ATLANTIS PRESS

Advances in Computer Science Research, volume 87

3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019)

Pedestrian Recognition based on Human Semantics and PCA-HOG

Enyuan Yang [1, 2, a], Rong Xie [1, 2, b]

[1] Beijing Engineering Research Center for IoT Software and Systems, China.

[2] Beijing University of Technology, Beijing, China.

[a] yzyangenyuan@126.com, [b] xierong8989@163.com

Abstract. In real monitoring scenarios, pedestrian semantics, such as gender and clothing type, is very important for pedestrian retrieval and pedestrian recognition. Traditional pedestrian semantics attribute recognition algorithm adopts manual feature extraction and cannot express the association between pedestrian semantics features. This paper proposes a pedestrian semantics recognition method based on improved AlexNet convolution neural network to obtain pedestrian semantics features. Vector. At the same time, a large number of experiments show that HOG descriptors have a good effect in pedestrian recognition, but the number is too large. In this paper, PCA-HOG descriptors are used to express pedestrians and obtain low-dimensional PCA-HOG eigenvectors. Finally, PCA-HOG feature vectors and pedestrian semantic feature vectors are joined together, and LR model is used to predict pedestrian recognition. Compared with traditional methods, the algorithm is simpler, more practical and has higher recognition accuracy.

Keywords: Human semantics, PCA-HOG, Feature mosaic, Loss function.

1. Introduction

Target detection is an early research direction in the field of computer vision, especially pedestrian detection. Because of its high-level semantic information, it can establish the relationship between human bottom features and high-level cognition [1]. Therefore, it is a very popular research direction in the field of computer vision. After decades of accumulation, it has achieved remarkable development. In recent years, with the concept of smart city proposed, tens of thousands of surveillance cameras are installed in every corner of the city, making everyone's life more intelligent. Therefore, pedestrian recognition in surveillance scenarios has important research value, and it also has great market prospects in intelligent video surveillance industry.

At present, pedestrian detection for surveillance can be basically divided into two categories: Background-Based modeling in continuous frames and human body-based modeling in single frame images. For example, Sun Qingjie, Key Laboratory of Computer Science, Chinese Academy of Sciences, et al. proposed a rectangular fitting human body detection algorithm based on side shadow human body model and its corresponding probability model. Tan Tieniu, Institute of Automation, Chinese Academy of Sciences, and others, are engaged in visual analysis of human motion. The core of the visual analysis is to detect, track, identify and understand human behavior from image sequences by using computer vision technology. It is mainly used in the field of visual monitoring and gait-based identification [8]. Zheng Nanning of Xi'an Jiaotong University and others have studied the method of recognizing pedestrians by using support vector machine. Through sparse Gabor filter, the pedestrian features in the pedestrian sample image are extracted. Then, the extracted sample features are trained by using support vector machine, and the window which may belong to pedestrians in the image is extracted by using the trained classifier through traversing the image[2].Although Gabor filter is relatively effective in extracting features, it takes a long time and is not suitable for real-time image processing.

Oren and C. Papageorgiou of Massachusetts Institute of Technology have established Haar wavelet template and applied it to pedestrian detection. Haar wavelet template is often used to express simple objects and has the characteristics of effective and fast detection. It has been widely used in image object detection. Similarly, Haar wavelet template pedestrian detection algorithm has become one of the classical algorithms in pedestrian detection field. One. Navneet Dalal and Bill Triggs of

France use gradient direction histogram (HOG) to represent human characteristics, and validate it on the INRIAPerson sample bank. This method has high detection rate and strong applicability in human detection. The same algorithm has strong performance in road pedestrian detection, which has attracted the attention of many scholars. Niebles. J. C. of the University of Illinois proposed a pedestrian recognition algorithm using AdaBoost cascade model, and applied it to the field of pedestrian detection to improve the effect of pedestrian detection and recognition [3]. However, most of the feature factors extracted by these methods are from a single angle, with high feature dimension and feature redundancy, and it is difficult to explain the reasons when the model is not effective. In this paper, a pedestrian recognition method based on human semantics and PCA-HOG in surveillance scene is proposed. The improved network structure based on AlexNet is used to generate human semantics feature vectors, then PCA-HOG feature vectors are stitched together as the feature vectors of pedestrian information and input into LR model for pedestrian recognition.

