



- (51) **International Patent Classification:**
A61B 8/08 (2006.01) A61B 5/06 (2006.01)
A61B 8/12 (2006.01)
- (21) **International Application Number:**
PCT/US2019/046364
- (22) **International Filing Date:**
13 August 2019 (13.08.2019)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/718,090 13 August 2018 (13.08.2018) US
- (71) **Applicant: RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY** [US/US]; 83 Somerset Street, New Brunswick, NJ 08901 (US).
- (72) **Inventors: MWIKIRIZE, Cosmas**; 171 Davidson Road, Apt. 825A, Piscataway, NJ 08854 (US). **NOSHER, John, L.**; 52 Stone Hedge Drive, Basking Ridge, NJ 07920 (US). **HACHALILOGLU, Ilker**; 206 Warren Street, Apt. 1L, Jersey City, NJ 07302 (US).

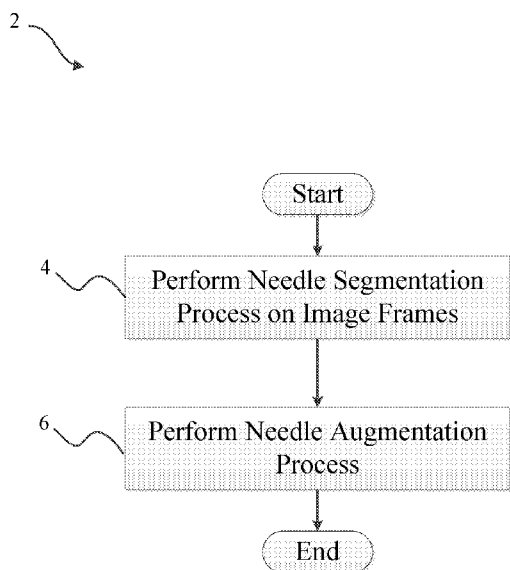
(74) **Agent: FRISCIA, Michael, R.** et al.; McCarter & English, LLP, Four Gateway Center, 100 Mulberry Street, Newark, NJ 07102 (US).

(81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** COMPUTER VISION SYSTEMS AND METHODS FOR REAL-TIME LOCALIZATION OF NEEDLES IN ULTRASOUND IMAGES

FIG. 1



(57) **Abstract:** Computer vision systems and methods for real-time localization of a needle in an ultrasound video are provided. The method includes a step of receiving at a processor a plurality of frames from the ultrasound video. The method also includes the step of identifying by the processor one of the plurality of frames as a background image and a frame adjacent to the background image as a current image. The method further includes the step of performing by the processor a bitwise complement operation on the background image to generate a complement image. Moreover, the method includes the step of performing by the processor a pointwise logical AND operation on the complement image and the current image to identify the location of the needle.



WO 2020/036968 A1

Published:

— *with international search report (Art. 21(3))*

COMPUTER VISION SYSTEMS AND METHODS FOR REAL-TIME LOCALIZATION OF NEEDLES IN ULTRASOUND IMAGES

SPECIFICATION

BACKGROUND

RELATED APPLICATIONS

[0001] This application claims the benefit of priority to United States Provisional Patent Application Serial No. 62/718,090 filed on August 13, 2018, the entire disclosure of which is hereby expressly incorporated by reference.

TECHNICAL FIELD

[0002] The present disclosure relates generally to the field of computer vision technology. More specifically, the present disclosure relates to computer vision systems and methods for real-time localization of needles in ultrasound images.

RELATED ART

[0003] Minimally invasive procedures such as regional anesthesia and interventional oncology involve the insertion of a needle toward target anatomy. In practice, image guidance is used to improve targeting accuracy. Of all imaging modalities, ultrasound imaging is ubiquitously used due to real-time, low-cost and radiation-free capabilities. However, with conventional 2D ultrasounds, aligning the needle with the ultrasound imaging plane at steep angles and deep insertions is difficult.

[0004] Real-time and accurate localization of hand-held needles is vital for the success of percutaneous ultrasound guided interventions such as biopsies and regional anesthesia. However, needle localization is a challenge when the needle shaft and tip are not fully visible. This could arise from difficulty aligning the needle with the scan plane and reflection of the ultrasound signal from the needle away from the transducer. The ensuing targeting errors may reduce procedure efficacy.

[0005] A two-dimensional (2D) ultrasound is the standard imaging modality used for guiding percutaneous interventions in clinics due to its real-time non-radiation-based imaging functionality. In a normal work-flow, a radiologist advances the needle by hand

and manipulates an ultrasound transducer, while concurrently observing needle motion vis-a-vis body anatomy. Prior art systems for improving needle localization during these procedures require the needle shaft and tip to have a high intensity, which is often not the case. Moreover, many prior art solutions involve computationally expensive processing routines. In addition, many of these solutions do not localize bending needles, which is a necessary feature as needles need to bend around critical anatomical structures in some procedures. Some prior art methods are not desirable because they involve computationally complex processing for computing motion vectors and are sensitive to noise and brightness variation. Moreover, prior art solutions cannot localize needles that are not wholly visible.

[0006] Therefore, there is a need for computer vision systems and methods for real-time localization of needles in ultrasound images, which is fast and accurate and can localize needles which cannot be seen. These and other needs are addressed by the computer vision systems and methods of the present disclosure.

SUMMARY

[0007] Computer vision systems and methods for real-time localization of a needle in an ultrasound video are provided. The method includes a step of receiving at a processor a plurality of frames from the ultrasound video. The method also includes the step of identifying by the processor one of the plurality of frames as a background image and a frame adjacent to the background image as a current image. The method further includes the step of performing by the processor a bitwise complement operation on the background image to generate a complement image. Moreover, the method includes the step of performing by the processor a pointwise logical AND operation on the complement image and the current image to identify the location of the needle.

[0008] A system for real-time localization of a needle in an ultrasound video is provided. The system includes a memory having computer instructions stored thereon, which when executed, causes a processor to perform a number of steps. The first step includes receiving a plurality of frames from the ultrasound video. The second step includes identifying one of the plurality of frames as a background image and a frame adjacent to the background image as a current image. The third step includes performing by the processor a bitwise complement operation on the background image to generate a complement image. The fourth step includes performing a pointwise logical AND operation on the complement image and the current image to identify the location of the needle.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The foregoing features of the invention will be apparent from the following Detailed Description of the Invention, taken in connection with the accompanying drawings, in which:

[0010] FIG. 1 is a flowchart illustrating processing steps for localizing a needle in ultrasound images;

[0011] FIG. 2 is a flowchart illustrating processing steps for a performing needle segmentation process of FIG. 1;

[0012] FIG. 3 is a flowchart illustrating processing steps for performing a needle augmentation process of FIG. 1;

[0013] FIG. 4 shows needle localization results for consecutive frames from both in-plane and out-of-plane insertion of a 17 gauge needle;

[0014] FIGS. 5A-5C show needle localization results for in-plane insertion of a 22 gauge needle;

[0015] FIG. 6 is a diagram illustrating hardware and software components capable of implementing the systems and methods of the present disclosure;

[0016] FIG. 7 is a drawing illustrating a system of the present disclosure;

[0017] FIG. 8 is a block diagram illustrating another embodiment of the system and method of the present disclosure, wherein a convolutional neural network localizes a needle tip in ultrasound images;

[0018] FIG. 9 shows needle tip localization results for consecutive frames from in-plane insertion of a 17 gauge needle and out-of-plane insertion of a 22 gauge needle;

[0019] FIG. 10 illustrates elimination by the system of false positives from a trajectory estimation;

[0020] FIG. 11 is a block diagram illustrating another embodiment of the system and method of the present disclosure, wherein a convolutional neural network localizes a

needle tip in ultrasound images;

[0021] FIG. 12 illustrates needle tip enhancement results according to step 244 of FIG. 11;

[0022] FIG. 13 is a block diagram illustrating a convolutional neural network capable of performing needle tip classification according to process 226 of FIG. 11;

[0023] FIG. 14 shows needle tip labels for a needle tip classification neural network and a needle tip regression convolutional neural network; and

[0024] FIG. 15 shows needle tip localization results for insertion of a 17 gauge needle and insertion of a 22 gauge needle.

DETAILED DESCRIPTION

[0025] The present disclosure relates to computer vision systems and methods for real-time localization of needles in ultrasound images as described in detail below in connection with FIGS. 1-15.

[0026] The systems and methods of the present disclosure provide solutions for real-time and accurate localization of needles (even if they are invisible) in an ultrasound scan. The systems and methods use a foreground detection algorithm for automatic needle segmentation from consecutive ultrasound image frames. The needle tip is then localized by solving a spatial total variation regularization problem using, for example, a Split Bregman method. The systems and methods of the present disclosure are not limited to the Split Bregman method as other regularization methods can be used on the data. For example, deep learning and reinforcement learning can be performed on ultrasound datasets to model the appearance of a needle tip in an ultrasound image thereby improving needle tip detection and reducing irrelevant background noise. Also, Split Bregman can be a general categorical formulation of the method, and other applications (e.g., noise model, sub-problem derivation, and solution, parameter optimization) can be specific to the moving device localization problem.

[0027] Although the systems and methods of the present disclosure provide for the localization of a needle, the systems and methods are not limited to localizing needles. In particular, the systems and methods can be used to localize any type of medical device or surgical instrument, including, but not limited to, forceps, stainless steel rods with conical tips, catheters, guidewires, radio frequency ablation electrodes, balloons (during angioplasty/stenting procedures), or any other medical devices. Moreover, although the systems and methods of the present disclosure provide for localization of any medical device in an ultrasound video, any type of imaging modalities can be used. In particular, the systems and methods of the present disclosure can localize any type of medical instrument in any type of video format or file because the logical differencing approach and Split Bregman method is compatible with any type of video.

[0028] The systems and methods of the present disclosure accomplish at least the following goals. First, the systems and methods perform needle segmentation based on a dynamic background model that can account for noise in an ultrasound. Second, the systems and methods can apply the Split-Bregman approach in a unique way to solve a spatial total variation (TV) problem for needle tip enhancement in 2D ultrasounds. Third, the systems and methods can accurately localize a needle tip in a computationally fast way with a high degree of accuracy. Fourth, the systems and methods can localize a needle for in-plane and out-of-plane insertions without the need for full needle visibility. Fifth, the systems and methods can localize bending needles since as there is no need for the needle to appear as a linear feature in the ultrasound image. Sixth, the systems and methods can be utilized in a smart computer assisted interventional system to facilitate needle localization in the presence of artifacts from anatomical features and other instruments.

[0029] The methods and systems of the present disclosure can be used with hand-held 2D ultrasound probes, during in-plane and out-of-plane needle insertion. For in-plane insertion, the method can handle insertion angles greater than 40° , but can also work with shallower angles. The method detects scene changes caused by needle motion in the ultrasound image scene. In each frame of the ultrasound sequence, the needle tip can be treated as the foreground, while the rest of the image can be designated as background data. The method does not require prior knowledge of the needle insertion side or angle. The needle localization process includes two phases: 1) needle segmentation from logical differencing of a dynamic reference ultrasound frame from the ultrasound frame of interest, and 2) needle tip augmentation in the ultrasound image with the segmented needle.

[0030] In a dynamic background model, the background of each frame (k) is modeled as the previous frame (k-1), based on logical differencing. The enhanced image is based on a point-wise operator of the complement of the background and the image (what is NOT in the k-1 frame but in the k-1 frame). This makes it possible to identify the objects that have moved between the two frames: the needle tip. To ensure that movement actually occurs between the two frames, the correlation coefficient of the background and current frame is calculated. The image is only used for tip localization if the difference exceeds a threshold.

[0031] FIG. 1 is a flowchart illustrating processing steps 2 for localizing a needle in an ultrasound video sequence. In step 4, the process performs needle segmentation, which will be explained in greater detail below. In step 6, the process performs needle augmentation which will be explained in greater detail below.

[0032] FIG. 2 is a flowchart illustrating processing steps 4 for performing needle segmentation as discussed in connection with FIG. 1. Consider an ultrasound frame sequence with temporal continuity, represented by the function $p(x, y, t)$, where t denotes the position in the time sequence and (x, y) are the spatial coordinates. The process 4 uses a dynamic background model which quickly adapts to changes in the ultrasound scene by employing a logical differencing approach for adjacent frames. In step 8, for the first frame, the background can be denoted as: $b(x, y, t_0) = p(x, y, t_0)$, and for all subsequent frames, the background is modeled as the previous frame in the sequence (which can be written as $b(x, y, t_n) = p(x, y, t_{n-1})$). In step 10, the method can then calculate the bitwise complement of the background image. Considering only spatial variation, for $b(x, y) = (x, y) \mid b(x, y) \neq 0$, the complement is $b^c(x, y) = (x, y) \in Z^2 \mid (x, y) \notin b(x, y)$. In step 12, the method can ensure that the background and the current frame are different. In step 14, for any current frame $p(x, y)$, the needle enhanced image equation is given by:

$$q(x, y) = b^c(x, y) \wedge p(x, y)$$

Equation 1

where \wedge denotes a point-wise AND logical operation. The needle enhancement equation yields only the objects in the ultrasound that moved between two successive frames, and thus gives an enhanced current needle tip location. The method can account for concurrent tissue movement surrounding the needle tip by taking into account collocated motion of the tissue and tip to be more significant than any other motion. Depending on the needle visibility profile, $q(x, y)$ can also contain an enhanced shaft. Since a short interval exists between successive frames, the method can minimize the effect of intensity changes in the rest of the ultrasound image, which leads to stable needle detection. In some cases, insignificant changes in needle motion will lead to an all zeros $q(x, y)$. This is expected since in typical ultrasound imaging (greater than 20 frames per second), the needle tip will not be at a different location in each frame, even with continuous motion. Before applying

the needle enhancement equation, the method can ensure that $b(x, y)$, the background, and $p(x, y)$, the current frame, are different (as discussed with regard to step 12). Accordingly, the method can calculate the correlation coefficient, p , between the two images, and increment to the next frame while maintaining the current frame position if p exceeds a predetermined threshold.

