Feature Study on Catheter Detection in Three-Dimensional Ultrasound

Hongxu Yang[a], Arash Pourtaherian[a], Caifeng Shan[b], Alexander F. Kolen[b], Peter H.N.de With[a]

[a]VCA Research Group, Eindhoven University of Technology, Eindhoven, The Netherlands
[b]In-body System, Philips Research, Eindhoven, The Netherlands

ABSTRACT

The usage of three-dimensional ultrasound (3D US) during image-guided interventions for e.g. cardiac catheterization has increased recently. To accurately and consistently detect and track catheters or guidewires in the US image during the intervention, additional training of the sonographer or physician is needed. As a result, image-based catheter detection can be beneficial to the sonographer to interpret the position and orientation of a catheter in the 3D US volume. However, due to the limited spatial resolution of 3D cardiac US and complex anatomical structures inside the heart, image-based catheter detection is challenging. In this paper, we study 3D image features for image-based catheter detection using supervised learning methods. To better describe the catheter in 3D US, we extend the Frangi vesselness feature into a multi-scale Objectness feature and a Hessian element feature, which extract more discriminative information about catheter voxels in a 3D US volume. In addition, we introduce a multi-scale statistical 3D feature to enrich and enhance the information for voxel-based classification. Extensive experiments on several in-vitro and ex-vivo datasets show that our proposed features improve the precision to at least 69% when compared to the traditional multi-scale Frangi features (from 45% to 76% at a high recall rate 75%). As for clinical application, the high accuracy of voxel-based classification enables more robust catheter detection in complex anatomical structures.

Keywords: Catheter detection, multi-scale feature, machine learning, 3D ultrasound

1. INTRODUCTION

In the last years, minimally invasive therapy such as interventional cardiac catheterization, has been widely used with the benefit of minimized risk and shorter recovery time compared to traditional open surgery. During interventional procedures, fluoroscopic imaging or ultrasound (US) are essential for guiding catheters, guidewires and other devices. With the introduction of 3D TEE (Transesophageal Echography), the number of ultrasound guided interventions have increased. However, 3D US suffers drawbacks like low signal-to-noise ratio and low resolution in imaging, which make catheter visualization in 3D US challenging. Given the increased number of US guided interventions, automated catheter detection will be particularlly helpful in 3D US imaging. This motivates the study of this paper.

US image-based medical tool detection has been studied over the last years. Unsupervised methods were used to detect the straight form factor of medical instruments in 3D US.[1–3] However, these methods lack discriminative information or require prior knowledge of the instrument's direction and diameter. Alternatively, Uherčík and Zhao applied the Frangi filter feature to detect the voxels indicating the instruments with supervised learning algorithms.[4,5] Based on the classified result, a RANdom Sample Consensus (RANSAC) was applied to fit the instruments in the ultrasound image. Although their results show the benefit of discriminative information, many issues still need to be addressed. For instance, the spatial parameters of their filters are decided empirically, which gives unstable behavior when the instrument's diameter has a large variance. In addition, the discriminative feature they described is limited to the tubular shape only. This feature cannot handle the complex background information in US. Recently, Pourtaherian et al. proposed a Gabor feature to detect needle voxels in 3D US.[6–8] The Gabor feature describes multiple orientations and frequencies as discriminative information, which is combined with voxel-based classification, giving a performance improvement from 30% to a precision of 61%

Further author information:
Hongxu Yang: E-mail: h.yang@tue.nl

using a silicon heart. Although the above instrument detection methods achieved valuable clinical results, many challenges remain for a cardiac intervention application. Firstly, the cardiac transducers have a lower spatial resolution compared to US transducers used in superficial needle detection just below the skin surface. Secondly, the heart has more complex anatomical structures, which makes the catheter detection become more challenging. Summarizing, the voxel-based classification methods are not sufficiently robust for our cases, so that new methods have to be considered.

In this paper, we study the voxel-based features for image-based catheter detection via supervised learning methods. To describe the discriminating information in complex 3D US signals, we extend the Frangi vesselness feature into multi-scale Objectness[9] and multi-scale Hessian features. Additionally, based on the catheter appearance, we introduce multi-scale *statistical* features, which enrich the discriminative information for supervised classification. When compared to the multi-scale Frangi feature on different datasets, our proposed features improve the precision to at least 69% (from 45% to 76% when recall is 75%) for voxel-based classification in 3D cardiac US.

