2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

ConvLSTM based Estimation Method of Incision Trajectory with Electric Knife by Connecting Restored Thermal Sources

Yuta Mizunuma[1], Itaru Kitahara[2], and Yoshihiro Kuroda[3]

***Abstract*— Surgical navigation for understanding the internal structure of an organ is being actively studied, and it is necessary to estimate the incision trajectory to update the structure information dynamically. In this study, we focused on the fact that the region incised by the electric knife becomes high in temperature. Thus, we propose an estimation method of incision trajectory by restoring thermal source from diffused thermal images using a ConvLSTM and connecting the restored thermal sources. We first verified the possibility of thermal source restoration, and confirmed that the method enabled to restore the thermal source with high PSNR equivalent to 42.61. Next, we verified the accuracy of the incision trajectory from proposed method by comparing with the traditional method. The results suggested a better performance compared with the traditional method.**

I. INTRODUCTION

Image-guided surgery, which allows for surgeons to see-through a hidden internal structure of organs that cannot be seen in nature[1], [2], has been intensively studied for the past quarter of a century. For adapting a dynamic change of the surgical scene, it is necessary to acquire the exact location of the incision and update the 3D model based on mechanical simulation, e.g. finite element method. A conventional approach uses feature points on the organ surface to estimate the incision. However, it is difficult to estimate correctly if the feature points cannot be extracted properly.

In principle, an electric knife rapidly raises tissue temperature, evaporates water inside the cell, and finally destroys the tissue. Thus, the incision area has high temperature just after incision. Therefore, it is possible to estimate the incision trajectory by observing high temperature region of tissue from thermal images. However, the heat spreads around so that the thermal diffusion increases estimation error. Therefore, it is important to correctly recognize incision area based on thermal source restoration. Note that simple tracking of the tip of the electric knife erroneously detects the knife tip in no contact with tissue as incision area.

*This work was supported by JSPS KAKENHI Grant Numbers JP21H03474, JP21K19778.

[1]Yuta Mizunuma is with Degree Programs in Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan mizunuma@le.iit.tsukuba.ac.jp

[2]Itaru Kitahara is with Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan

[3]Yoshihiro Kuroda is with the Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan

In this study, we propose an estimation method of incision trajectory by restoring thermal source from diffused thermal images using a ConvLSTM and connecting the restored thermal sources. In the experiment, we first verify the possibility of thermal source restoration from thermal image sequence. Next, we verify the accuracy of the incision trajectory from proposed method by comparing with the traditional method while the images that include an electric knife are excluded for basic verification. The main contributions of this paper are as follows:

- To propose an estimation method of incision trajectory from diffused thermal images.
- To verify the possibility of thermal source restoration from the diffused thermal images using ConvLSTM.

II. RELATED WORKS

For acquiring changes in organ shape, a conventional approach constructs a 3D organ model including internal structures from MRI or X-ray CT images in advance of surgery. It detects and tracks feature points using RGB images and updates the 3D model during surgery using a biomechanical model[1]. Paulus et al. proposed a method to estimate the cut indirectly from the changed distance of the surface points on the organ for adapting a topological change of anatomical structures[3]. Another possible approach is tracking the tip of the electric knife from image sequence. However, it is important to avoid detecting the tip of the electric knife that has no contact with the tissue. Also, it is important to correctly detect the incised region based on the time-series changes in thermal diffusion.

RNN (Recurrent Neural Network)[4] is a type of supervised machine learning method that learns features from time-series changes. ConvLSTM (Convolutional Long-Short Term Memory)[5], which is an extension of RNN and has a function of selecting and storing information in image sequences, gives good results in predicting images changes and tracking moving objects[6]. HybridNet is a method to estimate time-series changes by combining machine learning and physical models, and can estimate thermal diffusion. In such methods, it is learned to predict the information beyond the past information in time-series. On the other hand, in this study, the state before thermal diffusion is restored from the state after thermal diffusion by learning that goes back in the opposite direction to the time-series.

