Exercise classification using CNN with image frames produced from time-series motion data

Hajime Itoh

Division of Production Systems Engineering, Muroran Institute of Technology, 27-1 Mizumoto-cho
Muroran, 050-8585, Japan

Naohiko Hanajima

College of Design Manufacturing Technology, Muroran Institute of Technology, 27-1 Mizumoto-cho
Muroran, 050-8585, Japan

Yohei Muraoka, Makoto Ohata

Steel Memorial Muroran Hospital, 1-45 Chiribetsu-cho
Muroran, 050-0076, Japan

Masato Mizukami, and Yoshinori Fujihira

College of Design Manufacturing Technology, Muroran Institute of Technology, 27-1 Mizumoto-cho
Muroran, 050-8585, Japan
E-mail: 15042009@mmm.muroran-it.ac.jp, hana@mondo.mech.muroran-it.ac.jp

Abstract

Exercise support systems for the elderly have been developed and some were equipped with a motion sensor to evaluate their exercise motion. Normally, it provides three-dimensional time-series data of over 20 joints. In this study, we propose to apply Convolutional Neural Network (CNN) methodology to the motion evaluation. The method converts the motion data of one exercise interval into one gray scale image. From simulation results, the CNN was possible to classify the images into specified motions.

Keywords: CNN, Gray scale image, Exercises classification, Time-series data.

1. Introduction

Population of the elderly is increasing compared with that of the young in Japan[1]. It is estimated that the number of elderly care staff will be short of approximately 680,000 in 2035[2,3]. In order to contribute for solving this problem, exercise support systems has been developed[4]. It is expected that the elderly can become healthy by strengthening their muscles with gymnastics, and less elderly care staffs will be necessary.

Some exercise support systems have been equipped with a motion sensor to evaluate the elderly' exercise. Normally, the motion sensor provides three-dimensional time-series data of over 20 joints. In previous studies, the motion was judged by human-induced conditions, such as the height of a hand position or the angle of the arm motion[5]. However, the human-induced conditions tended to be experiential and intuition in some case. Sometimes, the judgement did not fit the human judgement[6].

In this study, we propose applying Convolutional Neural-Network (CNN) methodology to the motion

evaluation. The deep learning has succeeded in integrating big data. CNNs is the oldest deep neural-network architecture and is widely used as a means of handwriting recognition[7]. In recent years, image classification has been highly succeeded with CNN[8]. However, there are few examples that are applied to time-series data, and approaches to speech recognition have been proposed[9]. In this paper, we propose to apply CNN methodology to the motion evaluation. The method converts all joints of the motion data into one gray scale image. Then, CNN learns and classifies the converted images. We first explain a method to create motion data for CNN learning in Section 2. Section 3 proposes a conversion method from the motion data to the gray scale image. After that, some case studies are explained. Finally, we show the result of learning using CNN and discuss its ability of classification.

2. Creation of motion data for training data

Learning with CNN requires a lot of training data. However, exercise by humans cannot always do the right motion every time. Therefore, we created motion data by simulating human exercises using simple formulas. They were prepared with respect to what was recorded by a motion sensor in advance[4]. The motion sensor can acquire three-dimensional position data of 20 joints. To imitate the position coordinate, we utilized sinusoidal functions like Eq. (1). Eq. (1) is an example of the coordinate in the vertical direction of a hand position P.

$$P = R + l \times \sin(\theta) + \text{rand} \qquad (0 < \theta < 2\pi) \quad (1)$$

Where R is the coordinate of the rotating joint, l is the length of the arm from the rotating joint to the hand. The function rand is a uniform random number generator with a certain value.

3. Conversion of motion data to image

CNN cannot process time-series data because there is no input unit related to producing image frame from the time. In this section, we propose the conversion method from the time-series motion data to the gray scale image. Then they are learned with CNN.

