

# CASPER: computer-aided segmentation of imperceptible motion—a learning-based tracking of an invisible needle in ultrasound

Parmida Beigi<sup>1</sup> · Robert Rohling<sup>2</sup> · Septimiu E. Salcudean<sup>1</sup> · Gary C. Ng<sup>3</sup>

Received: 11 January 2017 / Accepted: 14 June 2017 / Published online: 24 June 2017  
© CARS 2017

## Abstract

**Purpose** This paper presents a new micro-motion-based approach to track a needle in ultrasound images captured by a handheld transducer.

**Methods** We propose a novel learning-based framework to track a handheld needle by detecting microscale variations of motion dynamics over time. The current state of the art on using motion analysis for needle detection uses absolute motion and hence work well only when the transducer is static. We have introduced and evaluated novel spatiotemporal and spectral features, obtained from the phase image, in a self-supervised tracking framework to improve the detection accuracy in the subsequent frames using incremental training. Our proposed tracking method involves volumetric feature selection and differential flow analysis to incorporate the neighboring pixels and mitigate the effects of the subtle tremor motion of a handheld transducer. To evaluate the detection accuracy, the method is tested on porcine tissue *in-vivo*, during the needle insertion in the biceps femoris muscle.

**Results** Experimental results show the mean, standard deviation and root-mean-square errors of 1.28°, 1.09° and 1.68° in the insertion angle, and 0.82, 1.21, 1.47 mm, in the needle tip, respectively.

**Conclusions** Compared to the appearance-based detection approaches, the proposed method is especially suitable for

needles with ultrasonic characteristics that are imperceptible in the static image and to the naked eye.

**Keywords** Needle tracking · Interventions · Optical flow · Ultrasound · Motion detection

## Introduction

Ultrasound (US) is a widespread imaging tool used to guide interventional procedures such as biopsies, treatment injections, and anesthesia [1]. During such applications, visualization of the needle and its accurate placement into the target is crucial. The visualization of the needle in the US image, however, is challenging due to specular reflection off the smooth surface of the needle [2]. Needle visualization is especially challenging with steep insertion angles, due to non-axial specular beam reflection off the smooth surface of the needle and similar line-like anatomical features. A sample of the literature is provided below to summarize the wide variety of approaches to enhance needle visibility.

(1) Signal and image processing methods such as line detection [3–5] and projection-based [6,7] approaches were used to augment and localize the needle in the ultrasound image. 3D imaging is also used to help with out-of-plane insertions [8,9]; the main issue of needle visibility due to poor echo still remains. Most of these types of approaches rely on the initial visibility of the needle in the US image as a long line-like structure. (2) Hardware-based enhancing methods, such as sensors and actuators [10,11], echogenic technology [12], needle guides [13], and robot-assisted needles [14,15] are used to detect the needle in US. Each of these have advantages and disadvantages; disadvantages are additional cost and complexity, and in the case of needle guides, restrictions on the freedom of needle trajectory paths

✉ Parmida Beigi  
parmidab@ece.ubc.ca

<sup>1</sup> Electrical and Computer Engineering Department, University of British Columbia, Vancouver, BC, Canada

<sup>2</sup> Electrical and Computer Engineering Department and Mechanical Engineering Department, University of British Columbia, Vancouver, BC, Canada

<sup>3</sup> Philips Ultrasound, Bothell, WA, USA

**Table 1** A sample of the literature summarizing the variety of approaches to enhance needle visibility

Signal/image processing	Hardware modification	Others
Line detection [3–5]	Sensors/actuators [10,11]	beam steering [16]
Projection-based [6,7]	Echogenic technology [12]	Tremor detection [17,18]
3D ultrasound [8,9]	Mechanical needle guides [13]	
	Robot-assisted [14,15]	

relative to the transducer, making them too cumbersome for some applications. In addition, the current clinical demand is inclined toward standard needles, and methods using customized apparatus represent a relatively small percentage of those used clinically. (3) Beam-steering approaches for linear transducers are also used to enhance the visibility of the needle shaft by adaptively steering the beam to a perpendicular degree to the needle [16]. In summary, previous approaches either rely on needle appearance as a high intensity line of pixels, require changes to the current workflow or are specific to linear transducers. A clinically suitable approach for curvilinear transducers to detect a steep needle when it is invisible or partially visible, is still needed.

The summary of the sample literature is in Table 1. Recently, the idea of analyzing the natural hand tremor motion to localize a handheld needle in ultrasound was proposed in [17, 18]. In [17], a multi-scale spatial decomposition followed by a temporal filtering was used to estimate the motion and detect pixels moving at the tremor frequency. In [18], a block-based approach is used to first select the regions with motion pattern similar to the motion pattern of the block at the puncture site. The time trace of the displacement was then computed along the spatiotemporal linear paths arising from the puncture site and the needle was identified as the path with maximum spectral correlation with the motion pattern of the puncture site. That method works best with curvilinear transducers because the initial portion of the shaft is usually detectable near the puncture site where the beam is perpendicular. Upon further testing *in-vivo*, the individual pixel-based analysis in these methods was found to be sensitive to noise and also resulted in localization errors due to the surrounding tissue that also moved with the needle.

The proposed system is called CASPER: Computer-Aided Segmentation of imPERceptible motion, which is a learning-based framework to track a needle by detecting variations of imperceptible features over time. The state of the art on using tremor motion for needle detection uses absolute motion and hence work well when the transducer is fixed [17,18]. We propose a tracking framework using differential optical flow and spatiotemporal micro-motion features to incorporate neighboring pixels and mitigate the effects of subtle tremor motion of a handheld transducer. Phase-based analysis of motion in complex-valued pyramids is used to extract spatial features as it is more robust to subtle changes

[17]. Our main novelty is incorporating the relative flow analysis and characteristics of the nearby regions in the detection framework. We also introduce a self-supervised tracking approach capable of improving the performance in the subsequent frames using spatial analysis and dynamic training update. Our contributions are: (1) including the surrounding spatiotemporal neighborhood in the analysis of the pixel data, (2) incorporating the direction of the flow field in addition to its magnitude, (3) tracking of the needle during insertion instead of detection at a fixed spatial position, and (4) extension to free-hand imaging: where both the needle and the transducer are handheld. Qualitative and quantitative analysis is performed *in vivo* on porcine subjects.

