POSTER: Segmentation Technique Based on Object Movement for Speech Production Simulation

Youhei Ohnishi
Osaka University
5-1 Mihogaoka Ibarakishi
567-0047 Osaka Japan
ohnishi.youhei@ais.cmc.osaka-u.ac.jp

Kazunori Nozaki
Osaka University
5-1 Mihogaoka Ibarakishi
567-0047 Osaka Japan
nozaki@ais.cmc.osaka-u.ac.jp

Ken-ichi Baba
Osaka University
5-1 Mihogaoka Ibarakishi
567-0047 Osaka Japan
baba@ais.cmc.osaka-u.ac.jp

ABSTRACT

In this paper, we propose the segmentation technique by using Snakes for 4dimentional magnetic resonance imaging (MRI) data of the movements of the oral tract shapes when pronunciation is performed. In the segmentation of the time-sequence MRI images, there is the specific problem that depends on a tongue shape quick deformation. We found that the Optical Flow of image sets is useful criteria for the decision of the geometry of control points in Snakes. Compared to the normal Snakes, our original method which modified Snakes by utilizing Optical Flow demonstrated superior accuracy of the segmentation of 4D MRI data.

Keywords

Segmentation, Snakes, 4dimensional MRI, Optical Flow

1. INTRODUCTION

For example, there are some studies of simulations for the blood flow by the heart beat and oral air flow by speech. In these simulations, the mesh for computation fluid dynamics is generated geometrical by using geometrical data. The geometrical data of the region of interest (ROI) is extracted image set such as MRI data. In these simulations, there are some temporal changes of ROI in time course. For example, the tongue moves when pronunciation is performed. Those organs movements should be taken into consideration. In order to represent movements of organs, we need to consider the changes of the geometrical data. These simulations often use Arbitrary Lagrangian Eulerian (ALE) method. In ALE method, the changes of the geometrical data are represented by the shift and deformation of the mesh. In order to shift and deform the mesh, we need to get the velocity of the mesh nodes. An approach to get the velocity uses a mathematical model that represent the movement of geometrical data in time course [Watanabe04]. However, in the speech production simulation, it is impossible to use the mathematical model, because the movement of a tongue is volitional. So, the other approach to get the velocity of the mesh nodes without the mathematical model is required, which is to get the velocity from time-sequence image sets. In this approach, the geometrical data is extracted from image sets at each time step. And then, the difference of the geometrical data is computed between at time step T_n and T_{n+1}. It is computed how much each mesh nodes move between adjacent time step by using difference of the geometrical data. The information about movement of the mesh nodes means the velocity of mesh nodes.

Some segmentation methods of using Active Contour Model (ACM) have been proposed, which extract the geometrical data from time-sequence image sets. In our initial implementation, we segment 4D MRI data of oral tract shape while speaking by using ACM simply. However, the problem is that the accuracy of segmentation depends on how much ROI changes between adjacent time steps. For example, if the tip of the tongue moves quickly between adjacent time steps, the result of segmentation tends to be unacceptable.

In this paper, we propose the segmentation technique by using Snakes with the specified criterion derived from Optical Flow in order to capture the quick motion of the tip of the tongue.

2. RELATED STUDY

Active Contour Model

Segmentation methods using ACM semi-automatically extract the geometrical data from time-sequence or spatial sequence image sets. When these segmentation methods are applied to an image in order to get the geometrical data of the boundary edge of ROI as a result contour (RC) of the segmentation, an initial contour (IC) is given on the image by the user and deformed based on some control rules such as some functions in order to close to the boundary edge of ROI. If these segmentation methods are used to segment time-sequence image sets, IC of the image at present time step is RC of the image at previous time step.

There are Snakes [Kass85] and Level Set Method (LSM) [Osher88] for two representative method using ACM. The main differences are how to represent active contour (AC) and control AC. In Snakes, AC is represented by control point (CP) set and deformed in a way that each CP is moved based on the evaluation function which is defined based on some features such as the shape of boundary edge of ROI. In LSM, AC is represented as the zero level set of an auxiliary function called the level set function (LSF), and deformed in a way that, the equation for the evolution of LSF is numerically solved.

