Proceedings of the 6th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/2026.34257

Anchor-Guided Non-Rigid Motion Estimation for the Tongue Using Optical Flow

Keun Ho Kim

Digital Health Research Division, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea
rkim70@kiom.re.kr

Abstract. Tongue movement is related to various neurological and physiological conditions and has recently attracted attention as a potential biomarker in digital healthcare research. However, analyzing tongue motion from sequential images is difficult because tongue deformation is highly non-rigid and motion patterns vary across different tongue regions. In this study, an anchor-guided optical flow method was applied to estimate tongue motion more stably under large deformation conditions. Anchor regions were defined in the upper tongue area and near the tongue tip, and these anchor points were used to construct a global deformation field. Residual local motion obtained from Farnebäck optical flow was additionally incorporated to preserve detailed surface deformation. Tongue image sequences acquired using a tongue imaging system developed at the Korea Institute of Oriental Medicine were used in the experiments. In conventional Farnebäck optical flow results, motion around the tongue tip was not consistently estimated when relatively large displacement occurred. The proposed framework showed more stable displacement estimation while preserving local deformation patterns on the tongue surface. The results indicate that combining anchor-based global deformation with residual optical flow may improve representation of non-rigid tongue motion. Future work will include comparison with deep learning-based optical flow approaches and further evaluation under various tongue motion conditions.

Keywords: Tongue motion estimation, optical flow, non-rigid deformation, anchor-guided motion estimation, Farnebäck optical flow

1. Introduction

Changes in tongue movement are observed in several neurological and systemic disorders and may provide clinically meaningful information about neuromuscular function. Tongue motion depends on coordinated neural and muscular activity, and disruption of these systems can affect movement patterns. In clinical practice, involuntary tremor, repetitive motion, spasms, or abnormal tongue positioning are sometimes identified in patients with neurological disease.

Altered tongue movement has been described in a number of disorders. In Parkinson's disease, rigidity and tremor symptoms may influence tongue motion as well [1]. Huntington's disease is associated with irregular involuntary muscle activity that can appear in tongue movement [2].

DOI: 10.54254/2753-8818/2026.34257

Tardive dyskinesia, which may occur after prolonged use of antipsychotic medication, is characterized by repetitive involuntary movement involving the mouth and tongue [3].

Abnormal tongue motion may also be observed after stroke or in motor neuron disorders. Patients with stroke can show impaired coordination or asymmetrical tongue movement due to neural damage. In amyotrophic lateral sclerosis, tongue weakness and fasciculation are frequently reported during clinical examination [4].

For these reasons, tongue motion has been investigated as a potential biomarker for neurological assessment. Recent developments in medical image analysis and computer vision have enabled quantitative evaluation of tongue movement from sequential image data. Optical flow-based approaches are particularly useful for tracking local displacement and deformation between consecutive frames. However, reliable tongue motion estimation remains difficult because tongue deformation is highly non-rigid and motion patterns vary considerably across different tongue regions.

1.1. Related work

Optical flow methods have been widely used to estimate motion between sequential images in computer vision applications. One representative classical approach is the Farnebäck algorithm, which estimates dense pixel-wise motion from local intensity variation between consecutive frames using polynomial approximation [5]. Because displacement is calculated at the pixel level, the method is suitable for analyzing detailed local motion while maintaining relatively low computational complexity. Despite these advantages, conventional optical flow methods often show reduced performance when large displacement or strong non-rigid deformation is present. Many optical flow algorithms assume brightness consistency and locally smooth motion between adjacent frames. However, these assumptions are not always satisfied in deformable anatomical structures such as the tongue. To address larger motion, coarse-to-fine pyramid approaches have commonly been introduced. In these approaches, motion is estimated first at lower image resolutions and then refined progressively at finer scales. Although this improves robustness in some cases, errors generated at coarse levels may influence later refinement stages, especially when abrupt motion or severe structural deformation occurs. Several studies have attempted to improve non-rigid motion estimation using deformation models that incorporate structural information or prior constraints. Anchor-guided or feature-guided methods stabilize deformation estimation by using predefined reference regions or corresponding feature points [6]. Such approaches can preserve overall structural movement more effectively under large deformation conditions. However, fine local surface motion may still be insufficiently represented when only anchor-based deformation is used. Accurate tongue motion estimation remains challenging because tongue movement includes both large-scale displacement and subtle local deformation simultaneously. As a result, conventional optical flow approaches often have difficulty representing global deformation and local surface motion together in a stable manner.

