RTIP: A FULLY AUTOMATED ROOT TIP TRACKER FOR MEASURING PLANT GROWTH WITH INTERMITTENT PERTURBATIONS

Deniz Kavzak Ufuktepe[1] *Kannappan Palaniappan*[1,⋆] *Melissa Elmali*[2] *Tobias I. Baskin*[2]

[1] Electrical Engineering and Computer Science Dept., University of Missouri, Columbia, MO, USA
[2] Biology Department, University of Massachusetts, Amherst, MA, USA

ABSTRACT

RTip is a tool to quantify plant root growth velocity using high resolution microscopy image sequences at sub-pixel accuracy. The fully automated RTip tracker is designed for high-throughput analysis of plant phenotyping experiments with episodic perturbations. RTip is able to auto-skip past these manual intervention perturbation activity, *i.e.* when the root tip is not under the microscope, image is distorted or blurred. RTip provides the most accurate root growth velocity results with the lowest variance (*i.e.* localization jitter) compared to six tracking algorithms including the top performing unsupervised Discriminative Correlation Filter Tracker and the Deeper and Wider Siamese Network. RTip is the only tracker that is able to automatically detect and recover from (occlusion-like) varying duration perturbation events.

Index Terms— Plant growth, Shi-Tomasi features, Kanade-Lucas-Tomasi, Radon Transform, Video Tracking

1. INTRODUCTION

Observing growth kinematics of roots under various experimental conditions is an important component of plant physiology studies. In some experiments, root needs to be tracked before and after some experimental plant manipulation of varying duration. In this work the aim is to track growth rate, root tip needs to be tracked accurately at subpixel resolution before and after manipulations. Various image analysis and single-object tracking methods may be suitable to track the root tip depending on the microscopy and experimental setup.

There has been extensive work done on plant image analysis for understanding root kinematics. Van der Weele et al. [1] used computational image analysis of deformable motion [2] at high resolution for root growth in *RootflowRT*. Jiang et al. [3] estimated the non-rigid motion of plant roots by using a combination of matching and tensor methods. Wangenheim et al. [4] developed a confocal microscope setup for vertical sample mounting and integrated directional illumination, and a tracker called *TipTracker*, a custom software for automatic tracking of diverse moving objects usable on various microscope setups. Baskin and Zelinsky [5] did a study on kinematic characterization of root growth by using *Stripflow*, which was introduced in a previous study by Yang et al. [6], leverages the prior knowledge that movement at nearby points are similar. Other tools have been developed to analyze the growth of root tips in experimental setups with roots grow vertically at lower resolution. The root tip analysis tool called *ARTT* combined a segmentation algorithm with additional imaging filters for 2D tip detection was proposed by Russino et al. [7]. French et al. [8] implemented *RootTrace*, combined a particle-filtering algorithm with a graph-based method to trace the root tip.

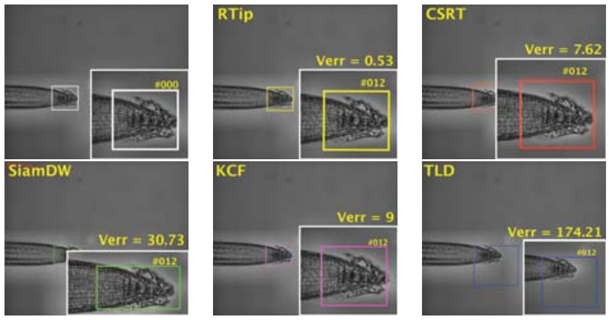

Fig. 1. Initial BB in frame 0 and tracking results on frame 12, of RTip, CSRT, SiamDW, KCF and TLD with velocity errors.

Complex biological protocols for characterizing the growth kinematics of plant phenotypes now have additional requirements for accurate video tracking. In this case the experiments required manual manipulation of the plant shoot during growth; perturbations that involved moving the plant so that the root tip becomes "occluded" (*i.e.* outside the microscope field-of-view) for an unknown duration. The plant is manipulated then placed back on the stage and the root can reappear in a new position. During this interval imaging continues, to preserve the time-base, so frames may be empty or blurred, and the continuity of the root tip trajectory is broken.

