**TELKOMNIKA Indonesian Journal of Electrical Engineering**
Vol. 12, No. 11, November 2014, pp. 7772 ~ 7777
DOI: 10.11591/telkomnika.v12i11.6030


# Robust Multiple Ship Tracking in Inland Waterway CCTV System

**Fei Teng*[1], Qing Liu[2]**
[1]School of Energy and Power Engineering, Wuhan University of Technology, Wuhan, 430070, China
[2]Wuhan University of Technology, Wuhan, 430070, China
*Corresponding author, email: changeerhao_love@126.com

***Abstract***

*In recent years, single object tracking has been extensively studied and achieved much development. However, multiple objects tracking is still an issue that remains to be addressed. Generally speaking, existing multiple objects tracking methods employ a manner of simultaneously tracking each object respectively. In this paper, we develop a multiple ship tracking algorithm based on deformable part model to accomplish multiple ship tracking in inland waterway CCTV (Closed-Circuit Television) automated surveillance. Our method utilizes HOG features to construct the appearance models of ships. Then by taking full advantages of the spatial constrains between ships, we can successfully explore mutual relations for multiple ships, thus accomplishing multiple ship tracking in its true sense. Moreover, structured learning method is used to learn how to update the model parameters. Numerous experimental results on challenging inland waterway CCTV video sequences demonstrate that our method can effectively and accurately perform robust multiple ship tracking.*

***Keywords:*** *multiple ship tracking, CCTV system, Deformable part model, mutual relations*



## 1. Introduction

In recent years, CCTV automated surveillance system has made great contribution to inland waterway management. Compared with artificial monitoring, CCTV system makes it possible to keep full 24 hours' surveillance and cruise forensics process records. In the application of electronic cruise in inland waterway, CCTV system plays as an eye's role in cruise check, thus guaranteeing ship trajectory tracking, illegal disposal, safety forewarning and so on.

Object tracking is a fundamental challenge in computer vision with applications in a wide range of domains such as human-computer interaction, moving recognition, video index and so on. By now, object tracking has achieved remarkable progress. In general, tracking methods fall into two categories: generative tracking method and discriminative tracking method [1]. Generative tracking method models the object of interest by just describing the object appearance [2-6]. Discriminant tracker models both the object of interest and the background. It focuses on finding a decision boundary to separate the object from the background [7-10]. Due to the considered background information, discriminative tracking method always outperforms the generative ones.

Discriminative tracking method accomplishes tracking by detection. In the field of object detection, many progresses and achievements have been demonstrated. SVM classifier with the HOG feature is a typical detector for detecting specific human [11], face [12] and so on. In recent years, another detector based on the deformable part model [13] also draws public attention due to its performable detection [14]. It is known as one of the best detectors. Recently, Lu Zhang et al. [15] propose a structure preserving object tracking method to tracking arbitrary objects based on a single (bounding box) annotation of object of interest in the first frame. It demonstrates favorable performance by combining the success of the Dalal-Triggle detector and the deformable part model. Genquan Duan et al. [16] introduce a mutual relation model to group multiple objects together.

According to the online algorithm they proposed, they can track multiple unseen objects. Motivated by the method in [16], we propose a multiple ship tracking method to be applied into CCTV surveillance video sequences in inland waterway.

*Received March 29, 2014; Revised June 20, 2014; Accepted July 10, 2014*

The remainder of this paper is organized as follows: In Section 2, we describe our tracker in details. In Section 3, we perform experiments to show how our multiple ship tracker works in inland waterway CCTV videos. Finally, the conclusion is given in Section 4.

## 2. Proposed Tracker

Generally, a tracking system comprises three main components [17]: an appearance model, location model and a search strategy. In the proposed multiple ship tracker, we model the appearance of our ships of interest with the Dalal-Triggs detector [11]. Other than tracking multiple ship respectively, we utilize the mutual relation models [16] based on deformable part model to describe a configuration of ships' state. Then, an online structured SVM [18] is adopted to learn and identify the configurations of ships. As for search strategy, a sliding-window exhaustive search is an advisable choice.

### 2.1. Appearance Model

Feature representation is critical to the performance of a tracker. Our tracker uses the HOG features to represent each ship of interest and SVM to obtain their appearance model respectively.

HOG features measure the magnitude and the orient of the image gradient. Actually, we first assume to divide the image patch into many 8*8 pixel cells without overlapping. Then we calculate the gradient in each cell using the simple [-1, 0, 1] masks for horizontal direction and [-1, 0, 1]' for vertical direction. Each pixel calculates a weighted vote for an edge orientation histogram channel based on the orientation of the gradient element centered on it, and the votes are accumulated into (unsigned) orientation bins over cells. Thirdly, each two adjacent cells comprise a block. The histogram of gradient of those blocks would be normalized with the L2-norm respectively. Finally, by simply jointing all histograms of blocks, we can obtain the HOG features of the object.