1.2. Main contributions

In this paper, we propose a pixel-level approach for tracking the motion of non-rigid objects, specifically focusing on the tongue, which exhibits large and complex movements. The main contributions of this study are summarized as follows:

1) We propose an anchor-guided non-rigid deformation framework for tongue motion estimation that reflects the structural characteristics of tongue motion.

DOI: 10.54254/2753-8818/2026.34257

2) The proposed method integrates global deformation modeling with Farnebäck residual optical flow to simultaneously capture large-scale displacement and fine local deformation.

3) Qualitative experimental results indicate that the proposed method achieves more consistent and structurally coherent displacement estimation around the tongue tip compared with the conventional Farnebäck optical flow method.

2. Methods

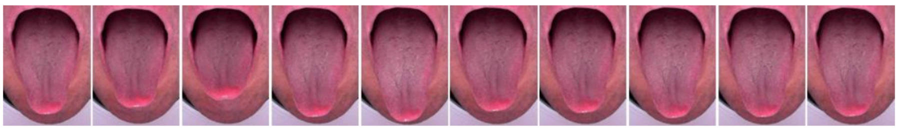
2.1. Data acquisition

Tongue motion images used in this study were collected using a tongue imaging system developed at the Korea Institute of Oriental Medicine (Figure 1(a)). During acquisition, 10 sequential tongue images were recorded at 0.2-second intervals, and the total recording time was approximately 2 seconds (Figure 1(b)). For motion analysis, two consecutive images were selected from the acquired image sequence.

Imaging conditions were controlled to reduce variations in tongue appearance during acquisition. Participants were instructed to undergo imaging at least 2 hours after meals and approximately 1 hour after tooth brushing in order to minimize temporary changes caused by food residue or oral conditions.

Because prolonged tongue exposure can dry the tongue surface through saliva evaporation and may influence tongue color, the exposure time during image acquisition was kept as short as possible. Continuous tongue exposure was limited to less than one minute during the examination.

(a)



(b)

Figure 1. A tongue imaging device and captured images

2.2. Conventional optical flow estimation using Farnebäck method

Given two segmented tongue images without background regions, denoted as $I_1(x,y)$ and $I_2(x,y)$, the motion between consecutive frames is represented by the optical flow vector $F(x,y) = (u(x,y), v(x,y))$ at each pixel location. In this study, dense optical flow was estimated using the Farnebäck method. The Farnebäck algorithm models local image neighborhoods with a

DOI: 10.54254/2753-8818/2026.34257

quadratic polynomial approximation and estimates pixel-wise displacement between adjacent frames based on local intensity variation.

To improve motion estimation under relatively large displacement, a coarse-to-fine pyramid strategy was applied [5]:

$$I_0 \rightarrow I_1 \rightarrow \cdots \rightarrow I_L \tag{1}$$

where L **represents** the number of pyramid levels. **In the present study, six pyramid levels (L=6) were used for optical flow estimation.**

2.3. Proposed anchor-guided non-rigid deformation method

The proposed anchor-guided method utilizes anchor points defined in the upper and lower regions of the tongue to generate a global deformation field. The deformation between these anchor regions is smoothly interpolated, and the resulting global motion is combined with local motion estimated using Farnebäck optical flow. This approach enables stable estimation of non-rigid deformation on the tongue surface.