3.1. Representation of pixels

Fig. 1 shows the image converted from a time-series motion data. Each row expresses X-Y-Z coordinate value for each joint, and the value is converted thereby
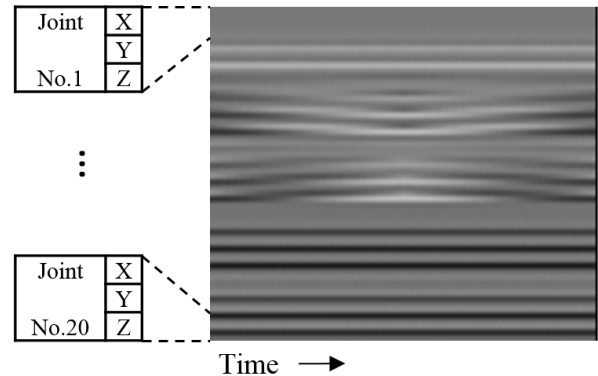

Fig. 1. The image converted from a motion data. One of the stripes represents time-series of the X-Y-Z change of each joint and it is represented in rows in order

to 255 tone gray scale. Time is ordered from left to right for the column. The origin of the X-Y-Z coordinate system is placed 1200 mm downward from the joint at the center of the body, the luminance value changes every 10 mm and can express up to 2550 mm.

3.2. Arrangement of order of rows

We can define several arrangements of order of the row. In this paper, we try two kinds of arrangement. First one is a joint oriented order as shown in Fig. 1. Second one is a coordinate oriented order, that is $x_1, x_2, x_3, \ldots, y_1, y_2, y_3, \ldots, z_1, z_2, z_3, \ldots$, where subscript is the number of the joints.

4. Case study for learning with CNN

We chose five kinds of motion for learning experiment. They are shown in Fig.2 and their descriptions are shown as follows[4].

  i.   Raise and drop arms from sideways to top.
  ii. No action
  iii. Raise and drop your arms from front to top.
  iv. Raise the left leg 90 degrees forward.
  v.   Bend trunk 90 degrees forward

We used Caffe to learn CNN. It is a deep learning framework developed by BVCL. The number of layers of CNN is six, the first layer has kernel size 5 and output is 32, the second layer has kernel size 5 and output is 32, the third layer has kernel size 5 and output is 64. Also, each pooling layer has kernel size 3. The number of output units varies depending on learning

Table 1.

Results when processing 100 images of the motion i. As the image being processed approaches 1, the possibility is higher, and as it is closer to 0, it indicates that there is a high possibility that the image is different.

	Comparative item	i.	ii.	iii.	iv.	v.	vi.	vii.	viii.	ix.	x.
Confirmation of exercise image	Average	1	2.58E-09								
	Standard deviation	0	1.06E-11								
Learning for big motion	Average	1	1.87E-11	2.94E-19	1.39E-23	4.2E-15					
	Standard deviation	0	1.65E-13	2.80E-21	1.00E-25	3.3E-17					
Learning for small change	Average	0.999781	5.54E-17	0	1.88E-29	2.76E-44	0.000219	2.34E-22	0	1.1E-14	0
	Standard deviation	7.07E-06	1.66E-18	0	5.75E-31	8.49E-46	7.07E-06	6.9E-24	0	2.98E-16	0

problem setting. We prepared 100 images for each kind of motion using Eq. (1). The random number contributes the variety of data set. The number of training iterations was set to 4000. We performed three kinds of case study to confirm the ability of classification of the motions using proposed method.

4.1. *Detection of exercise*

In order to confirm whether the exercise movement can be detected correctly, two kinds motions, i and ii were learned with CNN. Their images are greatly different. There are two inputs / outputs in the CNN composition.

4.2. *Classifying for large difference*

There is much difference over the prepared exercises. In order to confirm the ability to classify completely different actions, 5 kinds of exercises, i to v were learned with CNN for the classification of the exercises.
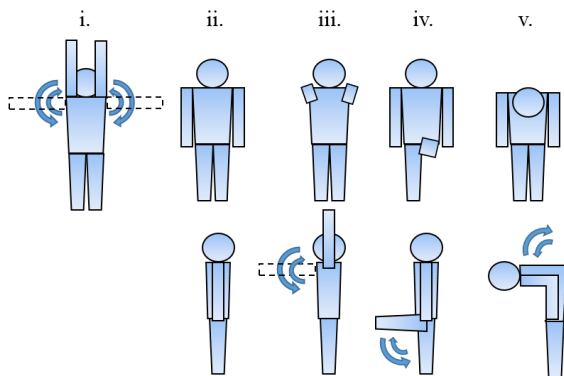

Fig. 2. Exercise to learn at CNN. The direction of the arrow indicates the movement direction, the top is the front view and the bottom is the horizontal.

There are five inputs / outputs in the CNN composition.

