Journal of Fiber Bioengineering and Informatics 8:4 (2015) 791–800
doi:10.3993/jfbim00196

Part Recognition Method Based on Visual Selective Attention Mechanism and Deep Learning ⋆

Dan Zhou, Nanfeng Xiao*

College of Computer Science and Engineering, South China University of Technology Guangzhou 510006, China

Abstract

In order to enable the industrial robots to recognize the specific targets quickly and accurately on the assembly line, an object recognition method driven by visual selective attention mechanism is proposed. With mass training data and a machine learning model containing a number of hidden layers, deep learning can learn more useful features, and thus ultimately improve the classification or the prediction accuracy. The main idea of this method is as follows: for all part images, the visual attention mechanism is used to choose salient regions in an image, achieving the goal of target segmentation. Then an image recognition method based on deep learning is applied to recognize the chosen salient regions. Experimental results show the effectiveness of the proposed method and the cognitive rationality.

Keywords: Visual Selective Attention Mechanism; Part Recognition; Deep Learning; Feature Learning

1 Introduction

Part recognition has become one of the most important applications of computer vision and pattern recognition in the field of mechanical industry. As the basis of automatic machining, it not only helps to liberate workers from heavy work, but also improves productivity and reduces cost. Part recognition firstly gained attention of researchers abroad, which led to many mature technologies. In recent years, researchers at home also have put forward a lot of methods for these problems. Kaihua Yan and Zhenwei Su et al. proposed a method that takes hu_moments of parts as feature vector for classification using SVM [1]; Huanhuan Quan and Luoping Zhang presented a method which uses color and geometry shape features, constituting both meaningful picture characteristics, as gist to recognize component parts [2]; Chunxiang He and Bo Liu put forward a new part recognition technology based on wavelet multi-scale edge detection and BP neural network [3]. However, all these methods ignore the fact that the research object is only part of the image. Therefore, there is no need to take the whole image as the inputs of the classifier.

⋆Project supported by the National Natural Science Foundation of China (No. 61171141, 61573145), the Special Fund for Public Welfare Research and Capacity Building of Guangdong Province (No. 2014B010104001) and the Basic and Applied Basic Research of Guangdong Province (No. 2015A030308018).

*Corresponding author.
Email address: xiaonf@aliyun.com (Nanfeng Xiao).

What is more, features in these methods are extracted manually, which is challenging to choose suitable features for recognition tasks of different parts.

It is well known that visual selective attention mechanism is the essential feature of the primate processing external environment information, and it is also the key technology of human choosing a specific region of interest from a large number of external information. In computer image processing, what the task focus on are usually only research objects, therefore it is very important to find the areas that can easily attract the observer's attention. The targets interested can be located to reduce the search space via the saliency computation model of the visual selective attention mechanism. Visual feature extraction is a key step of image classification or recognition, and thus the quality of features can directly affect the recognition results. It is not a scalable way relying on artificial experience to choose features for image recognition problems. The essence of deep learning is that, with mass training data and a machine learning model containing a number of hidden layers, the algorithm can learn more useful features, thus ultimately improve the classification or the prediction accuracy. Therefore, using the deep learning-based image recognition method to understand and recognize the chosen salient regions can not only avoid time consuming of the manual feature extraction, but also greatly improve the recognition accuracy.

The remainder of this paper is organized as follows: Firstly, a part recognition method based on visual selective attention mechanism and deep learning is proposed for the part recognition application of the industrial robots; Secondly, the visual selective attention mechanism and its saliency computation model are introduced; Thirdly, the unsupervised feature learning and part image recognition method based on deep learning are studied; Lastly, the part recognition algorithm proposed in this paper is verified by experiments.

2 Visual Selective Attention Mechanism

This paper adopts the saliency computation model [4–7] based on the low-level visual features of an image. First of all, color, luminance and orientation features are decomposed from the input image by a multi-scale filter. Then the center-surround difference method is used to generate the corresponding conspicuity maps from the different feature information. And finally the multi feature map integration strategy [8] is applied to obtain the saliency map from the previous conspicuity maps.

2.1 Early Visual Feature Extraction

Each feature is computed by a set of linear "center-surround" operations akin to visual receptive fields. For an color image, nine spatial scales are created using dyadic Gaussian pyramids, which progressively low-pass filter and subsample the input image, yielding horizontal and vertical image-reduction factors ranging from 1:1 (scale zero) to 1:256 (scale eight) in eight octaves. The center-surround is implemented in the model as the difference between fine and coarse scales: the center is a pixel at scale $c \in [2, 3, 4]$, and the surround is the corresponding pixel at scale $s = c + \delta$, with $\delta \in \{3, 4\}$. The across-scale difference between two maps, denoted Θ below, is obtained by interpolation to the fine scale and point-by-point subtraction. Using several scales not only for c but also for $\delta = s - c$ yields truly multi-scale feature extraction, by including different size ratios between the center and surround regions.