Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (16): 3221-3226, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3221.3226

Study on an Intelligent Human Detection System for Unmanned Area Security in Ports

[1,2]Mi Chao, [1,2]Huang Youfang, [3]Liu Haiwei, [3]He Xin and [2,3]Mi Weijian
[1]School of Economics and Management, Shanghai Maritime University, China
[2]Container Supply Chain Technology, Engineering Research Center,
Shanghai Maritime University, China
[3]Logistics Engineering College, Shanghai Maritime University, China

Abstract: This study purposes an intelligent human detection system based on video cameras for unmanned area security in ports. The optimized Histograms of Oriented Gradients (HOG) descriptors are used to preprocess the image and the HOG features are then built. And then, a Support Vector Machine (SVM) classifier is applied to process those HOG features and determine whether there are humans in the image. At last, this paper describes a group of experiments of human detection in Coal Terminal of Tianjin Port. The result shows that the computer takes 500 m sec to analyze a 720p image and detect the human and the correctness is 94% which meets the requirements of the unmanned area security in ports.

Key words: Human detection, histograms of oriented gradients, support vector machine

INTRODUCTION

Intelligent video surveillance (Fan *et al*., 2008) uses computer vision to analyze video, image sources automatically in real time (Li *et al*., 2010) without manual intervention. With this technology, the useful key information can be identified and extracted, the goal in the dynamic environment will be located, identified and tracked and the behavior of target will be identified and understood by real-time monitoring of the real-time reflection. In intelligent video surveillance system, how to identify the human is the first problem to be resolved. Effective human behavior recognition algorithm is the core of the system's robustness and practicality.

After a period of development, the now available human detection technologies are the following: firstly, the use of infrared alarm (Chan, 2011), according to changes of the external infrared energy to determine whether someone is moving. Because body's infrared energy is different with the environment. But the unmanned area of a Port is a large scale, the cost would greatly increase and the installation will be very complicated arrangement in this way; Secondly, the idea using ground coil has also been propose. But this method is a certain lack of intelligence for it is not only sensitive for people but also for other moving objects like animals, so it can be unreliable on the whole; in addition, though the way based on human motion (Senst *et al*., 2011) has greatly improved efficiency, especially in the environment of fixed background such as indoor (Yang *et al*., 2007). However, there are many operating machinery and moving motor vehicles in ports' unmanned area, so it may also leads to a large number of false positives for the lack of morphological image feature extraction and intelligent classification. Based on above facts, we must find the appropriate method to detect human on the complex background. It will be capable of identifying human in still images effectively and made high demands to image processing technology on a single image. For example, Changyan put forward the combination between HOG and the Kalman filter to shorten the time required to detect the human and improve the accuracy (Li *et al*., 2010); Wu Dongmei and Li Junwei describe human's characters with Hu moments and Zernike moments which is a form of a matrix function instead of the usual vector form; Hujing Rong proposed a method based on SVM, using wavelet transform (Chand *et al*., 2007) to extract local shape mutations of the target characteristic, then combined with dynamic frame's gait features (Bouchrika *et al*., 2006) and use the Support Vector Machine (SVM) for small sample learning and recognition; Dalal and Triggs (2005) firstly proposed the HOG methods in 2005. Because the change of image geometry and optical has little influence on the detection and some subtle human body movements can be neglected, it particularly suited to do the pedestrian detection in a still image. But the computation time will be too long to meet the requirements of real-time performance. In summary,

Corresponding Author: Mi Weijian, Container Supply Chain Technology Engineering Research Center,
Shanghai Maritime University, Shanghai, China Tel: 0086-21-38282670

this study will do research work based on HOG algorithm and optimized it to decrease the computation time. Meanwhile using linear SVM to learn the samples collected from port's background to enhance the robustness. The Main purpose is to develop a human recognition monitoring system suitable for harbor's complicated background.

