Pedestrian Detection Using Objectness Information

Weidong Yao, Xiaohui Chen, Li Chen, and Weidong Wang

***Abstract*—In recent ten years, the research works about pedestrian detection using various methods are coming forth one after another, pushing the state of the art in this area. Diverse features and excellent classification technique are the key to get prime performance in these research activities. We are now motived to present a more effective approach from these two aspects. Be inspired by the research achievement in objectness estimation field, our pedestrian detection method integrates boosted classifier with multiple objectness features, including salient feature and edgebox feature. At the same time, we improve classifier structure to achieve a better performance, the influence of classifier threshold on the prediction is also analyzed in this paper. As a practice part, we extract data from several exist datasets, and append traffic scene images of our city to form a new dataset for studying about the generalization ability of pedestrian detection. Our novel approach for pedestrian detection demonstrates significant performance advantages on precision and generalization ability by a series of experiments.**

***Index Terms*—Pedestrian detection, aggregated channel features , objectness, real adaboost algorithm.**

I. Introduction

Pedestrian Detection is an attractive filed of object recognition, which has been widely used in various domains such as surveillance video, robotics assistant driving, and smart cameras. According to incomplete statistics, since HoG [1] was proposed for pedestrian detection, 1700+ methods(see Fig. 1) have been published, pushing the state of the art in this area. Existing methods of pedestrian detection can be divided into three families: DPM variants, Deep Network and Decision Forests. Arguably it is still unclear what are the key ingredients for good pedestrian detection and which architecture is optimal. DPM proposed by Felzenszwalb *et al*. [2] is a powerful model of object detection, and utilized cleverly in pedestrian detection area. Previous work on neural networks [3], [4] about pedestrian detection pays more attention to special-purpose detection problems. The method based on decision forests is another mainstream structure of pedestrian detection, which deploys fast and reliably decision forests on various and reasonable features to ensure the quality and speed of detector, and our approach proposed in this paper also falls into this category.

On the other hand, the benchmark of pedestrian dataset has also been generally improved. It is shown that the datasets usually used in articles include INRIA, ETH, TUD-Brussels, Daimler, Caltech-USA, and KITTI. The pedestrians in these datasets have different height, gesture, and color, the environment around is quite complex, at the same time, there is also a big difference about the data size of these datasets, some contain hundreds of images, while some contain tens of thousands.

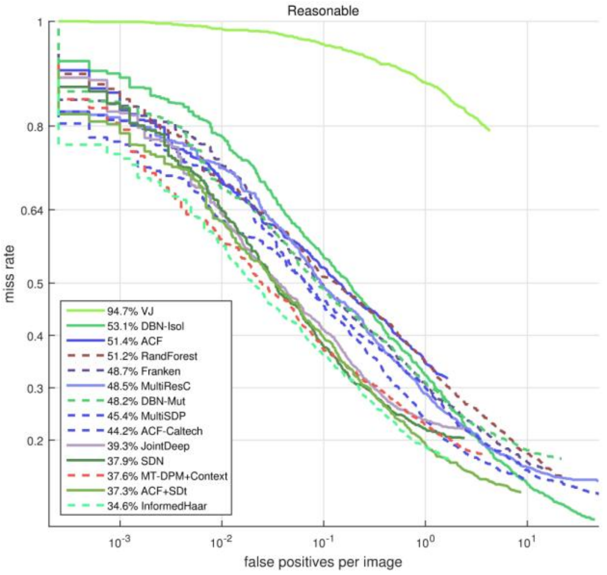


Fig. 1. The performance of various pedestrian detectors on Caltech-USA. Miss rate has been shown at 0.1 FPPI as a common reference point to compare results.

A. Related Work

Pedestrian detection is always a hot research topic in computer recognition, which enters a fast-developing period especially after 2005, datasets for training become larger and more complex, it demands that detector has higher precision and processing speed. Dalal N and Triggs B [1] studied the feature set of object detection, then they discovered the performance advantages of HoG for pedestrian detection via a series of contrast experiments between different feature descriptors about edge and gradient, as well as proposed a new method with liner SVM, which got quite good score on MIT Pedestrian Dataset. Dollar P *et al*. [5] introduced Caltech Pedestrian Dataset, and proposed ICF method based on VJ algorithm framework, then discovered the approximation among the adjacent scale of integral channel feature. Rodrigo Benenson [6], [7] introduced multi-scale pedestrian detector, whose core idea was migrating the computational work about multiscale image feature from detection phase to training phase. Costea [8] came up with integral word channel which aggregated HoG, LBP and Color feature, running at 16fps with the aid of GPU programming. On the basis of previous work, Dollar P [9] proposed ACF detection architecture, getting top performance on Caltech-USA dataset.