[0033] The output of the needle enhancement equation in step 14 of FIG. 2, $r(x, y)$, can also contain artifacts that do not belong to the needle. This is expected due to brightness variations in the ultrasound images, and the motion of tissues/organs that do not correspond to needle tip motion. Accordingly, the method of the present disclosure can enhance the output $r(x, y)$ such that the needle tip distinctively has the highest appearance. This can be thought of as an image restoration (de-noising) problem. The method can formulate a model as follows: $r(x, y) = e(x, y) + n(x, y)$. This model can represent a sum of two components: the desired needle enhanced image $e(x, y)$ and unwanted artifacts (noise), $n(x, y)$. The desired needle enhanced component $e(x, y)$ can be a function of bounded variation. Moreover, the images can be represented by vectors. The image restoration model becomes:

$$\mathbf{r} = \mathbf{e} + \mathbf{n}$$

Equation 2

where $\mathbf{e} \in \mathbb{R}^{mn \times 1}$ is the desired enhanced image (of size $m \times n$), $\mathbf{r} \in \mathbb{R}^{mn \times 1}$ is the corrupted image, while $\mathbf{n} \in \mathbb{R}^{mn \times 1}$ is the noise. In this notation, \mathbf{r} , \mathbf{e} , and \mathbf{n} are vectors containing all the pixel values in the respective image matrices in lexicographic order. The vector \mathbf{n} can represent stochastic or deterministic noise. Since pixels in the segmented image have spurious detail and possibly high total variation, the method can formulate a total variation regularization problem in the following form:

$$\min_{\mathbf{e}} \frac{\lambda}{2} \|\mathbf{r} - \mathbf{e}\|_2^2 + \|\mathbf{e}\|_{TV},$$

Equation 3

where λ is a regularization parameter and $\|\mathbf{e}\|_{TV} = \|\mathbf{D}_x \mathbf{e}\|_1 + \|\mathbf{D}_y \mathbf{e}\|_1$ is the anisotropic total variation norm, defined by \mathbf{D}_x and \mathbf{D}_y , which are the spatial first-order

forward finite difference operators along the horizontal and vertical directions respectively. The equation representing the total variation problem above can be a constrained formulation of a non-differentiable optimization problem. This problem can be efficiently solved with the Split Bregman approach, in which the main problem is reduced to a sequence of unconstrained optimization problems and variable updates. We first transform the equation representing the total variation problem into a constrained equivalent problem by introducing intermediate variables \mathbf{v} and \mathbf{w} , in the following equation:

$$\begin{aligned} \min_{\mathbf{v}, \mathbf{w}, \mathbf{e}} \quad & \frac{\lambda}{2} \|\mathbf{r} - \mathbf{e}\|_2^2 + \|\mathbf{v}\|_1 + \|\mathbf{w}\|_1 \\ \text{subject to} \quad & \mathbf{v} = \mathbf{D}_x \mathbf{e} \\ & \mathbf{w} = \mathbf{D}_y \mathbf{e} \end{aligned}$$

Equation 4

The method can then weakly enforce the constraints in the above equation by introducing quadratic penalties, and strictly enforce the constraints by applying the Bregman iteration as follows:

$$\min_{\mathbf{v}, \mathbf{w}, \mathbf{e}} \frac{\lambda}{2} \|\mathbf{r} - \mathbf{e}\|_2^2 + \|\mathbf{v}\|_1 + \|\mathbf{w}\|_1 + \frac{\nu}{2} \|\mathbf{v} - \mathbf{D}_x \mathbf{e} - \mathbf{b}_1\|_2^2 + \frac{\nu}{2} \|\mathbf{w} - \mathbf{D}_y \mathbf{e} - \mathbf{b}_2\|_2^2,$$

Equation 5

where ν is an additional regularization parameter, and the values of \mathbf{b}_1 and \mathbf{b}_2 are determined through Bregman iteration. Now a series of three sub-problems can be written as the following equations:

$$\min_{\mathbf{v}} \|\mathbf{v}\|_1 + \frac{\nu}{2} \|\mathbf{v} - \mathbf{D}_x \mathbf{e} - \mathbf{b}_1\|_2^2$$

Equation 6

$$\min_{\mathbf{w}} \|\mathbf{w}\|_1 + \frac{\nu}{2} \|\mathbf{w} - \mathbf{D}_y \mathbf{e} - \mathbf{b}_2\|_2^2$$

Equation 7

$$\min_{\mathbf{e}} \frac{\lambda}{2} \|\mathbf{r} - \mathbf{e}\|_2^2 + \frac{\nu}{2} \|\mathbf{v} - \mathbf{D}_x \mathbf{e} - \mathbf{b}_1\|_2^2 + \frac{\nu}{2} \|\mathbf{w} - \mathbf{D}_y \mathbf{e} - \mathbf{b}_2\|_2^2$$

Equation 8

Equations 6 and 7 can decouple over space and can have closed-form solutions as vectorial shrinkages given by the following equations:

$$\mathbf{v} = \text{sign}(\mathbf{D}_x \mathbf{e} + \mathbf{b}_1) \times \max\{|\mathbf{D}_x \mathbf{e} + \mathbf{b}_1| - \frac{1}{\nu}, 0\}$$

$$\mathbf{w} = \text{sign}(\mathbf{D}_y \mathbf{e} + \mathbf{b}_2) \times \max\{|\mathbf{D}_y \mathbf{e} + \mathbf{b}_2| - \frac{1}{\nu}, 0\}$$

Equation 9

From the third sub-problem equation 8 above, the following equation can be derived:

$$\mathbf{e} \{\lambda \mathbf{I} - \nu \{\mathbf{D}_x^T \mathbf{D}_x + \mathbf{D}_y^T \mathbf{D}_y\}\} = \lambda \mathbf{r} + \nu \mathbf{D}_x^T (\mathbf{v} - \mathbf{b}_1) + \nu \mathbf{D}_y^T (\mathbf{w} - \mathbf{b}_2)$$

Equation 10

Equation 10 can be solved using an iterative least squares solver by initializing \mathbf{b}_1 and \mathbf{b}_2 to zero and updating these variables after each Bregman iteration: $\mathbf{b}_1^{i+1} = \mathbf{b}_1^i + \mathbf{D}_x \mathbf{e} - \mathbf{v}$, $\mathbf{b}_2^{i+1} = \mathbf{b}_2^i + \mathbf{D}_y \mathbf{e} - \mathbf{w}$.

[0034] FIG. 3 is a flowchart illustrating the process 6 for needle augmentation as discussed above in connection with FIG. 1, in greater detail. The process 6 can be implemented in accordance with the above equations and disclosure discussed above in connection with performing needle enhancement. In step 16, the method can input data \mathbf{r} and parameters λ and ν . In step 18, the method can initialize the following variables: $\mathbf{e}=\mathbf{r}$, $\mathbf{v}=0$, $\mathbf{w}=0$, $\mathbf{b}_1 = 0$, $\mathbf{b}_2=0$. In step 20, the method can compute \mathbf{D}_x and \mathbf{D}_y in accordance with the above equations. In step 22, the method can make a determination as to whether $\|\mathbf{e}_{\text{curr}} - \mathbf{e}_{\text{prev}}\|_2 / \mathbf{e}_{\text{prev}}$ is greater than the *Tol* variable (a threshold). If a positive determination is made, the method can proceed to step 24 in which \mathbf{v} sub-problem, the \mathbf{w} sub-problem and the \mathbf{e} sub-problem are solved, and \mathbf{b}_1 and \mathbf{b}_2 are updated. If a negative determination is made in step 22, the method ends.

[0035] Data acquisition and experimental validation will now be explained in greater detail. 2D B-mode ultrasound data was collected using two imaging systems: SonixGPS (Analogic Corporation, Peabody, MA, USA) with a hand-held C5-2/60 curvilinear probe, and 2D hand-held wireless ultrasound (Clarius C3, Clarius Mobile Health Corporation, Burnaby, British Columbia, Canada). Experiments were performed on a freshly excised bovine tissue, and a porcine shoulder phantom, with insertion of a 17 gauge (1.5 mm diameter, 90 mm length) Tuohy epidural needle (Arrow International, Reading, PA, USA) and a 22 gauge (0.7 mm diameter, 90mm length) spinal Quincke-type needle (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). In-plane insertion was performed at 40° - 70° and the needle was inserted up a depth of 70 mm. 30 (15 in-plane, 15 out-of-plane) sequences of ultrasound images, each containing more than 100 frames were collected.

[0036] The method of the present disclosure was implemented in MATLAB 2017b on a 3.6 GHz Intel(R) Core™ i7 CPU, 16 GB RAM Windows personal computer. The present disclosure is not limited to this personal computer but it can be a good choice when the sub-problems of the method discussed above converge quickly and are numerically well conditioned. The variable ν can, in some cases, be neither be too large or too small. As an example only, $\nu = 2$ can be a good value, but in some cases, the de-noising output may not be sensitive to that specific value. The parameter λ can affect the balance between removing noise and preserving image content. Although, in some cases, it is preferable to use a value that matches the variance of the noise, the noise characteristics of $r(x, y)$ are unknown. Accordingly, a value of $\lambda = 5$ can be used. In some cases, the threshold of the correlation coefficient for skipping a frame due to insignificant motion can be 1. A value of $p > 0.995$ was feasible for most of the ultrasound data. Throughout the validation experiments, these values were not changed. Performance of the method was evaluated by comparing the automatically detected tip location to the ground truth determined by an expert sonographer. Tip localization accuracy was determined from the Euclidean distance between the two measurements.

[0037] The results and analysis will now be explained in greater detail. With regard to quantitative results, the method is computationally efficient, and achieves

convergence in less than 10 iterations, with $Tol = \left| \left| e_{curr} - e_{prev} \right| \right|_2 / e_{prev}$ is less than or equal to 10^{-3} . However, to save computational cost, the enhancement algorithm was run for only 3 iterations (0.06s), and achieves sufficient tip enhancement. The total computation time is 0.08s for each ultrasound frame: 0.01s for segmentation and 0.06s for enhancement and localization. This represents a substantial improvement in simplifying computational complexity of the needle localization problem. Error statistics for in-plane and out-of-plane needle insertion are shown in Table 1 below. Table 1 shows mean error with a 95% confidence interval. The term ‘‘RMS’’ means Root Mean Square. The term ‘‘SD’’ means Standard Deviation. All values shown in Table 1 are in mm. In some cases, localization for in-plane insertions can be more accurate than that for in-plane insertions. The overall tip localization is 0.50 ± 0.02 mm.

Needle insertion	Number of images	Mean	SD	RMS	Maximum
In-plane	650	0.46 ± 0.02	0.27	0.54	0.95
Out-of-plane	380	0.56 ± 0.03	0.29	0.63	1.20
Overall	1030	0.50 ± 0.02	0.28	0.58	1.20

Table 1

[0038] With regard to qualitative results, FIG. 4 shows needle localization results for consecutive frames from both in-plane and out-of-plane insertion of a 17 gauge needle. In both cases, the needle tip is accurately localized. For out-of-plane needles, the temporal window for needle tip visibility is limited, but the present method can account for this limitation by tracking small perturbations of the needle tip close to the target. As can be seen in FIG. 4, needle localization in 3 consecutive frames is shown with rows (A) illustrating in-plane insertion and rows (B) illustrating out-of-plane insertion of the 17 gauge needle. Column (a) illustrates an original image showing an annotated expert-localized tip. Column (b) illustrates a needle enhanced image, $e(x, y)$ (color coded), in which the rectangle surrounds the localized tip. Column (c) illustrates a zoomed in region of interest from column (b) containing the needle tip, and which also shows a colorbar for illustrating where the tip has the highest intensity in the image. Column (d) illustrates the automatically localized tip (shown as red dot) overlaid on the original image. As can be seen, the present method achieves accurate localization for both in-plane and out-of-plane needles.

[0039] FIGS. 5A-5C show needle localization results for in-plane insertion of a 22 gauge needle. Note that in the unaltered image, the needle tip and shaft are inconspicuous because this needle is thinner, and as a result, it reflects less of an ultrasound signal back to the transducer. Despite these challenges, the present method accurately localizes the tip. FIG. 5A shows an original image in which the annotated region contains the needle trajectory. As can be seen, the tip location is difficult to discern by the expert. FIG. 5B shows a needle enhanced image, $e(x, y)$ (color coded) in which a rectangle surrounds the localized tip (the highest intensity in the image). FIG. 5C shows the automatically localized tip (red dot) overlaid on original image.

[0040] The method of the present disclosure provides a novel approach for needle tip localization in 2D ultrasounds suitable for challenging imaging scenarios in which the needle is not continuously visible. The present method provides for the reduction of computational complexity, leading to a processing rate of 12 fps. The present method does not necessitate the needle to appear as a high intensity, continuous linear structure in the ultrasound image. Therefore, both in-plane and out-of-plane needle localization are achieved. A thin 22 gauge needle was used in the experiments to demonstrate the robustness of the present method to localize partially invisible and curved needles. Typically, such thin needles are prone to bending and the shaft has limited visibility, but this problem did not affect the accuracy of tip localization of the present method.

[0041] The method can be applied with the needle already in motion so that subsequent dynamics induce detectable scene changes for the algorithm to compute. The method can also overcome issues of probe re-alignment which might cause abrupt intensity changes unrelated to needle movement. Despite all of the above-mentioned issues, the present method provides a real-time method for needle localization in ultrasounds.

