Repetitive Motion Detection for Human Behavior Understanding From Video Images

Orrawan Kumdee
Department of Biomedical Engineering
Faculty of Engineering, Mahidol University
Nakhonpathom, Thailand
orrawan.kumdee@student.mahidol.ac.th

Panrasee Ritthipravat
Department of Biomedical Engineering
Faculty of Engineering, Mahidol University
Nakhonpathom, Thailand
panrasee.rit@mahidol.ac.th* (corresponding author)

***Abstract*— This paper aims to develop a technique for repetitive motion detection which is necessary for human behavior analysis particularly in children with autism spectrum disorders. Images from video sequences are mainly investigated. The technique uses image self-similarity measure, which is less sensitive to view changes, noise, and stable to low resolution images, as input data to multilayer perceptron neural network. Outputs of the network are composed of two classes, which are repetitive and non-repetitive motions. The classifier uses training data from a single person. The model is created by 10 fold cross validation. Trained network is tested with different data sets from seven normal subjects. The classification results show that the proposed technique provides an average accuracy of 0.9115 and can be used in real-time manner. In addition, the trained classifier is robust to images taken from different view.**

Keywords—repetitive motion detection; image self similarity measure; neural networks; children with autistic spectrum disorder.

I. INTRODUCTION

Children with autistic spectrum disorders (ASD) normally have three main impairments, i.e., social interaction, social communication, and imagination [1]. They are more likely to play alone in their world and do not understand social cues created by other people [2]. Some of them display stereotyped behaviors which are undesired repetitive motions, such as jumping, hand flapping etc. From the current report, the incidence rate is approximately 1 in 68 children in the United States [3]. Currently, there is no known cure for autism. Behavior intervention is generally used in order to teach each child to perceive and behave like normal children [4]. In order to measure the progress of behavior intervention, many research projects observe children's behavior from recorded video (VDO) data. The behaviors include joint-attention, eye contact, eye gaze, touch, handling, imitation, speech, vocalizations, social orienting, approaching, moving away, and stereotyped behaviors etc [5] – [9]. The frequencies of both positive and negative behaviors are evaluated by either therapists or re-searchers. Post-hoc analysis of behavior observations from VDO sequences is a tedious task because therapists have to evaluate each of the image sequences ranging from tenths of a second to the entire session and the results depended on prior knowledge of each evaluator [6]. Automatic behavior understanding from images is thus required.

Feil-Seifer and Mataric [10] produced automated detection and classification of positive and negative interactions between a robot and autistic children by using an overhead camera with two infrared (IR) emitters from their robot. Gaussian mixture models (GMM) and Naive-Bayes classifier were used. Various behaviors were able to be identified such as avoiding the robot, interaction with the robot or playing with bubbles, staying still, near a parent, against a wall and none at all.

Dorothée et al. [11] provided a concept of on-line behavior classification and adaptation of a robot's behavior according to human-robot interaction manners. Two interaction styles, i.e., 'gentle interaction' and 'strong interaction', were classified. The results showed that a dog robot, known as 'AIBO', was able to identify which interaction it was involved in.

Albinali et al. [12] used three wireless accelerometers placed on each wrist of children with autism to recognize stereotyped and repetitive motor movements for hand flapping and body rocking with six autistic children. The experiments which took place in the laboratory and classroom within real-time monitoring showed 88.6% accuracy could be achieved.

Westeyn et al. [13] also used wireless-bluetooth accelerometer in a proof of concept for continuous recognition of stereotyped behaviors for an autistic child. The activities included drumming, hand flapping, hand striking, pacing, and rocking. The data set was created from mimicking autistic behavior and HMM was used in the recognition part.

In addition, several review articles [14-17] for understanding normal human behavior from video images have presented a framework of human activity classification including walking, running, jumping, pulling, fighting and hugging etc. These behaviors are found in many applications, such as video surveillance, human-computer interface, video indexing and video retrieval.

From previous studies, many research projects have proposed various techniques for human behavior understanding from VDO sequences. Unfortunately, repetitive motion detection for children with ASD is rarely investigated.

