Fast Aerial UAV Detection Using Improved Inter-frame Difference and SVM

Li Xiaoping\textsuperscript{1,a}, Lei Songze\textsuperscript{2,b}, Zhang Boxing\textsuperscript{3,c}, Wang Yanhong\textsuperscript{4,d}, Xiao Feng\textsuperscript{5,e}

\textsuperscript{1}School of Computer Science and Engineering
Xi’an Technological University
86-15289387309, 710021

\textsuperscript{2}School of Computer Science and Engineering
Xi’an Technological University
86-18991896239, 710021

\textsuperscript{3}School of Computer Science and Engineering
Xi’an Technological University
86-17719755818, 710021

\textsuperscript{4}School of Science
Xi’an Technological University
710021

\textsuperscript{5}School of Computer Science and Engineering
Xi’an Technological University
710021

\textsuperscript{a}919083845@qq.com \textsuperscript{b}lei_sz@163.com \textsuperscript{c}1076781906@qq.com \textsuperscript{d}29314998@qq.com \textsuperscript{e}544070146@qq.com

ABSTRACT: In order to detect UAV in real time, the paper choose to use a dynamic detection method based on two consecutive inter-frame differences method to extract the region of interest. The position of the target appeared on the image was obtained by the method of two consecutive inter-frame difference, and the UAV was detected by the trained SVM classifier. UAV could be detected quickly and accurately in complex background and in different position and angle circumstances. Compared to the traditional HOG+SVM sliding window detection method, the experimental results show that the detecting speed with the methods is obviously improved when the recognition accuracy is invariable.

1. INTRODUCTION
In recent years, UAV(unmanned aerial vehicle) have been widely used in the military field. They could monitor troops in real time and detect important military targets and find all military regions and carry offensive weapons to attack hostile military regions. In order to avoid the destruction of troops and military regions by hostile UAV, it is necessary to detect UAV in real time.

The main difficulty detecting aerial UAV in real time is that the aerial non-cooperative UAV is easy to be affected by its own motion, which could change the angle, position and structure of the UAV in the video. As a result, the robustness of UAV features is low and it is difficult to detect accurately. The difficulty of target detection is similar to UAV detection, so it is worth studying and
practicing in order to obtain accurate and real-time detection method for aerial UAV. In recent years, the target detection method based on machine learning is easy to be used, which mainly includes two steps: region of interest segmentation and target detection. The methods segmenting regions of interest mainly includes moving object detection [2] and static object detection, including optical flow field method [1], background estimation, inter-frame difference method [2], and image segmentation method [17]. Image segmentation is insensitive to illumination and is suitable for scenes with simple background and clear target. The optical flow field method is suitable for all kinds of backgrounds. However, it’s calculation need to use too many pixels, large amount of computation and insufficient real-time performance. The method of inter-frame difference is easy to be affected by illumination. To sum up, it is difficult to detect small UAV by image segmentation when the UAV is far away from the camera and the background is complex and the illumination change is weak, and it is easy to misunderstand some complex background.

A real time detection method for aerial UAV using two continuous frame difference method combined with SVM is proposed in the paper. The pixels of the corresponding position of each successive two frame image in the video were subtracted and the absolute value was taken, and then the two adjacent values were performed or calculated, and the absolute value was compared with the predetermined threshold value, if the absolute value was greater than the threshold value, then the location of the absolute value of the frame is recorded and segmented in the gray image. Then it is sent to the support vector machine (SVM) model trained by gradient direction feature vector and Fisher linear discriminant analysis (HOG-FLD) in the methods.

2. SEGMENTING OF REGION OF INTEREST BASED ON INTER-FRAME DIFFERENTIAL METHODS

The improved inter-frame difference method is used in the paper. The inter-frame difference method [6] is to take the absolute value $D_n(x,y)$ for the gray value of the corresponding position of each successive frame in the video, and then the two adjacent values is performed OR operation, and the inter-frame differential image $P_n(x,y)$ is obtained. When the absolute value is greater than the pre-determined threshold value $T_3$, it is converted to binary image $R_n(x,y)$ to obtain the contour information of UAV. Its specific principles are as follows formula(1) , (2), (3):

$$\begin{align*}
P_n(x,y) &= |f_1(x,y) - f_{k-2}(x,y)| \quad (1) \\
D_n(x,y) &= P_n(x,y) | P_{n+1}(x,y) \quad (2) \\
R_n(x,y) &= \begin{cases} 
0, & D_n(x,y) < T_3 \\
1, & D_n(x,y) \geq T_3
\end{cases} \quad (3)
\end{align*}$$

Among them, “|” is used to OR operation, $k = 2, 4, 6, 8, \cdots 2n$, $n = 1, 2, 3, 4, \cdots n$.