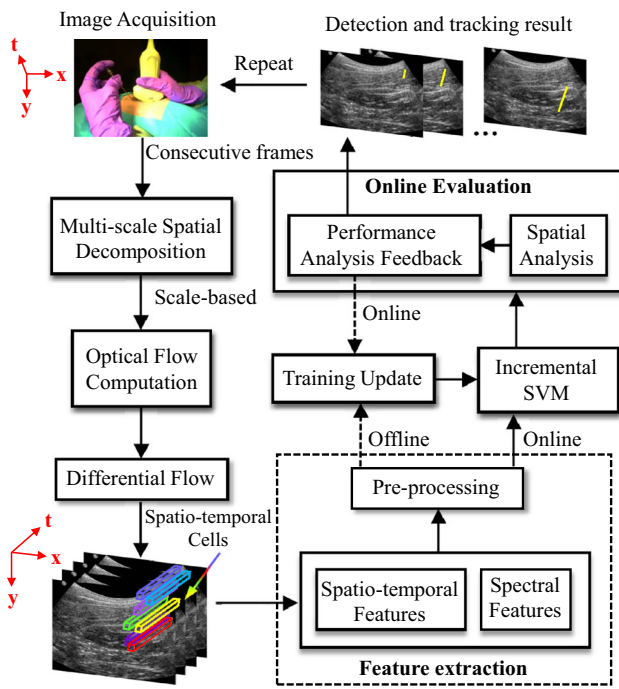
## Methods

### Method overview

An overview of our tracking method is shown in Fig. 1. It consists of three main steps: motion description using phase-based analysis and optical flow, spatiotemporal and spectral feature extraction from the micro-motion in cuboids, and needle tracking using incremental support vector machine (SVM). Steerable pyramids with oriented Gabor filter banks are designed in the range of insertion angles to isolate motion mainly at the orientation of the needle. The differential flow map of the magnitude-weighted phase of consecutive frames is computed from the optical flow analysis. Spatiotemporal and spectral features are then extracted for cells surrounding each pixel, and the feature vector is sent to an incremental SVM for classification. Classified pixels are analyzed based on the morphological estimate obtained from spatial analysis of the labels and their position. Mislabeled data are added as new training example and the model is updated, to enhance the prediction for subsequent frames. The tracking procedure continues, while the model as well as the classification results are updated iteratively. The details of the method will be described in the following sections.

### Motion descriptors

Needle insertion involves micro-motion of the imperceptible needle features and the tissue surrounding the needle.



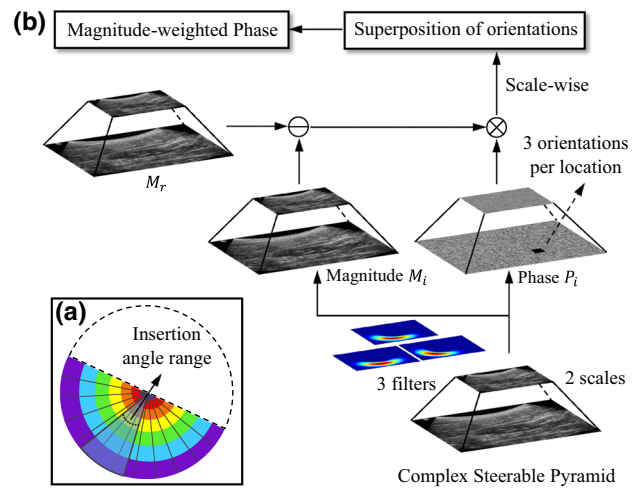
**Fig. 1** Block diagram of the proposed approach: Complex steerable pyramids are used for spatial decomposition, and motion descriptors are computed at each spatial scale. Spatiotemporal and spectral features are extracted within the spatiotemporal cells around each pixel and forwarded to an incremental SVM for classification. Classification results are spatially analyzed for continuity of the positive needle pixels and false-positive non-needle pixels are fed into the adaptive training for online learning. Detection result is overlaid on the current frame, and the procedure is repeated for the subsequent frames

Although the needle motion may not be perceptible in the B-mode data, it could be extracted from the analysis of low-level motion features. Since pixel-based motion descriptors are more sensitive to noise, we introduce an efficient feature representation from low-level motion descriptors in cuboids of pixel.

*Multi-scale spatial decomposition*

Being more robust to subtle intensity changes than magnitude, phase-based analysis of motion could be used to extract such micro-motion [17]. According to the shift property of the Fourier transform, the displacement in time/space induces a phase shift proportional to the displacement and frequency. Based on Fourier series, to extract motion, the displaced intensity profile ( $I$ ) of a frame in a B-mode sequence at time  $t$  and spatial position  $(x, y)$  can be decomposed into its spatial sub-bands  $sb_{\omega}(x, y, t) = A_{\omega}e^{i\omega(x+y+\Delta_{xy}(t))}$  as follows:

$$I(x + \Delta_x(t), y + \Delta_y(t)) \triangleq \sum_k A_k e^{i2\pi k(x+y+\Delta_{xy}(t))} \quad (1)$$



**Fig. 2** Multi-scale spatial decomposition using oriented Gabor filters. **a** Half-octave bandwidth filters: three filters in the insertion angle range. **b** Multi-scale phase-based decomposition of each frame ( $i$ ) according to the reference frame ( $r$ )

where each band contains a complex sinusoid at spatial frequency  $\omega = 2\pi k$ .

