Pedestrian Object Detection by Using Centroid Neural Network

Thao Nguyen, Kheon-Hee Lee, Chang-Sun Kim, Dong-Chul Park, and Soo-Young Min

***Abstract*—**This paper proposes a method for pedestrian object detection by using Centroid Neural Network (CNN). SIFT(Scale Invariant Feature Transform) is used to produce keypoint feature extracted from image data and the keypoints are used to discriminate a scene with pedestrian objects from a scene without pedestrian objects. Experiments on INRIA Person dataset show that the keypoint features extracted by using SIFT are useful for pedestrian object detection problems and the proposed CNN classifier can detect pedestran object effectively.

***Keywords*—**pedestrian object, neural network, feature, image data.

I. Introduction

TYPICALLY, there are two different procedures for various pattern recognition problems including pedestrian objects: feature extraction procedure and classifier design procedure. For feature extraction procedure for pedestrian object detection, there are two different categories of approaches in order to extract valuable information on pedestrian objects from image data [1]-[5]. The first category of approaches require two different procedures: detecting parts of a pedestrian object and combining them for detecting an entire pedestrian object[1]. The second category of approaches first require finding low level features within a target window and then determining if the target window contains a pedestrian object by using some statistical characteristics of the features [2].

A pedestrian object detection method proposed in this paper is based on CNN(Centroid Neural Network) and SIFT (Scale Invariant Feature Transform) [4] features and this method can be considered as one of the second category methods. The SIFT introduced by Lowe is invariant to scale, orientation, and view point. The SIFT has been widely accepted since it performs well especially in image matching, stereo matching and motion tracking[1]-[5].

For classifier design procedure, Centroid Neural Network (CNN) is adopted in this paper [6]-[8]. When compared with SVM(Support Vector Machine), CNN itself is an unsupervised clustering algorithm with a stable and fast clustering feature.

Thao Nguyen, Kheon-Hee Lee, Chang-Sun Kim and Dong-Chul Park are with the Department of Electronics, Myongji University, YongIn, Rep. of KOREA (phone: +82-31-330-6756, fax: +82-31-3306977, parkd@mju.ac.kr)

Soo-Young Min is with Software Device Research Center at Korea Electronics Technology Institute, SongNam, Rep. of KOREA (e-mail: minsy@keti.re.kr)

The organization of this short paper is constructed as follows: Section II introduces a keypoint extraction method by using SIFT and a review on CNN. Feature extraction with keypoints is proposed in Section III. Experiments on INRIA Person dataset are given in Section IV. Section V concludes this paper.

II. Feature Extraction and Classification

A. Scale Invariant Feature Transform

Since SIFT can provide features invariant to scale, rotation, illumination and viewpoint, it has been widely used for obtaining important invariant features from image[1]-[5]. By using the features by SIFT, an object matching operation images can be achieved[5]. The following procedures are required when SIFT is adopted:

1)Detection of Scale-space extrema: searches over all scales and image locations: DoG(Difference of Gaussian method)

2)Localization of Keypoints: finds a detailed model to determine location and scale and finds stable one by passing through a contrast and edge test.

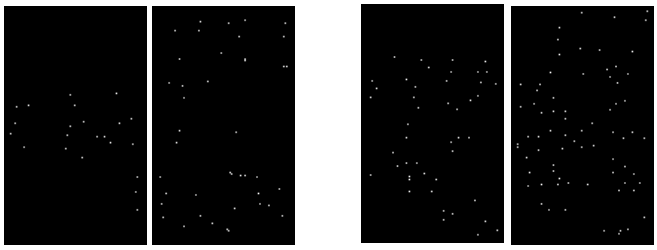
3)Assignment of Orientation: finds dominant orientations for keypoint in order to archive the rotation invariant.

4) Creation of Keypoint descriptor: finds a descriptor based on the histogram of gradient to represent each keypoint. Finally, the descriptor is used for alleviating illumination changes.