2. Method

2.1 Human Semantic Features

When identifying whether the target is the same person or not, human beings usually describe a person through some specific appearance or biological attributes, such as clothes, color, gender, height, hair length, and the details of clothes, such as the texture and pattern of clothes, or pedestrian accessories, including backpacks, hats, glasses, which have strong prior knowledge characteristics [5]. These features do not change with pedestrian posture and illumination. High-level semantic features are independent of the camera horizon and have good robustness [6].

Pedestrians are represented by semantic features rather than by underlying features. In applications, problems encountered are better interpretable. Semantic representation of images is actually a representation of knowledge [7]. From the semantic attributes set by experts in the field of pedestrian recognition, representative multi-semantic attributes suitable for pedestrian recognition are selected, including clothing-related attributes and human biological-related attributes.

In this paper, the attributes of human, clothing and pedestrian appendages are selected in reference [9]. Related attributes of human biological characteristics include: gender, short sleeves, lunch break, jacket, clothing pattern, shorts, trousers, skirts and so on; accessory attributes: backpack, handbag, etc[10]. N attributes that can describe pedestrians are selected as labels. The value of image attributes I in training set is 0 or 1, 0 means that the image does not contain attributes i, and 1 means that the image contains attributes I. Let the training set model the improved AlexNet network. The test image is input into the trained model, and the probability of each image containing attributes is obtained. Then the pedestrian image is represented as a N-dimensional vector, and the pedestrian attributes feature vector (1) is obtained. The improved AlexNet network can transform the high-dimensional bottom visual features into low-dimensional semantic feature vectors.

$$f_a = (a_1, a_2, \cdots, a_N), a_i \in [0,1] \tag{1}$$

2.2 Network Structure of Human Semantics

Suppose there are N pictures in the pedestrian sample, each of which is marked with L pedestrian attributes. For example, gender, hair length, age, etc. Each pedestrian picture can represent xi, i∈[1, 2 .....N]. Each picture Xi corresponds to the pedestrian attribute label vector yi. The attributes corresponding to each label vector yi are Yil, Yil∈[0,1], l∈[1,2.... L]. If yil = 1, it indicates that the training sample Xi has this attribute; yil = 0, it indicates that the training sample Xi does not have this attribute. This paper presents a pedestrian semantic recognition model based on convolution neural network. The model is fine-tuned based on AlexNet network model. The basic network structure is the same as AlexNet, and the number of layers is 8 (the first five layers are convolution layers, and the last three layers are full connection layers) [4]. In the training stage of the model, the input of the model is a pedestrian picture and the corresponding pedestrian attribute label vector. In the testing

stage, the output of the model is the attribute category of the pedestrian sample image prediction. Generally, attributes are related. Most attribute recognition methods will separate each attribute, ignoring the association information between the attributes [11]. For example, the length of hair can improve the accuracy of gender recognition. In order to make better use of the association between attributes and improve the recognition accuracy of pedestrian attributes. In this paper, a new loss function is proposed, so that all pedestrian attributes can be learned simultaneously in the training process. Loss functions such as formulas (2) and (3)

$$h_\theta(x) = g(\theta^{\mathrm{T}} x) = \frac{1}{1 + e^{\theta^T x}} \tag{2}$$

$$Loss = \sum_{i=1}^{N} \sum_{j=1}^{m} (y_{ij} \log h(x_{ij}) + (1 - y_{ij}) \log(1 - h(x_{ij}))) \tag{3}$$

2.3 PCA-HOG Features

The general idea of HOG descriptor is to describe the shape of local objects in the image with gradient intensity or edge direction distribution [12]. The process of description is by dividing the whole tracking area into non-overlapping cells, calculating the direction histogram information in each grid, and combining the histogram information into descriptors [13]. In order to improve the performance of the descriptor, the whole tracking region should be divided into blocks, and the sum of gradient intensity of all pixels in several grids in the block should be calculated to normalize all grids in the block and get the normalized descriptor.

