PEDESTRIAN DETECTION VIA PART-BASED TOPOLOGY MODEL

Wen Gao, Xiaogang Chen, Qixiang Ye, Jianbin Jiao[+]

Graduate School of Chinese Academy Science, Beijing, China
+Corresponding Author: Fax: +86-010-88256278, Email: jiaojb@gucas.ac.cn

ABSTRACT

In this paper, we propose a part-based topology model and a pedestrian detection method, which obviously improve the detection accuracy. In Our method, pedestrian is divided into several parts. Firstly, histogram of oriented gradients (HOG) features and linear support vector machine (SVM) classifier are used to detect pedestrian parts. Secondly, a novel binary descriptor called log-polar pattern (LPP) is proposed to represent the spatial relation of a part pair. Then multiple LPPs are combined as a log-polar topology pattern (LTP) to model the global topology of a pedestrian. Finally, we put the LTP into One-Class SVM (OC-SVM) to determine whether the detected parts indicate a pedestrian or not. Experiments in INRIA dataset show that our method is robust to occlusion and multi-postures, which obviously reduces the miss rate.

Index Terms—Pedestrian detection, Support Vector Machine, Log-polar Topology Pattern

1. INTRODUCTION

Pedestrian detection has been a very important research topic in image analysis with a number of potential applications, such as video surveillance，robotics and driver assistance. Although extensively investigated in recent years, pedestrian detection is still a challenging problem due to the complex backgrounds, varied illumination conditions and various views and postures.

The existing work of pedestrian detection can be divided into two strategies. The first strategy is to define a global representation for pedestrians, then use a classification procedure to perform detection [1, 2, 3, 4]. The global representation is usually defined as a group of local features (such as Haar-like [5], HOG [6] or LBP [7] features) extracted from fixed locations. These features from different locations are assigned different importance by classifiers in the training process. These approaches with global representations report a good performance when detecting pedestrians of small view/posture variation. However, when facing pedestrians of large view/posture variation in images, these approaches often fail since the feature locations are fixed and cannot adapt to the view changes or deformations of pedestrians.

The second strategy of pedestrian detection is to use a part-based representation, and perform detection by evaluating both the parts' responses and the topology of each part [8, 9, 10]. In [8], Haar-like features are trained with SVMs to detect parts, and then an empirically defined topology is used to represent pedestrians. In [9], Wu et al. introduced maximum a posterior (MAP) estimation to locate each part of pedestrian and model their topology. In [10], a deformable part-based model (DPM) is proposed for pedestrian detection, pedestrian parts and their spatial bias are modeled with a structure SVM of latent variables. In the training and detection phase, a local searching operation is applied to optimize the location deformation of each part. In [11], an extension to the DPM is proposed, which allows for sharing of object part models among multiple mixture components as well as object classes. This results in more compact models and allows training examples to be shared by multiple components, ameliorating the effect of a limited size training set. DPM methods report the state-of-the-art performance. But they suffer from the loss of global topology when performing part location, which will bring out false detection results in clutter background.

The work of this paper is inspired by the part-based detection model which is powerful and general for articulated body. We segment a pedestrian object into parts, and then the HOG feature is adopted to describe each part. We construct a descriptor, called log-polar topology pattern (LTP), to represent the topology of a pedestrian. On the extracted HOG and LTP descriptors, two kinds of SVMs, linear and One-Class SVMs [12], are employed respectively to train part models and a global topology model for pedestrians. Compared with the DPM methods, our method depends on global topology description in spite of deformation or move of local parts to deal with the view and posture problem, which brings a novel way for pedestrian detection.

2. PEDESTRIAN DETECTION METHOD

2.1 Overview of the method

Our method is based on a Discriminative model. Fig.1 shows the flowchart of the proposed method, which includes two phases: part detection and topology model. We first perform

978-1-4673-2533-2/12/$26.00 ©2012 IEEE

a local dense search and optimize the locations of each part by referencing the method of [10]. On the optimized part locations of samples, linear SVM models are trained for parts detection, which will be detailed in section 2.2. Since the locations of parts are not accurate, we extract a LTP descriptor for each sample according to its topology and a pre-learned OC-SVM classifier is used to classify pedestrian or non-pedestrian, which will be detailed in section 2.3.