HOG OPTIMIZATION ALGORITHM

As used herein, this study describes the human body with HOG (Histogram of Oriented Gradients) features by extracting the shape information and motion information. Thereby it forms a rich set of features which can effectively describe the broad contours of the human body and be not easily affected by the complicated background. So it has strong robustness. Specific feature extraction algorithm process as shown in Fig. 1.

The HOG feature is extracted in the following way:

- **Step 1:** Because the port's illumination is very strong, so it is necessary to use the square root gamma correction operation to suppress image noise

$$f(x,y)=\sqrt{f(x,y)} \tag{1}$$

- **Step 2:** Calculate the gradient intensity $\|\nabla f\|_2$ (Jung *et al*., 2011) and orientation $\theta(x, y)$ using Eq. 2 and Eq. 3:

$$\|\nabla f\|_2=\sqrt{\left[G_x^{\,2}+G_y^{\,2}\right]}=\sqrt{\left(\frac{\delta f}{\delta x}\right)^2+\left(\frac{\delta f}{\delta y}\right)^2} \tag{2}$$

$$\theta(x,y)=\arctan\frac{G_x}{G_y} \tag{3}$$

- **Step 3:** Scanning the whole image with a block of size 16×16 and step N_s from top to bottom, left to right and dividing each block into four cells. The gradient orientation is equally divided into nine parts, namely, accounting for 20 degrees each bin. Consequently 4×9 = 36 dimensional HOG feature vector can be extracted from each block
- **Step 4:** Each cell (8×8 pixels) also be divided into four sub-regions (4×4 pixels) (Pang *et al*., 2010). As shown below It can be discovered from Fig. 2 that these 16 regions are divided into four groups: Group 1: A, D, M, P; pixels within the group only contribute to the belonging cells when calculating the HOG; Group 2: B, C, N, O; pixels not only contribute to the belonging cells but also the left or right cells adjacent. Group 3: E, I, H, L; pixels contribute not only to the belonging cells but also the up or down cells adjacent. Group 4: F, G, J, K; pixels contribute to four cells around it. After this simplification, the calculation time of each cell's gradient orientation calculation is 9 = 1+2+2+4, while the original trilinear interpolation's times is 16 = 4+4+4+4. Therefore, the optimizing method in the calculation of block's HOG feature will decrease the calculation time to 56.25% = 9/16
- **Step 5:** In order to reduce the impact of the background light and the edge of mutation, the feature vector of each block can use the L2-Norm normalization calculation Eq. 4

$$\text{L2-norm},\quad v\leftarrow v/\sqrt{\|v\|^2+\varepsilon^2} \tag{4}$$

- **Step 6:** Save all blocks' feature vectors, then from top to bottom and left to right, using a 64×128 (accord to human body proportion) (Lee *et al*., 2010) detection window to scan the entire image. In each detection window, reading all ((128-16)/8+1)* ((64-16)/8+1) = 105 blocks, 105×36 = 3780 dimensional HOG feature vectors

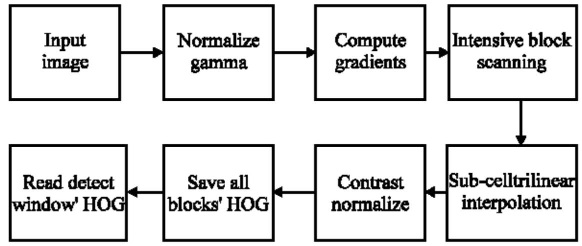

Fig 1: HOG optimization algorithm

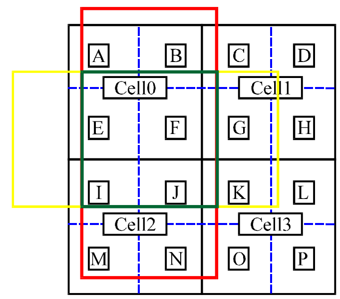

Fig. 2: Schematics of cells' sub-regions

Figure 3 is a graphical representation of the HOG feature vector, the abscissa 1~105 indicates total 105 blocks, the vertical coordinates represents 36 feature