2.3.1. Anchor-guided non-rigid deformation model

Tongue motion exhibits spatially varying deformation characteristics. In particular, the upper region of the tongue tends to exhibit relatively small motion, whereas the tongue tip often undergoes larger displacements. To account for this behavior, two reference regions are defined in the image. The top anchor is selected by the user based on visually stable features located in the upper or left region of the tongue, while the bottom anchor is also selected based on features located in the lower or right region, including the highlight region observed at the tongue tip. The motion vectors of each anchor are computed as follows:

$$T_{\text{top}} = (u_{\text{top}}, v_{\text{top}}) \tag{2}$$

$$T_{\text{tip}} = (u_{\text{tip}}, v_{\text{tip}}) \tag{3}$$

The global deformation at each position in the tongue region is linearly interpolated according to its relative position between the two anchors:

$$F_{\text{anchor}}(x, y) = (1 - t(y))\, T_{\text{top}} + t(y)\, T_{\text{tip}} \tag{4}$$

where $t(y)$ is a weighting function that increases from the top to the bottom of the tongue, defined as:

$$t\left(y\right) = \frac{y - y_{\text{top}}}{y_{\text{tip}} - y_{\text{top}}} \tag{5}$$

This process generates a non-rigid deformation field that reflects the structural deformation from the root to the tip of the tongue.

DOI: 10.54254/2753-8818/2026.34257

2.3.2. Farnebäck residual flow integration

Although the anchor-based deformation field effectively models global deformation, it has limitations in capturing fine-grained motion on the tongue surface. To address this, the local deformation computed using Farnebäck optical flow is incorporated in a residual form [7]. The final optical flow is defined as:

$$F(x,y) = F_{\text{anchor}}(x,y) + \alpha(x,y)\, F_{\text{FB}}(x,y) \tag{6}$$

where F_{FB} denotes the Farnebäck optical flow, and $\alpha(x,y)$ is a residual weighting function. To ensure stability near the anchor regions, $\alpha(x,y) \approx 0$ is applied around both the top and bottom anchors. In other regions, $\alpha(x,y) \approx 0.4$ is used to incorporate local deformation while preserving the global structure.

2.4. Optical flow vector visualization

To ensure a stable representation of optical flow magnitude, a 95th percentile-based normalization scheme is employed. First, the magnitude of the optical flow is computed as

$$|\, F\left(x,y\right) \,| = \sqrt{u(x,y)^2 + v(x,y)^2} \tag{7}$$

Then, the 95th percentile value of the flow magnitude distribution is defined as p_{95} . The optical flow vectors are visualized by mapping each pixel location (xy) to $(x+su, y+sv)$, where s is a scaling factor. In this study, the scaling factor is set to $s = 0.2$.

3. Experimental results

Figure 2 presents tongue motion between two consecutive frames. In the image sequence, relatively small downward displacement can be observed near the tongue root, while larger movement appears near the tongue tip. This difference suggests that tongue deformation is not spatially uniform and that motion near the tip region is generally larger than motion in the upper region.

Figure 3 shows tongue region segmentation results obtained using a U-Net-based model. Segmentation was successfully performed in both frames. The selected anchor points, T_{top} and T_{tip} , are connected between the two images using cyan and yellow lines. Motion between the anchor regions was estimated according to the relative distance between the two points through interpolation.

Figure 4(a) illustrates the anchor matching results obtained from the proposed method in the upper tongue region and near the tongue tip highlight area. Figure 4(b) shows optical flow estimated using the conventional Farnebäck method with pyr_scale=0.5, levels=6, winsize=45, and iterations=10. Figures 5(a) and 5(b) present another tongue motion sequence in which the lower tongue region moves mainly toward the right side. In this example, Figure 5(a) corresponds to the proposed method and Figure 5(b) corresponds to the conventional Farnebäck method. The conventional method shows relatively weak motion estimation in the lower tongue area, whereas the proposed framework preserves both the small displacement in the upper region and the larger motion observed in the lower region.

DOI: 10.54254/2753-8818/2026.34257

The Farnebäck method estimates motion from intensity variation between consecutive frames. Because of this characteristic, representing overall tongue deformation becomes difficult when large displacement occurs, particularly around the tongue tip region. Although the coarse-to-fine pyramid structure improves robustness, estimation errors may still appear under strong non-rigid deformation. In the proposed method, global deformation is first estimated using anchor regions defined in the upper and lower tongue areas. Residual local motion from Farnebäck optical flow is then incorporated to preserve fine surface deformation.

