

Chapter 43

Feature Extraction of Video Using Artificial Neural Network

Yoshihiro Hayakawa

National College of Technology, Japan

Takanori Oonuma

National College of Technology, Japan

Hideyuki Kobayashi

National College of Technology, Japan

Akiko Takahashi

National College of Technology, Japan

Shinji Chiba

National College of Technology, Japan

Nahomi M Fujiki

National College of Technology, Japan

ABSTRACT

In deep neural networks, which have been gaining attention in recent years, the features of input images are expressed in a middle layer. Using the information on this feature layer, high performance can be demonstrated in the image recognition field. In the present study, we achieve image recognition, without using convolutional neural networks or sparse coding, through an image feature extraction function obtained when identity mapping learning is applied to sandglass-style feed-forward neural networks. In sports form analysis, for example, a state trajectory is mapped in a low-dimensional feature space based on a consecutive series of actions. Here, we discuss ideas related to image analysis by applying the above method.

INTRODUCTION

According to the 2015 Telecommunications White Paper from the Japanese Ministry of Internal Affairs and Communications, the number of “things” related to the Internet of Things (IoT) is expected to increase tremendously; a wide variety of data will be collected and analyzed via IoT and movements to link these are expected to enhance work efficiency, among other things. Naturally, this includes the treatment of big data, and when analyzing data collected in such large quantities, their analysis could be performed interactively, allowing interpretations from many different angles. In particular, security camera images, while considering privacy, are linked to a large volume of data collected using IoT. The reason for this is

DOI: 10.4018/978-1-7998-0414-7.ch043

exceptionally simple. For example, to monitor the daily behavioral patterns of elderly people and children, it is convenient to construct surveillance systems using data collected from IoT because it is portable and allows us to analyze and extract data in order to evaluate whether their behaviors have changed or are in danger of doing so. There are the researches in the computer network for the realization of such a system (Shintaro, Mariko, Mingrui, Yoshikazu, & Toshimitsu, 2013). However, with this simple idea, the main point is how to construct a system in which the behavior pattern data collected from IoT can be linked to actual behavior. Although the search engine to connect sensor data with an event has been already proposed (Koji, Yutaka, Takuya, Yasue, Yasushi, & Takeshi, 2010; Takeshi, Yasue, Takuya, Koji, Yutaka, & Yasushi, 2010), we believe that if a video is captured at the same time as the IoT data are collected, specification of these data is a simple matter.

In relation to the treatment of images, deep learning using deep neural networks (DNNs) has attracted a lot of attention in recent years because it successfully demonstrated automatic organization of neurons selectively reacting to images of a cat (Quoc, 2013) and high recognition performance in relation to difficult tasks, such as general object recognition. In other words, this success engendered several methods to mitigate model complexity including weight decay (Stephen, & Lorient, 1989) and weight sharing (Le, Bernhard, John, Donnie, Richard, Wayne, & Lawrence, 1990; Steven, & Geoffrey, 1992) and to roughly form DNN at an early stage through prior learning (Geoffrey, Simon, & Yee-Whye, 2006). Furthermore, by creating an ensemble of multiple neural networks, such as DropOut (Geoffrey, Nitish, Alex, Ilya, & Ruslan, 2012) or DropConnect (Li, Matthew, Sixin, Yann, & Rob, 2013), which is a model assumed to have information deficiency in it, it has greatly contributed to the development of various parameter tuning for keeping generality.

Analyzing the action parameters from videos captured at the same time as data is collected from IoT, it is clear that although human eyes can specify actions easily by directly viewing them, due to issues such as their processing ability and fatigue, they soon encounter limitations when processing large quantities for long periods. Therefore, there is a demand for the automation of action pattern analysis using such images, and DNN can be used to achieve this.

In DNN image processing, if a convolutional neural network (CNN) (Alex, Ilya, & Geoffrey, 2012) or sparse coding is used, the response neurons corresponding to image input in the feature extraction layer learn to respond and become able to recognize objects. By utilizing this process on a video frame by frame, it is possible to consider a discrete feature vector corresponding to this video (Yasuyuki, Takashi, & Kuniaki, 2015). However, we have considered utilizing a continuous trajectory of an object or person moving consecutively. These trajectories are formed from a nonlinear mapping into a low-dimensional feature space by handling feed-forward neural networks without CNN or sparse coding.

In this paper, we handle form analysis of the table tennis as one example. In such a case, the series of movements from the backswing to hitting the ball and following through are important. Subsequently, if this video is provided as identity map learning data using sandglass-type feed-forward neural networks, the feature extraction function acquired in the middle layer would allow the movements to be expressed as a state trajectory of corresponding state as time proceeds. As a result, it is expected that nonlinear conversion and generality of the neural networks will work effectively in action pattern analysis. In this paper, we include the preliminary results of the data analysis and discuss the possibility of action analysis using DNN.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/feature-extraction-of-video-using-artificial-neural-network/237905

Related Content

Toward Automatic Answers in User-Interactive Question Answering Systems

Tianyong Hao, Feifei Xu, Jingsheng Lei, Liu Wenyin and Qing Li (2011). *International Journal of Software Science and Computational Intelligence* (pp. 52-66).

www.irma-international.org/article/toward-automatic-answers-user-interactive/64179

Measuring Textual Context Based on Cognitive Principles

Ning Fang, Xiangfeng Luo and Weimin Xu (2009). *International Journal of Software Science and Computational Intelligence* (pp. 61-89).

www.irma-international.org/article/measuring-textual-context-based-cognitive/37489

Combining Ontology with Intelligent Agent to Provide Negotiation Service

Qiumei Pu, Yongcun Cao, Xiuqin Pan, Siyao Fu and Zengguang Hou (2010). *International Journal of Software Science and Computational Intelligence* (pp. 52-61).

www.irma-international.org/article/combining-ontology-intelligent-agent-provide/46146

RA-CNN: A Semantic-Enhanced Method in a Multi-Semantic Environment

Zhiwei Zhan, Guoliang Liao, Xiang Ren, Guangsi Xiong, Weilin Zhou, Wenchao Jiang and Hong Xiao (2022). *International Journal of Software Science and Computational Intelligence* (pp. 1-14).

www.irma-international.org/article/ra-cnn/311446

Using Computational Modelling to Understand Cognition in the Ventral Visual-Perirhinal Pathway

Rosemary A. Cowell, Timothy J. Bussey and Lisa M. Saksida (2011). *Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications* (pp. 15-45).

www.irma-international.org/chapter/using-computational-modelling-understand-cognition/49228