Liver Ultrasound Tracking Using Kernelized Correlation Filter With Adaptive Window Size Selection

Satoshi Kondo

Konica Minolta Inc., Osaka, Japan
satoshi.kondo@konicaminolta.com

Abstract. We propose a method to track tissues in long ultrasound sequences of liver. The proposed method is based on kernelized correlation filter (KCF) and we introduce two extensions to KCF; adaptive window size selection and motion vector refinement with template matching. We compare KCF and the proposed method by using some training sequences of 2D ultrasound and the mean tracking error can be improved with the proposed method by up to nearly 3 pixels. The tracking performance is also assessed on 19 test sequences of 2D ultrasound with 62 regions of interests. Mean tracking error is 1.09 mm.

Keywords: Ultrasound, Liver, Tracking, Kernelized correlation filter, Template matching

1 Introduction

It is important to track a region of interest (ROI) to compensate motion to ensure accuracy of robot-assisted diagnosis [1], focused ultrasound surgery [2] and dose delivery in radiation therapies [3]. Ultrasound is one of potential imaging modalities for image guidance and has some advantages such as real-time imaging, noninvasive and cheap comparing to other imaging modalities such as CT and MRI.

Various methods have been proposed for tracking a moving object in a video sequence. In recent years, tracking methods using discriminative approach have been proposed and reported to exhibit high performance [4–6]. Especially, Kernelized Correlation Filter (KCF) is known to show high performance despite its high speed processing [7, 8].

KCF shows high tracking performance, but it has some problems such as, (1) the user has to specify a region enclosing the target object to track, (2) KCF emphasizes robustness than accuracy. For example, the criteria of true positive is that the tracked position is within 20 pixels relative to the ground truth in [7].

On the other hand, in medical applications, both robustness and accuracy are required for tissue tracking. In this paper, we propose a tracking method of tissues in long ultrasound sequences of liver. The proposed method is based on

KCF and we introduce two extensions to KCF. The first one is adaptive window size selection and the second one is motion vector refinement.

2 Overview of Kernelized Correlation Filter

In this section, we briefly explain KCF, which is a basis of our proposed method. In KCF, when a target object to track is specified, a discriminative function is calculated by kernel ridge regression using the image inside the region of the target to track as a positive sample and the images in the surrounding region of the target as negative samples. Since the positive sample and the negative samples are expressed with a circulant matrix in KCF, the regression coefficient vector in the kernel space is obtained by using Discrete Fourier Transform (DFT),

$$\hat{\boldsymbol{\alpha}}^* = \frac{\hat{\mathbf{y}}}{\hat{\mathbf{k}}^{\mathbf{xx}} + \lambda}, \tag{1}$$

where $\mathbf{y}$ is a regression target vector, 1 for an element corresponding to a positive sample and 0 for an element corresponding to a negative sample, $\mathbf{x}$ is an image patch in a tracked region, λ is a regularization weight in ridge regression, a hat $\hat{}$ and a star * denote the DFT of a vector and complex-conjugate, respectively. In the case of two dimensional data and the dimension of $\mathbf{x}$ is $M \times N$, the dimension of $\mathbf{y}$ is also $M \times N$. Note that the tracking window, which is the size of $\mathbf{x}$, has 2.5 times the size of the target to track in the implementation of KCF.

In case of Gaussian kernel, $\hat{\mathbf{k}}^{\mathbf{xx}'}$ is

$$\hat{\mathbf{k}}^{\mathbf{xx}'} = \exp\left(-\frac{1}{\sigma^2}\left(\|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2 - 2\mathcal{F}^{-1}(\hat{\mathbf{x}} \odot \hat{\mathbf{x}}'^*)\right)\right), \tag{2}$$

where $\mathcal{F}^{-1}$ and $\odot$ denote inverse Fourier transform and element-wise multiplication, respectively.

