Bibliographical reference

Is Speckle Tracking Feasible for Ultrasound Tongue Images?, Kele Xu, Pierre Roussel and Bruce Denby, Acta Acustica **vol. 103** (Number 3), 2017, pp. 365-368

DOI

https://doi.org/10.3813/AAA.919065

Acta Acustica united with Acustica
Vol. 103 (2017) 365 – 368

DOI 10.3813/AAA.919065

Is Speckle Tracking Feasible for Ultrasound Tongue Images?

Kele Xu[1,2], Pierre Roussel[1], Bruce Denby[3]
[1] ESPCI Paris, PSL Research University, CNRS, Institut Langevin,
1 Rue Jussieu, Paris, 75005 France. France. pierre.roussel@espci.fr
[2] Pierre and Marie Curie Univ, 4 Place Jussieu, Paris, 75005 kelele.xu@gmail.com
[3] Tianjin Univ, Tianjin, 300000 China. bruce.denby@upmc.fr

Summary
Ultrasound is widely used in speech production research. In this paper, we explore the feasibility of speckle tracking in ultrasound tongue image sequences to obtain tissue point correspondence. The speckle tracking approaches explored include: deformation registration; optical flow; and local invariant feature based methods. Moreover, an image similarity-based automatic tracking re-initialization method is incorporated into the speckle tracking to improve robustness. Results demonstrate that deformation registration provides the best performance, and that for short-duration sequences, speckle tracking is feasible and can provide point correspondence with millimeter accuracy, which is useful for practical applications in speech production research such as 3D tongue dynamic modeling and strain analysis of the tongue tissue.

1. Introduction

In speech production research, substantial effort has been devoted to the interpretation of articulator configurations [1, 2]. As the tongue is one of the major components of the vocal tract, accurate tongue motion tracking is of great importance in understanding the production of speech. B-mode ultrasound imaging is widely used to visualize the motion of the tongue. Non-invasive and easy to implement experimentally, it enables tracking algorithms to follow the motion of the tongue via the highly visible contour of its upper surface. Such contours, however, do not allow individual "tissue point" landmarks to be followed over a sequence of tongue images, thus limiting the utility of ultrasound in speech production research.

Recently, speckle tracking has drawn attention in ultrasound assessment of cardiac deformation, as the technique appears to provide additional information beyond what contours can offer [3]. By tracking unique localized speckle patterns, or "virtual" tissue points, in an ultrasound image, motion information on the underlying tissue can be deduced. At present, few attempts have been made to explore whether speckle tracking is feasible in ultrasound tongue images. In this paper, we aim to answer this question by making a quantitative comparison between different speckle tracking methods to explore the feasibility of speckle tracking in ultrasound tongue images. Several speckle-tracking approaches were tested, which can be divided into three main types of methods: deformation registration; optical flow; and local invariant feature based methods. Virtual tissue points obtained on the tongue surface can then be used to evaluate the performance of the different speckle tracking algorithms. Here, we use the mean sum of distances (MSD) metric to make a quantitative comparison between tracked tissue point contours (linking several virtual tissue points) and contours hand-labeled by a domain expert. Moreover, an image similarity-based tracking re-initialization method is also incorporated into the speckle tracking algorithms to improve the robustness of the tracking. The organization of the paper is as follows. In section 2, a brief summary is given of the different speckle tracking implementation methods. In section 3, the performance comparison of different speckle tracking approaches is conducted on the ultrasound tongue image sequences. Section 4 provides conclusion and discusses future work.

Received 30 December 2016,
accepted 1 April 2017,
published online 13 April 2017.

2. Speckle tracking approaches

In this section, we explored three kinds of speckle tracking methods, including deformation registration, optical flow and Scale Invariant Feature Transform (SIFT) flow (the local invariant feature based method).

