3D tongue motion visualization based on ultrasound image sequences

Kele Xu[1,2], Yin Yang[3], A. Jaumard-Hakoun[1,2], M. Adda-Decker[4], A. Amelot[4], S. K. Al Kork[1,2], L. Crevier-Buchman[4], P. Chawah[4], G. Dreyfus[2], T. Fux[4], C. Pillot-Loiseau[4], P. Roussel[2], M. Stone[5], B. Denby[1,2]

[1]Université Pierre et Marie Curie, Paris, France
[2]Signal Processing and Machine Learning Laboratory, ESPCI-ParisTech, Paris, France
[3]Electrical and Computer Engineering Department, University of New Mexico, USA
[4]Phonetics and Phonology Laboratory, Université Sorbonne Nouvelle, Paris, France,
[5]Vocal Tract Visualization Lab, University of Maryland Dental School, Baltimore, USA

kelele.xu@gmail.com

Abstract

The article proposes a real-time technique for visualizing tongue motion driven by ultrasound image sequences. Local feature description is used to follow characteristic speckle patterns in a set of mid-sagittal contour points in an ultrasound image sequence, which are then used as markers for describing movements of the tongue. A 3D tongue model is subsequently driven by the motion data extracted from the ultrasound image sequences. The "modal warping" technique is used for real-time tongue deformation visualization. The resulting system will be useful in a variety of domains including speech production study, articulation training, educational scenarios, etc. Some parts of the interface are still being developed; we will show preliminary results in the demonstration.

Index Terms: speech production, ultrasound image, tongue visualization, modal warping.

1 Introduction

Real-time 3D human tongue motion visualization is of importance in the field of speech production, and potential applications are apparent in several fields. For example in a silent speech interface (SSI) system [1], a realistic simulation of tongue movement can provide a direct and effective representation of speech production, which may help to improve the performance of an SSI.

Measuring tongue motion directly, however, is difficult since the tongue lies within the oral cavity and is thus inaccessible for most types of sensors. As a result, various imaging systems have been used to analyze tongue movement indirectly. Magnetic Resonance Imaging (MRI) systems capture tongue movement with good resolution, but require summation of many repetitions or static positions (~10ms) to get good spatiotemporal resolution. MRI is also captured in supine position, which is atypical for speech. X-ray imaging has better temporal resolution, but exposes subjects to radiation and is a through-transmission technique, which projects the entire 3D head onto a single 2D image. Ultrasound imaging of the tongue is attractive because it is non-invasive and can provide real-time images of tongue surface motion [2]. 3D tongue modeling based on Finite Element Modeling (FEM) is also a widely used technique. Much of the literature focuses on tongue motion modeling driven by muscle activations; however, treatment of muscle activation in the tongue is still a challenging problem. In [3], Y. Yang proposed a 3D tongue motion visualization that drives the tongue model with EMA. However, EMA's invasive property affects normal speech production. Here, we present a novel 3D tongue motion visualization framework, which explores the use of midsagittal ultrasound image sequences to drive a 3D tongue model during speech production. Speckle tracking, based on local invariant feature descriptions, is used to follow tongue surface motion in ultrasound image sequences. This motion information is subsequently transmitted to surface nodes of a mid-sagittal tongue model in order to drive the 3D model at the acquisition rate of the ultrasound images.

2 Speckle tracking based on local invariant feature description

In an ultrasound image, interference of scattered acoustic waves produces a gray level pattern commonly referred to as speckle. Although often considered an artifact or noise, speckle can also provide a signature for motion tracking applications. For a high enough frame rate, local speckle patterns are preserved between subsequent frames. The characteristic speckle pattern can thus provide markers for tagging the movement of soft tissue in the image, which can be useful for analyzing articulatory movement [4].

The classical techniques for speckle tracking, such as optical flow and block-matching, work well for small inter-frame displacement and simple deformation; however this is not always the case in ultrasound of the tongue. In this paper, we use local invariant feature description as a method of speckle tracking in ultrasound images to follow speckle signatures with relatively high accuracy and efficiency.

The Scale-Invariant Feature Transform (SIFT) algorithm [5], a state-of-the-art method of local invariant feature description, can be divided into three main steps: feature detection; feature description and feature match. Here, feature description corresponds to speckle pattern description, while the feature match can be regarded as a speckle pattern similarity measure. More specifically, SIFT keypoints are detected as local extrema in scale-space, and keypoints are assigned to one or more orientations based on local image gradient directions. Each keypoint is described as a vector using neighborhood oriented gradient histogram information. In contrast to the original SIFT descriptor, Gaussian Filtering is not used in our method in order to preserve the speckle pattern in the ultrasound image sequences. In the feature matching step, the Euclidean distances between vectors are calculated, to obtain keypoint correspondences.

In Figure 1, the correspondences of two keypoint groups are represented by lines joining regions in two adjacent frames in the sequence. The keypoints on the contour are used as makers tagging the movement of the tongue surface.

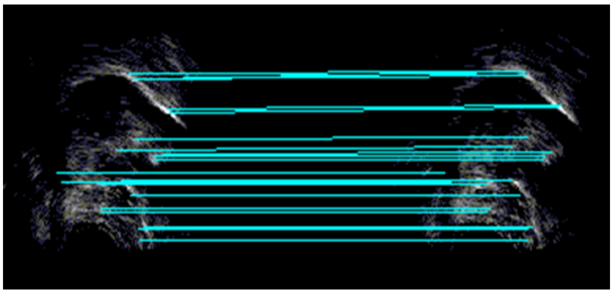

Figure 1: Speckle Tracking based on SIFT in ultrasound tongue image sequences (left: previous frame; right: next frame; the lines represent the correspondences of keypoints in two frames).