2. METHODS

The processing steps towards catheter detection in 3D US are illustrated in Figure 1. The acquired 3D US volume is processed to extract the feature vector of each voxel. These vectors are passed into a pre-trained machine learning model, which classifies the voxels as catheter or not. Based on this, a RANSAC catheter-fitting algorithm is performed to localize the catheter.[10] In this work, we primarily focus on selecting and improving features to enhance the voxel classification performance.

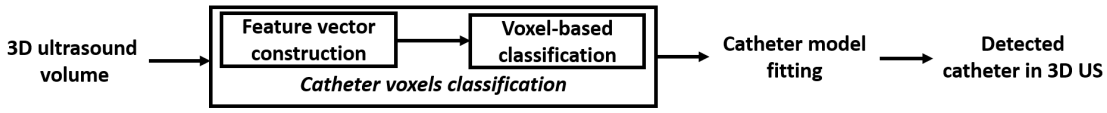


Figure 1: Catheter detection steps in 3D US

For each voxel v in 3D volume V, the Frangi feature vector is constructed from voxel intensity $I(v)$ and Frangi response $J(v,\sigma)$, i.e. $F_L(v,\sigma) = [I(v), J(v,\sigma)]$ with $v \in V$ and at scale σ. The Frangi response extracts tubular shape information around the voxel based on a Gaussian-filtered image. This filter is assigned a kernel scale, see the third shape example in the Objectness image shown at the left part of Figure 2. However, this type of shape description is not sufficient to describe the catheter voxels inside a complex anatomical environment. To address this, we adopt the Objectness description from Antiga.[9] Under a given scale σ, the Objectness vector is constructed as $F_O(v,\sigma) = [I(v), Obj_{ball}(v,\sigma), Obj_{tube}(v,\sigma), Obj_{plate}(v,\sigma)]$. The definition of Objectness (including the Frangi feature) is given as follows.

The Objectness is constructed based on the Hessian matrix for a given V, which is described in Eqn. (1). The partial derivatives in the Hessian matrix of Eqn. (1) are computed from a 3D Gaussian-filtered V at scale σ:

$$H_\sigma = \begin{bmatrix} f^\sigma_{xx} & f^\sigma_{xy} & f^\sigma_{xz} \\ f^\sigma_{yx} & f^\sigma_{yy} & f^\sigma_{yz} \\ f^\sigma_{zx} & f^\sigma_{zy} & f^\sigma_{zz} \end{bmatrix}. \tag{1}$$

As a result, based on the Eigenvalues of the Hessian matrix, i.e. $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$, the Objectness response is constructed by applying Eqns. (2) and (3). Parameter M indicates the shape, e.g. $M = 0$ for blob, $M = 1$ for tube and $M = 2$ for plate (Frangi equals to $M = 1$). Parameters α, β and γ are chosen empirically. Objectness (Obj_M) is specified by

$$Obj_M(v,\sigma) = \begin{cases} 0, & \text{else} \\ (1 - e^{-\frac{R_A^2}{2\alpha^2}}) \cdot e^{-\frac{R_B^2}{2\beta^2}} \cdot (1 - e^{-\frac{S^2}{2\gamma^2}}), & \text{for} \quad M < i \leq 3 \quad \text{and} \quad \lambda_i < 0. \end{cases} \tag{2}$$

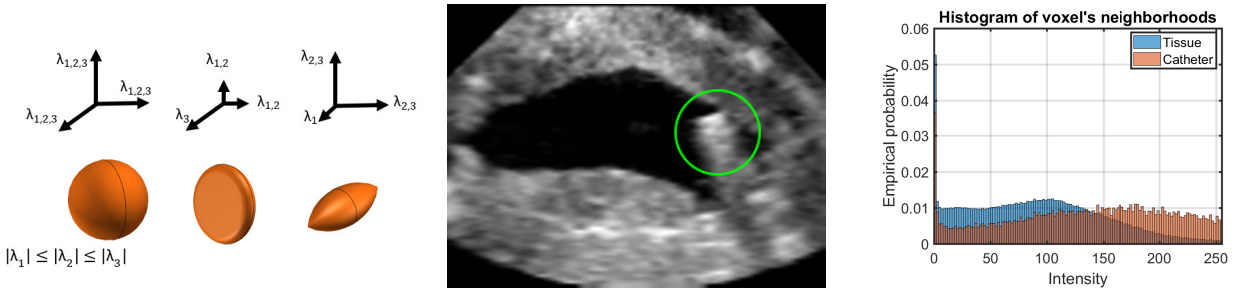
In Eqn. (2), the parameters in the exponentials R_A, R_B and S are defined by

