

Robust Inland Waterway Ship Tracking via Orthogonal Particle Filter-based Predator

Fei Teng, Qing Liu and Lin Zhu

*School of Energy and Power Engineering, Wuhan University of Technology
changeerhao_love@126.com*

Abstract

Tracking-Learning-Detection (TLD), also known as Predator, has become one of the most popular state-of-the-art algorithms in the domain of visual tracking in recent years. It has demonstrated outstanding performance in the application of long-term tracking of a human face in unconstrained videos. In this paper, we address the problem of tracking a single ship in inland waterway CCTV videos given its location in the first frame and no other prior information. Firstly, we deeply analyze Predator in a perspective of control system and point out the search strategy in detection procedure is the most time-consuming part in Predator system. Secondly, in order to speed up the whole pipeline, we propose a novel motion model based on extended particle filter with orthogonal design. Due to the power of particle optimization and re-combination with orthogonal design, we can relate the motion of object of interest better and obtain the most likely candidates of object regions more effectively. Finally, both qualitative and quantitative evaluations on numerous challenging CCTV videos demonstrate that the proposed algorithm achieves favorable performance in terms of efficiency and accuracy.

Keywords: Control system, extended particle filter, Predator, orthogonal design

1. Introduction

Visual tracking has been extensively studied in computer vision due to its crucial importance in many practical applications such as HCI (human-computer interaction), automated surveillance, traffic monitoring and video indexing, to name a few [1]. Although much success has been demonstrated in building some domain specific trackers (*e.g.*, mice [2], faces [3], humans [4], *etc.*), real-time robust visual tracking has remained challenging due to various factors such as the limited amount of priori knowledge about the object of interest, the number and type of parameters being tracked (*e.g.*, contour, location, scale) and so on [5].

The recent years have witnessed great developments of CCTV system in the domain of inland waterway automated surveillance [6-7]. With the combined efforts of automatic identification system (AIS), global position system (GPS) and inland waterway electronic chart, CCTV acts as an eye for observers to tell whether a ship is retrograding, overtaking or parking without regulation. To keep long-term robust ship tracking is the core of inland waterway CCTV automated surveillance. Despite that numerous tracking algorithms [8-12] have been proposed in the literature over the past couple of decades, ship tracking remains challenging due to factors such as pose variation, illumination change, occlusion, and motion blur.

Recently, Kalal *et al.* [13] propose a novel tracking framework “Predator” that explicitly decomposes the long-term tracking task into three components: tracking, learning and detection. Then the system tracks, learns and detects a specific face in unconstrained videos.

The numerous reliable experimental results demonstrate that Predator achieves superior performance. However, we observe that it is problematic to apply Predator into the domain of inland waterway CCTV ship tracking. Because in the detection procedure, in order to track both the location and scale of the object of interest, each frame is scanned at all predefined scales by employing a sliding window strategy. As a result, the greedy search strategy brings comparatively high computational load and makes the overall system of low efficiency in terms of time and space complexity.

Motivated by the above-mentioned discussions, in this paper, we firstly analyze Predator in a perspective of control system and point out the search strategy in detection procedure is the most time-consuming part. Moreover, we further propose a novel motion model based on extended particle filter with orthogonal design to speed up the complete pipeline. Experimental results demonstrate that particle optimization and re-combination with orthogonal design significantly improves the efficiency in terms of both time and space complexity.

The remainder of this paper is organized as follows: Section 2 deeply analyzes the Predator in a perspective of control system and explains why we need to develop a more effective and efficient motion model. Section 3 describes extended particle filter with orthogonal design in details. Section 4 conducts numerous experiments on challenging CCTV video sequences to illustrate the effectiveness of the proposed method. Finally, Section 5 presents a brief summary.

