Seminar paper summary

Group 7: Tian Xie

Section I. Introduction

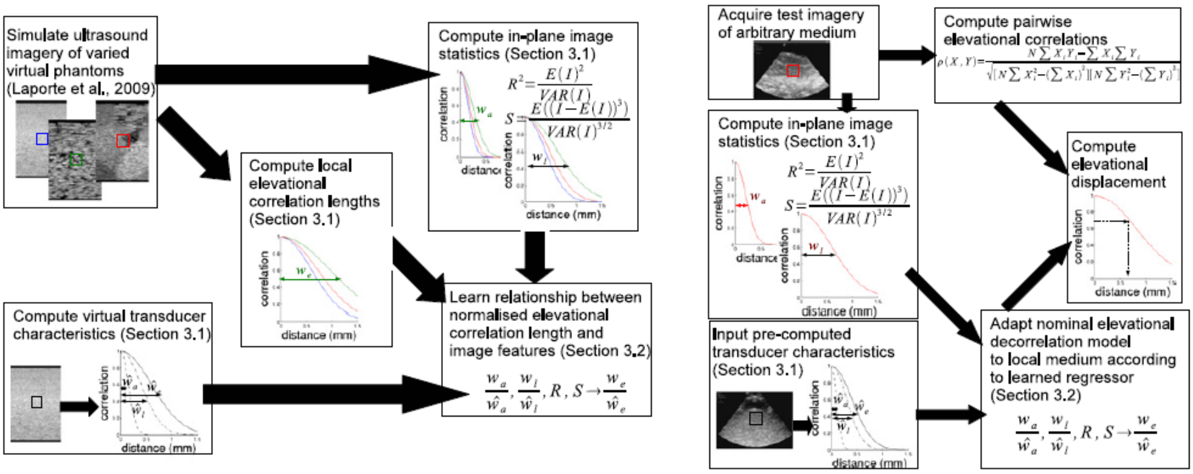
In my project, one essential problem is to estimate the transformation between two neighboring ultrasound images. In-plane motion can be estimated by conventional 2D image registration while it is especially difficult to estimate the out-of-plane transformation for ultrasound images. Luckily, the speckle noise of the images provides useful information to circumvent this problem. Therefore, here I would like to summarize the contents of two papers in which learning-based methods were used to estimate out-of-plane motion.

The first paper is *"learning to estimate out-of-plane motion in ultrasound imagery of real tissue"* (Arbel et al., 2010) and the second paper is *"deep learning for sensorless 3D freehand ultrasound images"* (Wein et al., 2017). The two papers all refer to speckle decorrelation. And both are estimating the out-of-plane motion. Section II will summarize and assess the contents of the former paper. Section III will deal with the latter.

Section II. *"Learning to Estimate Out-of-Plane Motion in Ultrasound Imagery of Real Tissue"*

The correlation of patches decreases as function of distance, which is called speckle decorrelation. Under the condition of Rayleigh scattering, which means that the medium contains dense and randomly-spread micro-structures, the speckles are considered fully developed speckles (FDS). The decorrelation curve of FDS is approximately a Gaussian function whose characteristics determined merely by transducer properties. However, in real tissues, it is unlikely to have Rayleigh scatterers. The correlation decreases not only more slowly but also differently with different medium. Therefore, the goal of this paper is to accurately estimate the elevational translation between two ultrasound images for different real tissues. The paper makes use of first and second order statistics, plus in-plane motion information, of the ultrasound images to learn a regression to fit the "stretching" effect of the Gaussian curve in real tissues. It successfully developes a locally adaptive, tissue-invariant model by using machine learning. The performance of their model is comparable to the state-of-art heuristic adaptive model (Gee et al., 2006).

This paper is of importance to my project because first it suggests that there are useful features which can be learned by the neural network. Second, the features used in this paper may also help solve our problem. In stead of blindly refer to end-to-end learning, now I can have a more solid understanding of the math behind the scenario. Third, the data acquisition procedure of my project is very similar to theirs. Therefore, knowing some details of their experiments helps me better plan my own data acquisition procedure.



