Human activity recognition by using convolutional neural network

Hankil Kim[1], Sungock Lee[2], Hoekyung Jung[3]
[1]Department of Music & Sound Technology, Korea University of Media Arts, South Korea
[2]Global HRD Inc. 34838, South Koea
[3]Department of Computer Engineering, PaiChai University, South Korea

Article Info

Article history:

Received Apr 24, 2019
Revised Jul 9, 2019
Accepted Jul 17, 2019

Keywords:

Activity Recognition
Convolutional Neural Network
Deep Learning
Intelligent Surveillance System
Thermal Camera

ABSTRACT

In recent years, many researchers have studied the HAR (Human Activity Recognition) system. HAR using smart home sensor is based on computing in smart environment, and intelligent surveillance system conducts intensive research on peripheral support life. The previous system studied in some of the activities is a fixed motion and the methodology is less accurate. In this paper, vision-based studies using thermal imaging cameras improve the accuracy of motion recognition in intelligent surveillance systems. We use one of the deep learning architectures widely used in image recognition systems called Convolutional Neural Networks (CNN). Therefore, we use CNN and thermal cameras to provide accuracy and many features through the proposed method.

Corresponding Author:

Hoekyung Jung,
Departement of Computer Engineering,
Paichai University,
155-40 Baejae-ro, Seogu, DaeJeon, South Korea.
Email: hkjung@pcu.ac.kr

1. INTRODUCTION

The HAR system, a widely used pattern recognition system[1-3], can be divided into several modules such as sensing, feature extraction classification, segmentation and post-processing [4]. HAR systems can be categorized into two types: time-based and acceleration-based. Acceleration-based methods require multiple accelerometers to be used for data collection, but time-based methods typically require the use of one or more cameras to collect data. The disadvantage of the acceleration method is that it can cause discomfort to the human body when performing activities such as walking, running, and lying down.

However, the various human activities to be monitored in this study include hand waving, punching, kicking, lying down, walking, running, and standing. The advantage of a vision-based system is that the sensor works without sticking to the body. However, recognition performance depends on lighting conditions, viewing angle, and other factors. In this paper, we propose a system that uses a time-based data set [5-8] captured by a thermal camera [9,10] and a CNN structure [11-14] to solve this problem. This can reduce the procedures of the handicraft process and increase the accuracy.

2. PROPOSED METHOD AND FEATURE EXTRACTION

Explaining This section introduces the proposed method and feature extraction. First, it explains how to perform the function extraction step-by-step. The first step is to cut the original image by hand. This is because the shape of the image is cut to some shape. The next step is to perform background subtraction based on the ROI coordinates between the background images. CNN [15] Sets the final binary image size, 200 × 200 image, for the input image, to a fixed size. The threshold is then used to obtain the binary image. The threshold is defined as 50 points. At the end of this process, a morphological operation is used and a GEI image of a binary image is obtained. Figure 1 also shows the system architecture for us to understand the system easier.

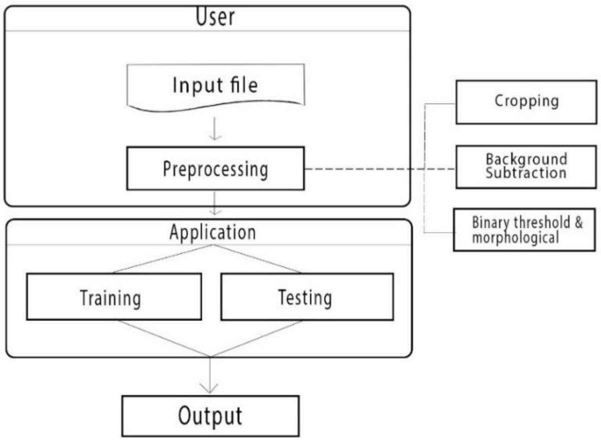

Figure 1. Effects of selecting different switching under dynamic condition

In Figure 1, our main system architecture is that we preprocess the user section and cut some work into handicrafts. The application section is our main proposed method used by the CNN (VGG16-Net) architecture [16-18], and part of the training is to use our model. We use the Keras model and test it as a part of the test. Our dataset is by a trained model. And our recognition system is illustrated in the flow chart of Figure 2. Our feature image is shown in Figure 3.