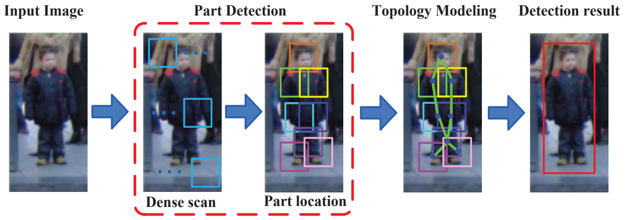

Fig.1. Flowchart of the proposed detection method

The final discriminative functions are shown in Eq.(1-2) to determine the detection results:

$$S(x)=\sum_{i=1}^{P}\alpha_i\cdot F_{i,HOG}+\beta\cdot F_{LTP} \tag{1}$$

$$f(x)=\begin{cases}0, & if\ S(x)<\tau\\ 1, & otherwise\end{cases} \tag{2}$$

where $F_{i,HOG}$ denotes a HOG feature vector extracted from part *i*, while α_i represents the weight vector of a linear SVM. F_{LTP} denotes a LTP descriptor, and β is the weight vector for OC-SVM. The first term of Eq.(1) computes the scores of part detection. The second term is the score of topology. In Eq.(2), $f(x)$ denotes the final classifier, when $f(x)=1$ denotes a pedestrian, and $f(x)=0$ is a negative. τ is the threshold determined empirically.

2.2 Part detection

In the part detection procedure, HOG features [6] are used to represent each pedestrian part. We first calculate the gradient orientation of each pixel. In a 8×8 pixels sized cell we calculate a 9-dimensinal HOG features by calculating the 9-bin histogram of gradient orientations of all pixels in this cell. Each block contains 4 cells, and 36-dimensional features are extracted from a cell. Each part is represented by *K* blocks, and 36*K*-dimensinal HOG features $F_{i,HOG}$ are extracted.

In the training phase, we reference to the work in [10] to calculate the weight vector α_i in Eq.(1) with an iterative algorithm, which alternates between fixing latent values for positive samples and optimizing the latent SVM objective function. Then a total of *P* parts' locations can be obtained in this stage. At the same time, *P* linear SVMs are trained corresponding to the first term of Eq. (1). The training set we adopt is INRIA [6].

2.3 Topology Modeling with LTP Descriptor

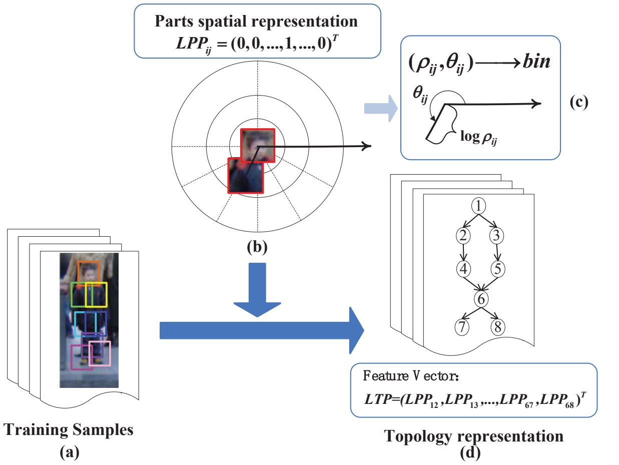
Parts spatial representation
$LPP_{ij}=(0,0,...,1,...,0)^T$
$(\rho_{ij},\theta_{ij})\longrightarrow bin$
θ_{ij}
$\log\rho_{ij}$
(c)
(b)
1
2
3
4
5
6
7
8
Feature Vector:
$LTP=(LPP_{12},LPP_{13},...,LPP_{67},LPP_{68})^T$
Training Samples
(a)
Topology representation
(d)

