Pedestrian Monitoring and Identification Method Based on Multi-feature Synergetic Double-layer Composite Structure

Qing Xia[1]

1 Sanjiang University Department of engineering
Nanjing, Jiangsu, 210012, China

Abstract **— In process of the pedestrian volume statistics for public places, in order to effectively remedy the defects of low detection efficiency, high false alarm rate and insufficient timeliness caused by pedestrian occlusion and congestion to pedestrian identification, a pedestrian identification model with a double-layer composite structure is designed in this article to coordinate and extract aggregated B-Haar features and Edgelet features, wherein the upper layer of the model has a complete binary tree structure which is integrated with Haar features (called aggregated B-Haar features) improved by local binary pattern and is mainly responsible for extracting the candidate pedestrian targets to ensure relatively high detection and identification rate; the lower layer is in a four-branch serial tree structure and meanwhile adopts Edgelet features and Bayesian principle to establish the tree decision structure so as to carry out multi-feature detection for the candidate pedestrians and further judge whether the candidate targets are pedestrians or not, thus to realize low false alarm rate and ensure the timeliness of the targets. Experimental analysis shows: compared with traditional tree structures and serial-parallel structures, the pedestrian identification method based on multi-feature synergetic double-layer composite structure in this article has significant advantages in the aspects of timeliness, detection rate and false alarm rate.**

Keywords - pedestrian monitoring; identification; double-layer; multi-feature; composite structure

I. INTRODUCTION

Due to the advantages of low cost, high precision and convenient installation during application, etc., the pedestrian identification and detection method based on video picture processing becomes the research hotspot in the field of pedestrian identification [1].

Existing pedestrian detection algorithms mainly include the pedestrian monitoring and identification methods based on background difference [2], frame difference [3], template matching [4], light stream [5], etc. Therein, the machine learning based method is currently a hot topic [6-8] in this field, but the machine learning based classification and detection algorithm has strict requirement for the establishment of the classification and detection model in practical application, because single classification model usually has low feature classification efficiency. At present, several simple classification models are usually combined in series or in parallel, but such combined model is still difficult to realize the real-time application in complicated environment due to its extremely low efficiency [9]. In order to remedy above deficiencies, a current popular solution is to combine serial model and parallel model to form a cascade model [10-11], but no implementation strategy is practicably provided for the feature extraction in different scenarios and accordingly the multi-feature synergetic detection cannot be achieved.

Meanwhile, only the recursive serial strategy is practicably implemented for the combined detection, without the detection strategy from “rough detection” to “fine detection”, namely from the human body detection to key part detection. As a result, such pedestrian identification excessively depends on the detection algorithm, thus to be unfavourable for the technology extension in the field of pedestrian volume statistics.

Therefore, in order to solve the efficiency problems in pedestrian identification models, an efficient double-layer composite identification model is established in this article on the basis of referring to the multi-feature detection thought mentioned in literature [12], wherein this model adopts a progressive multi-feature synergetic detection method and meanwhile is combined with Bayesian comprehensive decision principle. Specifically, this model aims at rapidly finding the candidate pedestrian targets as well as comprehensively and accurately judging the key parts of the candidate pedestrian targets so as to finally identify whether the targets are pedestrians or not.

II. DOUBLE-LAYER TREE-STRUCTURE COMBINED CLASSIFICATION AND IDENTIFICATION MODEL

During research, the author has proposed a double-layer combined classification model (Fig. 2) based on multi-feature synergetic detection and with the basic thought as progressive detection and identification. This model mainly includes two layers of tree structures with certain differences, wherein the upper layer is a complete binary tree combined with aggregated B-Haar features and is used to extract the features of the candidate pedestrians, and the main function of this layer is to roughly screen the unshielded targets to find the candidate pedestrians and accordingly determine the possible candidate areas; the lower layer is composed of four branched serial tree structures and aims at adopting Edgelet features to extract the features of such main parts of human body as leg, torso, arm, head and shoulder, etc.; and meanwhile, this layer also needs to adopt multi-feature comprehensive identification strategy based on Bayesian decision in order to comprehensively judge the detection and identification results of the four paths and finally identify the target pedestrians.

