Robust Liver Ultrasound Tracking using Dense Distinctive Image Features

Andre Hallack[1], Bartłomiej W. Papież[1], Amalia Cifor[2], Mark J. Gooding[2], Julia A. Schnabel[3]

[1] Institute of Biomedical Engineering (Department of Engineering Science), University of Oxford, UK
[2] Mirada Medical, Oxford, UK
[3] Division of Imaging Sciences & Biomedical Engineering, King's College London, UK

Abstract. A new framework for tracking of anatomically relevant landmarks on 2D liver ultrasound sequences is presented in this work. It combines logDemons nonlinear registration, which estimates the motion within ultrasound sequences, with a moving window tracking method, that propagates the estimated motion around the region of interest to subsequent frames. Robust and accurate nonlinear registration is obtained by employing the dense Scale Invariant Feature Transform as a similarity measure. The proposed method was evaluated on 24 sequences from the CLUST 2015 challenge. On a total of 62 landmarks within these sequences, a mean target error of 0.91mm was achieved, surpassing the previous challenge best performance, 1.44mm on CLUST 2014.

1 Introduction

Ultrasound provides real-time, safe and affordable imaging of soft tissues, making it one of the most popular techniques for tracking of internal body structures. This however comes at the expense of considerable amounts of noise. Hence, the challenges of using this imaging technique lie very often on the image interpretation. For real time tracking of structures, such as vessels and tumours, this translates to the need of robust and equally efficient image processing methods.

In this work, an efficient solution for tracking of anatomical structures in liver ultrasound sequences is presented. These images are subject to high amounts of motion due to breathing, as well as noise and shadowing effects, causing significant intensity changes in ultrasound structure and appearance [1]. Though not persistent over time, high levels of nonlinear deformation are observed in liver ultrasound. Consequently, motion correction approaches for anatomical structure tracking cannot rely only on rigid registration methods, which are much faster than nonlinear methods.

To solve this problem, we employed a tracking framework where only the regions of interest are analysed, which greatly contributes to the system's efficiency. In this framework, nonlinear motion correction was performed using logDemons diffeomorphic registration, which has already been successfully applied to liver ultrasound tracking [2]. The main contribution of this work is the use of the

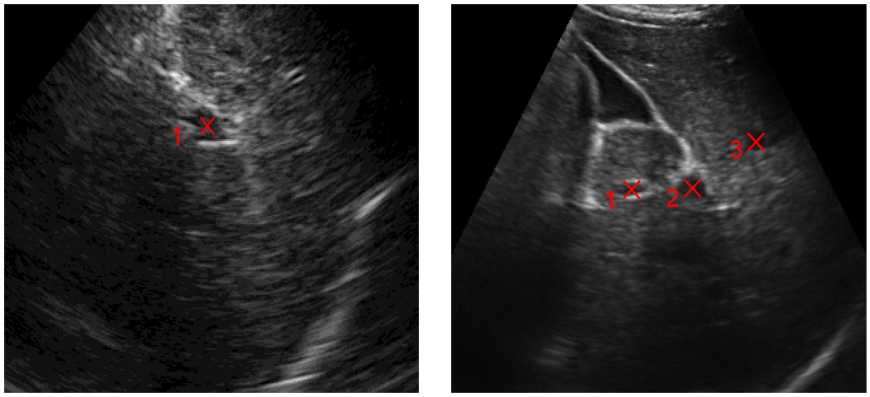

Fig. 1. Sample reference ultrasound images and landmarks from the liver ultrasound tracking challenge CLUST 2015.

dense Scale Invariant Feature Transform (dense SIFT) [3], a similarity metric which has not yet been widely explored for medical imaging registration, but has shown promise in computer vision problems to enhance distinctive features of images and to be very effective on 2D image registration.

This paper is structured as follows: in Sec. 2 the method developed to track annotations by registering liver ultrasound images is presented. This is followed in Sec. 3 with a description of the CLUST 2015[4] data used evaluate the proposed method, as well as the experiments conducted and their respective results. Finally, Sec. 4 concludes this work.

2 Methods

The task addressed in the CLUST 2015 registration challenge can be described as: given a sequence of temporal 2D images I_t, to estimate a set of annotation positions $\boldsymbol{x} = M(t)$ over time based on the initial position $M(1)$ of a relevant structure. Here, $\boldsymbol{x}$ is a 2D spatial location and t the frame index in the sequence. Fig. 1 presents examples of ultrasounds images used in this work and the landmarks being tracked.

