Preliminary Study on Intervertebral Disk Segmentation from Videofluorography by Multi Channelization and CNN

Ayano Fujinaka[a], Kojiro Mekata[b], Hotaka Takizawa[c], Hiroyuki Kudo[c]

[a]Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki, 305-8573, Japan
[b]Department of Rehabilitation, Kobe Red Cross Hospital, 1-3-1 Wakinohamakaigandori, Chuo-ku, Kobe-shi, Hyogo, 651-0073, Japan
[c]Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki, 305-8573, Japan

*Corresponding Author: a.fujinaka@mibel.cs.tsukuba.ac.jp

Abstract

Dysphagia has a large impact on individual patients and the society. However, the whole mechanism has not been analyzed. In order to understand dysphagia, it is essential to describe the anatomical features of cervical structures during swallowing. This study aims to segment cervical intervertebral disks (IDs) in videofluorography (VF) by multi channelization (MC) and convolutional neural network (CNN). The frame images of VF are gray-scale images. In the MC process, feature images are generated by applying image filters, such as the sobel filter and morphological tophat transform filter, to the frame images of VF. Among the feature images, three images are selected, and then color images are generated by setting the selected images to the RGB channels of the color images. The color images are input into CNN for segmentation. The proposed method is applied to actual VF, and experimental results are shown.

Keywords: Dysphagia, Videofluorography, Cervical intervertebral disk, Multi channelization, CNN.

1. Introduction

Swallowing is a vital reflex in human life, but the entire relationship between bones and cartilages during swallowing has not been analyzed yet. If the functions of cervical intervertebral disks (IDs) are impaired due to the reduction of cervical mobility and the compression of cervical spinal cords, dysphagia is likely to occur. However, there is no clear consensus on the shape analysis of IDs for dysphagic patients. There are some preceding studies to investigate dysphagia based on the medical analysis of cervical vertebra diseases and surgical forms [1,2]. There are engineering studies to segment the IDs of dysphagic patients [3,4].

There are methods to input multiple images and modality data to neural networks. Wang et al. proposed joint learning for person re-identification [5]. With this network, single image representation (SIR) and cross-image representation (CIR) are achieved in order to label not only trained people (probes) but also unknown people (galleries). Three layers are input into CNN-based tripled comparison model, thereafter loss functions are calculated. Ngiam et al. proposed multimodal deep learning into which several modalities such as audio and video are input [6]. These inputs are separately set in the Restricted Boltzmann Machines (RBMs), and are trained by a bimodal deep belief network model (DBM). By using a deep autoencoder model, both audio and video are reconstructed from video input.

This paper proposes a segmentation method of IDs in videofluorography (VF) by inputting multiple handcrafted image features to window-based CNN.

2. Multi channelization

Each frame of VF is a gray-scale image as shown in Fig. 1. By applying N image filters, such as the sobel filter and morphological tophat transform filter, to the gray-scale image, N feature images are obtained. Windows of the same

DOI: 10.12792/icisip2019.050

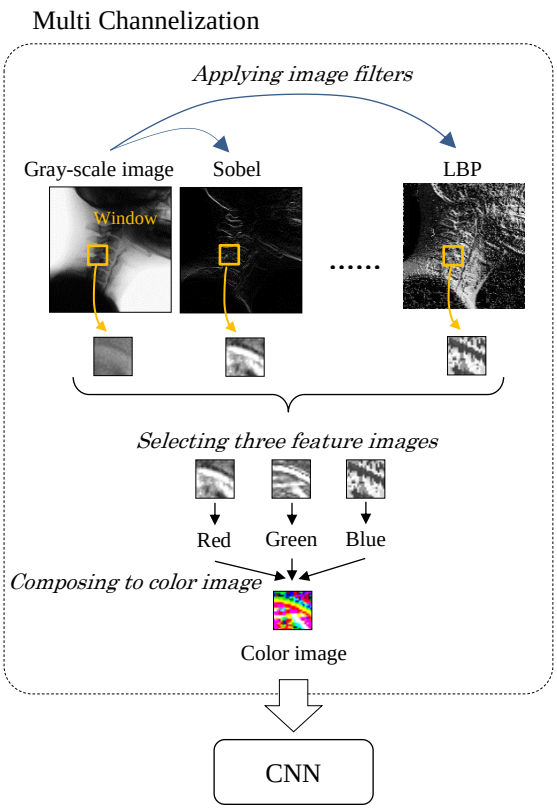


Fig. 1. Multi Channelization and CNN

size are set to the same position in the gray-scale and feature images. Three window images are selected from the *N+1* images including the gray-scale image, and then one color image is generated by setting the three window images to its RGB channels, respectively. This image generation process is called *multi channelization* (*MC*) in this study.

3. Segmentation of IDs by CNN

In the learning phase, cervical masks are extracted from the frame images of VF used for learning [4]. Window-sliding technique is used in the cervical masks, and color images are generated by the multi channelization. The color images are classified into positive and negative data, and are input into CNN.

In the test phase, color images are generated from VF used for test, and are input into the CNN as query data. If the probabilities of the query data are larger than a certain threshold, it is determined to be positive, and vice versa.