[0042] FIG. 6 is a diagram showing a hardware and software components of a computer system 50 on which the system of the present disclosure can be implemented. The computer system 50 can include a storage device 52, computer vision software code 54, a network interface 56, a communications bus 58, a central processing unit (CPU)

(microprocessor) 60, a random access memory (RAM) 62, and one or more input devices 64, such as a keyboard, mouse, etc. The computer system 50 could also include a display (e.g., liquid crystal display (LCD), cathode ray tube (CRT), etc.). The storage device 52 could comprise any suitable, computer-readable storage medium such as disk, non-volatile memory (e.g., read-only memory (ROM), erasable programmable ROM (EPROM), electrically-erasable programmable ROM (EEPROM), flash memory, field-programmable gate array (FPGA), etc.). The computer system 50 could be a networked computer system, a personal computer, a server, a smart phone, tablet computer etc. It is noted that the computer system 50 need not be a networked server, and indeed, could be a stand-alone computer system.

[0043] FIG. 7 is a drawing illustrating a system 100 of the present disclosure. The system can include an ultrasound device 101 including an ultrasound receiver 102 and an ultrasound base 104. The ultrasound receiver 102 is a device which can be used by a medical professional for placement on a patient to generate ultrasonic waves and receive a signal for generation of an ultrasonic video/image at the ultrasound base 104. The ultrasound device 101 can transmit data of the ultrasound video over the Internet 106 to a computer system 108 for localizing a medical device in an ultrasound video via a computer vision processing engine 110. It should be noted that the processing engine 110 can be located at the ultrasound base 104 obviating the need for the ultrasound device 101 to send data over the Internet 106 to a remote computer system.

[0044] FIG. 8 is a block diagram illustrating another embodiment of the systems and methods of the present disclosure, wherein a convolutional neural network (CNN) localizes a needle tip in ultrasound images. As discussed above, the systems and methods of the present disclosure provide for a needle tip enhanced image $e(x, y)$ in which the needle tip exhibits a high intensity.

[0045] It is noted that the needle tip may not always exhibit the highest intensity in the needle tip enhanced image $e(x, y)$ because of needle actuation speed and noise interference. For example, generally, the needle tip will not move in each ultrasound

frame because the speed of needle actuation may not match the ultrasound frame rate or the operator may intermittently stop moving the needle. As such, it is necessary to identify frames in which significant motion has not occurred. In addition, despite the above discussed needle enhancement process, high-intensity interfering artifacts not associated with needle motion could be present. Accordingly, a deep learning framework (e.g., a CNN) can be utilized for efficient needle tip detection.

[0046] FIG. 8 illustrates the CNN architecture and is built based on You Only Look Once (YOLO) real time object detection. YOLO is a state of the art single-shot object detection CNN architecture. The framework outputs two-dimensional bounding box predictions consisting of five components: x , y , w , h and η , where (x, y) coordinates represent the center of the bounding box, w and h are the width and the height of the bounding box, respectively, and η is the confidence value that the bounding box contains an object and that the object is the needle tip. As shown in FIG. 8, the CNN consists of B-mode ultrasound data 202, enhanced needle image 204, 256 x 256 fused image input layer 206, a convolution + exponential linear unit (ELU) + max pooling layer 208, a convolution + ELU layer 210 and a convolution layer 212. The CNN is configured to yield a confidence score 214 indicative of needle tip classification and localization accuracy.

[0047] The CNN utilizes eight convolutional layers to further reduce computational complexity toward real-time performance, but of course, other numbers of layers could be utilized. The 256 x 256 fused image input layer 206 is a pixel-level fusion layer in which the current ultrasound image $p(x, y)$ (obtained from the B-mode ultrasound data 202) and its tip enhanced needle image 204 counterpart $e(x, y)$ are concatenated before being input to the CNN. Since the needle tip is a fine-grained feature, the convolution layers are configured to maintain spatial dimensions of the respective inputs, thus mitigating reduction in resolution. It is noted that CNN neurons at deeper layers may have large receptive fields that can ensure incorporation of image level context pertinent to needle tip appearance.

[0048] Uniquely, each of the first seven convolution layers is followed by an

Exponential Linear Unit (ELU) activation layer, with $\alpha = 0.5$. The ELU provides for activations close to zero mean and unit variance to converge toward zero mean and unit variance even under the presence of noise and perturbations. The first five convolutions layers are followed by a 2 x 2 max pooling layer with a stride of 2. The CNN model is malleable to any input size and provides for two advantages. First, needle tip features will be learned end to end, thus eliminating the need to explicitly encode them. Second, it is expected that frames where no needle tip has moved will exhibit no detectable features, while the learned CNN model will accurately extract the tip when it is present.

[0049] The CNN model can be initialized with weights derived from training on the PASCAL Visual Object Classes (VOC) dataset. The ground-truth bounding box labels can be defined utilizing an electromagnetic (EM) tracking system and a radiologist experienced in interventional radiology. The ground-truth top location becomes the center of the bounding box (x, y) and the thickness $w \times h$ is chosen to be at most 20 x 20 pixels in all images. It is noted that the CNN model can utilize varying learning rates, batch sizes, training durations and optimizers. For example, the CNN model of FIG. 8 utilizes an initial learning rate of 10^{-4} , a batch size of 4, a training duration of 60 epochs and an Adam optimizer.

[0050] Data acquisition and experimental validation of the CNN model of the present disclosure will now be discussed. To train and evaluate the CNN model, a dataset of two-dimensional B-mode ultrasound images was collected using materials and settings specified in Table 2.

Imaging system	Needle type, dimensions and insertion profile	# of videos	Pixel size (mm)
<i>Bovine tissue</i>			
SonixGPS	17G SonixGPS (1.5 mm, 70 mm), IP	5	0.17
	17G Tuohy (1.5 mm, 70 mm), IP	10	0.17
	17G Tuohy (1.5 mm, 70 mm), OP	3	0.17
	22G BD (0.7 mm, 90 mm), IP	5	0.17
	22G BD (0.7 mm, 90 mm), OP	7	0.17
<i>Porcine tissue on spine phantom</i>			
Clarius C3	17G SonixGPS (1.5 mm, 70 mm), IP	5	0.24
	17G Tuohy (1.5 mm, 70 mm), IP	5	0.24
	22G BD (0.7 mm, 90 mm), IP	5	0.24
	22G BD (0.7 mm, 90 mm), OP	5	0.24

IP in-plane insertion, *OP* out-of-plane insertion

Table 2

Specifically, two imaging systems including the SonixGPS (Analogic Corporation, Peabody, MA, USA) with a handheld C5-2/60 curvilinear probe and the Clarius C3 two-dimensional handheld wireless ultrasound (Clarius C3, Clarius Mobile Health Corporation, Burnaby, BC, Canada) were used. In addition, experiments were performed on freshly excised bovine tissue, a porcine shoulder phantom and chicken breast, with insertion of a 17 gauge (1.5 mm diameter, 90 mm length) Tuohy epidural needle (Arrow International, Reading, PA, USA), a 17 gauge SonixGPS vascular access needle (Analogic Corporation, Peabody, MA, USA) and a 22 gauge spinal Quincke-type needle (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). During experimentation, the probe was handheld. Small amplitude perturbations not associated with needle motion were simulated by manually pressing the probe against the imaging medium and rotating it slightly about its longitudinal axis. Further, the chicken breast was overlaid on a lumbosacral spine model and immersed in a water bath during needle insertion to simulate fluid motion in the imaging medium. Ground-truth needle tip localization data using an EM tracking system (Ascension Technology Corporation, Shelburne, VT, USA) was collected with the SonixGPS needle. Specifically, in-plane insertion was performed at 40°-70°, and the needle was inserted up a depth of 70 mm. Fifty (35 in-plane, 15 out-of-plane) sequences of ultrasound images, each containing more than 400 frames, were collected.

[0051] Performance of the CNN model was evaluated by comparing the automatically detected needle tip location (i.e., the center of the detected bounding box) to the ground-truth determined from the EM tracking system for data collected with the SonixGPS needle. For data collected with needles without tracking capability, the ground-truth was determined by a radiologist. The ground-truth needle tip location was determined from intensity changes and tissue deformation (this is more difficult in the real-time clinical setting) by retrospectively inspecting the frame sequences. To account for large EM tracking errors (since the sensor does not reach the needle tip), the radiologist performed manual labeling of the dataset obtained with the SonixGPS needle and compared the EM data with the manual data. The EM system provides annotation on the ultrasound frames which acts as a visual cue to the approximate needle tip location in scenarios where the needle tip intensity is low. The radiologist used this information to label the needle tip. If the difference in needle tip localization was 4 pixels (~0.7 mm) or greater, the localizations were not included in the computation. Needle tip localization accuracy was determined from the Euclidean distance between the ground-truth and the localization from the CNN model.

[0052] The CNN model was implemented on an NVIDIA GeForce GTX 1060 6 GB GPU, 3.6 GHz Intel(R) Core™ i7 16 GB CPU Windows PC. The needle tip enhancement and augmentation methods were implemented in MATLAB 2018a. For the sub-problems in equations 9 and 10, the optimum values of $\nu = 2$ and $\lambda = 5$ were empirically determined. These values remained constant throughout the validation experiments. The needle tip detection framework was implemented in Keras 2.2.4 (on the Tensorflow 1.1.2 backend). In total, 5000 images from 20 video sequences were utilized for training the CNN model, while 1000 images from 10 other video sequences were used for validation. Additionally, 700 images from 20 video sequences that were not utilized in training or validation were utilized for testing the CNN model. The images were selected from continuous sequences having needle motion.

[0053] FIG. 9 illustrates needle tip localization results for consecutive frames from in-plane insertion of a 17 gauge needle and out-of-plane insertion of a 22 gauge needle. Specifically, FIG. 9 illustrates needle tip detection and localization in four consecutive

frames with (A) in-plane insertion of the 17 gauge Sonix GPS needle into chicken breast and (B) out-of-plane insertion of the 22 gauge needle into the porcine shoulder phantom. (I) illustrates the original image. The white box denotes the annotated ground-truth label. The ground-truth label can be determined with an EM tracking system for (A) and an expert sonographer for (B). (II) illustrates a detection result with a white bounding box overlaid on the enhanced image $e(x, y)$. The inset value positioned alongside the white bounding box annotation denotes the detection confidence. (III) illustrates the localized needle tip wherein the center of the detected bounding box (e.g., the star) is overlaid on the original image.

[0054] It is noted that the needle tip is accurately localized despite the presence of other high-intensity interfering artifacts in the B-mode ultrasound data 202. If there is a point cloud arising from partial enhancement of the needle shaft, the CNN learns to automatically identify the needle tip at the distal end of the cloud in the enhanced image $e(x, y)$. The temporal window for the needle tip visibility can be limited for out-of-plane insertion. However, the method of the present disclosure can be useful for tracking small movements of the needle tip close to the target. Additionally, the method of the present disclosure is agnostic to the needle type and the needle size utilized if the needle tip appears in the enhanced ultrasound image and needle motion is available in the B-mode ultrasound data 202. Increasing the training data size for each needle type can improve the performance of the systems and methods of the present disclosure.

[0055] CNN model comparisons provide for maximizing needle tip detection efficiency. For example, ablation studies alter the structural configuration of a deep learning framework to assess the impact on CNN model performance and are utilized to justify design choices. As such, the efficiency of the needle tip detection CNN framework was compared to that from alternative implementation approaches. The accuracy of needle tip detection can be determined using the mean average precision (mAP) metric on the validation dataset. The mAP metric is calculated as the average value of the precision across a set of 11 equally spaced recall levels, yielding one value that depicts the shape of the precision-recall curve. Table 3 shows the mAP metric for different configurations of the detection CNN. As shown in Table 3, the Detection CNN (with ELU) + $p(x, y) + e(x,$

y) (in bold) of the present disclosure yields the highest mAP metric compared to alternative detection CNN frameworks.

Method	mAP
Detection CNN + $p(x, y)$	0.202
Detection CNN + $e(x, y)$	0.867
Detection CNN (with leakyReLU) + $p(x, y) + e(x, y)$	0.914
Detection CNN (with ELU) + $p(x, y) + e(x, y)$	0.946

Table 3

[0056] Evaluation of the detection CNN frameworks will now be discussed. First, the performance of the CNN with the raw ultrasound image $p(x, y)$ as an input was examined. The detection efficiency of the CNN with the raw ultrasound image $p(x, y)$ as an input is very low (20.2%) because, without the needle tip enhancement algorithm, the needle tip features are not clearly discernible and are overshadowed by other high intensity artifacts in the cluttered ultrasound image. The detection efficiency of the CNN with the enhanced image $e(x, y)$ as the input yields a high mAP metric of 86.7% thereby demonstrating that the needle tip enhancement algorithm is efficient. Furthermore, the fusion of $e(x, y)$ and $p(x, y)$ yields the highest mAP metric of 94.6%.

[0057] The performance of the detection CNN model of the present disclosure was also compared against a similar model with Leaky Rectified Linear Unit (LeakyReLU) activation layers instead of an ELU wherein the fusion input 206 and other hyper parameters were maintained constant. As shown in Table 3, the detection CNN model with ELU of the present disclosure outperforms the detection CNN model with Leaky ReLU. It is noted that a batch size of 4 was utilized in the evaluation of the detection CNN frameworks because of the memory constraints of the GPU. It is expected that a larger batch size would have yielded an even higher mAP metric for the detection CNN model of the present disclosure.

[0058] Runtime performance of the CNN framework of the present disclosure will now be discussed. The CNN framework runs at 0.094 ± 0.01 seconds per frame on the

NVIDIA GeForce GTX 1060 GPU. Specifically, the CNN framework runs at 0.014 seconds per frame for enhancement, 0.06 seconds per frame for augmentation and 0.02 seconds per frame for detection. The runtime performance is approximately 10 frames per second (fps) and thereby achieves a fast needle tip localization framework. It is noted that the processing speed can be increased with additional computing resources. In frames where the needle tip is salient, the augmentation step is unnecessary, and the runtime speed increases to 29 fps.