$$\begin{cases} R_A &= \frac{|\lambda_{M+1}|}{\prod_{i=M+2}^{3}|\lambda_i|^{\frac{1}{3-M-1}}}, \\ R_B &= \frac{|\lambda_M|}{\prod_{i=M+1}^{3}|\lambda_i|^{\frac{1}{3-M}}}, \\ S &= \sqrt{\sum_{i=1}^{3}\lambda_i^2}. \end{cases} \tag{3}$$

The Eigenvalues λ_i are determined by the above Hessian matrix. By considering different shapes in the voxel neighborhoods, the Objectness features extract more information and describe the shape information using the Eigenvalue analysis of the Hessian matrix. However, this Eigenvalue analysis introduces an information loss. To preserve more information, we consider the Hessian matrix elements as features. The feature vector is constructed by $F_H(v,\sigma) = [I(v), \mathcal{H}(v,\sigma)]$ under a given kernel scale. To construct the part $\mathcal{H}(v,\sigma)$, the upper right 6 elements from Eqn. (1) are first extracted into a vector with pre-defined order, as shown in Eqn. (4), giving

$$\mathcal{H}(v,\sigma) = [f_{xx}^{\sigma}(v), f_{xy}^{\sigma}(v), f_{xz}^{\sigma}(v), f_{yz}^{\sigma}(v), f_{zz}^{\sigma}(v), f_{yy}^{\sigma}(v)]. \tag{4}$$

Then a circular shifting operation is applied to $\mathcal{H}(v,\sigma)$ for shifting the maximum value to the first position, which ensures the orientation-invariance property and preserves response distribution around the dominant direction.



(a) (b) (c)

Figure 2: (a): Objectness describes shapes by Eigenvalue analysis. (b): Example of catheter appearance in 2D slice. (c): Example of intensity distributions of the neighborhoods of catheter voxels and tissue voxels.

The above Frangi, Objectness and Hessian features are constructed based on Gaussian-filtered volumes, which are controlled by the scale size of the Gaussian kernel. The conventional method uses a specific scale size based on prior knowledge of the tool diameter, so-called *single-scale*. Nevertheless, this pre-designed filter only focuses on information of tools, while ignoring the information of the tissue or background at different scales. Consequently, we consider a *multi-scale* method for our problem. This involves different scales simultaneously, i.e. a scale size ranging from 1 to N voxels.

Finally, the observation of catheter appearance shows that the neighborhoods of catheter voxels have a different intensity distribution than the neighborhoods of tissue voxels, see e.g. the middle and the right examples in Figure 2. Therefore, we further employ *statistical* metrics to measure the distribution difference. For any voxel v at the center of a cube C, such as a cube with $7 \times 7 \times 7$ voxels with $v \in C \in V$, the proposed *Statistical* feature vector is constructed as $F_S(v) = [I(v), Max(C), Min(C), Mean(C), STD(C), Entropy(C)]$. Based on our proposed features, their combination is defined as $F_C = [F_O, F_H, F_S]$.

After feature vector extraction, a classifier is trained on annotated datasets to separate the catheter voxels from the rest when observing a new US image. In our study, Linear Discriminant Analysis (LDA), Linear Support Vector Machine (LSVM), Random Forest (RF) and Adaptive Boosting (AdaBoost) are employed to evaluate the feature performance for voxel classification.[11] Based on the classified volume, the SPD-RANSAC[10] algorithm is applied to localize the catheter in the US image.

