



Article Pedestrian Detection under Parallel Feature Fusion Based on Choquet Integral

Rong Yang ^{1,2}, Yun Wang ^{1,2,*}, Ying Xu ^{1,*}, Li Qiu ¹ and Qiang Li ³

- ¹ College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518000, China; ryang@szu.edu.cn (R.Y.); qiuli@szu.edu.cn (L.Q.)
- ² Guangdong Key Laboratory of Electromagnetic Control and Intelligent Robots, Shenzhen University, Shenzhen 518000, China
- ³ Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Science, Shanghai 200050, China; qiang.li@mail.sim.ac.cn
- * Correspondence: wangyun@szu.edu.cn (Y.W.); yxu@szu.edu.cn (Y.X.)

Abstract: Feature-based pedestrian detection method is currently the mainstream direction to solve the problem of pedestrian detection. In this kind of method, whether the appropriate feature can be extracted is the key to the comprehensive performance of the whole pedestrian detection system. It is believed that the appearance of a pedestrian can be better captured by the combination of edge/local shape feature and texture feature. In this field, the current method is to simply concatenate HOG (histogram of oriented gradient) features and LBP (local binary pattern) features extracted from an image to produce a new feature with large dimension. This kind of method achieves better performance at the cost of increasing the number of features. In this paper, Choquet integral based on the signed fuzzy measure is introduced to fuse HOG and LBP descriptors in parallel that is expected to improve accuracy without increasing feature dimensions. The parameters needed in the whole fusion process are optimized by a training algorithm based on genetic algorithm. This architecture has three advantages. Firstly, because the fusion of HOG and LBP features is parallel, the dimensions of the new features are not increased. Secondly, the speed of feature fusion is fast, thus reducing the time of pedestrian detection. Thirdly, the new features after fusion have the advantages of HOG and LBP features, which is helpful to improve the detection accuracy. The series of experimentation with the architecture proposed in this paper reaches promising and satisfactory results.

Keywords: pedestrian detection; Choquet integral; feature fusion; genetic algorithm

1. Introduction

Pedestrian detection is the key technology of intelligent transportation [1–3]. In addition, the core technologies included in pedestrian detection are also indispensable for other applications, such as, robotics, video surveillance and behavior prediction [4–6]. In recent years, researchers have proposed many different pedestrian detection methods and successfully applied them in commercial and military fields [7–11]. Feature based pedestrian detection method is the mainstream method at present. Although they are different in the processing of raw data and the training of classifier, they basically follow the similar path as shown in Figure 1. The input of this path is the original image in the form of pixel representation, while the output includes a set of rectangular borders with different sizes. Each rectangular border corresponds to a pedestrian identified in the image. A typical pedestrian detection scheme mainly includes three steps: selection of detection region, extraction of feature and classification of detection region.

In the stage of selection of detection region, the input is usually the original image, and the output is a group of regions with different sizes and ratios. Sliding window method is the simplest method among all the region selection algorithms. It can be used to obtain regions with multiple proportions and aspect ratios. As the number of candidate



Citation: Yang, R.; Wang, Y.; Xu, Y.; Qiu, L.; Li, Q. Pedestrian Detection under Parallel Feature Fusion Based on Choquet Integral. *Symmetry* **2021**, *13*, 250. https://doi.org/10.3390/ sym13020250

Academic Editor: Theodore E. Simos Received: 20 January 2021 Accepted: 30 January 2021 Published: 2 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). regions has a great influence on the speed of the whole pedestrian detection system, more complicated approaches analyze the original images in advance to filter out the regions in which no target objects are believed to contain, and therefore the number of candidate regions to be tested is reduced.



Figure 1. Pipeline of a general feature-based pedestrian detection.

For the extraction of features, the input is the candidate region that may or may not contain a pedestrian, and the output is a feature vector in the form of real-valued or binary-valued. The criteria of which features should be extracted are whether they can classify pedestrians and non-pedestrians. Feature extraction can be clustered as single and multifeature extraction, respectively. Single features typically include HOG (histogram of oriented gradient) [12], LBP (local binary pattern) [13] and Haar-like [14], while the representatives of multifeature are HOG-LBP [15], HOG-Harr-like [16] and HOG-SIFT (scale-invariant feature transform) [17,18].

In the stage of classification of detection region, the main task is to identify whether there is a human shape in the candidate region. The feature vector obtained in the feature extraction stage for a candidate detection region is input into the classifier, and a binary label is output after classification calculation to indicate whether the area is positive (that means containing pedestrian) or negative (that means not containing pedestrian). The classifier comprises SVM (support vector machine) [19,20], AdaBoost [21,22] and CNN (convolutional neural network) [23,24].

Compared to the methods of multi-component combination, the implementation process and structure of feature-based methods of pedestrian detection is relatively simple. When using different feature methods, it does not need to change the original architecture and consequently guarantees better portability. The critical steps of the feature-based pedestrian detection pipeline mentioned above are feature extraction and region classification. Therefore, a novel and efficient feature extraction algorithm is essential, and which inspires the origin of this paper.

