A kind of Rapid Robust Human Detection Algorithm for the Interested Area

Chen Xiaoting, Shen Yuan, Yin Jing, Liu Dongcai, Zhao Xiaohu, Chang Feng

Hefei Normal University, Electronic and Information Engineering College, Hefei, 230000

Abstract: A kind of rapid two-level human detection algorithm based on the interested area was proposed in the paper, the coarse-level classifier was obtained via using the cascade training of multi-scale direction feature and Adaboost to extract the possibly human interested area, the decomposition and dimension reduction were realized via nonnegative matrix, and the précised classifier was obtained via the cascade training of dimension reduction feature and intersection kernel support vector machine to detect the human precisely. The experiment on INRIA public test set showed that the algorithm proposed in the paper improved the detection accuracy and greatly reduced the detection time when compared with the currently representative algorithm.

Keywords: Human detection, Expanded multi-scale direction feature, Nonnegative matrix decomposition, Intersection kernel support vector machine

Introduction

Human detection is to orient the pedestrian in video sequence or static image, it has become the hot research topic on computer vision, and also the difficulty due to the rigid and soft features of human and influenced by the posture, appearance, light, shade and other factors[1-4].

The object[5] was described via Haar small potter feature at the early extraction of human feature, while it is easily affected by the light, posture and perspective. Dalal, etc.[6] proposed to extract the object contour feature with HOG, it had certain robustness on the local deformation, light, color change and others, while the feature dimension was huge and the computing complexity was high. Felzenszwalb, etc.[7] proposed to use the deformable component model, the method can better solve the shade problem, while the computing amount when matching the components was huge, thus leading to the low detection efficiency. Wang, etc.[8] proposed to solve the human shade problem via combining HOG and LBP (Local Binary Pattern), it still had the shortcomings of huge feature dimension and high computing complexity.

The expanded MSO feature was proposed in the paper. EMSO feature was obtained via adding the features in two directions based on MSO feature, which can describe the human contour information, and reduce the computing amount with the lower feature dimensions than others. The coarse classifier was trained via the cascade of EMSO and Adaboost, NMF-MHOG feature was obtained via reducing the dimension of Multi-Scale HOG feature, and the précised classifier was trained after combined with IKSVM. The experiment in the public test set of INRIA and TUD-Brussels indicated that the detection speed and accuracy in the paper were improved greatly than those in the literature [6-8], and it can be applicable to the scale change, complex background and posture change.

Pedestrian detection proposed in the paper

The detection accuracy was guaranteed with the strategy of two-level detector in the paper, and the human feature was extracted via EMSO and NMF-MHDG feature.

Extraction of EMSO feature

The featured cells of MSO are two squares in the vertical and horizontal directions [9], which can describe the upright human contour information or those with little posture change, while it can not describe the contour information effectively for the human with obvious posture change. Inspired by Extended Haar-like feature, EMSO proposed in the paper increased two featured blocks based on MSO. The EMSO was calculated as follows:

1) Divided the image into n cell blocks;

2) Divided the featured cells into four sub-blocks, and calculated the color integral values in each sub-block. The calculation for the horizontal gradient and dextral gradient were as follows respectively:

$$D_l = \left| \sum_{X \in Left\,subunit} I(X) - \sum_{X \in Right\,subunit} I(X) \right| \quad (1)$$

$$D_{dr} = \left| \sum_{X \in Downright\,subunit} I(X) - \sum_{X \in Upleft\,subunit} I(X) \right| \quad (2)$$

Wherein, I(X) was the pixel value of X point.

The vertical gradient Dd and sinistral gradient Ddl were calculated in formula (1) and (2).

3) Calculated the direction of expanded MSO cell block:

$$H_e = \arctan((D_{dl} + D_l)/(D_{dr} + D_d)) \quad (3)$$

4) The direction of quantitative block:

$$F_i = Q(H) \quad (4)$$

Wherein, Q was dispersed into $\{0,1,\ldots,8\}$, according to the continuous direction of

$0° \sim 180°$.

5) Connected the cell block direction value after the quantification to be the EMSO feature.

Extraction of NMF-MHDG feature

For the image with the given size of (W,H), the multi-scale image pyramid was consisted by K degrees, and the scale factor of $\sigma = 2^{1/k}$ was adopted for the quantification of each scale in k mutually independent level. In the paper, K=3, k=4, the image in each level was divided into the cell of 8×8, and HDG feature was extracted in the overlapped block of 2×2. The cascade of multi-scale HDG features can form the Multi-scale HDG feature, and the dimension was $(105+21+3) \times 36 = 4644$.