3. EXPERIMENTAL RESULTS

Data acquisition was applied in in-vitro and ex-vivo situations, see Table 1 for more details. During the recording, two RF-ablation catheters from different manufacturers were used, but both of them have the same 2.3-mm diameter. The volumes are re-sampled to obtain the same resolution in each direction. The datasets are manually annotated and confirmed by an expert. For each dataset, Leave-One-Out Cross-Validation (LOOCV) is performed to fully exploit the limited datasets. During the training stage, the non-catheter voxel samples are down-sampled to obtain the same size as the catheter voxel samples. In the testing stage, the testing volume is maintained at original size. In order to evaluate the discriminating power of the proposed features on the imbalanced test volume, we employ precision, recall, F-1 score and specificity as evaluation metrics. Although these metrics evaluate the voxel-based classification, which does not reflect the ultimate catheter detection performance, the comparison of these metrics enables to identify the optimal feature and classifier combination. After the classification step, the catheter model is fitted by the SPD-RANSAC algorithm to localize the catheter.[10] This algorithm becomes more challenging when the catheter voxels are falsely classified, i.e. the less catheter voxels are detected, the more difficult to localize the catheter. As a result, we expect to obtain a high recall value in the voxel-based classification, which is set to be 75% in this work. In our experiments, the parameters α, β and γ are defined as follows and given by

$$\begin{cases} M = 0: & \alpha = 0.01, \beta = 1.0, \gamma = 0.5; \\ M = 1: & \alpha = 0.05, \beta = 1.0, \gamma = 0.5; \\ M = 2: & \alpha = 0.05, \beta = 0.5, \gamma = 0.5. \end{cases} \tag{5}$$

Table 1: Recording conditions used to acquire 3D US data.

Data	Recording Condition	Number of Acquisitions	US System	Transducer Type & Frequency Range	Voxel Size per Dimension	Volume Size (lat.×az.×ax.)	Catheter Voxel #. vs. Non-Catheter Voxel #.
Dataset 1 (*in-vitro*)	Rubber (PVA) Heart Phantom	20	Philips EpiQ 7	3D Phase Array 2-7 MHz	0.6 mm	$141 \times 168 \times 101 \sim 145 \times 185 \times 101$	$\sim 1/1250$
Dataset 2 (*ex-vivo*)	Porcine Heart	10	Philips CX 50	3D Phase Array 2-7 MHz	0.4 mm	$179 \times 175 \times 92$	$\sim 1/2500$
Dataset 3 (*ex-vivo*)	Porcine Heart	12	Philips EpiQ 7	3D Phase Array 1-5 MHz	0.7 mm	$137 \times 130 \times 122$	$\sim 1/1800$

Figure 3 depicts the average precision value under different features and different scale settings, where only AdaBoost is shown here due to its high performance. From these images, several conclusions are made. Firstly, the Objectness has a higher performance than the traditional Frangi feature, especially in the multi-scale case. This improvement occurs because Objectness extracts more shape information from its definition. Secondly, due to the fact that Hessian features lose less information, they perform better than the Frangi and Objectness features. As for statistical features, they show less variance in size variations when compared to the Objectness and Hessian features in the single-scale cases. This is due to the Gaussian-based features, which only extract the information at a specific frequency range. This is reflected by the curve degradation in single-scale condition (catheter radius is around 1.2 mm, giving limited scale size 2∼3). Hence, the catheter information is extracted properly and the curve shows a maximum. When the scale size is exceeding 6, the features are focusing on tissue information rather than the catheter, so that the performance increases again. When using a larger range of scales, all features exhibit a saturation behavior in the multi-scale case.

Table 2 shows more details with optimized multi-scale range, while including the saturation phenomenon, i.e. multi-scale from 1 to 10 in Gaussian-based features and multi-scale from 1 to 7 in statistical features, while keeping a high precision value. The parameter F_C represents the combination of the proposed three multi-scale features. Additionally, Figure 4 depicts a comparison of the classifiers, when adopting F_C. Several observations are made from Table 2. Firstly, the performances in ex-vivo datasets are worse than the in-vitro values. This is because a real heart has more complex data structures featuring more challenging information for extraction. Secondly, due to the lower spatial resolution (i.e. lower frequency range) in Dataset 3, the traditional Frangi feature cannot extract sufficient information for classification, even under multi-scale condition. Thirdly, even the Objectness feature performs better than the Frangi feature, but due to information loss by Eigenvalue extraction,

