Pedestrian Detection Using Self-Organizing Map Neutral Network

Shi Yong-biao, Guo Feng, Long Xiang

College of Information Science & Technology

Hainan University

Haikou, China

hainu2008@163.com

Abstract-**This paper presents a method of detecting pedestrian's side in video frames of cluttered scenes. This detection technique is based on the novel idea of wavelet template and SOM neutral network. In order to make detection results more accurate and reduce computation cost, we process video frames for interpolation before wavelet transformation and we use background subtraction and frames difference to decide where pedestrians stand in a frame.**

Keywords-object detection ; image interpolation ; Haar wavelet；SOM；side views detection

I. INTRODUCTION

The problem of object detection has seen a high degree of interest over years. The basic problem is how to characterize an object class or how to distinguish objects from background. People detection is a hot point that attracts many researchers.

Currently, main detection algorithms include frame difference, background subtraction, and optical flow method.

Optical flow method results in large calculation, and it has a bad real-time effect. James H. Duncan and Tsai-Chia Chou have proposed such a detection method. In the method, intensity history at each pixel is convolved with the second derivative in time of a temporal Gaussian smoothing function. The zero crossings in a single frame of the resulting function indicate the positions of moving edges [1].

Commonly used temporal difference method is frame difference. Davide A. Migliore, etc, has proposed a method of frame difference in fast and robust motion detection [2].

There have been a lot of works about background subtraction method. Ahmed Elgammal, etc, has proposed a non-parametric model for background subtraction [3]. Anurag Mittal and Nikos Paragios have proposed a motion-based background subtraction method [4].

Besides, Michael Oren, etc, has proposed pedestrian detection method using wavelet template [5]. Anuj Mohan, etc, has proposed a method of examples-based object detection [6].

Optical flow, background subtraction and frames difference method are heavily based on motion and handcrafted models. Unlike these methods, an important aspect of our technique is that the model is automatically learned from examples and avoids the use of motion and explicit segmentation in training period. This method uses wavelet template to select pedestrians' feature and we train our system with self-organizing map. It proves more effective and accurate.

In order to improve the tempo of our algorithm, background subtraction and frames difference methods are used to decide where pedestrians stand in a video frame. Image interpolation makes wavelet transformation more accurate. That is to say, video frames are processed for interpolation before wavelet transformation.

This paper is arranged as followings:

In the first part, we simply present how to detect the location of pedestrians in an image. Background subtraction and frames difference method are simply explained. In the second part, image interpolation is introduced. In the third part, Haar wavelet transformation is demonstrated. In the fourth part, SOM is used to detect object. The last two parts show the experiment results and conclusion.

II. LOCATION DETECTION

Human body is highly non-rigid object, and different bodies are distinct in size, shape, clothes, texture and other aspects. Figure 1 demonstrates some representative images in the samples.

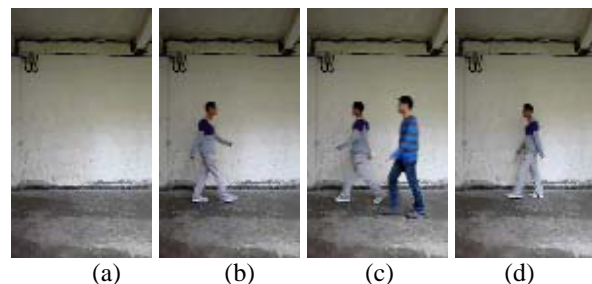|			
(a)	(b)	(c)	(d)

Figure 1: (a) background frame (b)-(d) different postures of pedestrians

Motion object detection means extracting moving humans' region from complex scenes. In our detection system, object detection influences feature selection, so object detection is very important.

In the system, we use three continuous frames to decide object of the middle frame. Background in three continuous frames is highly similar to each other. We use the middle frame minus the first one to get Result 1, and the third frame minus the middle frame to get Result 2. Result 1 conducts AND operation with Result 2, and then we get Result 3.

Result 3 conducts threshold operation, and then we get the object region of the middle frame.

Due to the similarity between frames, frames difference method often can't detect the whole region of motion objects. Moreover, there exists motion tail in the result.

Background subtraction is used when camera is fixed and background changes little. Consequently, we need to update the background frame when it is necessary. We use motion frames minus background frame to get Result 4, and conduct Result 4 for threshold operation, and then we get the object region.

