PROCEEDINGS OF SPIE

Feature tracking for automated volume of interest stabilization on 4D-OCT images

Max-Heinrich Laves, Andreas Schoob, Lüder A. Kahrs, Tom Pfeiffer, Robert Huber, et al.

Feature tracking for automated volume of interest stabilization on 4D-OCT images

Max-Heinrich Laves[a], Andreas Schoob[a], Lüder A. Kahrs[a], Tom Pfeiffer[b], Robert Huber[b], and Tobias Ortmaier[a]

[a]Institute of Mechatronic Systems, Leibniz Universität Hannover, Appelstraße 11A, 30167 Hanover, Germany
[b]Institute of Biomedical Optics, Universität zu Lübeck, Peter-Monnik-Weg 4, 23562 Lübeck, Germany

ABSTRACT

A common representation of volumetric medical image data is the triplanar view (TV), in which the surgeon manually selects slices showing the anatomical structure of interest. In addition to common medical imaging such as MRI or computed tomography, recent advances in the field of optical coherence tomography (OCT) have enabled live processing and volumetric rendering of four-dimensional images of the human body. Due to the region of interest undergoing motion, it is challenging for the surgeon to simultaneously keep track of an object by continuously adjusting the TV to desired slices. To select these slices in subsequent frames automatically, it is necessary to track movements of the volume of interest (VOI). This has not been addressed with respect to 4D-OCT images yet. Therefore, this paper evaluates motion tracking by applying state-of-the-art tracking schemes on maximum intensity projections (MIP) of 4D-OCT images. Estimated VOI location is used to conveniently show corresponding slices and to improve the MIPs by calculating thin-slab MIPs. Tracking performances are evaluated on an *in-vivo* sequence of human skin, captured at 26 volumes per second. Among investigated tracking schemes, our recently presented tracking scheme for soft tissue motion provides highest accuracy with an error of under 2.2 voxels for the first 80 volumes. Object tracking on 4D-OCT images enables its use for sub-epithelial tracking of microvessels for image-guidance.

Keywords: 4D imaging, maximum intensity projection, optical coherence tomography, feature tracking

1. INTRODUCTION

Optical coherence tomography (OCT) is a three-dimensional imaging modality, which provides high-accuracy depth information of light scattering properties of biological tissue.[1] Recent advances have enabled live processing and volumetric rendering of four-dimensional (3D plus time as fourth dimension) OCT images.[2] The integration of such an OCT into an operating microscope enables its use for high-accuracy intra-operative guidance with visibility of subcutaneous structures up to 3 mm in depth.[3][4] Within this setup, redundant combination or fusion of RGB camera and OCT data brings additional information to the live camera viewer.[5] An advanced future application is envisioned by soft tissue laser ablation with compensated tissue motion.[6]

By extracting three orthogonal slices from the regarded volume, the triplanar view (TV) as common representation of volumetric images is generated. As this works well for static single-shot 3D volumes, in time-varying volumes the surgeon can easily lose track of the volume of interest (VOI) due to observer or tissue movement. A manual adjustment of the TV to desired slices is highly inconvenient and limits the potential of 4D-OCT as a tool for intra-operative guidance. An alternative to this is the maximum intensity projection (MIP) and the sliding thin-slab MIP (STS-MIP). MIPs do not have to be adjusted, but the resulting 2D image is low in contrast, noisy, and occlusions by brighter overlying structures can occur.

To the author's knowledge, feature tracking has not been done on 4D-OCT images yet. Registration of two 3D-OCT volumes using 3D SIFT has already been achieved, but without any investigation of long-term tracking,

---

Further author information: (Send correspondence to M.H. Laves)
M.H. Laves: E-mail: laves@imes.uni-hannover.de, Telephone: +49 511 762 19617

which is necessary for usage on 4D images.[7] Recent work has covered OCT pose estimation on 2D B-scans for visual servoing purposes.[8] However, this is limited to planar pose estimation with 3 degrees of freedom.

The purpose of this paper is to provide a convenient view of 4D-OCT data by automatically and continuously selecting slices, based on a user defined, pre-selected VOI. Therefore, we present suitable image processing frameworks and investigate different vision-based tracking algorithms for their use on MIPs of 4D-OCT data. The estimated location of the VOI is used to select corresponding slices.