Although the Multi-scale HDG method can better solve the shade and scale change and other problems of human object, it has the problems of huge feature dimension and high computing complexness, it is hard to break through the shackles of real-time property. The paper used the method of dimension reduction[10] to solve these problems, and the specific steps were as follows:

1) Got the absolute value of Multi-scale HDG feature with the length of l, and converted into the matrix A of $m \times n$, including, $l = m \times n$, $m > n$;

2) NMF decomposition was made for the order r of the matrix A, r m, A=WHT, W and H were the nonnegative basis matrix and coefficient matrix of $m \times r$ and $n \times r$ respectively;

3) The vector ui at each column for W and H was normalized,

$$u_i = \frac{u_i}{\|u_i\|};$$

4) Connected all ui scales to be NMF-MHOG feature.

Experiment result and analysis

1200 positive samples and 1200 negative samples were trained and used, all positive samples came from the public detection set INRIA Person Dataset, all negative samples were composed by INRIA Person Dataset and Caltech Pedestrian Dataset. All detection samples came from INRIA Person Dataset.

The False Positive Per Image and Miss Rate were used in the paper as the standard to evaluate and measure the algorithm, and the comparison result was shown in Figure 1. Seen from Figure 1, when FPPI was the same, the Miss Rate in the paper was lower than that in literature [6,8] significantly, while higher than that in literature [7]. When FPPI was 0.1, the Miss Rate in the paper was only 0.26, reducing by 0.24, 0.03 and 0.16 than those in literature [6-8] respectively. This was because that the method based on the interested area can locate the human more accurately, and the multi-feature description can fully describe the shaded and multi-scale human targets, the detection result was better than that in literature [6-8].
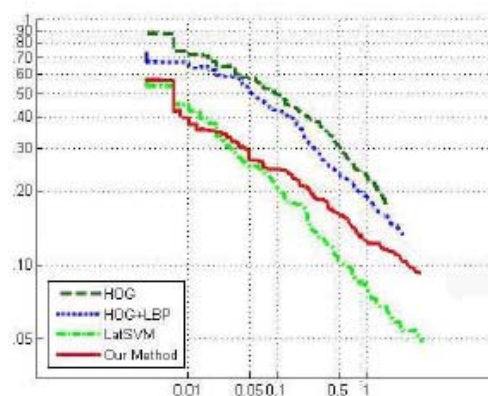

Fig.1 Performance compared with [6-8] on INRIA Person Dataset

Summary and outlook

A kind of human detection algorithm based on multi-feature and cascade classifier was proposed in the paper, and it had the following features: 1) It expanded the multi-scale direction feature (EMSO), the feature can fully describe the human edge information and reduce the false alarm; 2) It used NMF-MHOG feature, thus effectively removing the

redundant information, and reducing the computing amount; The algorithm was tested in two recognized detection sets with larger difficulty, thus proving the algorithm had improved the detection accuracy when compared with the current typical algorithm.

The shortcoming of the algorithm was removing some useful information when removing the redundant information at the dimension reduction, thus making the algorithm had a poor performance in the case with higher shade rate, high background complexness and intensive group. How to solve the problem will be the focus in the next work.

References:

[1] Wang Jinhua, Cao Jie, Li Yu, et al. Object tracking algorithm based on the feature fusion points [J]. Journal of Electronic Measurement and Instrument, 2010, 24(006):536-541.

[2] Su Songzhi, Li Shaozi, Chen Shuyuan, etc. A survey on Pedestrian Detection, Chinese Journal of Electronics, vol.40, 2012, pp.814-820.

[3] Li WT, Chang HS, Lien KC, et al. Exploring Visual and Motion Saliency for Automatic Video Object Extraction[J]. IEEE Transactions on Image Processing, 2013, 22(7): 2600-2610.

[4] Gong W G, Gui Z H, Li ZH H, et al. Face detection and tracking based on fusion AdaBoost and optical flow algorithms[J]. Chinese Journal of Scientific Instrument, 2008, 29(7):1398-1402.

[5] Mohan A, Papageorgiou C, Poggio T. Example-Based object detection in images by components[J]. IEEE Transactions. on Pattern Analysis and Machine Intelligence, 2001, 23(4):349-361.

[6] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. Proceeding of the IEEE International Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005: 886-893.

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

[8] Wang X Y, Han Tony X, Yan S C. An HOG-LBP human detector with partial occlusion handling[C]. Proceeding of the IEEE International Conference on Computer Vision, Kyoto, 2009: 32 – 39.

[9] Ye Q, Jiao J, Zhang B. Fast pedestrian detection with multi-scale orientation features and two-stage classifiers[C]. Proceeding of the IEEE International Conference on Image Processing, Hong Kong, 2010:881-884.

[10] Li Z, Liu J, Lu H. Structure Preserving Non-negative Matrix Factorization for Dimensionality Reduction[J]. Computer Vision and Image Understanding, 2013,9(117): 1175-1189.