Fig.2. LPP and LTP descriptor extraction

LPP descriptor: As shown in Fig.2, (a) is the training sample set, and according to (b) a log-polar pattern (LPP) can be formed by using a log-polar histogram. The polar radius and polar angle between the centers of a part pair are used to represent the spatial relation between them. Assuming using *m* bins for $\log\rho$ and *n* bins for θ, the spatial relation of a part pair can be represented as a $m\times n$ dimensional vector of binary values. We use the Euclidean distance to represent the spatial relation between part *i* and *j*, and computed as Eq.(3). The polar angel can be computed from Eq.(4), and the angle scope of each bin is non-uniform in our experiment, which can be modulated for proper pedestrian postures.

$$\rho_{ij}=\sqrt{|X_i-X_j|^2+|Y_i-Y_j|^2}\,/\,max(\rho_{ij}),i\neq j \tag{3}$$

$$\theta_{ij}=\arctan\frac{X_i-X_j}{Y_i-Y_j}+\lambda\pi,\lambda=1,2.\theta_{ij}\in[0,2\pi) \tag{4}$$

where X_i and Y_i denote the center point of a part, and the parameter λ is used to limit the angle scope in $[0,2\pi)$. $max(\rho_{ij})$ denotes the maximum distance of each pair of parts.

As shown in Fig.2, the relation is denoted by the directed connection between a part pair. Given a pair of relative parts in topology, such as head and right shoulder (shown in Fig.2(b)), we consider head as the pole of the log-polar coordinate system, so that the right shoulder is projected in a particular bin by computing the parameters ρ and θ, as shown in Eq.(3-4). By filling in the bin with 1 and others with 0, a $m\times n$ dimensional vector of binary values is constructed as a LPP descriptor.

LTP descriptor: After gaining the LPPs of each relative pair of parts, a log-polar topology pattern (LTP) is formed to represent the part topology of a pedestrian. As shown in Fig.2(d), the directly connected parts are empirically considered as the topology of pedestrians. We extract the LPP descriptors in each pair of parts and combined them together as the LTP descriptor for a pedestrian, as shown in Fig.2(c). It is known that when a pedestrian part is moving, it leads to the movement of the relative part. Taking left foot and upper leg movements for an instance, in log-polar coordinated system, upper leg is taken as the origin and left foot will be projected in a bin which covers a broader angle scope, so that the relative local deformation problem can be well solved.

Topology modeling: After calculating the LTP feature vector, an OC-SVM is adopted to train the pedestrian topology model and distinguish pedestrian from non-pedestrian. OC-SVM does not require negative samples, which is suit for topology relation description because of the difficulties in collecting all the possible non-pedestrian relations. The training phase is formulated as the following optimization problem:

$$min \quad \frac{1}{2}\|\beta\|_2^2 \tag{5}$$

$$s.t. \begin{cases} \beta \cdot \phi(F_{i,LTP}) \geq R \\ R > 0 \end{cases}, i = 1,...,l \tag{6}$$

Eq.(6) is the constraint to Eq.(5), which ensures that training samples should be correctly classified, where $\beta \in R^l$ is the weight vector gained from OC-SVM, $F_{i,LTP}$ is the LTP feature vector of the i^{th} sample, and l is the number of training samples. After solving Eq.(5-6), we obtain weights and the weighted feature vectors of samples.

Algorithm 1 Topology modeling procedure

```
1. Input Data:
   Pedestrian samples with part locations;
2. For $j$=0 to $M$
      For $i = 1$ to $P$
         Extract LPP descriptor $LPP_{ij}$;
      End for
      Construct topology descriptor,
      $F_{LTP}^j = \left[LPP_0^j,...,LPP_P^j\right]^T$;
   End for
   Construct feature matrix: $\{F_{LTP}^0, F_{LTP}^0, ..., F_{LTP}^M\}$;
3. Training model by solving optimization in Eq.(5-6);
4. Output:
   Weight vector $\beta$ and the topology model for the second part of Eq.1: $\beta \cdot F_{LTP}$.
```

The topology modeling procedure is summarized as Algorithm 1. Here we provide an example of training topology model. Given M training samples each of which contains P parts. LPP_{ij} is the LPP feature represent the relation between part i and j. F_{LTP}^i is the LTP feature of the i^{th} sample.