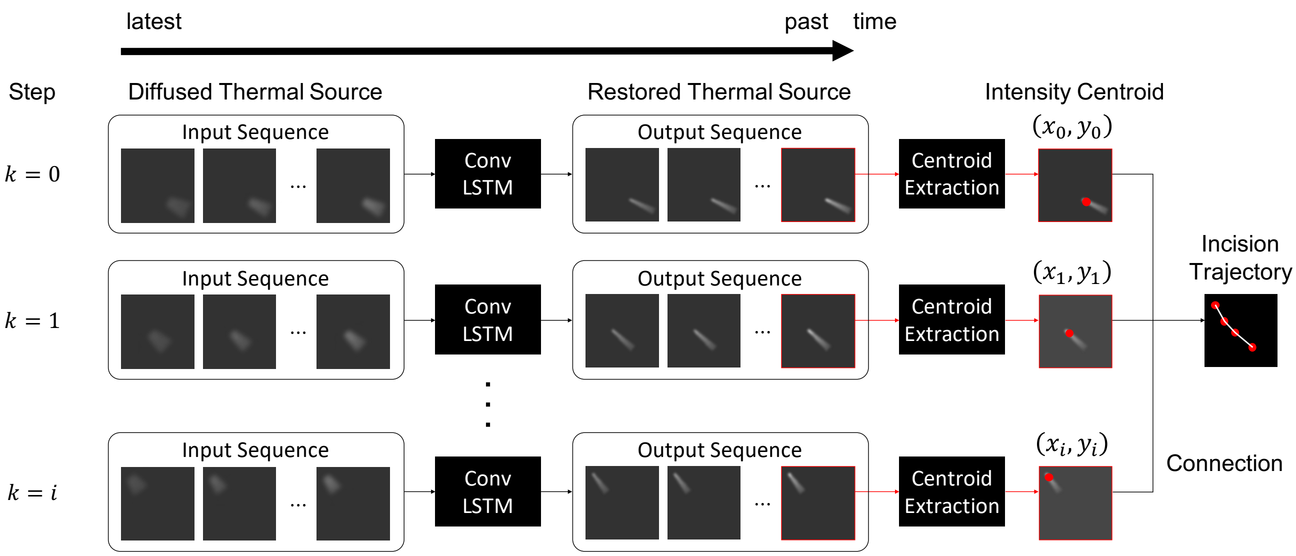

Fig. 1. Outline of the proposed method

III. METHOD

In this paper, we propose an estimation method of incision trajectory by connecting the intensity centroid of the restored thermal sources from diffused thermal images using ConvLSTM. Figure 1 shows the outline of the proposed method. Since restored thermal source can be obtained by tracing back thermal diffusion, the thermal images in the reverse order of time from the latest to the past are used as the input to the ConvLSTM. Then, the image sequence prior to input sequence, namely before thermal diffusion, is estimated. Note that we rely on the latest diffused thermal sources, not on the past images because we need to avoid detecting the tip of electric knife that has no contact with the tissue. Then, the coordinates of the intensity centroid on the final frame of the output sequence are extracted as:

$$(\tilde{x}_k, \tilde{y}_k) = \left(\left[\frac{\sum_{h=1}^{n} \sum_{w=1}^{m} I(w,h) w}{\sum_{h=1}^{n} \sum_{w=1}^{m} I(w,h)} \right], \left[\frac{\sum_{h=1}^{n} \sum_{w=1}^{m} I(w,h) h}{\sum_{h=1}^{n} \sum_{w=1}^{m} I(w,h)} \right] \right) \tag{1}$$

where $\tilde{x}_k$ and $\tilde{y}_k$ are the x, y coordinates of the intensity centroid at step k, respectively. $I(w,h)$ is the intensity at the pixel $(w,\ h)$, and n and m are the height and width of the image, respectively. The calculated intensity centroids are connected to form the incision trajectory. Equation (2) represents MSE (Mean Squared Error), which calculates difference of the intensities between the ConvLSTM output and Ground Truth, at step k.