In our method, we opted to use an image registration approach. Hence, by computing a nonlinear transformation field T_t that registers I_1 to I_t, $\hat{M}(t)$ can be estimated as $T_t(M(1))$.

One of the difficulties of applying registration methods for such tasks is that over long periods of time, large amounts of complex deformation and displacement will occur in the images, hindering the registration process between far apart acquisitions. On the specific case of liver ultrasound, nonlinear deformations are present for the observed structures, and thus nonlinear registration is necessary to correctly identify the structures over time. However, over short

[4] *CLUST 2015*, http://clust.ethz.ch/

acquisitions, the main persistent type of global motion is rigid. Hence, we opted first for a tracking method which propagates the estimated rigid motion at each time point to the next frame, followed by accurate deformable registration.

The proposed method consists of a tracking framework (similar to [4]), where the images are cropped around the expected annotation location (Sec. 2.1), and a logDemons diffeomorphic nonlinear registration framework (Sec. 2.2), using dense SIFT, which is a highly descriptive image transform (Sec. 2.3).

2.1 Tracking

The main concept of the tracking method used here was to perform image registration between cropped patches of the image sequence. Hence, from each frame I_t, a square region $W_{t-1}(I_t)$ is extracted. W_{t-1} determines the position around where this patch should be extracted. For the first frame, I_1, this is straightforward, since W_0 is the position of the initial annotation $M(1)$. For each subsequent frame I_t ($t > 1$), W_{t-1} is extracted around the previously estimated annotation location $\hat{M}(t-1)$. An overview of this method is shown on Algorithm 1. Each cropped patch has w x w pixels centred on the estimated annotated position. This method does not propagate the whole nonlinear transformation from frame to frame, but only the translation found for the annotation.

Fig. 2 presents an example of how the tracking framework progresses by propagating the previous estimated annotation location to each subsequent frame. This method was based on the work by König *et al.* [4]. Unlike that work, here, for the cases where several different annotated structures are present in the same imaging sequence, each annotation was tracked independently. Another difference to that framework is that no upper motion bounds were applied to the obtained transformations.

Data: Liver ultrasound 2D sequence I_t and the initial landmark position $M(1)$.
Result: Sequence of estimated landmark positions $\hat{M}(t)$ for each ultrasound sequence frame.

$W_0(I_1) \leftarrow crop(I_1, M(1))$
$\hat{M}(1) \leftarrow M(1)$
$t \leftarrow 2$
while $t <$ *number of frames* **do**
  $W_{t-1}(I_t) \leftarrow crop(I_t, \hat{M}(t-1))$
  $T_t \leftarrow register(W_{t-1}(I_t), W_0(I_1))$
  $\hat{M}(t) \leftarrow T_t(M(1))$
  $t \leftarrow t + 1$
end

Algorithm 1: Tracking method for liver ultrasound landmarks.

2.2 Nonlinear Registration

For each image I_t of the sequence, its cropped subregion around the expected annotation position $W_{t-1}(I_t)$ is registered to the reference cropped image $W_0(I_1)$,

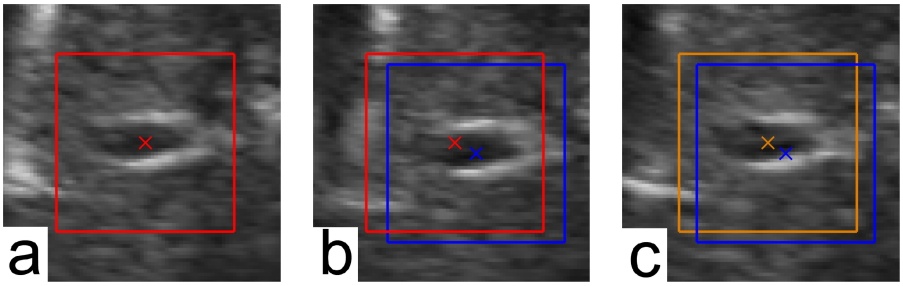

Fig. 2. Sequence of images exemplifying the tracking framework used in this work. On the first (reference) frame (a), a region of interest (red square) is cropped around the known landmark ($M(1)$ - red cross). This location is used as the region of interest for the next frame (b), indicated by the red square. Image registration between these two cropped regions (a and b - red squares) leads to a new estimation of the annotation position (blue cross). For the next frame (c), the latest estimated annotated position is used as the centre of the region of interest (blue square and cross) which will be used to register to the reference region of interest (a), and after motion correction finds a new position for the tracked structure (orange cross). This process is repeated for all frames of an ultrasound sequence.