2. Predator Closed-loop Control System

2.1. Predator Overview

Predator consists of a median flow tracker and a cascade of object detectors. The median flow tracker is based on a LK short-term tracker which recursively tracks the object of interest over time [14]. The cascade of detection procedure includes three components: a variance filter, an ensemble classifier and a nearest-neighbor classifier [15]. The detector localizes all appearances that have been observed so far. The detector runs in parallel with the tracker and corrects the tracker if necessary. The semi-supervised P-N online learning estimates the error of the tracker and detector and updates them to avoid these errors in the future [16]. The combined efforts of all components mentioned above make the system robust to pose variation, illumination change and occlusion. The basic flow of Predator is illustrated in Figure 1.

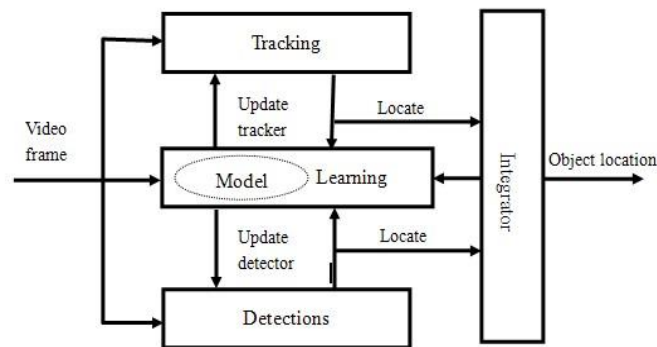


Figure 1. Block Diagram of Predator

2.2. Closed-loop Control System

According to the definition in control theory, traditional tracking methods (mean-shift tracker, particle filter, compressive tracking, *etc.*) belong to the open-loop controller, i.e., a type of controller that computes its output according to the current state and its model of the system. The input is the video frame (with the associated state) and the tracker acts as the controller. The tracker then makes response to the corresponding change of the input. However, it is commonly known that open-loop control system has its intrinsic limitations due to its design. For example, the response of the system may appear to be unstable because of some specific parameter variations caused by disturbance. That is to say, in an open-loop tracking system, slight inaccuracy of locations may lead to errors. These errors may accumulate over time, resulting in tracking drift or even tracking failures [17].

On the contrary, Predator is a typical closed-loop control system which extends open-loop control system with a mechanism of feedback. Closed-loop control system demonstrates advantage over open-loop controllers with reduced sensitivity to parameter variations. The Predator closed-loop control system is illustrated in Figure 2.

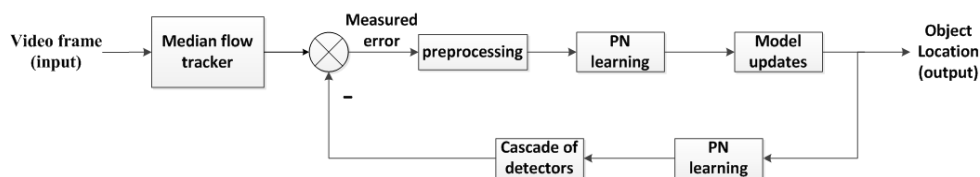


Figure 2. Predator Closed-loop Control System

Different from open-loop tracking system, Predator employs a cascade of object detectors to serve as its feedback basis. The feedback object detectors observe the performance of the tracker and correct the response of the tracker if it suffers any drift. Moreover, once the short-term median flow tracker completely loses track of the object of interest, the feedback object detectors will then reinitialize the tracking process. As a result, the feedback mechanism significantly stabilizes the tracking system, making it possible to perform long-term robust tracking in unconstrained videos.

According to classic control theory, an effective and efficient feedback factor is of crucial importance to the overall performance of a control system [18]. However, in the procedure of object detection in Predator, each frame is scanned at multiple predefined scales by using a sliding window. About a hundred thousands of likely candidates of object windows will be generated, depending on the size of the image and the size of the original object bounding box (in fact, for typical size of 320*240 input images, if the size of the original object bounding box is 100*80, this will generate about 70000 likely patches). The greedy search strategy in such a large search scope is problematic because of its comparatively high computational load. Hence the remedy power of cascade of detectors is decreased, which in turn has an uncertain effect on the short-term tracker or brings new error into the system. Based on the above discussions, we can safely draw the conclusion that a more effective and efficient motion model can improve the performance of the entire system. This motivates us to design a more effective and efficient motion model.