We developed RTip, a fully automated tracking tool to measure root tip motion in sequences where the root is imaged

*This work was partially supported by awards from U.S. NIH R01NS110915 and ARL W911NF-1820285. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U. S. Government or agency thereof.

at high enough resolution and where there are episodic perturbation discontinuities in the position of the root. RTip estimates a smooth displacement between bounding boxes (BB) while tracking, producing accurate velocity estimates. Experiments done on two root tip image sequences show that RTip outperforms six tested trackers in velocity accuracy, including Discriminative Correlation Filter Tracker (CSRT) [9] and SiamDW tracker [10]. Although both are high quality trackers, the shape and size of the BBs change during tracking, which induces centroid jitter errors and leads to imprecise velocity estimates compared to RTip. Fig. 1 shows initial BB for frame 0 and tracking results on frame 12 for RTip, CSRT, SiamDW, Kernelized Correlation Filters (KCF) [11] and Tracking, Learning and Detection (TLD) [12] with corresponding velocity errors, where changes in BB size and location can be seen. RTip is designed to automatically detect perturbed frames for skipping (pause tracking), when a valid frame comes after the perturbed ones to reinitialize and restart the tracker (continue tracking).

There are several novel contributions in this study. RTip, is a fully automated high-throughput root tip tracker for kinematic measurement and phenotype screening developed for use with high resolution microscopy imaging and a complex experimental setup involving episodic perturbations. To our knowledge, these conditions cause other state-of-the-art trackers to fail. RTip handles automatic detection and skipping of the perturbed frames in the sequence by using a fast and efficient root tip detection algorithm. Occlusion handling uses the Radon transform and a robust Kanade-Lucas Tomasi tracker feature matching to detect and localize a distorted root tip and restart tracking based on forward backward error and robust outlier handling. Root displacements are estimated with sub-pixel accuracy and minimal centroid jitter noise.

2. RTIP FOR TRACKING WITH PERTURBATIONS

2.1. Automatic Initialization and Reinitialization (AIR) with Bounding Box (BB) Improvement

For initialization, the root tip is found and localized roughly in the image with an initial BB, using Normalized Cross Correlation (NCC) [13] for template matching with an example cropped root tip image [14]. To improve accuracy, BB candidates around this initial BB are sampled and Shi-Tomasi structure tensor compact corner-like features extracted [15]. For each pixel, the minimum eigenvalue of the tensor above a threshold are selected for matching.

A robust version of Kanade-Lucas Tomasi (KLT) feature tracker [16], rKLT is used to track features between consecutive frames. KLT computes the displacement in textured windows between sequential images. Root tip motion is modeled as a translation plus noise and outliers to handle root perturbations, aiming to minimize the error in the motion vector. The rKLT uses the Forward-Backward Error (FBE) [17] to determine the inliers while tracking the features between frames.

To find a good translation, rKLT tracks feature points, forward and backward between consecutive frames. The distance between feature point p, and corresponding point obtained by the backward tracking of p is measured as FBE. If the distance is larger than a threshold for any p, then it is filtered out as an outlier. The number of inliers are stored for each candidate BB, and the candidate with the highest number of inliers is selected as the next BB. If the maximum number of inliers is smaller than a threshold, the current frame is not a valid, high quality frame and no good transformation can be estimated, hence it is skipped. The next frame is used to initialize or reinitialize the tracker using the same procedure.

2.2. Automatic Invalid Frame (AIF) Detection-Recovery

The Radon transform (RT) is an operator to calculate projections of an object (2D or 3D) along specified directions, *i.e.* rotation angles, using line integrals. The coordinates of the original image are rotated by an angle θ, and a set of parallel lines that are perpendicular to the rotated coordinates are used for integration as shown [18],

$$R_\theta(x') = \int_{-\infty}^{\infty} f(x'\cos\theta - y'\sin\theta, x'\sin\theta + y'\cos\theta)dy'$$
$$\text{where, } \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad (1)$$

The RT computes projection vectors for each $\theta \in [0, 180]$, but a simplified version with only 0 and 90 degree vectors can be used for segmentation of suitable images [19]. For each frame, the simplified RT vectors are calculated using 0 and 90 degree projections. The RT projection vectors of current, $\mathbf{u}$, and previous frame, $\mathbf{v}$, are compared using the Chebyshev distance, defined as,

$$D^{\mathrm{RT}}_{i,\theta}(\mathbf{u}, \mathbf{v}) := \max_p |\mathrm{u}_p - \mathrm{v}_p| \qquad (2)$$