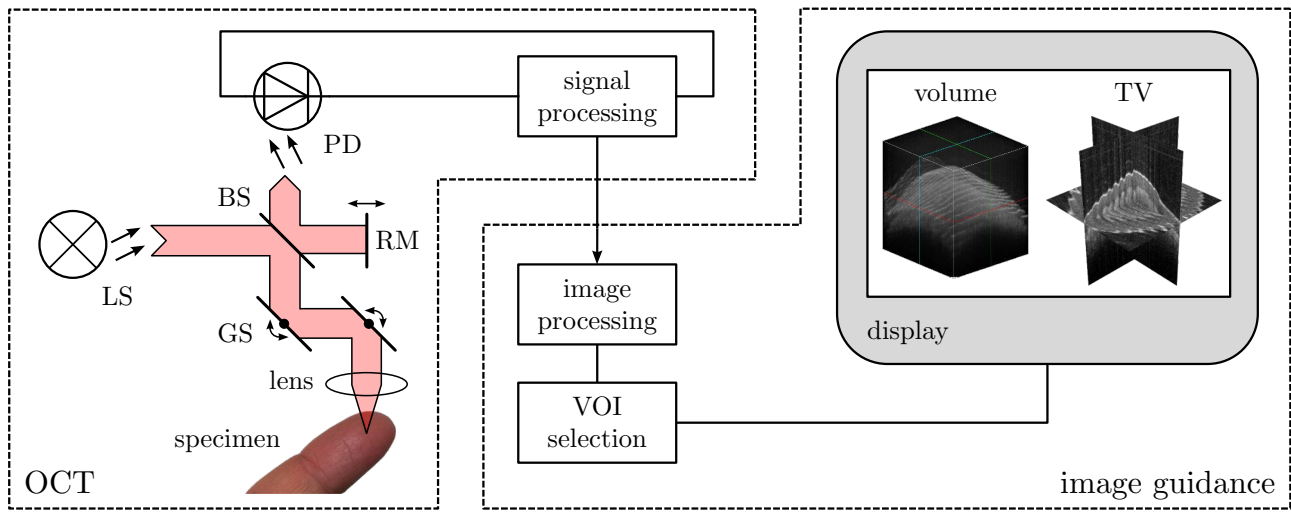# 2. METHODS

2.1 Experimental setup

Figure 1. Experimental setup with 4D-OCT image acquisition (left) and image guidance interface (right), with LS–light source, BS–beam splitter, PD–photo diode, RM–reference mirror, GS–two galvanometer scanners.[2]

The experimental setup (see Fig. 1) is composed of a 4D-OCT image acquisition unit—acquiring at a rate of 26 volumes per second—and an image guidance interface. Each volume consists of 320×320 depth scans, each with 400 pixel in depth, resulting in a rate of 1 GVoxel per second.[2] This study addresses the image processing required for tracking a user-selected VOI. The motion estimate is used to adapt corresponding slices and STS-MIPs subsequently for view stabilization.

2.2 Triplanar view of the OCT volume

The triplanar view shows three orthogonal slices (axial, coronal, and sagittal, see Fig. 2) of 3D image datasets, usually referred as anatomical planes. The coronal or sagittal correlates with the OCT B-scan, depending on the scan direction. To visualize a certain VOI, e. g. cancerous tissue under the epithelium, the user has to manually select slices, which intersect this structure.

Without any slice selection, the maximum intensity projection calculates 2D planar projections from 3D volumetric image data.[9] Parallel rays from a certain direction are casted through the OCT volume and intensities along the rays are analyzed. For each ray, only the maximum intensity value is projected on the resulting image plane. As OCT images represent light scattering properties, the maximum value corresponds either to matter interface or background noise. As a result of this, MIP images are low in contrast but have less speckle noise than single B-scans (see Fig. 2).

To overcome these drawbacks, the sliding thin-slab MIP (STS-MIP) is a technique for improved visualization of tomographic data. As the name suggests, MIPs from thin-slab subvolumes are only a small number of voxel thick.[10] The slabs can vary in thickness and resulting projections have higher contrast than conventional MIPs. Selecting the slabs containing the tracked VOI will be done based on the tracking result.