$$\mathrm{MSE}_k = \frac{1}{nm} \sum_{h=1}^{n} \sum_{w=1}^{m} (I_{\mathrm{GT}}(w,h) - \tilde{I}_k(w,h))^2 \tag{2}$$

where $I_{\mathrm{GT}}(w,h)$ is the intensity at (w,h) of Ground Truth, and $\tilde{I}_k(w,h)$ is the intensity at (w,h) ConvLSTM output at step k. MSE is used both for the error function in machine learning and an evaluation metrics in the experiments. In this study, in order to estimate the past from the latest, the image sequence prior to input sequence, namely before thermal diffusion, is used as the Ground Truth.

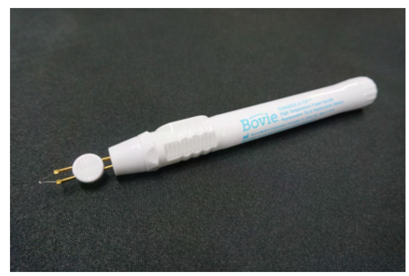

Fig. 2. Electric knife used to prepare the dataset

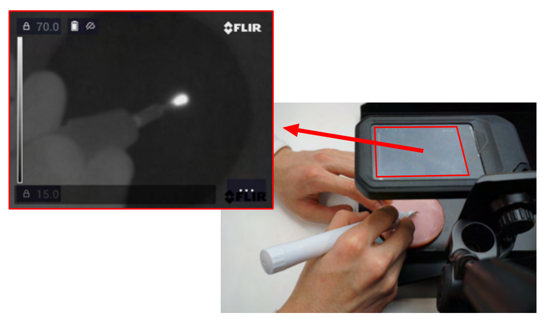

Fig. 3. Preparing dataset

IV. EXPERIMENT 1

We first verified the possibility of thermal source restoration from the diffused thermal images using ConvLSTM. The evaluation metrics for restoration of the thermal image were MSE and PSNR (Peak signal-to-noise ratio), which is the ratio between the square of the

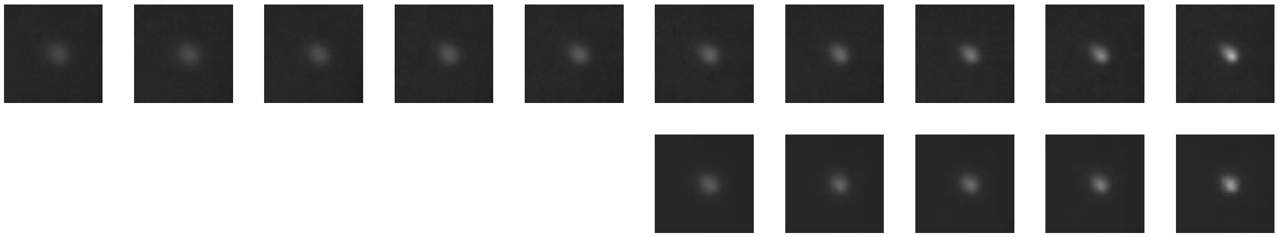

Fig. 4. An example of thermal source restoration: (Upper) left 5 frames are input to ConvLSTM, right 5 frames are Ground Truth, (Lower) ConvLSTM output

maximum intensity in the sequence, namely MAX_k, and MSE_k as:

$$\mathrm{PSNR}_k = 10\log_{10}\frac{{\mathrm{MAX}_k}^2}{\mathrm{MSE}_k} \tag{3}$$

In addition, as a metrics for the intensity centroid, Euclidean distance, $\mathrm{distance}_k$, between the coordinates of intensity centroid from ConvLSTM output and Ground Truth was calculated. Note that $\mathrm{distance}_k$ was calculated for the final frame of ConvLSTM output and Ground Truth images.