In this paper, repetitive motion detection is mainly focused. Image self-similarity measure is used in behavior classification. It is defined as an absolute correlation score of a pair of images in the VDO sequence. From previous studies, the image self-similarity measure was effectively used to recognize different types of periodic motions. It is relatively robust in real-time situation [18-21]. However, it has never been used with hand flapping behavior frequently found in children with autistic spectrum disorders. The hand flapping is a periodic motion as can be seen in Fig. 1. In this study, data from a video sequence are solely used. Special sensors are not required..

Fig. 1. Example of flapping hand of children with autism.

In section 2, image self-similarity measure is described. Experiments are then presented in section 3. Finally, discussion and conclusion are provided.

II. Method

Detection and classification of repetitive motions are mainly investigated in this paper. The proposed technique uses image self-similarity measurement and feed forward neural network. They are presented as follows.

A. *Image self-similarity measurement*

Cutler and Davis [18,19] proposed new technique to detect and analyze a periodic motion. They calculate an object's self-similarity as it changes over time. Their work is motivated by a human ability that can recognize periodic movements even in very low resolution images. Their study mainly focused on repetitive motions of humans or animals, such as walking and running, dogs running, and birds flying etc. In signal processing, if the signal $\vec{X}(t)$ is periodic it repeats itself with a constant of time in a period p as defined by.

$$\vec{X}(t) = \vec{X}(t+p) \tag{1}$$

Fig. 2. A walking pattern

Fig. 2 shows a walking pattern. Image self-similarity can be measured by many techniques such as normalized cross-correlation, Euclidean distance, absolute correlation from the image sequence, $I = I_{t_1}, I_{t_2}, I_{t_3}, \ldots, I_{t_n}$. In this study, the image self-similarity, S_{t_1,t_2}, is computed from an absolute correlation of the images at frame t_1 and frame t_2 as.

$$S_{t_1,t_2} = \sum \left| I_{t_1} - I_{t_2} \right| \tag{2}$$

Two-dimensional self-similarity matrix can be created by the whole image sequence as.

$$M(I) = \begin{bmatrix} S_{t_1,t_1} & S_{t_1,t_2} & S_{t_1,t_3} & \cdots & S_{t_1,t_n} \\ S_{t_2,t_1} & S_{t_2,t_2} & S_{t_2,t_3} & \cdots & S_{t_2,t_n} \\ S_{t_3,t_1} & S_{t_3,t_2} & S_{t_3,t_3} & \cdots & S_{t_3,t_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{t_n,t_1} & S_{t_n,t_2} & S_{t_n,t_3} & \cdots & S_{t_n,t_n} \end{bmatrix} \tag{2}$$

The self-similarity matrix can be plotted in 2D as seen in Fig. 3. In this figure, the lattice structure of repetitive motions can be clearly observed.

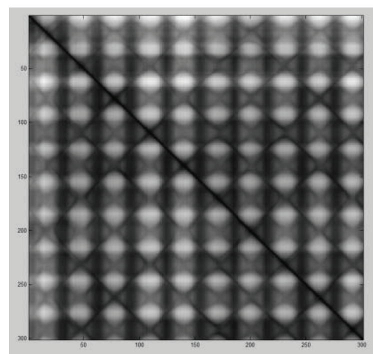
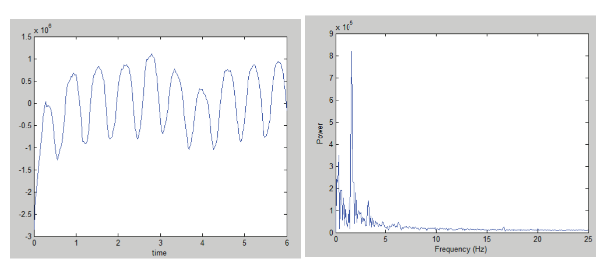