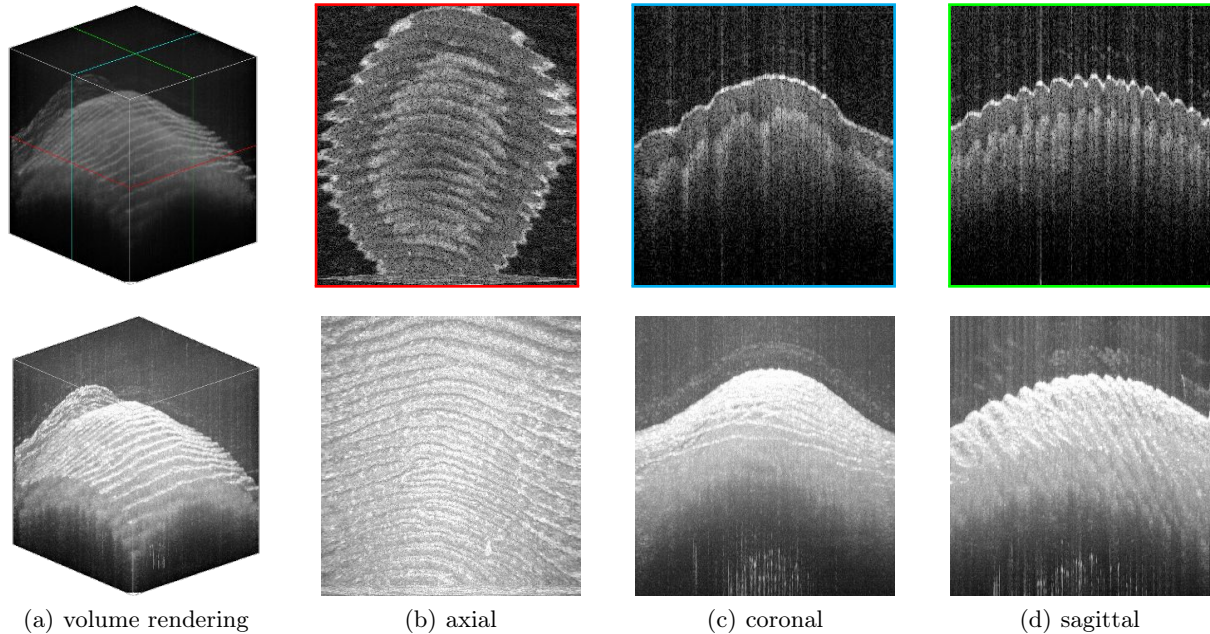(a) volume rendering   (b) axial   (c) coronal   (d) sagittal

Figure 2. Visualization of OCT data of finger tip. Top line: volume rendering with corresponding triplanar slices. Bottom line: Associated maximum intensity projections. (a) Volume rendering. (b–d) Triplanar views.

Table 1. A summary of tracking schemes evaluated in this paper.

#	Scheme	Description
1	SIFT	Scale-Invariant Feature Transform uses extremal values of differences of repeatedly Gaussian filtered image as feature points. These points are described with a histogram of local image gradients.[11]
2	SURF	Speeded-Up Robust Features are inspired by SIFT. It approximates the Gaussian filtering for computational efficiency and detects feature points by calculating the determinant of the Hessian matrix. Descriptors are obtained by using spacial wavelet responses.[12]
3	BRIEF	Binary Robust Independent Elementary Features use pixel binary tests in a smoothed image patch as feature descriptor. Due to its lack of feature detection, we combine BRIEF with SIFT detector.[13]
4	ORB	Oriented FAST and Rotated BRIEF combines FAST[14] feature detector with an orientation component and oriented BRIEF feature descriptors.[15]
5	BRISK	Binary Robust Invariant Scalable Keypoints uses a novel FAST-based[14] feature detector and a bit-string descriptor from intensity values by sampling each keypoint neighborhood.[16]
6	FREAK	Fast Retina Keypoints are inspired by the human visual system and compare image intensities over a retinal sampling pattern. Again, we combine it with the SIFT detector.[17]
7	BMATCH	Simple Block Matching slides an image patch over the reference image and calculates the normalized correlation coefficient. The best match is found at maximum correlation.
8	SOFT	Soft Tissue Motion Tracking is a recently proposed tracking scheme combining a piecewise affine deformation model with the epipolar geometry of stereo images.[6]