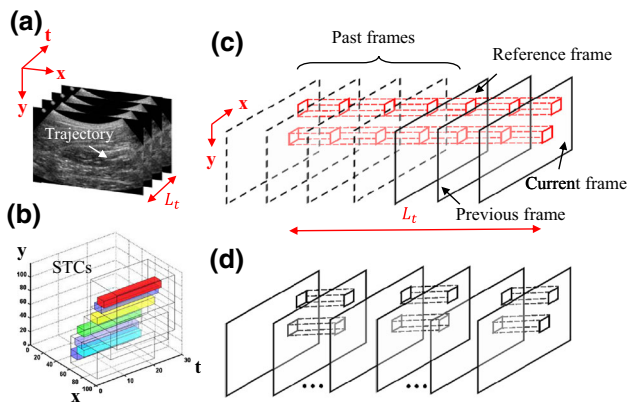
Upon testing, the phase of each sub-band, denoted by  $P_{\omega}(x, y, t) = \omega(x + y + \Delta_{xy}(x, y, t))$ , was found to be more robust, than intensity, to compute the motion map. As shown in Fig. 2, a complex steerable pyramid of two scales and three orientations ( $50^{\circ}$ ,  $65^{\circ}$  and  $80^{\circ}$ ) is used to decompose the signal into its spatial frequency bands. The tuned complex steerable pyramid, isolates motion mainly at the angle of interest, i.e., the expected range of insertion angles. To further reduce noise, phase responses are attenuated at stationary regions where intensity variation is low. This is obtained by computing the magnitude-weighted phase with respect to the reference frame as shown in Fig. 2b.

*Optical flow maps*

To characterize moving regions, we begin by computing the optical flow of the superposition of magnitude-weighted phase responses  $P_{\omega}$  at multiple orientations and scales, for consecutive frames:

$$I_t + uI_x + vI_y = 0 \quad (2)$$

where  $u = dx/dt$  and  $v = dy/dt$  are the lateral and axial components of the optical flow and  $(I_x, I_y)$  and  $I_t$  are spatial and temporal gradients of the phase images with respect to position  $(x, y)$  and time  $t$ , respectively. Spatial gradients are approximated in the lateral and axial directions of the phase image using the first derivative of Gaussian masks along the rows and columns of the phases of the filtered frames  $I_{x/y}(x, y, t) = I(x, y, t) \otimes g_{x/y}$ . The temporal gradient is simply estimated as the difference of the Gaussian



**Fig. 3** **a** Captured frames containing the needle, *arrow* is perpendicular to the invisible needle shaft. **b** Several spatiotemporal cells (STCs) constructed for the captured frames. **c** Frame selection for spatiotemporal analysis at the current frame, and the spatiotemporal cells for each candidate pixel (two candidate pixels in this example). **d** Optical flow is computed for each pair of the previous consecutive frames

smoothed images at the consecutive frames  $I_t(x, y, t) = I(x, y, t) \otimes g - I(x, y, t - 1) \otimes g$ .

To estimate the flow parameters  $(u, v)$ , we use the regularized Lucas–Kanade approach [18], based on the constant flow constraint on the neighboring pixels within a defined window  $w$ . Regularized least square is then used to find the best estimate by minimizing the error term:

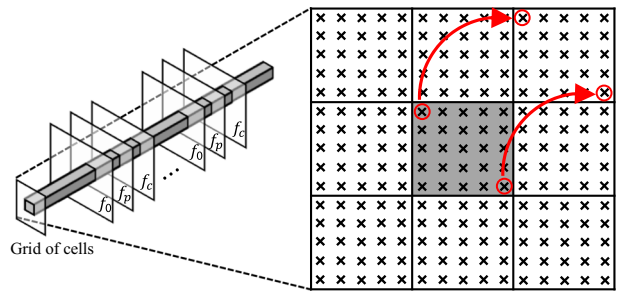
$$E(u, v) = \sum_{\forall i \in w} w(x_i, y_i) \left[ (u, v)^T \cdot \nabla I(x_i, y_i, t) + I_t(x_i, y_i, t) \right]^2 + Tik_c |u, v|^2 \tag{3}$$

where  $\nabla I(x, y, t) = (I_x, I_y)$  and  $w$  is the Gaussian weighting function and  $Tik_c$  is the constant in the Tikhonov regularization term [18].

Given a sequence of  $L_t$  frames up to the current frame  $\{f_1, \dots, f_r, f_p, f_c\}$ , the collection of the optical flow from all consecutive frames is defined as a time dependent flow field  $F$ .

$$F = \{ \langle u(I_1), v(I_1) \rangle, \dots, \langle u(I_r), v(I_r) \rangle, \langle u(I_p), v(I_p) \rangle, \langle u(I_c), v(I_c) \rangle \}. \tag{4}$$

where  $I_r, I_p$ , and  $I_c$  are the reference, previous and current images, respectively. The reference frame is also updated with each new frame and the optical flow is computed for each pair of the previous consecutive frames. Figure 3 describes the frame selection for phase-based analysis and optical flow computation. Features are obtained for several cuboids, called spatiotemporal cells (STCs), constructed around each pixel. STCs are of size  $L_n \times L_n \times L_t$ , in which  $L_n$  is approximately the same size as the needle width and  $L_t$  is the window size for temporal analysis.



**Fig. 4** Differential flow scheme shown for a  $3 \times 3$  grid of cells as an example. The center cell is the spatial cross section of the image with the corresponding STC of size  $5 \times 5 \times L_t$  as an example. *Arrows* represent the diagonal direction of the relative flow

*Differential flow maps*

To use motion analysis for needle detection in US data, captured by a handheld transducer, features are required to characterize needle motion well while being resistant to transducer and intrinsic body motion. The main issue with previous approaches is that, absolute motion detection works best when the transducer is approximately stationary in place and direct analysis of the motion cannot well distinguish the needle from the surrounding tissue moving with the needle.

We introduce differential flow features to better represent the needle in an US image, captured by a handheld transducer. While spatial gradients of the flow capture the sharp changes, the relative flow is basically the large-scale spatial differences at various locations. In more detail, the tremor motion on a handheld transducer is globally distributed along the image with the additive motion vector  $(u_{probe}, v_{probe})$ . Depending on the image intensity, a scale of the transducer’s motion vector is added to the optical flow field of each pixel. Therefore, assuming that the intensity is relatively uniform around the needle (invisible needle in the tissue), differential flow computation cancels out most of the transducer’s tremor effect.