2.1. Deformation registration

Non-rigid deformation registration methods are often applied to estimate the correspondences between landmarks in adjacent frames. In this method, the correspondence between the tissue points is obtained by minimizing a similarity measure. The local similarity measure is usually based on L^1 or L^2 norms comparisons on the intensities of image blocks, such norms are appropriate for the image sequences, which are characterized by Gaussian statistics. In [4], the author proposed to use sum of square difference (SSD) to measure the difference between the deformed previous frame and the next frame. However, the ultrasound image is contaminated by the multiplicative speckle noise, and the traditional approaches may not be suitable in this case as these methods assume that the imaging system is contaminated by the Gaussian noise. Commonly, researchers proposed to replace the common assumptions of a Gaussian distribution by ultrasound specific noise models, and a Rayleigh distribution is assumed. In [5], the author proposed a novel similarity measure (denoted by CD2) using the maximum likelihood method based on the assumption that two consecutive ultrasound images are both corrupted by multiplicative Rayleigh noise and the probability density functions of the noise are independent of each other. In this work, both CD2 and SSD are tested during the comparison. An example was given in Figure 1 to show the result of deformation registration-based speckle tracking methods. The original tissue points are manually selected in the previous frame. Before the deformation registration, the locations of the tissue points are shown in the next frame. After the registration, the tracked markers are shown in the deformed previous frame. As can be seen from the figure, all of the tracked markers lie near the tongue surface, which demonstrates the feasibility of the deformation registration-based speckle tracking.

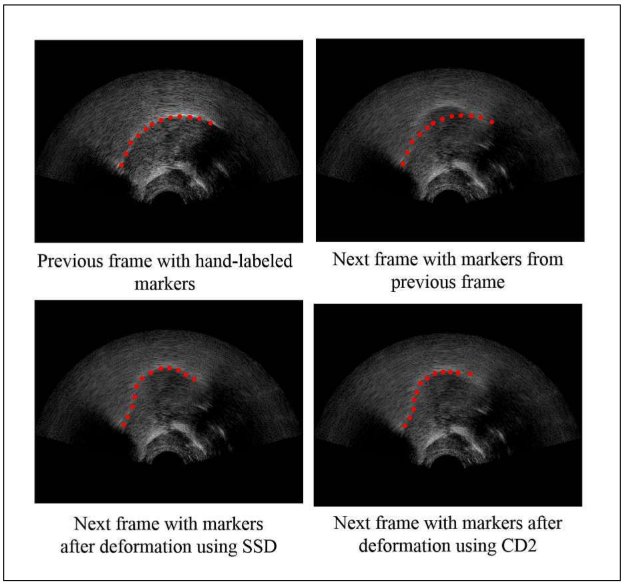

Figure 1. (Colour online) Deformation registration-based virtual land markers tracking in ultrasound tongue images.

2.2. Optical flow

Optical flow is another widely-used motion tracking method in the image sequences. Here, we use the Lucas-Kanade method [6] for the tracking task. Lucas-Kanade optical flow method sub-divides the original image into smaller sections and assumes that velocity in each section is a constant. Then, this method performs a weighted least-square fit of the constraint equation. By solving the optical flow constraint equation, the displacement along vertical and horizontal directions can be calculated (as shown in Figure 2c), thus, the speckle can be tracked.

2.3. Local invariant feature

Since last decade, local invariant feature has drawn many attentions due to its superior performance in computer vision. Despite the numerous publications of studies on the image feature descriptors, the descriptors representing the distribution of small-scale features within the interest point neighborhood, which was introduced in [7], have shown robust performance to obtain point correspondences. However, only sparse feature correspondences can be obtained using standard SIFT (as shown in Figure 3). Here, we use SIFT flow [8] to build dense correspondence between the pixels in adjacent frames based on the SIFT descriptors. The SIFT descriptor is extracted at each pixel in the image, describing local ultrasound image and encoding the pattern information. Similarly to the optical flow approach, the motion vector is calculated by searching the correspondences of the descriptors in the two frames (as shown in Figure 2d).