3 Deformable tongue motion visualization

Many attempts have been made to simulate the motion of deformable objects so as to speed up simulation and increase the realism. It has been suggested that the deformation of the tongue could be large and the hypothesis of small deformations may not be appropriate [6]. To address relative rotational deformations within the tongue, the modal warping technique has been adopted to obtain high computational efficiency in real-time computed tongue deformations for our framework, thus uniting the benefits of both modal reduction and stiffness warping [7].

In our simulations, the ArtiSynth [8] 3D finite mesh model of the tongue is used, consisting of 1803 nodes and 8606 tetrahedral elements. In Figure 2, six nodes (yellow points) are chosen as anchor nodes corresponding to the hyoid bone and short tendon as discerned in the ultrasound image. These anchor nodes are held fixed during the simulation. We must then set several constraint nodes (green points) lying on the mid-sagittal contour in the ultrasound images, in order to drive tongue motion. The assumption is that the behavior of the 3D model will be approximately correct if the movement of the driven 2D sagittal points is correct. Linear constraints on 3D movements can be integrated into an *Euler-Lagrange equation* using the *Lagrange multiplier method*.

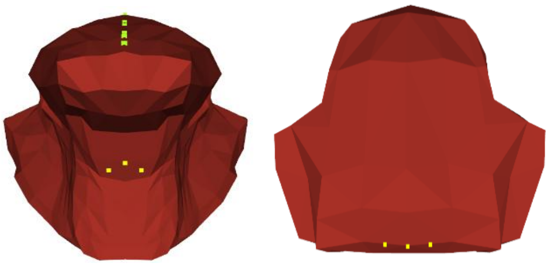

Figure 2: Six Anchor nodes (yellow nodes) on the tongue mesh at the rest position and Constraint Nodes (green nodes) on the mid-sagittal contour at the rest position of tongue model

Figure 3 shows an overview of our interface. Next to the main 3D display, smaller images show two ultrasound frames, from which displacements are calculated. These displacements are then passed to the constraint nodes on the 3D tongue model mid-sagittal surface in order to drive the model.

The interface developed will provide a tool to adapt the tongue rest shape to a specific user. Subject-specific tongue shape information can be extracted by combining mid-sagittal and coronal ultrasound images, using, for example, active contour models (e.g. Snake, [9]) to modify the tongue rest shape in a subject-specific way.

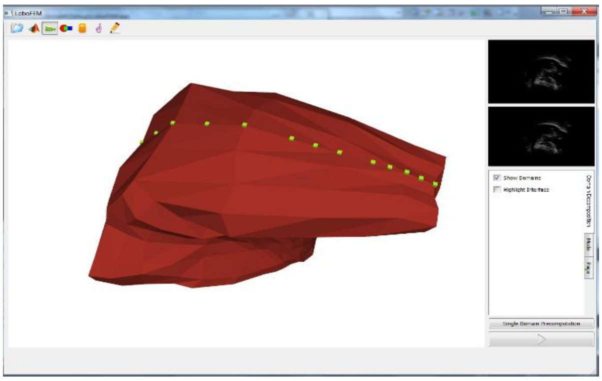

Figure 3: An overview of interface for the proposed system.

4 Conclusions

We use the local invariant feature description method to obtain displacements of reference points in ultrasound image sequences. The motion of a 3D tongue model is driven by the motion of the nodes on the mid-sagittal plane. The modal warping technique is used to obtain 3D tongue deformation visualization in real-time. The demonstration will show 3D tongue model motion visualization based on ultrasound images and the edit tool to modify the tongue rest shape in a subject-specific way, and the original tongue motion based on EMA will be also presented.

At the time of this writing, the system still has some limitations. The performance of speckle tracking is not yet completely stable, due to poor image quality in certain segments of the ultrasound sequence, leading to poor-tracking. Secondly, registration between 2D ultrasound points and the 3D model nodes is still under development.

5 Acknowledgements

This work is funded by the European Commission via the i-Treasures project (FP7-ICT-2011-9-600676-i-Treasures).

6 References

[1] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, “Silent speech interfaces”, Speech Commun., 52(4):270–287, 2010.

[2] M. Stone, "A Guide to Analysing Tongue Motion from Ultrasound Images", Clinical Linguistics and Phonetics 19 (6-7): 455-502, 2005.

[3] Y. Yang, X. Guo, J. Vick, L. G. Torres, and T. F. Campbell, “Physics-based deformable tongue visualization”, IEEE Trans. Vis. Comput. Graph., 19(5):811–23, 2013.

[4] J. D’hooge, “Principles and different techniques for speckle tracking”, in Myocardial Imaging: Tissue Doppler and Speckle Tracking, Blackwell Publishing Ltd., 17-25, 2007.

[5] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, Int. J. Comput. Vis., 60(2):91–110, 2004.

[6] J. M. Gérard, J. Ohayon, V. Luboz, P. Perrier, and Y. Payan, “Indentation for estimating the human tongue soft tissues constitutive law : application to a 3D biomechanical model”, ISMS, Lect. Notes Comp. Sci., 3078:77-83, Springer, 2004.

[7] M. G. Choi and H. S. Ko, “Modal warping: real-time simulation of large rotational deformation and manipulation”, IEEE Trans. Vis. Comput. Graph., 11(1):91–101, 2005.

[8] J. E. Lloyd, I. Stavness, and S. Fels, “ArtiSynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation”, Soft Tissue Biomechanical Modeling for Computer Assisted Surgery, Stud. Mech. Tiss. Eng. Biom.; 11: 355-394 Springer, 2012.

[9] M. Kass, A. Witkin, D. Terzopoulos, “Snakes : Active Contour Models,” Int. J. Compuer Vision, 33(1): 321–331, 1988.