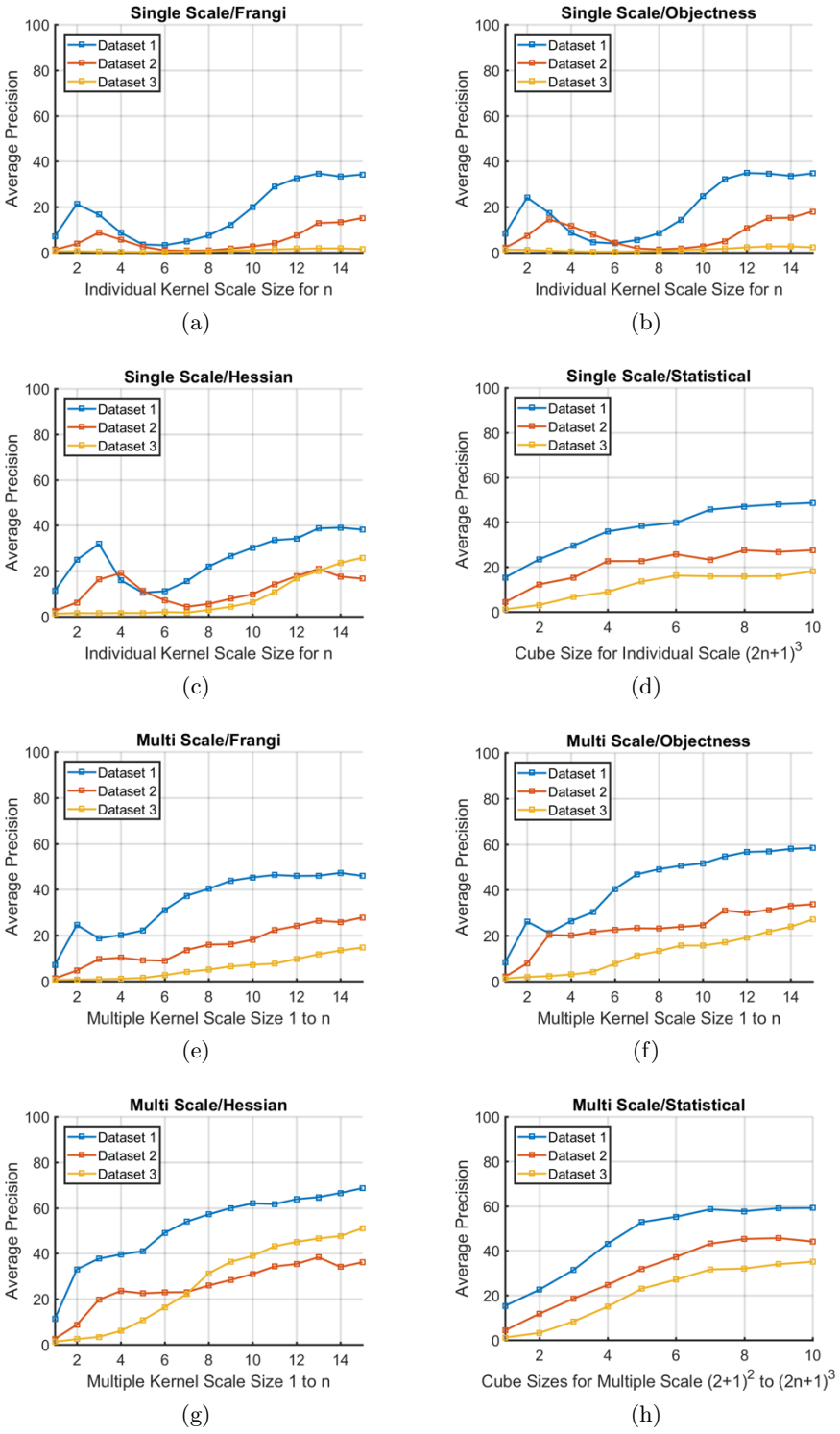


Figure 3: Average precision values (in %) in different settings. The (a)-(d) show a single-scale method, while (e)-(h) shows a multi-scale method. The x-axis units are voxels. Only AdaBoost results are illustrated.

Table 2: Average classification performance when recall=75%, only the best classifier is shown. Numbers are mean and (standard deviation), and partly formatted for the table alignment, not accuracy.