3. Extended Particle Filter Motion Model with Orthogonal Design

Particle filters [19-21] have been the most frequently used powerful tools in dealing with nonlinear and non-Gaussian state estimation of the stochastic system in the past decades. This

technique has been widely used in various fields such as robotics, signal processing, and visual tracking.

Particle filter is a type of Monte Carlo executive methods of recursive Bayesian state estimation. A typical state space representation of dynamic system can be written as follows:

$$x_k = f(x_{k-1}, v_{k-1}) \quad (1)$$

$$y_k = h(x_k, w_k) \quad (2)$$

where $x_k \in \mathbb{R}^{n_x}$, $y_k \in \mathbb{R}^{n_y}$ are the state vector and measurement vector at time k respectively. $v_{k-1} \in \mathbb{R}^{n_v}$, $w_k \in \mathbb{R}^{n_w}$ are the process noise and measurement noise with the known probability density function (PDF). n_x, n_y, n_v, n_w are the dimensions of vectors, and f, h are known possibly nonlinear functions. v_k and w_k are independent of the past and current states. Meanwhile, v_k is independent of the measurement noise w_k . The recursive Bayesian state estimation requires to construct the PDF of the state x_k , given all the available information $p(x_k | y_{1:k})$ where $y_{1:k} = \{y_i : i = 1, \dots, k\}$. This PDF can be obtained recursively in two steps: prediction and update. If the PDF $p(x_{k-1} | y_{1:k-1})$ at time $k-1$ has been obtained, the prior PDF $p(x_k | y_{1:k-1})$ of the state at time k can be obtained by using the system model [20]:

$$p(x_k | y_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | y_{1:k-1}) dx_{k-1} \quad (3)$$

Once the observation y_k at time k is available, the required posterior PDF $p(x_k | y_{1:k})$ can be obtained by using y_k to update the prior via Bayes rule [20]:

$$p(x_k | y_{1:k}) = \frac{p(y_k | x_k) p(x_k | y_{1:k-1})}{p(y_k | y_{1:k-1})} \quad (4)$$

where

$$p(y_k | y_{1:k-1}) = \int p(y_k | x_k) p(x_k | y_{1:k-1}) dx_k \quad (5)$$

The $p(y_k | x_k)$ is defined by the observation equation and the known state of w_k :

$$p(y_k | x_k) = \int \sigma(y_k - h(x_k, v_k)) p(w_k) dw_k \quad (6)$$

where $\sigma(\cdot)$ is the Dirac delta function and $p(w_k)$ is the PDF of w_k . The pseudocode description of generic particle filter is shown in Table 1.

Table 1. Pseudocode Description of Generic Particle Filter

For $i=1: N$

Draw $x_k^i \sim q(x_k | x_{k-1}^i, y_k)$

Evaluate the importance weights:

$$\omega_k^i \leftarrow \omega_{k-1}^i \frac{p(y_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, y_k)} \quad (7)$$

End For

Calculate total weight:

$$t = \text{sum} \{ \omega_{k,i=1}^{iN} \} \quad (8)$$

For $i=1: N$

 Normalize:

$$\omega_k^i \leftarrow \frac{\omega_k^i}{t} \quad (9)$$

End For

Calculate $N_{eff} : N_{eff} = \left(\sum_{i=1}^N W_k^{i2} \right)^{-1}$

If $N_{eff} < N_{thr}$

$$\left[\{ x_k^i, \omega_k^i, - \}_{i=1}^N \right] = \text{RESAMPLE} \left[\{ x_k^i, \omega_k^i \}_{i=1}^N \right] \quad (10)$$