Dalal-Triggs detector shows optimal performance in the practical applications by many scholars. In addition, HOG features also have many theoretical advantages. Firstly, it counts for not only the horizontal and vertical directions but also many other directions so as to describe the ships more precisely. Secondly, it is illumination-invariant due to the normalization and that it works on the relatively small patches that called cells. Both of them are helpful for us to model those ships appearance effectively.

Next, we take advantage of the typical linear SVM to model each ship. Linear SVM is the simplest support vector machine. It classifies large number of samples into two types, positive or negative that labeled with 1 or 0, with a linear discriminative function. According to maximum margin principle, we can finally obtain the corresponding weight output as our appearance models by a certain amount of training samples with known labels.

### 2.2. Mutual Relation Model Based on Deformable Part Model

The multiple ship tracking framework we adopt is the mutual relation model [16]. We denote all the ships of interest as V. Each ship patch i∈V is represented as $O_i = \{\mathbf{x}_i, w_i, h_i\}$ with center location $\mathbf{x}_i=(x_i, y_i)$, width $w_i$ and height $h_i$. Then the model can be intuitively described using (1).

$$f(O_i) = \sum_{i\in V} w_i \cdot \phi\left(O_i\right) + \sum_{(i,j)\in E} w_{i,j} \cdot \varphi\left(O_i, O_j\right) \quad (1)$$

Where $\phi\left(O_i\right)$ is the HOG feature vector exacted from image **I** corresponding to ship patches. $\varphi\left(O_i, O_j\right)$ is the mutual relation vector between $O_i$ and $O_j$ that describes how much impact one object has on another one. $w_i$ and $w_{i,j}$ are the relational weights that balance all interactions of objects and the responses of each object. *G*= (*V, E*) represents the relational graph whose nodes *V* are objects and edges $e_{ij}$ indicate mutual related objects.

Actually, we choose the spring change between two objects as the mutual relation vectors. Then we can define a score of an associated configuration C= ($O_1$, $O_2,\dots O_{|V|}$) as follows:

$$s(\mathrm{C},\ \beta) = \sum_{i\in V} \mathbf{w}_i \cdot f\left(O_i\right) - \sum_{(i,j)\in E} w_{i,j} \cdot \left\|(\mathbf{x}_i - \mathbf{x}_j) - \boldsymbol{e}_{ij}\right\|^2 \tag{2}$$

Where the first term in the right side of (2) is the appearance score that measures the compatibility between the observed image patches and the object patches. Especially, $w_i$ are the HOG feature weights. The second term is the deformable score that calculates the spring change between the observed objects locations and the ground truth location of those objects. Here we consider the $w_{i,j}$ as a hyper-parameter and set $\forall$ *i, j* : $w_{i,j} = w$. Then all parameters can be represented by $\beta$=( $w_i$ , $w_{|V|}$ , $e_1,\dots,e_{|E|}$ ).

The tracking output is the configuration that maximizing (2) and this can be found efficiently using dynamic programing. Obviously, the smaller the spring change is, the closer our tracking output is to the true location.

**2.3. Learning**

In our tracker, learning aims to update parameters β with positive (i.e. object of interest selected from the first frame and the output configurations in previous frames) and negative configurations. We adopt the favorable structured SVM to update model parameters with a gradient decent method. Similar to [17], we define the structured SVM loss ι as follows:

$$\ell(\beta, C) = \max_{\hat{C}}(s(\hat{C},\ \beta) - s(\mathrm{C},\ \beta) + \Delta(\mathrm{C}, \hat{C})) \tag{3}$$

Where $C$ is the ground truth and $\hat{C}$ is a particular configuration. $\Delta(C,\hat{C})$ is the loss function that indicates the accuracy of the prediction. $\Delta(C,\hat{C})$=0 if $C = \hat{C}$, or $\Delta(C,\hat{C})$>0. Equation (3) attempts to learn a set of weight parameters β so that the configuration score of the ground truth part locations is greater than the score of any other possible configurations of part location by at least $\Delta(C,\hat{C})$, which encodes the penalty of predicting configurations $\hat{C}$. Obviously, the structured SVM loss ι is convex in parameter β, because it is the maximum of a set of affine functions.