Dataset		Adaptive Boosting				
		F_L	F_O	F_H	F_S	F_C
Dataset 1 (*in-vitro*)	Precision	45.3 (25.7)	51.6 (6.50)	62.0 (20.8)	58.5 (24.1)	**75.6 (15.7)**
	F-1 Score	52.4 (20.2)	57.1 (19.2)	65.8 (14.0)	62.6 (18.4)	**74.4 (9.00)**
	Specificity	99.9 (0.10)	99.9 (0.06)	99.9 (0.05)	99.9 (0.06)	**99.9 (0.01)**
Dataset 2 (*ex-vivo*)	Precision	18.1 (7.20)	24.5 (13.0)	30.9 (18.8)	43.1 (11.3)	**48.8 (14.5)**
	F-1 Score	28.5 (0.60)	35.3 (14.4)	40.6 (20.2)	53.9 (10.0)	**57.9 (11.3)**
	Specificity	99.8 (0.10)	99.9 (0.10)	99.9 (0.20)	99.9 (0.03)	**99.9 (0.03)**
Dataset 3 (*ex-vivo*)	Precision	7.20 (4.30)	15.7 (7.30)	38.9 (14.5)	31.6 (12.2)	**52.9 (6.60)**
	F-1 Score	12.8 (6.90)	25.2 (9.80)	49.4 (15.7)	43.1 (12.7)	**61.8 (4.60)**
	Specificity	99.2 (0.70)	99.7 (0.20)	99.8 (0.30)	99.9 (0.10)	**99.9 (0.01)**

the Objectness feature is worse than the Hessian elements feature. In all three datasets, the Hessian elements feature has the best performance with Gaussian-based features. Finally, the statistical feature performs similar to the Hessian elements feature, but it is independent of Gaussian-based features, so that the combination of the proposed features, i.e. F_C, shows a further performance improvement. Figure 4 indicates that AdaBoost is the optimal classifier. The LSVM and RF classifiers (worse variance) perform similar and close to AdaBoost, but are sub-optimal classifiers. The LDA classifier has a large performance drop, which reflects that the features are not following the normalized distribution, when compared to the other three classifiers.

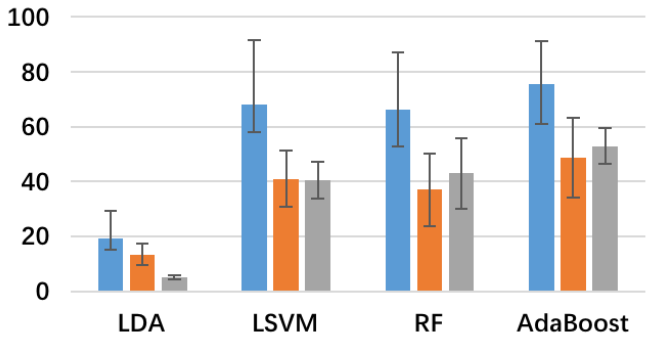
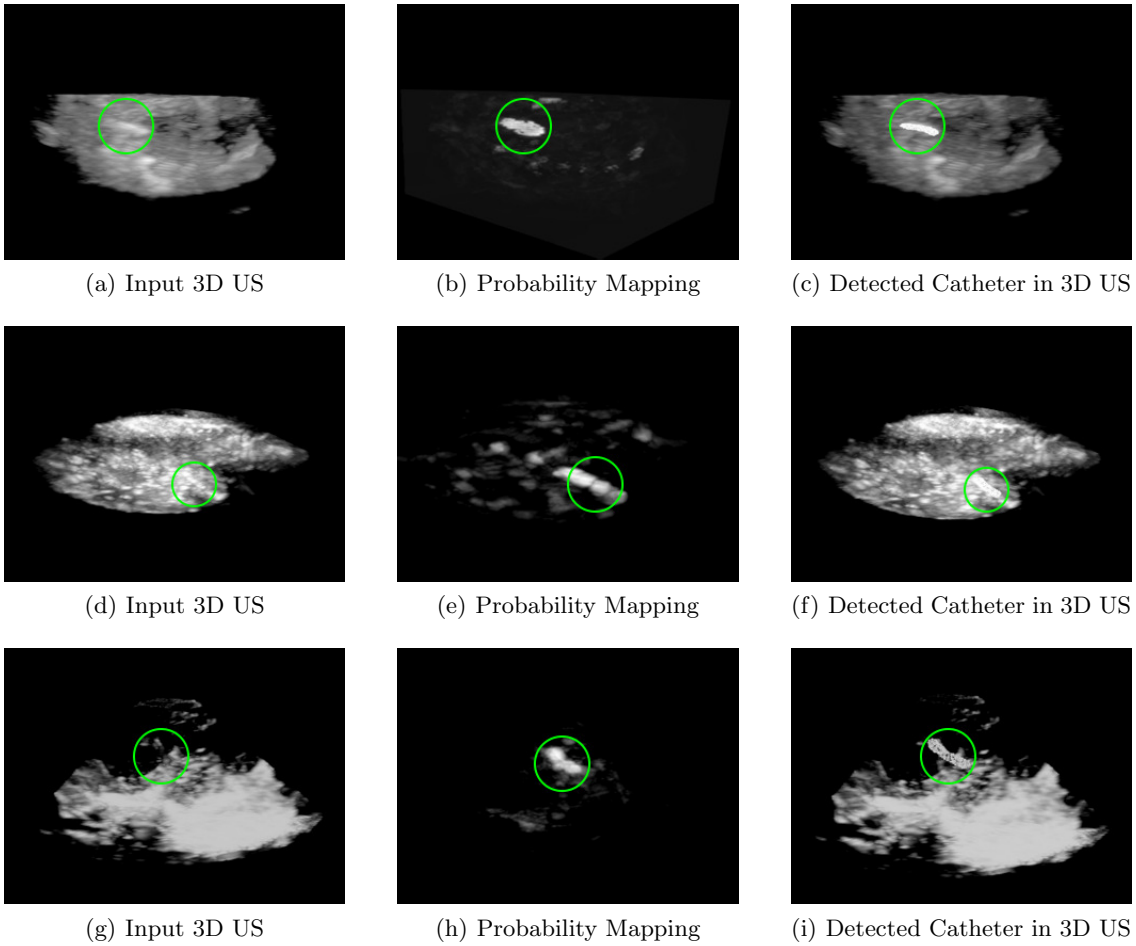