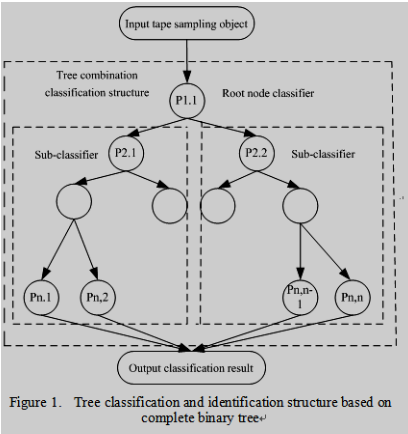


Figure 1. Tree classification and identification structure based on complete binary tree

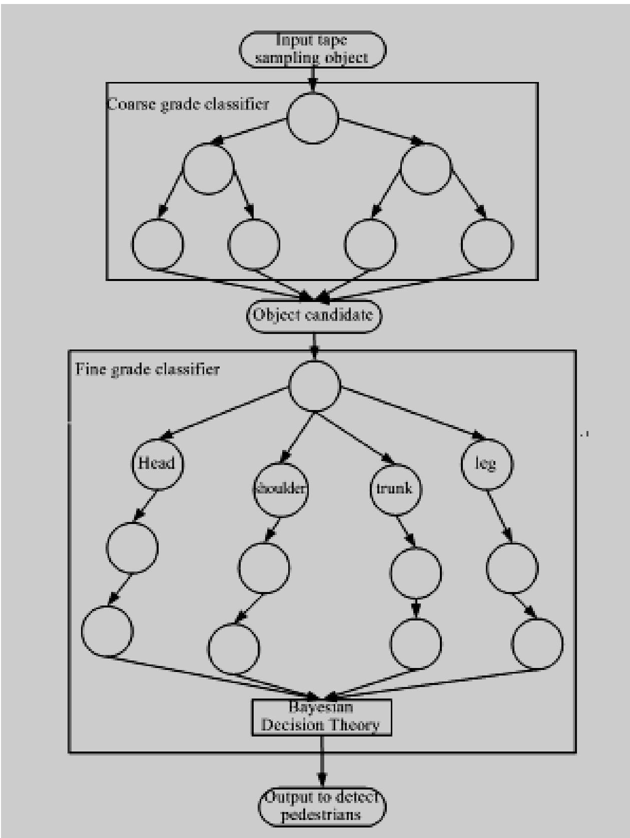


Figure 2. Multi-feature synergetic double-layer composite structure classification diagram

(1) Aggregated B-Haar features

During research, for the consideration of the robustness deficiency of Haar features, the author has adopted local binary pattern to binarize Haar features and accordingly obtain the binarized Haar features (hereinafter referred to as B-Haar features) for describing pedestrians. Due to the combination with local binary pattern thought, the illumination invariance feature among traditional Haar features is included therein.

Specifically speaking, binarized Haar-like feature value can be calculated according to the following formula (1):

$$b_j = \begin{cases} 1 & (s_1)_j - (s_2)_j > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

In the above formula, symbols $(s_1)_j$ and $(s_2)_j$ orderly denote the total brightness values of the pixels respectively in the dark area and the white area of the extracted object features. According to relevant analysis, we can find that compared with traditional Haar features, the optimized single B-Haar feature has obvious information flag bit, and the illumination invariance feature thereof will have unchanged feature information irrelevant to illumination change.

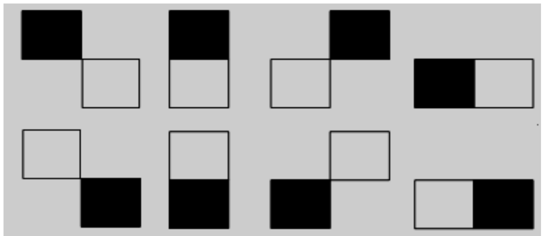


Figure 3. B-Haar feature combination

However, the pedestrian identification system may not meet the robustness precision requirement due to extremely indistinctive single B-Haar feature. Therefore, in order to improve the identification capability of single B-Haar feature, it is necessary to combine a group of B-Haar features. Meanwhile, local binary pattern principle is also introduced into B-Haar feature combination in order to continuously increase the robustness of the combined features and accordingly obtain the B-Haar feature set with 8 structures, as shown in Fig. 3. Such features are called as aggregated B-Haar features and the feature value calculation is as shown in Fig. 4.