The captured video is captured to a small video containing 90 frames in order to complete the real-time detection in the paper. After the video is processed by inter-frame differential methods, the binary video is saved as a file in avi format and the region of interest is segmented in the later stage. The results processed by the two consecutive frames difference method are shown in figure 1:

![Comparison Image](image)

(a) Original Image  (b) Two Successive Inter-Frame Differential Image

Figure 1. Comparison Image.

Because there are many noise points in the binary image and the target is prone to cavitation, the target location to be tagged is too wide and the number of targets to be tagged is not practical. So it is necessary to be processed used the morphological methods on the threshold image. The cavitation is filled by used the closed operation and the adjacent pixels is connected, and the open operation is used to remove the noise and the size of the target is not changed, which makes the target more full and
helps to save the location and detect the target in this paper. According to the constructed disk radius of 10 structural elements $se$, the binary image is first closed and then opened in the paper. The principle is respectively shown in expressions (4) and (5).

$$Xo se = (X \ominus se) \Theta se$$  (4)

$$Xo se = (X \Theta se) \oplus se$$  (5)

In the formula, $X$ is the binary image to be processed; $\Theta$ is the operator for the morphological corrosion operation [6]; and $\oplus$ is the operator for the morphological expansion operation.

Then the connected region marking function (bwlabel) in matlab is selected to save the position of eight connected regions with a threshold of 1, and the region of interest is segmented from the corresponding position of the gray image of the corresponding frame, and the region of interest is normalized to a pixel of 128*64. The segmentation results are shown in figure 2:

(a) Region 1                (b) Region 2

(c) Region 3                (d) Region 4

Figure 2. Area of Interest Image.

3. TARGET UAV DETECTION BASED ON HOG-FLD FEATURE FUSION AND SVM DETECTION

3.1 Feature Extraction From HOG-FLD Feature Fusion

Because of the strong stability and extensibility of the edge contour feature of the rotor UAV, the HOG [3](Histogram of Oriented Gradient) feature describing the edge contour feature of the object is calculated from the picture. The HOG feature is not easy to be affected by the local minor deformation of the object and its dimension is large, which is not conducive to the training of the classifier and to meet the real-time requirements of the subject. HOG-FLD [19] feature fusion method is used for feature extraction of image files in the paper. On the basis of the definite contour features, the computation amount is reduced to improve the speed of the algorithm, and the features that are favorable to the classification can be extracted, and the precision rate and processing time of the algorithm can be greatly improved.

The basis of feature extraction is to extract the characteristic of HOG. The idea of the algorithm is to calculate the edged gradient of image to be entered; each image is divided into rectangles with fixed size and equal size as cell containing pixel of $m \times m$; and all cells are divided into 18 directional channels or 9 non-directional channels, and the gradient histogram of all directions is all voted, and the weight used to vote is the gradient value calculated in previous step. Then the fixed blocks of the same size is combined by above-divided units, that contains $n \times n$ cells. And the local feature vectors corresponding to each block are normalized so that it make the effect of the experimental from the light from the image reduced, and the HOG feature vectors of the image are combined by the feature vectors of the blocks.

To extract HOG feature, the positive sample image of UAV is firstly grayed and filtered with Gamma correction method to make the image meet the standard requirement that the image preprocessed could make the influence of local shadow and light change reduced. Then the image is divide into a number of cells, they could formed a number of fixed blocks of the same as size. For blocks of the same size in the previous paragraph, the gradient range is classified by the above rules.
can calculate the cell characteristics in these blocks and eventually connect all the blocks and the feature vector of image is obtained. The formula normalizing the image is as Formula (6); the formula calculating the gradient component of each pixel is as formula (7) and formula (8); the formula calculating the size and direction of the gradient is as formula (9) and formula (10).

\[ v_g^* \leftarrow \sqrt{v_g^2/(\parallel v_g^2 \parallel +\varepsilon)} \]  

\[ G_{x(x,y)} = p_{i(x+1,y)} - p_{i(x-1,y)} \]  

\[ G_{y(x,y)} = p_{i(x,y+1)} - p_{i(x,y-1)} \]  

\[ S_{i(x,y)} = \sqrt{G_{x(x,y)}^2 + G_{y(x,y)}^2} \]  

\[ \theta_{i(x,y)} = \arctan(G_{y(x,y)}/G_{x(x,y)}) \]