Figure 4: Average precision (in %) for 4 different classifiers. The colors blue, orange and gray refer to Datasets 1, 2 and 3, respectively.

Although the voxel-based classification is improved, it is not sufficient to directly localize the catheter. However, due to the knowledge of the catheter's global shape inside the 3D volume, the SPD-RANSAC algorithm can quickly and robustly localize the catheter in the 3D US volumes. In our case, the inlier versus outlier threshold is set to be 3 voxels. Figure 5 depicts examples from three datasets. The images are: input 3D US, probability mapping of voxels from classifier, and detected catheter inside the volume, visualized from left to right, respectively. Figure 6 illustrates some example slices from processed volumes, in which red color refers to the ground truth while the green color depicts the detected catheter.

4. CONCLUSION

Image-based catheter detection supports cardiac catheterization interventions and potentially improves clinical outcome. For this purpose, we have developed an image-based catheter detection algorithm, exploiting three multi-scale features, i. e. Objectness, Hessian elements and Statistical features, all suited for 3D US volumes.

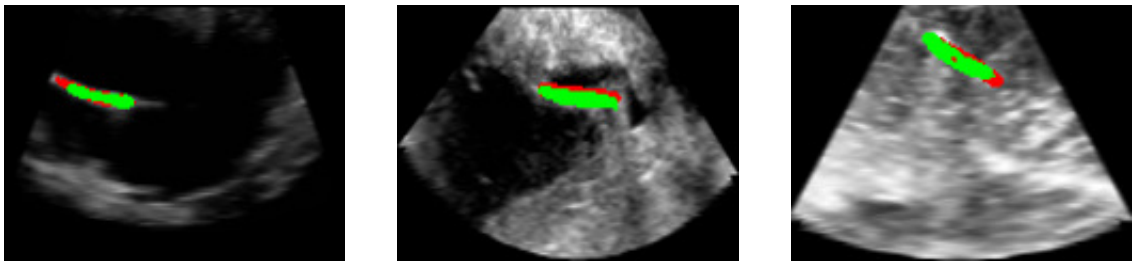


(a) Input 3D US (b) Probability Mapping (c) Detected Catheter in 3D US

(d) Input 3D US (e) Probability Mapping (f) Detected Catheter in 3D US

(g) Input 3D US (h) Probability Mapping (i) Detected Catheter in 3D US

Figure 5: Example of catheter detection. (a)-(c) are from dataset 1, (d)-(f) are from dataset 2 while (g)-(i) are from dataset 3. The catheter is detected and highlighted inside the volume (in the green circles). Images are rendered by MIP.[12]



(a) Dataset 1 (b) Dataset 2 (c) Dataset 3

Figure 6: Example slices from three different datasets. The red represents the ground truth while the green represents the detected results.