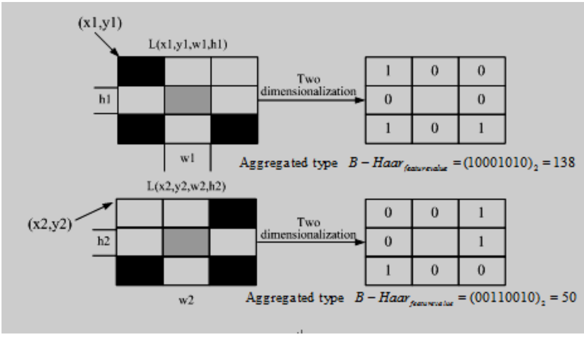


Figure 4. Aggregated B-Haar feature value calculation example

In form, the aggregated B-Haar features can be expressed by quaternion function, namely $L(\mathrm{x}, \mathrm{y}, \mathrm{w}, \mathrm{h})$, wherein (x, y) denotes the coordinate value of the leftmost upper point in the aggregated model and (w, h) denotes the length and the width of the rectangle of a single B-Haar. B-Haar feature value is within the interval of $[0, 255]$, and each value stands for a specific aggregated B-Haar feature form. Compared with traditional Haar-like features, the aggregated B-Haar features have relatively high calculation complexity, but only a few of features are needed to describe the pedestrian, thus to partly balance the calculation quantity.

(2) Edgelet features

Edgelet features have certain advantages and can identify the shielded pedestrians. This algorithm needs to match the edges of the similar shapes in the picture, so if the pedestrian identification range is not properly narrowed down in advance or the prior knowledge is not added therein, the timeliness of the system will be severely influenced due to large calculation quantity. In the multi-feature synergetic detection explained by the author, relatively good robustness can be obtained by taking Edgelet as the sub-features for pedestrian detection process. Specifically, the aggregated B-Haar features are firstly adopted in order to locate the pedestrians as fast as possible and accordingly narrow down the pedestrian detection area, then Edgelet features are adopted to extract the features of the key parts of human body, thus to complete pedestrian detection. Please refer to literatures for the specific implementation process of Edgelet features.

A. *Multi-Part Comprehensive Identification Strategy of Bayesian Decision*

After the above operations are completed, Bayesian decision strategy is adopted to comprehensively judge the confidence coefficient of each part so as to judge whether the targets are pedestrians or not.

As mentioned in literature, Bayesian decision algorithm is adopted to comprehensively judge the key parts of the candidates, but this algorithm can be used to analyze only three key parts of human body, so this algorithm is inapplicable to the complicated scenarios with large pedestrian volume and severe occlusion. Meanwhile, not being combined with double-layer tree structure, this algorithm has insufficient timeliness in practical application. In the Bayesian decision based multi-part detection strategy designed in the article, human body is divided into four parts, namely torso, arm, head and shoulder, leg. According to the designated prior knowledge regarding the physiological structure of human body, the occurrence of different parts in corresponding area can be realized, thus to significantly narrow down the detection range and meanwhile ensure the timeliness of the system. Subsequently, the likelihood probability is adopted to judge the detection result, and the specific implementation process is as follows:

Among the pedestrians detected by the aggregated B-Haar features, the existence and detection possibility of each part is extracted in Edgelet behavior features, as shown in following formula (2):

$$P\{y_{hs}=1, y_{arm}=1, y_{torso}=1, y_{leg}=1, y_{fb}=1, X\} \quad (2)$$

In the formula, $y_{hs}=1, y_{arm}=1, y_{torso}=1, y_{leg}=1, y_{fb}=1$ orderly denote that head and shoulder, arm, torso and leg can be identified; $y_{fb}=1$ denotes the candidate pedestrian target identified by the aggregated B-Haar features; letter X denotes the candidate pedestrian identification area and is mainly divided into four areas according to prior knowledge, namely X_{hs} (head and shoulder), X_{arm} (arm), X_{torso} (torso) and X_{leg} (leg); identifiers are independently provided for detecting each part, and the following formula (3) can be obtained according to the above formula (2):

$$\begin{aligned} &P\{y_{hs}=1, y_{arm}=1, y_{torso}=1, y_{leg}=1 \\ &| y_{fb}=1, X_{hs}, X_{arm}, X_{torso}, X_{leg}\} \\ &= P\{y_{hs}=1 | X_{hs}\} \bullet P\{y_{arm}=1 | X_{arm}\} \\ &\bullet P\{y_{torso}=1 | X_{torso}\} \bullet P\{y_{leg}=1 | X_{leg}\} \end{aligned} \quad (3)$$

If $F_i(X_i)$ is assumed as the response value of the identifier for detecting part i ($i \in \{hs, arm, torso, leg\}$), then the posterior probability of each part can be obtained, specifically as shown in the following formulae (4) and (5):

$$P\{y_i=1 | X_i\} = \frac{e^{F_i(X_i)}}{e^{-F_i(X_i)} + eF_i(X_i)} \quad (4)$$

$$P\{y_i=0 | X_i\} = \frac{1}{e^{-F_i(X_i)} + e^{F_i(X_i)}} \quad (5)$$

Formulae (4) and (5) respectively denote the posterior probabilities of identification and failed identification of the parts of human body, the possibility of identifying all parts of a candidate pedestrian target can be expressed by the following formula (6):

$$P\{y_{hs}=1, y_{arm}=1, y_{torso}=1, y_{leg}=1 \mid y_{fb}=1, X\} = \frac{e^{\sum_{i\in\{hs,arm,torso,leg\}} F_i(X_i)}}{\prod_{i\in\{hs,arm,torso,leg\}} e^{-F_i(X_i)}+e^{F_i(X_i)}} \tag{6}$$

If $c_i \in \{hs, arm, torso, leg\}$ is true, then the following probability formula (7) can be established:

$$\begin{aligned} &P\{y_{hs}=hs, y_{arm}=arm, y_{torso}=torso, y_{leg}=leg \\ &\quad \mid y_{fb}=1, X\} \\ &= \frac{e^{\sum_{i\in\{hs,arm,torso,leg\}} F_i(X_i)}}{\prod_{i\in\{hs,arm,torso,leg\}} e^{-F_i(X_i)}+e^{F_i(X_i)}} \end{aligned} \tag{7}$$

In formula (7), ω_i can be calculated according to the following formula:

$$\omega_i = \begin{cases} e^{-\{[\frac{(x_{si}-x_{dj})^2}{\sigma_{xi}^2}+\frac{(y_{si}-y_{di})^2}{\sigma_{yi}^2}]+\mu_i[\frac{H_{si}-H_{di}}{\sigma_{Hi}^2}+\frac{(W_{si}-W_{di})^2}{\sigma_{Wi}^2}]\}/2} & c_i=1 \\ 0 & c_i=0 \end{cases} \tag{8}$$

In formula (8), (W_{si}, H_{si}), (W_{di}, H_{di}), (x_{si}, y_{si}) and (x_{di}, y_{di}) respectively denote the size of actually identified part, the size of the assumed part of the pedestrian, the mass center of actually identified part and the mass center of the assumed part of the pedestrian; μ_i is a weight value and denotes the significance of the identification of different key parts.