4. Experiments

4.1 Experimental Conditions

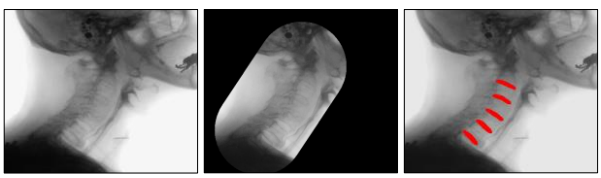
The participants of this study consist of twenty-seven patients and eleven healthy participants. This study is performed based on the documents published by the Japanese Society of Dysphagia Rehabilitation [7]. The frame rates of VF are 0.5 frame per second (fps). The resolution is 0.5 mm / pixel. The thirty-five participants are assigned to learning cases and the other three participants are assigned to test cases. A medical doctor decides the frames with the highest positions of hyoid bones, and manually makes the training data of all visible IDs in these frames. LeNet is used as CNN. The window size is set to be 21×21 pixels.

Gaussian, mean, median, sobel (horizontal and vertical derivations, gradient direction, and gradient magnitude), morphological top-hat transform, pixel values normalization, local binary pattern, and laplacian filters are used as image filters, and are denoted by GAU, MEA, MED, SBX, SBY, SBD, SBM, TOP, NML, LBP, and LPL, respectively (*N=11*). A gray-scale image is denoted by GRY.

Two kinds of experiments are conducted in this study. In Experiment 1, color images are generated by the same feature images. In Experiment 2, on the other hand, color images are generated from different feature images. In our previous study, the Gaussian image is the most promising, therefore it is always used as the R channel component in the second experiment. The RGB channels of a color image are denoted by, for example, (GAU, MEA, MED). The threshold of the probabilities is set to be 0.5. The segmentation accuracy is evaluated by the mean F measure of the three test cases.

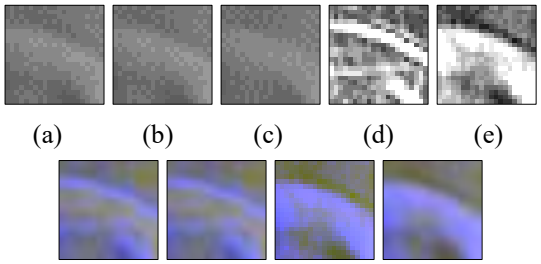
4.2 Results

Fig. 2(a) to (c) show a gray-scale frame image, its cervical mask, and the positive data of a learning case, respectively. Fig. 3(a) to (i) show the window images of (GAU, GAU, GAU), (MEA, MEA, MEA), (MED, MED,



(a) (b) (c)

Fig. 2. A Learning Case – (a) gray-scale frame image; (b) cervical mask; (c) positive data.

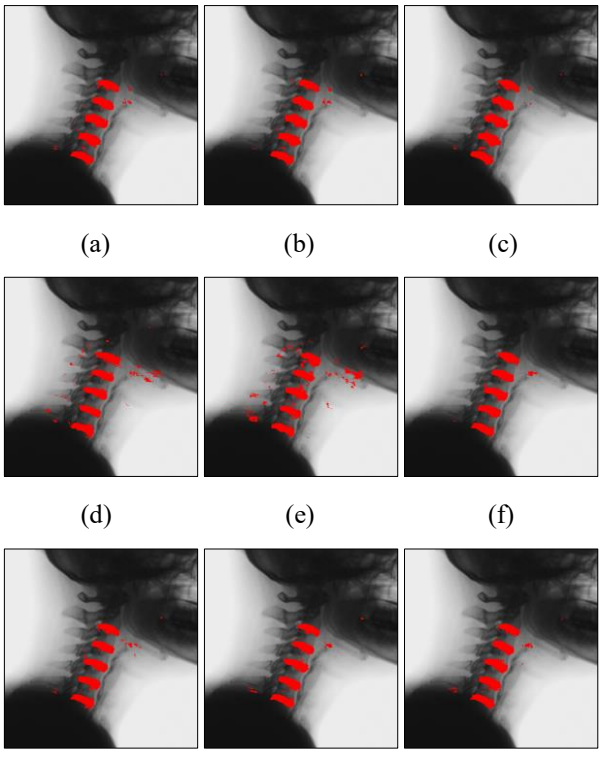


(a) (b) (c) (d) (e)

(f) (g) (h) (i)

Fig. 3. Window Images of IDs – (a) (GAU, GAU, GAU); (b) (MEA, MEA, MEA); (c) (MED, MED, MED); (d) (SBM, SBM, SBM); (e) (TOP, TOP, TOP); (f) (GAU, MEA, SBM); (g) (GAU, MED, SBM); (h) (GAU, MED, TOP); (i) (GAU, MEA, TOP).