Distances between corresponding RT projection vectors (profiles), $D^{\mathrm{RT}}_{i,0}$ and $D^{\mathrm{RT}}_{i,90}$, between current i^{th} and previous frames is used to detect the root tip object. A threshold on the distance value is used to decide whether there is an in-focus root tip present in the image. When one of the D^{RT}_{i} values is greater than a given threshold, the current frame is identified as a perturbed frame, *i.e.* no root tip object in frame. The tracker is then stopped and the previous frame along with its root tip object is stored and marked as the safe frame, so that this frame can be used to check subsequent frames until the root tip becomes visible again. Once a safe or reliable frame is marked, then for each new frame, the distance between RT vectors at each projection between current frame and marked frame are denoted as $D^{\mathrm{RT}}_{s,0}$ and $D^{\mathrm{RT}}_{s,90}$. To detect a valid frame after the perturbation, $|D^{\mathrm{RT}}_{i,0} - D^{\mathrm{RT}}_{s,0}|$ and $|D^{\mathrm{RT}}_{i,90} - D^{\mathrm{RT}}_{s,90}|$ are compared to a threshold, and if both are smaller, then tracking is reinitialized using the AIR module.

2.3. rKLT Tracking With MSAC Outlier Filtering

After the initial BB is determined by the AIR module, rKLT tracking is initialized using the features within the selected BB extracted by Shi-Tomasi feature detector. For each frame, rKLT tracks the points from the previous frame to the current frame by using the given FBE threshold. The old inliers are updated from previous feature points with the new inliers selected by rKLT. A similarity transformation between the old and new inliers is estimated using a variation of Random sample consensus (RANSAC); M-estimator sample consensus (MSAC) [20]. MSAC gives a value to the inliers with respect to how well they fit, in addition to giving a penalty to outliers when calculating the cost function.

A maximum distance threshold is used in transformation estimation, to find the number of inliers with best fit, where the error measure is the distance between a feature point p_i and the back transformation p'_i of the corresponding transformed point $T(p_i)$. In order to estimate a similarity transformation, the inliers must consist at least 2 points. If the number of inliers is smaller than 2, the AIR module is used. The control of the rKLT tracker's state is determined by the AIF module. The centroids of BBs are used to estimate velocity based on displacement along the x-axis.

3. EXPERIMENTAL RESULTS

RTip was tested on two root sequences, in which seedlings are grown on square plastic plates containing a transparent substrate allowing roots to be imaged using transmitted light. The plate was placed on the stage of a horizontal, compound microscope and imaged through a $10\times$ objective with a spatial resolution of 2.5 microns per pixel, and frame size of 2048×2448 pixels [6]. The interval between frames is 30 seconds with a total imaging time of 40 minutes for an 80 frame video. After imaging starts, during a perturbation, the plant is removed from the microscope to cut off the plant shoot in order to investigate its effect on root growth, then the root is placed back under the microscope. Images were acquired continuously even while the plant was being manipulated outside the field-of-view in order to have an accurate time base. Consequently there will be several frames without the root tip being visible and sometimes the root tip is not re-placed properly. RTip identifies and skips past perturbation activity to automatically track root tips and quantitatively measure kinematic behavior.

The threshold for the RT distance comparisons is selected as 4 pixels for each sequence, and maximum number of pixel distances for the FBE used in rKLT is 2 pixels. In the similarity transformation estimation, the threshold of maximum distance for MSAC is set to 2 pixels. A quality level threshold of at least 0.01 for the Shi-Tomasi feature detector was used. RTip is tested on the timelapse videos in automated mode. Some intermediate results of the RTip tracker with velocity estimations are shown in the Fig. 2. Six other well-known trackers were also tested on the same sequences including top performing CSRT [9] and SiamDW deep learning tracker with pre-trained weights [10]. Other trackers tested include KCF [11], Median Flow [17], Multiple Instance Learning (MIL) [21] and TLD [12]. For these six trackers, the BB was manually reinitialized after a perturbation occurred.