The possibility of judging that the candidate target is a real pedestrian can be specifically expressed by the following formula:

$$\begin{aligned} &P\{y=1|X\} = P\{y_{fb}=1, y_{hs}=c_{hs}, y_{arm}=c_{arm}, \\ &y_{leg}=c_{leg} \mid X\} = P\{y_{hs}=c_{hs}, y_{arm}=c_{arm}, y_{torso}=c_{torso}, \\ &y_{leg}=c_{leg} \mid y_{fb}=1, X\} \bullet P\{y_{fb}=1 \mid X\} \\ &= \frac{e^{\sum_{i\in\{hs,arm,torso,leg\}} \omega_i F_i(X_i)+F_{fb}(X)}}{\prod_{i\in\{hs,arm,torso,leg\}} e^{-F_i(X_i)}+e^{F_i(X_i)}} \end{aligned} \tag{9}$$

The possibility of judging that the candidate target is not a pedestrian can be expressed by the following formula (10):

$$\begin{aligned} &P\{y=0 \mid X\} = P\{y_{fb}=0, y_{hs}=0, \\ &y_{arm}=0, y_{leg}=0 \mid X\} = P\{y_{hs}=0, \\ &y_{arm}=0, y_{torso}=0, y_{leg}=0 \mid y_{fb}=0, X\} \\ &\bullet P\{y_{fb}=0 \mid X\} \\ &= \frac{1}{\prod_{i\in\{hs,arm,torso,leg\}} e^{-F_i(X_i)}+e^{F_i(X_i)}} \end{aligned} \tag{10}$$

The confidence coefficient of the candidate pedestrian is the likelihood ratio between the possibility that the candidate target is a real pedestrian and the possibility that the candidate target is not a pedestrian:

$$\begin{aligned} l(X) &= \log\frac{P\{y=1 \mid X\}}{P\{y=0 \mid X\}} \\ &= \sum_{i\in\{hs,arm,torso,leg\}} w_i F_i(X_i) + F_{fb}(X) \end{aligned} \tag{11}$$

$F_i(X_i)$ is normalized to obtain the following formula (12):

$$F_i(X_i) = \frac{F_i(X_i) - T_i}{TR_i - T_i}, \quad i \in \{hs, arm, torso, leg\} \tag{12}$$

In formula (12), T_i denotes the classification threshold value of the identifier for part detection; TR_i is the weight value sum of the positive samples of the weak classifier of the identifier for part detection.

$$TR_i = \sum_{j=1}^{M_i} a_{ij}, \quad i \in \{hs, arm, torso, leg\} \tag{13}$$

In formula (13), M_j denotes the quantity of weak classifiers and $F_i(X_i)$ denotes the edge of the response value of part i to the decision surface. The final output of the identification model can be expressed by the following formula (14):

$$H(X) = \begin{cases} 1 & l(X) > T \quad pedestriam \\ 0 & otherwise \; nonpedestriam \end{cases} \tag{14}$$

Therein, T denotes the threshold value of the confidence coefficient of the candidate pedestrian and is within the interval of $[0.6, 0.8]$, and shall be determined according to experiment results in specific practical application.

B. Implementation Process of Double-Layer Combined Classification and Identification Model

As shown in Fig. 2, according to the progressive identification thoughts, the double-layer combined classification and identification model includes rough sub-model and fine sub-model, wherein *Edgelet* features are adopted for the four branches of the fine sub-model to identify the main parts of the candidate object and further comprehensively analyze and judge the identification results of the four branches through Bayesian decision, thus to eliminate the negative samples of possible pedestrians. Few candidate objects need to be processed, so the running time of the single classifier of the fine local identification sub-model can be prolonged in order to maximally ensure relatively low false alarm rate but relatively high pedestrian detection rate.

The specific process of adopting the double-layer combined classification model to identify pedestrians is as shown in Fig. 5.