4.3. *Classifying for small difference*

There is a possibility that the arm does not rise to the top because of a disorder of the body. It may cause a very small difference in the actions. To investigate ability of classification for such small difference, we prepare five additional motions, vi to x with respect to the motions, i to v. The motions, vi. and viii raise the arms to the lower position from the top by 10 degrees than the motions, i and iii. The motions, ix and x raise the legs up to 80 degrees other than the montions, iv and v. Also, the motion vii raises the arms lower by 10 degrees than the motion i. There are ten inputs / outputs in total. The small change of 10 degrees set here is an exercise that cannot actually be judged by the exercise support system. According to the exercise supervisor, the small change of 10 degrees is judged to be a bad practice, so we will verify whether it can be judged or not in this study.

5. Experimental result and discussion

After learning, classification test was performed. We prepare data set for the classification separately from the learning data, and the number of data is 100. Table 1 shows the probability and standard deviation of classification for result of each motion with test data. At first, the row was arranged for the joint oriented order.

5.1. *Detection of exercise*

It can be seen from first 2 rows in Table 1 that they are correctly classified, that is average of i is 1, and ii is almost zero. The average value of 1 shows possibility of excessive learning. The reason is considered that the kind of the motions is just two and the images used for learning are very similar.

5.2. *Classification for large difference*

From middle 2 rows in Table 1, the motion i is detected correctly. An average value of 1 was output for the motion i. It is considered that the reason is same as in Section 5.1. And, since other average values are very small, it is considered that the correct classification is achieved.

5.3. *Classification for small difference*

In classification including the small change, it is observed from bottom 2 rows Table 1 that the average value of the motion i is almost 1 and the others are quite small values. Therefore, we should say that the classification was achieved. The average of the motion i is smaller than the previous section. The average value of the motion vi is second largest in the motions. It suggests that there is some correlation between the motion i and vi. Actually, the motion vi is a variant of the motion i. These values seem to represent their relationship. Because of this, small changes in luminance values can also be regarded as feature amounts, therefore there is a high possibility of judging the difference in small change accurately.

5.4. *Rearrangement for rows in the coordinate oriented order*

Next, we show the result when the row was arranged in the coordinate oriented order. Table 2 shows a part of the result of the rearrangement for the rows order and learning with CNN.

Table 2. Result of rearranging rows of pixels. Except for the motions, i and vi, since it was less than E - 24, it is excluded in this table.

	i.	vi.
Average	0.999892	0.000108
Standard deviation	3.15E-06	3.15E-06

It is observed that there is no significant difference between two kinds of row orders. It was found that the average value of the motion i in Table 2 was slightly larger than that in Table 1. It indicates that the possibility of obtaining the feature is higher for the motion i.

6. Conclusions

In this paper, we proposed a method to apply time-series motion data to CNN. We showed that learning and classification by CNN are possible by converting the motion data of one exercise interval into one gray scale image. It was found that not only large difference but also small difference in the motion data was correctly classified. Arrangement for rows did not affect the results of classification in this case. We will apply the proposed method to analyze the human exercise data in our future work.

References

1. Cabinet office, government of Japan, *Heisei 28th edition Aged Society White Paper* (Japan, 2016), pp. 2-8.
2. Ministry of Economy, Trade and Industry, *"Study Group on Providing Long-Term Care Service According to Future Nursing Care Demand" Report* (Japan, 2016), pp. 22-39.
3. M. Kasahara et al., Relationship between One Leg Standing Time and Knee Extension Strength in Elderly Patients, *Tairyoku Kagaku*. 50(3) (2001), pp. 369–373.
4. Y. Takahashi, Study on exercise support robots providing individualized instruction, *Presentation of Graduation Theses.* (2014).
5. M. Hirano et al., Development of an Exercise Support System for the Elderly Which Use a small Humanoid Robot, *Journal of Robotics and Mechatronics .* 25(6) (2013) 939–948.
6. K. Urita et al., Precision Analysis of Center-Of-Gravity Points Estimation Based on a Motion Sensor - Comparison with a Force Plate, SICE annual conference 2014 conference proceedings, In *Ternational Conference on Instrumentation*, Information Technology and System Integration (Japan, Sapporo, 2014), pp. 2020–2024.
7. D. Ciresan et al., Multi-column Deep Neural Networks for Image Classification, In *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on (2012), pp. 3642-3649.
8. A. Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks, In *Advances in Neural Information Processing System*s 25 (2012).
9. O. A. Hamid et al., Convolutional Neural Networks for Speech Recognition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 22(10) (2014) 1533–1545.