B. CNN(Centroid Neural Network)

The CNN algorithm[6] is an unsupervised competitive learning algorithm based on the classical k-means clustering . It finds the centroids of clusters at each presentation of the data vector. The CNN first introduces definitions of the winner neuron and the loser neuron. When a data x is given to the network at the epoch (k), the winner neuron at the epoch (k) is the neuron with the minimum distance to x. The loser neuron at the epoch (k) to x is the neuron that was the winner of x at the epoch (k-1) but is not the winner of x at the epoch (k). The CNN updates its weights only when the status of the output neuron for the presenting data has changed when compared to the status from the previous epoch.

When an input vector x is presented to the network at iteration n, the weight update equations for winner neuron j and loser neuron i in CNN can be summarized as



(a) (b)

Fig. 1 Distribution of keypoints: (a) Negative keypoints (b)Positive key points

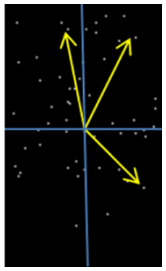


Fig. 2 Histogram feature components

$$w_j(n+1) = w_j(n) + \frac{1}{N_j+1}\left[x(n) - w_j(n)\right] \quad (1)$$

$$w_i(n+1) = w_i(n) + \frac{1}{N_i-1}\left[x(n) - w_i(n)\right] \quad (2)$$

where $w_j(n), w_i(n)$ are the winner and loser neurons with N_i and N_j data, respectively.

The CNN has several advantages over conventional clustering algorithms such as SOM or k-means algorithm when used for clustering and unsupervised competitive learning. The CNN requires neither a predetermined schedule for learning gain nor the total number of iterations for clustering. It always converges to sub-optimal solutions while conventional algorithms such as SOM may give unstable results depending on the initial learning gains and the total number of iterations. More detailed description on the CNN can be found in [6]-[8].

III. Keypoints Feature

With extracted keypoints and locations of keypoints for backgrounds (negative images) and pedestrian objects (positive images) as shown in Fig. 1 and Fig.2, respectively, we found that there exists distinctive difference between distributions of keypoints of positive images and those of negative images. The keypoints of positive images mostly spread out more widely in the image space than those of negative image. Based on this observation, we are able to quantify this difference with the feature descriptor as shown in Fig. 3. The extracted feature is the orientation histograms for accumulated magnitudes of Euclidean distance from each keypoint to the center point at the relative angle between each keypoint and the center point. In experiments, the best number of histogram bins is found to be 12. In our experiments, the numbers of keypoints for negative and positive images are set to 112 and 158, respectively.

Fig. 3 Positive training data images[10]

Fig. 4 Negative training data images[10]

IV. Experiments And Results

INRIA Person dataset is used for experiments[10]. 1,500 images for positive data and 1,500 images for negative data are obtained. SIFT is applied to each of the fixed set of 20,000 patches randomly sampled from images. Then these SIFT keypoint locations are processed to obtain the histogram of orientation for keypoints feature. These histogram features for each positive and negative data are passed through CNN for clustering process. CNN is set to produce two clusters which are corresponding to negative and positive.

In this procedure, when we utilize the fact that the number of keypoints for positive imags is much larger than those for negative image, we eliminate the windows which do not have enough number of keypoints.

In order to evaluate the proposed method, the detection accuracy and training speed for the proposed CNN method are compared with the conventional SVM method The results are summarized in Table I. As can be seen from Table I, detection accuracy of CNN on the testing dataset is about 90.12% while SVM shows almost perfect 99.96% accuracy. These results come from the fact that CNN is a unsupervised learning algorithm while SVM is a supervised learning algorithm. On the other hand, however, CNN has a very important advantage in training time over SVM: CNN finishes its training almost instantly while SVM requires more than 3.28 hours for the same training data on our PC environment (Intel Core Quad 2.33GHz, 4GB of RAM).

Table I
Detection Accuracy And Training Speed

	CNN	SVM
Accuracy	90.12%	99.96%
Training Time	0.92 s	3.28 h

Fig. 5 Experimental results on INRIA dataset

V. Conclusions

In this paper, we propose a classifier model for pedestrian object detection using CNN. The method combines SIFT and CNN to produce an automatic pedestrian detection system. The proposed method utilizes the observation of keyponts distributions for positive data and negative data. Based on the observation, we formulate a histogram feature for the orientation and magnitude of feature points. The proposed method is evaluated with INRIA data set. The results show that the proposed method can detect pedestrian objects with acceptable detection accuracy. The proposed method has the advantageous features of both SIFT and CNN: extremely fast training time and scale invariant feature. The proposed method is acceptable for system that requires real time performance or needs to update the training database quickly.