The Difference between Snakes and LSM

There are two differences between Snakes and LSM. The first difference is that topology of AC can change only in LSM. The second difference is that Snakes uses the parametric representation for the AC, and LSM uses implicit function.

Optical Flow

Optical Flow is an approximation of the local image motion and specifies how much each pixel moves between adjacent images. We can get the information about velocity of object in the image by using Optical Flow. Generally speaking, Optical Flow computation methods can be classified into two categories: the dense Optical Flow method (the dense method) and sparse Optical Flow method (the sparse method). Block matching method [Huang95] and Horn & Schunck algorithm [Horn81] are the dense method. In the dense method, the velocities of all pixels are calculated as Optical Flow. Lucas-Kanade (LK) method [Osher88] is also the dense method but it can be the sparse method. Pyramidal implementation of the LK (pyramidal LK) method [Bouguet00] is sparse method. In the sparse method, Optical Flow is computed only for an exclusive number of features. By computing Optical Flow only for some features, the accuracy of Optical Flow is able to be higher than the dense method. In order to obtain Optical Flow of the object which moves significantly, Pyramidal LK method is more effective than LK method.

3. PROPOSED METHOD

4D MRI Data of Oral Shape

The MRI data of oral tract shape in pronunciation of “/u/-/s/-/u/-/i/” has been taken by 3 tesla MRI system for a second, and covers before and after pronunci-ation. The partial resolution of the MRI data is 128*128 pixels and the temporal resolution is 16 images for a second. 7 slice images are taken for every time steps. So, there are 121 slice images in the MRI data.

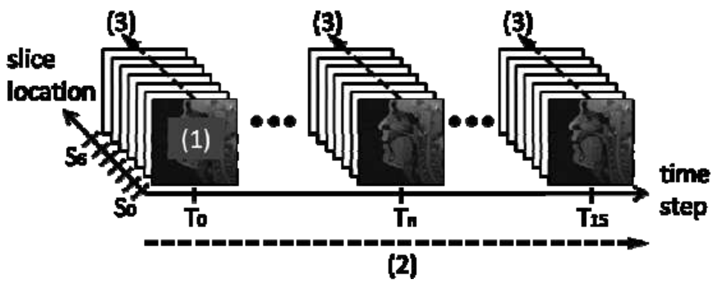

Fig.1: The segmentation process of 4D MRI data

Snakes to Segment 4D MRI Data

In the order to obtain the velocity of the mesh nodes from the difference of the geometrical data sets, there are two requirements for the segmentation to perform the simulations by using ALE method. The former one is keeping the same topology of the AC at every time step. The oral tract shape sometimes splits in two, develops holes, or the reverse of these operations by the motion of the tongue in pronunciation. In fact, the topology changes like these means that the oral tract gets thin or thick. In order to represent the movement of the tongue, the topology of AC must not change. The latter one is keeping the correspondence of the each point of the geometrical data between adjacent time steps. Because the geometrical data is used to compute the velocity of the mesh nodes. Those requirements indicate that Snakes is superior to LSM for our purpose.

The 4D MRI data can be segmented by simply using Snakes. In simply using Snakes, the process of the segmentation of the MRI data consists of 3 phases ((1), (2) and (3) in Fig.1).

In the first phase ((1) in Fig.1), the user gives $IC_{T0,S0}$ which is IC at time step T_0 and at slice location S_0. $IC_{T0,S0}$ is obtained in a way that Img $_{T0,S0}$ which the image at T_0 and at S_0 is smoothed by the Gaussian filter and binarized by threshold.

In the second phase ((2) in Fig.1), the time-sequence images are segmented by using Snakes. $RC_{T0,S0}$ is obtained by segmenting $Img_{T0,S0}$. Then, $RC_{T0,S0}$ is used as $IC_{T1,S0}$, and $Img_{T1,S0}$ is segmented. Thus, the segmentation of $Img_{Tn,S0}$ ($n \in 1\sim15$) is performed by using $RC_{Tn-1,S0}$ as $IC_{Tn,S0}$.