Surprisingly, most approaches compare their performance on specific test dataset whose training parts are the source

Manuscript received December 3, 2015; revised March 5, 2016.

The authors are with the Dept. of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China (e-mail: {yao1990, clbyx}@mail.ustc.edu.cn, {cxh, wdwang}@ustc.edu.cn).

doi: 10.18178/ijmlc.2016.6.2.578

data generating the whole detector, but paying no attention to generalization ability, which may be biased especially if the train part and test part have very close correlation. If we really hope to apply the high-quality algorithm to real-life scenario and get the performance described by their papers, generalization ability should be an important evaluation reflecting ability to adapt to environmental changes.

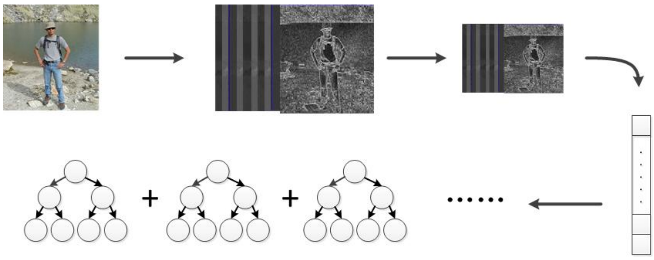


Fig. 2. ACF structure diagram. For a given image, compute channels, aggregate channels, channels vectorize, apply boosted trees.

B. Contribution

Our main result is a novel pedestrian detector, which is an extension of ACF detection architecture. Inspired by the research about object proposal, we combine salient feature and edgebox feature into our ACF architecture, and adjust the classifier structure via self-adaption of cascade threshold and cascade calibration. Rigorous test result(shown in Section V) is based on three core contributions.

Feature Selection: Common pedestrian detectors always choose low-level features as descriptor, such as LUV, HoG or gradient magnitude, which are used to draw color, texture, or margin information into features pool. These features certainly reflect the characteristics of pedestrian under particular scenario, but ignore the fact that pedestrian is also object. This may result in the condition that the detector generated by these low-level features relies too much on the specific dataset, whose train part and test part have close correlation. To avoid this situation, we select salient feature and edgebox feature to bring in objectness information, driving the detector pay more attention to objectness for getting better generalization.

Classifier Optimization: On account of the advantage about performance and speed, adaboost method gradually becomes a popular approach in pedestrian detection area. In this paper, we propose modified real adaboost to replace discrete adaboost, on the other hand, a series of research work has also been done to explore the change rules of cascade threshold and cascade calibration, which helps us come up with an adaptive scheme for different detection environment.

Dataset Extension: By comparing several common datasets, we find that the content and characteristic of these datasets have a big difference. We anew extract data from several exist datasets, and append traffic scene images to form a new dataset, called "Caltech++". We expect that this dataset can provide convenience for follow-up study about generalization ability of pedestrian detection.

The remainder of this paper is organized as follows. In the next section, we will describe the baseline detector. Section III discusses what measures we have adopted to improve detection performance. Section IV introduces our new dataset "Caltech++". Section V shows the result of our novel pedestrian detector. Section VI concludes the paper with some general considerations.

II. Baseline Detector

By considerations of the performance, speed and expandability, we choose ACF [9] as our baseline. Aggregated Channel Features(ACF) architecture can be described as Fig. 2. For an input image, ACF architecture will compute 10 channels of the given image, including gradient histogram of 6 direction, 3 color channels(LUV), and 1 normalized gradient magnitude, then sum every block of pixels in the different channel, and smooth the resulting lower resolution channels. Features are single pixel lookups in the aggregated channels. Boosting is used to learn decision trees over these features(pixels) to distinguish object from background. The entire framework employs feature pyramid to accelerate detection process. The whole detector consists of 2048 binary trees through 3 times of bootstrap. ACF detector gets 43% MR(0.1 fppi) on Caltech pedestrian dataset.