(a) Training (b) Motion estimation

Figure 1 Workflow of the paper (retrieve from the paper)

The workflow of the paper can be divided into two parts. The first part is about training, and the second part is motion estimation. For the training part, sparse Gaussian process is used to learn the relationship between normalized elevational correlation length and image features. In motion estimation step, the idea is the same as registration. The transformation between two images is fitted by using least-of-median-squares to all the patches. The authors also give great details of their experiment including data acquisition procedures. Finally, they compare their method with four baseline approaches to prove the advantage in generalization of their model.

Their model based on the following math assumptions: 1) the speckle decorrelation curve is a Gaussian function; 2) for different tissues, the standard deviation of the Gaussian curve is scaled. Then based on these assumptions and some previous work, the authors extract four features from the ultrasound imagery as learning input:

1. The squared signal-to-noise ratio, $R^2 = \frac{E\{I\}^2}{VAR\{I\}}$, where I is the echo intensity signal;
2. The skewness, $S = \frac{E\{I-E\{I\}^3\}}{VAR\{I\}^{3/2}}$;
3. Normalized Gaussian approximator along axial axis, $r_a = \frac{w_a}{\widehat{w_a}}$;
4. Normalized Gaussian approximator along lateral axis, $r_l = \frac{w_l}{\widehat{w_l}}$.

In feature 3 and 4, w denotes the standard deviation of the Gaussian curve. The hat denotes the variables obtained from the reference phantom. One brilliant idea here is that the author takes the ratio of the Gaussian approximators. By doing so, now the features become transducer invariant. Also, all the calculations are carried out patch-wise, so the inputs are also image-location invariant. Then sparse Gaussian process (Snelson et al.,2005) is used to learn a regression for $r_e = \frac{w_e}{\widehat{w_e}}$. And finally, a globally $r_e{}^*$ is estimated by taking into consideration the two successive images:

$$r_e^* = \frac{\frac{\bar{r}_{e_1}}{\xi_1} + \frac{\bar{r}_{e_2}}{\xi_2}}{\frac{1}{\xi_1} + \frac{1}{\xi_2}}.$$

The stretched curve is now $\rho_q(r_e^*\delta) = \hat{\rho}_q(\delta)$.

The next step is to estimate the motion. After the first step, now for the corresponding patches in two consecutive images, a transformation can be estimated because both the patch location and the elevational motion are known. Then least-median-of-squares is used to find out the registration. They used this method because it is robust to outliers.

$$\phi_{\text{LMedS}} = \underset{\phi}{\text{argmin}} \ \underset{q}{\text{med}} \, \|\mathscr{R}\left(\tilde{\mathbf{x}}_0^{[q]}, \phi\right) - \tilde{\mathbf{x}}_i^{[q]}\|^2.$$

Here x^q denotes the (x,y,z) of each patch. Φ represents the six-dimension of transformation.

The paper collects data from Field II ultrasound simulator. By adjusting the density and regularity of the simulator, images of different phantoms can be simulated. And the training is based on these data. They also collecte data for pork, chicken, turkey and beef, and use them to test the capability to generalize of their locally adaptive model. Images are collected with a resolution of 0.01 – 0.03mm. Each image is divided into patches accordingly. The details of the experiment are not explained here. Their procedure is stringent.

The paper compares their method with three baseline approaches: 1) nominal decorrelation model without adaptation; 2) Prager et al.'s speckle detector; 3) Gee et al.' s heuristic adaptive method. Their results with synthetic phantoms and real tissues both show that their method performs better than the first two baseline methods. And it is also comparable to the state-of-art heuristic approach in generalization.