Differential flow is computed for  $L_g \times L_g$  grid of cells, centered at each pixel. As shown in Fig. 4, in each of the surrounding cells, the relative flow is computed for each pixel, as the flow difference with respect to the corresponding pixel in the center cell. The spatial size of the cell is defined to approximately match the needle width, therefore the flow differences relative to the neighboring tissue mainly detects the motion due to insertion.

**Feature extraction**

To represent the micro-motion due to needle insertion, we introduce spatiotemporal and spectral features, extracted from the optical flow and differential flow maps of the flow field  $F$  in the constructed STCs. Features are obtained for

each scale separately and the classification results on both scale outputs are superimposed to get the final result. Care is taken in selecting the STC temporal dimension to ensure that the window size includes a complete cycle for temporal analysis. The computed flow field of the last  $L_t$  frames are stored in a buffer, where the oldest component is replaced by the flow field of the previous frame, with parsing of each new frame.

*Spatiotemporal feature selection*

Spatiotemporal features are first extracted from the sequence of phase images. At every pixel within the STC of the candidate pixel, the spatiotemporal motion descriptors of optical flow ( $u, v$ ), and differential flows ( $\Delta u, \Delta v$ ) are computed. The spatiotemporal features, introduced to detect the needle, are to address the effects of tremor motion induced on a handheld transducer, detect the needle from neighboring tissue more accurately and distinguish needle motion from the intrinsic body motion more robustly.

The optical flow parameters are represented in polar coordinate with  $A_{OF} = \sqrt{u^2 + v^2}$  and  $\theta_{OF} = \arctan(u/v)$ . A set of seven features are used for every scale of the sequence. Our first pair of features  $f_{1,2} = (\bar{A}_{OF}, skewness(\theta_{OF}))$  is the average of magnitude and the skewness of phase of the superposition of the optical flow along the temporal length  $L_t$  of an STC. For all pixels within an STC, we also compute the average of the optical flow field along  $L_t$ . Magnitudes of the average flow vectors for all pixels are concatenated, normalized, and histogrammed for the distribution of the average optical flow magnitudes in an STC. Statistical moments, median and skewness of the histogram of magnitude of the average optical flow then form  $f_{3,4}$ .

Pixel-wise average of differential flow field is computed along  $L_t$ , for all outer cells within the grid around each STC. The average differential flow for each pixel is described by  $A_{\Delta OF_i} = \sqrt{\Delta u_i^2 + \Delta v_i^2}$  and  $\theta_{\Delta OF_i} = \arctan(\Delta u_i/\Delta v_i)$ ,  $\forall i \in grid$ .  $A_{\Delta OF_i}$  values are concatenated for all corresponding pixels within all of the outer cells  $A_{tot}$  (Fig. 4). Auto-correlation function is then computed for  $A_{tot}$  obtained from the concatenated amplitude vectors, to find the motion patterns of pixels within the grid. The lag  $t$  auto-correlation function is defined as:

$$r_t(A_{tot}) = \frac{\sum_{s=1}^{T-t} (A_{tot}(s) - \bar{A}_{tot}) (A_{tot}(s+t) - \bar{A}_{tot})}{\sum_{s=1}^T (A_{tot}(s) - \bar{A}_{tot})^2} \tag{5}$$

Auto-correlation function is computed for every lag and the median of the auto-correlation value of the lags is used as  $f_5$ . The skewness of the differential angle  $\theta_{\Delta OF_i}$  is computed, to (1) extract the direction of movement of the neighboring

cells and (2) distinguish the needle using its different angular pattern. Skewness of  $\bar{\theta}_{\Delta OF_i}$  is used as  $f_6$ .

*Spectral feature selection*

The trace of the optical flow ( $F$ ) from all consecutive frames within an STC is computed for all pixels within the STC. The key concept is that pixels moving together has greater spectral correlation at the corresponding frequency of motion. Therefore considering the fact that STC width is approximately the same as the needle width, if the candidate pixel (center of the STC) is on the needle, pixels within the corresponding STC have similar spectral coherency pattern especially around the tremor frequency [18].

The flow field along a sequence of frames contains variations resulting from insertion and intrinsic body motion, and due to their constant statistical parameters over time, the sequence is considered a stationary process. Spectral coherence is obtained between the flow field of the center pixel  $F_0 = F(p_0)$  at STC versus the flow fields of all the neighboring pixels  $F_i = F(p_i)$  within the STC as follows:

$$C_{p_0 p_i}(f) = \frac{|PSD(F_0 F_i)|^2}{PSD(F_0) PSD(F_i)}, \quad \forall p_i \in \mathbf{STC} - \{p_0\} \tag{6}$$

where  $PSD(p_0)$ ,  $PSD(p_i)$  and  $PSD(p_0 p_i)$  are the power spectral densities (PSD) (*Fourier transform of the auto-correlation*) of the flow field of  $p_0$  and  $p_i$  and their cross PSD. Spectral coherence is computed for the magnitude of the flow field between the center pixel and all other pixels in the STC. The frequency component is quantized, the weighted votes for coherence components are computed from the spectral coherence, and locally histogrammed to produce the feature vectors. The accumulated value at each bin is normalized based on the maximum value at the corresponding bin. Low frequency components contain the insertion and the intrinsic frequency components we need for the analysis [18]. Skewness of the first five components of the normalized histogram are then used as  $f_7$ .

**Pixel-based classification**

SVM was chosen as the classifier due to its easy integration of the hand-crafted features, and its use of kernels to modify the feature space. In addition, it is formulated as a convex optimization, so that a tractable computation is obtained using the unique solution. SVM is chosen at two steps in our framework, it is first used to train the initial model offline using our training data. The initial trained model is then used in the online training to classify the unseen dataset and update the

trained model. Needle and background pixels are represented as +1 and −1 observations in the binary classifier.

