

## Moving Object Detection and Identification Method Based on Vision

TAO Jian-Ping<sup>1</sup>, LV Xiao-lan<sup>1</sup> and LIU Jun<sup>1</sup>

*1. Institute of Agricultural Facilities and Equipment, Jiangsu Academy of  
Agricultural Sciences, Nanjing 210014, China*

### **Abstract**

*In traffic monitoring system, moving object monitor is the key part of monitoring system. This paper proposed a non-reference background detection method based on the reference background model for the poor detection effect of Gaussian background modeling. The model utilizes a series of sampling values to estimate the probability model of observed pixels before current pixels; Then binaryzation moving object detection based on the probability model. In terms of moving object identification, this paper proposed several features, and being trained and identified through BP neural network. The experimental result indicated that this background model can detect the foreground moving object effectively, and achieved satisfied effect on pedestrians and vehicles target recognition rate.*

**Keywords:** *Traffic monitoring; Background model; Vehicle recognition; Machine vision; Probability model*

### **1. Introduction**

With the development of economy and the progress of the society, the traffic scene monitoring has become a very important topic. Through the camera to monitor the traffic scene has been widely used in social life, but in many cases, to analyze the video sequences by the people is not an appropriate way. Visual monitoring technology based on computer vision and image processing provides a more advanced and feasible method, its wide application prospect and potential commercial value has attracted the attention of the academic and business at home and abroad. This paper studied the moving object detection and recognition in traffic monitor [1~5], and proposed some improved method. For the defect of poor anti-interference performance of traditional detection to background illumination, [6~8]etc., this paper proposed a kind of non-reference background modeling to detect the foreground object effectively. In the pedestrians and vehicles target recognition, it proposed several characteristic of easy to be identified, and achieved highest recognition rate through BP neural network training and testing.

### **2. Moving Object Detection**

In traffic monitoring system, generally it conducts real-time monitor of specific sites of outdoor by static camera. For the camera is static, so usually use a kind of typical moving region detection method, background difference method, namely through comparing the current image and the background model to detect moving area. But in the outdoor environment, due to the effects such as change of leaves, shadows, and illumination, etc., the background area is not completely static. Therefore, how to build the background model becomes the key to background detection method.

It is the first step to extract moving object in video sequences obtained by camera in moving object identification. The most simple background model is a pixel gray value imitated by a Gaussian distribution [9], this model is suitable for the slowly changing

background, and cannot handle multiple background, such as there is shaking tree in the background. For traffic control applications, the pavement, shadows, and the vehicle can be imitated respectively by a gaussian distribution, background pixel gray value imitated by the mixed weighting of the three Gaussian distribution [10], the parameters of the model can be learnt and updated by increasing Expectation-Maximization Clustering. However, despite the pixel gray value imitated by three Gaussian mixture, but for the pavement, *etc.* it is still a single model assumption. For complex and dynamic background introduced multiple Gaussian mixture model[6,7,11] , but this method for rapidly changing background cannot accurately detect the moving region through a mixture of several Gaussian distributions. In addition, for background learning rate, multiple Gaussian mixture model has to be compromised. When the learning rate is low, this model can not response to a sudden change in the background; On the other hand, the slow foreground object would be mistaken as the background. This is the foreground aperture problem [12]. In order to overcome these problems above, this paper adopted a kind of non-reference background model. The experimental results show that the model can well adapt to changes in the background, and can accurately detect the foreground objects.

### 3. The Establishment of Non-Reference Model

#### 3.1. Density Estimation

Model design goal is to capture the latest information in image sequence, and to continuously update the information to capture the rapid changes of the background. So if you want to get the exact detection, at any time, we must estimate the distribution density function through the recent historical information.