Since background subtraction asks for an invariable background, many factors can influence detecting results, such as illumination intensity, wind blowing leaves, and shadows.

In order to make detection results more accurate, the result gotten from background subtraction conducts AND operation with that gotten from frames differences [7].

Figure 2 shows detecting results which are gotten via combining background subtraction with frames difference:
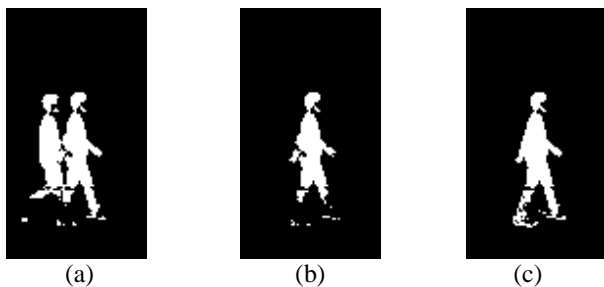

(a) (b) (c)

Figure 2: (a) gotten via continuous frames subtraction; (b) gotten via background subtraction; (c) result (b) logic AND with result gotten via background subtraction.

III. IMAGE INTERPOLATION

After finding the region of pedestrians, we set a rectangular window to shear the region out of a whole frame. Since the region is not too large, in order to make the results of wavelet transformation more accurate, we process the pedestrian region for interpolation before wavelet transformation.

Interpolation is used to enlarger an image and improves an image's resolution ratio.

Mainly used interpolation algorithms include Nearest Neighbor, Bilinear and Cubic Convolution. Nearest Neighbor is the fastest algorithm, Bilinear retrains more spectral information than the other two and Cubic Convolution is much better than the other two methods in retaining spatial information of image. Therefore, we use Cubic Convolution.

Cubic Convolution uses the weighted average of 16 pixels (including sampling pixel and its neighboring 15 pixels in the original image) to get interpolated pixel $f(i+u, j+v)$. The formula is as following [8]:

$$f(i+u, j+v) = A \times B \times C \qquad (1)$$

Array A, B, C respectively has the following format:

$$B = \begin{bmatrix} f(i-1,j-1) & f(i-1,j) & f(i-1,j+1) & f(i-1,j+2) \\ f(i,j-1) & f(i,j) & f(i,j+1) & f(i,j+2) \\ f(i+1,j-1) & f(i+1,j) & f(i+1,j+1) & f(i+1,j+2) \\ f(i+2,j-1) & f(i+2,j) & f(i+2,j+1) & f(i+2,j+2) \end{bmatrix} \qquad (2)$$

$$A = \begin{bmatrix} S(u+1) & S(u) & S(u-1) & S(u-2) \end{bmatrix} \qquad (3)$$

$$C = \begin{bmatrix} S(v+1) & S(v) & S(v-1) & S(v-2) \end{bmatrix} \qquad (4)$$

Function $S(x)$ is interpolation function basis. Figure 3 shows interpolation result.
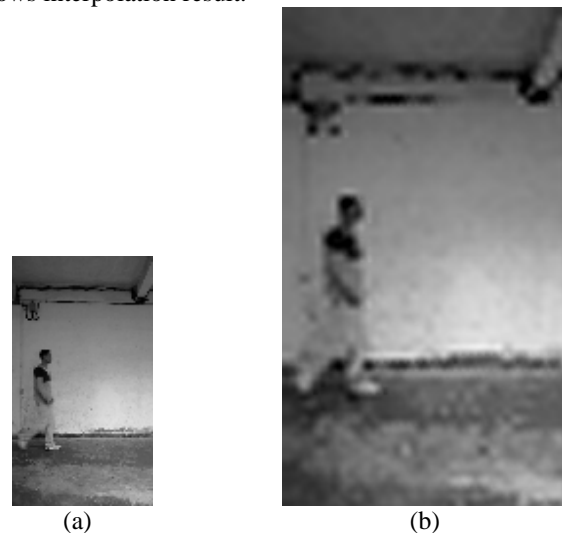

(a) (b)

Figure 3: (a) original image; (b) image after interpolation.

IV. HAAR WAVELET TRANSFORMATION

Haar wavelet transformation is a commonly used method for feature selection in object detection. In this paper, we use Haar wavelet transformation to obtain the horizontal, vertical and diagonal details of an image. We select some wavelet coefficients from these details to serve as feature vectors.