HOG [12] is widely considered as one of the best features to obtain edge or local shape information. It has achieved great success in target recognition and detection [25–27]. In [28], Zhu et al. integrate the cascade rejections method to accelerate the HOG extraction process in human detection. In [29], HOG-DOT algorithm with L1 normalization technique and SVM is used. Although this method can get good TPR (true positive rate), its FPR (false positive rate) is very high. HOG using discrete wavelet transform is proposed in [30], but its detection rate is not high, only 85.12%. In [31], selective gradient self-similarity (SGSS) feature is applied for feature extraction with HOG. The addition of SGSS significantly improves the accuracy of pedestrian detection, and the detection ability of cascade structure based on AdaBoost is better than linear SVM or HIKSVM.

It should be noted that the performance of HOG is poor in the background clustered with noisy edges. In such a situation, LBP [32] can play a very good complementary role. LBP feature has been widely used in different applications and has presented satisfactory performance in face recognition. It is a very effective feature to distinguish images because of its invariance to monotonic gray level changes and high efficiency of computation.

Based on the above reasons, it is natural to think that the combination of edge/local shape information with texture information can capture the appearance of pedestrians more efficiently. In [15], aiming at the problem of partial occlusion, a feature descriptor based on serial fusion of HOG and LBP features is proposed. Although the detection rate

has been improved, the detection efficiency is sacrificed owing to the increase of dimension. In [33], Jiang et al. also use concatenation method to combine HOG and LBP features to for a new feature vector, and then send it to XGBoost (eXtreme Gradient Boosting) classifier. Therefore, the problem of sacrificing detection efficiency to improve detection accuracy still exists. In the feature fusion of pedestrian detection, the current mainstream method is to concatenate several features in series. This approach may raise two problems. First, serial feature concatenation leads to an extreme increase of the number of features for the whole image being processed and consequently affects the processing speed. Secondly, the processing method of concatenating features does not consider the possible interaction between features and the possible impact of this interaction on the final classification decision. Therefore, the current feature fusion based on concatenation cannot be regarded as the feature fusion in a strict sense.

In this paper, Choquet integral based on the fuzzy measure is applied to realize the parallel fusion of HOG and LBP feature descriptors. This methodology is expected to improve the detection accuracy without increasing the feature dimension. The Choquet integral based on fuzzy measure is a very effective feature fusion method. When it is applied to the fusion problem, the fuzzy measure in the integral can well reflect the importance of each feature to the fusion target and the influence of the interaction among features on the fusion target. At the same time, it may help us to mine the possible interaction between different pedestrian descriptors, which has a positive research significance for the development of pedestrian detection technology.

The procedure of pedestrian detection based on the parallel fusion of HOG and LBP features is demonstrated in Figure 2. Here, HOG features and histogram of LBP descriptors of each cell are extracted from original image, respectively. They are parallelly fused by Choquet integral with its internal parameters, i.e., values of fuzzy measure, being optimized by a genetic algorithm. This fusion results in a new set of features, called parallel-HOG-HOLBP (histogram of gradient—histogram of local binary patterns), which is consequently transmitted to SVM for classification.



Figure 2. The demonstration of parallel fusion of HOG (histogram of oriented gradient) and LBP (local binary pattern) features.

The intervention of Choquet integral makes the two features (HOG and LBP) merge in parallel. The resulting feature, parallel-HOG-HOLBP, not only contains the original advantages of HOG and LBP, but also avoids the unavoidable dimension disaster in traditional serial fusion. Genetic algorithm is used to optimize the interval coefficients of fuzzy measure in Choquet integral. It is a more rational way to retrieve these parameters through a global optimization algorithm compared to through trial and error method. We designed a series of experiments to verify the effectiveness of the proposed method. The experimental results show that the proposed method has better comprehensive performance than the existing methods. The organizational structure of this paper is as follows. In Section 2, typical features used in this paper are presented. Aggregation of feature fusion based on Choquet integral is introduced in Section 3. An adaptive algorithm based on genetic algorithm is implemented in Section 4 to optimize the internal coefficients of fuzzy measure in Choquet integral. In Section 5, experimental results and analysis are demonstrated. Finally, Section 6 summarizes and prospects this paper.

2. Features Realignment

When using Choquet integral to fuse two kinds of features in parallel, the fused features should have the same dimension. Therefore, this section discusses the mechanism of feature realignment of HOG and LBP features.

2.1. Histogram of Oriented Gradient Feature Extraction

HOG feature calculates the distribution of gradient in local image, so it can describe the edge or local shape information of object well. A typical HOG feature extraction process includes four steps:

1. Standardize gamma space and color space.

In order to reduce the influence of illumination, the whole image needs to be normalized. In the texture intensity of the image, the local surface exposure contributes a large proportion, so this kind of compression can effectively reduce the local shadow and illumination changes of the image. As the color information has little effect, the original RGB image is usually converted to a gray image, and the gamma correction is used to normalize it by formula

$$I(x, y) = I(x, y)^{\gamma}$$
⁽¹⁾

where I(x, y) represents the intensity of the pixel at coordinates (x, y), and γ represents the parameter of gamma correction. Generally, the value of γ is set to 0.5.

2. Calculate image gradient.

The gradient of horizontal direction $G_x(x, y)$ and the gradient of vertical direction $G_y(x, y)$ are, respectively, calculated for the normalized image.

$$G_x(x, y) = I(x+1, y) - I(x-1, y)$$
(2)

$$G_{y}(x, y) = I(x, y+1) - I(x, y-1)$$
(3)

The gradient value G(x, y) and gradient direction $\theta(x, y)$ of each pixel are calculated from the gradient of the two directions, respectively.