In the detection phase, a regression function in Eq. (3) is calculated and the position where the regression value is maximum is the tracked position.

$$\hat{\mathbf{f}}(\mathbf{z}) = \left(\hat{\mathbf{k}}^{\mathbf{xz}}\right)^* \odot \hat{\boldsymbol{\alpha}}, \tag{3}$$

where $\mathbf{z}$ is a image patch in a frame to track which has the same size with $\mathbf{x}$.

Then, $\hat{\boldsymbol{\alpha}}$ is re-calculated at the new tracked position in the next frame using Eq. (1). In the implementation, however, $\hat{\boldsymbol{\alpha}}$ is gradually updated as in Eq. (4).

$$\hat{\boldsymbol{\alpha}}'_{t+1} = \beta\hat{\boldsymbol{\alpha}}_{t+1} + (1-\beta)\hat{\boldsymbol{\alpha}}_t, \tag{4}$$

where β is a weight for the interpolation.

3 Proposed Method

3.1 Overview

It is desirable that a target area in the object tracking is set to enclose the target object in the first frame by a user. However, in some cases, the user may specify

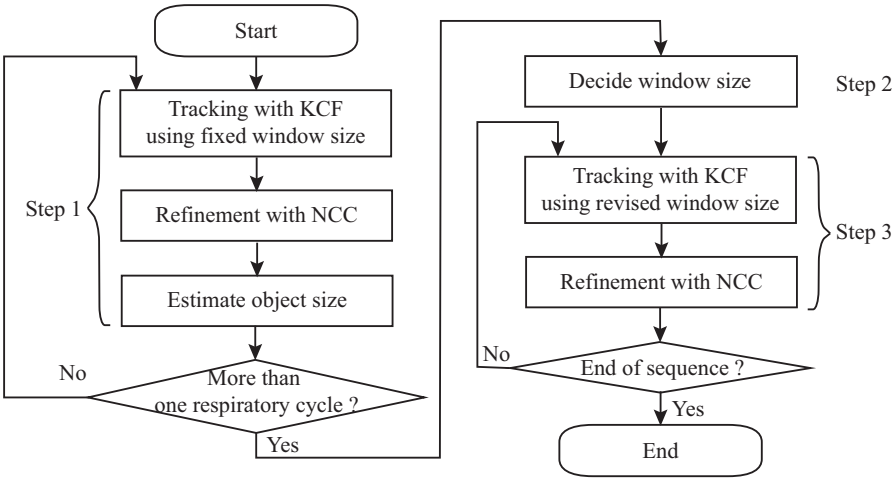

Fig. 1. A flowchar of the proposed method.

only the center position of the target object. In such cases, the tracking system has to decide the region of the target object to successfully track the object. Our proposed method is based on KCF and the size of the tracking window should be set in consideration of the following two aspects. The first one is the size of the tracking target object and the second one is the amount of motion of the target object.

In KCF, a discriminative function is determined using the image inside the region of the target object as a positive sample and the image in the surrounding region of the target object as negative samples. Therefore, it is desirable that the image of the target object and the image in the surrounding region have different texture. KCF calculates the correlation in the Fourier domain in the tracking process as described in Section 2. That means the amount of the motion should be within the area of the Fourier transform which is the same as the size of the tracked window. Also, the ultrasound images of the liver have the characteristic that the motion is approximately periodic which is induced by respiration.

In the proposed method, the size of the target object and the maximum amount of motion are obtained by using initial frames during about one breathing cycle (Step 1). This is a kind of calibration phase. The size of the tracking window is decided by using the size of the target object and the maximum value of the motion vectors (Step 2). In the subsequent frames, tracking is performed with the tracking window (Step 3).

Fig. 1 is a flowchart of the proposed method. We will describe the details of each process in the following sections.

3.2 Tracking using fixed window size for initial frames (Step 1)

In Step 1, tracking is performed by KCF with a predetermined size of the tracking window at first for each frame. The predetemined window has a rectangular area

centered at the point designated by the user to the target object. Then, in order to improve the tracking precision, the template matching is performed for a narrow search region around the position detected by KCF. We use normalized cross correlation to evaluate the template matching.