Set  $x_1, x_2, \dots, x_N$  as the previous pixel value of the current pixel,  $x_t$  as the pixel value of this pixel at time t. where, each pixel value is composed of three color channels. Its probability density function can be non-reference estimate according to the following core estimation function K.

$$P_r(x_t) = \frac{1}{N} \sum_{i=1}^N K(x_t - x_i)$$

If the core estimation function K is Gaussian function, its probability density can be estimated by the following equation.

$$P_r(x_t) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - x_i)^T \Sigma^{-1} (x_t - x_i)}$$

Where,  $\Sigma$  is covariance matrix. If  $\sigma_j^2$  represents the variance of jth color channel, and each channel are independent of each other, then

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

The density estimation can be simplified as

$$P_r(x_t) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^d \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2} \frac{(x_{ij} - x_{ij})^2}{\sigma_j^2}}$$

It can determine the pixel belongs to foreground point . If  $P_r(x_t) < th_1$ , then the pixel is reduced to foreground point, otherwise, as background point. Here  $th_1$  is threshold value.

### 3.2. Variance Estimation

The change of the brightness of the pixel values for two reasons. One reason is due to the different goals at different times projected lead to that point on the same pixel brightness values make a great leap. Another reason is that in a short period for projection in the same pixel points on the same target, due to the change of image fuzzy produce local luminance values. Covariance matrix  $\Sigma$  should response due to image fuzzy local luminance values change, rather than reaction jumping change brightness values. For different color channel, due to a minor change is different, so the variance is also different.

To estimate the variance of jth color channel  $\sigma_j^2$ , calculate the difference value of each pair of adjacent sampling value  $|x_{i,j} - x_{i+1,j}|$ , then solve the mid-value  $m_j$  of this series difference values. Due to  $x_{i,j}, x_{i+1,j}$  are two continuous luminance values, so they usually submit to same distribution, only a few submit to cross-distribution. Here, we assumed luminance value submit to normal distribution  $N(\mu, \sigma^2)$ , then  $(x_{i,j} - x_{i+1,j})$  also submit to normal distribution  $N(0, 2\sigma^2)$ . Therefore, the variance of jth color channel  $\sigma_j^2$  can be estimated through the following equation, that is

$$\sigma_j^2 = \frac{m_j}{0.68\sqrt{2}}.$$

Due to the difference value are all integral values, so it can calculated more accurate mid-value by linear fitting.

### 3.3. Restrain Error Detection

For the changing background of the outdoor environment, there are two main sources of error detection. First, uniform random noise in the whole image; second, background movement does not express in background model. In addition, the offset of the camera can also cause many error detections. If some part of the background due to movements occupied new pixel point, and has not been included in the model, so these pixel points are detected as foreground points. But the background moving parts in the original pixel points belong to the background distribution has a high probability. In general, only a few small offset between two adjacent frames, so we used early detect the distribution of pixel neighboring region of background to decide whether to order due to error detection due to changes in the background.

Set  $x$  as the foreground point obtained by early background detection,  $x_{t,j}$  is the jth color channel value at  $t$  time. Then the pixel deflection probability can be defined as follows:

$$P_N(x_t) = \max_{y \in N(x)} \prod_{j=1}^d \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2} \frac{(x_{t,j} - y_{b,j})^2}{\sigma_j^2}}$$

Where,  $N(x)$  is the background point adjacent pixel  $x$ ,  $y_{b,j}$  is the jth color channel value of the background point  $y$ . if  $P_N > th_2$ , this foreground pint is melt into background, thus it is eliminated the error detection due to small background movement.

## 4. BP Neural Network Classifier Design

### 4.1. Feature Extraction

For pedestrians and vehicles classification, the most significant difference between both pedestrians is non-rigid moving object, while the vehicle is rigid moving object. Because the vehicle has more regular appearance, at the same time, and the outline of pedestrian images are usually more complex, so to distinguish pedestrians and vehicles can be expressed by complexity of the image contour features  $X_1$ , the formula as follows:

$$X_1 = \frac{l_f}{Area_f}$$

Where  $l_f$  is the perimeter of binaryzation image,  $Area_f$  is the area of binaryzation image. In general, pedestrians binary image contour complexity is higher than vehicle contour.

Reflect the shape difference between pedestrians and vehicles, another is characterized by pedestrians and vehicles binary image of the ratio of length and width, referred to as the aspect ratio  $X_2$ . The equation description as follows:

$$X_2 = \frac{Width}{Length}$$

In general, the pedestrian has larger length-width ratio than vehicle.

