FAST PEDESTRIAN DETECTION WITH MULTI-SCALE ORIENTATION FEATURES AND TWO-STAGE CLASSIFIERS

Qixiang Ye1 Jianbin Jiao1 Baochang Zhang2
1Graduate University of Chinese Academy of Sciences, Beijing, China
2School of Automation Science and Electrical Engineering, BUAA, Beijing, China
{Qxye, Jiaojb}@gucas.ac.cn

ABSTRACT

In this paper, we propose an approach for fast pedestrian detection in images. Inspired by the histogram of oriented gradient (HOG) features, a set of multi-scale orientation (MSO) features are proposed as the feature representation. The features are extracted on square image blocks of various sizes (called units), containing coarse and fine features in which coarse ones are the unit orientations and fine ones are the pixel orientation histograms of the unit. A cascade of Adaboost is employed to train classifiers on the coarse features, aiming to high detection speed. A greedy searching algorithm is employed to select fine features, which are input into SVMs to train the fine classifiers, aiming to high detection accuracy. Experiments report that our approach obtains state-of-art results with 12.4 times faster than the SVM+HOG method.

Index Terms— Pedestrian detection, Orientation features, Cascade classifier

1. INTRODUCTION

Pedestrian detection in images has been a hot research topic for its variety of applications, such as visual surveillance, robot vision, and smart room [2-5] etc.

In the past years, the most representative object detection method in images is about Haar-like features and cascade classifiers, e.g. Adaboost, which can obtain excellent results when the object pattern has small variance [1]. SVM is another representative method, especially when the object has large pattern variation [12]. Despite of the success of the above methods, the Haar-like features are not very competent for pedestrian representation when appearances vary with cloth, view and deformation, etc.

In recent works, contour features, like histogram of oriented gradient (HOG) features [3], are widely investigated for pedestrian detection. HOG features are the orientation statistics of image pixels in defined blocks and are robust to grey and view variance and some local deformations. Besides the HOG features, LBP [4], HOG-LBP [5] and Edgelet [6] are also representative local contour features, which are usually combined with a cascade Adaboost, feed-forward classifiers [14] or a SVM classifier for pedestrian detection.

Despite of the success of above contour features, they are mostly in fixed scale and do not consider the multi-scale characteristics of the pedestrian pattern. Although multi-scale orientation features, like variable-size HOG (v-HOG) [7], Gabor wavelet features [8], Shapelet features [9] and COV [10] features etc., have been developed, the problem of how to select and organize the orientation features in proper location and scale need to be further investigated. On the other hand, lots of state-of-art methods, like HOG+SVM [3], LBP+SVM [4], Gabor wavelet +SVM [8] methods, suffer from high computation cost of either features or classifiers.

This paper proposes a kind of new features for pedestrian representation. We first select square image blocks of different scales as units. Then, orientation features on the selected units are extracted as features. In the detection procedure, two-stage classifiers are constructed for pedestrian detection, Adaboost classifiers are set in front, followed by SVM classifiers. In both feature extraction and classifier design, the efficiency and accuracy problems are comprehensively considered.

The main difference between MSO features and the well known HOG features are as follows. 1) MSO features consist of coarse and fine features. 2) MSO features are multi-scale in both feature units and pixel orientations, while HOG is in fixed size block. 3) The spatial locations of MSO units are determined by the feature selection while those of HOG are fixed.

The rest of this paper is organized as follows. The feature extraction procedure is presented in section 2, feature selection and classifier construction in section 3. We present the experiments in section 4 and conclusions in section 5.

2. MSO FEATURE EXTRACTION

The MSO features are deduced from Haar-like and HOG features, by taking the advantages of them. We use square blocks as units instead of rectangle blocks for the reason that coarse orientation can only be calculated on square blocks. Another reason is that any rectangle can be assembled with two or more square blocks. The fine MSO
features are the pixel orientation statistics in the basic square units, which are in multi-scales as shown in Fig.1.