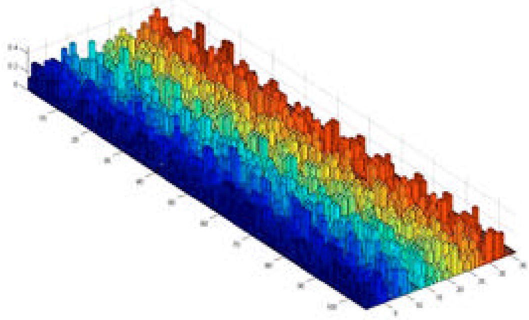

Fig. 3: HOG visual representation

vectors of each block, the z axis indicates the normalized results of block 'HOG feature vectors.

SVM CLASSIFIER

SVM is first purposed by Cortes and Vapnik (Mueen *et al*., 2007), it has obvious advantages when solving the problems of small sample, nonlinear and high dimensional pattern recognition. Since the processed data is non-linear, it is necessary to use the kernel function (Yamashita *et al*., 2011) to accept two low-dimensional space vector and then calculatethe high-dimensional space inner product value through some transform. This study mainly uses its ability of learning and classifying.SVM classification is shown in Fig 4.

The hyper plane's equation can be expressed as Eq. 5:

$$w^T x + b = 0 \quad (5)$$

$$w = \alpha_1 x_1 y_1 + \alpha_2 x_2 y_2 + \cdots + \alpha_n x_n y_n \quad (6)$$

The SVM used in this study is libsvm.

Pedestrian libraries: INRIAPerson and the actual background samples of the port;

The main process of using SVM training is as follows:

- **Step 1:** Complete the sample library according to images captured by the camera of the port
- **Step 2:** Classify and label the positive and negative samples in the completed sample library. All positive samples are marked as +1 and negative are marked as -1 and call hog function for each sample to extract 3780 d hog feature
- **Step 3:** The extracted features are put into SVM to train to get the initial classifier

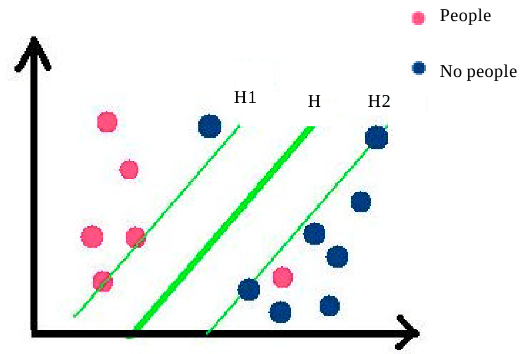

Fig. 4: SVM classification

- **Step 4:** Initial detection model has been obtained and then be used to detect negative samples of the harbor actual background. It can be found that many test results are labeled +1, but actually they are negative samples , so it is wrong
- **Step 5:** Extract the hog feature of the wrong samples in the last step and integrate the obtained hog feature with the hog feature obtained in Step 2 together. Then training to generate the final classifier

FUSION ALGORITHM

For a binary detector, it will generally have a plurality of overlapping rectangular box near the position where the target is in the image, so the results need to be fused. A mode detection program based on mean shift (Loum *et al*., 2007) is used to point to the position of maximum probability density.

- **Step 1:** First select all positive results into a list and project them to the three-dimensional space, expressed as point y_i, saving coordinate information and scale information

$$\begin{cases} y_i & \text{result} = 1 \\ \text{ignore} & \text{result} = -1 \end{cases} \quad (7)$$

- **Step 2:** The bandwidth matrix can be calculated

$$\text{diag}[H_i] = \left[\left(\exp(s_i)\sigma_x\right)^2, \left(\exp(s_i)\sigma_y\right)^2, (\sigma_s)^2 \right] \quad (8)$$

- **Step 3:** Iteratively calculate the mean shift vector for each point until they are merged into a modal

$$m(y) = H_h \frac{\nabla \hat{f}(y)}{\hat{f}(y)} \equiv y_m - y \quad (9)$$

$$y_m = H_h(y_m)\left[\sum_{i=1}^{n} \bar{w}_i(y_m) H_i^{-1} y_i\right] \quad (10)$$

- **Step 4:** The modals obtained above are the final fusion results

Figure 5 shows the fusion process in the image. The yellow rectangles are the results before fusion, it can be found that these rectangles distribute in two areas, one covers a people while the other is not people. After image processing, all rectangles aggregate into only one red rectangle. Combined with the current scale, it is a bit smaller than the yellows to fit the human body.

