Saliency-Based Deformable Model for Pedestrian Detection

Xiao Wang[1], Jun Chen[1,2], Wenhua Fang[1], Chao Liang[1,2,*], Chunjie Zhang[3], Kaimin Sun[4], and Ruimin Hu[1,2]

[1] National Engineering Research Center for Multimedia Software,
School of Computer, Wuhan University, Wuhan, 430072, China
[2] Research Institute of Wuhan University in Shenzhen, China
[3] School of Computer and Control Engineering,
University of Chinese Academy of Sciences, Beijing, 100190, China
[4] School of Remote Sensing Information Engineering,
Wuhan University, 129 Luoyu Road, Wuhan, 430079, China
cliang@whu.edu.cn

Abstract. Pedestrian detection, which is to identify category (pedestrian) of object and give the position information in the image, is an important and yet challenging task due to the intra-class variation of pedestrians in clothing and articulation. Previous researches mainly focus on feature extraction and sliding window, where the former aims to find robust feature representation while the latter seeks to locate the latent position. However, most of sliding windows are based on scale transformation and traverse the entire image. Therefore, it will bring computational complexity and false detection which is not necessary. To conquer the above difficulties, we propose a novel Saliency-Based Deformable Model (SBDM) method for pedestrian detection. In SBDM method we present that, besides the local features, the saliency in the image provides important constraints that are not yet well utilized. And a probabilistic framework is proposed to model the relationship between Saliency detection and the feature (Deformable Model) via a Bayesian rule to detect pedestrians in the still image.

Keywords: Pedestrian detection, Saliency-Based Deformable Model, Saliency Detection, Bayesian rule.

1 Introduction

Recently, pedestrian is an important component part over a large public areas, such as visual surveillance, robotics, and automotive safety. In these scenarios, pedestrian detection is becoming a hot research spot in the computer vision community. During the past three years, a lot of research effort [13, 4, 5, 14, 15, 17] has been devoted to this field. However, the latent scale and the position of the pedestrian are unknown, so they usually resize the sliding window or

[*] Corresponding author.

W.T. Ooi et al. (Eds.): PCM 2014, LNCS 8879, pp. 203–210, 2014.

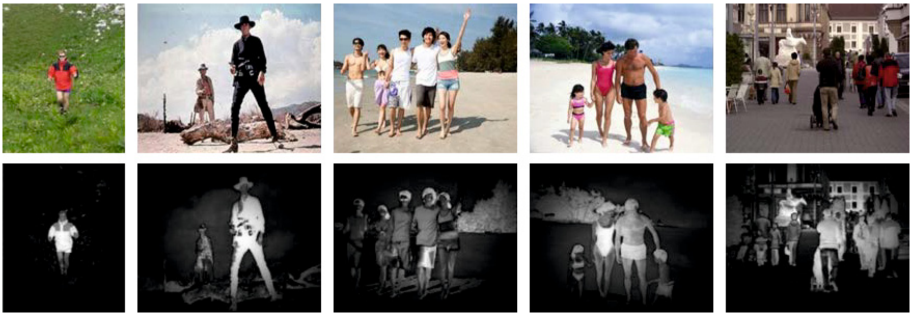

Fig. 1. Examples of salient regions extracted from original color images in different scenarios. Upper part of the figure shows pedestrian in different scenarios and the lower part shows corresponding salient regions. We can see from the above image that pedestrians have been included in salient map.

/and image many times when detecting by sliding window, and this is so-called multiple scales detection [6]. The selection of a proper scale and the step width of the sliding window will greatly affect the algorithm's precision and efficiency. This is more obvious in the case of high-resolution images. Fortunately, we can get the approximate location of pedestrians by salient regions and avoid the above problems. As we can see from the Fig .1, salient regions contain objects (such as pedestrians) of the image in different scenarios.