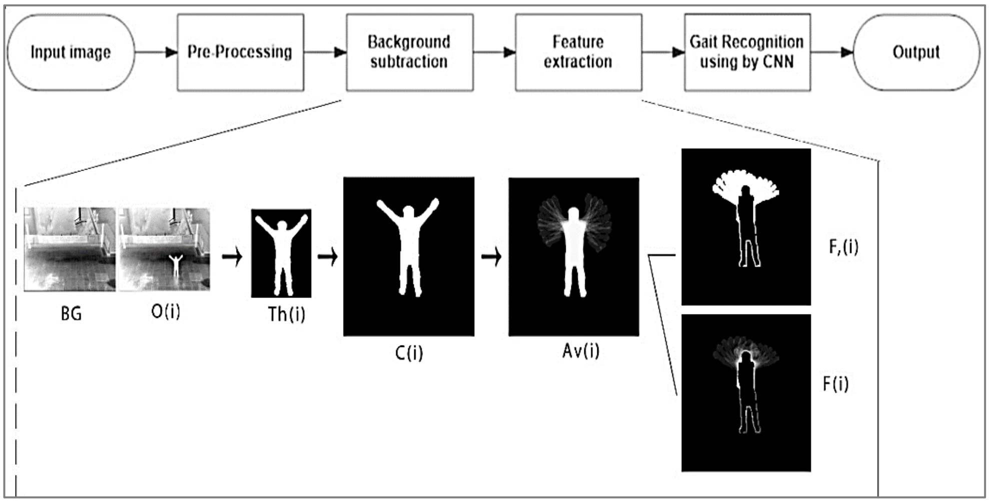

Figure 1. Flow chart of our proposed method

In Figure 2, the Th (i) image is a thermal image by BG and O (i) is a background subtraction. BG image is a background image and O (i) image is an original image. As shown in Figure 2, it cuts the various sizes and appears as a Th (i) image. For C (i) images, create a 224 × 224 blank image fixed size and center the cropped image of the blank image. It then creates an Av (i) average image of the GEI image. F, (i) is the EGEI edited state image and F (i) is as an EGEI image. The last three channels shown in Figure 3 combine Av (i), F, (i), and F (i) images. Figure 3 selects a sample from a final function, such as a) shaking by hand, b) punch c) walking, d) kick, e) lie down, and f) running.

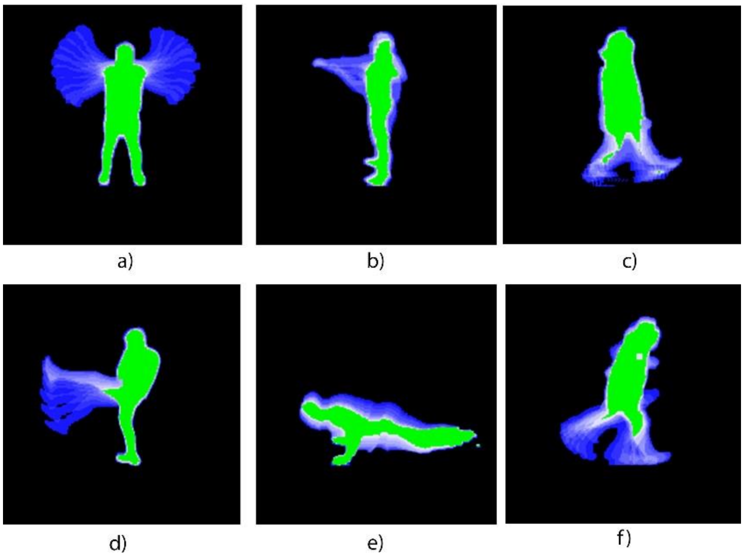

Figure 2. Example of our final features

3. EXPERIMENTAL RESULT

In this section, we briefly describe our databases, comparison and accuracy. First, it's about database 1, collected in multiple environments with objects that are different from people images taken with a thermal camera in a dark environment. So, as mentioned earlier, we have combined all databases with database 1. There are six different people and topics caught in different environments. There is a sample of the database image shown in Figure 4. We can visualize the database description in more detail in Table 1.

