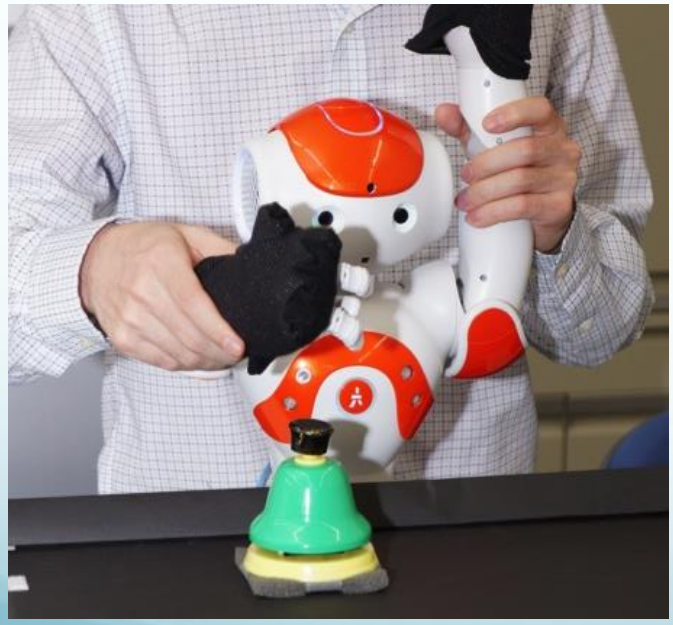
Multimodal Integration Learning of Object Manipulation Behaviors using Deep Neural Networks



Kuniaki Noda, Hiroaki Arie, Yuki Suga and Tetsuya Ogata

Waseda University

November 4, 2013 17:30-17:45
IROS2013 @ Tokyo, JAPAN

Background

- Robot control under open-ended environment
  - Noise robust environment recognition
  - Adaptive behavior control

Real-time large-scale sensory-motor information processing is essential

- Deep learning
  - Trained with large scale data
  - Higher-order representation is self-organized
  - Breakthrough in machine learning
    - Large scale visual recognition challenge (ILSVRC2012)

(Google official blog, 2012)

Applications for robots have yet to be investigated

MIT Technology Review 10 BREAKTHROUGH TECHNOLOGIES 2013

Research objective

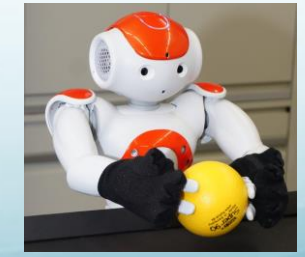
Sensory-motor integration of robot behaviors
utilizing deep neural network

Large scale real world data processing

Crossmodal memory retrieval and temporal sequence prediction

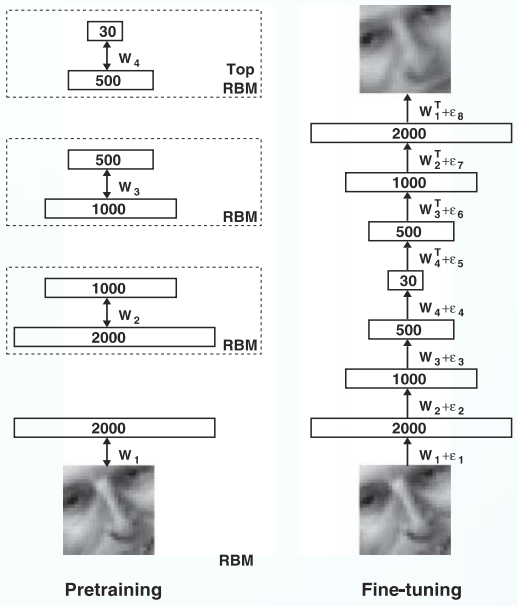
Adaptive behavior control of a humanoid robot
depending on the environmental changes

Verification experiment
by object manipulation task

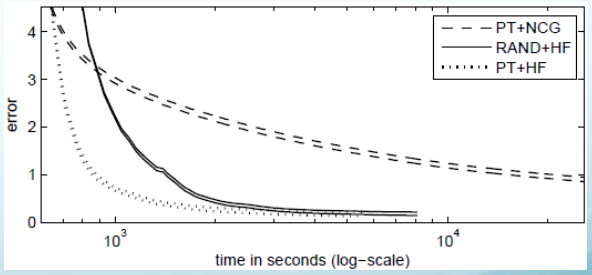


Deep learning

- *G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, 2006.*
  - Epoch-making article which leads to the current trends for the deep learning
  - Utilize RBM for training single layer network in the pre-training phase, followed by the entire layer training in the fine-tuning phase
- *J. Martens, "Deep learning via Hessian-free optimization," ICML, 2010.*
  - Utilize quadratic programming
  - Pre-training is not required
  - Optimization algorithm based on the Newton's method contributes in faster convergence



(Hinton, 2006)



(Martens, 2010)