Generally speaking, pedestrian detection can be considered as a very important part of visual retrieval problem [11], given a query pedestrian image taken in one image, the algorithm is expected to search the same pedestrian captured by other image. Typically, it consists of two stages feature extraction and scale transformation, where the former aims to find robust feature representation while the latter seeks to give the bounding box precisely. The combination of the histograms of oriented gradients (HOG) features and linear SVM learning machine, proposed by Dalal et al. in [3], has been proven as a competitive method to detect pedestrians. Farenzenaet al. [8] has divided the image of person into 5 regions by exploiting symmetry and asymmetry perceptual principles, and then combine multiple color and texture features to represent the appearance of people. Wanli Ouyang [12] has proposed pedestrian detector which is learned with a mixture of deformable part-based models to effectively capture the unique visual patterns appearing in pedestrians. However, a majority of detectors surveyed in [7] remain complex. Because they detect pedestrian from full image, rather than effective local regions where the object exists.

In this paper, we propose a novel Saliency-Based Deformable Model (SBDM) which combines the salient detection model and the Deformable Model combined by Bayesian rule to detect the latent pedestrians in the still image. It is easy to see from Fig. 2 that we tend to focus on effective local regions of the image rather than the entire. More precisely, we can get saliency degree by histogram based contrast method [2]. Therefore, we can get the effective local regions by combining saliency regions and original image. And deformable model feature [8] is extracted from the effective local regions at the reference scale, and finally a logistic regression classifier is adopted to detect the pedestrian object.

Summarizing, the contribution of our work is two-fold. (1) it takes full advantage of the local regions where the objects exist rather than the whole image which contains a lot of interference information, thus greatly reduces the false alarm rate and the computational complexity. (2) a new probabilistic framework is proposed to model the configuration relationship between results of the salient detection model and the deformable model via Bayesian rule.

2 The Proposed Pedestrian Detection Framework

Our framework has fused the salient region detection model and the traditional deformable model descriptor by Bayesian rule. Rather than summing these votes, we model all the random variables S_R (rectangular window R of the image is salient region) and P_{SR} (we can detect pedestrian from the S_R) in a probabilistic method so that we can determine the final probabilistic values via the Bayesian inference process. Thus, we are interested in modeling the joint posterior of S_R and P_{SR} given local region R, and apply the Bayesian theorem then gives:

$$p(S_R, P_{SR}|R) \propto p(R|S_R, P_{SR})p(S_R, P_{SR}) \tag{1}$$

Our framework is illustrated in Fig. 2. We now focus on the prior and likelihood terms separately.

