Liver Ultrasound Tracking Using Long-term and Short-term Template Matching

Satoshi Kondo

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

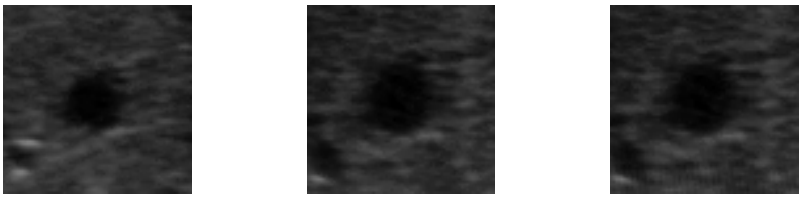
Abstract. We propose a method to track tissues in long ultrasound sequences of liver. The proposed method is based on template matching and uses multiple templates called long-term template and short-term template. A template to track the target tissue is adaptively selected from the long-term template and the short-term template. The tracking performance is assessed on 21 sequences of 2D ultrasound with 54 regions of interests. Mean tracking error is 1.71 mm. We also confirm that tracking can be performed in about 84 msec per frame using a personal computer.

Keywords: Ultrasound, Liver, Tracking, Template matching, Multiple templates

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. Template matching is widely used because it is relatively simple and gives high performance. Template matching is also used for ultrasound video sequences. One of the important items to design template matching is a method for selecting templates. There are two major methods for selecting templates. One is to use a surrounding area of a tissue specified at the first frame as a template and the template is never updated until the end of the sequence [5]. And the other is to use a tracked region at the most recent frame as a template and the template is updated at every frame [6]. The former has a disadvantage that tracking fails when the shape of the tracking target changes. The latter can overcome the problem of the former, it has a disadvantage that small tracking errors are accumulated. Fig. 1 shows examples of a target vessel in ultrasound images. As can be seen in Fig. 1, Fig. 1(b) is better than Fig. 1(a) to be used as a template when template matching is performed for Fig. 1(c). A method to solve these problems has been proposed in [7] and the method adaptively selects



(a) (b) (c)

Fig. 1. Examples of ultrasound images in a sequence. These images show the target vessel in different frames. (a) 1st frame. (b) 37th frame. (c) 38th frame.

a template from the first frame or the most recent frame. Moreover, the idea in [7] is applied to ultrasound liver tracking in [8].

While it is possible to perform template matching with high correlation when the template is selected from the most recent frame, it still has a drift problem if it cannot estimate the motion in high accuracy, i.e. sub-pixel order.

In this paper, we propose a tracking method of tissues in long ultrasound sequences of liver. In the proposed method, we pay attention to the following characteristics that liver ultrasound video sequences have: 1) Each region of interest moves to almost same direction, and 2) Motion of each region has a high periodicity. Under these observation, we adopt the following methods in the proposed method. 1) We estimate a global motion for the whole frame and utilize the estimated global motion for estimating motion of each ROI. 2) To avoid drift, templates obtained at the first frame are used preferentially. 3) To track ROI even when texture and shape of the ROI are changed from the first frame, we select additional templates from past neighborhood frames. 4) Search ranges of template matching changes adaptively depending on the motion of the past frames.

Though the proposed method is the same as the methods proposed in [7] and [8] in terms of using a plurality of templates selected from both the first frame and the most recent frame, we propose a method to select a template from a plurality of recent frames by paying attention to the characteristic that the movement of the liver tissue by breathing is periodic.

2 Proposed Method

2.1 Overview

Fig. 2 is an overview of the procedure of our proposed method and Fig. 3 is a schematic diagram of the proposed method. Our proposed method is based on template matching. We use multiple templates called long-term and short-term templates. We will describe the details about each process in the following sections.

- At the first frame
  - Select global and long-term templates (Step 1)
- At each following frame
  - Estimate global motion (Step 2)
  - For each tracker (ROI)
    - Estimate long-term motion (Step 3)
    - Estimate short-term motion (Step 4)
    - Obtain the final tracking result (Step 5)

Fig. 2. Overview of the procedure of the proposed method.