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$$
(4)

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right)$$
(5)

3. Construct the histogram of gradient direction for each cell.

The image is divided into several cells, as shown in Figure 3a, each cell is 8×8 pixels. The gradient direction of 360 degrees is divided into nine ranges averagely (Figure 3b), and the histogram corresponding to these nine bins is constructed to count the gradient information of the 8×8 pixels. The horizontal axis of the histogram is the nine bins in gradient directions, while the height of each bin is the superposition of the gradient value of those pixels whose gradient directions belong to the bin.

Each cell gets a 9-dimensioanl vector. As shown in Figure 3a, four adjacent cells constitute a block, and the vectors of four cells in a block are connected in serial to obtain a 36-dimensional vector. The block is used to scan the image with the scanning step as a cell. Finally, the vectors of all blocks are connected in serial to get the HOG feature of the image. For example, for an 128×64 image, every 8×8 pixel constitutes a cell and every 2×2 cells constitute a block. As each cell has nine features, there are $4 \times 9 = 36$ features in each block. Taking eight pixels as the step size, there will be seven scanning windows in the horizontal direction and 15 scanning windows in the vertical direction. In other words, 128×64 images have $36 \times 7 \times 15 = 3780$ features.



Figure 3. Illustration of HOG feature extraction: (**a**) cells and blocks of HOG feature; (**b**) nine ranges of gradient direction in 360 degrees.

2.2. Histogram of LBP Descriptor

The LBP descriptor shows the difference of gray-level between a pixel in center and its neighbor pixels in a specific size region. If we denote the gray value of a pixel as I(x, y), then the LBP value of this pixel is a decimal calculated by

$$L(x, y) = \sum_{i=0}^{K-1} f(I_i(x, y) - I(x, y)) \cdot 2^i$$
(6)

where

$$f(I_i(x, y) - I(x, y)) = \begin{cases} 1, & \text{if } I_i(x, y) - I(x, y) \ge 0\\ 0, & \text{if } I_i(x, y) - I(x, y) < 0 \end{cases}$$
(7)

Here, *K* is the number of neighbor pixels around the center pixel. Figure 4 shows an LBP feature extraction process with K = 8 and radius as 1.



Figure 4. LBP feature extraction process with *K* = 8 and radius as 1.

In order to fuse HOG and LBP in parallel by Choquet integral, the two features extracted from a candidate image should have the same dimension. Referring to the construction method of HOG features, we realign the LBP feature for the candidate image. The gradient value and gradient direction of LBP value for each pixel are calculated, respectively. The gradient direction of 360 degrees is divided into nine ranges averagely and the histogram corresponding to these nine bins is constructed to count the gradient directions, while the height of each bin is the superposition of the gradient value of those pixels whose gradient directions belongs to the bin, as shown in Figure 5. Similarly, for an 128×64 image, a new feature vector with length 3780 is constructed. We called this new feature vector histogram of local binary patterns (HOLBP).



Figure 5. Histogram of local binary patterns in a cell.

3. Feature Fusion in Parallel by Choquet Integral

Since the dimensions of HOG and HOLBP are consistent, it is possible to fuse these two feature descriptors in parallel. In this paper, Choquet integral based on fuzzy measure [34,35] is utilized as an aggregation tool to perform the fusion task.

3.1. Signed Fuzzy Measure

Denote $X = \{x_1, x_2, \dots, x_n\}$ as a set of feature attributes being considered. The set of all the subsets of *X* is called the power set of *X* and is denoted by *P*(*X*).

Definition 1. A signed fuzzy measure is a set function μ : $P(X) \to (-\infty, \infty)$ with $\mu(\emptyset) = 0$.

A signed fuzzy measure μ assigns a real value for each single element and each possible combination of elements in *X*. If we regard the elements in set *X* as a set of features to be fused, then the signed fuzzy measure values corresponding to each feature and the signed fuzzy measure values corresponding to each possible combination of each feature can be clearly explained as their influence on the fusion target. Due to the non-additivity of the signed fuzzy measure, the influence of any combination of features to be fused on the fusion target is not the simple sum of their respective influences. Therefore,

the signed fuzzy measure defined on set *X* has interpretable physical meaning, indicating the possible interaction between the features to be fused.

A signed fuzzy measure has advantages of describing the individual and joint contribution rates from features to be fused toward the fusion target flexibly. A signed fuzzy measure μ is called subadditive if it satisfies $\mu(A \cup B) \leq \mu(A) + \mu(B)$ whenever $A, B \subset X$, while a signed fuzzy measure μ is called super-additive if it satisfies $\mu(A \cup B) \geq \mu(A) + \mu(B)$ whenever $A, B \subset X$ and $A \cap B = \emptyset$.

3.2. Choquet Integral as Aggregation Tool

Definition 2. Let μ be a signed fuzzy measure defined on P(X). For a real-valued function $f: X \to (-\infty, \infty)$, its Choquet integral is defined as

$$\int f d\mu = \int_{-\infty}^{0} \left[\mu(F_{\alpha}) - \mu(X) \right] d\alpha + \int_{0}^{\infty} \mu(F_{\alpha}) d\alpha$$
(8)

where $F_{\alpha} = \{x | f(x) \ge \alpha\}$ is a set whose elements have their function values greater or equal to α , $\alpha \in (-\infty, \infty)$.