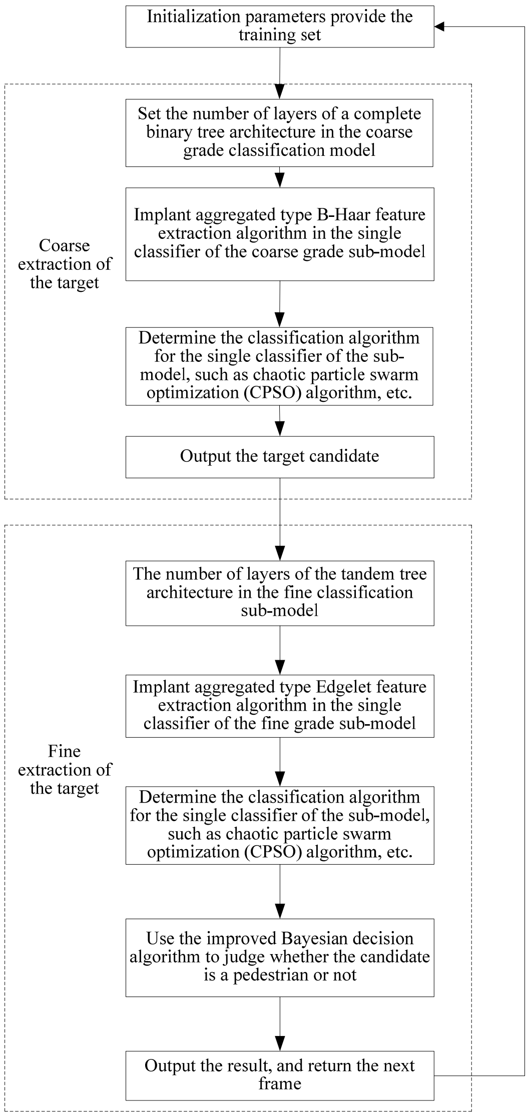


Figure 5. Aggregated multi-feature synergetic double-layer identification and detection flow chart

III. EXPERIMENT RESULT AND ANALYSIS

In order to verify the feasibility and the effectiveness of the method illustrated by the author, it is necessary to execute the following steps: firstly, determine the source of the experiment samples, namely open source for pedestrian sample set library and meanwhile add a lot of shot pictures stored in standard library format; secondly, determine corresponding classification algorithm and take it as the classification algorithm of the single classifier in the double-layer combined monitoring and identification method; thirdly, analyze and summarize the detection result.

A. Selection of Sample Set and Classification Algorithm

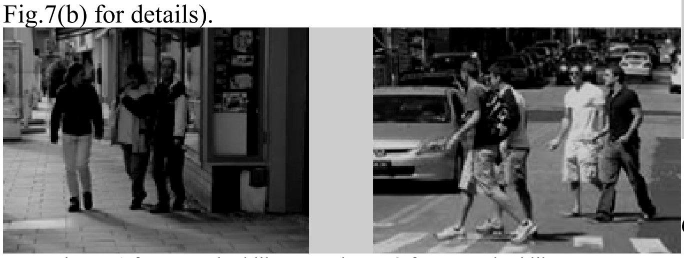
Select the experiment sample library from the open source for pedestrian sample set library ($INRIA$, MIT, $Daimler$), clip the pedestrian pictures in the positive sample set to obtain the part pictures and meanwhile add them to the sample set library. In this article, the author has allocated other pictures and the pedestrian pictures at the proportion of 2:3 into the testing sample set and the training sample set, wherein the negative and positive samples of the testing sample set are used to evaluate the detection rate and the false alarm rate of the detection results.

In consideration of the disequilibrium between the quantity of pedestrian samples and the quantity of other samples, the algorithm search capability, the stability, etc. During the sample classification process, the improved chaos particle swarm optimization algorithm based on T-distribution variation function in literature is taken as the single classifier algorithm in order to solve the above mentioned disequilibrium problem through setting price sensitivity parameter and improve the global search capability of particles through T-distribution variation function and chaos algorithm. Moreover, the optimal compromise of the correct classification of positive and negative samples is taken as the optimization guiding thought, and parameters are dynamically optimized within the price sensitivity weight value interval. The advantages of the above algorithms are relatively consistent with the expected features of the single classifier algorithm.