III. Improved Handling

Since ACF was proposed, several new researches have further pushed the development of this branch. Hwang [10] structured multispectral ACF, which combined the thermal channel, which further reduces the MR in night traffic scenes. Zhang [11] discovered that multiple top performing pedestrian detectors can be modelled by using an intermediate layer filtering low-level features in combination with a boosted decision forest, and discussed the influence of different filter families on performance. In this section, we will show our improvement about features and classifiers based on ACF detection architecture.

A. Feature Selection

Salient Feature: Reliable estimation of visual saliency allows appropriate processing of images without prior knowledge of their contents. Ming-Ming [12] promised a novel salient region detection method based on global contrast, which simultaneously evaluates global contrast differences and spatial coherence. In our opinion, salient feature well describes the objects which human eyes are more likely to notice. In Section V, we will prove with experiments that SalientACF shows predominant performance in pedestrian detection, especially in the scene of large scale and partial occlusion.

EdgeBox Feature: Inspired by object proposals, we realize that good features based on object proposal can improve the generalization ability of object detectors since objectness pays more attention to the general characteristics of different objects. Dollar P [13] utilizes edge information to generate object bounding box proposals using edges. And given just 1000 proposals this method achieves over 96% object recall at overlap threshold of 0.5 and over 75% recall at the more challenging overlap of 0.7. In consideration of this excellent performance, we draw EdgeBox thinking into our own detection architecture, combine with our optimized classifier, make up our own detector "EdgeBoxACF". At the same time, we remove normalized gradient magnitude from base architecture since the EdgeBox feature already has combined better edge information than gradient magnitude. In Section V, we will demonstrate the performance of EdgeBoxACF on Caltech-USA and INRIA dataset. Experiment results have shown that EdgeBoxACF has better generalization ability than common ACF, in line with our expectations (Fig. 3).

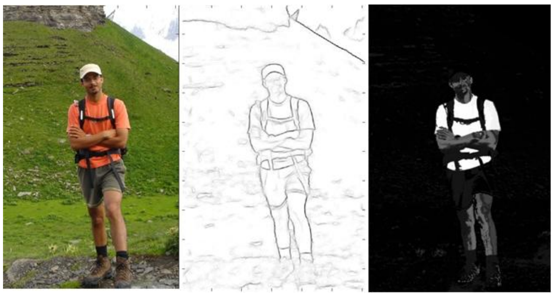


Fig. 3. New features examples. Given an input image, followed by EdgeBox feature channel and salient feature channel.

TABLE I: THE ALGORITHM OF REAL ADABOOST

Input: the training data, given as N pairs (x_i, y_i), where x_i is the attributes vector, and y_i is the desired output, either +1 or -1.

Output: A function $H(x)$ that can be used to classify an attributes vector x.

Initialization: Associate a probability $p_{(i)} = \frac{1}{N}$ with the i th example.

Iterate: The number of weak classifiers T, for $t = 1, 2, ..., T$ compute the hypothesis h_t, the function g_t, and an update to the probabilities $p_1, ..., p_N$ by the following steps:

(a) Select(at random with replacements) a subset S_t of the training examples. The i th example is selected with probability p_i.

(b) For each classifier h_j compute the following values:

$$P_r^+, P_r^-, P_w^+, P_w^-, G(j) = \sqrt{P_r^+(j)P_w^-(j)} + \sqrt{P_w^+(j)P_r^-(j)}$$

Choose as h_t a classifier h_j that minimizes $G(j)$.

If $G(t) \geqslant 0.5$ go back to Step a.

(c) Calculate the weights c_t^+, c_t^- and the function g_t as follows:

$$c_t^+ = \frac{1}{2}\ln\frac{P_t^+(t)+\epsilon}{P_W^-(t)+\epsilon},\ c_t^- = \frac{1}{2}\ln\frac{P_W^+(t)+\epsilon}{P_r^-(t)+\epsilon}$$

$$g_t(x) = \begin{cases} c_t^+ & if\ h_t(x) = 1 \\ c_t^- & if\ h_t(x) = -1 \end{cases}$$

where ϵ is a small positive number.

(d) Update the probabilities.

new pre_normalized $p_i = p_i e^{-y_i g_t(x_i)}$

normalized $p_i = \frac{pre\_normalized\ p_i}{Z_t}$

where Z_t is a normalization factor chosen so that $\sum_i p_i = 1$.