Note that the learned discriminative function is gradually updated in KCF as mentioned in Section 2. Also, we do not update the template and use the same template for each frame in the template matching. The template is surrounding area of the target object in the first frame. The reason is to use the template obtained in the first frame for each frame is to avoid drift. Template matching is performed only in the region of around $\pm$ 2 pixels of the positions obtained by KCF. When the maximum normalized correlation value is greater than a threshold C_{th}, the result of the template matching is adopted. Otherwise, the tracking result of KCF is adopted without refinement. The area of performing the template matching is decided empirically.

Also, we estimate the size of the target object in every predetermined frame. Since the target object is a tissue such as blood vessels or tumors, the size of the target object is determined by the ellipse fitting in the proposed method.

We repeat the above tracking process for each frame and analyze the temporal history of the tracked positions. When the target object reaches the right (or left) end two times, we determine that one respiratory cycle has been passed. When it is determined that one respiratory cycle has been passed, the process proceeds to Step 2.

3.3 Refinement of region size (Step 2)

In Step 2, the size of the tracking window is determined using the amount of the motion obtained from the tracking results and the size of the target object. The size of the tracking window is a larger value of γ_1 times of the maximum value of the amount of motion between adjacent two frames and γ_2 times of the median value of the object size (major axis). The values are decided for width and height of the tracking window separately. The width and the height of the tracking window are multiples of 8, the minimum value of the width and the height is 16 pixels, and the maximum value is the initial tracking window size. When there are multiple tracking targets in a sequence, the sizes of the tracking windows are decided for each target.

3.4 Tracking using refined window size (Step 3)

In the subsequent frames, tracking with KCF and template matching is performed using the tracking window size determined in Step 2. When the target window size is changed in Step 2, $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{x}}$ in KCF and the size of the template used in the template matching should be changed. This is performed using the template used in Step 1, which is an image patch around the target object in the first frame. Specifically, for the template matching, the template is revised by extracting the center area of the initial template. The updated template is

Table 1. Tracking results for the 2D point-tracking training data. The numbers show the tracking errors in pixels.

SequenceName	KCF		Proposed	
	Mean	Maximum	Mean	Maximum
CIL-02 #1	3.00	6.50	2.52	5.81
ETH-05-2 #2	6.51	28.98	3.67	27.06
ICR-04 #2	2.38	8.50	1.77	6.50
MED-05-1 #1	5.93	14.64	5.30	12.71

converted to $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{x}}$ for KCF. Thus, $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{x}}$ is reset at the beginning of Step 3.

The subsequent process, tracking with KCF and the template matching, is the same as in Step 1.

4 Experimental Results

We evaluated the performance of the proposed method using the 2D point-tracking training data of liver ultrasound. The training data was provided by organizers of CLUST 2015, MICCAI Challenge on Liver Ultrasound Tracking.

Our implementation is based on the open source MATLAB code (version 2) at http://home.isr.uc.pt/∼henriques/circulant/. In the experiment, we used the following values for parameters. The feature is gray scale pixel value. Note that we compared the tracking performance with gray scale feature and Histogram Oriented Gradient (HOG) feature [9] as a preliminary experiment and the gray scale feature showed better performance for liver ultrasound sequences, while HOG feature shows much better performance than gray scale feature for surveillance and sport videos in [7]. The kernel type in KCF is Gaussian kernel with $\sigma = 0.2$. We selected a Gaussian kernel based on preliminary experimental results. λ in Eq. (4) is 10^{-4}. β in Eq. (4) is 0.0075, which is one tenth of the default parameter in case the feature vector is the gray scale feature in KCF. The initial tracking window and template size and the threshold C_{th} are 96×96 pixels and 0.8, respectively. The object size is estimated every 5 frames in Step 1. In Step 2, γ_1 and γ_2 in step 2 are 8 and 4, respectively. These values were decided empirically. We used the same parameters for all the sequences.