Incremental learning is considered as an online method to update the model by adding one example to an existing solution at a time. The key is to update the weights to keep the Kuhn–Tucker conditions satisfied on the enlarged dataset in addition to the previous examples [19]. In detail, given an observation  $f \in \mathbb{R}^m$  and a mapping function  $\Phi$ , an SVM discriminant function is given by:

$$\langle \omega, \Phi(f) \rangle + b \quad (7)$$

where  $\langle \cdot \rangle$  is the inner product operator and  $(\omega, b)$  are the linear separator parameters. The weight vector  $\omega$  can be written as linear combination of the training examples ( $\omega = \sum_{i=1}^n \alpha_i l_i \Phi(f_i)$ ), Eq. (7) is written as:

$$\sum_{i=1}^n \alpha_i l_i K(f_i, f) + b \quad (8)$$

The optimal discriminant parameters are chosen in order to maximize the margin  $\Gamma = \frac{1}{\|\omega\|}$ , the distance between the hyperplane and the closest training vectors  $f_i$ . Minimizing is usually formulated in dual quadratic form using weightings  $\alpha$  and offset  $b$  (Lagrange multipliers) as follows:

$$\min_{\alpha_i, b} : \omega = \sum_i \alpha_i \left( \frac{1}{2} \sum_j l_i l_j K(f_i, f_j) \alpha_j + b l_i - 1 \right), \quad (9)$$

subject to  $0 \leq \alpha_i \leq C$

where  $f$  is the  $m$ -dimensional feature vector for  $n$  pixels,  $l \in \mathbb{R}^n$  is the label vector and  $b$  determines the offset. The radial basis function is used as the kernel  $K$  due to its ease of initialization (requiring only one parameter) and classification accuracy for nonlinear patterns:  $K(f_i, f_j) = e^{-\Gamma \|f_i - f_j\|^2}$ .

The saddle point of Eq. (9) is given by Kuhn–Tucker conditions:

$$g_i = \frac{\partial \omega}{\partial \alpha_i} = \begin{cases} \alpha_j = 0 \Rightarrow |g_i| \geq 1 \\ 0 < \alpha_j < C \Rightarrow |g_i| = 1 \\ \alpha_j = C \Rightarrow |g_i| \leq 1 \end{cases} \quad (10)$$

During online training, the margin vector coefficients change value at each incremental step to keep Kuhn–Tucker conditions satisfied for all examples in the updated training set. The optimum values for the classifier's parameters  $C$  and  $\gamma$ , the regularization term and the inverse of the RBF variance are obtained using cross-validation over the initial training set and grid-search.

## Online evaluation

We analyze the spatial distribution of the classification results of at each iteration. Our developed self-supervising step involves a voting procedure to automatically determine the cluster of points belonging to the needle. Due to the fact that the imaged tissue does not change drastically in consecutive frames, the estimation is improved for each new frame with addition of the new classification result. The aim of the spatial distribution and online update is also to account for false positives such as reverberation artifacts, copies of the needle with similar motion pattern as the needle. These are less likely to be detected with longer training and are specific to a sequence, which could be used in a self-supervisory framework to enhance the localization within the sequence.

### Spatial distribution analysis and online update

The spatial distribution analysis and online update are performed in an iterative approach (Fig. 5). Considering the continuity of the needle, we use a parametric representation of a line  $\rho = x \cos(\varphi) + y \sin(\varphi)$ , to represent the line using parameters  $\rho$  and  $\varphi$ , where  $\rho$  is the orthogonal distance from the origin to the line and  $\varphi$  is the angle between the line trajectory and x-axis.  $\varphi$  is computed for all lines formed by pixels classified as +1 to identify the lines' angles. A needle cluster is selected from the histogram analysis of the populated needle pixels contributing to lines with angle  $\varphi$  within the insertion range ( $50^\circ$ – $80^\circ$ ). Outliers are further removed by histogram analysis of the  $\rho$  values to keep the lines with more populated  $\rho$ . Nearby +1 pixels within the  $L_n$  distance of the needle cluster and  $L_n$  distance from the selected lines are considered as true positives. All other +1-classified pixels farther from the cluster are considered as false positives and are added to the dynamic training set for online training. During online training, a new false-positive sample  $f_c$  is first added with initial weight  $\alpha_c = 0$ . If  $f_c$  is supposed to be a support vector, however, all the weights change value at incremental steps to keep Kuhn–Tucker condition satisfied for all examples.

## Experimental analysis and set-up

Ultrasound images were obtained using an iU22 ultrasound machine and a handheld C5–1 (1–5 MHz) curvilinear transducer (Philips Ultrasound, Bothell, WA, USA). The acoustic and imaging parameters were kept constant suitable for general abdominal imaging. The imaging depth was varied within 50–70 mm and the insertion depth was within  $50^\circ$ – $80^\circ$ . A standard 17 gauge Tuohy epidural needle (Arrow International, Reading, PA, USA) was used for insertion. The porcine trial was conducted at Jack Bell Animal Research Facility, Vancouver (UBC animal care #A14 – 0171). 60

**Fig. 5** Summary of the spatial distribution analysis and online update.  $I$  is the input image,  $fr$  is the frame number,  $p_i$  is pixel  $i$  and  $f^{s1}/f^{s2}$ : feature vectors at both scales.  $I_{tot}$  is the classifier's output binary image that gets updated at each iteration.  $p_{fp}$  is the false-positive pixel set, i.e., non-needle pixel misclassified as needle. Outlier is defined based on the histogram analysis described in Sect. 2.5.1

1. Initialize  $I_{tot} = 0_{size(I)}$ ,  $fr = 0$ ,  $pixel = \{\}$
2. **repeat**
3.    $\{C_{-1}, C_{+1}\} \leftarrow SVM_{test}(f_r, f^{s1}, f^{s2})$
4.    $I_{tot}(C_{+1}) \leftarrow +1$
5.    $l \leftarrow possible\ lines\ in\ I_{tot}$
6.    $l_o \leftarrow outliers(l, \rho, \varphi)$
7.    $l \leftarrow l - l_o$  (remove the outliers)
8.    $\forall p_i \in I_{tot} > 0, \forall l_j \in l: d_i \leftarrow distance(p_i, l_j)$
9.    $l \leftarrow \{l_j\ with\ d_i < L_n\}$  (update the line set)
10.    $p_{fp} \leftarrow p_i \in C_{+1} \ \& \ p_i \notin l$
11.    $I_{tot}(p_i \in l) \leftarrow +1$  (update the classification output)
12.    $(P, \theta) \leftarrow lines(I_{tot})$
13.    $SVM_{train}(f^{s1}(p_{fp}), f^{s2}(p_{fp}), y(p_{fp})) \leftarrow -1$  (re-label false positives)
14.    $fr \leftarrow fr + 1$
15. **until**  $fr \leq fr_{max}$

sequences of US images were captured for independent insertions in the biceps femoris muscle of an anesthetized pig. For a realistic scenario, a pulsating vessel was present in the field of view when possible.