2.3 Investigated tracking schemes

The different tracking schemes evaluated in this study are summarized in Tab. 1. Implementations of schemes 1–7 are used from computer vision library "OpenCV 3.2".[18] The first six uses a brute force matcher to find feature correspondences. BMATCH is used with normalized correlation coefficient as similarity measure (refer to OpenCV documentation[19] for further explanation). SOFT aims on minimizing the tissue model energy function

$$\varepsilon(\boldsymbol{S}) = \varepsilon_C(\boldsymbol{S}) + \lambda_D \varepsilon_D(\boldsymbol{S}) \qquad (1)$$

of tissue model $\boldsymbol{S}$. ε is defined by the sum of correspondence energy ε_C and deformation energy ε_D providing regularization. The weight parameter is empirically set to $\lambda_D = 6.0$. This makes SOFT robust towards outliers.
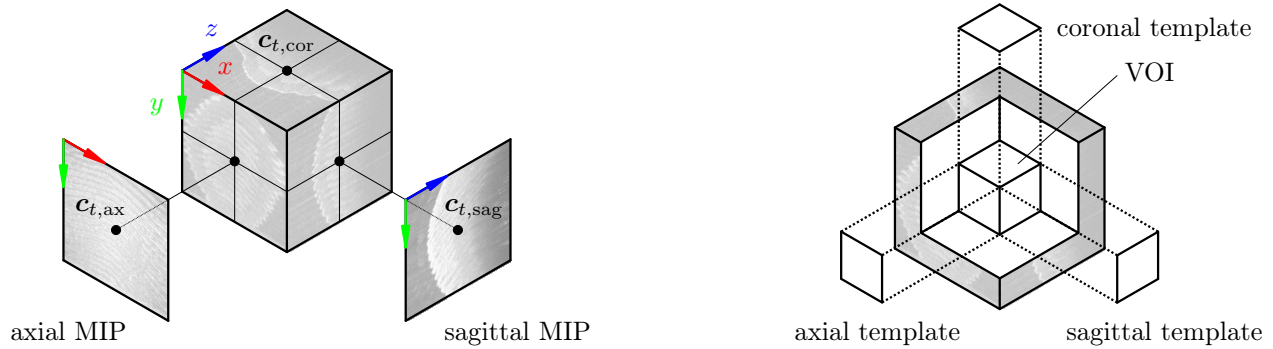
2.4 Tracking setup

Figure 3. Tracking setup. Left: The axial and sagittal MIPs are used to generate an artificial stereo view. Right: VOI within the OCT volume and corresponding 2D templates used for the tracking schemes. The VOI is tracked throughout the sequence with respect to image acquisition unit (OCT) and tissue motion.
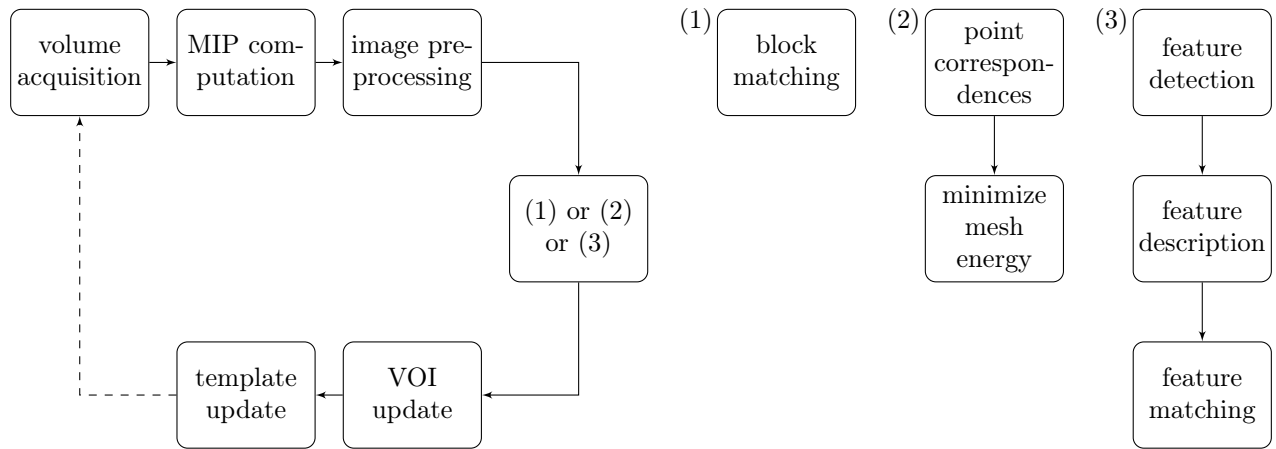