We adopt Hessian-free optimization as the training algorithm

Multimodal integration mechanism

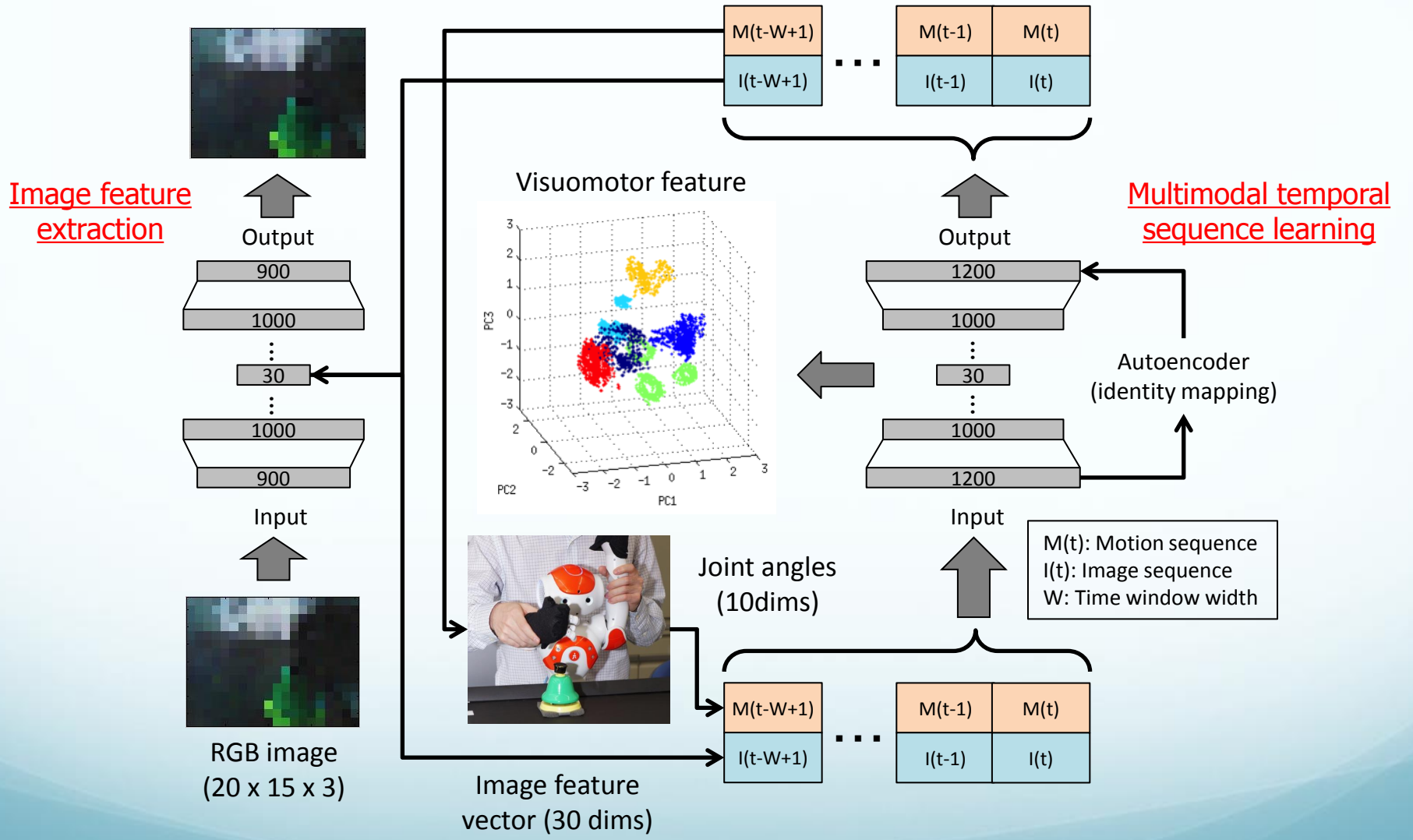
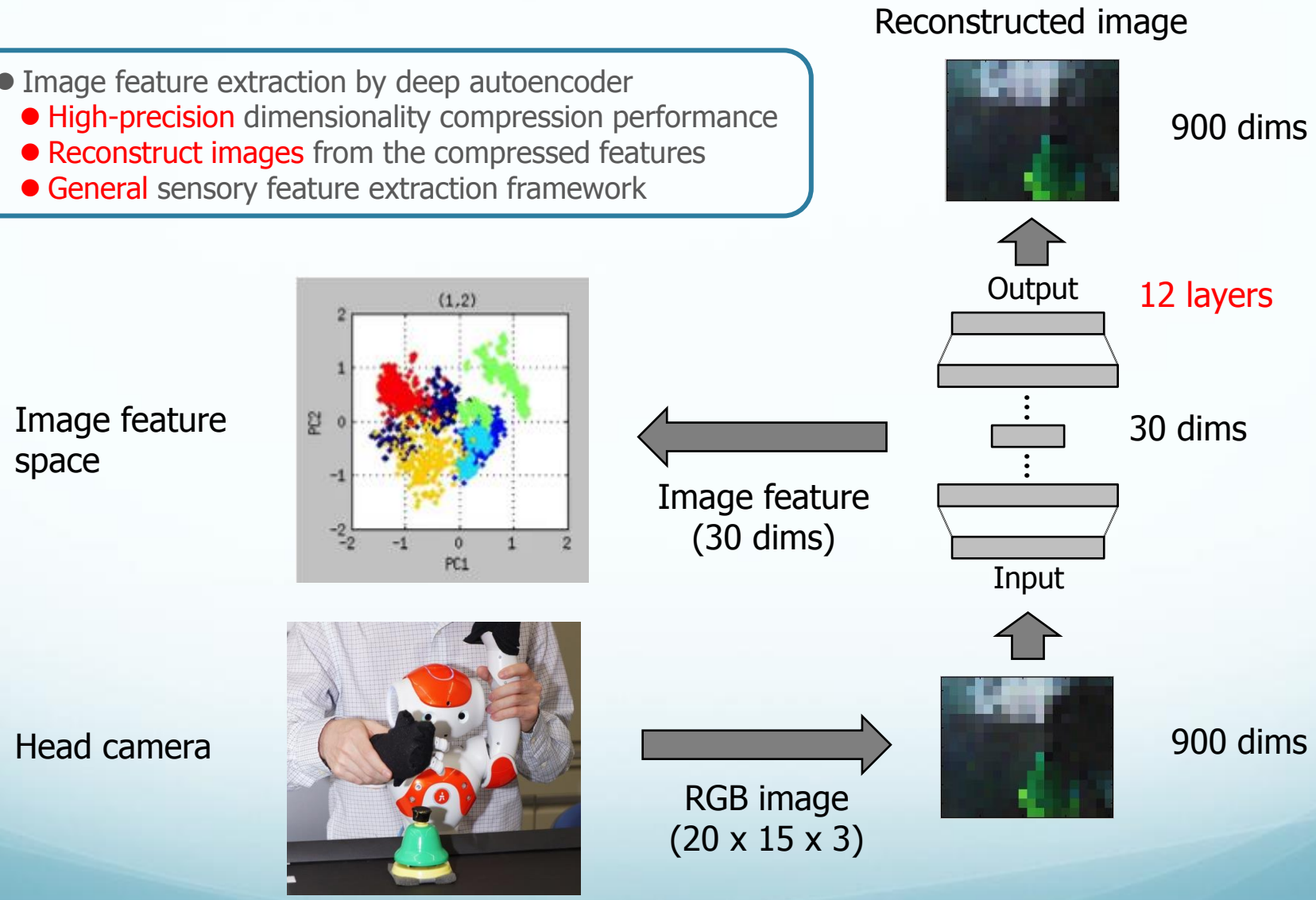