B. Experiment calculation Result Analysis

This section mainly aims at verifying the effectiveness of the double-layer combined classification and monitoring structure for the statistical identification of pedestrians. The steps are as follows: firstly, determine and train a certain quantity of positive and negative samples (from pedestrian sample set database); then, combine the single classifiers to form various identification structures, wherein the rough identifiers mainly adopt the aggregated B-Haar features to extract pedestrian features while the fine local identifiers mainly adopt Edgelet features to extract the features of the key parts of human body and meanwhile combine the Bayesian decision for identification. Namely, such combined model is used for pedestrian identification. Additionally, the method illustrated in literature is used as the algorithm for each single classifier.

The experiment results of the identification strategy proposed in this article are as shown in Fig. 6 and Fig. 7. The two figures both include the pictures from the standard library and the pictures shot by mobile phone, and these pictures are relatively classic traffic scenario pictures (including many pedestrian). According to the two figures, for the rough pedestrian identifier based on complete binary tree structure, the detection result thereof (refer to Fig. 6(a) and Fig. 6(b) for the details) has relatively high identification rate but fairly large false detection rate. According to relevant analysis, we can know that the false detection rate is gradually reduced along with the screening in the serial tree structure at the second layer for fine identification, so the relatively satisfactory pedestrian identification results are obtained (refer to Fig. 7(a) and Fig.7(b) for details).



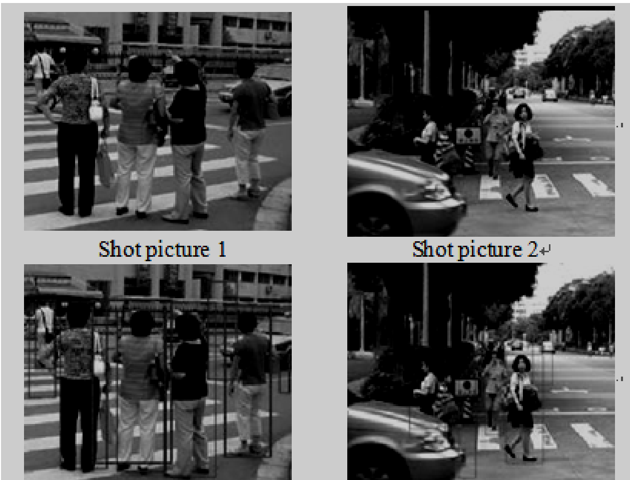
Picture 1 from standard library   Picture 2 from standard library

Experiment result 1 of rough detection   Experiment result 2 of rough detection

Figure 6(a) Rough Detection of Pictures from Standard Library

Experiment result 1 of Fine Detection   Experiment result 2 of Fine Detection

Figure 6(b) Fine detection of pictures from standard library



Rough detection for shot picture 1   Rough detection for shot picture 2

Figure 7(a) Experiment result of rough detection for shot pictures

Fine Detection for shot picture 1 Fine Detection for shot picture 2

Figure 7(b) Experiment result of fine detection for shot pictures

C. *Contrastive Analysis of Detection and Identification Results*

TABLE I COMPARISON OF DETECTION RESULTS OF DIFFERENT METHODS

Method	ADR	AFAR	ADT
Method 1	92.8%	9.1%	0.092 (frame/s)
Method 2	89.7%	8.8%	0.142 (frame/s)
Method 3	91.7%	9.4%	1.21 (frame/s)

In this section, the double-layer combined strategy designed in this article is compared with the combined classification method mentioned in the representative literatures , and the specific detection results are as shown in Table 1. Methods 1, 2 and 3 are respectively corresponding to the double-layer combined classifier model designed in this article, the hierarchical pedestrian detection structure and the serial-parallel combined pedestrian detection structure respectively illustrated in literatures. According to the analysis of average detection rate (abbreviated as ADR), average false alarm rate (abbreviated as AFAR) and average detection time (abbreviated as ADT) in Table 1, we can find: among the three typical methods, method 1 has best timeliness and detection rate; although its false alarm rate is 0.3% higher than that of method 2, its average detection rate is 3% higher than that of method 2; therefore, method 1 has better comprehensive performance than method 2. Although method 3 has relatively high detection rate and acceptable false alarm rate, it still has an obvious disadvantage, namely poor timeliness, and meanwhile it is greatly difficult to apply method 3 in specific detection environment. Obviously, the double-layer composite structure designed in this article has relatively strong comprehensive performance. If this structure is introduced into pedestrian detection system, the performance of the system will be significantly improved, thus to be greatly favorable for the market promotion of the whole detection system.