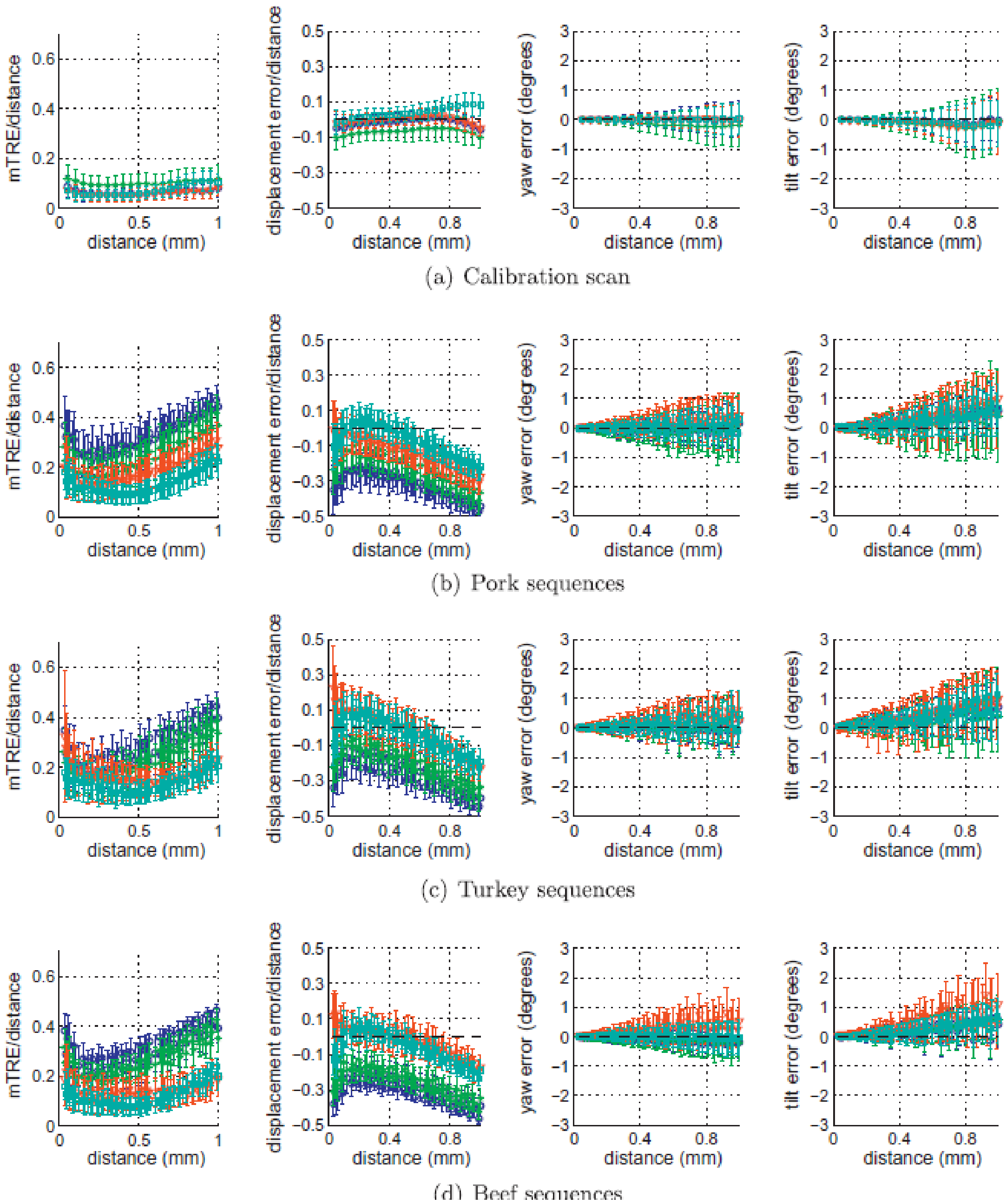


Figure 2 Result for different tissues. Cyan: the model proposed by the paper. (retrieve from the paper)

I think this paper is of high quality. First, it builds all the learning on a solid math model. Second, the normalized features allow them to generate a locally adaptive model, which is of great importance to the generalization of speckle decorrelation method. Finally, the paper is very stringent about all the experimental details. However, there are several drawbacks of this paper. First, the assumption that the Gaussian curve is only scaled might be too simple. Second, the figures they used to present their results look very chaotic. It is not a good idea to put all the information in a single chart. It is even worse when readers are reading the paper in a black and white version.