In the experimental results, the proposed framework showed more stable displacement estimation near the tongue tip and reduced unnecessary motion in relatively stable tongue regions. Local surface deformation was also better preserved while maintaining the overall tongue structure. Compared with the conventional Farnebäck approach, displacement patterns produced by the proposed method appeared more spatially consistent.

These observations suggest that the proposed framework is suitable for representing non-rigid tongue deformation. Recently developed deep learning-based optical flow models may provide additional improvement, and future work will include comparison with such approaches as well as further investigation of their applicability to tongue motion analysis.

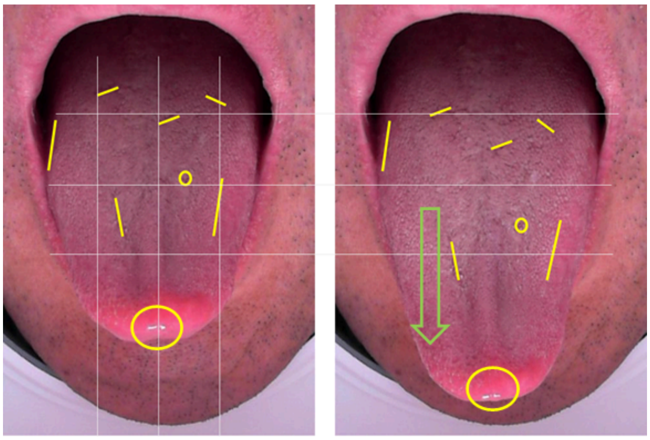


Figure 2. Movement of features on a tongue

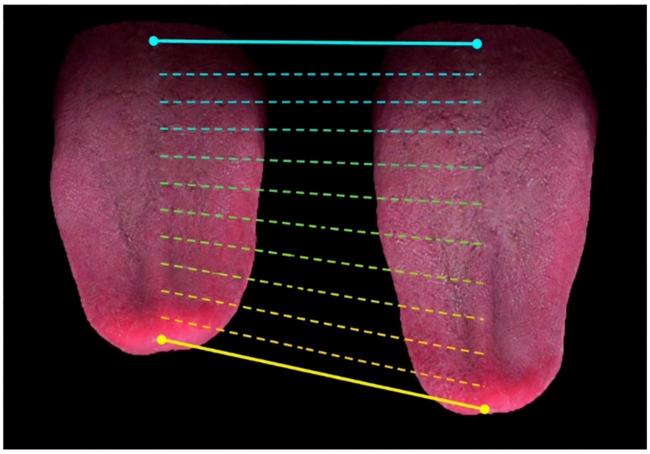
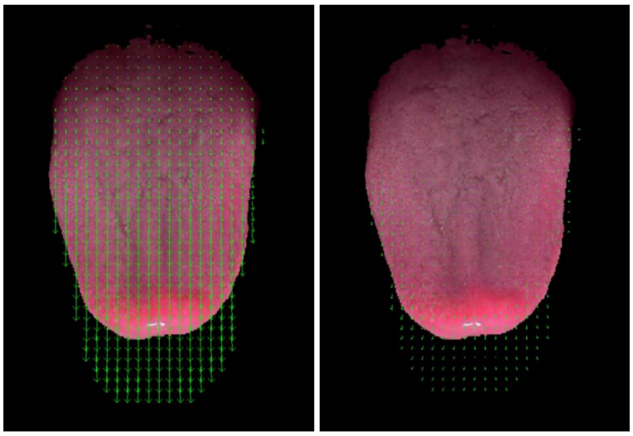


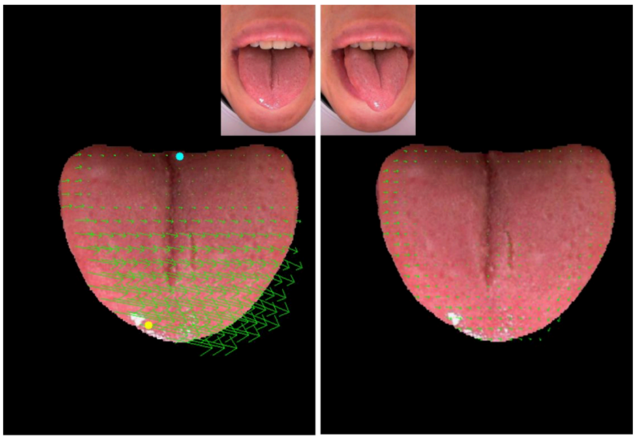
Figure 3. Tongue region extraction using the U-Net algorithm and interpolation of optical flow between the selected anchor vectors T_{top} and T_{tip}