In general pedestrian image shape is complex, and vehicles with relatively regular rectangular shape, so area of pedestrians binarization image accounted for the proportion of the area of the adjacency rectangle relative to the vehicle is smaller, so it can be used as a feature. Its definition as follows:

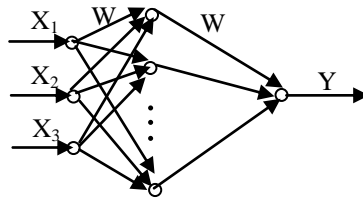
$$X_3 = \frac{Area_f}{Area_m}$$

Where  $Area_f$  is the area of moving object,  $Area_m$  is the area of the adjoin rectangular moving object.

### 4.2. Network Structure

This paper adopted three layers of BP network, the transmission function of input layer to implicit layer, and implicit layer to output layer, both selects non-linear *Sigmoid* function, to complete the nonlinear mapping of input model to output model. Put the extracted feature as the input vector of network  $X = \{X_1, X_2, X_3\}$ , where  $X_1$  is the contour complexity,  $X_2$  is duty ratio,  $X_3$  is the length-width ratio. Network output vector is  $Y$ . The number of hidden nodes selection has great influence on network performance in neural network based on BP algorithm, too less hidden nodes, network may not be learnt, for the number of hidden nodes is too less, local minimum is more, or robustness is not strong, the previous untrained sample can not be identified, poor fault tolerance. Increase hidden nodes may improve the matching accuracy of network and training group, but for the hidden layer nodes are too many and the learning time is too long, the error may not be the optimum. It still can not find a good analytic expression to express the selection of hidden layer nodes, according to previous experience, to design the hidden nodes refer to the following formula:  $n = \sqrt{n_i + n_o} + a$  ( $n_i$  is the input nodes,  $n_o$  is the output nodes,  $a$  is the constant between 1 to 10), Experiments indicates that when the

number of hidden layer nodes selected as 6, the network performance is optimal. Figure 1 is the BP neural network topology, where  $W$  is the weight.



**Figure 1. BP Neural Network Topology**

#### 4.3. Learning Algorithm

Before the network training, you must first measure network convergence to the extent of the expected value. The quantity of this metric is the network error. Because it is a supervised learning, so for a given training set, the expectations are known. Again assuming that can properly training set, can be reasonable description of the problem space, is it can be to train the network.

First define the network error. For back propagation algorithm, network error is known as the mean square error. Mean square error is defined as

$$E = \frac{1}{2} \sum_{j=1}^N (y_j - \hat{y}_j)^2$$

In order to simplify the formulas, define the local gradient as

$$\delta_j = \frac{\partial E}{\partial net_j}$$

Considering the effect of weight  $w_{ij}$  on error, can obtain

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} = \delta_j O_i$$

Weight correction should reduce the error fast, and the modify quantity as

$$\Delta w_{ij} = -\eta \delta_j O_i$$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$$

Where,  $\Delta w_{ij}$  indicates the changes of nerve cell  $j$  weight connecting  $L-1$  layer nerve cell  $i$  and target nerve cell of  $L$  layer, parameter  $\eta$  is the learning rate. If node  $j$  output unit, then

$$\delta_j = \frac{\partial E}{\partial \hat{y}_j} \frac{\partial \hat{y}_j}{\partial net_j} = -(y_j - \hat{y}_j) f'(net_j)$$

if the node  $j$  is not the output unit, then  $O_j$  will affect the sequent nodes. Therefore,

$$\delta_j = \frac{\partial E}{\partial net_j} = \sum_k \frac{\partial E}{\partial net_k} \cdot \frac{\partial net_k}{\partial O_j} \cdot \frac{\partial O_j}{\partial net_j}$$