Figure 4. Tracking pipeline bases either on (1) path for BMATCH, (2) path for SOFT, (3) path for tracking schemes 1–6

The following tracking setup is used to assess the performance of the tracking schemes with respect to precision and computational cost. First, a volume of interest with the dimension of $100 \times 100 \times 100$ voxels and a starting center voxel $\boldsymbol{c}_0 = (150, 120, 150)^{\mathrm{T}}$ within the first OCT volume is defined. In the tracking pipeline (see Fig. 4), the MIPs of current volume are computed and $\boldsymbol{c}_t$ is projected onto them. Due to low contrast, the MIPs are preprocessed using histogram equalization. The actual tracking is performed on the resulting MIPs and therefore, the former three-dimensional tracking problem turns into multiple two-dimensional tracking problems. To reconstruct all components of $\boldsymbol{c}_t$, it is sufficient to track in only two of three possible MIPs. This results in an artificial stereo view with a rectangular epipolar geometry (see Fig. 3 left). The location of pixels which correspond to the same voxel share one component, e.g. let $\boldsymbol{c}_t = (x, y, z)^{\mathrm{T}}$, then the projection onto axial MIP is $\boldsymbol{c}_{t,\mathrm{ax}} = (x, y)^{\mathrm{T}}$ and onto sagittal MIP is $\boldsymbol{c}_{t,\mathrm{sag}} = (z, y)^{\mathrm{T}}$. SOFT enforces consistency between axial and sagittal y-coordinate, whereas the others just merge the common dimension of the results by arithmetic mean. As shown in Fig. 4, the tracking pipeline splits into three paths, depending on the selected method. The result of either of these paths is used to update the current VOI and to extract new tracking templates for next iteration.

As ground truth generation in soft tissue motion is still a problem, the forward-backward error metric[20] is used to quantitatively evaluate tracking precision. The FB error for tracked VOI center $\boldsymbol{c}_t$ is defined as

$$e_{\mathrm{FB},t} = \|\boldsymbol{c}_0 - \boldsymbol{c}_{0,\mathrm{end}}\|_2 \qquad (2)$$

with initial position c_0 and final position $c_{0,\text{end}}$ after tracking forward to volume $t = n$ and then backward to first volume $t = 0$. The test sequence counts 200 OCT volumes and errors are calculated at every 10th volume, which means $n = \{10, 20, \ldots, 200\}$.
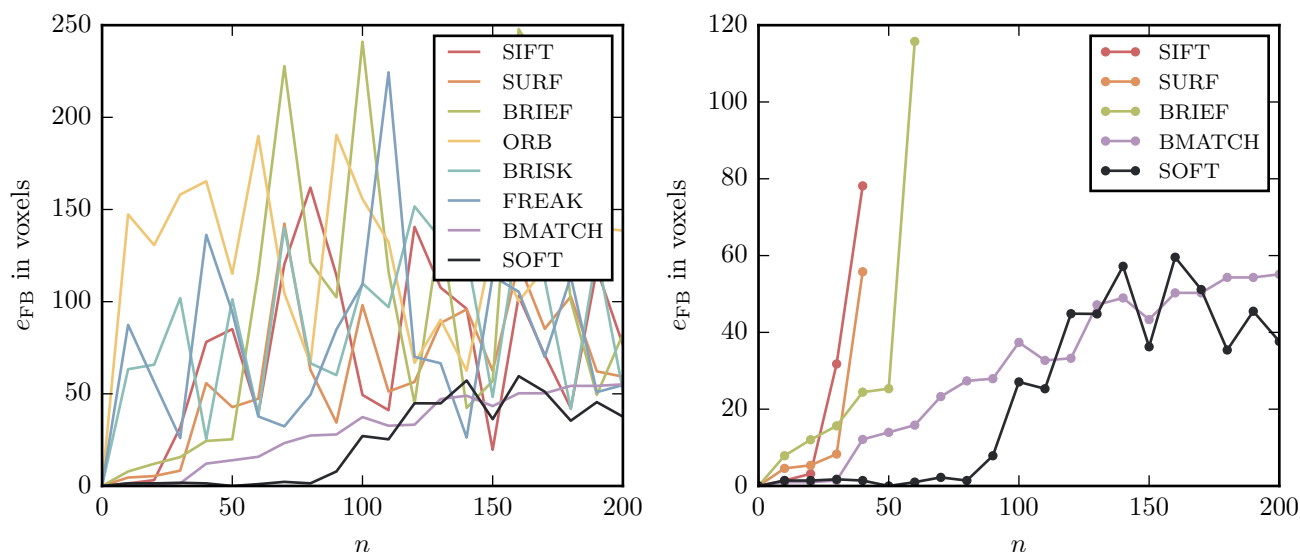
3. RESULTS