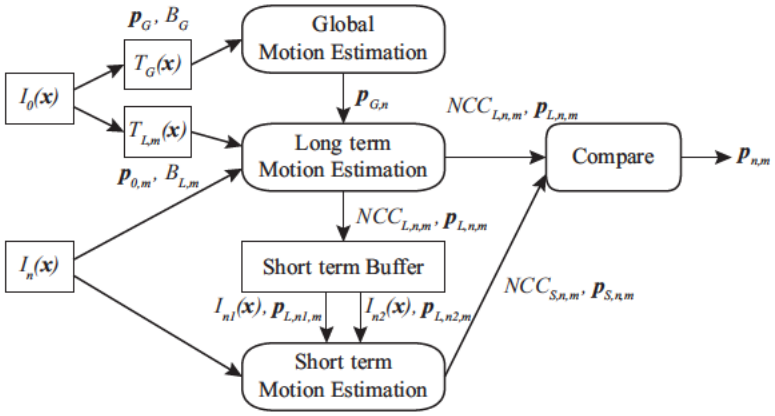


Fig. 3. A schematic diagram of the proposed method.

2.2 Selection of global and long-term templates (Step 1)

Suppose we have a video sequence of ultrasound images $I_n(\mathbf{x})$ where $\mathbf{x} = (x, y)^T$ are the pixel coordinates and $n = 0, 1, 2, \cdots$ is the frame number. We also have m annotations at pixel positions $\mathbf{p}_{0,m}$ which are the positions of tracking targets at the first frame $I_0(\mathbf{x})$, where $\mathbf{p} = (p_x, p_y)^T$.

We select templates of two types at the first frame $I_0(\mathbf{x})$. One is used to determine a motion of the entire frame and it is referred to as a 'global template' $T_G(\mathbf{x})$. The other is a template used in each tracker (ROI) and it is referred to as a 'long-term template' $T_{L,m}(\mathbf{x})$. In addition to a long-term template, each tracker has short-term templates. We will describe the short-term templates later.

A long-term template is obtained as a square area around an annotation point. Since the size of the target tissue, e.g. vessel, is different for each annotation, we decide the size of the long-term template based on variance of the pixel values inside ROI. We first locate a smallest rectangular ROI, which has $B_{Lmin} \times B_{Lmin}$ pixels, around an annotation. The size of the ROI is gradually increased until the maximum size $B_{Lmax} \times B_{Lmax}$ and we calculate a variance of pixel values in the ROI. When we first find a local minimum of the variance,

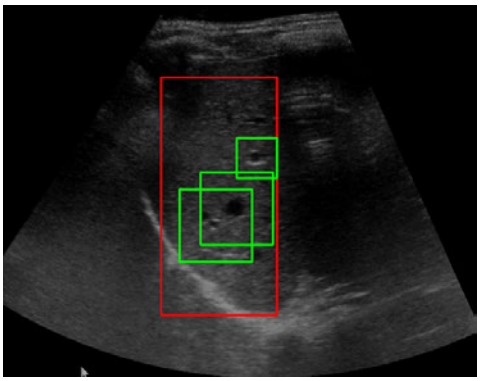


Fig. 4. An example of a set of a global template and long-term templates. The red rectangle shows a global template and the green rectangles show long-term templates.

we stop increasing the size of the ROI and the ROI is the long-term template. We denote a set of pixel coordinates included in the long-term template for m-th annotation as $B_{L,m}$.

A global template is intended to be used to determine a motion of the entire frame. The proposed method extracts a large rectangular area as much as possible in the ultrasound image area by excluding low brightness (shadow) area at the first frame. In the case the long-term template ROIs at the first frame are not included in the global template ROI, we expand the region to include the long-term template ROIs. We denote a set of pixel coordinates included in the global template as B_G.

The global template and the long-term templates are never updated after those are set at the first frame, i.e. all subsequent frames use the same global template $T_G(\mathbf{x})$ and long-term templates $T_{L,m}(\mathbf{x})$, selected at the first frame, for template matching.

Fig. 4 shows an example of a set of a global template and long-term templates. The red rectangle shows a global template and the green rectangles show long-term templates.

2.3 Global motion estimation (Step 2)

Global motion estimation is performed at each of the second and subsequent frames. Template matching is performed using the global template $T_G(\mathbf{x})$. The template matching is evaluated based on normalized cross correlation (NCC). The displacement giving the maximum NCC, which is found by exhausted search, is the tracked position $\mathbf{p}_{G,n}$ by the global motion estimation as Eq. (1),

$$\mathbf{p}_{G,n} = \arg\max_{\mathbf{p}} \frac{\sum_{\mathbf{x}\in B_G} T_G(\mathbf{x}) \cdot I_n(\mathbf{x}+\mathbf{p})}{\sqrt{\sum_{\mathbf{x}\in B_G} T_G(\mathbf{x})^2 \cdot \sum_{\mathbf{x}\in B_G} I_n(\mathbf{x}+\mathbf{p})^2}}. \quad (1)$$