estimating the nonlinear transformation T_t. This transformation can then be used to estimate the annotation position for the current frame, $\hat{M}(t) = T_t(M(0))$. For this nonlinear registration step we employed a diffeomorphic logDemons framework [5]. Demons is an iterative registration method which finds a non-regularised deformation field T_c by minimizing a similarity measure (Sim), but at the same time computes a smooth regularised version (T_s) by applying a Gaussian filter to it [6]. This is described by the following optimization problem:

$$T = \arg\min_{T_s}\Big(\mathrm{Sim}\left(W_0(I_1), T_c(W_{t-1}(I_t)\right) + \mathrm{Dist}(T_s, T_c) + \mathrm{Reg}(T_s)\Big) \qquad (1)$$

where usually $\mathrm{Dist}(T_s, T_c) = ||T_s - T_c||^2$ and $\mathrm{Reg}(T_s) = ||T_s||^2$.

The logDemons version of this method ensures that the obtained transformation is invertible by restricting T_s to a subspace of diffeomorphisms (see details in [5]).

2.3 Dense SIFT

Due to the intensities distortions mentioned in Sec. 1, most intensity-based measures of similarity lack the robustness to accurately register such ultrasound image [2]. Hence, in this work, image intensities were not directly compared during the logDemons registration steps. At each registration iteration, these images were transformed using the dense Scale Invariant Feature Transform (dense SIFT), a modified version of SIFT for dense image analysis [3, 7, 8]. This method computes at each voxel a descriptor vector based on the histogram of

gradients around its neighbourhood, generating a vector-valued images. This enhances distinctive characteristics of the ultrasound images and therefore allows for more accurate registration.

Transformed images using dense SIFT (SIFT($I(\boldsymbol{x})$) can then be locally compared as the sum-of-square-differences (SSD) of the dense SIFT feature vectors at each voxel $\boldsymbol{x}$:

$$\mathrm{Sim}_{\mathrm{SIFT}}(I_1(\boldsymbol{x}), I_t(\boldsymbol{x})) = \left|\left|\mathrm{SIFT}(I_1(\boldsymbol{x})) - \mathrm{SIFT}(T(I_t(\boldsymbol{x})))\right|\right|_2^2 \tag{2}$$

This similarity measure is used within the logDemons framework (Eq. 1).

2.4 Parameters

The cropped region around each annotation was of 51 by 51 (w x w) pixels, the size of this region was chosen to safely contain the whole annotated structure. For nonlinear registration, the logDemons framework was applied with three resolution levels with 20 iterations at each level and transformation field smoothing $\sigma_{\mathrm{diff}} = 2$ pixels, these parameters were not optimised for this problem. The SIFT Flow library was used for dense SIFT with the standard parameters: cell size = 2 and 8 bins [3].

3 Experiment, Results and Discussion

The proposed framework was evaluated on 24 2D+t liver ultrasound sequences (from 4 different scanners) designated for the CLUST 2015 challenge. These sequences showed a spatial resolution between 0.30mm and 0.55mm, number of frames ranging from 895 to 5586 and image rate from 11Hz to 23Hz. A total of 62 landmarks were provided (between 1 and 4 per sequence) at the initial frame to be tracked over the whole sequence. The method was assessed by the challenge organizers in terms of the mean target error (MTE) and standard deviation (σ) between the computed and ground truth landmarks at selected frames, as well as its 95th percentile. For each of these sequences, the detailed results are presented in Tab. 1 and the overall outcome for each different scanner is shown in Tab. 2.

The MTE over all the sequences and landmarks was 0.91mm with standard deviation of 1.66mm. These results were below the ones obtained on the previous ultrasound liver tracking competition, CLUST 2014, where the best reported result was 1.44mm MTE and 2.04mm stadard deviation on a similar dataset [9].