A. Dataset and Experimental Conditions

We recorded scenes when just touching a porcine tissue (commercially available processed ham) with an electric knife (DEL1, Bovie Medical Corp.) without moving the electric knife as shown in Fig. 2 and 3. After the tip of the electric knife was touched with tissue for about 2 seconds and separated, a moving image was recorded for about 30 seconds with a thermal camera (FLIR C5). Then, 0.5 fps thermal image sequence was captured from moving image, and the images were normalized to [0,255] in the range of 15 °C to 70 °C. Ten frames were used for learning with cropped in the center, and resized to 64×64, that is, both values of n, m in (1) and (2) were 64. The latest 5 frames of the sequence were used as input, and the past 5 frames were used as Ground Truth. We prepared 31 sequences for training data and 16 sequences for validation data.

When inputting to ConvLSTM, intensities in the image were normalized to [0, 1] with a floating-point number, and the images were sorted in the reverse order of time from the latest to the past. Since a large amount of data is required for learning, data augmentation was performed by randomly flipping left and right. For the number of epochs, we adopted an early-stopping criteria, which stops learning when the MSE for the validation data does not improve for 20 consecutive epochs. For optimization, we adopted Adam[8] with a learning rate of 1.0×10^{-4} and reduced the learning rate by 0.5 times if the accuracy does not improve for 4 consecutive epochs. The batch size was set to 1. Thermal image tends to have a low contrast between the thermal source and the surroundings due to thermal diffusion, resulting that the calculated intensity centroid gets closer to the image center. Thus, only pixels with a temperature of 34.25 °C or higher were used for calculating the intensity centroid.

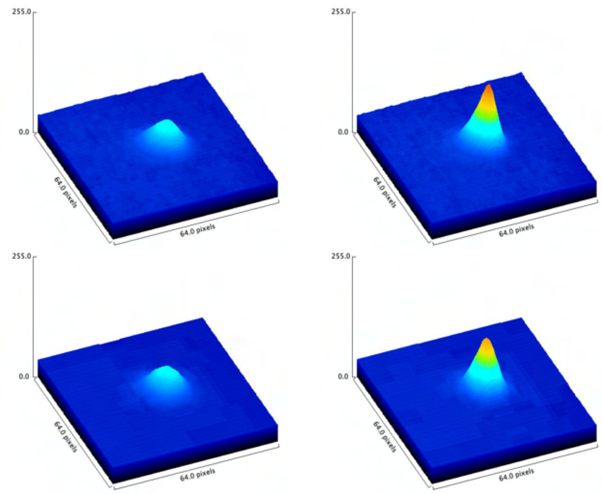

Fig. 5. (Upper) Ground Truth, (Lower) ConvLSTM output

This temperature threshold was empirically determined. The implemented system consists of PC with Intel Core i7-9750H CPU, NVIDIA GeForce RTX 2070, 16 GB main memory, and PyTorch 1.6.0 framework.

B. Results and Discussion

Table I shows the mean and standard deviation of MSE_k, PSNR_k, and $\mathrm{distance}_k$ in the last epoch. The mean MSE was less than 1.0×10^{-4}. The mean PSNR 42.61 confirmed that the estimation error was small to the signal intensity. The mean distance showed that the intensity centroid of the ConvLSTM output was close to the centroid in Ground Truth within about a pixel. Also, mean distance was calculated directly from the input without restoring the thermal source. In particular, a value such as 0.48 ± 0.60 and 2.17 ± 1.68 were obtained in case of threshold value 28.75 °C and 23.25 °C, respectively.

TABLE I

MSE, PSNR, and distance in Experiment 1

Metrics	Value
MSE [a.u.]	$2.81\times10^{-5} \pm 9.52\times10^{-6}$
PSNR [a.u.]	42.61 ± 1.37
distance [px]	0.79 ± 0.49

An example of the ConvLSTM output is shown in

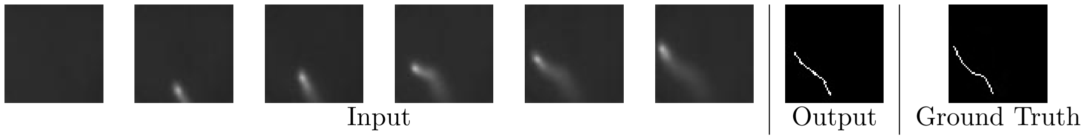

Fig. 6. Example of input and output images to estimate incision trajectory. Left six images for the input, middle one image for the output, and right one image for the Ground Truth