2.4 Long-term motion estimation (Step 3)

Long-term motion estimation is executed for each tracker (ROI). The long-term motion estimation is performed using a long-term template $T_{L,m}$ which is obtained at the first frame as in Eq. (2),

$$\mathbf{p}_{L,n,m} = \arg\max_{\mathbf{p}\in S_L} \frac{\sum_{\mathbf{x}\in B_{L,m}} T_{L,m}(\mathbf{x}) \cdot I_n(\mathbf{x}+\mathbf{p})}{\sqrt{\sum_{\mathbf{x}\in B_{L,m}} T_{L,m}(\mathbf{x})^2 \cdot \sum_{\mathbf{x}\in B_{L,m}} I_n(\mathbf{x}+\mathbf{p})^2}}, \quad (2)$$

where S_L is a set of pixel coordinates in the search area for the long-term motion estimation.

The search range S_L depends on the results of the global motion estimation in Step 2. When the maximum NCC value of the global motion estimation is higher than a threshold $Th_{NCC,G}$, the long-term motion estimation is performed in the vicinity of $\mathbf{p}_{G,n}$. Otherwise, we use a default value of the search range for the motion estimation S_d. The tracked position $\mathbf{p}_{L,n,m}$ and the maximum NCC value $NCC_{L,n,m}$ are stored in the short-term buffer shown in Fig. 3.

2.5 Short-term motion estimation (Step 4)

Short-term motion estimation is performed using short-term templates. Short-term templates are selected from images of the long-term tracked positions until the most recent frame. Here, we use the tracking results with the long-term motion estimation to select short-term templates and the tracking results with the short-term motion estimation are not used at the subsequent frames to avoid drift.

We estimate a cycle of motion from the past tracking results. Fig. 5 shows an example of temporal changes of tracked positions from 300th frame to 500th frame in a ultrasound video sequence. As can be seen in Fig. 5, motions of tissue in liver have periodicity. Two short-term templates for n-th frame are selected from the frames during the recent cycles. Suppose the cycle of the motion is L_c, one short-term template is selected from $(n - L_c)$th frame to $(n - 1)$th frame (period #1 in Fig. 5) and the ROI with the maximum $NCC_{L,n,m}$ is selected as the first short-term template. Another short-term template is selected from $(n - L_c/2 \times 3)$th frame $(n - L_c/2)$th frame (period #2 in Fig. 5) and the ROI with the closest position to the tracked position at $(n - 1)$th frame is selected as the second short-term template.

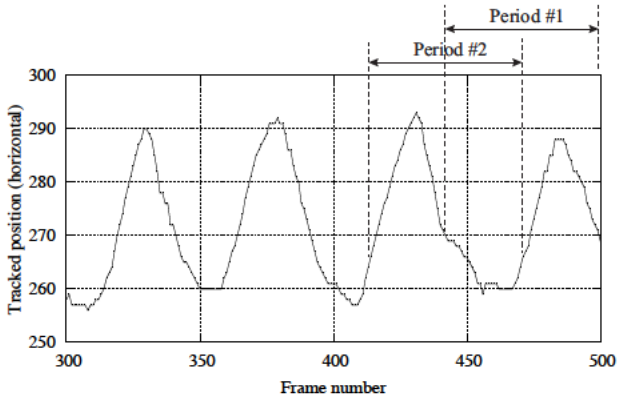


Fig. 5. An example of tracked positions in an ultrasound video sequence.

When we fail to obtain the period of the motion, e.g. in the beginning of the sequence, the short-term motion estimation is performed for a default search range S_d. Otherwise, we apply principal component analysis to the trajectory of the tracking positions. And we perform the motion estimation only for the direction of the eigen vector with the biggest eigen value. The displacement which gives the maximum NCC, $NCC_{S,n,m}$, in the search range is the result of the short-term motion estimation, $\mathbf{p}_{S,n,m}$

2.6 Final tracking result (Step 5)

We compare the maximum NCC value of the long-term motion estimation $NCC_{L,n,m}$ obtained in Step 3 and the maximum NCC value of the short-term motion estimation $NCC_{S,n,m}$ obtained in Step 4. The motion estimation result $\mathbf{p}_{n,m}$ is obtained with Eq. (3),

$$\mathbf{p}_{n,m} = \begin{cases} \mathbf{p}_{L,n,m} \ (NCC_{L,n,m} \geq \alpha \times NCC_{S,n,m}) \\ \mathbf{p}_{S,n,m} \ (\text{otherwise}) \end{cases}, \tag{3}$$

where α is a weight less than 1.0 which prioritizes the long-term motion estimation.