IV. CONCLUSION

The combination feature thought of literature is taken as the reference to optimize the double-layer combined identification model designed in this article for classifier, the feature extraction methods with maximum identification efficiency are respectively adopted in different stages for feature identification and extraction, and meanwhile the aggregated B-Haar features and Edgelet features are adopted to coordinate feature extraction. Furthermore, Bayesian decision principle is also adopted for comprehensive judgment to realize fine local identification and detection. According to the experimental contrastive analysis, compared with traditional tree structure and serial-parallel structure, the multi-feature synergetic double-layer combined classification structure has overall advantages in the aspects of timeliness, detection rate and false alarm rate. In the consideration that the state of the pedestrians in the picture with continuous frames is estimated in literature [25], this article mainly focuses on identifying single-frame picture, and due to the significant difference in application scenarios, the above two methods cannot be directly compared in practical applications. In subsequent research, this literature will be taken as the basis for the analytical research on video tracking algorithm so as to obtain certain innovative breakthrough on the basis of absorbing the research achievement of literature.

REFERENCES

[1] Alex Tek, Benoist Laurent, Marc Piuzzi, et al., “Matthieu chavent, marc baaden, olivier delalande et al. advances in human-protein interaction-interactive and immersive molecular simulations”. InTech, 2012.

[2] Gangyi Zhu, et al. “SciCSM: Novel Contrast Set Mining over Scientific Datasets Using Bitmap Indices”. In Proceedings of the 27th international conference on scientific and statistical database management, ACM, 2015.

[3] Guanxiong Liu, Yishuang Geng, Kaveh Pahlavan, “Effects of calibration RFID tags on performance of inertial navigation in indoor environment”, 2015 International Conference on Computing, Networking and Communications (ICNC), 2015.

[4] Jie He, Yishuang Geng, Fei Liu, Cheng Xu, “CC-KF: Enhanced TOA Performance in Multipath and NLOS Indoor Extreme Environment”, IEEE Sensor Journal, Vol. 14, No. 11, pp. 3766-3774, 2014.

[5] Jie He, Yishuang Geng, Kaveh Pahlavan, “Toward Accurate Human Tracking: Modeling Time-of-Arrival for Wireless Wearable Sensors in Multipath Environment”, IEEE Sensor Journal, Vol. 14, No. 11, pp. 3996-4006, 2014.

[6] Jie He, Yishuang Geng, Yadong Wan, Shen Li, Kaveh Pahlavan, “A cyber physical test-bed for virtualization of RF access environment for body sensor network”, IEEE Sensor Journal, Vol. 13, No. 10, pp. 3826-3836, 2013.

[7] Liguo Zhang, Binghang He, Jianguo Sun, et al., “Double image multi-encryption algorithm based on fractional chaotic time series”. Journal of Computational and Theoretical Nanoscience. 2016.

[8] Na Lu, Caiwu Lu, Zhen Yang, Yishuang Geng, Modeling Framework for Mining Lifecycle Management, Journal of Networks, Vol. 9, No. 03, pp. 719-725, 2014.

[9] MA, Ruina, Zhihan LV, Yong HAN, et al., “Research and implementation of geocoding searching and lambert projection transformation based on WebGIS”. Geospatial Information, vol. 5, pp. 013, 2009.

[10] Shuang Zhou, Liang Mi, Hao Chen, Yishuang Geng, “Building detection in Digital surface model”, 2013 IEEE International Conference on Imaging Systems and Techniques (IST), 2012.

[11] Song Yu, et al. “A Rapid and High Reliable Identify Program for Nighttime Pedestrians.infrared physics & technology”. 2015.

[12] Tianyun Su, et al. “3D seabed: 3D modeling and visualization platform for the seabed”. 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). pp. 1-6. IEEE, 2014.