MED), (SBM, SBM, SBM), (TOP, TOP, TOP), (GAU, MEA, SBM), (GAU, MED, SBM), (GAU, MED, TOP), and (GAU, MEA, TOP), respectively. Fig. 4(a) to (i) show the segmentation results of a test case when the corresponding filters are applied to.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 4. Segmentation Results of a Test Case – (a) (GAU, GAU, GAU); (b) (MEA, MEA, MEA); (c) (MED, MED, MED); (d) (SBM, SBM, SBM); (e) (TOP, TOP, TOP); (f) (GAU, MEA, SBM); (g) (GAU, MED, SBM); (h) (GAU, MED, TOP); (i) (GAU, MEA, TOP).

Table 1. Results of Experiment 1.

Filters	Orders of accuracy	F measures
(GAU, GAU, GAU)	1	0.611
(MEA, MEA, MEA)	2	0.607
(MED, MED, MED)	3	0.605
(SBM, SBM, SBM)	4	0.600
(TOP, TOP, TOP)	5	0.600
(NML, NML, NML)	6	0.598
(LBP, LBP, LBP)	7	0.585
(SBY, SBY, SBY)	8	0.573
(SBD, SBD, SBD)	9	0.569
(LPL, LPL, LPL)	10	0.560
(SBX, SBX, SBX)	11	0.543

Table 2. Results of Experiment 2.

Filters	Sums of orders	F measures
(GAU, MEA, SBM)	7	0.628
(GAU, MED, SBM)	8	0.622
(GAU, MED, TOP)	9	0.620
(GAU, MEA, TOP)	8	0.613
⋮		

Table 1 shows the combinations of the image filters, the orders of segmentation accuracy, and the mean F measures in Experiment 1. Table 2 shows the combinations of the image filters, the sums of the accuracy orders in Table 1, and the mean F measures.

5. Discussion

The highest F measure in Experiment 2, 0.628, is higher than that in Experiment 1, 0.611. The segmentation accuracy increases when different feature images are combined. It implies that there are synergy effects between these different feature images. GAU, MEA and MED are the top three in Experiment 1, but their combination, (GAU, MEA, MED), is not ranked in the top four in Experiment 2. Instead, the other combinations are ranked in. When GAU is fixed, the F measures are likely to be high if the second filter is a smoothing filter such as MEA or MED, and the third filter is an edge and convex detecting filter such as SBM and TOP.

We are planning to increase the variations of feature images. This would be effective for accuracy improvement,

but requires more calculation time for feature selection. It is necessary to create an efficient method to select the optimal combination of feature images.

In this study, we used LeNet, which was a pioneer CNN. The heights of IDs are approximately 10 pixels (5mm) in general, and LeNet can use windows of suitable size. In this paper, we used the window of 21×21 pixels, which was determined to be the optimal in our previous study. In the future, we should consider using more sophisticated CNNs, such as AlexNet, VGG-Net, GoogLeNet, and ResNet.

6. Conclusions

This study proposes a segmentation method of IDs in VF based on MC and CNN. The experimental results demonstrate that the MC technique is effective to increase the segmentation accuracy.

Acknowledgment

We are grateful to Dr. Jun Matsubayashi in Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Dr. Tomoyuki Takigawa in Department of Orthopaedic Surgery, Okayama University Hospital, Dr. Kazukiyo Toda and Dr. Yasuo Ito in Department of Orthopaedic Surgery, Kobe Red Cross Hospital for helpful discussion.

References

(1) Vaccaro Alexander, Beutler William, Peppelman Walter, Marzluff Joseph M. et al. : “Clinical Outcomes With Selectively Constrained SECURE-C Cervical Disc Arthroplasty: Two-Year Results From a Prospective, Randomized, Controlled, Multicenter Investigational Device Exemption Study”, Spine, Vol.38, Issue 26, pp. 2227–2239, 2013

(2) Hardik Sardana, Hitesh Inder Singh Rai, Amandeep Kumar, Deepak Agrawal, S.S. Kale : “Dysphagia, dysphonia & dyspnoe caused by ostrich beak-like anterior C1-C2 cervical osteophyte”, Interdisciplinary Neurosurgery, Vol.16, pp.132-134, 2019

(3) Yuki Saito, Kojiro Mekata, Hotaka Takizawa, Hiroyuki Kudo : “Preliminary study on segmentation of intervertebral disks in VF images by use of SVM,” The 36th Japanese Society of Medical Imaging Technology (JAMIT), Vol. 1, No. 1, pp. 292-294, 2017.

(4) Ayano Fujinaka, Yuki Saito, Kojiro Mekata, Hotaka Takizawa, Hiroyuki Kudo : “Segmentation of intervertebral disks from videofluorographic images using convolutional neural network”, International Forum on Medical Imaging in Asia (IFMIA), Proceedings Vol. 11050; 110501I, 2019

(5) Faqiang Wang, Wangmeng Zuo, Liang Lin, David Zhang, Lei Zhang : “Joint Learning of Single-Image and Cross-Image Representations for Person Re-identification”, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Electronic ISSN: 1063-6919, 16541210, 2016

(6) Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee et al. : “Multimodal deep learning”, ICML'11 Proceedings of the 28th International Conference on International Conference on Machine Learning, pp. 689-696, 2011

(7) The Japanese society of dysphagia rehabilitation : http://www.jsdr.or.jp/wp-content/uploads/file/doc/VF18-2-p166-186.pdf. Accessed 2 July 2019