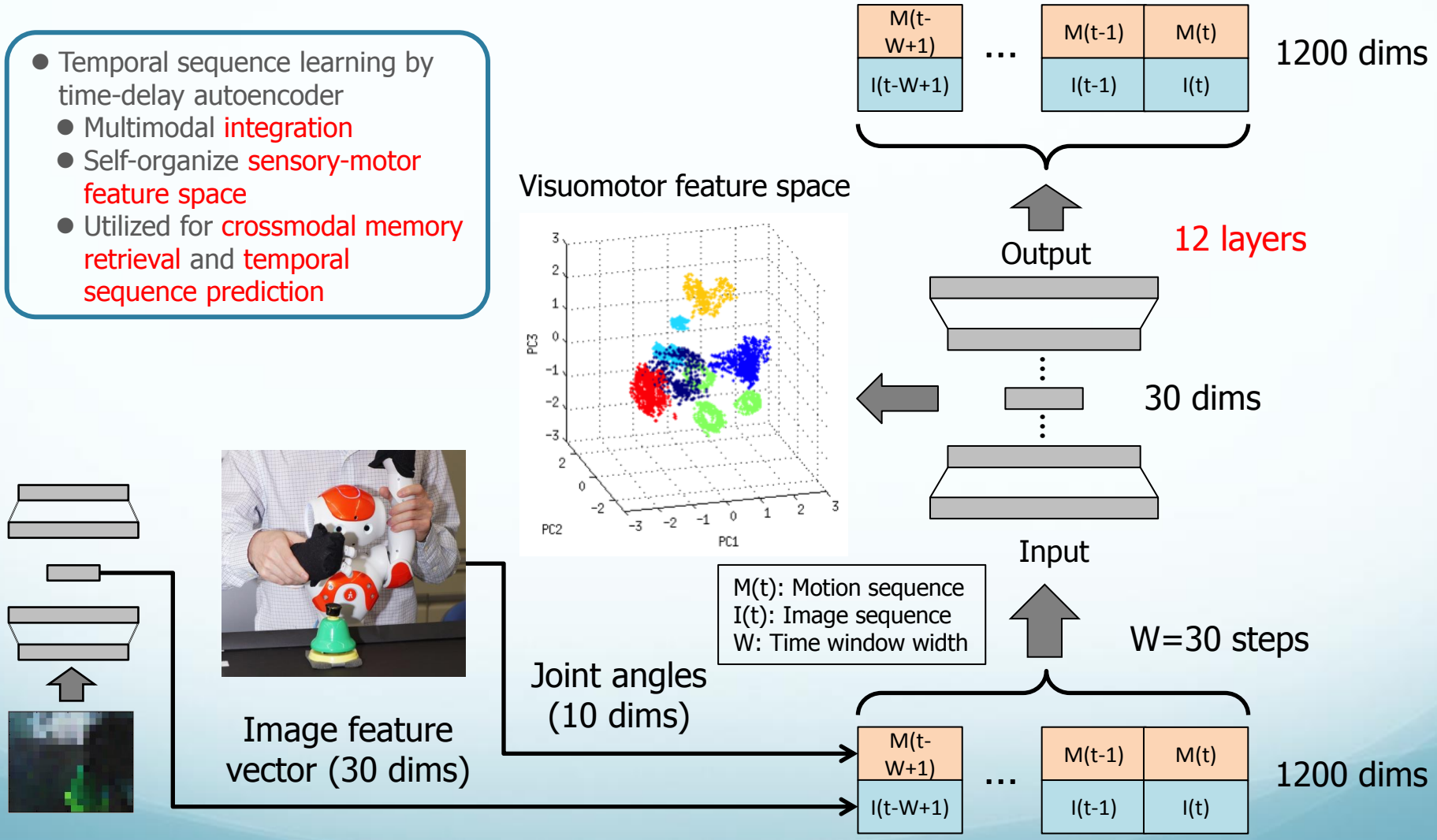
Image feature extraction network

- Image feature extraction by deep autoencoder
  - High-precision dimensionality compression performance
  - Reconstruct images from the compressed features
  - General sensory feature extraction framework