The values of f for each element are denoted as $f(x_1), f(x_2), \dots, f(x_n)$. To calculate the value of a Choquet integral with a given function f, they are usually sorted in a nondecreasing order such as $f(x'_1) \leq f(x'_2) \leq \dots \leq f(x'_n)$. Here, $(x'_1, x'_2, \dots, x'_n)$ is a certain permutation of $\{x_1, x_2, \dots, x_n\}$. Then, the value of the Choquet integral can be obtained by

$$\int f d\mu = \sum_{i=1}^{n} \left[f(x'_i) - f(x'_{i-1}) \right] \cdot \mu(\{x'_i, x'_{i+1}, \cdots, x'_n\})$$
(9)

where $f(x'_0) = 0$.

In real programming, it is inconvenient to perform such a sorting operation in Equation (9). Actually, the value of the Choquet integral of f with respect to μ can be calculated as a linear form as follows.

$$\int f d\mu = \sum_{j=1}^{2^n - 1} z_j \mu_j$$
(10)

in which

$$z_{j} = \begin{cases} \min_{i:frc(\frac{j}{2^{i}}) \in [\frac{1}{2}, 1)} f(x_{i}) - \max_{i:frc(\frac{j}{2^{i}}) \in [0, \frac{1}{2})} f(x_{i}), & if it is > 0 \text{ or } j = 2^{n} - 1 \\ 0, & otherwise \end{cases}$$

$$for j = 1, 2, \dots, 2^{n} - 1.$$
(11)

The definition of Choquet integral shows that it is actually a mapping from n-dimensional space to a real value, so it is usually regarded as a powerful tool to aggregate different features, and the result of aggregation are furthermore to be used for the solution of data classification or regression problems.

3.3. Feature Fusion by Choquet Integral

For a sliding window, each corresponding singleton dual of HOG feature and HOLBP feature constructs a set of feature attributes, denoted as $X = \{x_1, x_2\}$. A real-valued function f is defined on X by assigning $f(x_1)$ and $f(x_2)$ the numerical value of the corresponding singleton of HOG feature and HOLBP feature, respectively. A signed fuzzy measure μ is defined on P(X) to describe the influence of each individual feature as well as each possible combination of features to the fusion target. Since $\mu(\emptyset) = 0$, in this case,

three values of μ are required to be set, that is, $\mu(\{x_1\}), \mu(\{x_2\} \text{ and } \mu(\{x_1, x_2\}))$. Figure 6 illustrates the process of feature fusion of HOG and HOLBP by Choquet integral.



Figure 6. Feature fusion of HOG and HOLBP by Choquet integral.

4. Pedestrian Detection Framework with Parameters Retrieved by Genetic Algorithm

To accomplish the parallel feature fusion between HOG and HOLBP via Choquet integral based on signed fuzzy measure, a series of interval parameters, i.e., $\mu(\{x_1\})$, $\mu(\{x_2\} \text{ and } \mu(\{x_1, x_2\})$, need to be retrieved. Of course, we can use trial and error method to predict the values of these parameters. However, a more scientific way is to retrieve these parameters through a global optimization algorithm.

As shown in Figure 7, genetic algorithm is an adaptive optimization algorithm which can ensure global search. It includes the process of initialization of new generation, evaluation of each individual of population, selection, reproduction (crossover), and mutation.



Figure 7. Mind mapping of a genetic algorithm.

4.1. Parameters Retrieving under Genetic Algorithm Framework

In the genetic algorithm of parameter retrieving, each individual of a chromosome represents a set of signed fuzzy measures. Due to binary coding, each chromosome is composed of 30 genes (10 genes corresponding to a parameter to be optimized). The value of each gene is a binary number. Each chromosome is decoded as three real values between 0 and 1, corresponding to the normalized value of $\mu(\{x_1\}), \mu(\{x_2\} \text{ and } \mu(\{x_1, x_2\}))$. The fitness value of each chromosome is evaluated by the AUC (area under curve) of ROC (receiver operating characteristic) curve. Since the probability that a chromosome in a population can be selected to generate offspring depends on its fitness value, the pedestrian detection parameter optimization algorithm based on genetic algorithm takes the maximum AUC as the criterion to optimize.

Figure 8 shows the process diagram of parameter retrieving process based on a GA structure under the application of pedestrian detection. The algorithm starts with a randomly generated initialization population. Each individual of chromosome in the population is decoded into a set of values, which is actually a representation of a specific signed fuzzy measure. Typical HOG feature extraction and HOLBP feature construction presented in Section 2 are performed. The two sets of features are fused by the Choquet integral with respect to the specific signed fuzzy measure represented by the corresponding individual chromosome in the current population. The fusion results are a new set of features, called parallel-HOG-HOLBP, which is consequently transmitted to SVM for classification. The same process is done for each sliding window of the images in INRIA data set [36]. AUC is calculated from the ROC curve which is constructed for each individual of the population.



Figure 8. Algorithm frame parameter retrieval based on GA.

The value of AUC is used to evaluate the fitness value of the chromosome being considered. Then, a tournament selection is conducted. Better individuals (with higher AUC values) have more opportunities to perform several randomly chosen genetic operators to produce offspring. The population is updated by the newly created offspring. This process is repeated until the number of individuals generated exceeds the preset maximum size of population. In the process of program iteration, in order to keep the global search space, some special operations are used when the best fitness value remains unchanged for successive generations (the default value is 20). The individuals in the original population are divided into three parts according to the ascending order of fitness value. The excellent individuals in the first part are all retained, the individuals in the second part produce new offspring by random mutation, and the individuals in the third part are randomly replaced by the new individuals produced by previous genetic operations. As a result, the population is updated and the program continues to iterate.