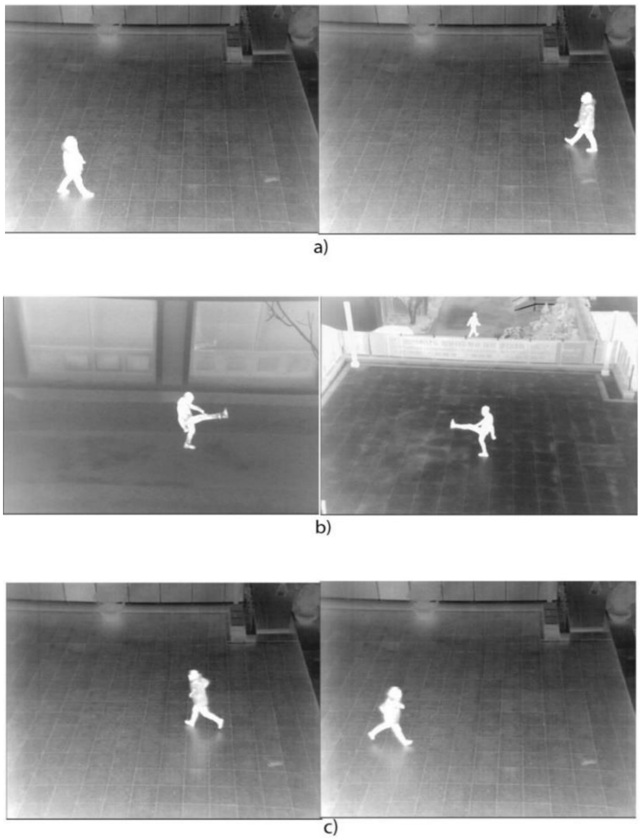

Figure 3. Example of the database of different objects

Table 1. Description of database

Datasets	Detailed description
Figure 4a	- outside next to road
	- man is walking right and left
Figure 4b	- outside and front of building
	- the man is kicking
Figure 4c	- outside next to road
	- the man is running

Table 2 shows the number of images in our database. Table 3 shows the numbers of images and the types of motion in each dataset. The source database represents a three-channel feature image. The augmented database displays artificially augmented feature images. It then divides into one or two sets of data and the original augmented database. For example, an augmented data set was used to study the VGG-16Net [19, 20] and the original data set was used to test the VGG-16Net. In addition, we use the inverse to estimate double cross validation. A description of the HAR system database sample is given above.

Table 1. Number of images in database

Datasets	Total
Original database	2101
Augmented database	42020
Number of people	6
Different environment	5

Table 2. Numbers of images and the types of motion in each dataset

Activity	Frame	Behavior
Hand waving	1163	120
Punching	954	317
Kicking	807	99
Walking	362	165
Running	142	63
Sitting	44	21
Standing	56	11
Laying down	580	207
Total	12108	1003

Second, we introduce the comparison and accuracy. In our study we use the CNN method in HAR systems to improve these problems. CNN allows some features that you do not need to track to detect human leg or hand location information. You can also recognize many activities. Because if we train the data set ready for the inputs of the module, many functions will be recognized as we expected. Table 4 shows the pros and cons between the previous method and our method. In addition, we tested our method and constructed it to estimate double-cross validation in Figures 5-6.

In Figures 5 and 6, we trained ten epochs and it shows the good accuracy end of the last epoch. In NN training, a stroke represents one complete step through a given set of data. The upper direction of the figure is the accuracy, the lower left accuracy is 0 ~ 1 and the number of epoch is 0 ~ 10. Accuracy values are between 0 and 1. Measure 0 means that the feature values match and does not recognize the input image. Measured value 1 indicates the function matching value, and the corresponding compensation value is displayed as true. From 0 to 10 for epoch we have trained 10 sets of data. After double-crossing training, our data set is ready for testing. Table 5 shows the summary of comparisons and accuracies.

Table 3. System of method comparative analysis

Method	Description
Previous methods	- Fixed tasks event handling
	- Complicate to extends, such as add more motions
	- Drawback of recognition about same movements
Our method	- Easy to extend as movements as re-train modules
	- Minimal drawback of recognition more than previous method
	- Easy to use our system