After interpolation, the rectangular window is double its original size in both rows and columns. Therefore, we conduct wavelet transformation for twice. Then we divide the result into some blocks and select wavelet coefficients for horizontal, vertical and diagonal details. We select the largest number of each block to make up feature vectors, which are the input data of SOM. Table (a) and Table (b) respectively show the average coefficient values for the set of horizontal and vertical coefficients of pedestrian class.

The feature vector uses a set of 20 coefficients that are consistent along the ensemble either as indicators of

"change" or "no-change". There are 48 vertical and 48 horizontal coefficients at scales of 12×16 pixels. We select 12 values from the horizontal and 8 from the vertical coefficients.

15	44	50	41	18	30
47	8	8	26	33	72
37	64	19	12	7	20
7	11	10	21	38	26
5	30	1	3	6	35
20	7	5	17	4	5
6	22	20	22	10	52
2	3	5	23	15	20

Table (a): horizontal coefficients

177	38	26	22	22	39
44	15	222	38	45	83
28	84	65	17	248	69
5	4	5	97	57	15
252	40	2	2	4	75
29	37	250	32	2	2
4	64	31	37	246	50
2	2	3	58	27	25

Table (b): vertical coefficients

V. RECOGNITION STEPS

A. Introduction of SOM

Self-Organizing Maps (SOM) is proposed by Teuvo Kohonen, a Finnish scientist, in the year of 1981. SOM, which imitates how human brain deals with signals, is a kind of unsupervised learning algorithm.SOM uses WTA (winner takes all) competing mechanism. And it has been used in many fields, such as image segmentation, pattern recognition, etc.

SOM has two layers: one is input layer, and the other is output layer or competing layer. Unlike other neutral networks, SOM has no hidden layers or middle layers. There are n neutral nodes in input layer, and there are $m \times m$ nodes in output layer, and all of the output nodes make up a one-dimension or two-dimension array. Each of the nodes in input layer has a connection with every output node, just like Figure 4 shows.
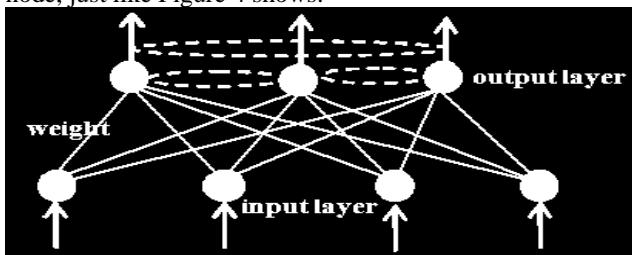

Figure 4: structure of SOM.

In SOM network, input data are mapped to the output layer, and the topology structure of output layer doesn't change. With the network being trained via continuously learning from the input data, the weight vector space gradually reach accordance with the input patterns in probability distribution. In the mean while, output nodes compete for the response to input patterns, and the winner node adapt its weights, as a consequence, it encourages its neighboring nodes and restrains the nodes which are far from it.

The input layer receives input vectors. The number of input nodes is decided by the dimension of the input vector, Output layer is used to receive the results and the number of output nodes is decided by the number of patterns of what we are studying at. To some degree, it can be a little more than actual account. In this paper, we receive two patterns: pedestrian or not.

B. Taining of SOM

SOM neutral network is trained as below [9]:

1) Initialize the network. We use random numbers to initialize the weights of SOM network; set a pattern rate neighbors region; and set the training times.

2) Input a new pattern X to the network. Array X has a format like this:

$$X = (x_1, x_2, \cdots, x_n)^T \qquad (5)$$

3) Figure up each Euclidean distance D between input data and weight vector. D is calculated according to the following formula:

$$D = \sqrt{\sum_{i=1}^{n} \left(x_i - w_{ij} \right)^2} \qquad (6)$$

w_{ij} is the weight between node i of the input layer and node j of the output layer.

4) Select the output node (winner) which has the smallest distance in Step 3). We make a sign of the node as j^*, and find out its neighboring nodes.

5) Train the weights. The winner node and its neighboring nodes update their weights according to the following formula:

$$\Delta w_{ij} = \eta \exp\left(-\frac{|j - j^*|^2}{\sigma^2}\right) \left(x_i - w_{ij}\right) \qquad (7)$$

η is a constant between 0 and 1.

6) Judge whether it reaches the training step. If "yes", we start a new round of learning; or we go back to Step 2).

VI. THE EXPERIMENT RESULTS

In our experiment, we use 318 positive examples and 327 negative examples to evaluate the system.