We compared the tracking performance of the proposed method with KCF in which the tracking window size is fixed at 96×96 pixels. Table 1 shows the results for some the 2D point-tracking training sequences. In Table 1, mean and maximum tracking errors are shown for KCF and the proposed method. Note that the ground truth is not given for all frames for the training sequences and the errors are calculated only for the frames the ground truth is given.

As can be seen in Table 1, the proposed method shows better tracking performance comparing to the original KCF. The proposed method can improve the mean errors by up to nearly 3 pixels and the maximum errors by nearly up to 2 pixels.

Table 2. Tracking results for the 2D point-tracking test data. The numbers show the tracking errors in millimeters.

SequenceName	Mean	Standard deviation	95th percentile	Minimum	Maximum
CIL	2.21	1.82	6.00	0.10	8.05
ETH	0.83	0.77	2.55	0.01	10.94
ICR	0.90	0.64	1.99	0.01	7.80
MED1	1.64	2.27	5.47	0.02	17.29
MED2	1.37	1.19	4.26	0.01	7.64
All	1.09	1.35	3.07	0.01	17.29

The tracking results for each sequence group (CIL, ETH, ICR, MED1 and MED2) in the test dataset are shown in Table 2.

As for computational time, we measured the processing time using a computer with an Intel Core i7 3.3 GHz CPU (6 cores) and 64 GB memory. We implemented the proposed method with MATLAB. The average processing time per target object and per frame in Step 1 and Step 3 for each sequence in the proposed method was from 106 to 155 msec and from 75 to 120 msec, respectively. For comparision, The average processing time per target object and per frame for each sequence in the original KCF was from 23 to 58 msec. The average additional time in the proposed method comparing to the original KCF was 59 msec per target object and per frame. We think the processing time can be improved if we implement the template matching in the proposed method by C++, and it's a future item.

5 Conclusion

In this paper, we proposed a tracking method of target tissues in long ultrasound sequences of liver. The proposed method is based on kernelized correlation filter (KCF) and we introduce two extensions to KCF for improving the tracking accuracy. The experimental results showed the proposed method had better accuracy comparing to the original KCF. Mean tracking error with the proposed method for test sequences of 2D ultrasound was 1.09 mm.

Items for future research are to improve the accuracy of tracking tissues near the boarder and expand the proposed method to 3D ultrasound.

References

1. Abolmaesumi, P., Salcudean, S. E., Zhu, W. H., Sirouspour, M. R., DiMaio, S. P.: Image-Guided Control of a Robot for Medical Ultrasound. IEEE Trans. Robotics and Automation, 18(1), 11–23 (2002)
2. Kopelmana, D., Inbarb, Y., Hanannelc, A., Freundlichc, D., Castelb, D., Pereld, A., Greenfeldd, A., Salamona, T., Sarelie, M., Valeanue, A., Papae, M.: Magnetic resonance-guided focused ultrasound surgery (MRgFUS): Ablation of liver tissue in a porcine model. European Journal of Radiology, 59(2), 157-162 (2006)

3. Bouchet, L. G., Meeks, S. L., Goodchild, G., Bova, F. J., Buatti, J. M., Friedman, W. A.: Calibration of three-dimensional ultrasound images for image-guided radiation therapy. Physics in Medicine and Biology, 46(2), 559 (2001)
4. Avidan, S.: Support vector tracking. IEEE Trans. Pattern Analysis and Machine Intelligence, 26(8), 1064–1072 (2004)
5. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Trans. Pattern Analysis and Machine Intelligence, 34(7), 1409–1422 (2012)
6. Hare, S., Saffari, A., Torr, P. H.: Struck: Structured output tracking with kernels. In IEEE International Conference on Computer Vision (ICCV), 263–270 (2011)
7. Henriques, J. F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Analysis and Machine Intelligence, 37(3), 583–596 (2015)
8. Kristan, M., Matas, J., Leonardis, A., Vojir, T., Pflugfelder, R., Fernandez, G., Nebehay, G., Porikli, F., Cehovin, L.: A novel performance evaluation methodology for single-target trackers. arXiv preprint arXiv:1503.01313 (2015)
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 886–893 (2005)