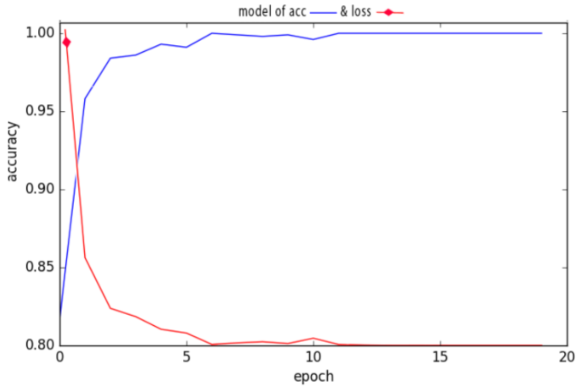

Figure 4. Model of first training

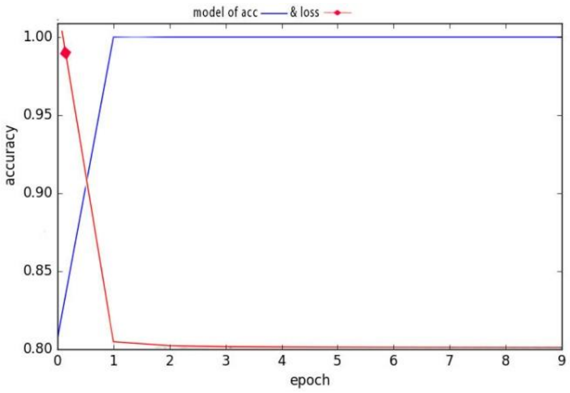

Figure 5. Model of second training

Table 4. Summary of comparisons and accuracies (unit: %)

Method	Accuracy
Fourier descriptor based method	50.4
GEI based method	65.1
Our method	95.9

4. CONCLUSION

Recently, many researchers have studied on the activity recognition[21-25]. There are several types of HAR systems. For example, HAR system can be used for sport activities, daily life activities in the hospital use, patient monitoring after surgery, care for elder people and etc. Intelligent surveillance systems are recognized when a patient falls without people around him.

In this study, however, we proposed this approach to solve the above mentioned problems. Human activity has several unique characteristics that do not require a subject. In this study, human motion using CNN is analyzed and thermography - based human activity recognition is used. The main problem was to resolve the perception of daytime and nighttime human activities in this study that could not recognize previous studies. Vision-based awareness is useful for activity recognition systems. However, the thermal camera is operating during the day and night described above. We can see many studies that have been researched and developed by methods used in HAR systems. Our results were very good and worked well after we used the CNN method in the HAR system. The result is shown as 95.9%, and is more recognizable than the other methods.

We plan to increase the number of activities for the experiment in the future. We will also add more camera types that will allow us to increase our data sets in other dimensions. We will also be presenting research on people who own personal items such as their wallets, bags and cell phones. We also plan to study the accuracy and layout of real-time HAR systems.

ACKNOWLEDGEMENTS

This work was supported by the research grant of Paichai University in 2019.

REFERENCES

[1] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural networks*, vol. 61, pp. 85-117, 2015.

[2] M. I. Jordan and T. M. Mitchell, “Machine learning : Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255-260, 2015.

[3] C. H. Hwang, K.C. Sin and H. K. Jung, “NoSQL Database Design Using UML conceptual Data Model Based on Peter Chen’s Framework,” *International Journal of Applied Engineering Research*, vol. 12, no. 10, pp. 632-636, 2017.

[4] D. C. Ciresan, U. Meier U, and L. M. Gambardella, “Deep, big, simple neural nets for handwritten digit recognition,” *Neural computation*, vol. 22, no. 12, pp. 3207-3220, 2010.

[5] R. Poppe, "A survey on vision-based human action recognition," *Image and vision computing*, vol.28, no.6, pp.976-990, 2010,.

[6] J. Wang, M. She, S. Nahavandi and A. Kouzani, "A review of vision-based gait recognition methods for human identification," *Digital Image Computing: Techniques and Applications (DICTA)*, pp. 320-327, 2010.

[7] P. K. Pisharady and S. Martin Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," *Computer Vision and Image Understanding*, vol. 141, pp. 152-165, 2015.

[8] D. Weinland, R. Remi Ronfard and B. Edmond, "A survey of vision-based methods for action representation, segmentation and recognition," *Computer vision and image understanding*, vol. 115, no. 2, pp. 224-241, 2011.

[9] W. K. Wong, Z. Y. Chew, C. K. Loo and W. S. Lim, "An effective trespasser detection system using thermal camera," in *Computer Research and Development, 2010 Second International Conference,* pp. 702-706.

[10] W. K. Wong, P. N. Tan, C. K. Loo and W. S. Lim, "An effective surveillance system using thermal camera," in *Signal Acquisition and Processing*, ICSAP 2009, pp. 13-17, 2009.