Fig. 3. Image self-similarity matrix plot for a walking pattern

For classification of a periodic motion, Cutler and Davis [18,19] used one dimensional signal of image self-similarity values generated by comparing a fixed image I_{t_1} with the subsequent images, $I_{t_2}, I_{t_3}, \ldots, I_{t_n}$, as presented in Fig. 4 (a). The signal was used to calculate the power spectrum as seen in Fig. 4(b). Periodic motion was detected by finding a peak of the power spectrum that was more than a certain threshold, K. The threshold was set to 3 times of standard deviation and plus with the average power spectrum as.

$$P(f_i) > \mu_P + K * \sigma_P \tag{3}$$



(a) (b)

Fig. 4. (a) Image self-similarity values of image I_{t_1} compared with subsequent images $I_{t_2}, I_{t_3}, \ldots, I_{t_n}$ (b) Average power spectrum of the signal

However, this technique is impractical when the image self-similarity signal contains non-Gaussian noises and when the repetitive period is not constant. Moreover, setting a specific threshold is not generalizable when the image self-similarity signal is highly nonlinear.

Later on, 2D lattice structures in the similarity matrix are employed to classify a periodic motion instead of a 1D power spectrum in order to minimize the number of false periodicities while maximizing the number of true periodicities.

BenAbdelkader et al. [20, 21] applied the image self-similarity measure to recognize gait patterns of people. Their first paper used PCA in the classification and a recognition rate of 93% was gained. In their second approach, they applied the so-called subspace linear discriminant analysis (s-LDA) to improve recognition performance. Image self-similarity was used because it is robust to classification in noisy data and works well with low-resolution images. Junejo et al. [22] explored the self-similarity of action sequences over time. The actions were bending, jumping, running and waving etc. In addition, they showed the self-similarity matrix of golf swinging from two different views. The self-similar structure of these images were similar. Jonathan [23] also used self-similarity to visualize and identify the structure and rhythm of music and audio for orchestral, jazz, and popular music. Their study was mainly used for musicological analysis.

Nevertheless, previous studies cannot be used directly with our study because an unclear lattice structure of image sequences is generated. The simple classification technique is not practical particularly when small repetitive motions exist. In this paper, a feed forward neural network is employed for behavior classification. One dimensional image self-similarity is used as the input data of the network.

B. Artificial neural network (ANN)

Artificial neural networks can be considered as a highly complex, nonlinear and parallel processing system that can learn from experiences. A standard feed forward neural network can be separated into three layers including an input layer, a hidden layer and an output layer. Each layer is connected to the subsequent layer by weighted connections as shown in Fig. 5. ANN can learn through adjustments of the weights based on the error signal. In this paper, multi-layer perceptrons with backpropagation training algorithm (MLP) method is used for repetitive and non-repetitive motion classification.

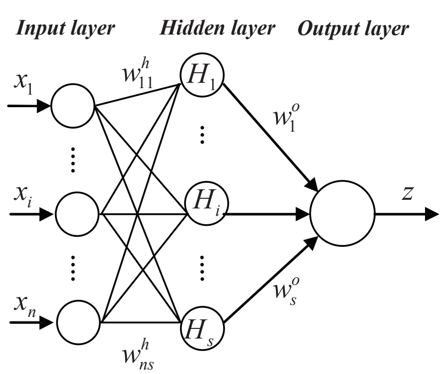


Fig. 5. Structure of multilayer perceptrons with backpropagation learning (MLP) [24]