Fig. 4. The upper row is the dataset, where the left 5 frames were input images to ConvLSTM and the right 5 frames were Ground Truth. The lower row is the ConvLSTM output. Figure 5 shows the three-dimensional plot of the temperature distribution, especially in the 6th and 10th frames in the upper row of Figure 5, i.e. Ground Truth, and the 1st and 5th frames in the lower row, i.e. ConvLSTM output. The results suggested that the distributed heat was concentrated in the ConvLSTM output, similar to Ground Truth. The calculation time was less than 10 minutes to train the network.

V. EXPERIMENT 2

We verified the accuracy of the incision trajectory from proposed method by comparing with a traditional method that connects the intensity centroid that was obtained from the raw image without thermal source restoration. The evaluation metrics for accuracy of the incision trajectory was extended DTW (Dynamic Time Warping)[9] to 2D, which calculates a sum of the minimum distances of the points between two data.

A. Dataset and Experimental Conditions

The path of incision trajectory was pre-determined, and the material whose heat conductivity is much different from the tissue phantom was temporarily placed and captured by thermal camera. Then, based on the captured image, Ground Truth of the incision trajectory was manually created. Note that there may be some variability due to manual labeling. We captured the scenes by a thermal camera when stroking the tissue phantom with an electric knife. More specifically, we stroked the tissue and kept away a knife for a while. We repeated this procedure a couple of times for making an image sequence. The images were cropped and resized as well as Experiment 1. The images containing the electric knife were excluded for basic verification to verify the performance without noise derived from the existence of a tool. 20 and 11 sequences were prepared for training and validation data, respectively, with 76 frames at the minimum length and 207 frames at the maximum length.

Since the dataset needs to be loaded with a fixed number of frames during training, the data loader loaded the fixed length sequence by shifting the original sequence by one frame. The number of loaded sequences were $f_{\text{seq}} - (f_{\text{len}} - 1)$, where f_{seq} is the number of original frames, and f_{len} is the number of frames to be set as a fixed length, in this study, it was set to 10 frames. k in Figure 1 represents the number of shifts to obtain a fixed length sequence from the original long sequence. For example, when $k = 0$, the 1st to 10th images from the original sequence are used, when $k = 1$, the 2nd to 11th images are used, and when $k = 3$, the 3rd to 12th images are used. As a result of loaded all sequences from original dataset, the training and validation data were 2487 and 997 sequences, respectively. The number of epochs, optimization method, batch size, and temperature threshold were the same as Experiment 1. Also, The execution environment for performing calculations was the same as in Experiment 1.

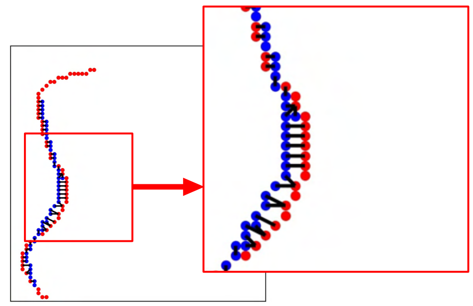

Fig. 7. An example of extended DTW

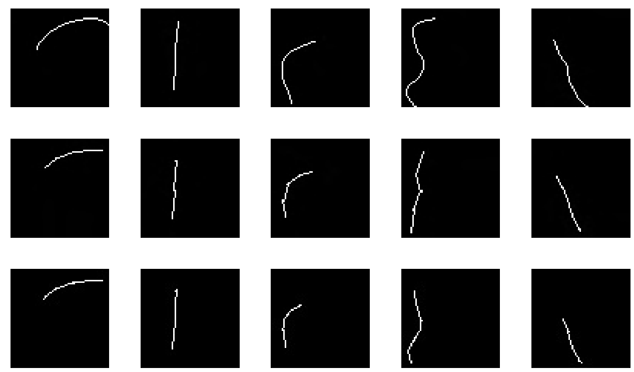

Fig. 8. Output example: (Upper) Ground Truth, (Middle) Traditional method, (Lower) Proposed method

