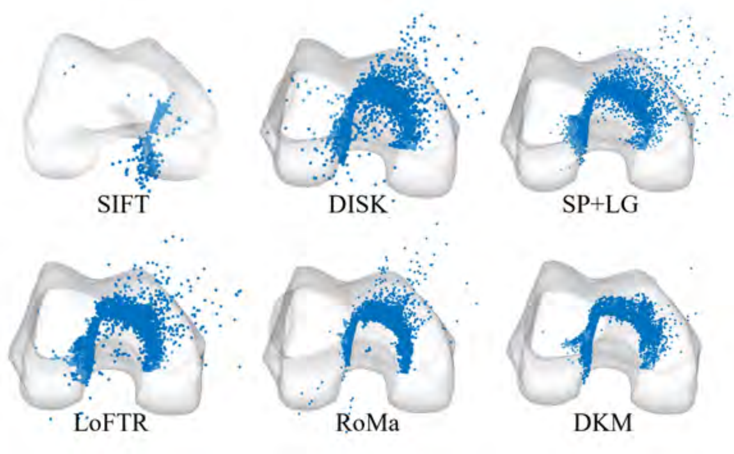

Fig. 4: Example of the alignment of 3D points reconstructed using the 6 considered methods with a 3D model obtained pre-operatively from CT image segmentation. This example represents points reconstructed from 30 image pairs combined with semantic segmentation of cartilage and bone of specimen 4. Each alignment is based on a randomly selected registration result.