Due to the deformable nature of the root tip, it is difficult to create a precise ground truth using manually drawn BBs in every frame [22]. So the ground truth BBs for both test sequences were drawn using linear interpolation between selected key frames. The reference or key frames were frequently updated to minimize the growth rate change in the interpolated intervening BBs. There is some noise in the ground truth since the root tip growth rate is not constant and the reference boxes are drawn manually. Reference BBs are drawn in the first frame and each 5th frame before perturbation with linearly interpolated BBs using these key frames. Similarly, a new initialization BB is drawn in the frame after perturbation and each 5th frame until the end of the sequence, with the BBs in-between estimated using linear interpolation.

Several different metrics are used to evaluate tracker performance (see Table 1). First, Velocity Estimation Error (V_{err}), is the average of all distances between the algorithm estimated velocity or pixel displacement and the velocity of the manual ground truth in each frame (i.e. centroid locations). The mean and standard deviation of the error distances for each method is shown in Table 1. Second, the gray level Root Mean Squared Error (RMSE) is used to measure the intensity differences between the initial root tip template and tracked BBs, is defined as,

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} \sum_{\mathbf{x} \in \mathcal{BB}} (c_i(\mathbf{x}) - t(\mathbf{x}))^2 \right]^{1/2} \quad (3)$$

$$\text{with, } c_i(\mathbf{x}) = I_i(\mathbf{x} + \Delta\mathbf{x}), t(\mathbf{x}) = I_0(\mathbf{x}) \quad (4)$$

where I_i is the i^{th} frame and I_0 is Frame 0 containing the template of the root tip. Finally, the average Structural Similarity Index (SSIM) over all BBs is used to measure the BB similarity compared to the ground truth for each tracker incorporating luminance, contrast and structure information,

$$SSIM = \frac{1}{n} \sum_{i=1}^{n} \frac{(2\mu_{c_i}\mu_t + s_1)(2\sigma_{c_i t} + s_2)}{(\mu_{c_i}^2 + \mu_t^2 + s_1)(\sigma_{c_i}^2 + \sigma_t^2 + s_2)} \quad (5)$$

where μ_{c_i} and μ_t are the mean intensities within the corresponding BBs, σ_{c_i} and σ_t the corresponding BB intensity variances and $\sigma_{c_i t}$ the covariance between BB pixel intensities, and s_1, s_2 are two constants used for numerical stability.

The measurements of RMSE and SSIM are averaged over all the images from both timelapse videos, and the mean and standard deviation values are given in Table 1. Ground truth values (first row) are also included in the table since the root tip structures have deformable motion. Tracker quality evaluated using RMSE and SSIM, together with the adaptability of each method to the perturbations are shown in Table 1.

Tracker	V_{err}	RMSE	SSIM	Adapt
GT	NA	6.23±1.68	0.81±0.05	NA
RTip	**0.49±0.34**	**6.37±1.58**	**0.80±0.05**	✓
MIL	0.56±0.61	6.41±1.90	0.80±0.06	×
MedFl	0.57±0.48	**6.35±1.88**	**0.80±0.05**	×
KCF	0.97±1.17	8.55±1.41	0.69±0.05	×
Siam	1.92±1.72	8.94±1.61	0.68±0.05	×
CSRT	2.64±2.66	8.04±1.87	0.71±0.07	×
TLD	13.01±30.09	10.23±1.78	0.68±0.05	×

Table 1. Experiment results for V_{err} averaged over all frames from both sequences S1 and S2. Mean and standard deviation of root tip velocity errors are in pixels per frame, RMSE (lower is better) and SSIM (higher is better) values with corresponding standard deviations (lower is better). Adaptability of tracker due to a perturbation (Col 5).

4. DISCUSSION

A comparison of velocity estimates V_{err} shows that RTip gives the highest accuracy results with the lowest mean error and standard deviation using the BB centroids. Evaluation using RMSE and SSIM show that, RTip has a small error with a small standard deviation in RMSE, and a high similarity value with a small standard deviation in SSIM, close to the ground truth. MIL and MedianFlow also have close mean values to RTip, but the standard deviations, *i.e.* jitter in the localizations, are higher than RTip. Other tested trackers have higher errors with lower similarity values. Therefore, KCF, SiamDW, CSRT and TLD do not perform well in such root tip image sequences. Moreover, none of the other tested trackers could adapt automatically to the perturbation activity in the videos, requiring a manual stop and restart with a manual reinitialization after the perturbation. RTip successfully detected the perturbations and automatically recovered from the episodic activity, without any manual intervention. Hence, RTip is effective for high-throughput phenotyping and screening studies of root tip kinematics.