3. EXPERIMENTAL RESULTS

The INRIA pedestrian dataset [6] is used to train models and evaluate the proposed detection approach. In the dataset, there are 2416 samples for training and 288 images for testing. Testing images cover pedestrians of diverse postures and complex backgrounds.

While extracting the LPP descriptor, we define the log-polar histogram as 5×8 dimensions, and formulate the LTP descriptor as a 320 dimensional vector. The angle scope of each bin is non-uniform. Especially, the widest pace covers 90 degrees and the narrowest covers 30 degrees in our experiment. For the final classifier, we adjust the threshold τ in Eq.(2) to pursuit a better tradeoff between miss rate and False Positive Per-Image (FPPI). We define a correct detection as the overlapping between the predicted region and the ground-truth region is more than 50 percent [6]. The non-maximum suppression value is set as 0.5.

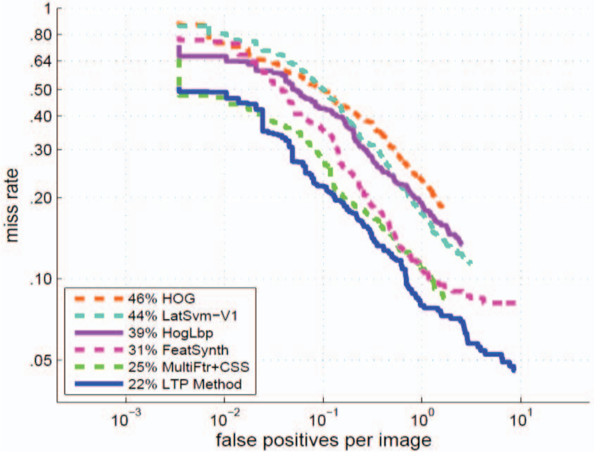

Fig.3 Performance and comparisons on INRIA dataset.

In Fig.3, we plot miss rate versus false positives per image (FPPI) as a common reference value and use log-average miss rate as a common reference for summarizing performance. We compare LTP method with some global representation method [6,8,15] and part-based method [11,14]. It is obvious that the LTP method performs better than others with a log-average miss rate of only 22%, while others achieve miss rate around 25-46% (the lower curves indicate better performance). Since restraining the parts combination in a trained topology model, some false detection without a regular topology can be rejected, our proposed method thus is more robust in reducing the false alarm rate. In particular, when FPPI is larger than 0.1, the miss rate is visibly reduced to less than 20%.

Table.1 reports the classifier and miss rate comparison of the methods mentioned above, ordered by descending log-average miss rate in INRIA dataset. LTP descriptor well represents the topology of pedestrian, and OC-SVM is more suitable for modeling with less false classification. It shows that the part-based LTP method performs at least 0.03 lower in Log-average miss rate than global-based method, and even 0.09 lower than other part-based method.

Method	Part based	Classifier	Log-average miss rate
HOG[6]	--	Linear SVM	0.459788
LatSVM-v1[11]	√	Latent SVM	0.438304
HOGLBP[8]	--	Linear SVM	0.390968
FeatSynth[14]	√	Linear SVM	0.308773
MultiFtr +CSS[15]	--	Linear SVM	0.247449
LTP Method	√	OC-SVM	**0.216689**

Table.1. Comparison of pedestrian detectors

Some detection results are shown in Fig.4. It can be seen that in (a) and (b) all pedestrians are correctly located in spite of the crowd scene and variety views with occlusions. In (c) and (d), humans of unusual postures (bending down or riding a bicycle) are also correctly detected, showing the effectiveness of the proposed topology model when facing posture variations.