$IC_{T,S}$ = IC used for segmentation of $Img_{T,S}$
$RC_{T,S}$ = the result contour of segmentation of $Img_{T,S}$
for n = 1..the number of time step **do**
    $RC_{n,S}$ = segment using Snakes($Img_{n,S}$, $IC_{n,S}$)
    $IC_{n+1,S} = RC_{n,S}$

Algorithm 1: Pseudo-code of the segmentation algorithm in the second phase.

In the final phase ((3) in Fig.1) is the segmentation of the spatial-sequence images. The segmentation $Img_{Tn,S1}$ is performed by using $RC_{Tn,S0}$ as $IC_{Tn,S1}$. Similarly, the segmentation of $Img_{Tn,Sk}$ ($k \in 1\sim6$) is performed by using $RC_{Tn,Sk}$ as $IC_{Tn,Sk}$. Thus, 4 dimensional geometrical data is obtained by the segmentation of the temporal and spatial sequence images involved in the MRI data.

---

for n = 1..the number of time step **do**
    for k = 1..the number of spatial slice **do**
        $RC_{n,k}$ = segment using snake($Img_{n,k}$, $IC_{n,k}$)
        $IC_{n,k+1} = RC_{n,k}$

Algorithm 2: Pseudo-code of the segmentation algorithm in the final phase.

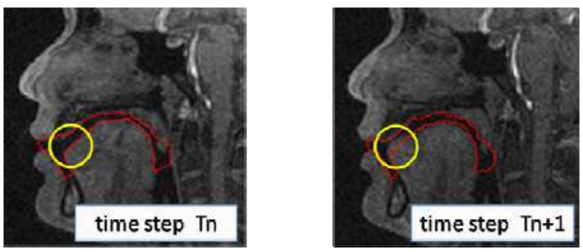

Fig.2: The problem depends on a tongue shape quick deformation.

There is the problem in the second phase of the method simply using Snakes. The problem is that the accuracy of segmentation depends on how much ROI changes between adjacent time steps. For example, if the tip of the tongue moves significantly between adjacent time steps, the result of segmentation is unacceptable (the yellow circle in Fig.2). In order to solve this problem, the criterion of tongue movement incorporated into segmentation process.

Segmentation Technique Based on the Tongue Movement

In order to obtain more accurate RC by alleviating the above-mentioned problem which depends on a tongue shape quick deformation, it is effective to move IC closer to the boundary edge of ROI. In our proposed method, IC is refined based on the object movement estimated between adjacent time steps, and the movement is estimated by using Optical Flow. Now, we are interested in only ROI, not all pixels. So, the pyramidal LK method is utilized with IC as the features to calculate Optical Flow. Fig.3(a), and (b) show the shape of the tongue at time step T_n, T_{n+1} respectively. In these images, black line denotes the boundary edge of the tongue, and green line denotes the RC at time step T_n. In Fig. 3(a), red arrow means Optical Flow of all pixels on the RC calculated by pyramidal LK method. In our proposed method, the refined IC at the time step T_{n+1} is obtained in a way that the information of Optical Flow add to the coordinate data of RC at time step T_n. The refined IC is shown as a red broken line in Fig.3 (b) .

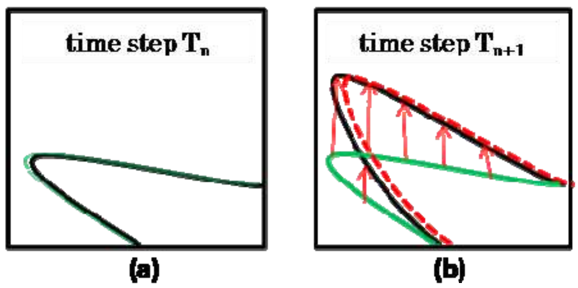

Fig.3: The shape image of tongue at time step of T_n , T_{n+1} respectively.