Section III. ***Deep Learning for Sensorless 3D Freehand Ultrasound Imaging***

This paper is trying to tackle a similar problem as the previous one. However, it estimates six degree of freedom at the same time. The objectives of this paper are to realize sensorless freehand 3D ultrasound imaging. The approach proposed is to circumvent the problem by end-to-end deep learning. Although the paper sees convolutional neural network an analogy of speckle decorrelation, the results given by them cannot prove that the network is trained on speckle characteristics. Instead, their network shows strong reliance on anatomical features. The quality of this paper may not be that satisfying, however the structure of their network gives me a starting point to build a network that will be usable in my project.

The paper first states the similarity of CNN and speckle decorrelation method in the following aspects. First, the local cross-correlation can be approximated by convolutional filters. Second, the patch-wise calculation of decorrelation approach is similar to pooling layers in CNN. Finally, the selection of reliable FDS features and regions can be done by activation layers. The analogy made by the authors might not be true but still, the analogy makes it reasonable to use deep learning to study speckles.

Then the paper introduces the architecture of their network. They experimented on two structures. One is a standard CNN. The other is the standard CNN with two additional channels containing vector field components, which are got from preprocessing with optical flow. The authors hypothesis is that because optical flow applies additional in-plane motion information, the network can better focus on elevational motions.

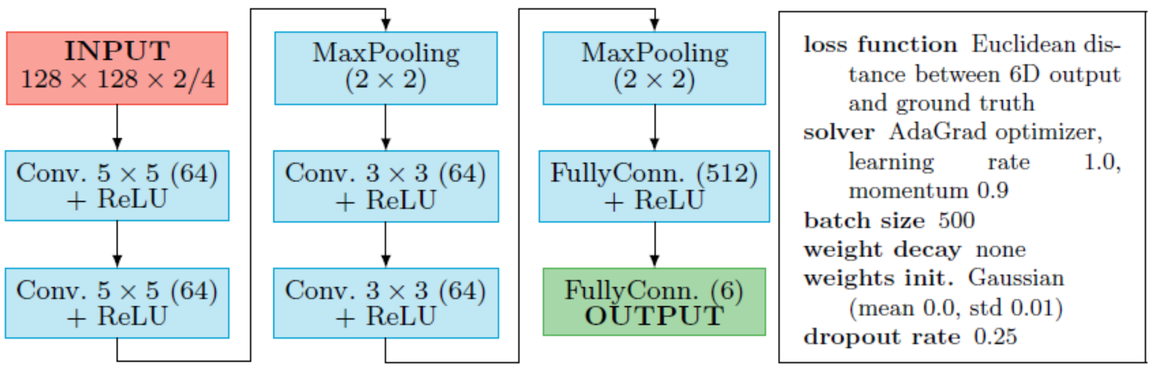


Figure 3 The architecture of the network (retrieve from the paper)

The dataset the authors used are a set of 20 US sweeps on Bluephantom biopsy phantom, a set of 88 in-vivo sweeps on volunteers' forearms and a set of 12 in-vivo sweeps on lower legs. All the sweeps were captured by a Cicada-64 research ultrasound machine. And a linear 128-element probe was used with the depth set to 5cm (focus at 2 cm). The images were later resampled by approximately every 0.3mm. They attached optical target on the probe, and obtained the ground truth value of transformation with absolute position accuracy around 0.2 mm. The authors made sure that for each dataset, there are two separate folds of images for training and testing respectively. However, other details of the experiment are not provided in the paper. For example, what is the maximum distance the network is trying to estimate is not mentioned in their paper. However, this actually affects the accuracy of the result. The authors should give more detailed explanations so that the reader can assess whether they are using the neural network in a proper way.

The paper presents their result by comparison with two baseline method. One is linear motion approach, which simply assumes that the probe is moving at a constant speed. The second is the conventional speckle decorrelation method. In the following tables, they draw the conclusions:

1. CNN with optical flow has the best performance. The use of optical flow helps the CNN to focus on the elevational estimation;
2. In Table 2, when speckles are filtered out, the elevational estimation becomes worse. Therefore, the speckles should help with this direction in some way;
3. In Table 3, the results with white background color are directly predicted with the network trained on forearms. The performance is poor compared with the network trained on lower legs (shown with blue background color). It proves that the network cannot be generalized.