Despite using the same parameters for all analysis, the results obtained over different scanners did not vary much; for the best case (dataset ETH) the MTE was 0.59mm and for the worse case (CIL) it was 1.74mm. A noteworthy result was for sequence MED 0.6-1, where for one of the annotations (2) this framework clearly lost track of the landmark and showed very large errors. This highlights one of the possible flaws of the proposed tracking framework: by reducing the analysis to cropped regions around the expected annotation location we are prone to failures if the motion between frames is close to the size of the region

of interest (this effect is further illustrated in Fig. 3). Increasing the size of the cropped regions could help reducing this error, however, it will lead to longer computation time.

This algorithm was implemented on a single thread C++ program and tested on a Intel i7-3770 computer with 3.40GHz, Ubuntu Linux 12.04 operating system. The average frame processing speed was of 4.8 images per second, which is close, but short of the acquisition rate of these sequences. Since the analysis for each of the landmarks in a sequence is done independently, this speed is directly proportional to the number of annotations, and the experiments showed a rate of 11.8 frames per second per annotation.

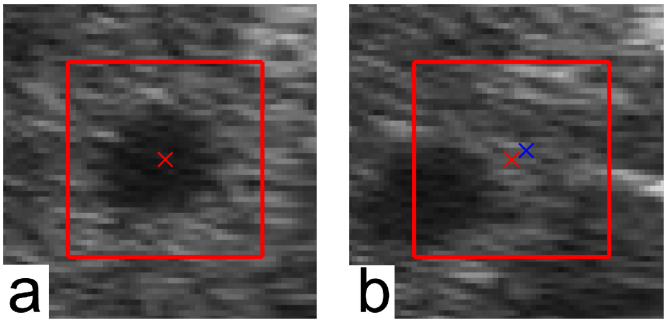

Fig. 3. Illustration of an specific sequence (MED-06-1 - landmark 2) where the proposed method loses track of the structure of interest. On (a) the estimated landmark position (red cross) is centred on the structure of interest. On the subsequent frame (b) there is a large displacement of the structure of interest, since now the estimated landmark position from the previous frame (red cross) and the cropped region of interest (red square) do not contain the whole the structure. Here, the nonlinear registration method fails to follow the observed displacement, and finds an erroneous new estimated landmark position (blue cross).

4 Conclusions

In this work a new method for liver 2D ultrasound tracking was proposed and evaluated within the CLUST 2015 challenge. Our method combined a nonlinear registration method with a tracking method which focused only on the region of interest around the tracked annotation, which then propagated the translational information from previous frames to the next frame. The nonlinear image registration between the cropped reference and moving regions of interest (at each frame) around the observed structure were performed using logDemons. One of the main advantages of the proposed solution is the use of dense SIFT as a similarity measure, a feature transform which led to better characterization of the observed structures than standard intensity based measures.

Table 1. CLUST 2015 2D tracking results of the 62 sequences. The outcome is measured in terms of the mean target error (MTE) and standard deviation (σ) in millimetres for each landmark in each sequence.

	Results per landmark (mm)							
Dataset	MTE_1	σ_1	MTE_2	σ_2	MTE_3	σ_3	MTE_4	σ_4
CIL 03	1.51	1.53	2.92	1.04				
CIL 04	1.35	0.50	0.95	0.47				
ETH 06-1	0.68	0.29						
ETH 06-2	0.73	0.36						
ETH 07-1	0.34	0.16	0.47	0.29				
ETH 07-2	0.43	0.23	1.08	0.46				
ETH 08-1	0.45	0.18	0.62	0.38				
ETH 08-2	0.62	0.21	0.90	0.44				
ETH 09-1	0.59	0.59	0.47	0.52	0.64	0.93	0.54	0.56
ETH 09-2	0.45	0.24	0.86	0.46	0.91	0.43		
ETH 10-1	0.40	0.23	0.47	0.25	0.41	0.18		
ETH 10-2	0.41	0.29	0.56	0.29	0.57	0.57		
ICR 05	0.66	0.22	1.02	0.54				
ICR 06	0.74	0.25	1.29	0.54				
ICR 07	0.66	0.27	0.84	0.63				
ICR 08	0.58	0.48	0.37	0.22	1.03	1.66		
MED 06-1	1.93	0.89	7.73	12.26	1.15	0.50	0.98	1.33
MED 06-2	1.64	0.92	0.73	0.51	1.59	1.05		
MED 07-1	0.98	0.48	0.86	0.31	1.25	0.70		
MED 07-2	1.00	0.48	0.65	0.43	0.79	0.30		
MED 07-3	3.06	1.92	0.86	0.44	0.79	0.42		
MED 07-4	2.85	1.67	0.60	0.38	0.83	0.39	0.36	0.18
MED 08-1	0.70	0.41	1.63	0.60	0.68	0.36		
MED 08-2	0.95	0.37	1.07	0.47	2.57	1.32		