The gradient of the structured SVM loss in (3) with respect to model parameters β is shown as (4).

$$\nabla_{\beta}\ell(\beta, C) = \nabla_{\beta}s(\mathrm{C}^*,\ \beta) - \nabla_{\beta}s(\mathrm{C},\ \beta) \tag{4}$$

Where the negative configuration $\mathrm{C}^*$ is given by:

$$C^* = \arg\max_{\hat{C}}(\bar{s}(\hat{C},\ \beta) + \Delta(\mathrm{C}, \hat{C})) \tag{5}$$

However, taking the trade-off between the appearance score and the deformation score into account, we redefine the negative appearance score as follows:

$$\bar{s}(\mathrm{C},\ \beta) = \sum_{i\in V} \mathbf{w}_i \cdot f\left(O_i\right) \tag{6}$$

As a result, the searching direction ***p*** for learning is defined as:

$$p = \nabla_{\beta}\bar{s}(\mathrm{C}^*,\ \beta) - \nabla_{\beta}s(\mathrm{C},\ \beta) \tag{7}$$

And a passing-aggressive algorithm [19] is used to perform the parameters update as shown in (8).

$$b \leftarrow b - \frac{i(b,C)}{\|p\|^2 + \frac{1}{2K}} p \qquad (8)$$

Where K ∈ (0 , + ∞) is a hyper-parameter that controls the "aggressiveness" of the parameter update.

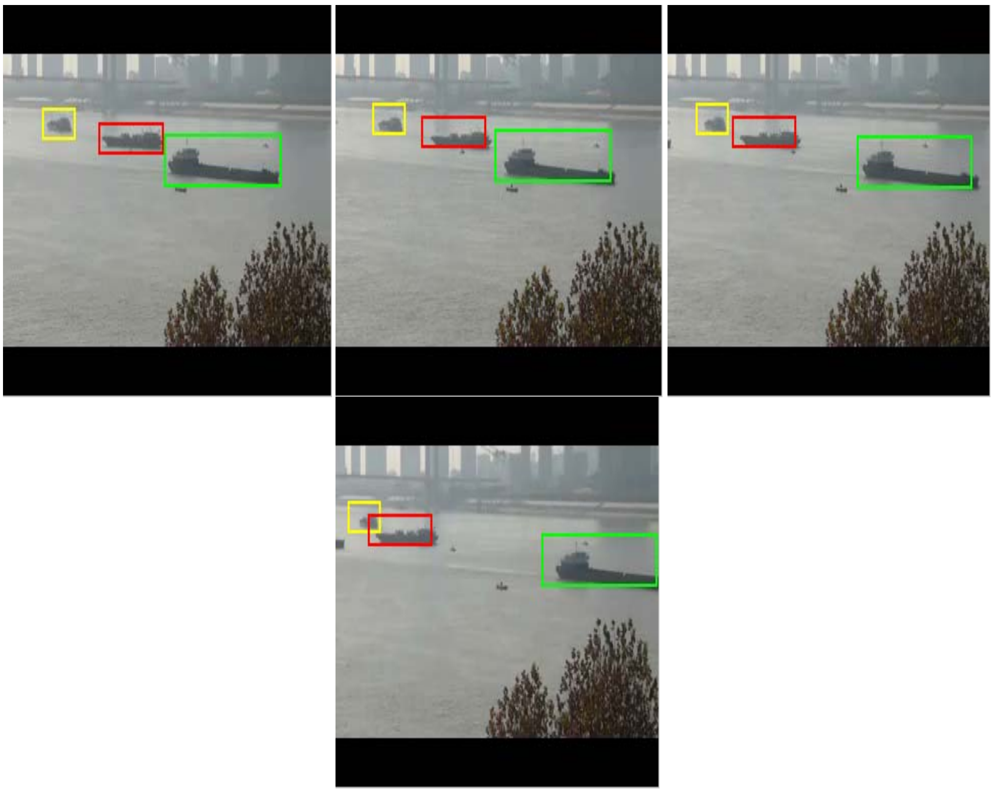
## 3. Experimental Results

To verify our multiple ship tracking method, we test on 3 CCTV videos in inland waterway. Figure 1 shows the experimental results.

In CCTV 1, three ships are selected with the colorful bounding boxes in the first frame. All the three ships are with different appearances and moving in the same directions. In the scene that is simple and without ship occlusion, in-plane rotation and so on, our tracker demonstrates excellent multiple ship tracking performance.

In addition, in CCTV 2, we show two ships tracking in the case of partial occlusion. The two target ships move in the opposite directions. Due to the application of spatial constraints, we can effectively identify the ship being partial occluded using the ships that are not being occluded.

