Title	MMSE Analysis for Generated Robot Motion Language with ALBERT
Author(s)	Sato, Koki; Ishima, Fumiya; Kobayashi, Ryo et al.
Citation	The 11th International Symposium on Adaptive Motion of Animals and Machines (AMAM2023). 2023, p. 89-90
Version Type	VoR
URL	https://doi.org/10.18910/92281
rights	
Note	

MMSE Analysis for Generated Robot Motion Language with ALBERT

Koki Sato [1], Fumiya Ishima [2], Ryo Kobayashi [2], Atori Ikeyama [2], Sho Yamauchi [2], Keiji Suzuki [2],

[1]Graduate School of Systems Information Science, Future University Hakodate, Japan

g2122031@fun.ac.jp

[2]School of Systems Information Science, Future University Hakodate, Japan

1 Introduction

In order to create a robot with a variety of appealing features, factors related to shape and behavior are important. This can be seen in the development process of LOVOT, a home robot, where the robot's behavior is designed to match its shape[1]. With the increase in the number of robots operating in the home, there is a need to create robots with a variety of appealing features that will attract human attention.

In the robot geometry element, Yamauchi et al. proposed a design method for robots that prioritizes the appearance of the robot shape [2]. This method enables the automatic design of robots in engineering.

Robot behavior elements require behavior that is easy for humans to understand. To achieve this, studies have been reported using a method called the animation principle [3]. In addition, there are studies that have applied the expression method of ballet to the behavior of robots using mathematical methods [4]. From these studies, knowledge already exists on how to represent robot behavior to humans in a way that is easy to understand. However, these robot behaviors have limited patterns of expressible behavior.

Therefore, the creation of robots with diverse appeal requires a system that automatically and continuously generates diverse and continuous actions that attract human attention and interest. The purpose of this paper is to analyze whether the data generated by the proposed system is a variety of actions.

2 Related Work

A related study of continuous motion is the reinforcement learning method proposed by Peng et al [5]. In this method, the behavior of a dog is reproduced in a simulation using motion capture, and this reproduced dog behavior is imitated with reinforcement learning to enable the robot to perform the agile and continuous behavior of a dog. However, the behavior is a continuation of a single dog behavior pattern. Therefore, a variety of behaviors is not possible.

In this study, a variety of motions on the robot, which have not been done in previous studies, realize by using our system.

3 MMSE Analysis for Generated Robot Motion Language with ALBERT

3.1 Overview of the Proposed System

The automatic robot motion generation system using ALBERT in this research is shown in Figure 1 [6]. The system consists of two types of learning: learning in ALBERT and learning on the robot. First, cat videos are collected from YouTube. In this study, cats are used as samples of different behaviors. 2D key points are extracted from the collected videos. The 2D key point data is converted to 3D key points. The 3D key point data is then converted into data consisting of joints and their relative angles. This data is called motion language in this paper. By converting the data into motion language, ALBERT (A Lite BERT), a natural language processing model, is used for pre-training and fine-tuning [7]. ALBERT is used for this purpose. In addition, the robot learns the actions of the robot by using ALBERT which has been trained and adapted to the joints of the robot.

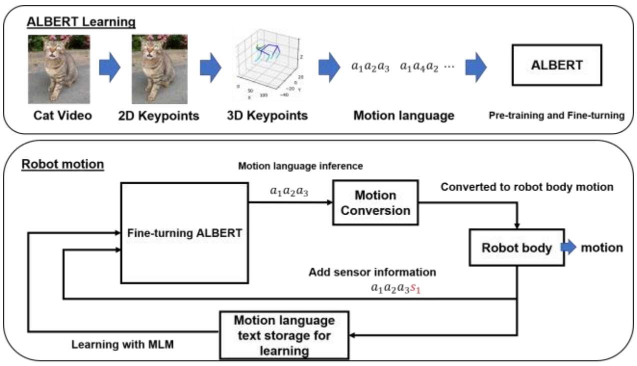

Figure 1: Automatic generation system for robot motion

3.2 Implementation

For 2D key point extraction, a method called AnimalPose was used [8]. In this method, an object is detected by object detection, followed by a key point extraction of the object. The 2D key points of 20 joints of the cat are extracted.