Where, \( v_g^* \) is the result of histogram normalization, and \( v_g \) is the extracted vector histogram; \( p_{i(x+1,y)} \), \( p_{i(x-1,y)} \), \( p_{i(x,y+1)} \), and \( p_{i(x,y-1)} \) respectively denote the location of 4 pixel points; \( G_{x(x,y)} \) and \( G_{y(x,y)} \) respectively denote the coordinate position in the horizontal and vertical directions of the two pixels. And \( S_{i(x,y)}, \theta_{i(x,y)} \) respectively denote the length of the gradient direction vector and the angle of the gradient direction vector.

In this article, a window including 64 * 128 pixel is used to scan the sample image and the image to be detected using a window including 64 * 128 pixel. The scanning step size is 8 pixels (scanning is in horizontal and vertical direction). The window is divided into cells including 8 * 8 pixel and forms 8 * 16 = 128 units. Then setting up four adjacent units up and down to the left and right as a block of pixels, a window contains 105 blocks of pixels. A 3780 dimensional feature vector named HOG feature description value is generated in a window containing 105 pixel blocks according to the calculation steps of HOG. Its specific HOG algorithm is shown in Figure 3:

![Figure 3: The Principle of HOG Feature](image)

On the basis of HOG feature, the linear subspace is constructed by using FLD[18]. By calculating the optimal projection matrix, the methods could obtain projection matrix for feature extraction of the training set, and the similarity of its projection vector is taken as the similarity degree of cosine similarity, which purpose is to reduce the intra-class dispersion \( S_W \) as much as possible and to increase the inter-class dispersion \( S_B \) as much as possible (or that is in the training concentration, to make the sample data of the same UAV as close as possible, and the sample data of different UAV to be far away). In this way, the features with classification ability are extracted. When dealing with a problem that type is \( c \), the mathematical formulas of inter-class dispersion \( S_B \) and intra-class dispersion \( S_W \) should be defined as follows:

\[ S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \]  

\[ S_W = \sum_{i=1}^{c} \sum_{x_i \in A_i} (x_i - \mu_i)(x_i - \mu_i)^T \]
Where $\mu_i$ denotes the mean value of class $\Omega_i$; $\mu$ means the mean value of total sample; $N_i$ means the number of samples of class $\Omega_i$. The optimal projection matrix $W_{opt}$ could be obtained by solving the optimization problem such as formula (13), where $Sw$ must be a nonsingular matrix (or that is the total number of training samples $N$ is greater than the characteristic dimension of UAV image):

$$W_{opt} = \arg \max_{W} \frac{|W^T Sw W|}{|W^T S \omega W|}$$

(13)

$W_{opt}$ could also be obtained by solving the generalized eigenvalue problem such as the formula (14):

$$Sw W = \Lambda S \omega W$$

(14)

In order to solve the problem that the intra-class dispersion matrix $Sw$ is singular, PCA principal component analysis (PCA) is used to reduce the dimension of the feature space (dimensionality reduction to N-cu), and then Fisher linear discriminant analysis (FLD) is used to deal with it. The projection vector $y$ of test sample $x$ is obtained according to formula (15):

$$y = W_{opt}^T x$$

(15)

Cosine similarity $Scos$ is used as the similarity measure of projection vector $y$. Where the cosine similarity $Scos$ of vector $A = \{a_1, a_2, \cdots, a_n\}$ and $B = \{b_1, b_2, \cdots, b_n\}$ is defined as follows:

$$Scos = \frac{(A, B)}{||A||_2 ||B||_2} = \sum_{i=1}^{n} a_i b_i$$

(16)

3.2 Support Vector Machine

Because the support vector machine (SVM) proposed by Vapnik has the advantages of simple system structure, global optimization, good generalization, and short training and prediction time [9], this paper uses SVM as a machine learning tool to calculate the rule of samples in order to achieve fast and efficient learning sample features and accurate classification purposes. The main idea of SVM is to deal with the linear inseparability of the original space by selecting the kernel function of Polynomial Kernel to correspond the data to the high-dimensional space. When the algorithm is used to realize the two-classification, the sample features such as HOG must be extracted from the original space first, and then the sample features in the original space are represented as a vector in the high-dimensional space. In order to minimize the error rate of the two-class classification problems, we need to find a hyperplane that is used to divide the two classes in the high-dimensional space.