3. Comparison between different speckle tracking methods in ultrasound tongue images

In ultrasound tongue images, due to the lack of strict tissue point correspondence, there is no ground truth to evaluate the performance of different speckle tracking methods, so a specific manual evaluation method is needed. As the researchers are more interested in the upper surface of tongue, an experiment has been carried out to compare the tracked speckles on the contour (consisting in speckles on the contour in the ultrasound tongue image) with hand-labeled curves. The idea behind this kind of evaluation is that: different speckle tracking method can be used to follow several tissue points on the tongue contour and the virtual contour can be compared to with the hand-labeled curves. If the distances between the tracked contour and the labeled contour is small, the speckle tracking method is of better performance to follow the motion of the tongue surface.

For validation, we follow the widely-used distance measurement, mean sum of distances (MSD) defined as

$$\mathrm{MSD}(\mathbf{U},\mathbf{V}) = \frac{1}{2n}\left(\sum_{i=1}^{n} min|v_i - u_j| + \sum_{j=1}^{n} min|u_i - v_j|\right),$$

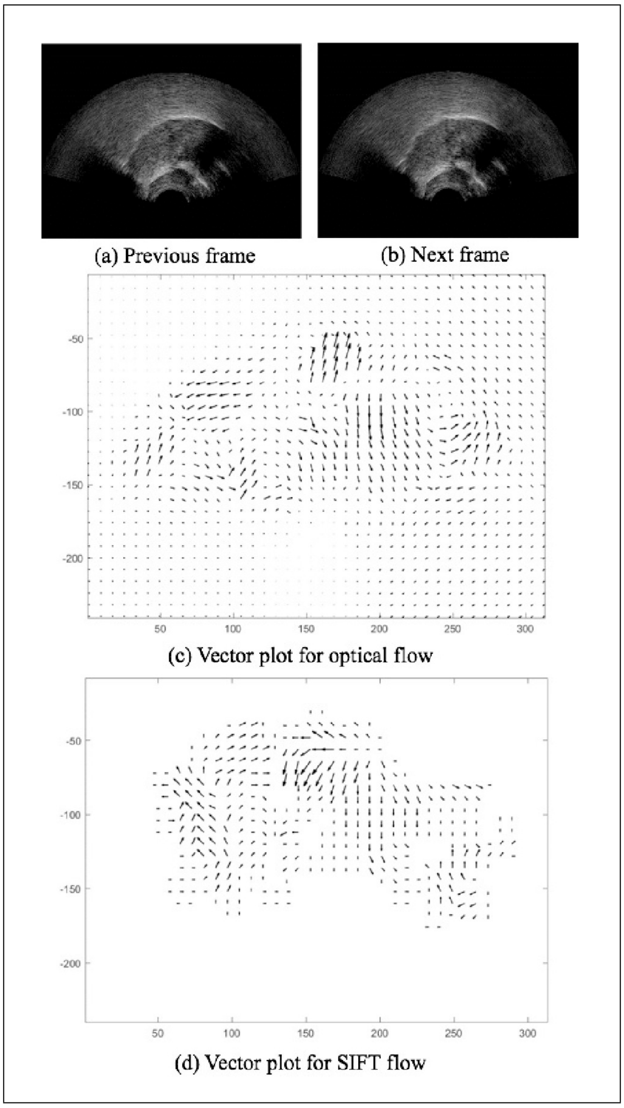

Figure 2. Optical flow and SIFT flow-based speckle tracking.