Termination:

$$H(x) = sign(\sum_{t=1}^{T} g_t(x))$$

Architecture Obviously, there are two ways to import new features into ACF architecture: in parallel or in tandem. Considering that in parallel scene, recall rate about proposal windows will seriously influence the performance, we use parallel connection way to combine different channels. According to this thought, we afresh organize the structure of SalientACF and EdgeBoxACF in more reasonable way.

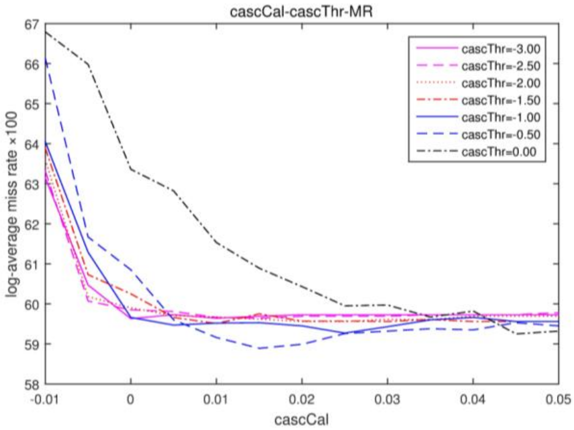


Fig. 4. MR performance on INRIA with the changing of cascCal and cascThr.

B. Classifier Optimization

RealAdaboost: Despite discrete adaboost has obtained quite good performance, we still attempt to use stronger method. Here we choose real adaboost [14] as base structure, and modify iteration rules of parameters. Our new real adaboost can be described as Table I. Our whole detector consists of 4096 binary trees through 4 times of bootstrap.

Self-adapting Parameters: In the process of training our classifier, we use two parameters to adjust threshold: cascThr(constant cascade threshold) and cascCal(cascade calibration), excessive threshold and calibration will seriously influence precision and processing speed. By experiments and comparison, we find that there is a great influence on pedestrian detection when cascThr and cascCal are assigned different values(see Fig. 4), unsuitable threshold and calibration may result in 10% performance loss. Generally speaking, larger cascThr values may lead to quite good precision, but much slower speed. Therefore, appropriate cascThr and cascCal values make a big contribution to actual performance. At the same time, in our actual experiments, the optimal values of these two parameters also have a big difference for disparate datasets. We further analyze the deep-seated reasons for these differences and discover the internal relation between parameter values and image characteristics(pedestrian size, image quality, etc.). Based on this discovery, a new strategy of adaptive threshold is used in our detection structure, automatically adjusting values in accordance with different conditions.

IV. NEW DATASET

By comparing several common datasets, we find that the content and characteristic of these datasets have a big difference. Many factors, such as capture device, actual scene, illumination intensity, etc., may cause this difference. It's really difficulty to structure a new dataset containing all kinds of pedestrians and scenes. Despite scenes differ from each other, the characteristic of pedestrians are quite semblable. There is still distinction of complexion, costume, etc. for different pedestrians, but for classifier, the pedestrian differences are much smaller than different scenes. According to this thought, we anew extract data from several part of exist datasets, and append traffic scene images of Hefei, China to form a new dataset, called "Caltech++". We expect that this dataset can provide convenience for follow-up study about generalization ability of pedestrian detection.

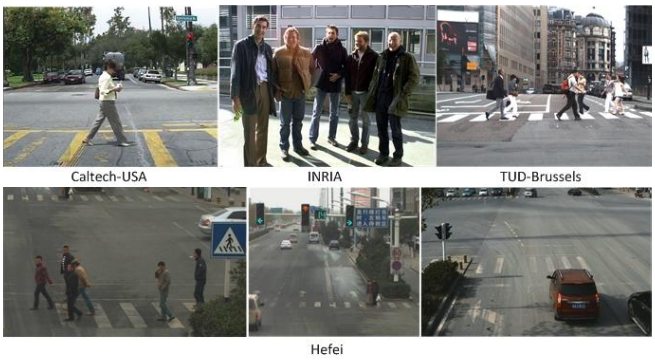


Fig. 5. Caltech++ DataSet.

Caltech++ dataset(see Fig. 5) consists of 4 parts. Caltech10× dataset(16310 pedestrians over 42500 frames in training, 10140 pedestrians over 40240 frames in testing), as the main body of Caltech++ dataset, provides positive samples and part of negative samples. INRIA and TUD-Brussels dataset supply a small amount, but abundant negative samples. In addition, we collect 60 video clips from traffic environment in Hefei, including 20 intersections, each one contains video data at morning, noon and evening.