Let the sample set be $x_i, y_i$ where $i=1,2,\ldots,N, x_i \in \mathbb{R}^n, y_i \in \{0,1\}$ is the class identifier. Then in each dimensional space, its linear discriminant function is as follows:

$$g(x) = w^* x + b$$

(17)

The formula of the classification surface equation is as follows:

$$w^* x + b = 0$$

(18)

After normalization of the discriminant function, the following conditions must be satisfied for the two types of samples:

$$g(x) \geq 1$$

(19)

The classification interval could be $2/||w||$, where the maximum requirement $||w||$ for the classification interval is kept to a minimum, all samples must be correctly classified, and the following conditions must be met:

$$y_i [(w^* x_i) + b] - 1 \geq 0$$

(20)

An SVM whose inner product function is $k(x_i, x_j)$ is constructed with the following formulas (which can be understood as the formula for calculating the extreme value of a quadratic function with conditional constraints):

$$Q(a) = \sum_{i=1}^{N} a_i y_i k(x_i, x_j)$$

(21)
Its constraints are expressed as: \( 0 \leq a_i \leq C, \sum_{i=1}^{N} a_i y_i = 0 \)

The formulas of the support vector machine that could be calculated are as follows:

\[
f(x) = \text{sgn} \left( \sum_{i=1}^{N} a_i y_i k(x_i, x) + b^* \right)
\]

(22)

Among them, \( b^* \) is a constant parameter, indicating the size of the threshold that needs to be classified.

3.3 Obtaining the Sample and Training Classifier Model

Firstly, 500 original images of positive samples are used to statistic the aspect ratio of UAV, and the statistics show that aspect ratio of UAV is 1:2, so the pixels of each sample image are normalized to 64 * 128 to avoid the effect of image size on the recognition effect of the algorithm. Finally 1400 positive sample images and 1400 negative sample images are used to train SVM model in the paper. As described in the HOG-FLD feature part above, the input image all could create 105 pixel blocks which could get a 36 dimensional feature vector respectively, so 36 dimensional feature vector finally could combined into a 3780 dimensional feature vector as HOG feature vector of the input image. Then the extracted HOG vector is used as the input vector of FLD analysis algorithm to cut back the dimension of the feature vector of the whole image. According to the experimental recognition efficiency of of the whole algorithm, the dimension of the vector is adjustable and determined with its parameter changed. Sending the final extracted HOG-FLD feature vector to the SVM model, and the SVM classifier that could detect the input image is if the UAV could be trained .

4. PARAMETER ANALYSIS AND EXPERIMENTAL RESULTS

4.1 Parameter Adjustment of HOG-FLD

In the experiment of extracting feature vector, the parameter \( k \) of code of FLD algorithm represent the dimension of the feature vector that needs to be cut back by FLD analysis algorithm, with the change of the parameter \( k \), the dimension of the feature vector is also changing. As shown from Figure 4, the contrast graph of recognition time of the algorithm is obvious with the change of the parameter \( k \) of the algorithm. It could be clearly seen from the line chart that when parameter \( k \) equals 50, the recognition time of the detected algorithm is the least in the paper.

![Figure 4. the Time Line Chart of with k Changed.](image)

4.2 Experimental Effect of Aerial UAV Detection Algorithm Based on Region of Interest

Experiments show that in the whole algorithm, when the number of pixels in the divided blocks is 8*8, SVM kernel function type is 2, when the threshold of segmentation is \( k=90 \) and \( \sigma=10 \), the algorithm has the best overall recognition effect on accuracy rate and time, and the recognition result is shown in Figure 5:
4.3 Comparison of Detection Effect of Traditional HOG-SVM UAV Detection Algorithm

The experimental results of the proposed method are compared with the experimental results of the HOG and SVM method based on image segmentation, and the experimental results are obtained by statistics so that it could be verified its own efficiency. According to the accuracy of one of the evaluation criteria in the machine learning algorithm, $N$ is used to represent the number of regions of interest per test image, $TP$ is used to represent the number of UAVs recognized as UAVs, and $TN$ is used to represent the number of non-UAVs recognized as non-UAVs. The ACC formula for accuracy is:

$$ACC = \frac{TP+TN}{N}$$ (23)