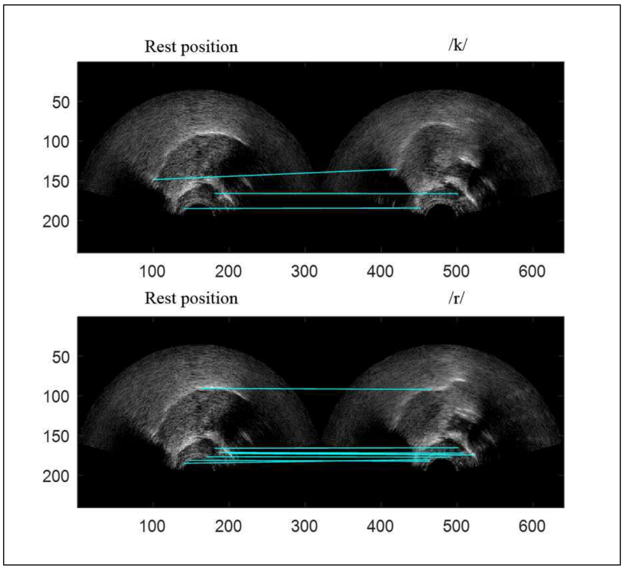

Figure 3. (Colour online) SIFT-based speckle tracking in ultrasound tongue images.

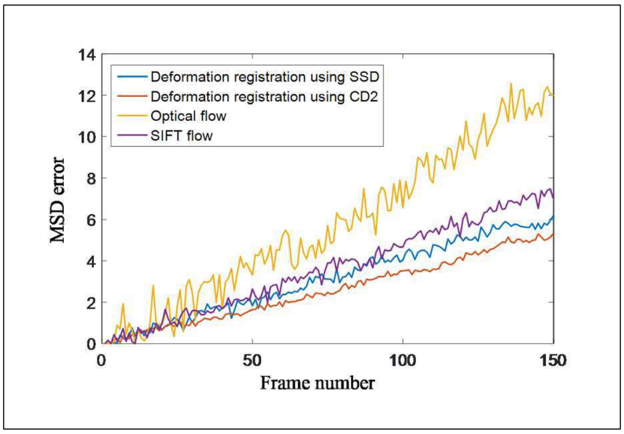

Figure 4. (Colour online) The MSD error (in pixel, 1 pixel = 0.295 mm) of different speckle tracking methods.

where **U** is the contour extracted automatically, which consists of the tracked tissue points. **V** is the result of hand-labeled contour and i and j are the index of the control points on the contour and n is the number of contraint points. The comparison is conducted on a continuous sequence, which contain 150 frames. The data recorded comes from a normal female subject, and the contour in all of the frames are hand-labeled. The MSD error for different approaches on the continuous sequences is given in the Figure 4. The virtual landmarks on the tongue contour are reset to those that were input for the hand-labeled frames (i.e., MSD = 0).

As can be seen from Figure 4, CD2-based local similarity measurement gives the best result in our experiment. Based on the experiment, we can see that optical flow has larger errors. Moreover, the deformation of the tongue is large and the motion of the tongue is quite fast, which poses even a greater challenge to follow the motion of the tongue by using the optical flow as standard optical flow cannot handle with large deformation. Moreover, it is worthwhile to notice that the tracking error was accumulated during the tracking processing.

If hand-refinement is done for the speckle tracking method, the performance can be further improved. Here, we followed the image similarity-based automatic tracking re-initialization method proposed in [9]. The main idea of automatic re-initialization is the following: before speckle tracking algorithms was applied to the ultrasound tongue images, the image similarity index is calculated, between current frame and the hand-labeled frames. If this coefficient exceeds a set threshold (here we set it as 0.85), the virtual landmarks on the tongue contour are reset to those that were input for the hand-labels frames. This provides a method to prevent accumulation of errors over long sequences, which can lead to erroneous tracking, and amounts to a sort of "automatic re-initialization" of the tracking points. In this paper, we used the complex-wavelet structural similarity (CW-SSIM) [10] to measure the similarity between two frames.