4.2. Classifier Training

Each chromosome corresponds to a signed fuzzy measure. Based on each signed fuzzy measure, a Choquet integral fuses HOG and LBP features in a candidate image, and the generated parallel-HOG-HOLBP features are sent to an SVM classifier to evaluate the performance of the chromosome according to the classification results on a set of testing images.

Based on the principle of structural risk minimization, support vector machine has a very powerful ability in dealing with nonlinear problems. The algorithm uses learning samples to find an optimal hyperplane in high-dimensional space, so as to separate different samples from two groups. First, the parallel-HOG-HOLBP features of positive and negative samples are calculated as input of the SVM classifier. Then the final decision function is calculated as

$$f(x) = \sum_{i=1}^{n} \omega_i \,\phi(x_i) + b \tag{12}$$

where $\phi : X \to \mathcal{F}$ is a nonlinear mapping from the input space to a high-dimensional feature space. f(x) is optimal in the sense of maximizing the distance between the hyper-plane and the nearest point $\phi(x_i)$. The following equation is usually used to solve optimization problem mentioned above.

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i$$
(13)

where ξ_i is a slack variable, which corresponds to the vertical distance from each wrongly classified sample point to the corresponding boundary hyperplane. Parameter *C* is the penalty coefficient. The larger this parameter is, the more severe the penalty is.

In this paper, INRIA data set [36] is selected as the training set to train SVM classifier, because INRIA data set is a benchmark data set which is widely used in pedestrian detection. We extract positive samples from INRIA training set according to the pedestrian coordinates marked in the dataset, and construct negative samples from the training set by randomly cropping. After extraction, the training set being used consists of 2416 positive samples and 12,180 negative samples from the INRIA dataset. The parameters of SVM classifier are shown in Table 1.

Parameter	Value	Mark	
С	0.5	Penalty parameters for wrongly classified samples	
Tol	1e-4s	Criteria for stopping iteration	
Multi_class	ovr	Multiclass classification strategy parameters	
Class_weight	balanced	Adjust weights based on frequency of each class	
Max_iter	1000	The maximum iteration number	
Loss	squared_hinge	Loss function type	

Table 1. The running parameters of SVM classifier algorithm.

4.3. Classifier Training and Evaluation Criterion

To evaluate the classifier constructed by each chromosome in the current iteration, 1126 positive samples and 453 negative samples are extracted from the INRIA testing set. For each chromosome, a confusion matrix is summarized by four indicators, as shown in Figure 9. The indicators represent four situations:



Figure 9. Interpretation chart of indicators in confusion matrix.

- 1. The actual value is true, and the classifier assigns it to be positive (True Positive = TP);
- 2. The actual value is true, and the classifier assigns it to be negative (False Negative = FN);
- 3. The actual value is false, and the classifier assigns it to be positive (False Positive = FP);
- 4. The actual value is false, and the classifier assigns it to be negative (False Negative = TN).

Three new indicators are sequentially calculated. They are

$$precision = \frac{TP}{TP + FP} \tag{14}$$

$$recall = \frac{TP}{TP + FN}$$
(15)

and

$$F1 \ Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{16}$$

Here, indicator *precision* and indicator *recall* describe the classifier's correct predictions as a percentage of all results, where indicator *F1 Score* takes into account that the destination of optimization is to find the best combination of precision and recall. In our algorithm, indicator *F1 Score* is utilized as the fitness value of each chromosome in iterations.

5. Experimental Results and Analysis

5.1. Data Construction

This paper uses INRIA pedestrian data sets to construct a training set and testing set. We extract positive samples from INRIA training set according to the pedestrian coordinates marked in the dataset, and construct negative samples from the training set by randomly cropping. After construction, the training set consists of 2416 positive samples and 12,180 negative samples, where the testing set consists of 1126 positive samples and 453 negative samples.

5.2. Experimental Results and Analysis

In order to validate the performance of parallel-HOG-HOLBP features and relevant GA-based pedestrian detection algorithm proposed in this paper, four classifiers with different combinations of features are selected to be tested on the same set of testing set. They are:

- 1. SVM classifier with HOG features, denoted as HOG-SVM;
- 2. SVM classifier with serial fusion of HOG and LBP features, denoted as HOG-LBP-SVM;
- 3. SVM classifier with parallel-HOG-HOLBP features whose fusion parameters are set by experience, denoted as HOG-HOLBP-SVM;

4. SVM classifier with parallel-HOG-HOLBP features whose fusion parameters are optimized by GA process, denoted as HOG-HOLBP-GA-SVM.

The experimental results of HOG-SVM and HOG-LBP-SVM are expressed as confusion matrices shown in Tables 2 and 3.

Using Choquet integral to fuse HOG and HOLBP in parallel, suitable values of signed fuzzy measure are extracted because they are the essential parameters in Choquet integral. Their values directly affect the effectiveness of the subsequent pedestrian detection. In experiments of HOG-HOLBP-SVM, 10 groups of signed fuzzy measure are chosen by experience. Their performances are validated by F1 scores and shown in Figure 10. The F1 score reaches the best result with $\mu(\{x_1\}) = 0.45$ and $\mu(\{x_2\}) = 0.10$. HOG-HOLBP-SVM experiment based on this best combination is conducted and the detection results are expressed as the confusion matrix shown in Table 4.