DOI: 10.54254/2753-8818/2026.34257



(a) The proposed method (b) the Farnebäck method

Figure 4. Optical flow estimation results for tongue motion sequence A using (a) the proposed method and (b) the conventional Farnebäck method



(a) The proposed method (b) the Farnebäck method

Figure 5. Optical flow estimation results for tongue motion sequence B using (a) the proposed method and (b) the conventional Farnebäck method

4. Conclusion

This study presented an anchor-guided optical flow framework for estimating tongue motion with non-rigid deformation. The method combined global deformation generated from anchor regions with residual local motion obtained using Farnebäck optical flow. By introducing anchor points in the upper tongue region and near the tongue tip, the proposed framework was designed to better represent spatially varying tongue movement.

In the experimental results, the method showed improved estimation of large displacement around the tongue tip while maintaining local surface deformation patterns. Compared with the conventional Farnebäck method, unnecessary motion in relatively stable regions was reduced and overall tongue deformation appeared more structurally consistent.

These results suggest that the proposed approach may be useful for quantitative tongue motion analysis in medical imaging and digital healthcare research. Further studies are needed to evaluate the method under more diverse tongue motion conditions and to compare its performance with

DOI: 10.54254/2753-8818/2026.34257

recent deep learning-based optical flow models. Additional improvement of preprocessing and postprocessing procedures may also help improve estimation stability and robustness.

Acknowledgement

The Korea Institute of Oriental Medicine (KIOM) under the main research program (Grant No. KSN2512012); the strategic research program (Grant No. SSN2611011)

References

[1] Kearney, E., Haworth, B., Scholl, J., Faloutsos, P., Baljko, M., & Yunusova, Y. (2018) Treating Speech Movement Hypokinesia in Parkinson's Disease: Does Movement Size Matter? Journal of Speech, Language, and Hearing Research, 61(11), 2703–2721. https: //doi.org/10.1044/2018_JSLHR-S-17-0439.
[2] Munhoz, L., Jabbar, A.Q., Silva Filho, W.J.E., Nagai, A.Y., & Arita, E.S. (2023) The oral manifestations of Huntington's disease: A systematic review of prevalence. Oral Diseases, 29(1), 62–74. https: //doi.org/10.1111/odi. 14076.
[3] Ruparelia, P., Ruparelia, P., Pandya, U., Gill, N., & Verma, O. (2022) Restless tongue: Lingual Tardive Dyskinesia – A rare case report. Journal of Oral Biology and Craniofacial Research, 12(1), 99–101. https: //doi.org/10.1016/j. jobcr.2021.10.006.
[4] Rocha, P.S., Folgado, D., Conceição, V.A., Santos, M.O., & de Carvalho, M. (2026) Quantification of tongue motor dysfunction in amyotrophic lateral sclerosis using a smartphone-based task and deep learning. Sensors, 26(5), 1498. https: //doi.org/10.3390/s26051498.
[5] Farnebäck, G. (2003) Two-frame motion estimation based on polynomial expansion. Proceedings of the Scandinavian Conference on Image Analysis (SCIA), Lecture Notes in Computer Science, 2749, 363–370.
[6] Diao, Z., Meng, Z., Li, F., Hou, L., Yamashita, H., Tohei, T., Abe, M., & Sakai, A. (2025) Anchor point based image registration for absolute scale topographic structure detection in microscopy. Scientific Reports, 15, 13486. https: //doi.org/10.1038/s41598-025-98390-5.
[7] Fleet, D., & Weiss, Y. (2006) Optical flow estimation. In Handbook of Mathematical Models in Computer Vision. Springer.