[0059] The accuracy of the CNN framework can be increased by mitigating false needle tip detections. Since YOLO is a multi-object detection framework, YOLO can detect several bounding boxes with different confidence scores on a single input image. These false positives can be minimized by selecting the bounding box with the highest confidence score and utilizing a hard threshold of 0.35 for the confidence score. The hard threshold of 0.35 is an empirically determined value and is kept constant throughout validation. With this threshold, the CNN framework can yield an overall sensitivity of 98% and specificity of 91.8%. It is noted that the robustness of needle tip detection can be further improved by utilizing a bigger training dataset.

[0060] FIG. 10 illustrates elimination of needle tip localization false positives from a trajectory estimation. As shown in FIG. 10, points *A* and *B* lie along the correct trajectory whereas the bounding box with center *F* is a false positive. It is assumed that the needle tip detection framework accurately localized two previous spatial positions $A(x_1, y_1)$ and $B(x_2, y_2)$ in successive frames that are at least 30 pixels (approximately 5 mm) apart. From *A* and *B*, the needle trajectory $\alpha_1 = \tan^{-1} (|(y_2 - y_1) / (x_2 - x_1)|)$ is approximated. Then, for each subsequent detection with a bounding box at $F(x_1, y_1)$, the trajectory of angle α_2 is estimated using points *A* and *F* with *A* as a static reference. If $|\alpha_1 - \alpha_2| > 10^\circ$, the new detection is determined to be skewed from the correct trajectory and thus indicative of a false positive. Accordingly, the localization result is not utilized in calculating the localization error. During the localization process, false positives and true negatives lead to maintenance of the current needle tip position. As such, the elimination of needle tip localization false positives from the trajectory estimation is robust to spatiotemporal redundancies.

[0061] Needle tip localization accuracy of the method of the present disclosure can be evaluated in view of the needle tip localization error and the performance of comparable methods. Overall, the needle tip localization error of the method of the present disclosure is 0.72 ± 0.4 mm. A direct comparison with comparable methods is difficult because the method of the present disclosure does not require initial needle visibility. Therefore, the dataset collected for the method of the present disclosure is tailored for the evaluation of the method of the present disclosure.

[0062] Table 4 shows the results of a comparison between the method of the present disclosure and comparable methods. Specifically, Table 4 shows the results of the method of the present disclosure with detection CNN, a method disclosed in white paper “Convolution Neural Networks for Real-Time Needle Detection and Localization in 2D Ultrasound” by Cosmas Mwikirize et al. (“method of [16]”), and a method utilizing the Hough transform with the random sample consensus (RANSAC).

Method	Success (%)	Localization error (mm)	Overall processing time (s)
Proposed method (with detection CNN)	100	0.72 ± 0.04	0.094
Method of [16]	44	1.04 ± 0.36	0.56
Hough transform + RANSAC	18	1.2 ± 0.32	0.74

Bold characters denote performance of the proposed method, which is the best when compared to alternative methods

Table 4

[0063] The embodiments of the present disclosure utilizing CNNs for detection and the method of [16] were compared by evaluating performance on the same set of 200 randomly selected ultrasound images with only in-plane needle insertion. It is noted that the method of the present disclosure with CNN detection outperforms the method of [16] in both needle tip localization accuracy and computational efficiency. Localization errors of approximately 2 mm (i.e., 56% of the data) were discarded for the purpose of the comparison. A one-tailed paired test shows that the difference between the localization errors from the method of the present disclosure and the method of [16] is statistically significant ($p < 0.005$). The localization accuracy obtained from the method of [16] is worse than previously reported because a more challenging dynamic dataset with very low

shaft intensity was utilized for validation.

[0064] The systems/methods of the present disclosure utilizing CNN for detection, and the systems/methods utilizing the Hough transform and the RANSAC, were also compared. Specifically, the systems/methods of the present disclosure utilizing CNN detection were compared with an intensity-based method that directly localizes the needle tip utilizing the Hough transform and the RANSAC. As shown in Table 4, the method utilizing the Hough transform and the RANSAC achieved success in only 18% of the dataset (neglecting errors > 2 mm), with an overall localization error of 1.2 ± 0.32 mm.

[0065] Table 5 shows the determined needle localization from the maximum intensity in $e(x, y)$, (i.e., the method of the present disclosure without the tip detection step).

Localization approach	Ground truth	# of images	Error (mm)
Proposed method (with detection CNN)	Tracking system	250	0.76 ± 0.03
Proposed method (with detection CNN)	Expert	450	0.69 ± 0.05
Proposed method (with detection CNN) (overall)		700	0.72 ± 0.04
Highest intensity in $e(x, y)$ (without detection CNN)	Tracking system	250	0.94 ± 0.04
Highest intensity in $e(x, y)$ (without detection CNN)	Expert	450	1.12 ± 0.05
Highest intensity in $e(x, y)$ (without detection CNN)(overall)		700	1.06 ± 0.04

Bold characters denote performance of the proposed method, which is the best when compared to alternative methods

Table 5

The results shown in Table 5 demonstrate that the localization accuracy is worse without the detection CNN framework. As such, without the benefit of implicitly learning heuristic features associated with the needle tip via deep learning, there is a higher likelihood of localizing artifacts with similar intensity to the needle tip.

[0066] FIG. 11 is a block diagram 220 illustrating another embodiment of the system and method of the present disclosure, wherein a convolutional neural network localizes a needle tip in temporal ultrasound data. The framework accommodates variations between images captured under different imaging scenarios and needle insertion

profiles. The method of the present disclosure targets both out-of-plane and in-plane inserted needles in which the needle tip exhibits low contrast compared to the rest of the ultrasound image.

[0067] As shown in FIG. 11, the systems and methods of the present disclosure consists of four processes based on input ultrasound frames F_0 to F_t 222. Specifically, the processes include needle tip enhancement 224, needle tip classification 226, needle tip location regression 228 and needle tip location 230. The needle tip enhancement process 224 comprises enhancement of the needle tip from two consecutive ultrasound frames. The needle tip classification process 226 determines, using a classifier network, whether the enhanced image contains substantial needle tip information. The needle tip classification process 226 is necessary because the hand-inserted needle tip does not move smoothly through space and as such frames will exist in the sequence where the needle tip has not changed spatial location. The needle tip location regression process 228 estimates the needle tip location using a keypoint regression CNN. These processes will be described in detail below.

[0068] The two-dimensional B-mode ultrasound images utilized in the learning experiments were collected using two imaging systems including the SonixGPS (Analogic Corporation, Peabody, MA, USA) with a handheld C5-2/60 curvilinear probe and the Clarius C3 two-dimensional handheld wireless ultrasound (Clarius C3, Clarius Mobile Health Corporation, Burnaby, BC, Canada). In addition, experiments were performed on freshly excised bovine tissue, porcine tissue and chicken breast overlaid on a lumbosacral spine phantom, with insertion of a 17 gauge SonixGPS vascular access needle (Analogic Corporation, Peabody, MA, USA) and a 22 gauge spinal Quincke-type needle (Becton, Dickinson and Company, Franklin Lakes, NJ, USA).

[0069] The needles were inserted both in-plane (30° to 70° insertion angle) and out-of-plane up to a depth of 70 mm. For the SonixGPS needle, tip localization data was

collected from an EM tracking system (Ascension Technology Corporation, Shelburne, VT, USA). During the experiments, minor motion events were simulated by exerting pressure on the tissue with the probe and probe rotation. In total, 60 volumes were collected (30 in-plane, 30 out-of-plane: 40 with Sonix GPS system and 20 with Clarius C3 system), with each video sequence having more than 600 frames. In training and validating the needle tip classifier, 7000 positive examples (enhanced images with the tip) and 5000 negative examples (enhanced images without tip information) were utilized. Regarding needle tip location regression, 7000 images (only positive examples) for training and validation were utilized and 500 images from 20 sequences not used for training and validation were utilized for testing.

[0070] In the needle tip enhancement process 224, the logical difference of consecutive frames $z(x, y)$ is computed from the ultrasound frame sequence $F_0, F_1 \dots F_{t-1}, F_t$ 222 where the subscript denotes the respective temporal position and F_0 is the first frame. For a current frame $n(x, y)$ and a previous frame $m(x, y)$, $z(x, y)$ can be calculated by equation 11 below:

$$z(x, y) = m^c(x, y) \wedge n(x, y)$$

Equation 11

[0071] In equation 11, $m^c(x, y)$ is the bitwise complement of $m(x, y)$, while \wedge is the bitwise AND logical operation. This logical differencing routine produces intensities T : [0, 255] representative of the difference between the two frames. Therefore, subtle motion between the two frames will be captured. It is noted that the needle tip will provide the most prominent feature in $z(x, y)$. To compensate for irrelevant motion events, a 12 x 12 median filter is applied to $z(x, y)$ and further tip augmentation is not performed. FIG. 12 shows the result of the tip enhancement process 224.

[0072] FIG. 13 is a block diagram illustrating a convolutional neural network 240 capable of performing processing steps for tip classification process 226 of FIG. 11. The network consists of a needle tip enhanced image 242, 6 blocks of convolution, Rectified Linear Unit (ReLU) activation layers, batch normalization and max pooling layers 244, 2

blocks of convolution, ReLU and batch normalization layers 246, and 3 fully connected layers 248. The input needle tip enhanced image 242 consists is a 256 x 256 image. All convolution layers utilize a 3 x 3 kernel, and a stride/padding of 1 pixel. The max pooling layers utilize a 2 x 2 kernel.

[0073] Positive and negative examples for the enhanced needle tip (see FIG. 14) can be labeled by a radiologist experienced in interventional radiology while referring to the corresponding original ultrasound sequences and EM tracking information for tracked needles. In positive examples, the needle tip exists as a distinct high intensity feature, a few pixels thick, against a low intensity background. This type of feature is lacking from negative training examples. If other high intensity noisy artifacts exist, the classifier can differentiate them from the needle. It is noted that the last feature map for the classification task, from the last fully connected (FC) layer, is a vector of dimension 1 x 1 x 2, (i.e., it is meant to differentiate images of two classes: those with an enhanced needle tip and those without it). A Softmax activation is applied to this feature map and the Log Loss is utilized to calculate deviations between the network output and the ground-truth. During training, a stochastic gradient descent with momentum (SGDM) optimizer, and initial learning rate of 10^{-2} is utilized.

[0074] The architecture of the tip regression CNN mirrors the classification CNN of FIG. 13, with the exception of terminating in a 4-output FC layer (the needle tip location is defined by two keypoints). In essence, there is a cascade of two twin CNNs which share all of the learned layers but differ only in the output. During training and testing, the needle tip classification process 226 and the needle tip location regression process 228 are run in series.

[0075] The input to the regression CNN is the needle tip enhanced image 242 and the corresponding location labels. Keypoints for the needle tip are labeled as shown in FIG. 14. Two keypoints on the needle tip are utilized (e.g., $t_p: (x_1, y_1), (x_2, y_2)$). The utilization of the two keypoints constrains the spatial likelihood of the needle tip since the enhanced tip feature is not geometrically definable for labeling purposes. The keypoint

labels are placed geometrically opposite along a line through the center of the enhanced needle tip feature and the pixel at the distal end.

[0076] The enhanced images are augmented by rotating them through 90° , 180° and 270° . The labels are computationally manipulated to match the rotated images. From the original 7000 enhanced images, this yields 28,000 training examples. Further, the labels are normalized to be in the range $Q: [-1, 1]$. Since the initial labels are in the range $[1, 256]$, scaling reduces the magnitude of the Mean Squared Error (MSE), which is utilized as the loss function metric, and the magnitude of the gradients. This quickens the training process and ensures stable convergence. At test time, the outputs are rescaled to match the original data.

[0077] The last feature map for the needle tip location regression process 228, from the last FC layer, is a vector of dimension $1 \times 1 \times 4$, containing the four coordinates of the two labels: (x_1, y_1) , (x_2, y_2) . During training, the parameters include the RMSprop optimizer, an initial learning rate of 10^{-3} , and a mini batch size of 32. At test time, the top location $T(x_t, y_t)$ is directly obtained from the CNN outputs as the average of the x and y outputs, i.e., $x_t = (x_1 + x_2) / 2$. $y_t = (y_1 + y_2) / 2$. Since a region proposal step or anchor boxes are unnecessary to tell the network where to look, the framework provides a single-shot approach for needle tip localization.

[0078] The qualitative results of the method of the present disclosure will now be discussed. FIG. 15 shows examples of needle tip enhancement and localization. It is noted that the CNN framework accurately localizes the needle tip despite that the needle tip has low contrast and the needle shaft information is not visible. This is because the method of the present disclosure detects subtle intensity changes arising from needle tip motion, that are otherwise not easily discernible with the naked eye. Further, localization accuracy is generally resilient to high intensity artifacts in the rest image.

[0079] The classification accuracy of the method of the present disclosure will now be discussed. The classification CNN achieves an overall sensitivity and specificity of

88% and 82% respectively on test sequences. By comparison, the method disclosed in white paper “Learning Needle Tip Localization from Digital Subtraction in 2D Ultrasound” by Cosmas Mwikirize et al. (“Digital Subtraction Method”) achieved sensitivity and specificity of 95% and 90% respectively. Specifically, the Digital Subtraction Method outperforms the method of the present disclosure in the presence of increased motion artifacts in the imaging medium due to an included computationally expensive step for noise removal. However, in relatively stable sequences, the classification accuracy between the method of the present disclosure and the Digital Subtraction Method was similar (98% vs 97%). Therefore, the method of the present disclosure is suitable for procedures where there is minimal motion from physiological events (e.g., lumbar facet injections and peripheral nerve blocks). It is noted that a larger training dataset would improve the classification accuracy of the method of the present disclosure even in the presence of motion artifacts.