For the conversion to 3D key points, a method called C3DPO was used [9]. This method uses two neural networks to convert 2D key points into 3D key points. The method is also suitable for cats, which are non-rigid models.

The motion language is defined by each joint and its relative angle. Table 1 shows an overview of the motion language. Figure 2 shows the position of each joint in the cat. The joints are the head, neck, elbow, and knee. The relative angles are calculated using inverse kinematics on the position of each joint output by C3DPO. This angle is discretized into 10 equal parts of 360° from 0 to 9.

For pre-training ALBERT, data generated by uniform random numbers for each joint and relative angle were used. This allows ALBERT to acquire the form of a motion language.

For fine-tuning, data converted from the video is used. This allows ALBERT to capture the cat's behavior and generate a variety of motions. An example output is shown in Figure 3.

Table 1 Overview of motion language

Alphabet (from a to t)	Roll, pitch, and yaw of each joint
Figures (from 0 to 9)	Position of joint angle
*	Beginning of sequence
#	End of sequence

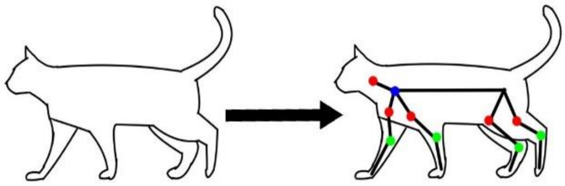

Figure 2 : Cat joint degrees of freedom (Number of degrees of freedom: red point 3, blue point 2, green point 1)

4 Motion Analysis with MMSE

The data generated by ALBERT are analyzed to determine if they are diverse in behavior. As an analysis method, Multivariate Multiscale Entropy (MMSE), an index that measures the degree of randomness on multiple time scales, is used to evaluate the data [10]. Using this method, we analyze whether the data generated by ALBERT is a variety of motion by considering a variety of motion as "motion complexity".

Three types of data are used for comparison in MMSE: random data generated by uniform random numbers, ALBERT-generated data, and cat transformation data from video. The number of channels was set to 4 to match each joint of the cat, and other parameters were implemented with reference to the proposed paper on MMSE [10].

The results of the MMSE experiment are shown in Figure 4. Each scale factor is analyzed separately. For scale factors 1-5, there is little difference between the three types of data. For scale factors 6-20, the ALBERT-generated data has a small difference from the data converted from the video, and a large difference from the random data. For scale factors 21-50, the ALBERT-generated data has a small difference from the random data and a large difference from the data converted from the video.

From the above, it can be seen that for scale factors 1-20, the complexity of the ALBERT-generated data and the data converted from the video are close, so a wide variety of motions can be generated. However, for scale factors 21-50, the complexity of the ALBERT-generated data is lower than that of the data converted from the video, so there are challenges in generating a variety of motions.

5 Conclusion and Future Work

In this study, we analyzed whether the output by ALBERT, trained on data converted from cat videos to motion language, was a variety of motion. As a validation method, the diverse motions were considered as motion complexity and MMSE was used. The results show that diverse motions can be generated when the scale factor is between 1 and 20.