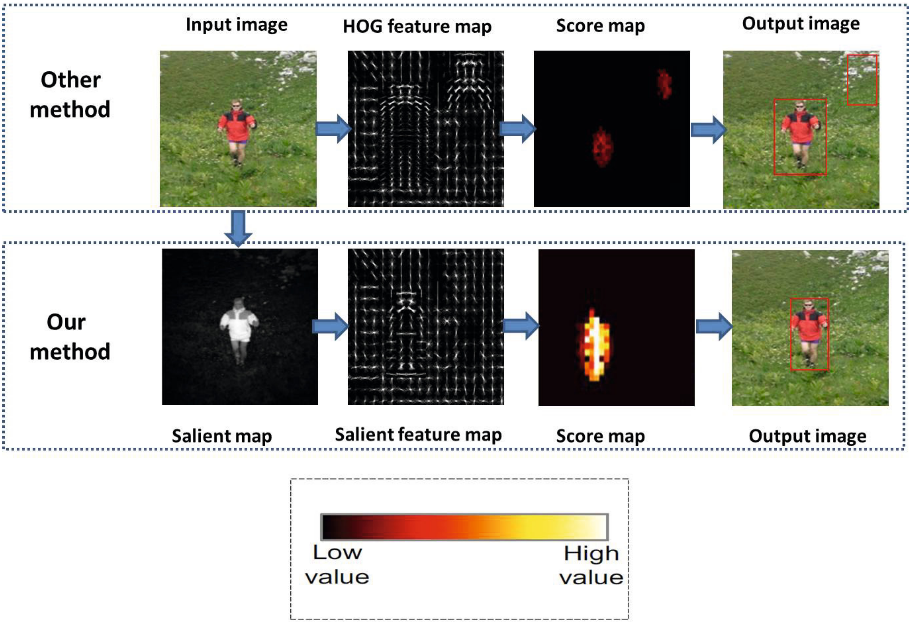

Fig. 2. The framework of our pedestrian detection system

2.1 Salient Region Prior Model

According to a histogram based contrast method [2] which defines saliency values for image pixels using color statistics of the input image. The pixels with the same color value have the same saliency value under this definition, since the measure is oblivious to spatial relations. Therefore, it is so easy to get saliency value for each pixel as,

$$V(k) = \sum_{i=1}^{N} d(k,i), \tag{2}$$

I is the image in Lab color space; $d(k,i)$ is the color distance metric between pixels k and i, and N is the number of pixels in image I. To be more precise, the new saliency degree can be seen as a weight, and the weight is calculated by

$$\omega_k = \frac{1}{N} \sum_{k \subset I} \frac{1}{1 + e^{-V_k}} \tag{3}$$

where ω_k is the weight of k in appearance model, V_k is the salience value of k pixel in image. And the image detected can be written as:

$$D(k) = I(k)\omega_k \tag{4}$$

where $D(k)$ is the new pixel of k in the new image detected. It is so easy to get $p(P_{SR}|S_R)$ from detecting portion, thus we get:

$$p(S_R, P_{SR}) = p(P_{SR}|S_R)p(S_R) \tag{5}$$

2.2 Deformable Likelihood Model

We follow the framework of deformable models [9, 16, 10, 1] and describe an object by a non-rigid constellation of parts location and appearance. In order to explicitly model occlusion, we use a binary part visibility term. Each part is defined by the location of a bounding box $p_i = (p_i^l, p_i^r, p_i^t, p_i^b)$ in the image and the binary visibility state v_i. The score of a model α in the image $\mathbf{I}$, which gives model parts locations $P = (p_0, \ldots, p_n)$ and visibility states $V = (v_1, \ldots, v_n)$, $v_1 \in \{0, 1\}$, is defined as follow:

$$S(I, P, V) = \max_{c \in \{1..C\}} S(I, P, V, \alpha_c) \tag{6}$$

Pedestrian parts are always occluded because of the presence of other pedestrians and self-occlusions. The locations of occluded parts may have consistent appearance, because occlusions often do not happen at random. We use occlusions by learning separate appearance parameters A^o for occluded parts. The bias terms b_i and b_i^o control the balance between occluded and non-occluded appearance terms in S_A. One mixture component of the model has a tree T and edges structure with nodes E corresponding to object parts and relations among parts respectively.

$$S(I, P, V, \alpha_c) = \sum_{i \in T} S_A(I, p_i, v_i, \alpha_i) + \sum_{(i,j) \in E} S_D(p_i, p_j, \alpha_i) \tag{7}$$

where the unary term S_A provides appearance score using image features $\phi(I, p_i)$,

$$S_A(I, p_i, v_i, \alpha_i) = v_i(A_i.\phi(I, p_i) + b_i) + (1 - v_i)(A_i^0 .\phi(I, p_i) + b_i^0) \tag{8}$$

and the binary term S_D defines a quadratic deformation cost $S_D(p_i, p_j, \alpha_i) = d_i.\psi(p_i - p_j)$ with $\psi(p_i - p_j) = \{dx; dy; dx^2; dy^2\}$ where $dx = p_i^{x_1} -(p_j^{x_1} + \mu_{ij}^x)$ and $dy = p_i^{y_1} -(p_j^{y_1} + \mu_{ij}^y)$). Notably, the score function (6) linearly depends on the model parameters $\alpha_c = \{A_0; \ldots; A_n; A_0^o; \ldots; A_n^o; d_1; \ldots; d_n; B\}$. To represent multiple appearances of an object, our full model combines a mixture of c trees described by parameters $\alpha = \{\alpha_1; \ldots; \alpha_c\}$.