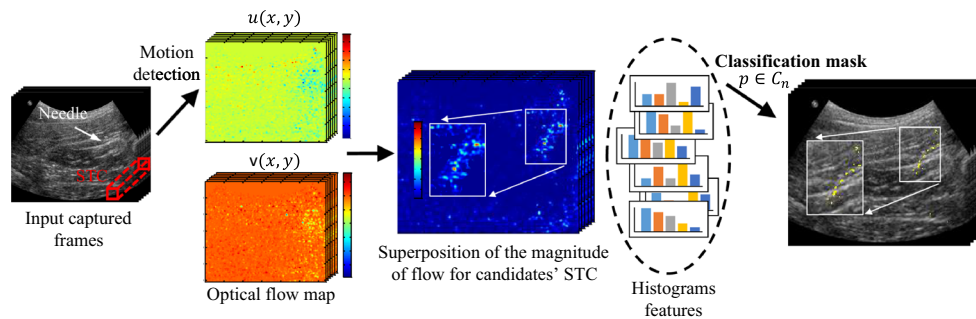
Based on the pixel resolution, width of the 17-gauge needle (1.149 mm), and different spatial scales, the histogram of the features was obtained for cross-validation dataset, while varying  $L_n$  and  $L_g$  in the range of 0.2–1.2 mm. We empirically found the appropriate values of  $L_n \approx 0.9$  mm and  $L_g \approx 0.5$  mm for the analysis. The gold standard was defined by the expert as a line passing through the needle axis, used to select observations for the training set. A channel with the same width as the needle was computed around the segmented trajectory for performance evaluation of the test set. For Hough transform analysis, 2 mm was chosen as the minimum connectivity of regions to form a line. Also, since each classification result contains the data from the past  $L_t$  frames, and considering the maximum insertion rate of 2 cm per second, we use the maximum connectivity of 20 mm to connect the line segments with the same angle. Care is taken to ensure that the window size for spectral analysis includes at least one complete cycle of the heartbeat and tremor, therefore  $L_t$  is made equal to the ultrasound frame rate. The histogram analysis for outlier removal was performed by obtaining the histograms of  $\varphi$  and  $\rho$  values of the lines and keeping the lines contributing to the most populated three bins.

To evaluate the performance, the needle trajectory calculated by the proposed method is compared against the gold standard manually annotated by a sonographer with 30 years of experience. The method performance was evaluated in terms of shaft and tip detection errors. Shaft accuracy was obtained according to the angular deviation between the needle direction calculated by the algorithm and the true needle direction of the gold standard  $\Delta\theta$ . Tip accuracy was obtained

by calculating the Euclidean distance between the needle tip at the closest channel boundary at the tip depth, and the detected trajectory  $\Delta p$ .

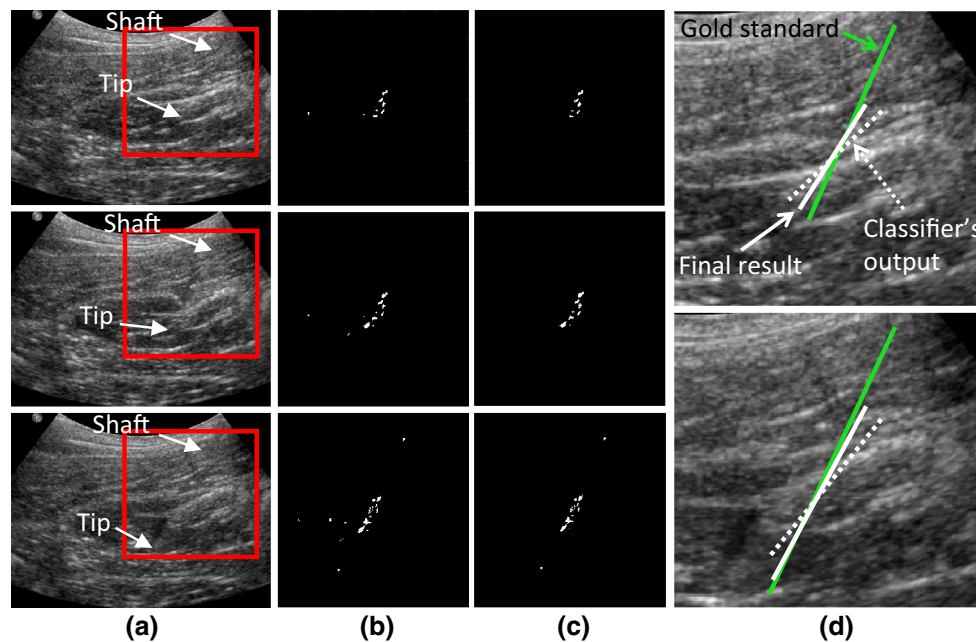
The needle is detected at each frame in the incremental training framework and the classification results are enhanced for the subsequent frames based on the spatial analysis. For the sake of validation, all insertions were made in plane such that the needle tip was as visible as possible. In many cases, however, the gold standard could not be obtained from the static image, and the expert had to analyze the sequence to label the needle based on moving regions. To annotate data for the offline training set, needle pixels with significant motion pattern and matching the gold standard are annotated as +1. This ensures that the neighboring pixels, within the STC of a pixel, are also a part of the needle, which aims to increase the classification accuracy. The expert also records their confidence level of the gold standard selection. For the initial offline training data, 30% of the sequences were randomly selected from images with high confidence (92% of the entire data). Over 10 permutations, the remaining 70% of the sequences were grouped into cross-validation and test sets. 10% of the data were used in cross-validation tests to find the optimum values for SVM hyper-parameters  $C$  and  $\gamma$ . The remaining 60% of image sequences were used in the online evaluation. Note that several pixels were obtained from each training image, and each image contributed to over 160 training samples to form the initial offline training dataset of 3000 observations. To have balanced classes, background pixels were randomly selected equal to the number of needle pixels for each training dataset, and repeated 10 times.