PCA-HOG descriptor can be summarized as follows: firstly, the tracking region is transformed into the grid HOG descriptor, and then the principal component analysis (PCA) PCA is used to transform the HOG descriptor into a linear subspace [14]. Using PCA-HOG descriptor to recognize pedestrians can make the algorithm focus on the shape information of pedestrians and be robust to some illumination changes. Moreover, PCA-HOG descriptor reduces the data dimension of the original HOG descriptor, which greatly reduces the amount of computation [15].

2.4 LR Pedestrian Recognition Network Structure

Based on the improved network structure of AlexNet, the feature vectors of human semantics are generated; the 30-dimensional feature vectors are obtained by PCA-HOG descriptors; and finally, the feature vectors of human semantics and PCA-HOG feature vectors are joined together to form a feature vector, which is input into the LR model as the feature vectors of pedestrian information, for the final pedestrian recognition. The network structure is shown in Fig.1.

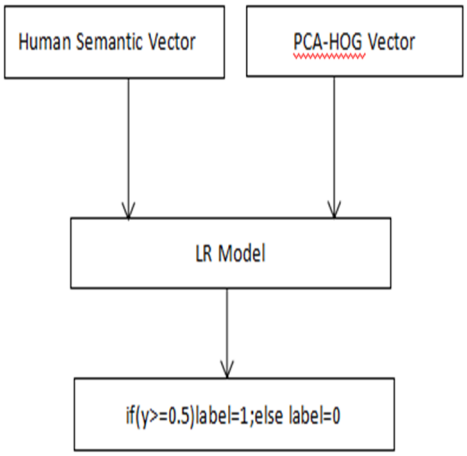

Fig 1. LR Pedestrian Recognition Network Structure

3. Results and Analysis

The experimental results of the deep convolution network model used in the human semantic network are obtained by using the Caffe deep learning framework. The experimental results are shown in TABLE 1. In the experiment, when identifying 10 kinds of attribute tags, the multi-classification is transformed into a two-classification task for a single category in each tag, and the last layer of software Max is changed into 10 LR models. It can be seen from the experiment that the average recognition accuracy of 10 kinds of attribute tags reaches 0.7845 in the multi-label classification and recognition task of pedestrian attributes by using the traditional manual design feature algorithm ikSVM. The convolution network structure used in this study has significantly improved the recognition results in pedestrian attribute multi-label classification and recognition tasks. The average recognition accuracy of 10 types of attribute labels is 0.805.

The dimension of HOG descriptor obtained by opencv is 400 dimensions, and it is reduced to 30, 40, 50, 60, 80, 100, 200 and 300 dimensions by PCA. The experimental results are shown in TABLE 2. By comparison, 30 dimensions are the best dimension.

INRIA contains many kinds of pictures of different clothes, postures and backgrounds. It is a rather complex static pedestrian detection database. The experimental results are shown in TABLE 3. By comparison, the pedestrian recognition method based on human semantics and PCA-HOG in this paper has higher detection rate than HOG+linear SVM and HOG+RBF-SVM models. With the increase of samples, the detection rate is higher, which shows that the model has good robustness. And the accuracy of pedestrian detection has been improved.