$$= \sum_k \delta_k w_{jk} f'(net_j)$$

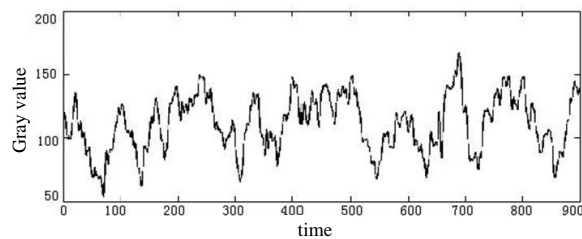
In practical, in order to fast the convergence rate, it often adds the previous weight correction into correction, commonly referred to as inertia item, namely

$$\Delta w_{ij}(t) = \alpha \Delta w_{ij}(t-1) + \eta \delta_j O_i$$

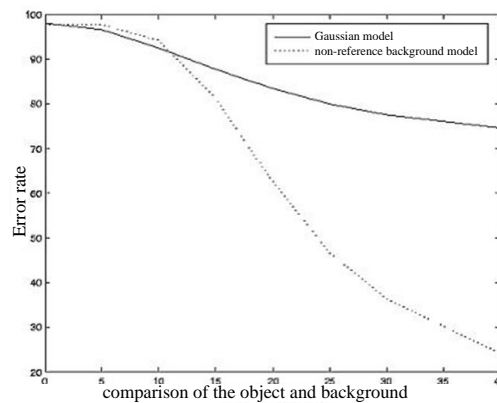
Where,  $\alpha$  is Inertia coefficient.

## 5. Experimental Result

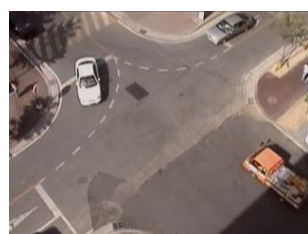
We are taking a series of images from the perspective of traffic intersection fixed sequence as experimental samples. Taking 15 frames image every second, the image size of  $240 \times 320$ . The experimental system established on PC of Intel PIV-3G CPU with 1G internal storage, adopted Matlab 7.1 for experimental simulation. For outdoor scene, background pixel value changing with time, Figure 2 shows a pixel point changes of the grey value in 900 seconds. Figure 3 provides the performance comparison of Gaussian model with the background model, it can be seen that with the increase of target and background contrast, the background model error rates significantly lower than the gaussian model. Through the background model and get the target detection of the results as shown in Figure 4.



**Figure 2. Condition of some Pixel Changing with Time**



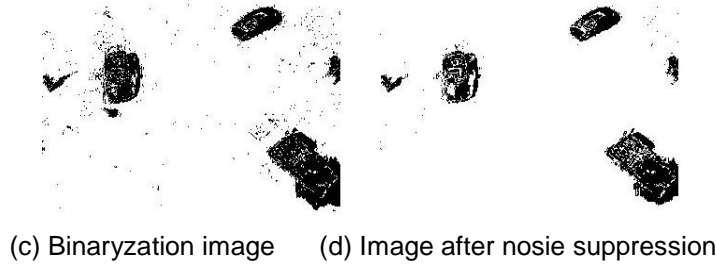
**Figure 3. Performance Comparison of Non-Reference Background Model and Gaussian Model**



(a) Original image

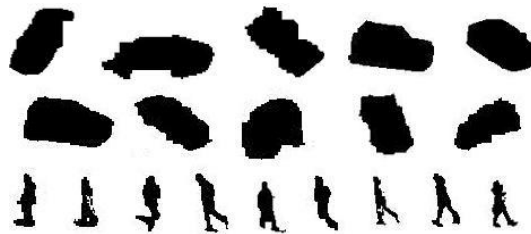


(b) Probability model



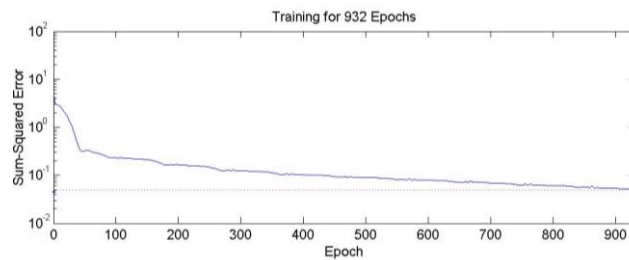
**Figure 4. Experimental Result of Non-Reference Background Model Detection**