B. Results and Discussion

Figure 6 shows an example of the input and output by the proposed method. Figure 7 shows an example of extended DTW, where blue points are predicted points by proposed method, red points are points of Ground Truth, and black lines are the minimum distances from each bluepoint to red point. In this paper, extended DTW was defined as the value of extended DTW divided by the number of blue points because each trajectory has a different number of data points. Table II shows the mean and standard deviation of extended DTW for each of the traditional and proposed methods. The proposed method suggested a better performance compared with the traditional method.

TABLE II
EXTENDED DTW IN EXPERIMENT 2

Metrics	Traditional method	Proposed method
Ext. DTW [px]	1.23 ± 0.67	0.96 ± 0.54

Figure 8 shows a couple of output examples. The upper row is the Ground Truth, middle low is the incision trajectory estimated by the traditional method, and the lower row is the incision trajectory estimated by the proposed method. From the results, it can be confirmed that the estimated incision trajectories have a similar shape to the trajectory of the Ground Truth. However, the estimated trajectory tends to become shorter because the image has a broad area of high intensity regions so that the intensity centroid tends to get closer to the center of the high intensity regions. The calculation time was less than 11 hours to train the network.

VI. CONCLUSIONS

In this paper, we proposed an estimation method of incision trajectory by restoring thermal source from diffused thermal images using a ConvLSTM and connecting the restored thermal sources. The results of Experiment 1 suggested that the distributed heat was concentrated in the ConvLSTM output, similar to Ground Truth. The results of Experiment 2 suggested that proposed method gave better value than the traditional method. However, the estimated trajectory tends to become shorter because the image has a broad area of high intensity regions so that the intensity centroid tends to get closer to the center of the high intensity regions.

REFERENCES

[1] N. Haouchine, J. Dequidt, I. Peterlik, E. Kerrien, M. O. Berger, and S. Cotin, Image-guided Simulation of Heterogeneous Tissue Deformation for Augmented Reality during Hepatic Surgery, 2013 IEEE International Symposium on Mixed and Augmented Reality, pp. 199-208, 2013.

[2] B. D. Lucas and T. Kanade, An Iterative Image Registration Technique with An Application to Stereo Vision, DARPA Image Understanding Workshop, pp. 674-679, 1981.

[3] C. J. Paulus, N. Haouchine, S. H. Kong, R. V. Soares, D. Cazier, and S. Cotin, Handling Topological Changes during Elastic Registration: Application to Augmented Reality in Laparoscopic Surgery, International Journal of Computer Assisted Radiology and Surgery, pp. 461-470, 2017.

[4] T. Mikolov, S. Kombrink, L. Burget, J. Černocký, and S. Khudanpur, Extensions of Recurrent Neural Network Language Model, 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 5528-5531, 2011.

[5] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. Advances in Neural Information Processing Systems, pp. 802-810, 2015.

[6] C. I. Nwoye, D. Mutter, J. Marescaux, and N. Padoy, Weakly Supervised Convolutional LSTM Approach for Tool Tracking in Laparoscopic Videos, International journal of Computer Assisted Radiology and Surgery, pp. 1059-1067, 2019.

[7] Y. Long, X. She, and S. Mukhopadhyay, HybridNet: Integrating Model-based and Data-driven Learning to Predict Evolution of Dynamical Systems, Proceedings of The 2nd Conference on Robot Learning, pp. 551-560, 2018.

[8] D. P. Kingma and J. L. Ba, Adam: A Method for Stochastic Optimization, 3rd International Conference on Learning Representations, 2015.

[9] E. J. Keogh and M. J. Pazzani, Derivative Dynamic Time Warping, Proceedings of the 2001 SIAM International Conference on Data Mining, pp. 1-11, 2001.