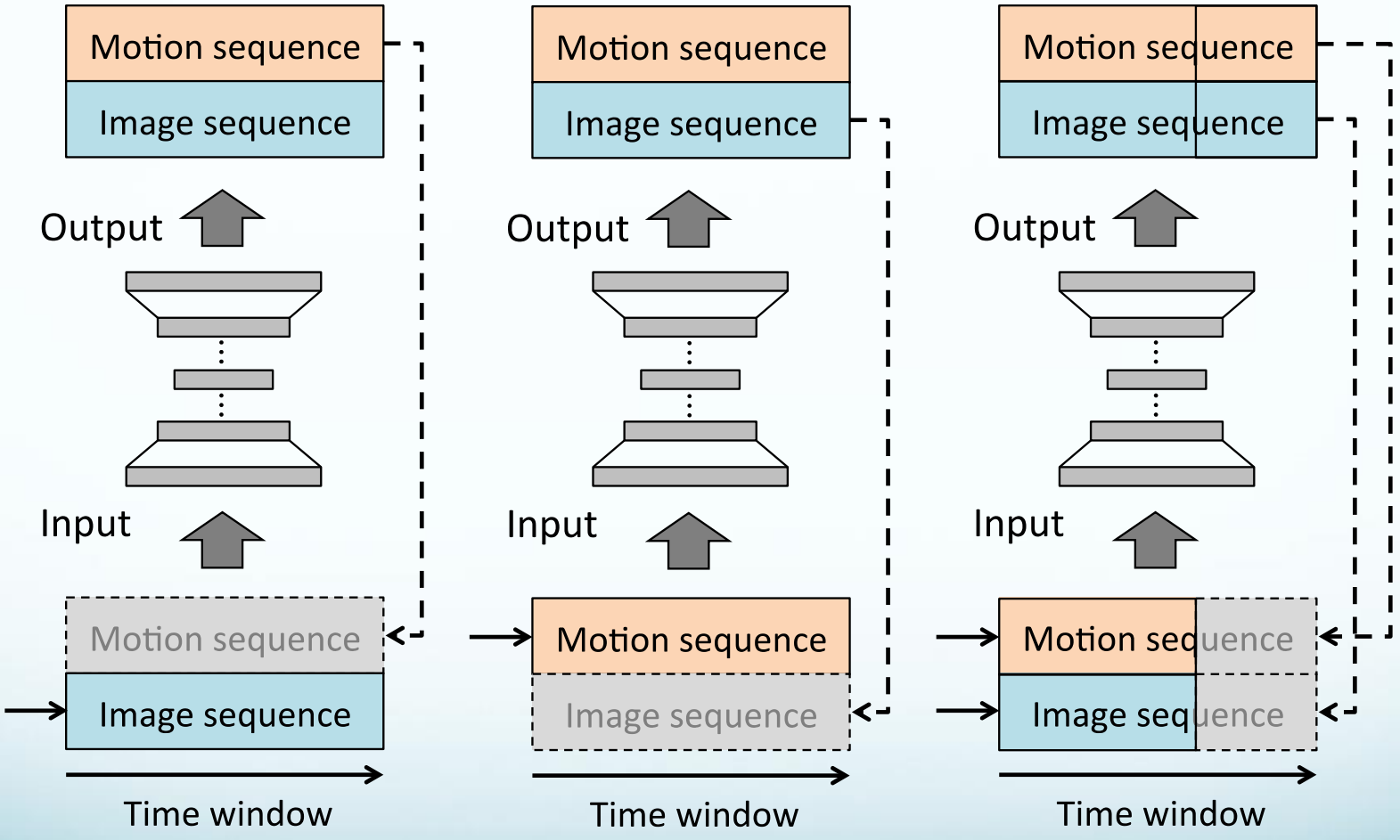


Multimodal temporal sequence learning network

- Temporal sequence learning by time-delay autoencoder
  - Multimodal integration
  - Self-organize sensory-motor feature space
  - Utilized for crossmodal memory retrieval and temporal sequence prediction

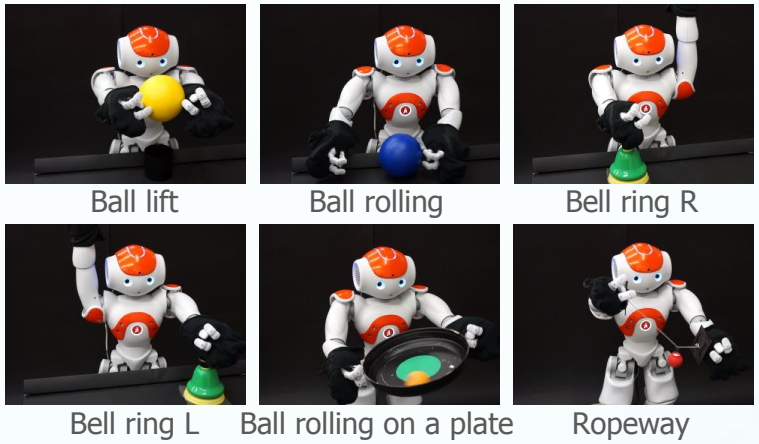


Crossmodal memory retrieval and temporal sequence prediction



Evaluation experiment

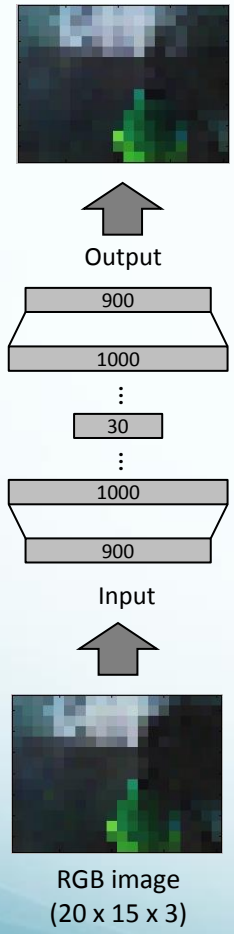
- Sensory-motor integration learning of object manipulation behaviors
  - 6 object manipulation behaviors
- Sensory-motor data
  - 20x15 RGB image: 900 dims
  - Arm joint angles: 10 DOF
  - Time window: 30 steps