This paper collected the pedestrian and the vehicles all 48 samples, each take 36 training samples, the other 12 samples for testing. Set goals network training error is 0.05, the initial weights randomly selected between (-1, 1), coefficient of inertia item 0.95, learning rate, the initial value of 0.1, learning rate increase the ratio of 1.05, learning rate to reduce the ratio of 0.7, maximum error ratio of 1.04. Figure 5 shows the parts of pedestrians and vehicles samples.



**Figure 5. Part Pedestrians and Vehicles Samples**

Learning for training sample set, time was 130 seconds, part of the training set of sample data are shown in table 1. With training after the network sets of test samples for testing, the mean square error is 0.0247, the maximum error component is 0.2164, the recognition rate of 92%. Concrete test set sample data and test results are shown in table 2. Convergence curve is shown in Figure 6.



**Figure 6. BP Network Learning and Training Convergence Curve**

**Table 1. Part Training Set Sample Data and Expectation Output Table**

type	model number	feature vector (sample input value)			expectations
		X1	X2	X3	y
pedestrians	1	0.2222	0.3472	2.3913	1
	2	0.2981	0.2752	1.9231	1
	3	0.2219	0.3240	2.2400	1
	4	0.3090	0.2423	1.8000	1

type	model number	feature vector (sample input value)			expectations
		X1	X2	X3	y
vehicle	5	0.3478	0.2423	1.9091	1
	6	0.3167	0.2496	1.7500	1
	7	0.2111	0.2979	2.1667	1
	8	0.2105	0.3065	2.2917	1
	9	0.4135	0.1663	1.7500	1
	25	0.0960	0.4447	0.7826	0
	26	0.0774	0.5243	1.0147	0
	27	0.0801	0.5857	0.3898	0
	28	0.0783	0.4316	0.8795	0
	29	0.0654	0.5758	0.6023	0
	30	0.0655	0.5444	0.7297	0
	31	0.0676	0.5925	0.5638	0
	32	0.0852	0.4399	0.7037	0
	33	0.0731	0.6618	0.9692	0

**Table 2. Test Sample and Result**

type	No.	feature vector (sample input value)			expectations	actual putput value
		X1	X2	X3	y	Y
pedestrians	21	0.2159	0.2845	2.0697		0.9091
	22	0.3419	0.2259	1.9846		1.0847
	23	0.3049	0.2147	1.9800	1	1.0258
	24	0.2368	0.3569	2.2167		0.9942
	45	0.0854	0.5548	0.4012		-0.0365
vehicle	46	0.0716	0.5644	0.8845	0	0.0219
	47	0.0997	0.3978	0.7719		0.0561
	48	0.0683	0.5913	0.7258		-0.0215

## 6. Conclusion

In the traffic monitoring system, moving target detection and recognition is the key technology of traffic monitoring system. Based on the conclusion of previous background modeling adopt a detection method based on the non- reference background model. The model utilizes a series of latest sampling values to estimate the probability model of observed pixel; then conduct binarization of the probability model to detect moving object, finally obtained the good detecting effect via suppress noise processing. In the aspect of moving object identification, this paper proposed several features, and trained and identified through BP neural network. The experimental results show that pedestrian and vehicle target classification through the characteristics has achieved a high recognition rate.



## Acknowledgements

This work was supported by the Independent Innovation of Agricultural Sciences Of Jiangsu Province, CX(13)5065.