[0080] The needle tip localization error of the method of the present disclosure will now be discussed. The ground-truth needle tip location is determined from the original ultrasound sequences by a radiologist, with augmentation from the EM tracking system. Needle tip localization error is determined from the Euclidean Distance between the automatically determined needle tip location from the method of the present disclosure and the ground truth. Based on an evaluation for both in-plane (IP) and out-of-plane (OP) needle insertions, the method of the present disclosure provides for an overall needle tip localization error of 0.55 ± 0.07 mm. Table 6 shows a comparative analysis between the method of the present disclosure (i.e., proposed method) and other needle tip localization methods.

Method	Test data	error (mm)	Processing time (s)	Success rate
Proposed method	500(250 IP,250 OP)	0.55 ± 0.07	0.015 (67 fps)	94%
Method in [1]	500(250 IP,250 OP)	0.78 ± 0.08	0.091 (11 fps)	97%
Method in [2]	250 (IP)	1.23 ± 0.45	0.42 (2 fps)	52%

Table 6

[0081] As shown in Table 6, the proposed method outperforms the other methods with a statistically significant improvement ($p < 0.005$). Since the method in [2] requires

needle shaft information, 250 test images were utilized from the sequences with in-plane needle insertion for its evaluation. For all methods, an error cap of 2 mm was imposed as a measure of needle tip localization success. By dropping the worst 3% cases of the data analyzed with the method in [1] in order to obtain a 94% success rate similar to the proposed method, the localization error dropped to 0.73 ± 0.07 mm which demonstrates that the proposed method provides improved results ($p < 0.005$). Most of the failed cases (6%) resulted from the top localization network regressing onto a false artifact in the enhanced image. It is noted that performance in this regard can be improved with a larger training dataset.

[0082] The systems and methods of the present disclosure provide for an improvement in overall processing time. The experiments were executed on an NVIDIA GeForce GTX 1060 6GB GPU, 3.6 GHz Intel (R) Core™ i7 16GB CPU Windows PC and the method of the present disclosure was implemented in MATLAB 2018a using the Deep Learning Toolbox. The learning framework is trained offline and tested with offline video sequences of ultrasound data. Average processing time for the enhancement process is 0.002 seconds per image. Classification of the enhanced image can be an average of 0.007 seconds while needle tip localization requires 0.006 seconds. As such, the overall processing time is 0.015 seconds (67 fps) which provides a 509% improvement, in fps, over the method in [1].

[0083] Different CNN structures consisting of different numbers of convolution and pooling layers were investigated for the regression component of the learning framework. In addition, relative performance using the normalized Root Mean Square Error (RMSE) on the validation dataset was evaluated. The method of the present disclosure provides for a normalized RMSE of 0.006 which is better than that achieved by the maximum network depth for the 256 x 256 image input consisting of 8 convolution layers with pooling after each layer (0.014) and a network with only 6 convolutions (0.05). In addition, the performance of different optimizers in the regression network was compared. The following optimizers achieved the correspondingly indicated normalized RMSE: RMSprop (0.006), Adam (0.05), and SGDM (0.14). It is noted that there was no significant improvement in performance when using dropout layers in the regression

network.

[0084] Having thus described the system and method in detail, it is to be understood that the foregoing description is not intended to limit the spirit or scope thereof. It will be understood that the embodiments of the present disclosure described herein are merely exemplary and that a person skilled in the art can make any variations and modification without departing from the spirit and scope of the disclosure. All such variations and modifications, including those discussed above, are intended to be included within the scope of the disclosure.

CLAIMS

What is claimed is:

1. A method for real-time localization of a medical device in a video, comprising the steps of:
 - receiving at a processor a plurality of frames from the video;
 - identifying by the processor one of the plurality of frames as a background image and a frame adjacent to the background image as a current image;
 - performing by the processor a bitwise complement operation on the background image to generate a complement image; and
 - performing by the processor a pointwise logical AND operation on the complement image and the current image to identify the location of the medical device.
2. The method of Claim 1, further comprising the step of enhancing the medical device using a Split Bregman process.
3. The method of Claim 1, wherein the medical device is a needle.
4. The method of Claim 3, wherein a tip of the needle is inserted in-plane.
5. The method of Claim 3, wherein a tip of the needle is inserted out-of-plane.
6. The method of Claim 3, wherein the needle's insertion angle is not known.
7. The method of Claim 3, wherein the needle is non-linear.
8. The method of Claim 1, further comprising the step of setting a threshold correlation coefficient for skipping a frame due to insignificant motion.
9. The method of Claim 1, wherein the video is an ultrasound video.

10. A system for real-time localization of a medical device in a video, comprising:
 - a processor; and
 - a memory in communication with the processor, the memory having computer instructions stored thereon which, when executed, cause the processor to perform the following steps:
 - receiving a plurality of frames from the video;
 - identifying one of the plurality of frames as a background image and a frame adjacent to the background image as a current image;
 - performing a bitwise complement operation on the background image to generate a complement image; and
 - performing a pointwise logical AND operation on the complement image and the current image to identify the location of the medical device.
11. The system of Claim 10, wherein the instructions further cause the processor to enhance the medical device using a Split Bregman process.
12. The system of Claim 10, wherein the medical device is a needle.
13. The system of Claim 12, wherein a tip of the needle is inserted in-plane.
14. The system of Claim 12, wherein a tip of the needle is inserted out-of-plane.
15. The system of Claim 12, wherein the needle's insertion angle is not known.
16. The system of Claim 12, wherein the needle is non-linear.
17. The system of Claim 10, wherein a threshold correlation coefficient is set for skipping a frame due to insignificant motion.
18. The system of Claim 10, wherein the video is an ultrasound video.

19. A non-transitory computer-readable medium having computer-readable instructions stored thereon which, when executed by a computer system, cause the computer system to perform the steps of:

receiving at a processor a plurality of frames from a video;

identifying by the processor one of the plurality of frames as a background image and a frame adjacent to the background image as a current image;

performing by the processor a bitwise complement operation on the background image to generate a complement image; and

performing by the processor a pointwise logical AND operation on the complement image and the current image to identify the location of a medical device.

20. The non-transitory computer readable medium of Claim 19, further comprising the step of enhancing the medical device using a Split Bregman process.

21. The non-transitory computer readable medium of Claim 19, wherein the medical device is a needle.

22. The non-transitory computer readable medium of Claim 21, wherein a tip of the needle is inserted in-plane.

23. The non-transitory computer readable medium of Claim 21, wherein a tip of the needle is inserted out-of-plane.

24. The non-transitory computer readable medium of Claim 21, wherein the needle's insertion angle is not known.

25. The non-transitory computer readable medium of Claim 21, wherein the needle is non-linear.

26. The non-transitory computer readable medium of Claim 19, further comprising the step of setting a threshold a correlation coefficient for skipping a frame due to insignificant motion.

27. The method of Claim 1, further comprising the step of enhancing image data representing the medical device using a convolution neural network.

28. The system of Claim 10, wherein the instructions further cause the processor to enhance image data representing the medical device using a convolution neural network.

29. The non-transitory computer readable medium of Claim 19, further comprising the step of enhancing image data representing the medical device using a convolution neural network.

FIG. 1

2

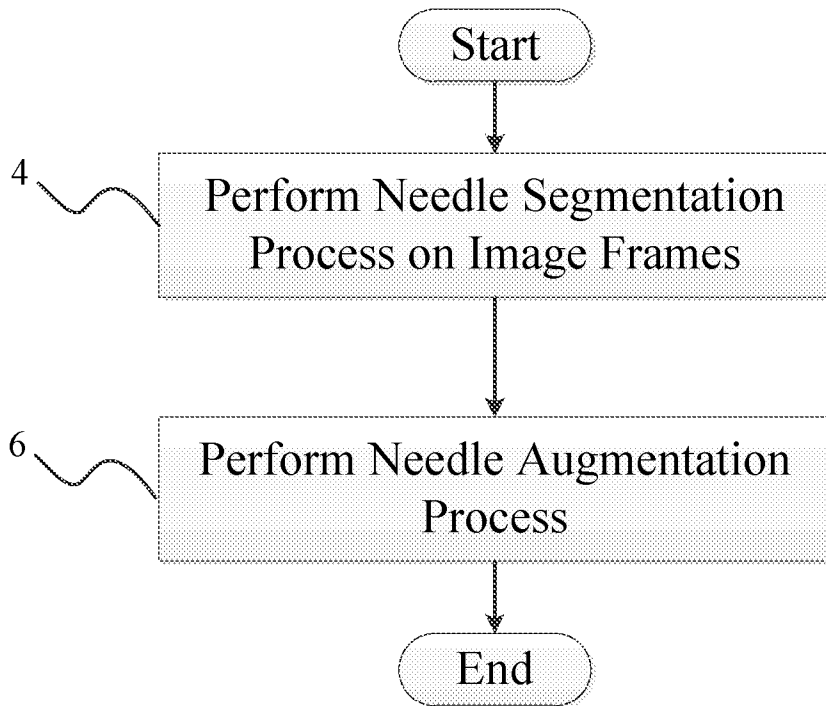



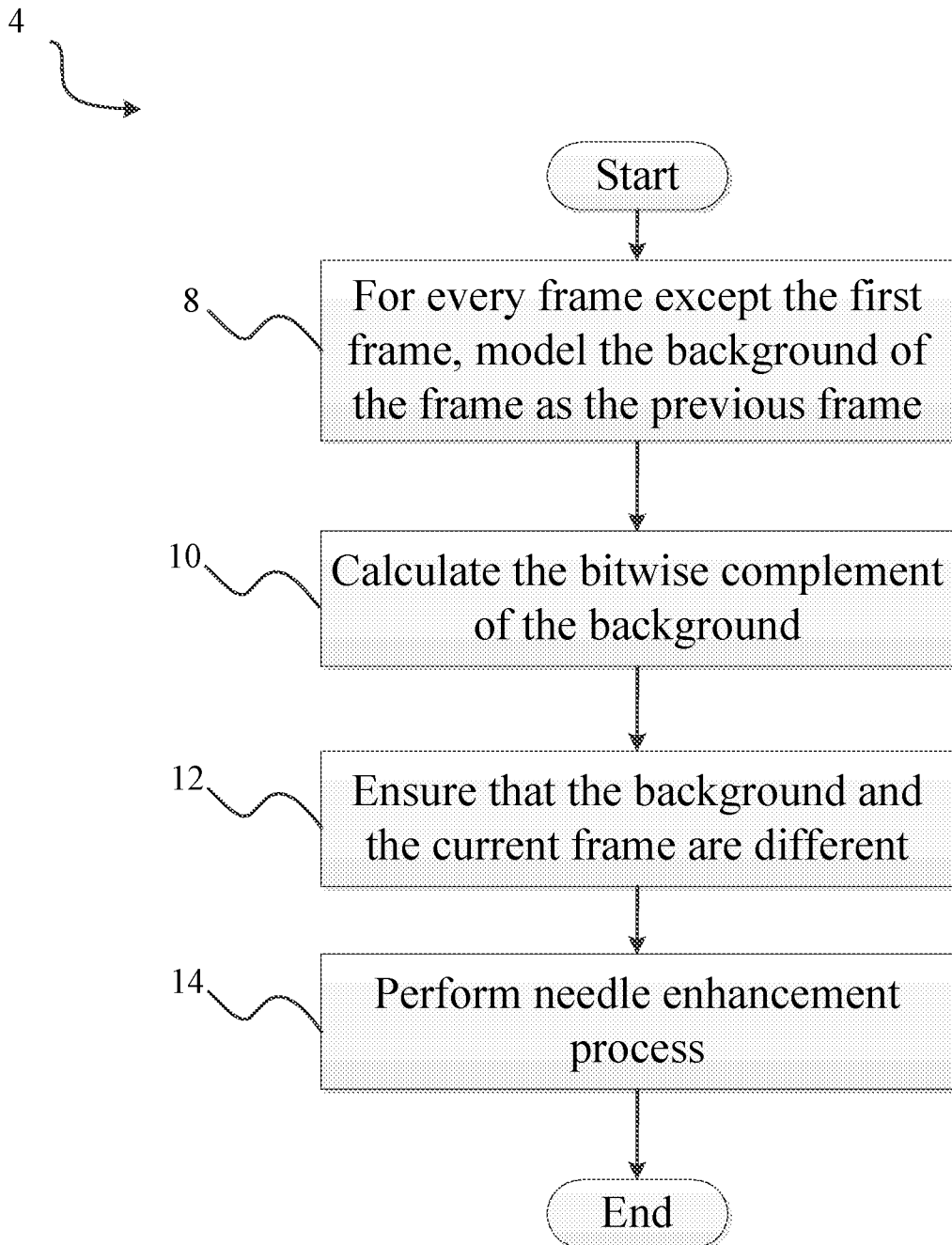
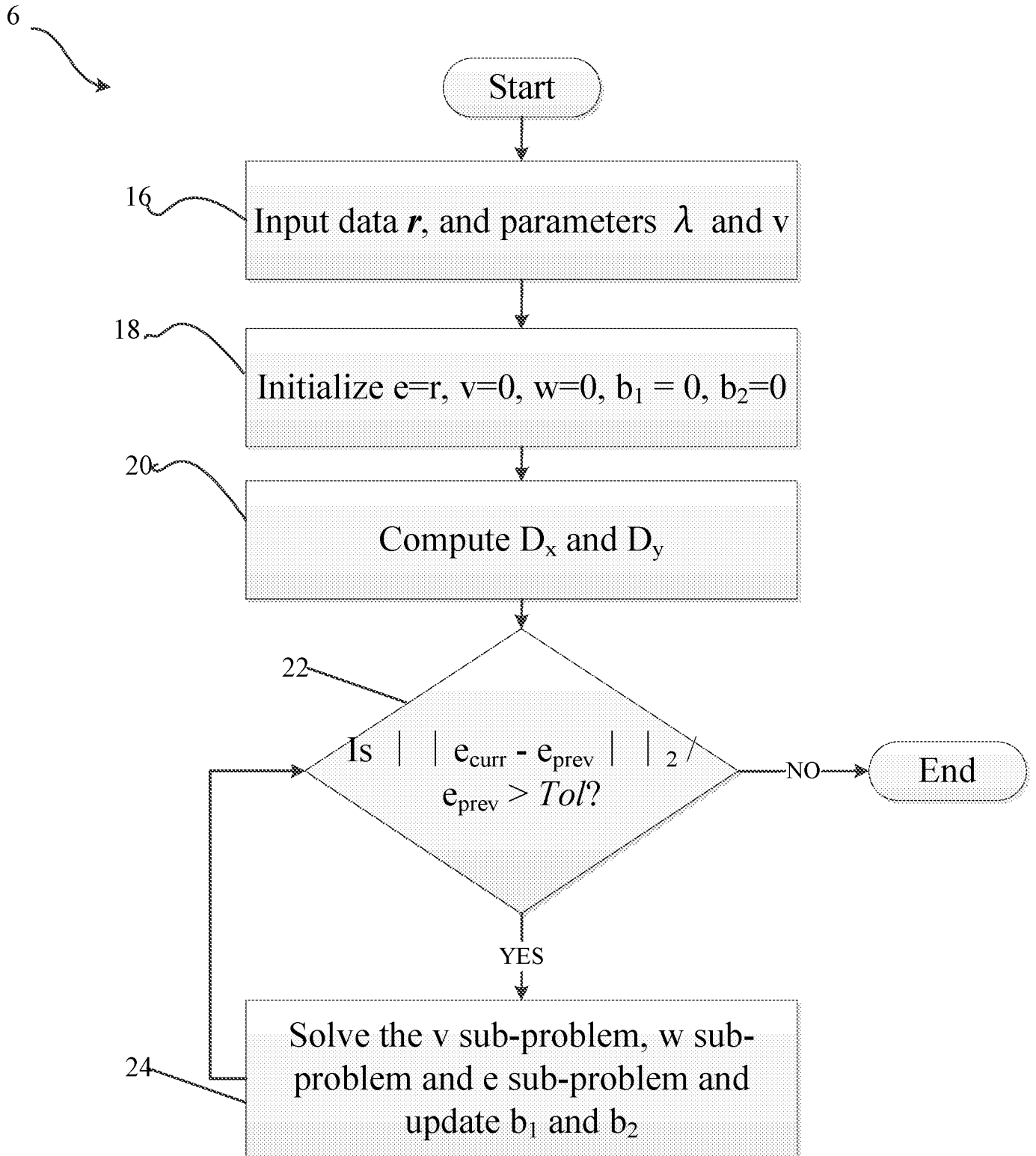
FIG. 2