III. EXPERIMENTS

In the study, hand flapping behavior are mainly focused. Video sequences from eight subjects were recorded by the use of a JVC Everio GZ-HM30SAG HD Camcorder. Images were captured at the rate of 50 fps. The captured images were then resized to 160x120 pixels for ease of manipulation [18]. VDO data from only one subject were used for training. The rest of the data were used for model validation. For the training data, the subject was asked to do hand flapping for 12 seconds. After that, she was asked to stay still for 12 seconds as seen in Fig. 6. Two views of these motions, i.e., front-view and right-side views, were recorded as seen in Fig. 7. One of the authors manually assigns observed behavior in each image as repetitive or non-repetitive motion. From these VDOs, both 1D and 2D image self-similarity values are created as seen in Figs. 8 and 9. For the staying still part, the signals are shown in Fig. 10. As can be seen in these figures, lattice structures in hand flapping behavior and staying still cannot be clearly observed. Direct implementation of the previous repetitive motion classification [19] is thus not practical. In this study, a 1D image self-similarity signal is taken into account. The signal is chopped for analysis with the window size of 50 frames. Windowing technique is preceded for creating the training data. Fig. 11 shows an example of an image self-similarity signal and the windowing process. The input data from both the hand flapping and staying still parts are collected. In this study, each class possesses 1100 input data taken from both views. For the transition period containing both the hand flapping and staying still part, the signals are assigned to be in the staying still pattern. In total, there are 2230 input data for the training. The learning rate is varied from 0.05, 0.1 – 0.5. Momentum is adjusted from 0.6-0.8. Hidden nodes are finely tuned from 2-10 nodes. Ten fold cross validation is applied. An averaged area under the ROC curve is used for model evaluation. The model is validated with VDO sequences from seven different subjects. VDO images of four subjects are taken from similar angles with the training data. The angles are front and right-side views. Images from the rest of the three subjects are taken from front and left-side, which are different from the training set. By doing so, robustness of the classification model to view changes can be tested. Classification results are shown in table 1, and Fig. 12 shows an example of the results.

Fig. 6. Example of hand flapping from front and side views

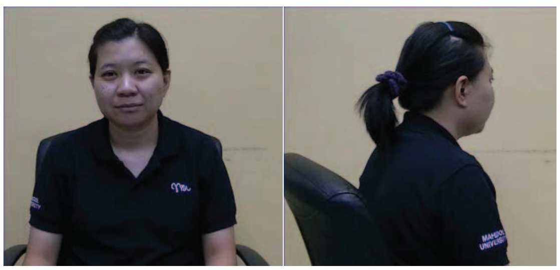
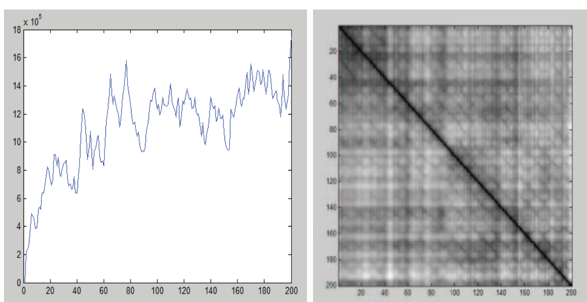
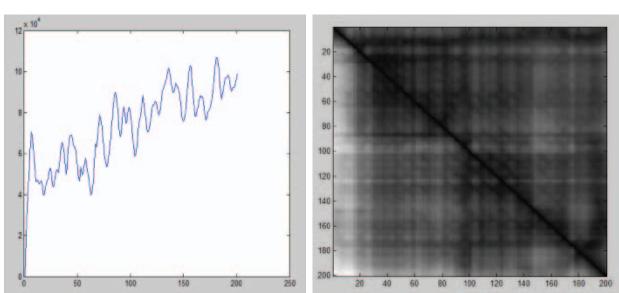


Fig. 7. Show the normal adult sitting on the chair with front and side view



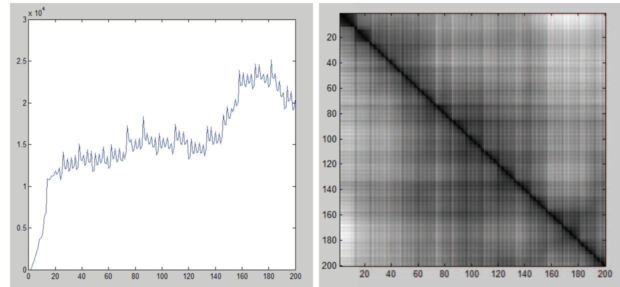
(a) (b)

Fig. 8. (a) 1D Image self-similarity signal and (b) lattice structure of image I_{t_1} compared with images $I_{t_2}, I_{t_3}, \ldots, I_{t_n}$ of hand flapping motion taken from the front view



(a) (b)