Figure 10. The results of 10 trials for different combinations of signed fuzzy measure in HOG-HOLBP-SVM.

Keeping the dataset unchanged, the GA-based feature fusion and pedestrian detection algorithm (HOG-HOLBP-GA-SVM) is used as a classifier for training and testing. Parameters of the signed fuzzy measure of Choquet integral to accomplish the feature fusion in parallel are optimized by the genetic algorithm during the iteration process, as explained in Section 4.1.

Under the premise of the same running parameters of genetic algorithm, HOG-HOLBP-GA-SVM was run for 10 trials. The results of these 10 trials are recorded in Table 5, in which the minimum fitness value, the maximum fitness value and the average value at the end of each trial are recorded in the corresponding rows of each run. As shown in Table 5, among the 10 randomly generated trials, Trial 3 gives the best optimization result, that is, at the end of running, the maximum fitness value reaches 0.9758. In Trial 3, an optimization set of parameters ($\mu(\{x_1\}) = 0.382, \mu(\{x_2\}) = 0.174$) is obtained at the end of iteration. The standard deviations of the three measurements for the 10 trials are also shown in the bottom row of Table 5. The optimization process of trial 3 is shown in Table 6. The confusion matrix of HOG-HOLBP-GA-SVM experiment on this trial is shown in Table 7. For the remaining trials, HOG-HOLBP-GA-SVM can also reach into the nearby space of the optimized point. This shows that HOG-HOLBP-GA-SVM has a satisfactory performance on the efficiency and effectiveness.

A strict Value	Prediction Value		
Actual value	Positive	Negative	
True	980	146	
False	50	403	

Table 2. Confusion matrix of HOG-SVM.

Table 3. Confusion matrix of HOG-LBP-SVM.

A sheel Value	Prediction Value		
Actual value –	Positive	Negative	
True	1056	70	
False	44	409	

Table 4. Confusion matrix of HOG-HOLBP-SVM.

A . (Predicti	on Value
Actual value	Positive	Negative
True	1090	36
False	28	425

Table 5. Results of 10 trials in the experiments of HOG-HOLBP-GA-SVM.

Trials	Minimum Fitness Value	Maximum Fitness Value	Mean Fitness Value
Trial 1	0.4012	0.9440	0.8063
Trial 2	0.5616	0.9249	0.7553
Trial 3	0.4928	0.9758	0.9019
Trial 4	0.6420	0.9535	0.8611
Trial 5	0.5170	0.9746	0.8740
Trial 6	0.6203	0.9515	0.8047
Trial 7	0.5417	0.9560	0.8518
Trial 8	0.7454	0.8752	0.7704
Trial 9	0.4404	0.9628	0.8958
Trial 10	0.5333	0.9747	0.7840
SD	0.1003	0.0501	0.0530

To compare the performance of four classifiers, the results in the confusion matrix (Tables 2–4 and 7) are converted to three indicators, i.e., precision, recall and F1 score. The experimental results of four methods are shown in Table 8, where precision, recall, F1 score, and feature extraction time are reported. As shown in Table 8, three indicators of detection rate of the two parallel feature fusion methods (HOG-HOLBP-SVM and HOG-HOLBP-GA-SVM) are superior to those of HOG-SVM and serial feature fusion method (HOG-LBP-SVM). In addition, there is an obvious slowdown in the feature extraction time, from 88.285 and 131.854 ms per frame to 10.075 and 10.126 ms per frame, respectively. The reduction of execution time shows that the parallel feature fusion algorithm has better real-time performance than the serial feature fusion algorithm. Figure 11 shows the comparison from the precision, recall and F1 score of the four algorithms on testing set and a summary of performance comparison among four algorithms is depicted in Figure 12.



Figure 11. Performance comparison among four algorithms on the testing set: (**a**) comparison from precision; (**b**) comparison from recall; (**c**) comparison from F1 score.



Figure 12. Summary of performance comparison among four algorithms.

Iterations	$\mu(\{x_1\})$	$\mu(\{x_2\})$	F1 Score
1	0.568	0.781	0.9034
2	0.529	0.743	0.9265
3	0.526	0.623	0.8947
4	0.498	0.592	0.9321
5	0.493	0.588	0.9325
6	0.474	0.434	0.9411
7	0.429	0.367	0.9419
8	0.447	0.533	0.9416
9	0.415	0.219	0.9530
10	0.427	0.348	0.9522
11	0.421	0.299	0.9535
12	0.412	0.315	0.9568
13	0.410	0.310	0.9566
14	0.344	0.221	0.9572
15	0.374	0.203	0.9570
÷	÷	÷	÷
4000	0.382	0.174	0.9758

Table 6. Optimization process of μ values of trial 3 in the series experiments on HOG-HOLBP-GA-SVM.

A strict Value	Predicti	on Value	
Actual value —	Positive	Negative	
True	1087	39	
False	15	438	

Table 7. Confusion matrix of HOG-HOLBP-GA-SVM.

Table 8. Performance evaluation of four classifiers.