2.3 The Probabilistic Framework

As a result of substituting (5) into (1), we get the posterior:

$$p(S_R, P_{SR}|R) \propto p(R|S_R, P_{SR})p(P_{SR}|S_R)p(S_R) \tag{9}$$

More specifically, we can also get the the posterior by joint distribution. We model the joint distribution over all the random variables (S_R,P_{SR}), so we can determine probabilistic values as follow:

$$p(S_R, P_{SR}|R) = \frac{p(S_R, P_{SR}, R)}{\sum_R p(R|S_R, P_{SR})p(S_R, P_{SR})} \tag{10}$$

where $p(S_R, P_{SR}|R)$ is the probability of pedestrian in a rectangular of image, and P_{SR} is results pedestrian detection from effective salient regions.

$$p(S_R, P_{SR}, R) = p(R|S_R, P_{SR})p(S_R, P_{SR}) \tag{11}$$

As a result of substituting (5) and (11) into (10), we get the final expression for the posterior:

$$p(S_R, P_{SR}|R) = \frac{p(R|S_R, P_{SR})p(P_{SR}|S_R)p(S_R)}{\sum_R p(R|S_R, P_{SR})p(S_R, P_{SR})} \tag{12}$$

3 Experiment

Different experiments were conducted to evaluate our method. In this section, we used the INRIA still image database to compare our approach (use of feature subsets and mean features) against previous works. The proposed approach is validated by comparing with several state-of-the-art pedestrian detection methods on INRIA datasets. The database contains 1774 pedestrian positive examples

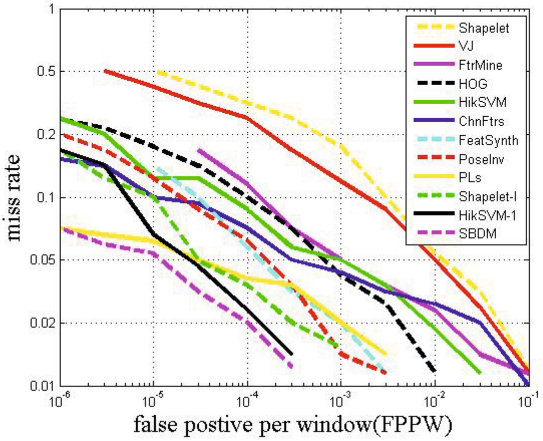

Fig. 3. Results of FPPW on the INRIA persons dataset

and 1671 negative images without pedestrian. The pedestrian annotations were scaled into a series of windows whose size is 64 × 128 and included a margin of 16 pixels around the pedestrians.

The dataset was divided into two, where 1,000 pedestrian annotations and 1,000 person-free images were selected as the training set, and 774 pedestrian annotations and 671 person-free images were selected as the test set. For each cascade level, the Logitboost algorithm [7] was trained using all the positive examples and $N_{\mathrm{n}} = 10{,}000$ negative examples generated by boostrapping. Detection on the INRIA pedestrian dataset is challenging since it includes subjects

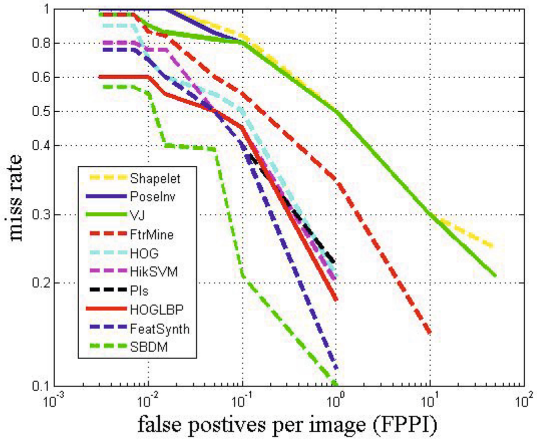

Fig. 4. Results of FPPI on the INRIA persons dataset

with a wide range of variations in pose, clothing, illumination, background, and partial occlusions.