Fig. 9. (a) 1D Image self-similarity signal and (b) lattice structure of image I_{t_1} compared with images $I_{t_2}, I_{t_3}, \ldots, I_{t_n}$ of hand flapping motion taken from the side view



(a) (b)

Fig. 10. (a) 1D Image self-similarity signal and (b) lattice structure of image I_{t_1} compared with images $I_{t_2}, I_{t_3}, \ldots, I_{t_n}$ of the staying still part

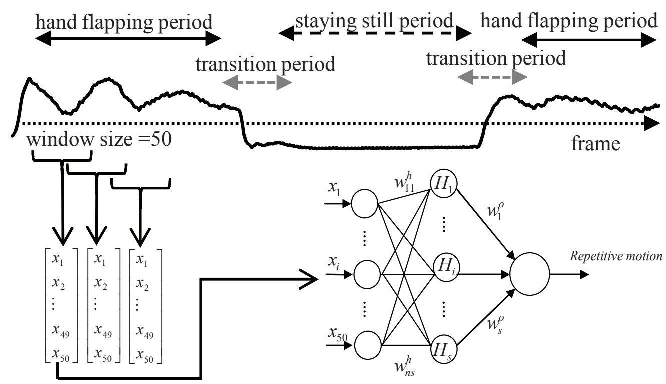


Fig. 11. An example of image self-similarity signal and windowing process

Fig. 12. Example of the classification result

TABLE I. RESULT OF CLASSIFICATION OF REPETITIVE MOTION

View	Classification		
	Sensitivity	*Specificity*	*Accuracy*
Front	0.986	0.815	0.908
Side	0.970	0.852	0.915
AVG.	**0.978**	**0.8335**	**0.9115**

As seen above, the performance from the front and side view is above 90% and the average values of sensitivity, specificity and accuracy are 0.978, 0.8335 and 0.9115, respectively.

IV. DISCUSSION

From the experiments, classification accuracies of images taken from front view and side view are not very different. The average accuracy of the front view is 0.908 whereas the accuracy of the side view is 0.915. The results show that the proposed algorithm can efficiently differentiate hand flapping and staying still. In addition, it is robust to images taken from different views. For the training data, images taken from the right-side view only are taken into account. In the testing, images of three subjects were intentionally taken from a different angle which is the left-side view. With good classification results, the system can be confidently used when view changes occur. The image self-similarity signals from these images were not seriously different. The trained classifier can be used without relearning. Testing errors mainly occur in the transition period where motion changes between hand flapping and staying still occur. Adding more training examples may increase the classification accuracy. From the results, the average accuracy is 0.9115. The algorithm is efficient for detecting hand flapping behaviors. However, in this study, the experiments were conducted in the controlled environment. The subjects mainly sit steadily and only move their hands. For real testing with autistic children, the results may be different because some children may not sit tidily. In these cases, human tracking may be added in order to constantly capture children movements.

V. CONCLUTION

This paper presents a repetitive motion detection technique. Hand flapping behavior is mainly investigated. This behavior is frequently found in children with autistic spectrum disorders. The technique uses image self-similarity values as inputs to the multilayer perceptron neural network. Experiments with eight normal subjects were conducted. The results showed that the technique was efficient for hand flapping detection and it was robust to images taken from different views. For future works, real implementation with autistic children should be tested. Different repetitive motions, such as jumping, body-rocking, and flapping of objects etc. may be further evaluated.

ACKNOWLEDGMENT

The research was supported by National Science and Technology Development Agency (NSTDA) and National Research Council of Thailand (NRCT).

REFERENCES

[1] American Psychiatric Association, “Diagnostic and Statistic Manual, Mental Disorders.4th Text Revision”, Washington D.C. American Psychiatric Association Press, 2000.

[2] S.Melinda, S. Jeanne and Ted Hutman, "Autism Symptoms and Early Signs", http://www.helpguide.org/articles/autism/autism-symptoms-and-early-signs.htm, last updated: July 2015.

[3] D. M. N. S. Y. Developmental, and P. Investigators, “Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010,” Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002), vol. 63, pp. 1, 2014.