Figure 5. Quantitative tracking results measured with forward-backward error in voxels over OCT volume count. Left: Overview of all tracking results. Right: Selected tracking results until tracking failure.

Quantitative results of the aforementioned tracking setup (see Sec. 3) are shown in Fig. 5. The left figure shows results of all tracking schemes—including tracking failures. ORB, BRISK, and FREAK fail immediately. For instance, after the first 10 volumes, ORB already has a tracking error of ~150 voxels, as a result of a high number of erroneously matched features. The right figure shows tracking schemes up to their failure. In the beginning, SIFT and SURF have quite similar error values, but SIFT fails earlier. BRIEF tracks the VOI robustly for up to 50 frames before failure, but is characterised with higher error values in the beginning. The only schemes tracking the whole sequence robustly are BMATCH and SOFT, whereas SOFT has the lowest errors for up 110 frames and BMATCH has an almost linear error behaviour. At the end of the sequence, SOFT has the lowest error with $e_{\text{FB}} = 37.64$ voxels.

Qualitative results are shown in Fig. 6. First line shows four consecutive axial MIPs beginning at an arbitrary volume k. The images contain motion of the specimen relative to the OCT as well as tissue deformation due manipulation with the needle. The green dots represent tracked features. The VOI center $c_{t,\text{ax}}$ is highlighted through orange reticle. The needle induces partial occlusion. Coronal and sagittal slices transecting c_t are automatically selected and shown in the lower lines of Fig. 6. SOFT is robustly stabilizing the triplanar view.

Table 2. Average execution time (in seconds) per OCT volume.