The method was implemented in MATLAB on a 4 GHz processor and 16 GB RAM. Total computation time is 1.18 s for each frame on average: multi-scale spatial decomposition and optical flow computation takes 0.53 s, feature selec-



**Fig. 6** Summary of the steps in the classification pipeline. Initial sequence of frames are shown on the left side, where a white arrow points to the (invisible) needle trajectory. Optical flow is computed and

the superposition of their magnitude is shown for the selected spatiotemporal cells. Finally, the classification result of the incremental training is shown as an overlay on each frame



**Fig. 7** Localization results shown for three subsequent frames. **a** B-mode ultrasound image, with two arrows pointing to the needle tip and the shaft locations. **b** and **c** zoom in of the classifier's output and the final output of the algorithm, respectively. **d** Gold standard (green line),

the Hough transform estimated trajectory from the classifier's output (dashed line) and the algorithm's output (white solid line) overlaid on two corresponding sample frames

tion step takes 0.17 s for both scales, SVM evaluation takes 0.003 s, and spatial distribution analysis and online training takes 0.01 and 0.46 s, respectively.

## Results

Figure 6 shows a summary of the classification pipeline. The flow map is computed for frames within an STC with respect to the reference frame, and their flow magnitude is superimposed to create a coarse motion mask. The features vector is computed for candidate pixels and sent to the SVM classifier to determine the needle pixels. The classification results

are then sent to the spatial distribution analysis and online training update to determine the needle location. As shown in Fig. 7, the spatial distribution analysis and online learning, removes the outliers and improves the classification results at each frame. The needle cluster is enhanced with respect to the final result of the previous frame and therefore both the classifier's output and the final result are improved in the new frame. The localization accuracy is further enhanced in subsequent frames when the needle cluster grows further and estimates the true trajectory more accurately.

Table 2 describes the accuracy of the detected needle using the mean, standard deviation (SD) and root-mean-square (RMS) of the error. Results are evaluated for angular devia-



**Table 2** Comparison of the needle localization results on porcine femoris muscle *in vivo*

Method	Success	$\overline{\Delta\theta}$	$\sigma_{\Delta\theta}$	RMS( $\Delta\theta$ )	$\overline{\Delta p}$	$\sigma_{\Delta p}$	RMS( $\Delta p$ )
CASPER	100%	1.28°	1.09°	1.68°	0.82 mm	1.21 mm	1.47 mm
Offline CASPER	100%	1.81°	1.34°	2.25°	1.53 mm	1.43 mm	2.09 mm
Method of [18]	84%	2.36°	1.57°	2.84°	1.78 mm	1.29 mm	2.20 mm
Hough	14%	0.93°	0.33°	0.99°	1.42 mm	0.67 mm	1.57 mm

tion against the gold standard  $\Delta\theta$ , and the tip offset from the detected trajectory  $\Delta p$ . Results are averaged over 36 sequences of images used during each online evaluation as well as the involved permutations.

We compared our method with respect to the state-of-the-art needle detection method based on the tremor motion [18] and an appearance-based detection method. As shown in Table 2, 16% of the data required manual identification of the initial portion of the shaft for [18], as the automatic method (presented in [18]), failed to select it correctly. For a fair comparison of fully automatic methods, these are ignored and 84% success rate was reported for [18]. Our proposed method outperforms the previous approach with statistically significant improvement in both angular ( $P < 0.0005$ ) and tip deviation ( $P < 0.002$ ), obtained using Mann–Whitney–Wilcoxon test. The method is also compared against an appearance-based approach relying on visible needle features, by applying a tuned Hough transform at the insertion angle. The Hough transform-based method was tested on all images and the error was reported only for cases with angular deviation  $\Delta\theta \leq 10^\circ$ . Although the average errors of shaft angle and tip were 0.93° and 1.42 mm respectively, the method only succeeded in five cases where the needle trajectory was totally visible in the static image. This shows the challenge of needle localization based on intensity features only.

Offline CASPER accounts for the classifier’s output (formulated by Hough), just before spatial distribution analysis and online learning update. As summarized in Table 3, comparison of offline CASPER against [18] shows the importance of feature selection, i.e., the spatiotemporal neighborhood and the direction of the flow. Comparison against CASPER shows the importance of spatial analysis and online learning update in the overall performance. Offline CASPER versus [18] shows highly significant improvement ( $P < 0.01$ ) in terms of the angle accuracy and significant improvement ( $P < 0.05$ ) in terms of the tip accuracy. CASPER versus offline CASPER shows highly significant improvement in both angle and tip, confirming the role of the online update.

Figure 8 demonstrates the performance of CASPER versus the method of [18] for three different frames. Frames were purposefully selected from the sequence where only tremor was present, i.e., for a window of time when the needle was

**Table 3**  $P$  values of Mann–Whitney–Wilcoxon test on the results of offline CASPER against CASPER and the method of [18]

	CASPER	Method of [18]
Angle	1.90E–03	0.0093
Tip	2.52E–03	0.0171

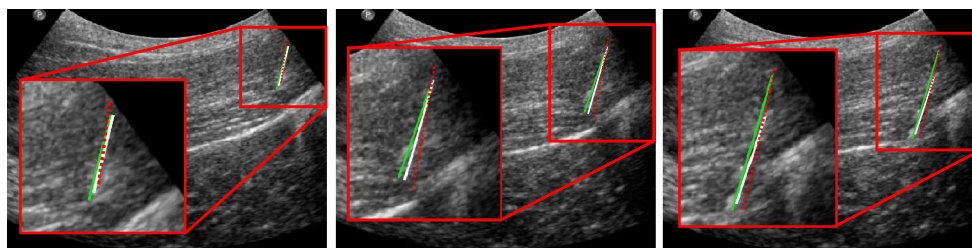
stationary and held by hand. Windows of previous frames were concatenated to produce a longer time span. The localization result of [18] is relatively close to that of CASPER initially, but it deviates further as the insertion in progressing. This is mostly due to the fact that CASPER analysis is performed on each new frame relative to the previous two frames, therefore changes during insertion are detected more reliably compared to the method in [18].