FIG. 3



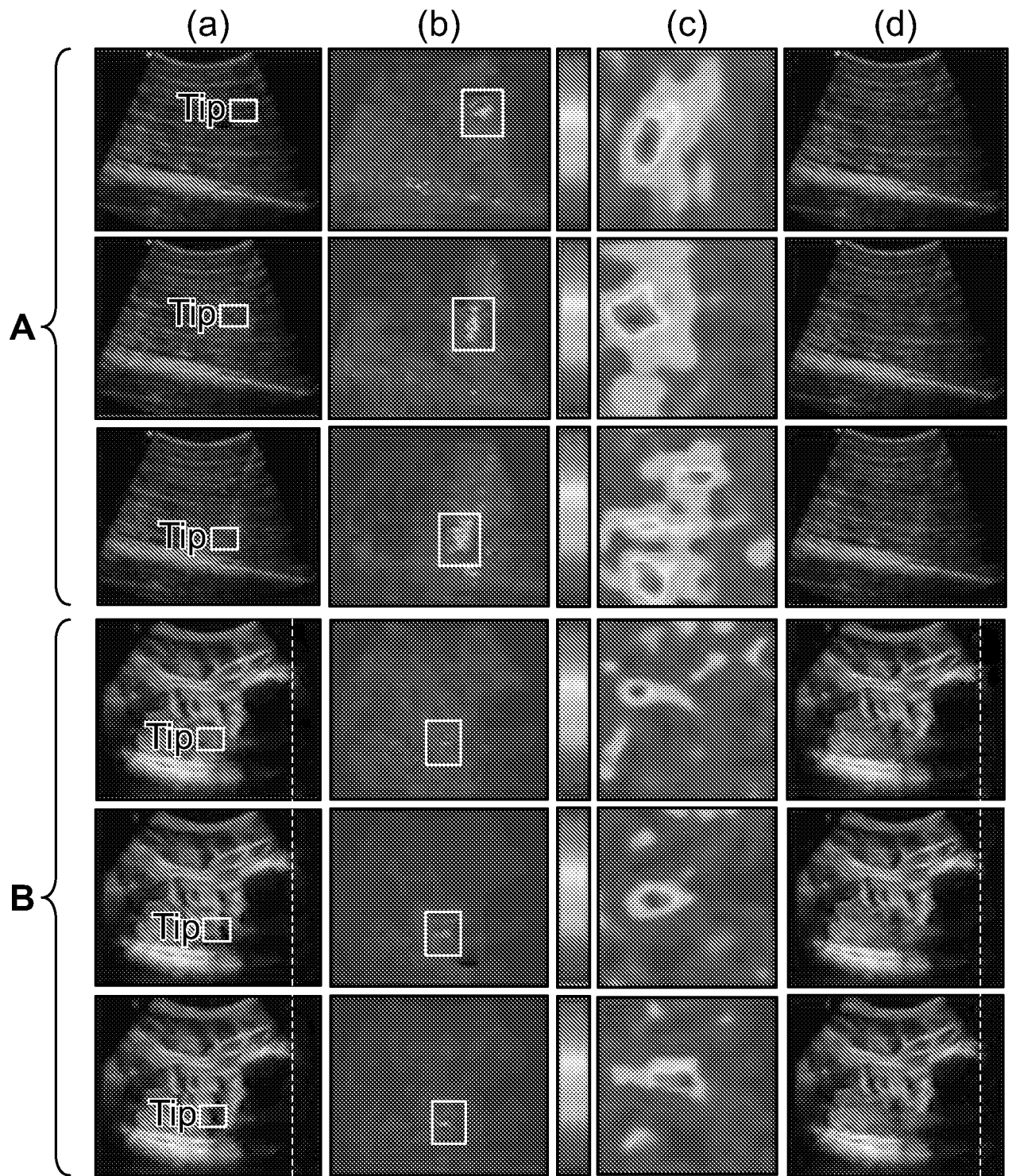


FIG. 4

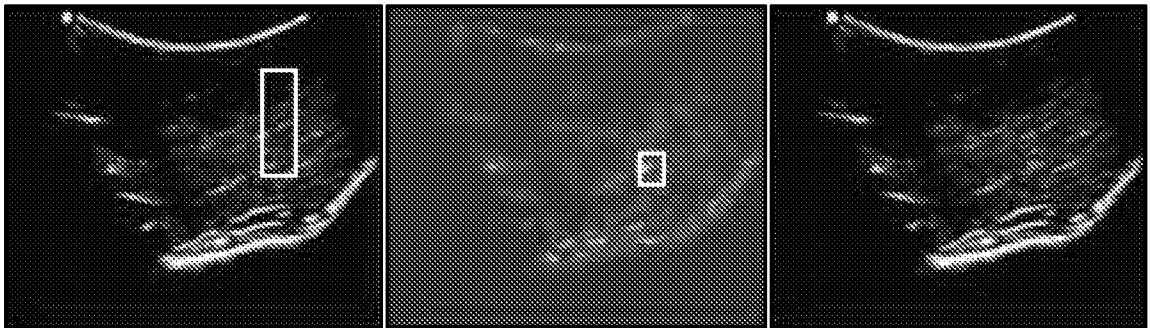
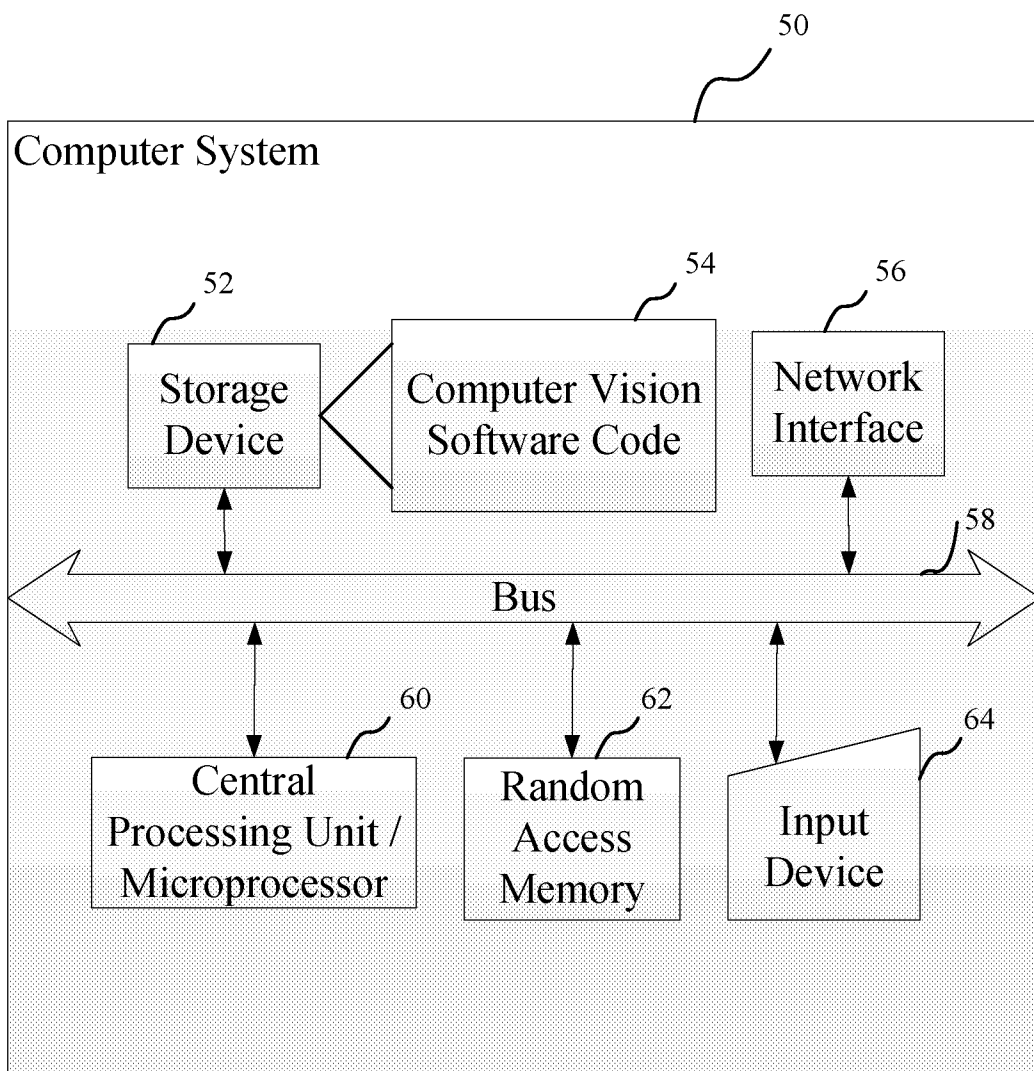