3 Experimental Results

We evaluated the performance of the proposed method using the 2D point-tracking test data. The test data was provided by organizers of CLUST 2014, MICCAI Challenge on Liver Ultrasound Tracking.

In the experiment, we used the following values for parameters: B_{Lmin} = one-tenth of smaller size of image width and height, $B_{Lmax} = 120$ pixels, $Th_{NCC,G} = 0.95$. $S_d = 15$ pixels for both horizontal and vertical directions and $\alpha = 0.95$.

Table 1. 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
ETH	0.79	1.13	1.83	0.00	33.47
MED	2.62	3.84	6.85	0.02	39.46
All	1.71	2.98	4.47	0.00	39.46

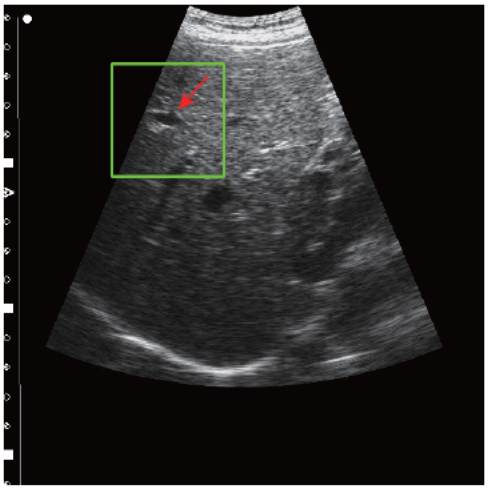


Fig. 6. An example of failure case. The red arrow points to a tracking target and the green box shows the ROI of a long-term template.

The tracking results for each sequence group (ETH and MED) are shown in Table 1. Although we do not show the result for each sequence, the mean errors for all ROIs in ETH sequence group are less than 2 mm and 30 out of 37 ROIs in MED sequence group are less than 3 mm.

In MED sequence group, there are some sequences which have tracking target near the boarder. Fig. 6 shows an example of a tracking target and a long-term template ROI in this case. As shown in Fig. 6, the long-term template includes non-ultrasound image area and it is the main reason to fail the tracking. Also, we confirmed that tracking sometimes failed when the tracking target is small tissue.

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 C++ using OpenCV and OpenMP. Note that the maximum number of threads in OpenMP is the same as the number of the annotations (trackers) for each sequence. The average processing time was about 84 msec/frame. Note that our implementation of the motion estimation utilizes an OpenCV function and it is not optimized for the proposed method yet.

4 Conclusion

In this paper, we proposed a tracking method of target tissues in long ultrasound sequences of liver. The proposed method uses multiple templates, i.e. long-term and short-term templates. The experimental results using 21 sequences of 2D ultrasound showed the proposed method had good accuracy. We also confirmed that tracking can be performed in about 84 msec per frame using a personal computer.

Items for future research are to improve the accuracy of tracking tissues near the boarder and small tissues, improve the processing speed by optimizing the motion estimation processing, 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. Bohs, L. N., Trahey, G. E.: A novel method for angle independent ultrasonic imaging of blood flow and tissue motion. IEEE Trans. on Biomedical Engineering, 38(3), 280–286 (1991)
5. Krupa, A., Fichtinger, G., Hager, G. D. Full motion tracking in ultrasound using image speckle information and visual servoing. In IEEE International Conference on Robotics and Automation, 2458–2464 (2007)
6. Revell, J., Mirmehdi, M., McNally, D: Computer vision elastography: speckle adaptive motion estimation for elastography using ultrasound sequences. IEEE Trans. on Medical Imaging, 24(6), 755–766 (2005)
7. Matthews, I., Ishikawa, T., Baker, S.: The template update problem. IEEE Trans. on Pattern Analysis and Machine Intelligence, 26(6), 810–815 (2004)
8. De Luca, V., Tschannen, M, Szekely, G., Tanner, C.: A Learning-based Approach for Fast and Robust Vessel Tracking in Long Ultrasound Sequences. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, LNCS, 8149, 518–525. Springer Berlin Heidelberg (2013)