## Discussion and conclusion

We have proposed a needle tracking approach for ultrasound-guided interventions based on novel spatiotemporal features and incremental training. Differential optical flow and spatiotemporal features, incorporate neighboring pixels and could mitigate the effects of subtle tremor motion of a handheld transducer. Micro-motion descriptors are computed from the magnitude-weighted phase of the spatially decomposed data using Gabor filters. The self-supervised tracking framework improves the performance in the subsequent frames and updates the adaptive training dataset.

Note that any method based on motion detection in general requires relatively small variations from frame-to-frame for its best performance. In this work, several steps were taken to mitigate the potential effects of other sources of motion in the detection. The tuned oriented filters in spatial decomposition, the specific frequency channels in spectral features computation, and the spatial distribution analysis in the final step, all aim to mitigate the effects of other sources such as intrinsic body motion.

In addition to the accuracy, online learning is especially helpful for tracking, where new samples (i.e., detected false positives) are adaptively added to the training, with each new frame. Investigation of decremental learning could be inter-



**Fig. 8** Localization results shown for three frames. CASPER output (*white solid line*), the method of [18] (*red dashed line*) and gold standard (*green line*) are overlaid on the B-mode image

esting, by adjusting the false positives added to the model, iteratively.

The current clinical demand is inclined toward 2D ultrasound imaging with standard needle/apparatus. Our focus in this study was specifically on curved array transducers as they are more challenging, with fewer solutions compared to linear array transducers which benefit from beam steering. Unlike the previous works on detecting tremor for needle detection, the proposed method relaxes the strict requirement of a portion of visible needle near the insertion site that is typical of curvilinear transducers, so it could possibly work on other transducer geometries as well.

**Acknowledgements** This work is jointly funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR). Thanks to Philips Ultrasound for supplying the ultrasound machine and research interface.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

## References

- Matalon TA, Silver B (1990) US guidance of interventional procedures. *Radiology* 174(1):43–47
- Chin Ki Jinn, Perlas Anahi, Chan Vincent W S, Brull Richard (2008) Needle visualization in ultrasound-guided regional anesthesia: challenges and solutions. *Reg Anesthesia and Pain Med* 33(6):532–544
- Ayvali E, Desai JP (2015) Optical flow-based tracking of needles and needle-tip localization using circular hough transform in ultrasound images. *Ann Biomed Eng* 43(8):1828–1840
- Uhrcik M, Kybic J, Liebgott H, Cachard C (2010) Model fitting using ransac for surgical tool localization in 3D ultrasound images. *IEEE Trans Biomed Eng* 57(8):1907–1916
- Zhao Y, Cachard C, Liebgott H (2013) Automatic needle detection and tracking in 3d ultrasound using an ROI-based RANSAC and Kalman method. *Ultrasonic imaging* 35(4):283–306
- Ding M, Fenster A (2004) Projection-based needle segmentation in 3D ultrasound images. *Comput aided surg : int Soc Comput Aided Surg* 9(5):193–201
- Wu Q, Yuchi M, Ding M (2014) Phase grouping-based needle segmentation in 3d trans-rectal ultrasound-guided prostate transperineal therapy. *Ultrasound in Med Biol* 40(4):804–816
- Beigi P, Salcudean T, Rohling RN, Lessoway VA, Ng GC (2015) Needle trajectory and tip localization in 3D ultrasound using a moving stylus. *Ultrasound in Med Biol* 41(7):2057–2070
- Zhao Y, Shen Y, Bernard A, Cachard C, Liebgott H (2017) Evaluation and comparison of current biopsy needle localization and tracking methods using 3D ultrasound. *Ultrasonics* 73:206–220
- Adebar TK, Fletcher AE, Okamura AM (2014) 3D ultrasound-guided robotic needle steering in biological tissue. *IEEE Trans Biomed Eng* 61(12):2899–2910
- Hakime A, Deschamps F, De Carvalho EGM, Barah A, Auperin A, De Baere T (2012) Electromagnetic-tracked biopsy under ultrasound guidance: preliminary results. *Cardiovasc interv radiol* 35(4):898–905
- Swiggum HP, Ahn K, Dilger JA, Smith HM (2013) Needle echogenicity in sonographically guided regional anesthesia: blinded comparison of 4 enhanced needles and validation of visual criteria for evaluation. *J Ultrasound in Med* 32(1):143–148
- Kim C, Chang D, Petrisor D, Chirikjian G, Han M, Stoinovici D (2013) Ultrasound probe and needle-guide calibration for robotic ultrasound scanning and needle targeting. *IEEE Trans Biomed Eng* 60(6):1728–1734
- Cowan NJ, Goldberg K, Chirikjian GS, Fichtinger G, Alterovitz R, Reed KB, Kallem V, Park W, Misra S, Okamura AM (2011) Robotic needle steering: Design, modeling, planning, and image guidance. *Surgical robotics*. Springer, New York, pp 557–582
- Boctor EM, Choti MA, Burdette EC, Webster RJ (2008) Three-dimensional ultrasound-guided robotic needle placement: an experimental evaluation. *Int J Med Robot Comput Assist Surg* 4(2):180–91
- Stephanie Cheung, Robert Rohling (2004) Enhancement of needle visibility in ultrasound-guided percutaneous procedures. *Ultrasound in Med Biol* 30(5):617–624
- Beigi P, Salcudean T, Rohling RN, Lessoway VA, Ng GC (2016) Automatic detection of a hand-held needle in ultrasound via phased-based analysis of the tremor motion. In: *SPIE medical imaging*, pp 97860I-1–97860I-6
- Beigi P, Salcudean T, Rohling RN, Ng GC (2016) Spectral analysis of the tremor motion for needle detection in curvilinear ultrasound via spatio-temporal linear sampling. *Int J Comput Assist Radiol Surg* 11(6):1183–1192
- Poggio T, Cauwenberghs G (2001) Incremental and decremental support vector machine learning. *Adv neural inform process syst* 13:409