We use ROC curve to evaluate the relation between pedestrian detection rate and false positives rate. From the ROC curve given in Figure 7, we can see that the SOM system has a pedestrian detection rate of 88.7%, with a false positive rate of 1 for every 15000 windows examined.

Though the system results in a good performance, there are still some disadvantages. In Figure 6, for example, some background's feature is similar to pedestrians', so it results in mistaken detection (the door in the background is detected as pedestrian). Besides, we train SOM with side-view images, some frontal and rear pedestrians are not wholly detected.
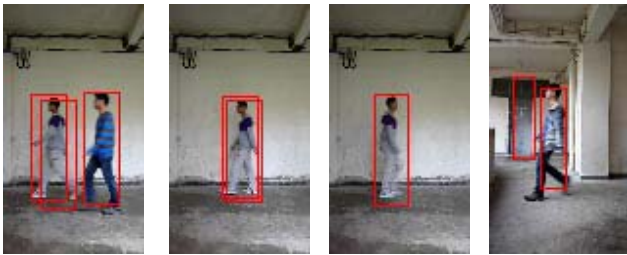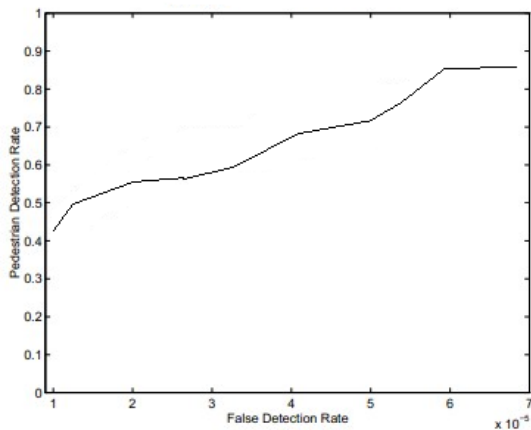

Figure 6: results of identification

Figure 7: ROC curve for SOM network detection system

VII. CONCLUSION

In this paper, we introduce a new idea of combining wavelet transformation with SOM to identify pedestrians. This method reduces calculation and improves detection accuracy. It focuses on the two important steps-- feature selection and training of classifier: 1) Compute Haar wavelet transformation for the image. Select some coefficients from vertical, horizontal and diagonal details, which serve as the feature vector for ensuing training of SOM and 2) Set up an SOM, and use several groups of pedestrian and background feature vectors which are obtained from Step 1) to train SOM. Since SOM is a kind of unsupervised network, it adjusts itself to suit different input patterns as an effective tool for detecting pedestrians.

The approach described in our paper represents a new framework that may be used to detect complex objects besides pedestrians.

REFERENCES

[1] James H. Duncan and Tsai-Chia Chou. "On the Detection and the Computation of Optical Flow ". IEEE transactions on pattern analysis and machine intelligence, 1992,pp.346-352.

[2] Davide A. Migliore, Matteo Matteucci and Matto Naccari. "A Revaluation of Frame Difference in Fast and Robust Motion Detection ". VSSN 2006. Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, 2006, pp. 215–218 .

[3] Ahmed Elgammal, David Harwood and Larry Davis. "Non-parametric Model for Background Subtraction". Computer Vision, ECCV 2000.

[4] Anurag Mittal and Nikos Paragios. "Motion-based background subtraction using adaptive kernel density estimation". Computer Vision and Pattern Recognition, 2004,pp. 302-309.

[5] Michael Oren, Constantine Papageorgiou, Pawan Sinba, Edgar Osuna, and Tomaso Paggio."Pedestrian Detection Using Wavelet Templates ". Computer Vision and Pattern Recognition, 1997, pp. 193-199.

[6] Anuj Mohan, Constantine Papageorgiou, and Tomaso Paggio. "Example-Based Object Detection in Images by Components ". IN: IEEE transactions on pattern analysis and machine intelligence, 2001.

[7] Xia Yongquan, Ning Shaohui, and Li Weili. "A Simple and Effective Moving Objects Detection Algorithm" . Computer Measurement & Control. 2011,pp. 356-358,.

[8] Shen Jia."Simulation of Classical Interpolation Algorithm"Science & Tecnology Information.2009,219-221

[9] Shi Xiaoyun. "Stationary satelnte image classification based on SOM neural network".Electrnic Design Engineering.2011,55-60