CCTV 3 suffers from the in-plane rotation. Generally speaking, ships in inland waterway are always sailing in a relatively slow speed. As a result, even suffering from the in-plane rotation, the appearance of ship is almost with little change. On the other hand, with the model parameters learning and updating, our tracker can effectively adapt to the appearance change, thus to perform robust multiple ship tracking.



CCTV 1 (Frame 1/38/287/554in sequence)

CCTV 2 (1/60/127/307 in sequence)

CCTV 3 (Frame 1/98/150/282 in sequence)

Figure 1. Experimental Results with Multiple Ship Tracking

**4. Conclusion**

In this paper, we propose a robust multiple ship tracking method in inland waterway. Benefitting from HOG features and the linear SVM, we can effectively model the ships appearances. Moreover, the mutual relations based on the deformable part model make great contribution to utilize the spatial constraints to accomplish multiple ship tracing. According to the theoretical discussion and the experimental results, we can conclude that our multiple ship tracker is suitable for the inland waterway and demonstrates outstanding multiple ship tracking performance. In future work, we aim to explore the filter method that can optimize our tracker with the robust performance in serious occlusion.

## Acknowledgements

The authors would like to thank the editor and reviewers for their valuable comments and suggestions that lead to an improved manuscript. This work is supported by the National Science Foundation of China (NSFC 51279152).

## References

[1] Julia A Lasserre, Christopher M Bishop, Thomas P Minka. *Principled Hybrids of Generative and Discriminative Models.* Proc. of CVPR, New York. 2006; 87-94.
[2] A Balan, M Black. *An adaptive appearance model approach for model-based articulated object tracking.* Proc. IEEE Conf. Computer Vision and Pattern Recognition, New York. 2006: 758–765.
[3] MJ Black, AD Jepson. EigenTracking: Robust matching and tracking of articulated objects using a view-based representation. *Intl J. Computer Vision*. 1998; 26(1): 63–84.
[4] D Ross, J Lim, RS Lin, MH Yang. Incremental learning for robust visual tracking. *Intl J. Computer Vision*. 2008; 77(1): 125–141.
[5] A Adam, E Rivlin, I Shimshoni. Robust fragments-based tracking using the integral histogram. *CVPR*, New York. 2006: 798–805.
[6] SMS Nejhum, J Ho, MH Yang. Visual tracking with histograms and articulating blocks. In *CVP*R, Anchorage, AK. 2008; 1-8.
[7] Fei Teng, Qing Liu. Multi-scale ship tracking via random projections. In *SIVP*. 2014; 8(3).
[8] B Babenko, MH Yang, S Belongie. Robust object tracking with online multiple instance learning. *IEEE Trans. Pattern Analysis and Machine Intelligence*. 2011; 33(8): 1619–1632.
[9] H Grabner, C Leistner, H Bischof. Semi-supervised on-line boosting for robust tracking. In *ECCV*, Marseille-France. 2008: 234–247.
[10] Z Kalal, J Matas, K Mikolajczyk. P-n learning: Bootstrappingbinary classifiers by structural constraints. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA. 2010: 49–56.
[11] N Dalal, B Triggs. Histograms of oriented gradients forhuman detection. *Proc. IEEE Conf. Computer Vision and Pattern Recognit*ion. San Diego, CA. June 2005; 886–893.
[12] Qiang Zhu, Yeh MC, Avidan S. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York. 2006; 2; 1491-1498.
[13] P Felzenszwalb, D McAllester. "A discriminatively trained, multiscale, deformable part model," In *CVPR*, Anchorage, AK. June 2008. pp. 1-8.
[14] S Branson, P Perona, S Belongie, "Strong supervision from weak annotation: Interactive training of deformable part models," In *Proc. IEEE Int. Conf. Computer Vision*, Barcelona. June 2011, pp. 1832–1839.
[15] Lu Zhang, Laurens van der Maaten. "Structure preserving object tracking. *IEEE Conference on Computer Vision and Pattern Recognition*. Portland, Oregon. 2013. pp. 1838-1845.
[16] Genquan Duan, Haizhou Ai, Song Cao, and Shihong Lao. "Group Tracking: "Exploring Mutual Relations for Multiple Object Tracking. In *ECCV*. Florence. October 2012; 129-143.
[17] A Yilmaz, O Javed, M Shah. Object tracking: A survey," *ACM Computing Surveys*. 2006; 38(4):
[18] I Tsochantaridis, T Joachims, T Hofmann, Y Altun. Large Margin Methods for Structured and Interdependent Output Variables. *JMLR*. 2005; 6: 1453-1484.
[19] K Crammer, O Dekel, J Keshet, S Shalev-Shwartz, Y Singer. Online passive-aggressive algorithms. *Journal of Machine Learn-ing Research*. 2006; 7: 551–585.