Classifier	Precision	Recall	F1 Score	Feature Extraction Time (ms/frame)
HOG-SVM	0.9515	0.8703	0.9091	88.285
HOG-LBP-SVM	0.9600	0.9378	0.9488	131.854
HOG-HOLBP-SVM	0.9712	0.9591	0.9651	10.075
HOG-HOLBP-GA-SVM	0.9864	0.9655	0.9758	10.126

A ROC (receiver operating characteristic) curve was drawn with false positive rate as horizontal coordinate and true positive rate as vertical coordinate to estimate the influence of sample distribution on the performance of the algorithm.

The larger the area surrounded by the ROC curve, the better the performance of the classifier. Figure 13 shows the corresponding ROC curves of HOG feature, HOG-LBP feature, HOG-HOLBP feature, and HOG-HOLBP-GA feature with SVM being used as classifiers, respectively. It can be seen that HOG-HOLBP-GA-SVM classifier has better performance than other classifiers apparently.

In addition, according to the detection results of each algorithm, FPPI (false positive per image) curves are drawn, respectively. FPPI curve represents the average number of correct retrievals in each image, and its value is closer to the practical application of the classifier. The lower the curve in the graph, the stronger the performance of the corresponding model. The seven FPPI curves shown in Figure 14 compare the proposed pedestrian detection algorithms with other popular pedestrian detection algorithms. It can be seen from the figure that HOG-HOLBP-GA-SVM classifier has achieved good performance.



Figure 13. The ROC curves of classifiers.



Figure 14. The FPPI curves of pedestrian detection algorithms.

6. Conclusions

The key issues of pedestrian detection are to extract efficient features so as to accomplish detection correctly and promptly. This paper attempted to present a novel parallel framework and solve these problems with Choquet integral being involved. The intervention of Choquet integral makes the two features (HOG and LBP) merge in parallel. The resulting feature, parallel-HOG-HOLBP, not only contains the original advantages of HOG and LBP, but also avoids the unavoidable dimension disaster in traditional serial fusion. Genetic algorithm is used to retrieve the interval parameters of fuzzy measure in Choquet integral. It is a more rational way to retrieve these parameters through a global optimization algorithm compared to through trial and error method. We conducted extensive experiments to demonstrate that the proposed method has more effective characteristics compared with the original methods. Our research reveals that feature fusion in parallel is an effective and promising way to improve pedestrian detection performance.

Author Contributions: Conceptualization, R.Y. and Y.W.; methodology, R.Y.; software, Y.X.; validation, L.Q. and Q.L.; formal analysis, R.Y.; investigation, Y.X.; resources, L.Q.; data curation, Q.L.; writing—original draft preparation, R.Y.; writing—review and editing, Y.W. and Y.X.; visualization, L.Q.; supervision, R.Y.; project administration, R.Y.; funding acquisition, R.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant numbers 61773266, 61873170 and U1813225, and by Science and Technology Research and Development Foundation of Shenzhen, grant numbers JCYJ20170818144254033, JCYJ20190808112605503 and JCYJ20190808144607400. The APC was funded by National Natural Science Foundation of China, grant number 61773266.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Zhao, Z.; Zheng, P.; Xu, S. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* 2019, 30, 3212–3232. [CrossRef] [PubMed]
- 2. Dollar, P.; Wojek, C.; Schiele, B.; Perona, P. Pedestrian detection: An evaluation of the state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 743–761. [CrossRef] [PubMed]