Fig. 3 shows that the performance of our method is comparable to the state-of-the-art approaches. We compare ours experiments results with the state-of-art approaches on INRIA dataset. The x-axis corresponds to false positives per window(FPPW), the y-axis corresponds to the miss rate, and we plot the detection error tradeoff curves on a log-log scale for INRIA dataset.

Fig. 4 compares our approache with other state-of-the-art methods. Our detector is competitive in terms of the detection quality with respect to ChnFtrs [7], provides significant improvement over HOG+SVM and others. The x-axis corresponds to false positives per image (FPPI), the y-axis corresponds to the miss rate.

4 Conclusion

Previous researches mainly focus on feature extraction and sliding window, where the former aims to find robust feature representation while the latter seeks to locate the latent position. However, most of sliding windows are based on scale transformation and traverse the entire image. Therefore, it will bring computational complexity and false detection which is not necessary. To conquer the above difficulties, we propose a novel Saliency-Based Deformable Model (SBDM) method for pedestrian detection. In SBDM method we present that, besides the local features, the saliency in the image provides important constraints that are not yet well utilized. And a probabilistic framework is proposed to model the relationship between Saliency detection and the feature (Deformable Model) via a Bayesian rule to detect pedestrians in the still image.

Acknowledgement. The research was supported by the National Nature Science Foundation of China (61303114, 61231015, 61170023, 61303154, F020503), the Specialized Research Fund for the Doctoral Program of Higher Education (20130141120024), the Technology Research Project of Ministry of Public Security (2014JSYJA016), the Fundamental Research Funds for the Central Universities (2042014kf0250), the China Postdoctoral Science Foundation funded project (2013M530350), the major Science and Technology Innovation Plan of Hubei Province (2013AAA020), the Guangdong Hongkong Key Domain Breakthrough Project of China (2012A090200007), the Key Technology R&D Program of Wuhan (2013030409020109), and the Special Project on the Integration of Industry, Education and Research of Guangdong Province (2011B090400601).

References

[1] Azizpour, H., Laptev, I.: Object detection using strongly-supervised deformable part models. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part I. LNCS, vol. 7572, pp. 836–849. Springer, Heidelberg (2012)

[2] Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 409–416. IEEE (2011)

[3] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)

[4] Ding, Y., Xiao, J.: Contextual boost for pedestrian detection. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2895–2902. IEEE (2012)

[5] Dollár, P., Appel, R., Kienzle, W.: Crosstalk cascades for frame-rate pedestrian detection. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 645–659. Springer, Heidelberg (2012)

[6] Dollár, P., Belongie, S., Perona, P.: The fastest pedestrian detector in the west. In: BMVC, vol. 2, p. 7 (2010)

[7] Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(4), 743–761 (2012)

[8] Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person reidentification by symmetry-driven accumulation of local features. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2360–2367. IEEE (2010)

[9] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1627–1645 (2010)

[10] Johnson, S., Everingham, M.: Learning effective human pose estimation from inaccurate annotation. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1465–1472. IEEE (2011)

[11] Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2288–2295. IEEE (2012)

[12] Ouyang, W., Wang, X.: Single-pedestrian detection aided by multi-pedestrian detection. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3198–3205. IEEE (2013)

[13] Ouyang, W., Zeng, X., Wang, X.: Modeling mutual visibility relationship in pedestrian detection. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3222–3229. IEEE (2013)

[14] Paisitkriangkrai, S., Shen, C., Hengel, A.V.D.: Efficient pedestrian detection by directly optimizing the partial area under the roc curve. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 1057–1064. IEEE (2013)

[15] Yan, J., Lei, Z., Yi, D., Li, S.Z.: Multi-pedestrian detection in crowded scenes: A global view. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3124–3129. IEEE (2012)

[16] Yang, Y., Ramanan, D.: Articulated pose estimation with flexible mixtures-of-parts. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1385–1392. IEEE (2011)

[17] Zeng, X., Ouyang, W., Wang, X.: Multi-stage contextual deep learning for pedestrian detection. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 121–128. IEEE (2013)