	SIFT	SURF	BRIEF	ORB	BRISK	FREAK	BMATCH	SOFT
t	0.284	0.169	0.126	0.115	0.696	0.228	0.0345	1.663

Table 2 shows overall execution times of all tracking schemes, including MIP calculation. The first six feature-based schemes all have an execution time of same order of magnitude. BMATCH has the lowest execution time and is the only one investigated, which is able to process the OCT volumes at image acquisition rate of 26 Hz. SOFT has the longest execution time of 1.66 s. It has to be mentioned, that schemes 1–7 are implemented efficiently by OpenCV, whereas SOFT has not been optimized, yet.
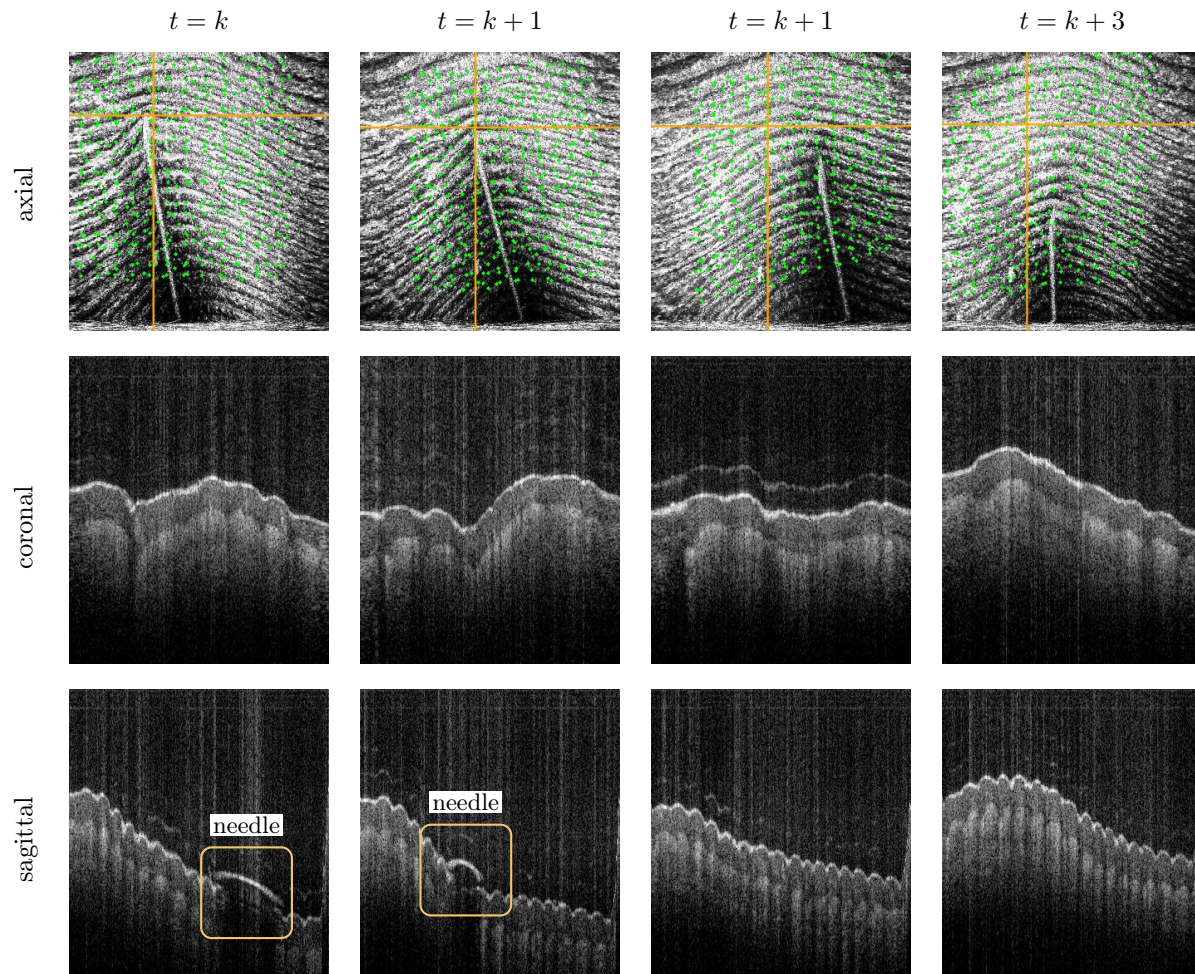
Figure 6. Automated slice selection on TV with SOFT algorithm. Top line: feature tracking on axial MIP. Lower lines: corresponding coronal and sagittal slices on which c_t is located. Note sagittal view transecting the needle for $t = \{0, 1\}$.

4. CONCLUSION

This study reveals first results of different feature tracking techniques for application in novel 4D-OCT imaging. It has been shown that conventional feature tracker—like SIFT or SURF—are not able to perform well on this kind of data. A simple block matching algorithm performed surprisingly good, but the tracking results state, that an algorithm especially tailored for soft tissue environments is mandatory for this dataset. The achieved accuracy is sufficient for intended use in 4D-OCT-guided interventions, even though the best performing algorithm has been developed for use on camera images. Therefore, future work will address the advancement of our soft tissue tracking scheme for its use on 4D-OCT images, including efficient GPU implementation.

Optical coherence tomography as relatively new and emerging medical imaging modality will attract more attention in the future. With recent advances in performance of graphics processing units (GPU), real-time processing of 4D-OCT data to extract additional information for intra-operative guidance and navigation will be possible and worth investigating. Object tracking on OCT MIPs for better visualization by VOI stabilization is an important step for enabling 4D-OCT for a whole new class of medical imaging products.

REFERENCES

[1] Huang, D., Swanson, E., Lin, C., Schuman, J., Stinson, W., Chang, W., Hee, M., Flotte, T., Gregory, K., Puliafito, C., and Et, A., "Optical coherence tomography," *Science* **254**(5035), 1178–1181 (1991). doi:10.1126/science.1957169.

[2] Wieser, W., Draxinger, W., Klein, T., Karpf, S., Pfeiffer, T., and Huber, R., "High definition live 3D-OCT in vivo: design and evaluation of a 4D OCT engine with 1 GVoxel/s," *Biomed. Opt. Express* **5**(9), 2963–2977 (2014). doi:10.1364/BOE.5.002963.