All convergent results are discrete points in three dimensional space. As shown in Fig. 6, yellow points are the discrete results of initial detecting. Starting from a random yellow point and then iterating its neighborhood points to calculate the convergent location using Eq. 9 red point in the figure. But it has certain position deviation with the red points, the original result in space. In order to guarantee the accuracy of the human's position, it is necessary to consider one of the discrete yellow points which nearest the convergent location as the convergence result:

Fig. 5: The human (from INRIAPerson) detect result before (yellow) and after fusion (red)

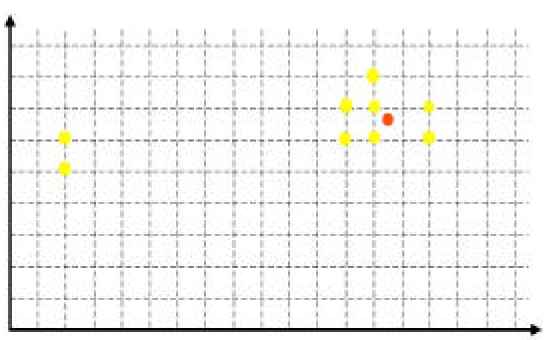

Fig. 6: Discrete points' convergence

$$\begin{cases} d = \min\left\{\sqrt{(x_i - x_r)^2 + (y_i - y_r)^2}\right. & (i \geq \text{threshold}) \\ \text{false positive} & (i < \text{threshold}) \end{cases} \quad (11)$$

In Eq. 11, i indicate the number of points in neighborhood oriented, (x_r, y_r) is the coordinate of the red point, so the smallest distance to it will be calculated and the point is the final result. Actually there are two diffident areas in the figure, but the left area only have two points, less than the threshold, therefore it is regard as a false positive.

EXPERIMENTAL RESULTS

As shown in Fig. 7 and 8, Hardware architecture of the artificial intelligence-based adaptive visual recognition system consists of four parts: A high-speed video network HD camera, mounting platform (including camera bracket, galvanized outdoor lamp posts, outdoor waterproof plug pole box), wireless communications equipment, a server used to display the monitor screen and do some simple operation.

The detection images in Coal Terminal field operation areas are shown in Fig. 9.

After 24 h of continuous observation with a mutable weather in Tianjin port, 235312 images were analyzed by the human detection system. It can be calculated that the computation period for every frame of 720p image is less than 500 m sec^{-1} which meets the real-time requirements of the ports.

Fig. 7: High-speed video network HD camera

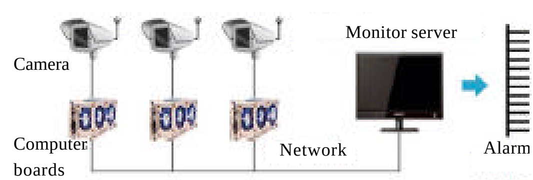

Fig. 8: Hardware structure of the system

Fig. 9: The result of the human detection system

Table 1: Experimental results of accuracy

Experimental environment	True detection	False detection	Accuracy (%)
Clear			
Day	22	1	95.70
Night	6	0	100.00
Rainy			
Day	16	2	89.00
Night	3	0	100.00
Total	47	3	94.00

And Table 1 shows the accuracy of the 24 h experiment. Statistically analysis on the accuracy results of the system: it can be obtained that the true detection rate of 94% which proved that this algorithm has a relatively strong ability to identify the human.