Based on multi-scale processing with different feature definitions, the proposed features extract more discriminative information than the traditional Frangi vesselness filter, which is substantiated as a precise shape definition, more accurate orientation information and local statistical properties. We have computed the voxel-based classification scores achieving a precision of 76% with the in-vitro phantom dataset and 50% with the ex-vivo porcine heart datasets, at a high recall rate of 75%. When compared to multi-scale Frangi features, the voxel-based precision performance improves to at least 69% with an adaptive boosting classifier (similarly, it boosts the results from 45% to 76% with the in-vitro dataset). As for clinical application, the high accuracy of voxel-based classification enables more robust catheter detection, which is particularly important for navigating through complex anatomical environments like the cardiac structure.

REFERENCES

[1] Linguraru, M. G., Vasilyev, N. V., Del Nido, P. J., and Howe, R. D., "Statistical segmentation of surgical instruments in 3-D ultrasound images," *Ultrasound in medicine & biology* **33**(9), 1428–1437 (2007).

[2] Aboofazeli, M., Abolmaesumi, P., Mousavi, P., and Fichtinger, G., "A new scheme for curved needle segmentation in three-dimensional ultrasound images," in [*IEEE International Symposium on Biomedical Imaging (ISBI).*], 1067–1070, IEEE (2009).

[3] Cao, K., Mills, D., and Patwardhan, K. A., "Automated catheter detection in volumetric ultrasound," in [*IEEE 10th International Symposium on Biomedical Imaging (ISBI)*], 37–40, IEEE (2013).

[4] Uherčík, M., Kybic, J., Zhao, Y., Cachard, C., and Liebgott, H., "Line filtering for surgical tool localization in 3D ultrasound images," *Computers in biology and medicine* **43**(12), 2036–2045 (2013).

[5] Zhao, Y., Cachard, C., and Liebgott, H., "Automatic needle detection and tracking in 3D ultrasound using an ROI-based RANSAC and Kalman method," *Ultrasonic imaging* **35**(4), 283–306 (2013).

[6] Pourtaherian, A., Zinger, S., de With, P. H. N., Korsten, H. H., and Mihajlovic, N., "Gabor-based needle detection and tracking in three-dimensional ultrasound data volumes," in [*IEEE International Conference on Image Processing (ICIP)*], 3602–3606, IEEE (2014).

[7] Pourtaherian, A., Zinger, S., Mihajlovic, N., de With, P. H. N., Huang, J., et al., "Multi-resolution gabor wavelet feature extraction for needle detection in 3D ultrasound," in [*Eighth International Conference on Machine Vision*], 987513–987513, International Society for Optics and Photonics (2015).

[8] Pourtaherian, A., Scholten, H., Kusters, L., Zinger, S., Mihajlovic, N., Kolen, A., Zou, F., Ng, G., Korsten, H., and de With, P. H. N., "Medical instrument detection in 3-dimensional ultrasound data volumes," *IEEE Transactions on Medical Imaging* (2017).

[9] Antiga, L., "Generalizing vesselness with respect to dimensionality and shape," *The Insight Journal* **3**, 1–14 (2007).

[10] Papalazarou, C., de With, P. H. N., and Rongen, P., "Sparse-plus-dense-RANSAC for estimation of multiple complex curvilinear models in 2D and 3D," *Pattern Recognition* **46**(3), 925–935 (2013).

[11] Caruana, R. and Niculescu-Mizil, A., "An empirical comparison of supervised learning algorithms," in [*Proceedings of the 23rd international conference on Machine learning*], 161–168, ACM (2006).

[12] Napel, S., Marks, M., Rubin, G., Dake, M., McDonnell, C., Song, S., Enzmann, D., and Jeffrey Jr, R., "CT angiography with spiral CT and maximum intensity projection.," *Radiology* **185**(2), 607–610 (1992).