5. CONCLUSIONS

The proposed RTip tracker uses the Radon transform for occlusion handling, and robust KLT (rKLT) feature matching to handle deformable tissue structures and outliers, providing accurate bounding boxes and subpixel motion estimates of the root tip with smooth trajectories. Jitter in the position or shape of the bounding box from KLT is minimized using the Forward-Backward Error (FBE) combined with using a robust M-estimator Sample Consensus (MSAC) method for rejecting outliers in the rKLT matches during similarity transformation estimation. Manual interaction is not needed

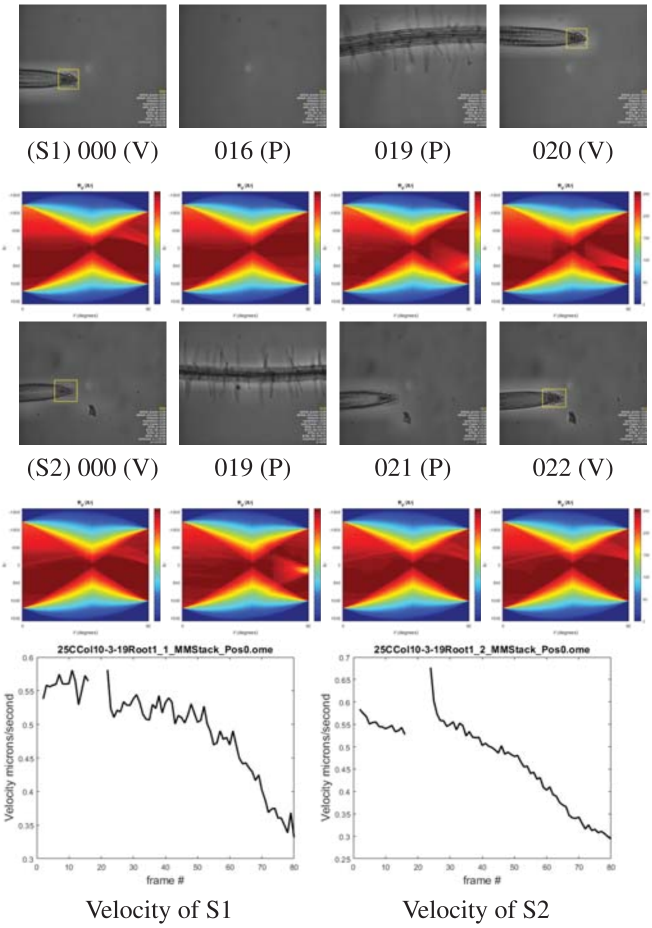

Fig. 2. RTip results for root sequences S1 and S2 shown in Rows 1 and 3 with corresponding RTs for each frame in Rows 2 and 4. Frames are marked as Valid (V) and Perturbed (P). Row 5 shows the root tip velocity vs frame number (time).

to recover from episodic perturbations when the root is removed from the imaging station for biological manipulation and then placed back on the microscopy stage, modeled as a specialized type of occlusion handling or activity detection.

The simplified Radon transform was used to identify the visibility of the root tip object of interest in the field-of-view, including filtering frames where the root may be visible but not the root tip. FBE assists in the robust matching of multiple bounding boxes near the root tip to recover from perturbations, motion blur, and illumination changes. RTip can be scaled to high-throughput biological studies with thousands of motion sequences since RTip can automatically detect the root tip object without manual interactive bounding box (re)initialization. RTip outperformed SiamDW, a state-of-the-art deep learning-based tracking algorithm and winner of the VOT-2019 RGB-D challenge. Compared to the six trackers tested, RTip had the lowest average velocity error of less than 0.50 pixel per frame and the lowest variance.

6. REFERENCES

[1] Corine M van der Weele, Hai S Jiang, Krishnan K Palaniappan, et al., "A new algorithm for computational image analysis of deformable motion at high spatial and temporal resolution applied to root growth. roughly uniform elongation in the meristem and also, after an abrupt acceleration, in the elongation zone," *Plant Physiology*, vol. 132, no. 3, pp. 1138–1148, 2003.