According to the experimental results in Coal Terminal of Tianjin Port, the human detection system with optimized HOG algorithm and the SVM classifier shows its real-time computation efficiency and high accuracy.

CONCLUSION

This study described the corresponding technology of human detection based on HOG descriptors and SVM classifier. For better computation efficiency, an optimized HOG calculation method was raised. After several uninterrupted field experiments, the results show that the human detection system with the optimized HOG descriptors and SVM classifier can analyze one 720p in 500ms. And the human detection accuracy rate remained at 94%. The experimental results show the system meets the real-time and accuracy requirement of the unmanned area security in ports.

ACKNOWLEDGEMENT

This research is supported by “Young University Teachers' Training Project” of Shanghai Education Commission and “Local University Capacity Promotion Special Programs (13510501800)” of Science and Technology Commission of Shanghai Municipality.

REFERENCES

Bouchrika, I. and M.S. Nixon, 2006. People detection and recognition using gait for automated visual surveillance. Proceedings of the IEE International Symposium on Imaging for Crime Detection and Prevention, June 13-14 2006, London, pp: 576-581.

Chan, A.L., 2011. Multi-stage infrared stationary human detection. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 22-27, 2011, Prague, pp: 1221-1224.

Chand, P., A. Davari, B. Liu and K. Sedghisigarchi, 2007. Feature extraction of power quality disturbances using adaptive harmonic wavelet transform. Proceedings of the 39th Southeastern Symposium on System Theory, March 4-6, 2007, Macon, GA., pp: 266-269.

Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and P a ttern Recognition, Volume 1, June 20-26, 2005, San Diego, CA., USA., pp: 886-893.

Fan, X., L. Xu, X. Zhang and L. Chen, 2008. The research and application of human detection based on support vector machine using in intelligent video surveillance system. Proceedings of the 4th International Conference on Natural Computation, October 18-20, 2008, Jinan, pp: 139-143.

Jung, H., J.K. Tan, S. Ishikawa and T. Morie, 2011. Applying HOG feature to the detection and tracking of a human on a bicycle. Proceedings of the 11th International Conference on Control, Automation and Systems, October 26-29, 2011, Gyeonggi-do, pp: 1740-1743.

Lee, K., C.Y. Choo, H.Q. See, Z.J. Tan and Y. Lee, 2010. Human detection using histogram of oriented gradients and human body ratio estimation. Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology, Volume 4, July 9-11, 2010, Chengdu, pp: 18-22.

Li, C., L. Guo and Y. Hu, 2010. A new method combining HOG and Kalman filter for video-based human detection and tracking. Proceedings of the 3rd International Congress on Image and Signal Processing, October 16-18, 2010, Yantai, pp: 290-293.

Loum, G., C.T. Haba, J. Lemoine and P. Provent, 2007. Texture characterisation and classification using full wavelet decomposition. J. Applied Sci., 7: 1566-1573.

Mueen, A., M.S. Baba and R. Zainuddin, 2007. Multilevel feature extraction and X-ray image classification. J. Applied Sci., 7: 1224-1229.

Pang, Y., Y. Yuan, X. Li and J. Pan, 2010. Efficient HOG human detection. Signal Process., 91: 773-781.

Senst, T., R.H. Evangelio and T. Sikora, 2011. Detecting people carrying objects based on an optical flow motion model. Proceedings of the IEEE Workshop on Applications of Computer Vision, January 5-7, 2011, Kona, HI., pp: 301-306.

Yamashita, A., Y. Ito, T. Kaneko and H. Asama, 2011. Human tracking with multiple cameras based on face detection and mean shift. Proceedings of the IEEE International Conference on Robotics and Biomimetics, December 7-11, 2011, Karon Beach, Phuket, pp: 1664-1671.

Yang, T., F. Chen, D. Kimber and J. Vaughan, 2007. Robust people detection and tracking in a muti-camera indoor visual surveillance system. Proceedings of the IEEE International Conference on Multimedia and Expo, July 2-5, 2007, Beijing, pp: 675-678.