Fig.1. MSO features, (a) Multi-scale units at a location, (b) a coarse feature, (c) fine features.

2.1 Coarse feature extraction

Calculation of the unit coarse feature contains two steps. 1) Calculating unit orientation, 2) quantilizing the orientation into feature bins. The unit is first divided into left-right sub-units as shown in Fig.1b, then we can calculate the unit vertical gradient with

\[ D_v = \left| \sum_{X \in \text{Left subunit}} I(X) - \sum_{X \in \text{Right subunit}} I(X) \right| \]  

where \( I(X) \) is the image color value. A similar operation is used to calculate the unit horizontal gradient \( D_h \). And then we can calculate the orientation of a unit with

\[ f(n) = Q(\arctan(D_v / D_h)) \]

where \( Q \) converts the continuous values into discrete feature bins in \{0, 1, ..., 7\}. The calculation of Eq.1 can be speeded up with “integral image” [7].

2.2 Fine feature extraction

Fine features are unit orientation histograms. To calculate the fine features, it is needed to determine the pixel's orientation firstly with a color and scale competition scheme.

Horizontal gradients of pixel \( i \) are calculated with

\[ dx = \max\left( dx^c, c = r, g \text{ or } b \right) \]

\[ dx^c = \left( \sum_{i \in \Omega_L} G_\sigma * I_c(i) - \sum_{i \in \Omega_R} G_\sigma * I_c(i) \right) \]

where \( c \) is the color components (r, g or b component), \( G_\sigma \) is a Gaussian filter with scale parameter \( \sigma \). \( \Omega_L \) and \( \Omega_R \) represent the left and right neighboring regions of pixel \( i \), respectively. Similar operations can be used to calculate the vertical gradient \( dy \). And then we can obtain the pixel orientation in scale \( \sigma \) by

\[ \theta_\sigma(i) = \arctan\left( \frac{dy}{dx} \right) \]

Existing researches have observed that the fraction of the gradient and intensity can affect the discrimination [14] as

\[ w = \frac{\Delta I}{I} \]

where \( I \) represents the initial stimulus intensity, \( w \) is a fraction. Larger the value is, larger the discrimination is between the background and foreground. By this observation, we calculate the fraction of pixel \( i \) in scale \( \sigma \) by

\[ w_\sigma(i) = \frac{\Delta I}{I} \cdot \frac{\sqrt{dy^2 + dx^2}}{\sum_{i \in T} G_\sigma * I(i)} \]

and then determine the pixel's orientation as

\[ \theta(i) = \arg \max_{\theta_\sigma(i)} \left( w_\sigma(i) \right) \]  

The orientations of continuous values in the range of \{0, 180\} can be converted into the discrete orientation feature bins in \{0, 1, ..., 7\} with a normalization procedure. Then the histogram \( h_n(u) \) of unit \( n \) can be calculated as fine features.

Given a sample, there are totally 11180 (units) \times 8 (orientations) features. It is necessary to do the feature selection to improve the efficiency and effectiveness.

3. FEATURE SELECTION AND CLASSIFIER CONSTRUCTION

The pedestrian detection is separated into coarse detection stage and fine one. Boosting and greedy search methods are employed respectively to select features.

3.1 Coarse feature selection

Fig.2. Unit selection for coarse features. (a) weak classifier and (b) selected units of a 10-stage Adaboost.

The feature selection is about unit selection. To select coarse features, a weak classifier is constructed, which is

\[ C(n) = \begin{cases} 0, & \text{if } T_{\text{lower}}(n) \leq f_n \leq T_{\text{upper}}(n) \\ 1, & \text{otherwise} \end{cases} \]

where \( C(n) \) is the classification result on a coarse feature \( f_n \). \( T_{\text{lower}} \) and \( T_{\text{upper}} \) are thresholds for the weak classifier determined by an exhaustive searching strategy in the learning process. The reason why we use an orientation scope as a weak classifier is that most of the local contour orientation of a pedestrian is in a scope. For example, the...
orientation of shoulder is about from 0 to 45 degree, and that of a leg is from 45 to 100 degree. Given the weak classifier, a boosting procedure is adopted to select features.