[2] Kannappan Palaniappan, Hai S. Jiang, and Tobias I. Baskin, "Non-rigid motion estimation using the robust tensor method," in *IEEE CVPR Workshop on Articulated and Nonrigid Motion*, 2004, vol. 1, pp. 25–33.

[3] Hai S Jiang, Kannappan Palaniappan, and Tobias I Baskin, "A combined matching and tensor method to obtain high fidelity velocity fields from image sequences of the non-rigid motion of the growth of a plant root," in *IASTED Int. Conf. on Biomedical Engineering*, 2003, pp. 386–010.

[4] Daniel von Wangenheim, Robert Hauschild, and Jiří others, "Live tracking of moving samples in confocal microscopy for vertically grown roots," *eLife*, vol. 6, pp. e26792, 2017.

[5] Tobias I Baskin and Ellen Zelinsky, "Kinematic characterization of root growth by means of Stripflow," in *Plant Cell Morphogenesis*, pp. 291–305. Springer, 2019.

[6] Xiaoli Yang, Gang Dong, Kannappan Palaniappan, Guohua Mi, and Tobias I Baskin, "Temperature-compensated cell production rate and elongation zone length in the root of *Arabidopsis thaliana*," *Plant, Cell & Environment*, vol. 40, no. 2, pp. 264–276, 2017.

[7] Andrea Russino, Antonio Ascrizzi, Liyana Popova, Anna Tonazzini, Stefano Mancuso, and Barbara Mazzolai, "A novel tracking tool for the analysis of plant-root tip movements," *Bioinspiration & Biomimetics*, vol. 8, no. 2, pp. 025004, 2013.

[8] Andrew French, Susana Ubeda-Tomás, Tara J Holman, Malcolm J Bennett, and Tony Pridmore, "High-throughput quantification of root growth using a novel image-analysis tool," *Plant Physiology*, vol. 150, no. 4, pp. 1784–1795, 2009.

[9] Alan Lukezic, Tomas Vojir, Luka ˇCehovin Zajc, Jiri Matas, and Matej Kristan, "Discriminative correlation filter with channel and spatial reliability," in *CVPR*, 2017, pp. 6309–6318.

[10] Zhipeng Zhang and Houwen Peng, "Deeper and wider siamese networks for real-time visual tracking," in *CVPR*, June 2019.

[11] Martin Danelljan, Fahad Shahbaz Khan, Michael Felsberg, and Joost Van de Weijer, "Adaptive color attributes for real-time visual tracking," in *CVPR*, 2014, pp. 1090–1097.

[12] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2011.

[13] Jae-Chern Yoo and Tae Hee Han, "Fast normalized cross-correlation," *Circuits, Systems and Signal Processing*, vol. 28, no. 6, pp. 819, 2009.

[14] Kannappan Palaniappan, Filiz Bunyak, et al., "Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video," in *13th Int. Conf. Information Fusion*, 2010.

[15] Jianbo Shi and Carlo Tomasi, "Good features to track," in *CVPR*, 1994, pp. 593–600.

[16] Carlo Tomasi and Takeo Kanade, "Tracking of point features," Tech. Rep., Tech. Rep. CMU-CS-91-132, Carnegie Mellon University, 1991.

[17] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas, "Forward-backward error: Automatic detection of tracking failures," in *International Conference on Pattern Recognition*, 2010, pp. 2756–2759.

[18] Rengarajan Pelapur, Filiz Bunyak, Guna Seetharaman, and Kannappan Palaniappan, "Vehicle detection and orientation estimation using the radon transform," in *Proc. SPIE Conf. Geospatial InfoFusion III (Defense, Security and Sensing: Sensor Data and Information Exploitation)*, 2013, vol. 8747, p. 87470I.

[19] Sema Candemir, Stefan Jaeger, Kannappan Palaniappan, et al., "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 577–590, 2013.

[20] Philip HS Torr and Andrew Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[21] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie, "Visual tracking with online multiple instance learning," in *CVPR*, 2009, pp. 983–990.

[22] Sumit Nath and Kannappan Palaniappan, "Adaptive robust structure tensors for orientation estimation and image segmentation," *Lecture Notes in Computer Science (ISVC)*, vol. 3804, pp. 445–453, 2005.