Acknowledgment

This work was supported by the IT R\&D program of The MKE/KEIT (10040191, The development of Automotive Synchronous Ethernet combined IVN/OVN and Safety control system for 1Gbps class).

References

[1] M. Brown and D. G. Lowe, "Recognising panoramas," in *Proc. IEEE Int. Conf. Computer Vision*, 2003, vol. 2, pp. 1218-1225.

[2] M. Brown and D. G. Lowe, "Invariant Features from Interest Point Groups," Dep. of Computer Science, University of British Columbia, Vancouver, Canada.

[3] A. Tetal, "Face Description with Local Binary Patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.28, no.12, pp.2037-2041, 2006.

[4] D. G.Lowe, "Distinctive Image Features from Scale-Invariant," *International Journal of Computer Vision*, vol. 60, no. 2 , pp. 91-110, 2004.

[5] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Computer Vision and Image Understanding*, vo. 113, no. 3, pp. 345-352, March 2009.

[6] Dong-Chul Park, "Centroid Neural Network for Unsupervised Competitive Learning," *IEEE Trans. Neural Networks*, vol. 11, no. 2, pp. 520-528, May, 2000.

[7] Dong-Chul Park and Young-Jun Woo, "Weighted centroid neural network for edge reserving image compression," *IEEE Trans. Neural Networks*, vol. 12, no. 5, pp.1134-1146, March 2001.

[8] Dong-Chul Park, Oh-Hyun Kwon, and Jio Chung, " Centroid neural network with a divergence measure for gpdf data clustering," *IEEE Trans. Neural Networks*, vo. 19, no. 6, pp. 948-957, June 2008.

[9] C. Cortes and V. Vapnik, "Support-Vector Network," *Machine Learning*, vo. 20, pp. 273-297, 1995.

[10] http://pascal.inrialpes.fr/data/human/

Thao Nguyen received the B.S. degree in Computer Engineering from Ho Chi Minh City University of Technology in 2010 and the M.S. degree in Electronics Engineering from MyongJi University. His research interests include pattern recognition, deep learning, neural networks, and object recognition

Kheon-Hee Lee received the B.S. degree in Electronics Engineering from MyongJi University, Korea, in 2013. He is pursuing his M.S. degree in Electronics Engineering at Intelligent Computing Research Lab at MyongJi University. His research interests include pattern recognition, deep learning, neural networks, and object recognition

Dong-Chul Park (M'90-SM'99) received the B.S. degree in electronics engineering from Sogang University, Seoul, Korea, in 1980, the M.S. degree in electrical and electronics engineering from the Korea Advanced Institute of Science and Technology, Seoul, Korea, in 1982, and the Ph.D. degree in electrical engineering, with a dissertation on system identifications using artificial neural networks, from the University of Washington (UW), Seattle, in 1990. From 1990 to 1994, he was with the Department of Electrical and Computer Engineering, Florida International University, The State University of Florida, Miami. Since 1994, he has been with the Department of Electronics Engineering, MyongJi University, Korea, where he is a Professor. From 2000 to 2001, he was a Visiting Professor at UW. He is a pioneer in the area of electrical load forecasting using artificial neural networks. He has published more than 130 papers, including 40 archival journals in the area of neural network algorithms and their applications to various engineering problems including financial engineering, image compression, speech recognition, time-series prediction, and pattern recognition. Dr. Park was a member of the Editorial Board for the IEEE TRANSACTIONS ON NEURAL NETWORKS from 2000 to 2002.

Soo-Young Min received the B.S. degree in Electronics Engineering from Inha University, Incheon, Korea, in 1987. Since 1993, he has been with Software Device Research Center at Korea Electronics Technology Institute, Song Nam, Korea, where he is a senior researcher. He has been involved in numerous research projects on wireless communication protocol and system software. His research interests include vehicular communication network, wireless communication protocol, and system software design.