## Reference

- [1] X Li, Z Lv, J Hu, L Yin, B Zhang, S Feng. Virtual Reality GIS Based Traffic Analysis and Visualization System. *Advances in Engineering Software*. (2015).
- [2] D Zeng, Y Geng, Content distribution mechanism in mobile P2P network, *Journal of Networks*, vol. 9, no. 5, pp. 1229-1236, Jan. (2014).
- [3] L Tonglin. "ZHT: A light-weight reliable persistent dynamic scalable zero-hop distributed hash table." *Parallel & Distributed Processing (IPDPS)*, 2013 IEEE 27th International Symposium on. IEEE, (2013).
- [4] W Ke. "Optimizing load balancing and data-locality with data-aware scheduling." *Big Data (Big Data)*, 2014 IEEE International Conference on. IEEE, (2014).
- [5] M Zhou, G Bao, Y Geng, B Alkandari, X Li, Polyp detection and radius measurement in small intestine using video capsule endoscopy, 2014 7th International Conference on Biomedical Engineering and Informatics (BMEI), Oct. (2014).
- [6] X Song, Y Geng, Distributed community detection optimization algorithm for complex networks, *Journal of Networks*, vol. 9, no. 10, pp. 2758-2765, Jan. (2014).
- [7] J Yang, S He, Yancong Lin, Z Lv. Multimedia cloud transmission and storage system based on internet of things. *Multimedia Tools and Applications*. (2016).
- [8] Y Wang, Y Su, G Agrawal. A Novel Approach for Approximate Aggregations Over Arrays. In *Proceedings of the 27th international conference on scientific and statistical database management*, ACM, (2015).
- [9] J Hu and i Gao. Modules identification in gene positive networks of hepatocellular carcinoma using Pearson agglomerative method and Pearson cohesion coupling modularity[J]. *Journal of Applied Mathematics*, 2012 (2012).
- [10] W Ke. "Next generation job management systems for extreme-scale ensemble computing." *Proceedings of the 23rd international symposium on High-performance parallel and distributed computing*. ACM, (2014).
- [11] L Tonglin. "Distributed Key-Value Store on HPC and Cloud Systems." 2nd Greater Chicago Area System Research Workshop (GCASR). (2013).
- [12] T Su, W Wang, Z Lv, W Wu, X Li. Rapid Delaunay Triangulation for Random Distributed Point Cloud Data Using Adaptive Hilbert Curve. *Computers & Graphics*. (2015).
- [13] J Hu and Z Gao. Distinction immune genes of hepatitis-induced hepatocellular carcinoma[J]. *Bioinformatics*, (2012), vol. 28, no. 24, pp. 3191-3194.
- [14] W Ke. "Overcoming Hadoop Scaling Limitations through Distributed Task Execution."
- [15] Z Su, X Zhang, and X Ou. "After we knew it: empirical study and modeling of cost-effectiveness of exploiting prevalent known vulnerabilities across iaas cloud." *Proceedings of the 9th ACM symposium on Information, computer and communications security*. ACM, (2014).
- [16] G Bao, L Mi, Y Geng, K Pahlavan, A computer vision based speed estimation technique for localizing the wireless capsule endoscope inside small intestine, 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Aug. (2014).
- [17] K Leng, W Shi, J Chen, Z Lv. Designing of a I-shaped less-than-truckload cross-dock: A simulation experiments study. *International Journal of Bifurcation and Chaos*. (2015).
- [18] Y Lin, J Yang, Z Lv, W Wei, H Song. A Self-Assessment Stereo Capture Model Applicable to the Internet of Things. *Sensors*. (2015).
- [19] J He, Y Geng, K 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, Nov. (2014).
- [20] W Ou, Z Lv, Z Xie. Spatially Regularized Latent topic Model for Simultaneous object discovery and segmentation. *The 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* 2015.
- [21] W Huang, Y Geng, Identification Method of Attack Path Based on Immune Intrusion Detection, *Journal of Networks*, vol. 9, no. 4, pp. 964-971, Jan. (2014).
- [22] Y Su. In-situ bitmaps generation and efficient data analysis based on bitmaps. In *Proceedings of the 24th International Symposium on High-Performance Parallel and Distributed Computing*, pp. 61-72. ACM, (2015).

## Authors



**TAO Jian-Ping**, He was born in 1982. Now he is a lecture in Jiangsu Academy of Agricultural Sciences. His research interest is mainly in the area of Computer Software, Mechanical and Electrical Integration. He has published several research papers in scholarly journals in the above research areas and has participated in several books.