Table 1 phantom dataset	avg. absolute error (mm/°)						final drift (mm)		
	t_x	t_y	t_z	θ_x	θ_y	θ_z	min	med.	max
linear motion	2.27	8.71	38.72	2.37	2.71	0.97	2.29	70.30	149.19
speckle decorrelation	4.96	2.21	29.89	2.10	4.46	1.93	12.67	47.27	134.93
standard CNN	2.25	5.67	14.37	2.13	1.86	0.98	14.31	26.17	65.10
CNN with optical flow	1.32	2.13	7.79	2.32	1.21	0.90	1.70	18.30	36.90

Table 2 forearms dataset	avg. absolute error (mm/°)						final drift (mm)		
	t_x	t_y	t_z	θ_x	θ_y	θ_z	min	med.	max
linear motion	4.46	6.11	24.84	3.51	2.59	2.37	10.11	46.23	129.93
speckle decorrelation	4.36	4.09	18.78	2.53	3.02	5.23	9.19	36.36	98.95
standard CNN	6.30	5.97	6.15	2.82	2.78	2.40	3.72	25.16	63.26
CNN with optical flow	3.54	3.05	4.19	2.63	2.52	1.93	3.35	14.44	41.93
after speckle filtering	3.57	3.59	8.56	2.56	2.64	2.01	5.14	22.04	44.15

Table 3 lower legs dataset	avg. absolute error (mm/°)						final drift (mm)		
	t_x	t_y	t_z	θ_x	θ_y	θ_z	min	med.	max
linear motion	4.49	4.84	39.81	4.39	2.18	2.46	37.35	73.40	143.42
speckle decorrelation	5.02	2.87	30.89	1.82	1.78	4.11	43.21	54.74	89.97
standard CNN	5.34	5.62	17.22	2.58	2.45	2.84	21.73	43.21	65.68
CNN with optical flow	4.14	3.91	17.12	1.94	2.58	2.15	25.79	40.56	52.72
CNN trained on legs	3.11	5.86	5.63	2.75	3.17	5.24	8.53	19.69	30.11

Figure 4 Result (retrieve from the paper)

From my point of view, there are several severe drawbacks in their result analysis. First, given the errors of t_z they reported, which is greater than 4mm for all the cases, it is already out of the range (usually no more than 1 or 2mm) in which speckle decorrelation is still valid. Hence, their suggestion that the neural network may make use of speckles become suspicious. And the way they used speckle decorrelation is not so clear either. Given the range they are trying to estimate, it is meaningless to compare the CNN result with speckle decorrelation. Second, linear motion is too coarse an estimation, probably any methods can easily compete with this baseline. Finally, even though this paper sees CNN an analogy to speckle decorrelation, from their result the CNN is definitely trained on anatomical structures instead. In all, I believe the authors should compare their model with some other approaches to justify its effectiveness, and they should reorganize their dataset to really make use of speckles.

Although this paper includes many flaws, it does shed light on the implementation of deep learning in this field to circumvent some tough problems. And it still worth starting with their architecture to explore the capability of CNN in motion estimation for ultrasound images.

References

Gee, A.H., Housden, R.J., Hassenpflug, P., Treece, G.M., Prager, R.W., 2006. Sensorless freehand 3D ultrasound in real tissue: speckle decorrelation without fully developed speckle. Med. Image Anal. 10, 137–149.

Laporte, C., & Arbel, T. (2011). Learning to estimate out-of-plane motion in ultrasound imagery of real tissue. Medical Image Analysis 15, 202-213. doi://doi.org/10.1016/j.media.2010.08.006

Prager, R.W., Gee, A.H., Treece, G.M., Berman, L.H., 2002. Analysis of speckle in ultrasound images using fractional order statistics and the homodyned k distribution. Ultrasonics 40, 133–137.

Salehi, R. M., Sprung, J., Bauer, R. & Wein, W. (2017). Deep Learning for Sensorless 3D Freehand Ultrasound Imaging. In: Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017. MICCAI 2017.

Snelson, E., Ghahramani, Z., 2006. Sparse Gaussian processes using pseudo-inputs. In: Proceedings NIPS 2005, pp. 1257–1264.