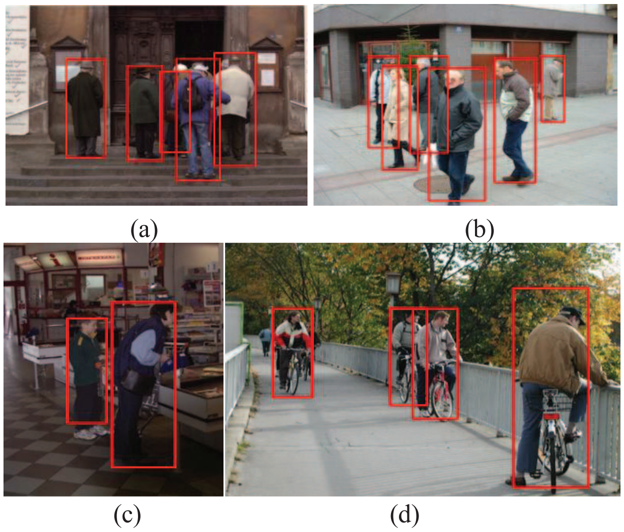

Fig.4. Detection results

4. CONCLUSIONS

In this paper, we proposed a part based pedestrian detection method, in which a novel topology descriptor LTP is applied to model the relationship between each part of pedestrian. Compared with existing global representation and deformable part-based models, experiments show that the proposed LTP descriptor is simpler but more effective. The performance of LTP method reaches state-of-the-art with even more robustness to view and posture variation, which reduces the miss rate to 22% in INRIA dataset. In the future, we will extend our topology model to detect human body with more complex postures.

5. REFERENCES

[1] Q. Zhu, S. Avidan, M. Yeh, and K. Cheng, “Fast Human Detection Using A Cascade Of Histograms Of Oriented Gradients,” *IEEE CVPR*, pp. 1491-1498, 2006.

[2] R. Xu, B. Zhang, Q. Ye, J. Jiao. “Human Detection in Image Via L1-Norm Minimization Learning,” *IEEE. ICASSP*, 2010, pp. 3566-3569.

[3] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio, “Object Recognition With Cortex-Like Mechanisms,” *IEEE Transactions on PAMI*, vol. 29, no.3, pp.411-426, 2007.

[4] O. Tuzel, F. Porikli, P. Meer, “Pedestrian detection via classification on riemannian manifolds,” *IEEE Transactions on PAMI*, vol.30,no.10, pp.1713-1727, 2008.

[5] P. Viola, M. Jones, and D. Snow, “Detecting Pedestrians Using Patterns Of Motion And Appearance,” *IJCV*, vol.63, no.2, pp. 153-161, 2005.

[6] Dalal, N, Triggs, B., “Histograms of Oriented Gradients for Human Detection,” *IEEE CVPR*, vol.1, pp.886-893, 2005.

[7] X. Wang, T.X. Han, S. Yan, “An HOG-LBP Human Detector With Partial Occlusion Handling,” in: Proc. *IEEE ICCV*, 2009.

[8] A. Mohan, C. Papageorgiou, and T. Poggio, “Example-based object detection in images by components,” *IEEE Transactions on PAMI*, vol.23(4) pp.349-360, 2001.

[9] B. Wu, Nevatia R. “Detection Of Multiple, Partially Occluded Humans In A Single Image By Bayesian Combination Of Edgelet Part Detectors,” *IEEE ICCV*, 2005, Vol.1, pp.90-97.

[10] P. Felzenszwalb, R. Girshick, D. McAllester and D. Ramanan. “Object Detection with Discriminatively Trained Part Based Models,” *IEEE Transactions on PAMI*, Vol. 32, No. 9, September, 2010.

[11] P. Ott and M. Everingham, “Shared Parts for Deformable Part-based Models,” *IEEE CVPR*, 2011.

[12] Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C. “Estimating The Support of A High-dimensional Distribution,” *IEEE. Neural Computation,* 2001 1443-1471.

[13] P.Dollar, C.Wojek, B.Schiele, and P.Perona, “Pedestrian detection: A benchmark,” *IEEE CVPR*, 2009.

[14] A, Bar-Hillel, D. Levi, E. Krupka, C. Goldberg, “Part-based feature synthesis for human detectioin,” *IEEE ECCV*, 2010.

[15] S. Walk, N. Majer, K. Schindler, and B. Schiele, “New features and insights for pedestrian detection,” *IEEE CVPR*, 2010.