Robustness Evaluation of Extracting Features Based on Self-organizing Map Neural Network

Binquan Li\textsuperscript{a}, Xiaohui Hu\textsuperscript{b}

Beijing University of Aeronautics and Astronautics, Beijing 100191, China
\textsuperscript{a}jz05022300@sina.com, \textsuperscript{b}binquanli@buaa.edu.cn

Keywords: robustness, extracting features, image changing, CNN, SOM neural network

Abstract. The descriptors should be robust to changes of images taken under various conditions in order to obtain correct recognition. HOG, LBP and convolutional neural network (CNN) are proved to perform better in extracting features of images. However, little attention was paid to the robustness of these methods. In this paper, we proposed a self-organizing map (SOM) neural network based descriptor evaluation method in order to assess the robust image features. It efficiently observes the features extracted by HOG, LBP and CNN with deep learning. The results show that CNN performs better and is more robust among the method of extracting features.

Introduction

Feature extracting for images are important in applications of pattern recognition, such as object recognition of images taken by unmanned aerial vehicle (UAV) and satellite utilizing remote sensing. Therefore, the feature extracting method should be robust to the changes of images obtained under different conditions in order to recognize the object correctly.

HOG, LBP and CNN have better performance in extracting features of images. HOG descriptor was extended for use in sketch based image retrieval (SBIR). The GF-HOG adaptation was shown to outperform existing gradient histogram descriptors [1]. LBP has been found to be a powerful feature for texture classification. When LBP is combined with the HOG descriptor, it improves the detection performance on some datasets [2-3]. Convolutional neural networks (CNNs) constitute models whose capacity can be controlled by varying their depth and breadth, which make strong and mostly correct assumptions about the nature of images [4-10]. However, little attention was paid to the robustness of image changing by the methods above.

The changes of images always exist in the real world that contain viewpoint change, light change, zoom and rotation, and image blur. It is valuable and important to evaluate the robustness of these feature extraction methods against image changes. In this paper, SOM neural network is proposed to evaluate the robustness of feature extraction by classifying the extracted feature vectors. We also design a frame of CNN and optimize the parameters in order to implement the feature extracting. Compared with HOG and LBP, the CNN shows the best performance and robustness.
Experimental Setup

**Data Set.** Fig. 1 shows example images of our data set used for the evaluation. Four image transformations are evaluated: viewpoint change (Fig. 1a); illumination (Fig. 1b); zoom and rotation (Fig. 1c); image blur (Fig. 1d).

![Example images of the data set](image)

(a) (b)

(c) (d)

Figure 1. Data set. Examples of images used for the evaluation: (a) viewpoint change, (b) light change, (c) zoom and rotation, and (d) image blur.

**HOG, LBP and CNN.** We convert the features extracted by the three methods into image, in order to compare the efficiency of them. We convert the vector to figures after extracting the features of each image. The cell size of HOG and LBP is selected as 2. The architecture of CNN is constructed as follows. CNN compose unsupervised learning method and classifier into one frame. We only utilize the unsupervised learning approach to extract the features.

Experimental Results

In this section, we present and discuss the experimental results of the evaluation. The performance is compared for viewpoint changes, illumination, zoom and rotation and image blur. **Features extracted by HOG, LBP and CNN.** We design a CNN with the $5c$-$2s$-$10c$-$2s$ architecture, where $5c$ represents convolution layer with feature maps as 5, $2s$ represents down sampling layer with scale as 2. We utilize the images as the input data. Finally, we concatenate all end layer feature maps into feature vector. The parameters of CNN are shown in Table 1.

<table>
<thead>
<tr>
<th>layer</th>
<th>Definition</th>
<th>Feature maps</th>
<th>Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Image input</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>Convolutional layer C1</td>
<td>6</td>
<td>5×5</td>
</tr>
<tr>
<td>2</td>
<td>Pooling layer S1</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Convolutional layer C2</td>
<td>12</td>
<td>10×10</td>
</tr>
<tr>
<td>4</td>
<td>Pooling layer S2</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1. Architecture and construction of CNN.
**SOM neural network.** We utilized self-organizing map (SOM) neural network to evaluate the robustness of feature extracted by HOG, LBP and CNN.

SOMs operate in two modes called training and mapping that is similar as most artificial neural networks. "Training" builds the map using input samples while "mapping" classifies a new input feature.

The feature vectors of images in dataset are extracted by HOG, LBP and CNN. We take feature vectors as the input samples of SOM neural network. The SOM can distinguish each feature vector after certain of training iterations. If the training iterations of feature vectors by a descriptor are more than another descriptor, it is concluded that the feature vectors is so similar that it is hard to distinguish. In other words, this descriptor is more robust than the other.

The training of SOM utilizes competitive learning. When a feature vector $X$ is fed to the network, its Euclidean distance to all weight vectors is calculated as follows:

$$d_j = \|X - W_j\| = \sqrt{\sum_{i=1}^{m}(x_i(t) - w_{ij}(t))^2}$$  \hspace{1cm} (1)

Where, $X$ is the feature vector extracted by different descriptors, $W_j$ represents the neuron $j$ of the competing layer, $m$ is the dimension of feature vector, $w_{ij}$ is the weight of neuron $i$ in the input layer and neuron $j$ in the competing layer. The neuron whose weight is most similar to the input feature vector is called the best matching unit (BMU). The update formula for a neuron with weight vector is described as follows:

$$\Delta w_{ij} = w_{ij}(t + 1) - w_{ij}(t)$$  \hspace{1cm} (2)

The algorithm is summarized in Table 2.

Table 2. Robustness evaluation by SOM.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Randomize the weight vectors.</td>
</tr>
<tr>
<td>2.</td>
<td>Get an input vector extracted by different descriptors.</td>
</tr>
<tr>
<td>3.</td>
<td>Traverse each node of the map</td>
</tr>
<tr>
<td></td>
<td>(1)Utilize the Euclidean distance formula to compute the similarity between the input feature vector and the map's weight vector.</td>
</tr>
<tr>
<td></td>
<td>(2)Track the node BMU producing the smallest distance.</td>
</tr>
<tr>
<td>4.</td>
<td>Update the nodes in the neighborhood of the BMU by Eq. 2.</td>
</tr>
<tr>
<td>5.</td>
<td>Repeat from step 2 until the iteration limit.</td>
</tr>
</tbody>
</table>
The iteration steps we choose are [30 50 100 200 400]. The iteration continues until feature vectors are classified into different classes. The iteration steps distinguishing feature vectors by different descriptors are shown in Fig. 2.

![Iteration steps distinguishing feature vectors by different descriptors.](image)

As shown in Fig. 2, CNN performs better and the feature vectors extracted by CNN are more robust to image changing. Zoom and rotation have great effect on the robustness of image changing. HOG, LBP and CNN are less sensitive to the illumination. HOG shows its stability in the experiments.

**Discussion and Conclusions**

In this paper, we proposed a SOM neural network based descriptor evaluation method. In the tests, CNN performs best, which proves that it is a powerful tool extracting features. CNN, followed by HOG, performs better than LBP, and HOG performs common in zoom, rotation and processing grey images. CNN has better adaptability because its architecture can be designed according to different situations. Furthermore, CNN can not only extract features, but also add a classifier to the frame of networks. However, CNN is time consuming. With the development of computation ability such as GPU and FPGA, CNN will be applied to more situations in extracting features and pattern recognition.

**Acknowledgments**

This work is supported by National Natural Science Foundation of China under Grant U1435220.

**References**