V. Experiments

Our experiments consist of two components, Caltech-USA test and INRIA test. It's worth mentioning that here we train different pedestrian detectors on Caltech-USA training dataset, not our Caltech++ dataset, in order to compare reasonably with the results of other methods. We use the evaluation methodology of FPPI, plotting miss rate versus false positives per-image in log-log scale(lower curves indicate better performance). We calculate the miss rate at 0.1 FPPI or 1 FPPI as a common reference point to compare results.

Caltech-USA Test We run our SalientACF and EdgeBoxACF detector on Caltech-USA test dataset. In this part, 3 kinds of contrast tests(reasonable, large scale, partial occlusion) have been done for better reflection of the performance of SalientACF and EdgeBoxACF. At the same time, we also show some results of ACF family methods on this test dataset, including ACF-Caltech and ACF+SDt, in addition, we have optimized ACF model parameters and classifier structures, the results of this detector(OurACF) also have been contained in the figure. The experiments indicate that EdgeBoxACF shows better performance for Reasonable Pedestrian(pedestrian size>50pixels) than other ACF family methods, SalientACF and OurACF also get quite good results(see Fig. 6). SalientACF shows top performance for large scale pedestrian(see Fig. 7) and partial occlusions(see Fig. 8).

INRIA Test To compare the generalization ability of our own detectors with baseline detector, we run SalientACF and EdgeBoxACF on INRIA test dataset. Because we train our test detectors(OurACF, SalientACF, EdgeBoxACF) on Caltech-USA dataset, so the performance on INRIA test dataset can objectively reveal the generalization ability. The experimental results(see Fig. 9) demonstrate the gain effect of our improved handling on generalization. Especially, EdgeBoxACF shows significant performance improvement in this part, which reflects unique generalization ability. Not only the advantage for performance curve, EdgeBoxACF detector also shows superior performance in terms of actually usage.

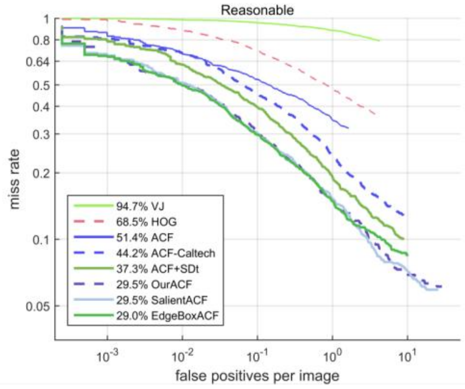


Fig. 6. Caltech-USA reasonable test.

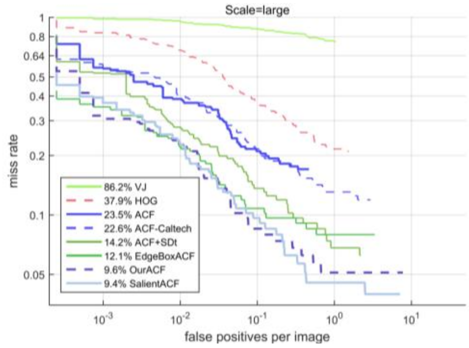


Fig. 7. Caltech-USA large scale test.

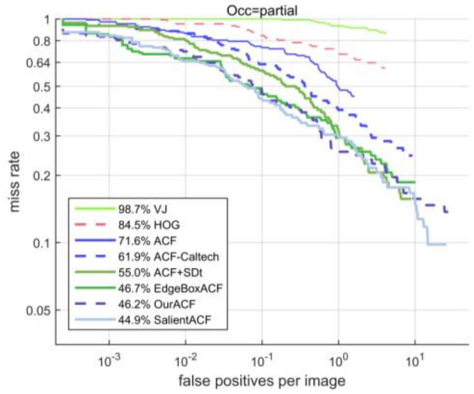


Fig. 8. Caltech-USA partial occlusion test.

Analysis Benefited by our improved measures about features and classifiers, SalientACF detector and EdgeBoxACF detector both get top performance on Caltech-USA test dataset, which shows the accurate expression ability of Salient and EdgeBox features, on the other hand, our test on INRIA test dataset(but train on Caltech-USA train dataset) confirms the generalized gain of SalientACF and EdgeBoxACF. High level Features, such as Salient Feature and EdgeBox Feature, well describe the character of objectness for pedestrians, combined with other low level features can provide more comprehensive and objective information. In addition, the amelioration of classifiers, including Real AdaBoost and threshold self-adjustment, also makes contribution to our powerful detectors with excellent generalization ability.