4. EXPERIMENTAL RESULT

To validate our proposed method, the method have been applied to 2 time-sequence images of the MRI data explained in Chapter 3. In these images, the tip of the tongue moves significantly. Fig.4 shows the results of this evaluation. Here, the window size is arbitrary in calculating Optical Flow using pyramidal LK method. The window of 13*13 pixels was suitable for estimation of the movement captured in the images used in this experimentation. Fig.4 (a) and (b) show respectively the picture at the time step T_n, and T_{n+1}, which are used in this experimentation. Fig.4 (c) is the picture extending the tip of the tongue

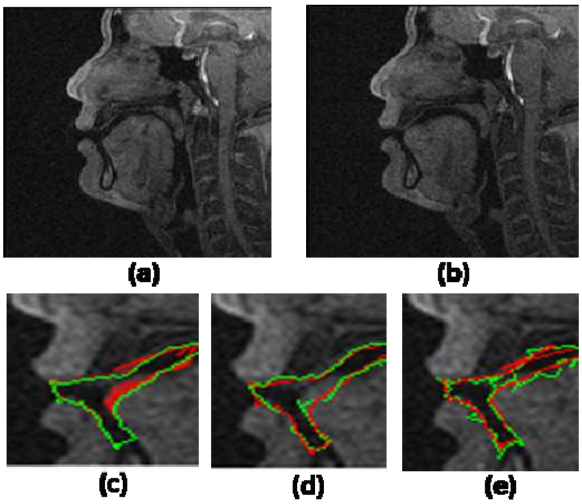

Fig4: The experimental result.

of Fig.4 (a). In Fig.4 (c), the green line and red line respectively show the RC at time step T_n and the Optical Flow calculated from these consecutive pictures. Fig.4 (d) and (e) are the pictures extending the tip of the tongue of Fig.4 (b). Fig.4 (d) shows the result of simply using Snakes and Fig.4 (e) shows the result of our proposed method. In Fig.4 (d) and (e), the red line plots the IC, and the green line plots RC. From these results, our proposed method is more effective than the method simply using Snakes, in order to extract the contour from the 2 time-sequence images in which the object in the images moves significantly.

5. CONCLUSION

In this paper, we proposed the segmentation technique by using Snakes for 4D MRI data of oral tract shape. Our proposed method overcame the oral tract shape deformation problem by utilizing the criterion of Optical Flow. Compared with the normal Snakes, the proposed method could improve the accuracy of segmentation. The proposed method is not taken into consideration of calculation cost so much, and so it is possible to be more rigorous segmentation by improving the accuracy of both the Snakes and Optical Flow.

6. FUTURE WORK

In our future work, we are going to segment all the oral shape MRI data by using our proposed method, and extract 4 dimensional geometrical data. In this work, the window size has been determined experimentally according to the movement of interest. In order to determine the window size that is suitable for several movements, we should analyze more cases in which the method simply using Snakes can not be used to segment. Furthermore, we will generate the mesh from the 4 dimensional geometrical data, and simulate the oral air Flow by using the mesh. Then, we consider the availability of the mesh from the point of the calculation accuracy.

ACKNOWLEDGMENTS

We have the special thanks to National Institute of Information and Communications Technology (NICT) Project of “The development of the high-end device and applications connected with Broad band network”, and Kobe Advanced ICT Research Center (KARC) to take the MRI data.

REFERENCES

[Kass88] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Models, International Journal of Computer Vision, pp. 321-331, 1988

[Lucas81] Lucas, B., and Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision, Proc. of 7th International Joint Conference on Artificial Intelligence (IJCAI), pp. 674-679..

[Bouguet00] Jean-Yves Bouguet. Pyramidal Implementation of the Lucas Kanade Feature Tracker.

[Horn81] Berthold K.P. Horn and Brian G. Schunck. Determining Optical Flow. Artificial Intelligence, 17, pp. 185-203, 1981

[Huang95] Y. Huang and X. H. Zhuang, “Motion-partitioned adaptive block matching for video compression”, 1995.

[Osher88] S.J. Osher, J.A. Sethian, Fronts propagating with curvature dependent speed. Algorithms based on Hamilton-Jacobi formulations. J. Comp. Physics, 79, 1988, 12-49, RESEARCH ARTICLE.

[Watanabe04] Watanabe,H,Sugano,T, Sugiura,S Hisada,T, Finite Element Analysis of Ventricular Wall Motion and Intra-Ventricular Blood Flow in Heart with Myocardial Infarction. JSME Int J(C), 2004; pp 1019-1026.