FIG. 5A

FIG. 5B

FIG. 5C

FIG. 6



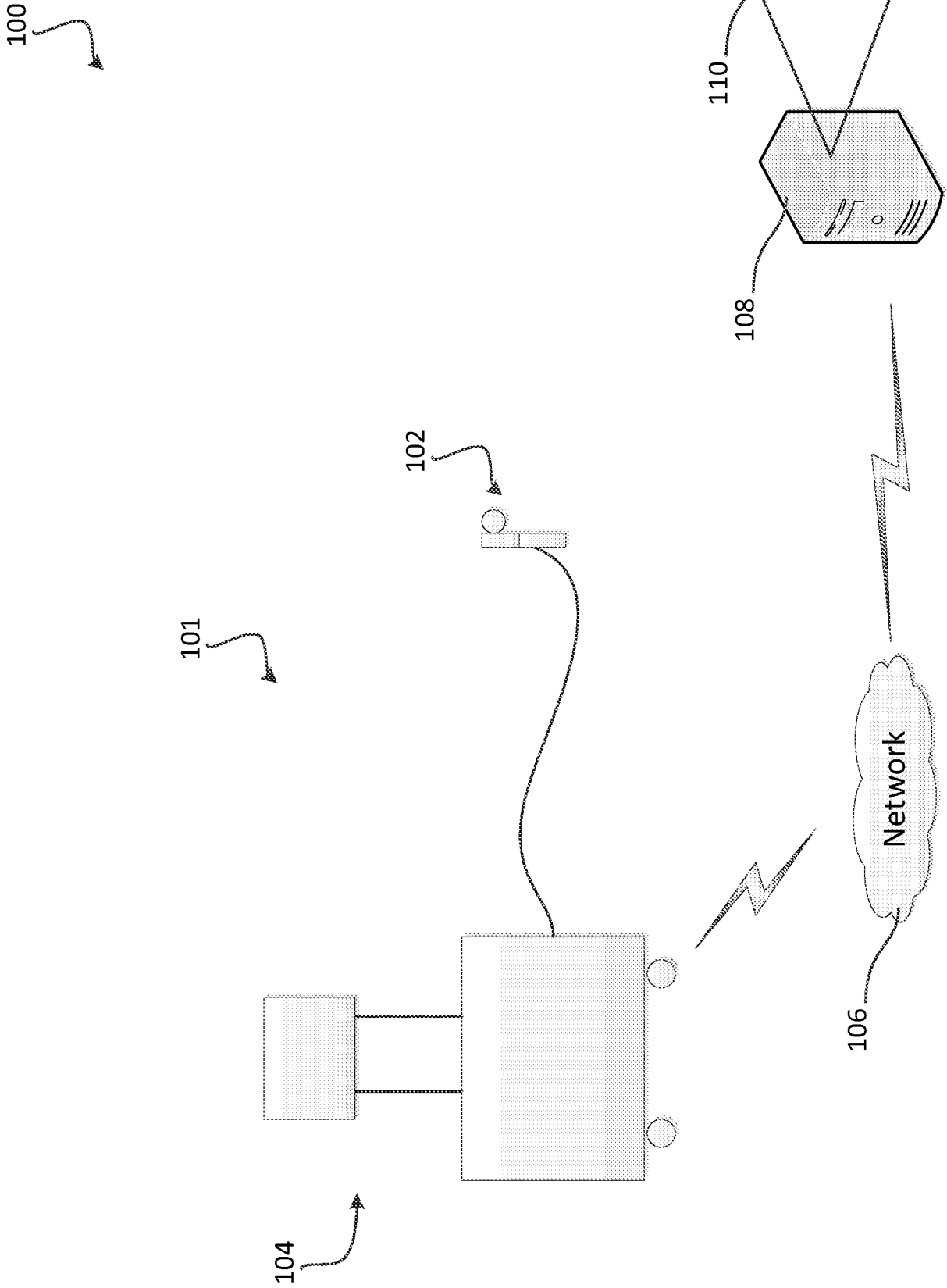


FIG. 7

200 →

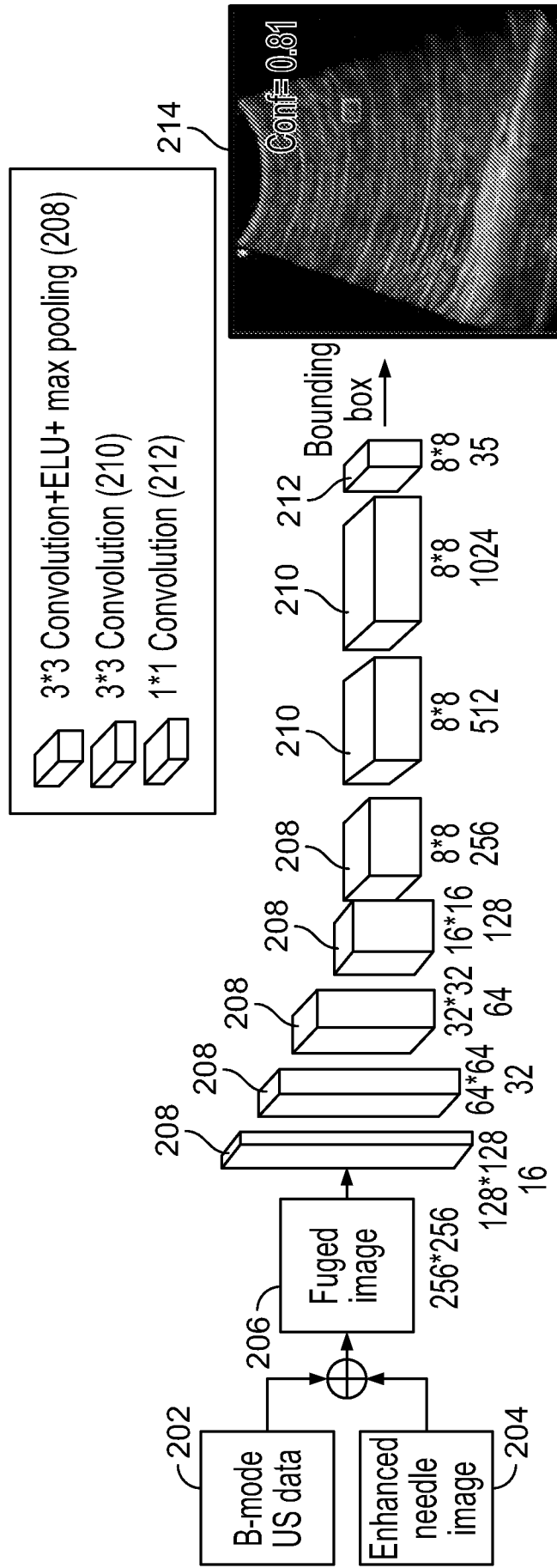


FIG. 8

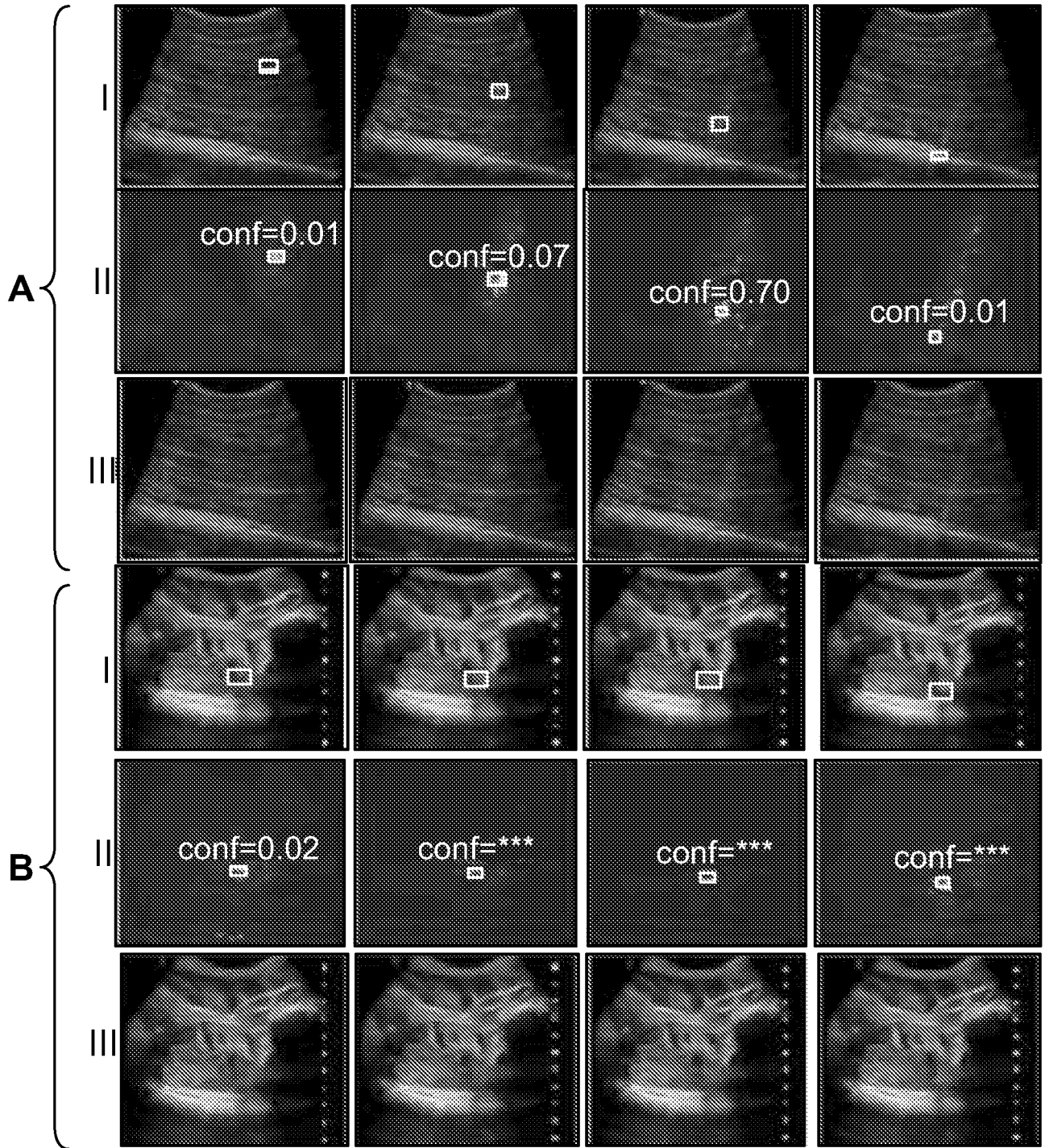


FIG. 9

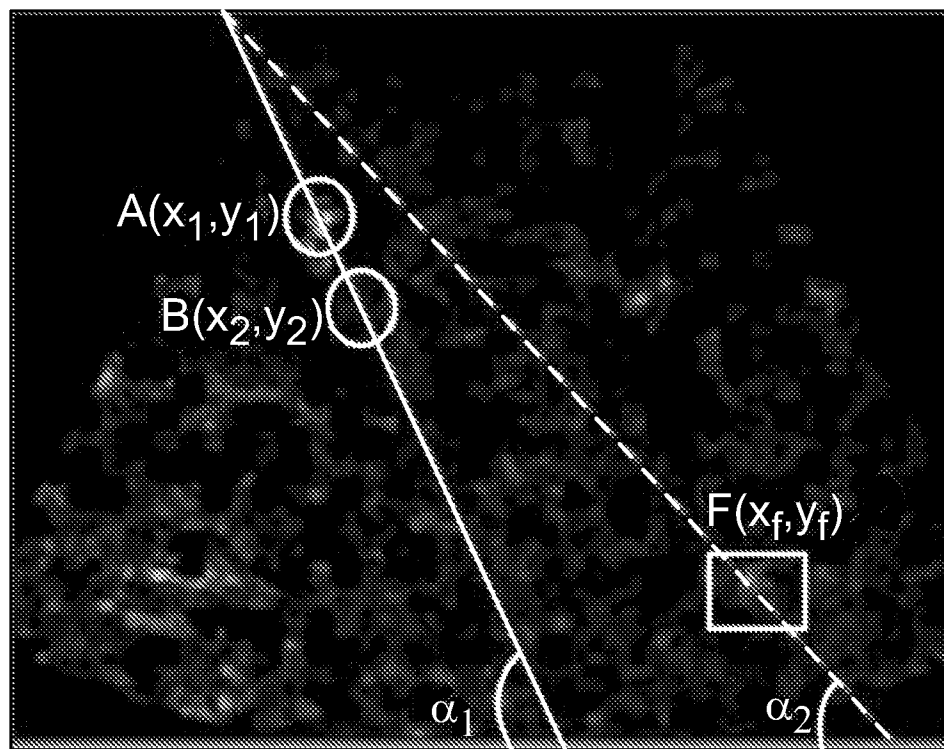


FIG. 10

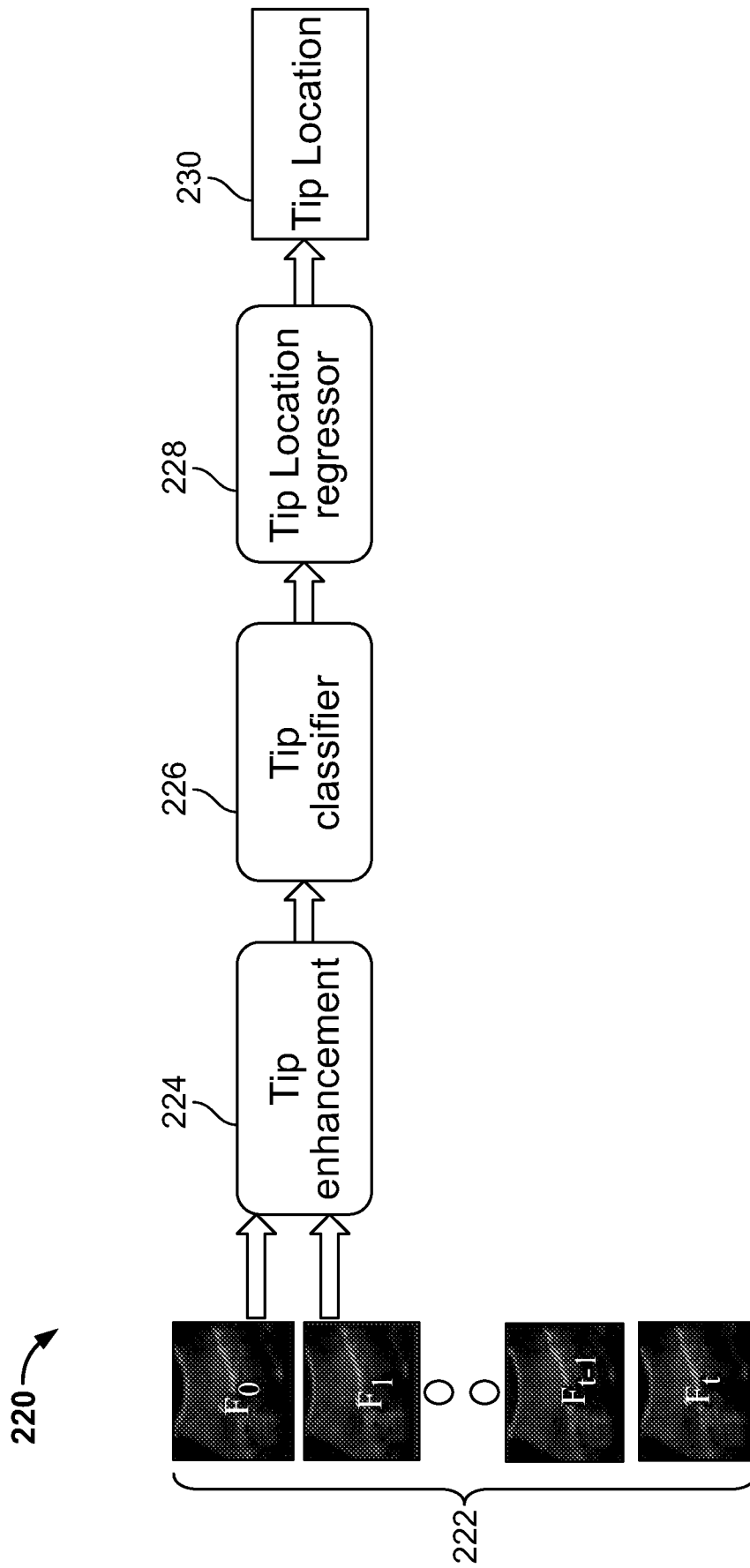


FIG. 11

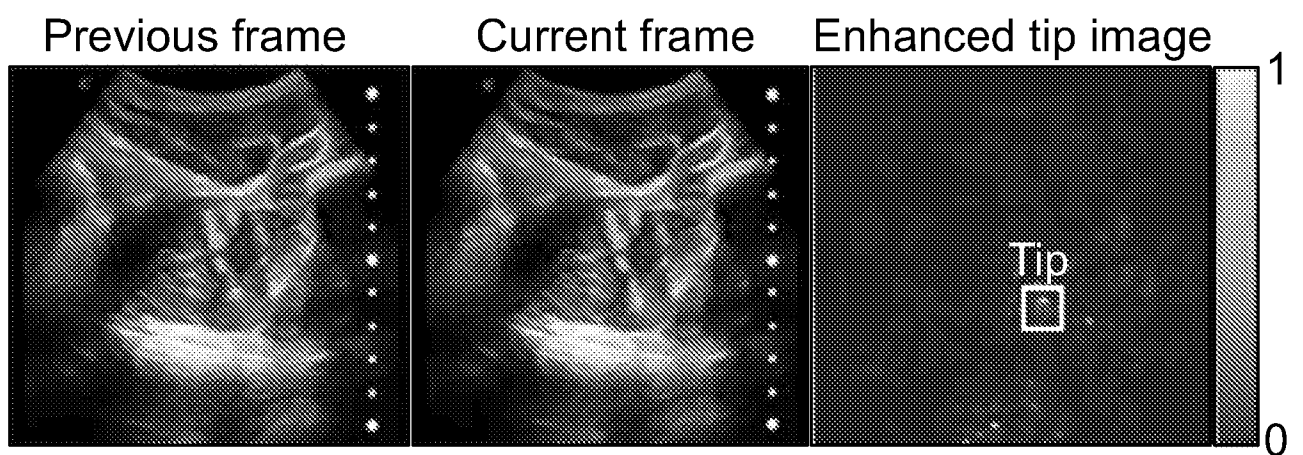


FIG. 12

240 ↗

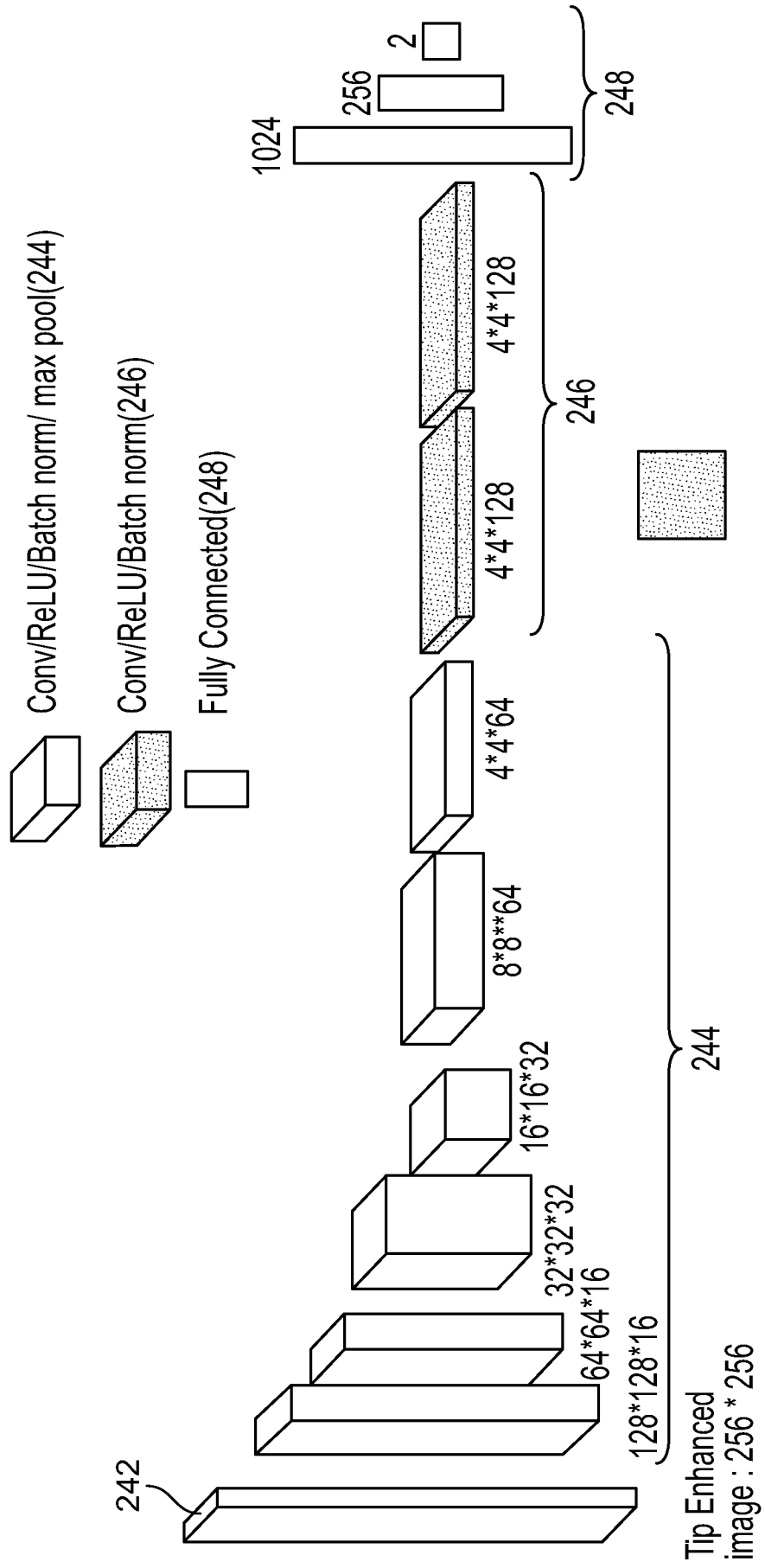


FIG. 13

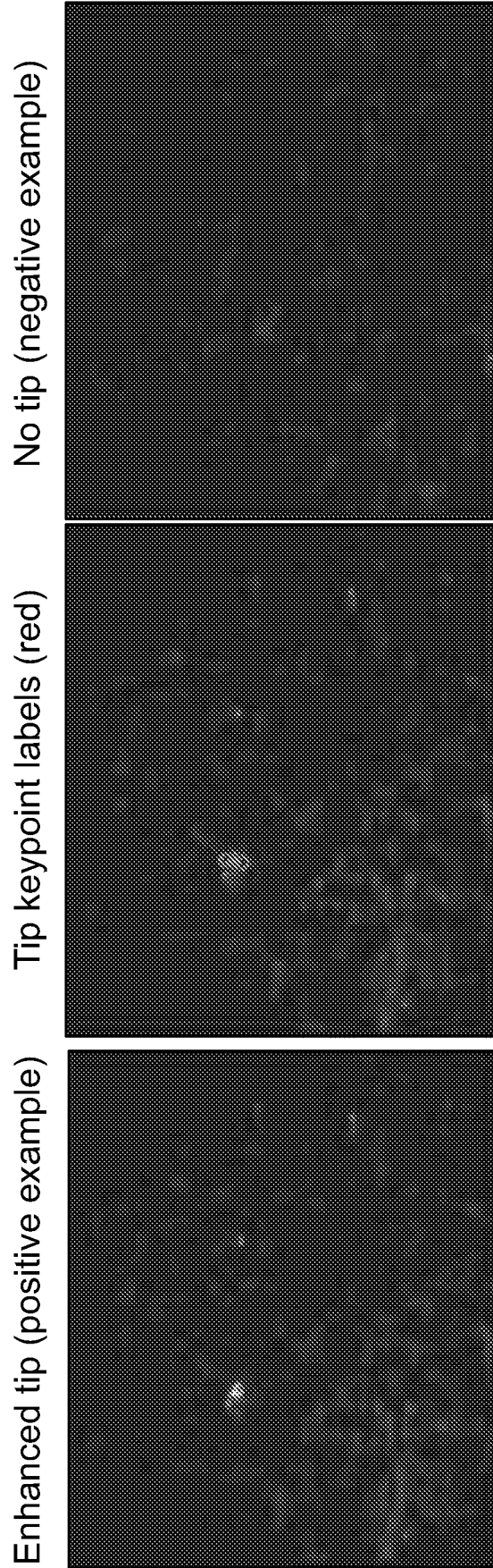


FIG. 14

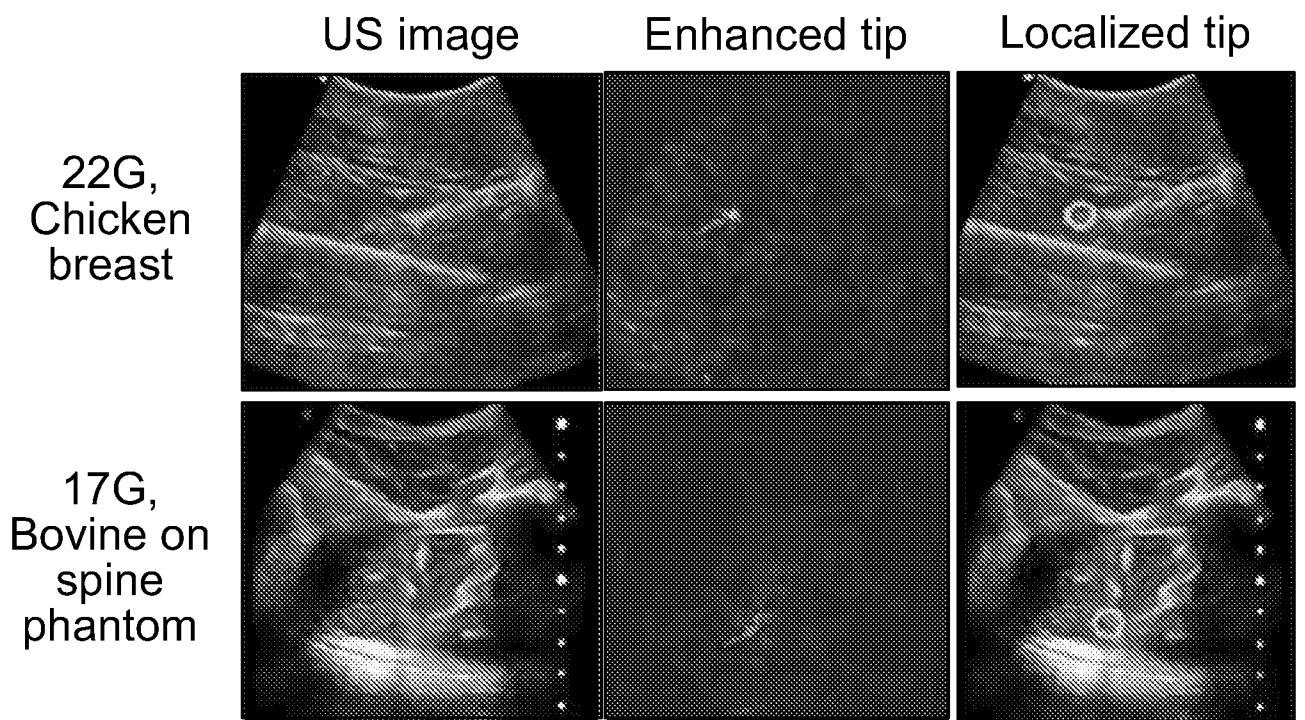


FIG. 15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US19/46364

A. CLASSIFICATION OF SUBJECT MATTER

IPC - A61B 8/08, 8/12, 5/06 (2019.01)

CPC - A61B 8/08, 8/0841, 8/12, 17/3403, 5/01

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	MWIKIRIZE, C. et al. "Learning needle tip localization from digital subtraction in 2D ultrasound" International Journal of Computer Assisted Radiology and Surgery (2019) 14:1017-1026, 25 Mach 2019 <DOI:0.1007/s11548-019-01951-z>.	1-29
A	US 2007/0167801 A1 (WEBLER, W. et al.) 19 July 2007; entire document.	1-29
A	DONG, B. Et al. "A NOVEL METHOD FOR ENHANCED NEEDLE LOCALIZATION USING ULTRASOUND-GUIDANCE" ISVC 2009: Advances in Visual Computing pp 914-923; 30 November 2009; <DOI: 10.1007/978-3-642-10331-5_85>	1-29
A	CN 104248454 B (HARBIN ENGINEERING UNIVERSITY) 29 June 2016; see machine translation; entire document.	1-29
A	RENFREW, M. et al. "Active Localization and Tracking of Needle and Target in Robotic Image-Guided Intervention Systems" Auton Robots; 42(1): 83-97; 12 June 2017 <doi:10.1007/s10514-017-9640-2>.	1-29
A	NEUBACH, Z. et al. "ULTRASOUND-GUIDED ROBOT FOR FLEXIBLE NEEDLE STEERING" IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 57, NO. 4, APRIL 2010; first bulished 25 August 2009; <DOI: 10.1109/TBME.2009.2030169>.	1-29

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"D" document cited by the applicant in the international application	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"E" earlier application or patent but published on or after the international filing date	"&" document member of the same patent family
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

15 October 2019 (15.10.2019)

Date of mailing of the international search report

29 OCT 2019

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

Telephone No. PCT Helpdesk: 571-272-4300