3.2 Fine feature selection

Fine feature selection is based on greedy searching algorithm. In the \( t \)th feature selection iteration, the addition of unit \( u_{i+1} \) should induce the maximum classification accuracy increase with respect to the selected feature set \( S_t \), defined as

\[
U_{i+1} = \arg \max_{u \in \mathcal{A}} \left( p(C | (S_t \cup \{u_i\}) \right) \tag{9}
\]

where \( p(\cdot) \) represents the corrected probability of two classes \( C \) (negative and positive class). Following this, we interactively select features and move them from unselected feature set \( A \) to selected feature set \( S \). The updates of the feature sets at iteration \( t+1 \) are defined as

\[
A_{t+1} = A_t \setminus u_i, \quad S_{t+1} = S_t \cup \{u_{i+1}\}. \tag{10}
\]

The feature selection process is illustrated in Fig.3.

![Feature selection process](image)

3.3 Two-stage classifiers for pedestrian detection

![Two-stage classifiers](image)

The MSO features on the selected units are extracted with coarse features for Adaboost and fine features for SVM. We train classifiers for different views (from frontal to side view, with 15 degree interval). A multi-view pedestrian detection framework is constructed. In this detection framework, the Adaboost using the selected coarse features aims to improve the detection efficiency, and the SVM classifiers using the fine features aim to improve the detection accuracy. In experiments, we found that the Adaboost with 10-15 stages is proper. A linear SVM based on selected 255 units (255 \( \times \) 8 = 2040 features) can achieve 99.85% fine classification accuracy.

When detecting pedestrians, an integral image is firstly calculated on which we calculate the coarse features and then use cascade of Adaboost classifier to locate candidates in various views, which are then inputted into the final SVM classifiers to do fine classification. Classified positives in various scales are merged to obtain final detection results.

4. EXPERIMENTS

To evaluate the performance of the proposed approach, we collected more than 3000 training positives of frontal and other views respectively [11], and 5000 negatives. Fig.5 shows some positive examples. It can be seen that the patterns of pedestrian have large variance and are influenced by their background, appearance and views.

![Positive samples](image)

The testing procedure is conducted on INRIA public dataset, which consists of a variety of pedestrian cases including different sizes, cloth-color and views [3].

![Performance and comparisons](image)

(a)

(b)

Fig.6. Performance and comparisons, (a) is comparison of four methods, and (b) is comparison of three kinds of classifiers.

Miss rate (Log value) and False Positives Per Window
(FPPW) are used to compare our approach with the existing works [3][10][13]. It can be seen in Fig.6a that the proposed approach outperforms the well-known HOG + SVM method and COV feature + LogitBoost method [10]. It should be mentioned that, the performance of our approach is lower than the reference [13] when the FPPW is smaller than $10^{-4}$. In [13], color and texture features are combined and a partial least square analysis method is used for feature dimension reduction, which brings out a higher performance. In Fig.6b, we compare three kinds of classification methods on MSO features. They are SVM, cascade of Adaboost and the preferred two-stage classifiers, which shows that the two-stage classifiers perform best.

5. CONCLUSION AND FUTURE WORKS

The multi-scale orientation features for pedestrian are proposed, which are comparable to the existing representative features. Then, a detection approach with two-stage classifiers is proposed on the features, which ensure the detection obtain a much higher speed with a state-of-art accuracy. It is found that the proposed approach is challenged by the occlusion, which usually exists in crowd scenes. This will be investigated in the future works.

6. ACKNOWLEDGEMENT

This work is supported by Major State Basic Research Development Program of China with No. 2010CB731800, National Science Foundation of China with No. 60872143.

7. REFERENCE