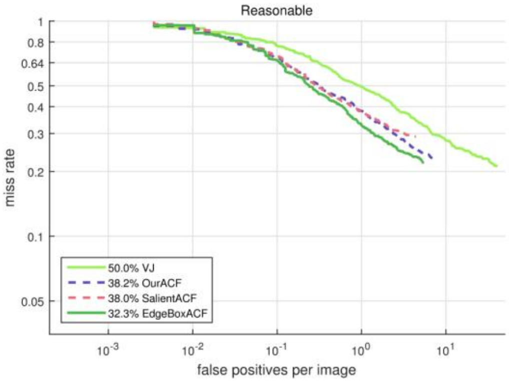


Fig. 9. INRIA reasonable test. Detectors are trained on Caltech-USA dataset, but tested on INRIA test dataset. Sorted by miss rate at 1 FPPI.

VI. Conclusions and Future Work

In order to utilize objectness property of pedestrian to improve the generalization and self-adaption ability, we propose a novel pedestrian detection architecture which respectively draws in salient and edgebox future, at the same time, we propose improved real-adaboost algorithm as classifier boosting method, and a new strategy of adaptive threshold is used in this paper, automatically adjusting values in accordance with different conditions. We have demonstrated that our new detector have excellent performance in particular scenario and better generalization ability, not just the advantage on performance curve.

In future work we plan to explore the influence of image quality(image brightness, motion blur) on pedestrian detection performance, and the deeply utilization with extra information, such as depth, movement information, etc.

References

[1] D. Navneet and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[2] F. Pedro, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[3] S. Pierre *et al.*, "Pedestrian detection with unsupervised multi-stage feature learning," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[4] W. L. Ouyang and X. G. Wang, "Joint deep learning for pedestrian detection," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2013.

[5] V. Paul, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proc. Ninth IEEE International Conference on Computer Vision*, 2003.

[6] B. Rodrigo *et al.*, "Pedestrian detection at 100 frames per second," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[7] B. Rodrigo *et al.*, "Seeking the strongest rigid detector," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[8] C. A. Daniel and S. Nedevschi, "Word channel based multiscale pedestrian detection without image resizing and using only one classifier," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[9] D. Piotr *et al.*, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1532-1545, 2014.

[10] H. Soonmin *et al.*, "Multispectral pedestrian detection: Benchmark dataset and baseline," *Integrated Computer-Aided Engineering*, vol. 20, pp. 347-360, 2013.

[11] S. S. Zhang, R. Benenson, and B. Schiele, "Filtered feature channels for pedestrian detection," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

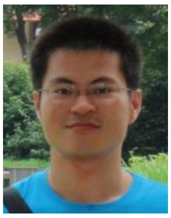
[12] C. Ming *et al.*, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569-582, 2015.

[13] Z. C. Lawrence and P. Doll ár, "Edge boxes: Locating object proposals from edges," *Computer Vision–ECCV*, Springer International Publishing, pp. 391-405, 2014.

W. Bo *et al.*, "Fast rotation invariant multi-view face detection based on real adaboost," in *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

Weidong Yao received the B.S. degree in communication and information engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2009. He is currently working toward the MS degree with the Wireless Information Network Laboratory, USTC. His current research interests include image processing, machine learning and video transmission.

Xiaohui Chen received the BS and MS degrees in communication and information engineering from University of Science & Technology of China (USTC), Hefei, China, in 1998 and in 2004, respectively. He is currently a lecturer at the Department of Electronic Engineering and Information System, USTC. His current research interests include wireless network QoS, mobile computing, MAC protocol, and traffic model.



Li Chen received the B.S. degree in electrical engineering from Harbin Institute of Technology, Harbin, China, in 2009, and the Ph.D. degree with the Wireless Information Network Laboratory, University of Science and Technology of China, Hefei, China. His research interests include wireless communications, with an emphasis on cooperative communications and energy-efficient communication.

Weidong Wang received the BS degree from Beijing University of Aeronautics & Astronautics, China, in 1989, and the MS degree from University of Science & Technology of China (USTC), in 1993. Currently, he is a full professor in the Department of Electronic Engineering and Information System, USTC. He is a member of Committee of Optoelectronic Technology, Chinese Society of Astronautics. His research interests include wireless communication, microwave and millimeter, radar technology.