[3] Lankenau, E., Klinger, D., Winter, C., Malik, A., Müller, H. H., Oelckers, S., Pau, H.-W., Just, T., and Hüttmann, G., "Combining Optical Coherence Tomography (OCT) with an Operating Microscope," in [*Adv. Med. Eng.*], 343–348 (2007). doi:10.1007/978-3-540-68764-1_57.

[4] Diaz Diaz, J., Kundrat, D., Goh, K.-F., Majdani, O., and Ortmaier, T., "Towards Intra-operative OCT Guidance for Automatic Head Surgery Guidance: First Experimental Results," in [*MICCAI 2013*], *Lect. Notes Comput. Sci.* , 347–354 (2013).

[5] Bergmeier, J., Kundrat, D., Schoob, A., Kahrs, L. A., and Ortmaier, T., "Methods for a fusion of optical coherence tomography and stereo camera image data," in [*Proc. SPIE*], **9415**, 94151C–94151C–7 (2015). doi:10.1117/12.2082511.

[6] Schoob, A., Laves, M.-H., Kahrs, L. A., and Ortmaier, T., "Soft tissue motion tracking with application to tablet-based incision planning in laser surgery," *Int. J. Comput. Assist. Radiol. Surg.* **11**(12), 2325–2337 (2016). doi:10.1007/s11548-016-1420-5.

[7] Niemeijer, M., Garvin, M. K., Lee, K., Ginneken, B. V., and Abr, M. D., "Registration of 3D spectral OCT volumes using 3D SIFT feature point matching," in [*Proc. SPIE*], **7259**, 72591I–72591I–8 (2009). doi:10.1117/12.811906.

[8] Ourak, M., Tamadazte, B., Lehmann, O., and Andreff, N., "Wavelets-based 6 DOF visual servoing," in [*ICRA 2016*], 3414–3419 (2016). doi:10.1109/ICRA.2016.7487518.

[9] Laub, G. A. and Kaiser, W. A., "MR Angiography with Gradient Motion Refocusing," *J. Comput. Assist. Tomogr.* **12**(3), 377–382 (1988). doi:10.1097/00004728-198805000-00002.

[10] Napel, S., Rubin, G. D., and Jeffrey, R. B., "STS-MIP: A New Reconstruction Technique for CT of the Chest," *J. Comput. Assist. Tomogr.* **17**(5), 832–838 (1993). doi:10.1097/00004728-199309000-00036.

[11] Lowe, D. G., "Distinctive image features from scale invariant keypoints," *Int. J. Comput. Vis.* **60**, 91–110 (2004). doi:10.1023/B:VISI.0000029664.99615.94.

[12] Bay, H. and Ess, A., "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.* **110**, 346–359 (2008). doi:10.1016/j.cviu.2007.09.014.

[13] Calonder, M., Lepetit, V., Strecha, C., and Fua, P., "BRIEF: Binary robust independent elementary features," in [*ECCV 2010*], *Lect. Notes Comput. Sci.* , 778–792 (2010). doi:10.1007/978-3-642-15561-1_56.

[14] Rosten, E. and Drummond, T., "Machine learning for high-speed corner detection," in [*ECCV 2006*], *Lect. Notes Comput. Sci.* **3951**, 430–443 (2006). doi:10.1007/11744023_34.

[15] Rublee, E. and Bradski, G., "ORB: An efficient alternative to SIFT or SURF," in [*ICCV 2011*], 2564–2571 (2011). doi:10.1109/ICCV.2011.6126544.

[16] Leutenegger, S., Chli, M., and Siegwart, R. Y., "BRISK: Binary Robust invariant scalable keypoints," in [*ICCV 2011*], 2548–2555 (2011). doi:10.1109/ICCV.2011.6126542.

[17] Alahi, A., Ortiz, R., and Vandergheynst, P., "FREAK: Fast Retina Keypoint," in [*CVPR 2012*], 510–517 (2012). doi:10.1109/CVPR.2012.6247715.

[18] Bradski, G., "The OpenCV Library," *Dr. Dobb's Journal of Software Tools* (2000).

[19] Itseez, "Open Source Computer Vision Library." http://opencv.org/ (accessed on 2017-01-10).

[20] Kalal, Z., Mikolajczyk, K., and Matas, J., "Forward-Backward Error: Automatic Detection of Tracking Failures," in [*ICPR 2010*], 2756–2759 (2010). doi:10.1109/ICPR.2010.675.