Table 1. Recognition Rate of Different Attributes in DeepMAR and Improved AlexNet

Attributes	DeepMAR	Improved AlexNet
Hat	62.3883	70.4545
Hair	70.3845	68.7856
Glasses	73.67	78.5789
The shoulder bag	73.5757	80.3456
Handbag	72.4595	74.3567
Paper bag	66.6855	68.3478
Gender	69.2357	78.67
Sweater	66.7341	70.6774
Short skirt	76.8846	82.845
Dress	79.0893	76.5678
Jeans	82.328	83.6745
Shoes	74.7954	76.4557

Table 2. Recognition Effect of PCA-HOG Vectors with Different Dimensions

PCA-HOG dimension	Recognition rate
400	**0.9757**
30	**0.9800**
40	**0.9802**
50	**0.9794**
80	**0.9826**
100	**0.9781**
200	**0.9779**
300	**0.9770**

Table 3. Pedestrian Detection Rate of Different Models

Database	Number of pictures	Method	Detection rate
INRIA	800	HOG+Linear SVM	74.8%
		HOG+RBF-SVM	82.3%
		Method of this paper	87.8%
	1500	HOG+Linear SVM	76.5%
		HOG+RBF-SVM	88.4%
		Method of this paper	91.2%

4. Conclusion

This paper presents a pedestrian recognition method based on human semantics and PCA-HOG. When using the improved AlexNet deep convolution neural network to extract human semantic features, the network structure is improved. The output layer is changed from soft Max to 10 LR models, which can simultaneously recognize multiple human attributes. At the same time, dimensionality reduction of HOG features is processed to select effective features. The model can be interpreted better by transforming the bottom features with more dimensions into high-level features with less dimensions. The experimental results show that the proposed method has better performance in the accuracy of pedestrian recognition and the interpretability of the model. At the same time, we can also expand the addition of Haar descriptor, LBP descriptor, to reduce the impact of illumination, occlusion and other issues, and further improve the accuracy of the model.

References

[1]. Viola P, Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features [C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. 2003.

[2]. Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. 2005.

[3]. Felzenszwalb P F, Huttenlocher D P. Pictorial Structures for Object Recognition [J]. International Journal of Computer Vision, 2005, 61 (1):55-79.

[4]. Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size [J]. 2016.

[5]. Farenzena M , Bazzani L , Perina A , et al. Person re-identification by symmetry-driven accumulation of local features [C]// Computer Vision & Pattern Recognition. IEEE, 2010.

[6]. Chen Y Huo Z, Hua C. Multi-directional saliency metric learning for person re-identification [J]. IET Computer Vision, 2016, 10 (7): 623-633.

[7]. Zheng W S, Gong S, Xiang T. Person re-identification by probabilistic relative distance comparison [C]// Computer Vision & Pattern Recognition. IEEE, 2011.

[8]. Zhao R, Ouyang W, Wang X. Unsupervised Salience Learning for Person Re-identification [C]// Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013.

[9]. Umeda T, Sun Y, Irie G, et al. Attribute Discovery for Person Re-Identification [C]// International Conference on Multimedia Modeling. 2016.

[10]. Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification [C]// Computer Vision & Pattern Recognition. 2015.

[11]. Nguyen N B, Nguyen V H, Duc T N, et al. Using Attribute Relationships for Person Re-Identification [M]// Knowledge and Systems Engineering. 2015.

[12]. Liu H, Xu T, Wang X, et al. Related HOG Features for Human Detection Using Cascaded Adaboost and SVM Classifiers [C]//The International Multimedia Modeling Conference. Springer Berlin Heidelberg, 2013.

[13]. Porikli F. Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces [J]. Proc Cvpr, 2005, 1:829-836.

[14]. Hoang V D, Le M H, Jo K H. Hybrid cascade boosting machine using variant scale blocks based HOG features for pedestrian detection [J]. Neurocomputing, 2014, 135 (C):357-366.

[15]. Gan G, Cheng J. Pedestrian Detection Based on HOG-LBP Feature. [C]// Seventh International Conference on Computational Intelligence & Security. 2012.