As shown from Table 1, the recognition accuracy rate of the algorithm is close to the latter algorithm and the recognition time of the algorithm is faster than that of the latter algorithm.

| Test Algorithm | ACC     | Test Time |
|----------------|---------|-----------|
| HOG_FLD+SV     | 92.45%  | 0.093s    |
| HOG+SVM        | 92.60%  | 0.162s    |

5. CONCLUSION

The method based on inter-frame difference to segment the region of interest is used in the paper. In the testing stage, the acquired regions of interest are input into the trained model of SVM classifier, which cuts back the recognition time of the whole algorithm in the extent. Extracting the HOG-FLD feature of the input image is easier for training SVM classifiers. In the platform of Matlab, compared traditional HOG+SVM detection algorithm to the algorithm in the paper, the experimental results show that the detection algorithm in the paper is better than the sliding window (HOG+SVM) algorithm to detect UAV in terms of time and accuracy.

ACKNOWLEDGMENTS

Fund projects: Key projects in the Industrial Field of Shaanxi (2016KTZDGY4-09), Scientific Research Program Funded by Shaanxi Provincial Education Department (17JK0364), National Natural Science Foundation of China (61572392), National Joint Engineering Laboratory of New Network and Detection Foundation (Grant No. GSYSJ2016008).

REFERENCES

[1] J. Barron, D. Fleet, S. Beauchemin. (1944). Performance of optical flow techniques. International Journal of Computer Vision, 12: 42-77.
[2] A. Lipton, H. Fujiyoshi, R. Patil. (1998). Moving target classification and tracking from real-time video. Proc IEEE Workshop on Applications of Computer Vision, 8-14.

[3] DALALN, TRIGGS B. (2005). Histograms of oriented gradients for human detection. Proc IEEE Conference on Computer Vision and Pattern Recognition, 1-8.

[4] Qiang Zhu, Shai Avidan, Mei Chen Yeh, and Kwang Ting Cheng. (2006). Fast human detection using a cascade of histograms of oriented gradients. Proc.IEEE international Conference on Computer Vision and Pattern Recognition.

[5] Suard F, Akotomamony A R, Bensrhair A, etal. (2006). Pedestra in Detection Using Infrared Images and Histograms of Oriented Gradients. Proceedings Intelligent Vehicle Symposium, 206-212.

[6] Wang Maosen, Chen Long, Dai Jinsong. (2016). Research on UAV Motion Detection Based on Improved Frame Difference Method. Electricity and Automation, 45: 165-168.

[7] Sun Ting, Qi Yingchun, Geng Guohua. (2016). Moving Target Detection Algorithm Based on Inter-frame Difference and Background Difference. Journal of University, 46: 1325-1329.

[8] Vapnik V N. (1995). The nature of statistical learning theory. Springer-Verlag, New York, 37-69.

[9] Guo Mingwei, ZhaoYuzhou, Xiang Jumping, etal. (2014). A Survey of Target Detection Algorithms Based on Support Vector Machine. Control and Decision, 29: 192-200.

[10] Zhang Han, He Dongjian. (2011). Cross-camera Moving Target Detection and Recognition. Automation Technology and Application, 30: 43-46.

[11] Zhai Jiyou, Zhuang Yan. (2017). Significant Detection of Boundary Prior and Adaptive Region Merging. Computer Engineering and Application.

[12] Chen Shanchao, Fu Hongguang, Wang Ying. (2012). Application of An Improved Graph Segmentation method in Tongue Image Segmentation. Computer Engineering and Application, 48: 201-203.

[13] Yan Yu, Song Wei. (2016). Color and Texture Mixed Descriptor Image Retrieval Method. Computer Science and Exploration, 1-8.

[14] Chan Qiwen. (2011). ROI Detection Algorithm for Small Infrared Target in Infrared Image Based on Hypothesis Testing. Computer and Modernization, 8: 135-137.

[15] Zhao Zhuxin. (2012). Estimation of Target’s Motion Parameters Using Line-scan Camera. Opto-Electronic Engineering, 36-40.

[16] Girshick R, Donahue J, Darrell T, etal. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Computer Science, 580-587.

[17] Wu Dapeng. (2010). Cam-shift Object Tracking Algorithm Based on Inter-frame Difference and Motion Prediction. Opto-Electronic Engineering, 1: 210-213.

[18] Belhumeur P, Kriegman D. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19: 711-720.