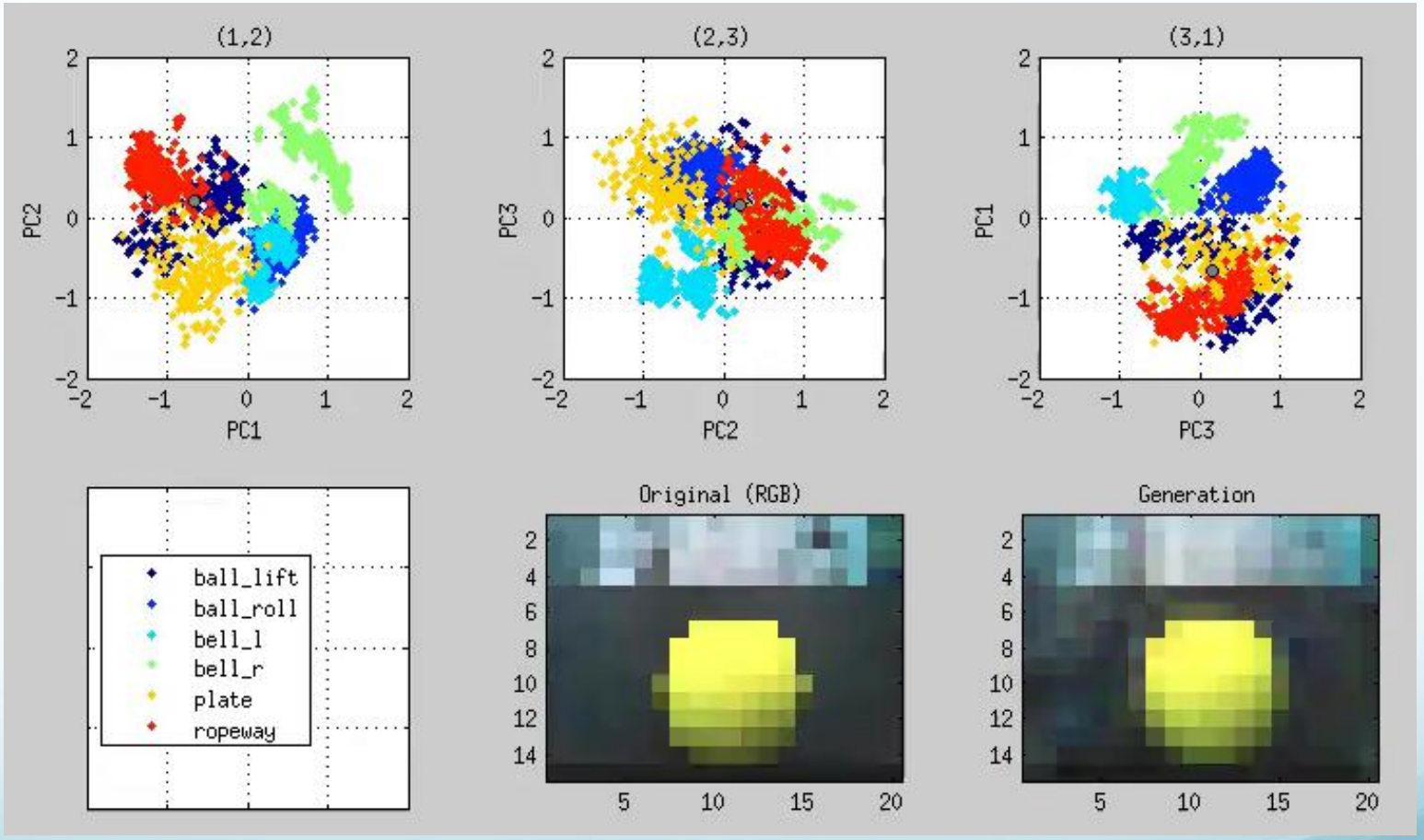
	Training	Test	I/O dim.	Network structure
Image feat.	8444	948	900	1000-500-250-150-80-30 -80-150-250-500-1000
Temp. seq.	6848	776	1200	1000-500-250-150-80-30 -80-150-250-500-1000

- Optimization utilizing GPGPU (CUBLAS)
  - 30 min. each for the feature extraction and the temporal sequence learning