The highly promising results under the CLUST 2015 challenge attested the validity of the proposed method, showing a performance at least comparable to state-of-the-art solutions and close to real-time speed. We also were able to identify the main conditions where our framework fails at this tracking problem, inciting the development of solutions which can handle well large deformations between adjacent frames. Further development can also be made in terms of the optimisation of the method's parameters both to improve its accuracy as well as to reach real-time processing of ultrasound sequences (including mult-thread computing).

5 Acknowledgments

We would like to acknowledge funding from the CRUK/EPSRC Cancer Imaging Centre at Oxford.

AH also gratefully acknowledges the support of the Research Council UK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doc-

Table 2. CLUST 2015 2D tracking results for each ultrasound scanner, as well as the final overall result. The outcome is given by the mean target error (MTE), standard deviation (σ) and minimum (Min) and maximum (Max) error in millimetres for each scanner group.

	Combined results (mm)					
Dataset	MTE	σ	95th	Min	Max	Scanner Type
CIL	1.74	1.27	4.23	0.04	6.58	Ultrasonix MDP
ETH	0.59	0.44	1.29	0.01	12.45	Siemens Antares
ICR	0.80	0.73	1.81	0.01	7.78	Elekta Clarity - Ultrasonix
MED 1	1.53	3.25	3.99	0.01	35.20	DiPhAs Fraunhofer
MED 2	1.27	0.95	3.48	0.02	6.48	Zonare z.one
Overall	**0.91**	**1.66**	**2.20**	**0.01**	**35.20**	

toral Training in Healthcare Innovation) and the funding provided by CAPES Foundation, process N^o BEX 0725/12-9.

References

1. J. A. Noble, N. Navab, and H. Becher, "Ultrasonic image analysis and image-guided interventions," *Interface focus*, vol. 1, no. 4, pp. 673–685, 2011.
2. A. Cifor, L. Risser, D. Chung, E. M. Anderson, J. Schnabel, *et al.*, "Hybrid feature-based diffeomorphic registration for tumor tracking in 2-d liver ultrasound images," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 9, pp. 1647–1656, 2013.
3. C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE T Pattern Anal*, vol. 33[5], pp. 978–94, 2011.
4. L. König, T. Kipshagen, and J. Rühaak, "A non-linear image registration scheme for real-time liver ultrasound tracking using normalized gradient fields," *Challenge on Liver Ultrasound Tracking CLUST 2014*, p. 29, 2014.
5. T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Symmetric log-domain diffeomorphic Registration: a demons-based approach." *MICCAI*, vol. 11[1], pp. 754–61, 2008.
6. J. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Med Image Anal*, vol. 2[3], pp. 243–60, 1998.
7. A. Hallack, B. W. Papież, J. Wilson, L. M. Wang, T. Maughan, M. J. Gooding, and J. A. Schnabel, "Correlating tumour histology and *ex vivo* mri using dense modality-independent patch-based descriptors," *Submitted Work*, 2015.
8. D. G. Lowe, "Object recognition from local scale-invariant features," in *ICCV*, vol. 2. IEEE, 1999, pp. 1150–7.
9. V. De Luca, T. Benz, S. Kondo, L. König, D. Lübke, S. Rothlübbers, O. Somphone, S. Allaire, M. a. Lediju Bell, D. Y. F. Chung, a. Cifor, C. Grozea, M. Günther, J. Jenne, T. Kipshagen, M. Kowarschik, N. Navab, J. Rühaak, J. Schwaab, and C. Tanner, "The 2014 liver ultrasound tracking benchmark." *Physics in medicine and biology*, vol. 60, no. 14, pp. 5571–99, July 2015.