[11] A. Sharif Razavian, H. Azizpour, J. Sullivan and S. Carlsson, "*CNN features off-the-shelf: an astounding baseline for recognition,"* in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 806-813, 2014.

[12] P. Robinson, "The CNN effect: The myth of news, foreign policy and intervention," *Routledge*, 2005.

[13] M. Matsugu, K. Mori, Y. Mitari and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," *Neural Networks*, vol. 16, pp. 555-559, 2003.

[14] J.Y. Kim, S.W. Heo and J.T. Lim, “A License Pl;ate Recognition Algorithm using Multi-Stage Neural Network for Automobile Black-Box Image,” *Journal of the Korea Information and Communication Engineering*, vol. 22, no.1, pp. 44-48, 2018.

[15] S. Dodge and L. Karam, "*Understanding how image quality affects deep neural networks*," in, Quality of Multimedia Experience(QoMEX), 2016 Eighth International Conference, pp. 1-6, 2016.

[16] Z. Liu, X. Li, P. Luo, CC. Loy and X. Tang, "*Semantic image segmentation via deep parsing network*," in Proceedings of the IEEE International Conference on Computer Vision, pp. 1377-1385, 2015.

[17] A.C. Najarro and S.M. Kim, “Nonlinear Compensation Using Artificial Neural Network in Radio-over-Fiber System,” *Journal of Information and Communication Convergence Engineering*, vol. 16, no. 1, pp. 6-11, 2018.

[18] H.K. Jung, J.Y. Kim and H,K. Jung, “Convolution Neural Network(CNN) based I,age Processing System,” *Journal of Information and Communication Convergence Engineering*, vol. 16, no. 3, pp. 160-165, 2018.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, et al, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, no.1, pp. 2825-2830, 2011.

[20] L. C. Jiao, S. Y. Yang, F. Liu, et al, “Seventy years beyond neural networks: retrospect and prospect,” *Chinese Journal of Conputers*, vol. 39, no. 8, pp. 1697-1716, 2016.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[22] M. Henrion, “Propagating uncertainty in Bayesian networks by probabilistic logic sampling,” *Machine Intelligence and Pattern Recognition*, vol. 5, pp. 149-163, 1988.

[23] G. J. J. Burg and P. J. F. Groenen, “GenSVM: A generalized multiclass support vector machine,” *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1-41, 2016.

[24] D. Chen, “Fault Classification Research of Analog Electronic Circuits Based on Support Vector Machine,” *Chemical Engineering Transactions*, vol. 51, pp. 1333-1338, 2016.

[25] C. Ma, C. Chen, Q. Liu, et al, “Sound Quality Evaluation of the Interior Noise of Pure Electric Vehicle Based on Neural Network Model,” *IEEE Transactions on Industrial Electronics*, vol. 64, no. 12, pp. 9442-9450, 2017.

BIOGRAPHIES OF AUTHORS

Hankil Kim, he received the B.S., M.S. degrees from the Department of Electronic Engineering of Hanbat National University, Korea, in 2006, and 2011, and the Ph. D. degree in 2015 from the Department of Computer Engineering of Paichai University, Korea. Since 2006, he has worked in the Department of Music & Sound Technology at Korea University of Media Arts as a professor. His current research interests include multimedia sound, broadcast sound, SR(Sound Reinforcement), Sound Engineering, and database.

Sungock Lee, she received the B.S. degree from the Department of English Language and Culture at Hanyang University, Korea, in 2008, the M.S. degree from the Department of English Language Education at Korea University, Korea, in 2012, and the Ph. D. degree from the Department of Computer Engineering at Paichai University, Korea in 2015. Since 2012, she has worked as an Executive Director at Daewoo Vocational College. Her current research interests include vocational education and training (VET) in computer science, image processing, machine learning, bigdata, VR and IoT.

Hoekyung Jung, he received the M.S. degree in 1987 and Ph. D. degree in 1993 from the Department of Computer Engineering of Kwangwoon University, Korea. From 1994 to 1995, he worked for ETRI as a researcher. Since 1994, he has worked in the Department of Computer Engineering at Paichai University, where he now works as a professor. His current research interests include multimedia document architecture modeling, information processing, information retrieval, machine learning, bigdata, and IoT.