Image feature space and image reconstruction



PC1-PC2 PC2-PC3 PC3-PC1



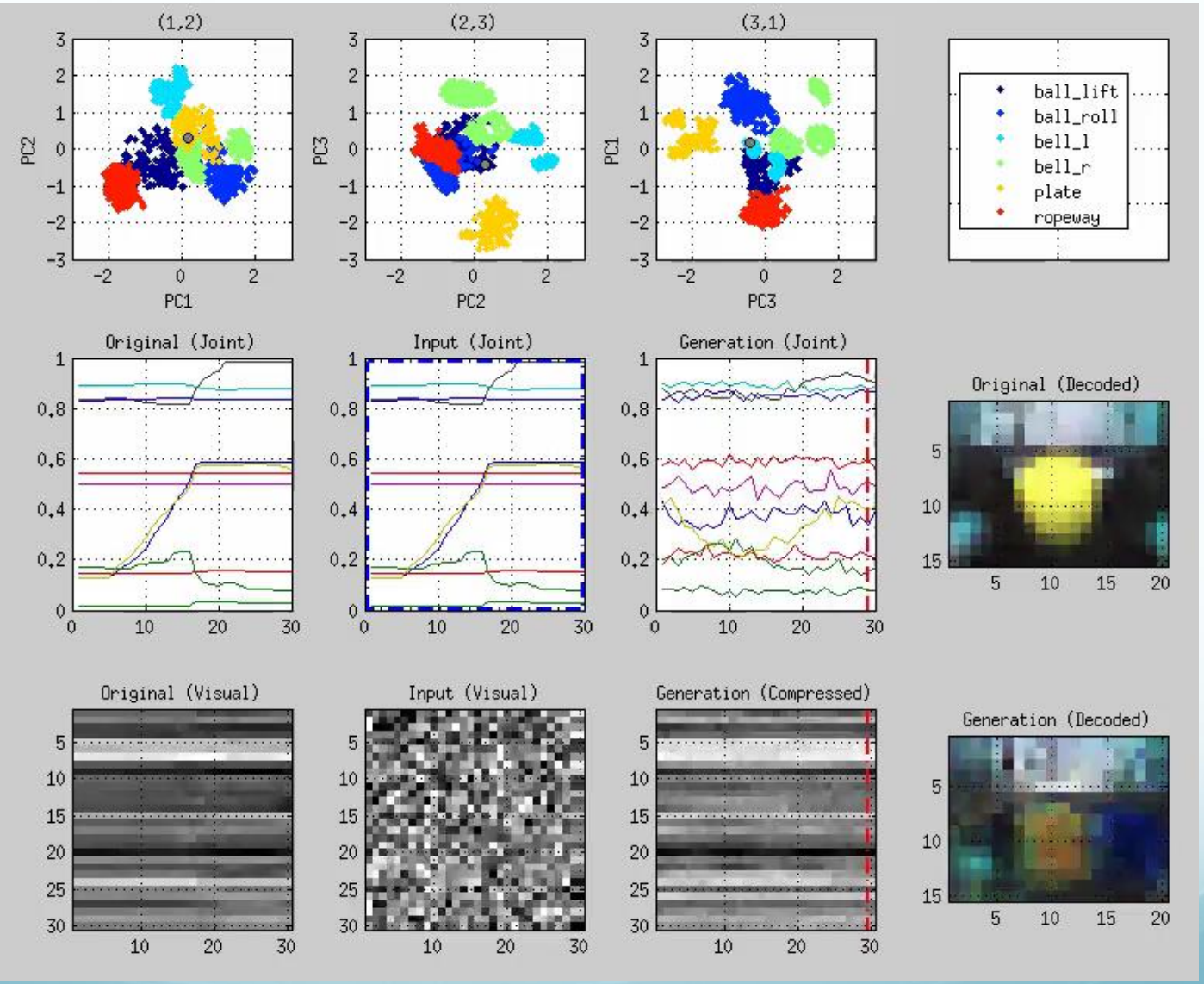
INPUT IMAGE RECONSTRUCTED IMAGE

Image retrieval from motion

Multimodal feature space

Joint angles sequence

Image feature sequence



ORIGINAL IMAGE

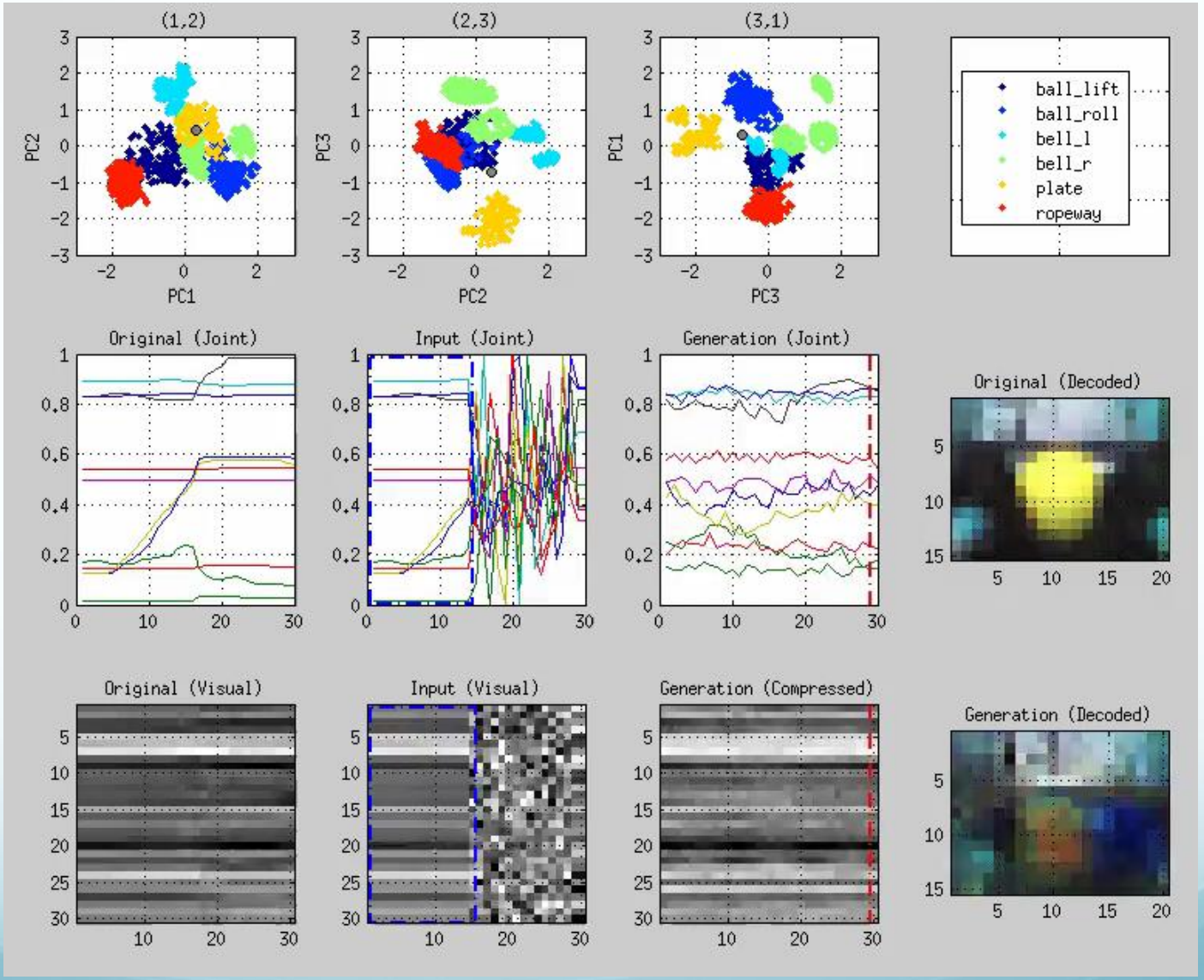
RETREIVED IMAGE

Temporal sequence prediction

Multimodal feature space

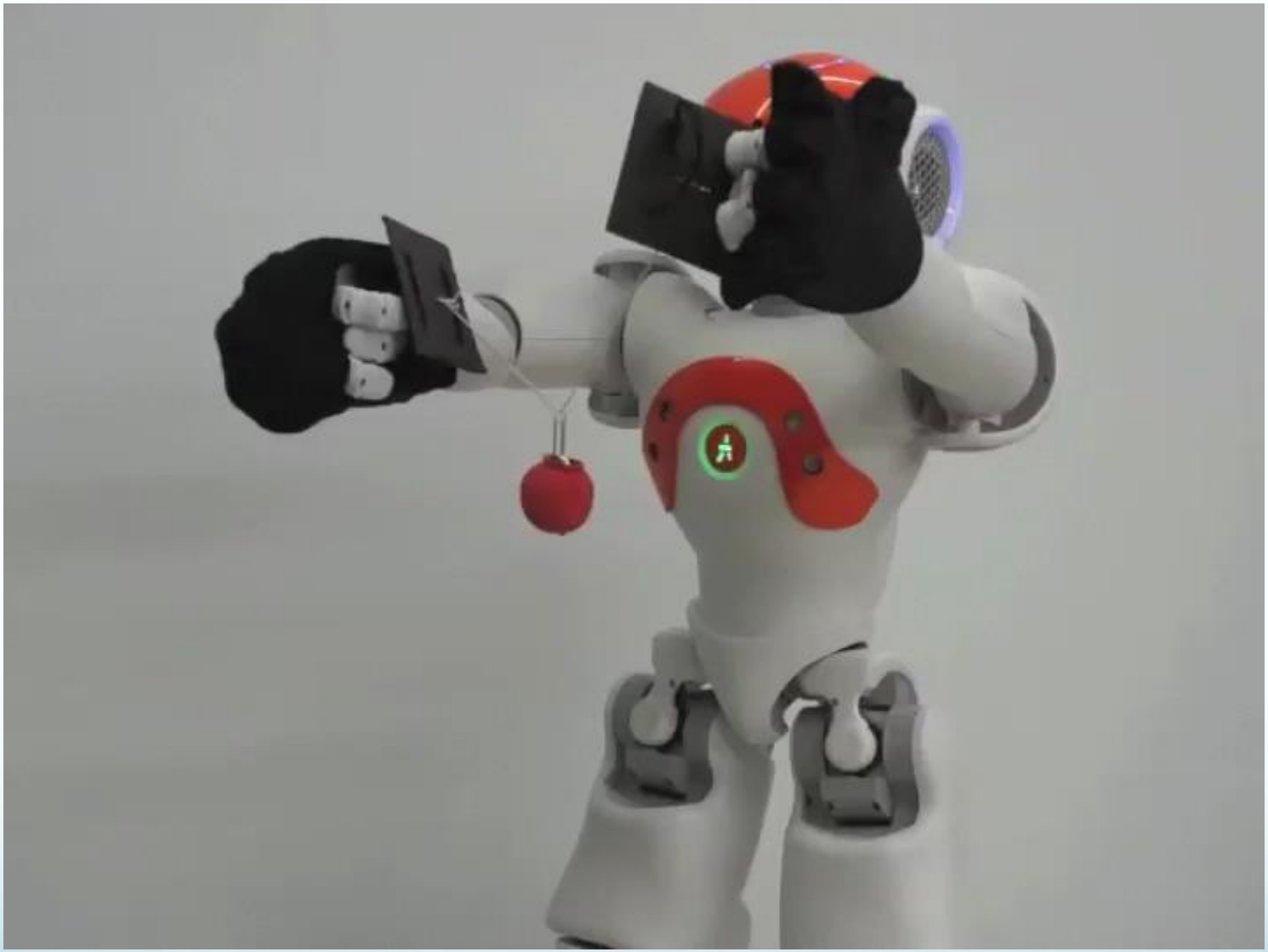
Joint angles sequence

Image feature sequence



ORIGINAL IMAGE

RETREIVED IMAGE



Ropeway → Bell ring R → Bell ring L → Bell ring R

Conclusion and future work

- Multimodal integration learning of robot behaviors utilizing deep neural networks
  - Large scale real-world sensory-motor information processing
  - Crossmodal memory retrieval and temporal sequence prediction
  - Adaptive robot behavior control regarding environmental changes

- Robust image recognition
  - Increase variations of the environment lighting condition
  - Local feature extraction networks (e.g. Convolution network)
- Analysis on the internal structure of the networks
  - Relationship between the network structure and the learning capability

Thank you!

The work has been supported by JST PRESTO "Information Environment and Humans" and MEXT Grant-in-Aid for Scientific Research on Innovative Areas "Constructive Developmental Science" (24119003).