In the future, we will work to improve the generation of a variety of motions on the long-term scale identified in this study. One candidate cause is that ALBERT has an input length of 512 for training, which may affect the motion language generation. In order to clarify the cause of this problem, we will visualize the motions generated with ALBERT to analyze the patterns of motions on a long-term scale and improve our system.

```
*a6b8c2d1e1f1g1h1j1m9n7o2p6q6r6s6t6u6#
*a9b1c9d5e9f9g4h4i4j5k6l8m6n6o6p0q5r0#
*a1b0c0d1e0f4g9h0i3j1k1l2m4n4o9p1q1u0#
*a6b9c6d4e4f9g4h5i6j9k6l0m6n6o3p3q0#
*a3b0c0d1e9f4g9h9i9j5k5l8m5n4o6p6q3r6s3u3#
*c1d3e4f0g0h4i5k6l5m0n0o5q6r0s6t2u0#
*a0b4c3d0e9g9h0i4j6k0l6n6o3p3q0r6#
*a3b0c1d4e4f4g4h4i0j4k5l8m6o6p4q3r0#
*a6b3c0d1e3f0g9h0i3j1k9l7m2n4o3p3q0#
*a9b6c3d4e4f4g4h4i4j0k4l5m4n5o8p4q4r6#
.
.
.
```

Figure 3: Example of ALBERT output (One line of the motion language shows the joint positions of one frame of video.)

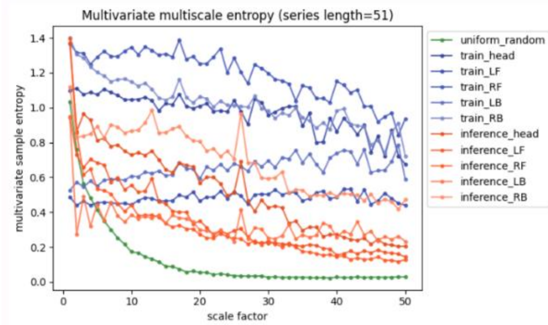

Figure 4 : MMSE results

References

[1] Naoto Yoshida, Shuto Yonemura, Masahiro Emoto, Kanji Kawai, Naoki Numaguchi, Hiroki Nakazato, Shunsuke Otsubo, Megumi Takada, Kaname Hayashi, "Production of Character Animation in a Home Robot: A Case Study of LOVOT" International Journal of Social Robotics, Vol.14, pp.39-54, 2022

[2] Sho Yamauchi, Keiji Suzuki, "Robot Design Method Using 3D Printed Inner Crystal Structure" IEEJ Transactions on Electronics, Information and Systems, Vol.139, No.9, pp.1051-1958

[3] A.J.N.van Breemen, "Bringing Robots To Life: Applying Principles Of Animation To Robots", https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0a568ccb6027f5dd160a1a89889d35e2db0beb06, Feb.13, 2023

[4] A.Bacula, A.LaViers, "Character Synthesis of Ballet Archetypes on Robots Using Laban Movement Analysis: Comparison Between a Humanoid and an Aerial Robot Platform with Lay and Expert Observation", Vol.13, pp1047-1062

[5] Xue Bin Peng, Erwin Coumans, Tingnan Zhang, Tsang-Wei Lee, Jie Tan, Sergey Levine, "L-earning Agile Robotic Locomotion Skills by Imitating Animals", Robotics: Science and Systems, Corvallis, Oregon, USA, Jul.12-16, 2020

[6] Koki Sato, Fumiya Ishima, Ryo Kobayashi, Atori Ikeyama, Sho Yamauchi, Keiji Suzuki, "Automatic Generation of Characteristic Movements for Robots Using ALBERT" Joint convention, The Hokkaido chapters of the institutes of electrical and information engineers, Japan, pp.201-202, Nov.5-6, 2022

[7] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations", Published as a conference paper at ICLR, 2020

[8] AnimalPose document, https://mmpose.readthedocs.io/en/latest/model_zoo/animal_2d_keypoint.html#animalpose-dataset, Feb.13, 2023

[9] David Novotny, Nikhila Ravi, Benjamin Graham, Natalia Neverova, Andrea Vedaldi, "C3DPO: Canonical 3D Pose Networks for Non-Rigid Structure From Motion", https://arxiv.org/pdf/1909.02533.pdf, Retrieved, Feb13.2023

[10] Mosabber Uddin Ahmed, Danilo P Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data" Physical Review E, Vol. 84, No. 6, p. 061918, 2011