[4] Centers for Disease Control and Prevention, “Autism Spectrum Disorder (ASD): Treatment”, Available from: http://www.cdc.gov/ ncbddd/autism/treatment.html , Updated February 24, 2015

[5] J. Hooker, “Observing Joint Attention in Children with Autism: The Development of a Video Coding Protocol”. Undergraduate Honors Theses, 2013.

[6] K.M. Tsui, D.J. Feil-Seifer, M.J. Matarić , H.A. Yanco, "Performance evaluation methods for assistive robotic technology", Performance Evaluation and Benchmarking of Intelligent Systems: Springer, p. 41-66, 2009.

[7] B. Robins, K. Dautenhahn, R. te Boekhorst, and A. Billard, “Effects of repeated exposure to a humanoid robot on children with autism ", In Designing a more inclusive world, Springer London, pp. 225-236, 2004.

[8] F. Michaud, T. Salter, A. Duquette, H. Mercier, M. Lauria, H. Larouche, and F. Larose. “Assistive Technologies and Child-Robot Interaction”,. In AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics, 2007.

[9] Baranek G.T., “Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age”. Journal of autism and developmental disorders, pp. 213-24, 1999.

[10] F.S.David and M. Mataric. "Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features." In Proceedings of the 6th international conference on Human-robot interaction, pp. 323-330. ACM, 2011.

[11] F.Dorothée, D. Polani , and K. Dautenhahn, "On-line behaviour classification and adaptation to human-robot interaction styles.", Proceedings of the ACM/IEEE international conference on Human-robot interaction. ACM, 2007.

[12] F. Albinali, M.S. Goodwin and S.S. Intille, "Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum", Int. Conf. Ubiquitous Comput., pp. 71-80, 2009.

[13] T. Westeyn, K. Vadas, X. Bian, T. Starner and G. D. Abowd, "Recognizing mimicked autistic self-stimulatory behaviors using hmms", In Wearable Computers, Proceedings. Ninth IEEE International Symposium on, pp. 164-167), 2005.

[14] Borges PVK, Conci N, Cavallaro A., “Video-based human behavior understanding: a survey”, Circuits and Systems for Video Technology, pp. 1993-2008, 2013.

[15] O.P. Popoola and K. Wang, "Video-based abnormal human behavior recognition—A review", Systems Man and Cybernetics, Part C: Applications and Reviews, pp. 865-78, 2012.

[16] M. Pantic , A. Pentland , A. Nijholt and T.S. Huang . "Human computing and machine understanding of human behavior: a survey."

Artifical Intelligence for Human Computing. Springer Berlin Heidelberg, pp. 47-71, 2007.

[17] L. Wang , W. Hu and T. Tan, "Recent developments in human motion analysis", Pattern recognition, pp. 585-601, 2003.

[18] R. Cutler and L.S. Davis, "Real-Time Periodic Motion Detection, Analysis, and Applications". Proc. Computer Vision and Pattern Recognition, pp. 326-332, 1999.

[19] R. Cutler and L.S. Davis, "Robust Real-Time Periodic Motion Detection, Analysis, and Applications," T-PAMI, 22(8), pp. 781-796, 2000.

[20] C. BenAbdelkader, R. Cutler, H. Nanda and L. Davis, "Eigengait: Motion-based recognition of people using image self-similarity", In Audio-and Video-Based Biometric Person Authentication, Springer Berlin Heidelberg, pp. 284-294, 2001

[21] C. BenAbdelkader, R. G. Cutler, and L. S. Davis. "Gait recognition using image self-similarity", EURASIP Journal on Applied Signal Processing, 2004(4), 2004.

[22] I.N. Junejo , E. Dexter, I. Laptev and P. Pérez, "Cross-view action recognition from temporal self-similarities", Springer Berlin Heidelberg, pp. 293-306, 2008.

[23] J. Foote, "Visualizing music and audio using self-similarity", In Proceedings of the seventh ACM international conference on Multimedia (Part 1), pp. 77-80, 1999.

[24] S. Kumar, "Neural networks: a classroom approach. Tata McGraw-Hill Education", (2004)