- Dollar, P.; Appel, R.; Belongie, S.; Perona, P. Fast feature pyramids for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 2014, 36, 1532–1545. [CrossRef] [PubMed]
- 4. Liu, S.Y.; Guo, H.Y.; Hu, J.G.; Zhao, X.; Tang, M. A novel data augmentation scheme for pedestrian detection with attribute preserving GAN. *Neurocomputing* **2020**, *401*, 123–132. [CrossRef]
- 5. Severino, J.V.B.; Zimmer, A.; Brandmeier, T.; Freire, R.Z. Pedestrian recognition using micro Doppler effects of radar signals based on machine learning and multi-objective optimization. *Expert Syst. Appl.* **2019**, *136*, 304–315. [CrossRef]
- 6. Liu, W.; Liao, S.; Ren, W.; Hu, W.; Yu, Y. High-level Semantic Feature Detection: A New Perspective for Pedestrian Detection. *CVPR* **2019**, 5187–5196.
- 7. Cao, Z.; Simon, T.; Wei, S.-E. Realtime multi-person 2D pose estimation using part affinity fields. CVPR 2017, 2017, 1302–1310.
- Yang, Z.; Nevatia, R. A multi-scale cascade fully convolutional network face detector. In Proceedings of the 23rd International Conference on Pattern Recognition, Cancun, Mexico, 4–8 December 2016; IEEE Computer Society Press: New York, NY, USA, 2016; pp. 633–638.
- 9. Sumi, A.; Santha, T. Frame level difference (FLD) features to detect partially occluded pedestrian for ADAS. J. Sci. Ind. Res. 2019, 78, 831–836.
- 10. Yan, Y.C.; Ni, B.B.; Liu, J.X.; Yang, X.K. Multi-level attention model for person re-identification. *Pattern Recognit. Lett.* **2019**, 127, 156–164. [CrossRef]
- 11. Li, G.; Yang, Y.; Qu, X. Deep Learning Approaches on Pedestrian Detection in Hazy Weather. *IEEE Trans. Ind. Electron.* **2020**, *10*, 8889–8899. [CrossRef]
- 12. Dalal, N.; Triggs, W. Histograms of Oriented Gradients for Human Detection. CVPR 2005, 1, 886–893.
- Ojala, T.; Pietikäinen, M.; Harwood, D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In Proceedings of the 12th IAPR International Conference on Pattern Recognition, Jerusalem, Israel, 9–13 October 1994; IEEE Computer Society Press: New York, NY, USA, 1994; Volume 1, pp. 582–585.
- 14. Zhang, S.; Bauckhage, C.; Cremers, A.B. Informed Haar-like features improve pedestrian detection. CVPR 2014, 947–954.
- 15. Wang, X.; Han, T.X.; Yan, S. An HOG-LBP human detector with partial occlusion handling. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Lisbon, Portugal, 5–8 February 2009; IEEE: New York, NY, USA, 2009; pp. 32–39.
- Wei, Y.; Tian, Q.; Guo, J. Multi-vehicle detection algorithm through combining Harr and HOG features. *Math. Comput. Simul.* 2019, 155, 130–145. [CrossRef]
- 17. Yang, Z.; Kurita, T. Improvements of local descriptors in HOG/SIFT by BOF approach. *IEICE Trans. Inf. Syst.* 2014, E96-D, 1293–1303. [CrossRef]
- 18. Wang, L.; Guo, C.; Liu, J.; Meng, D. A novel learning-balsed frame pooling method for event detection. *Signal Process.* **2017**, 140, 45–52. [CrossRef]
- 19. Alonso, I.P.; Fernández-Llorca, D.; Sotelo, M.A.; Bergasa, L.M.; Revenga, P.A.; Nuevo, J.; Ocaña, M.; Garrido, M.Á.G. Combination of feature extraction methods for SVM pedestrian detection. *IEEE Trans. Intell. Transp. Syst.* 2007, *8*, 292–307. [CrossRef]
- 20. Bilal, M.; Hanif, M.S. High performance real-time pedestrian detection using light weight features and fast cascaded kernel SVM classification. *J. Signal Process. Syst.* **2019**, *91*, 117–129. [CrossRef]
- 21. Cao, J.; Pang, Y.; Li, X. Leaning multilayer channel features for pedestrian detection. *IEEE Trans. Image Process.* 2017, 26, 3210–3220. [CrossRef]
- 22. Li, G.; Zong, C.; Liu, G.; Zhu, T. Application of convolutional neural network (CNN)-Adaboost algorithm in pedestrian detection. *Sens. Mater.* **2020**, *32*, 1997–2006. [CrossRef]
- 23. He, Y.-Q.; Qin, Q.; Josef, V. A pedestrian detection method using SVM and CNN multistage classification. J. Inf. Hiding Multimed. Signal Process. **2018**, *9*, 51–60.
- 24. Kim, J.H.; Batchuluun, G.; Park, K.R. Pedestrian detection based on faster R-CNN in nighttime by fusing deep convolutional features of successive images. *Expert Syst. Appl.* **2018**, *114*, 15–33. [CrossRef]
- 25. Albiol, A.; Monzo, D.; Martin, A.; Sastre, J.; Albiol, A. Face recognition using HOG-EBGM. Pattern Recognit. 2008, 29, 1537–1543. [CrossRef]
- 26. Tudor, B. Pedestrian detection and tracking using temporal differencing and HOG features. Comput. Electr. Eng. 2014, 40, 1072–1079.
- 27. Hoang, V.-D.; Le, M.-H.; Jo, K.-H. Hybrid cascade boosting machine using variant scale blocks HOG features for pedestrian detection. *Neurocomputing* **2014**, *135*, 357–366. [CrossRef]
- 28. Zhu, Q.; Yeh, M.C.; Cheng, K.T. Fast human detection using a cascade of histograms of oriented gradients. *CVPR* **2006**, 2, 1491–1498.
- 29. Maggiani, L.; Bourrsset, C.; Quinton, J.C.; Berry, F.; Sérot, J. Bio-inspired heterogeneous architecture for real-time pedestrian detection applications. *J. Real Time Image Process.* **2018**, *14*, 535–548. [CrossRef]
- Hong, G.S.; Kim, B.G.; Hwang, Y.S.; Kwon, K.K. Fast multi-featured pedestrian detection algorithm based on histogram of oriented gradient using discrete wavelet transform. *Multimed. Tools Appl.* 2015, 75, 15229–15245. [CrossRef]
- 31. Wu, S.; Payeur, R.P. Improving pedestrian detection with selective gradient self-similarity feature. *Pattern Recognit.* **2015**, *48*, 2364–2376. [CrossRef]
- 32. Ojala, T.; Pietikäinen, M.; Harwood, D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [CrossRef]
- 33. Jiang, Y.; Tong, G.; Yin, H.; Xiong, N. A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters. *IEEE Access* 2019, *7*, 118310–118321. [CrossRef]

- 34. Wang, Z.; Klir, G.K. Fuzzy Measure Theory; Plenum Press: New York, NY, USA, 1992.
- 35. Wang, Z.; Yang, R.; Leung, K.-S. Nonlinear Integrals and Their Applications in Data Mining; World Scientific: Singapore, 2010.
- 36. INRIA Data Set. Available online: http://pascal.inrialpes.fr/data/human/ (accessed on 20 November 2020).