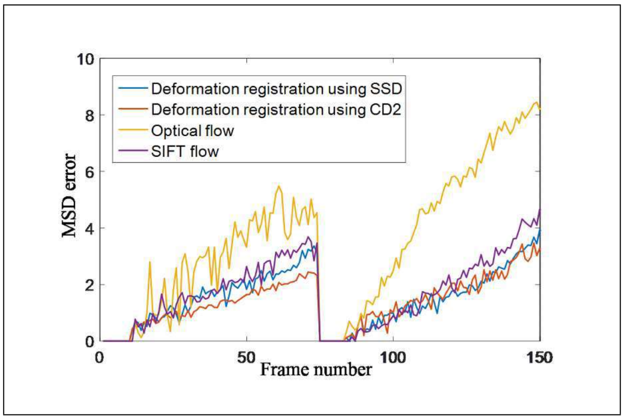

Figure 5. (Colour online) The MSD error (in pixel, 1 pixel = 0.295 mm) of different speckle tracking methods with automatic tracking re-initialization.

Following is the evaluation procedure; the MSD will be used to measure the performance of different speckle tracking methods with proposed re-initialization. To make the comparison more effective, the parameter and data will be kept same for all different methods. The results are given in Figure 5. As can be seen from the figure, the re-initialization method can dramatically reduce the accumulated error, thus improve the performance of the speckle tracking, which demonstrates the feasibility of proposed automatic re-initialization method. Moreover, except for the optical flow method, the maximum MSD error is about 1.2 mm in our experiment by using speckle tracking with automatic re-initialization for other algorithms.

4. Conclusions

The results demonstrated that speckle tracking can provide tissue point correspondence information with millimeter level accuracy. This kind of correspondence information is vital for several applications. Tracking error will accumulate during the tracking processing, but the CW-SSIM based automatic re-initialization method was tested on the speckle tracking methods and tracking performance was improved. Our work still retains certain limitations. Firstly, the performance comparison is conducted only by measuring the MSD distance on the tongue surface; a more robust and feasible quantitative comparison method is needed for the future study. Secondly, the performance of the speckle tracking methods still needs to be evaluated quantitatively by making use of other imaging modalities such as MRI and EMA.

References

[1] M. Stone: A guide to analyzing tongue motion from ultrasound images. Clin Linguist Phonet **19** (2005) 455–502.

[2] J. Stoffers, C. Neuschaefer-Rube, M. Kob: Comparison of vocal tract resonance characteristics using LPC and impedance measurements. Acta Acust united Ac. **92** (2006) 320–333.

[3] B. H. Amundsen, T. H. Valle, T. Edvardsen, H. Torp, J. Crosby, E. Lyseggen, A. Støylen: Noninvasive myocardial strain measurement by speckle tracking echocardiography: validation against sonomicrometry and tagged magnetic resonance imaging. J Am Coll Cardiol **47** (2006) 789-793.

[4] M. Jacob, H. L. LeHouillier, S. Bora, S. McAleavey, D. Dalecki, J. McDonough: Speckle tracking for the recovery of displacement and velocity information from sequences of ultrasound images of the tongue. 8th International Seminar on Speech Production, Strasbourg, France, 2008.

[5] B. Cohen, I. H. Dinstein: New maximum likelihood motion estimation schemes for noisy ultrasound images. Pattern Recogn **35** (2002) 455–463.

[6] B. D. Lucas, T. Kanade: An iterative image registration technique with an application to stereo vision. International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 1981.

[7] D. G. Lowe: Distinctive image features from scale-invariant keypoints. Int J Comput Vision. **60** (2004) 91–110.

[8] C. Liu, J. Yuen, A. Torralba: SIFT Flow: dense correspondence across scenes and its applications, IEEE T Pattern Anal **5** (2011) 978–994.

[9] K. Xu, T. G. Csapó, P. Roussel, B. Denby: A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization. J. Acoust Soc Am vol **139** (2016) EL154–EL160.

[10] M.P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, M. K. Markey: Complex wavelet structural similarity: A new image similarity index. IEEE T Image Process **11** (2009) 2385–2401.