End if

$$x_k \leftarrow \sum_{i=1}^N \omega_k^i \cdot x_k^i \quad (11)$$

From Table 1, we observe that some particles are eliminated and some particles are copied in RESAMPLE procedure, which results in sample impoverishment. Yiu-Wing Leung [22] proposes a genetic algorithm called orthogonal genetic algorithm with quantization for global numerical optimization with continuous variables. They utilize the orthogonal experiment design to enhance the genetic algorithm. Motivated by [22], we propose a novel motion model based on extended particle filter with orthogonal design. Our objective is to reduce the search scope of the cascade of detectors, thus we can significantly improve the efficiency of the process of detection in Predator. The detailed steps of our proposed algorithm are shown in Table 2.

Table 2. Extended Particle Filter with Orthogonal Experiment Design

Step1: Randomly sample N particles $\{x_k^i, \omega_k^i\}_{i=1}^N$ from the particle sets $x_k^i \sim q(x_k | x_{k-1}^i, y_k)$ and initialize parameters.

Step2: Evaluate and normalize the importance weights using formula (1), (7), (8) and (9).

Step3: Orthogonal experimental design. Suppose that there are Q groups of particles G_1, G_2, \dots, G_Q , where $G_i = (G_{i1}, G_{i2}, \dots, G_{id})$ and d represents the dimensionality of each group. Then each group is regarded as a level $l_i, i \in \{1, 2, \dots, Q\}$. In the next step, particles in each group are divided into F sub-groups, and each sub-group contains N particles equally. Then each sub-group is regarded as a factor $f_j = (G_{(j-1)N+1}, \dots, G_{jN})$, $j \in \{1, 2, \dots, F\}$. Once we obtain the level l_i and factor f_j , we construct our orthogonal array by the following procedure:

a) Construct the basic columns:

Compute the minimum J which satisfies $F \leq (Q^J - 1) / (Q - 1)$, and let $M = Q^J$;

FOR $k = 1$ TO J DO

$$j = \frac{Q^{k-1} - 1}{Q - 1} + 1 \quad (12)$$

FOR $i = 1$ TO Q^j DO

$$a_{i,j} = \left\lfloor \frac{i-1}{Q^{j-k}} \right\rfloor \bmod Q \quad (13)$$

END

END

b) Construct the non-basic columns:

FOR $k = 2$ TO J DO

$$j = \frac{Q^{k-1} - 1}{Q - 1} + 1 \quad (14)$$

FOR $s = 1$ TO $j - 1$ DO

FOR $t = 1$ TO $Q - 1$ DO

$$a_{j+(s-1)(Q-1)+t} = (a_s * t + a_j) \bmod Q \quad (15)$$

END

END

END

c) Increment $a_{i,j}$ by one for all $1 \leq i \leq M$ and $1 \leq j \leq N$.

d) Calculate the new state vector at time $k + 1$ and update the importance weights:

$$x_{k+1}^j = \frac{\sum(f * w)}{\sum(w)} \quad (16)$$

$$w_{k+1}^j = p(y_k | x_{k+1}^j) \quad (17)$$

e) Calculate the final state vector:

$$x = \sum_{j=1}^N x_{k+1}^j * w_{k+1}^j \quad (18)$$

Due to the power of particle optimization and re-combination with orthogonal design, we significantly reduce the scope of search region for the object of interest, which further facilitates the object searching procedure and improves the efficiency of whole pipeline.

4. Experiments

In order to demonstrate the superior performance of our proposed algorithm, we compare our tracker with three state-of-the-art trackers on 4 challenging CCTV videos. The reference trackers are Predator [13], incremental Learning for visual tracking (IVT tracker) [11] and multiple instance learning (MIL tracker) [5]. The 4 CCTV videos are challenging because the ship of interest suffers cluttered background, illumination variation, dramatic occlusion and extreme scale change. To make fair comparison, all trackers mentioned above are implemented in MATLAB. The qualitative results are illustrated in Figure 3 and the quantitative results are shown in Table 3 and Figure 4.

4.1. Qualitative Evaluations

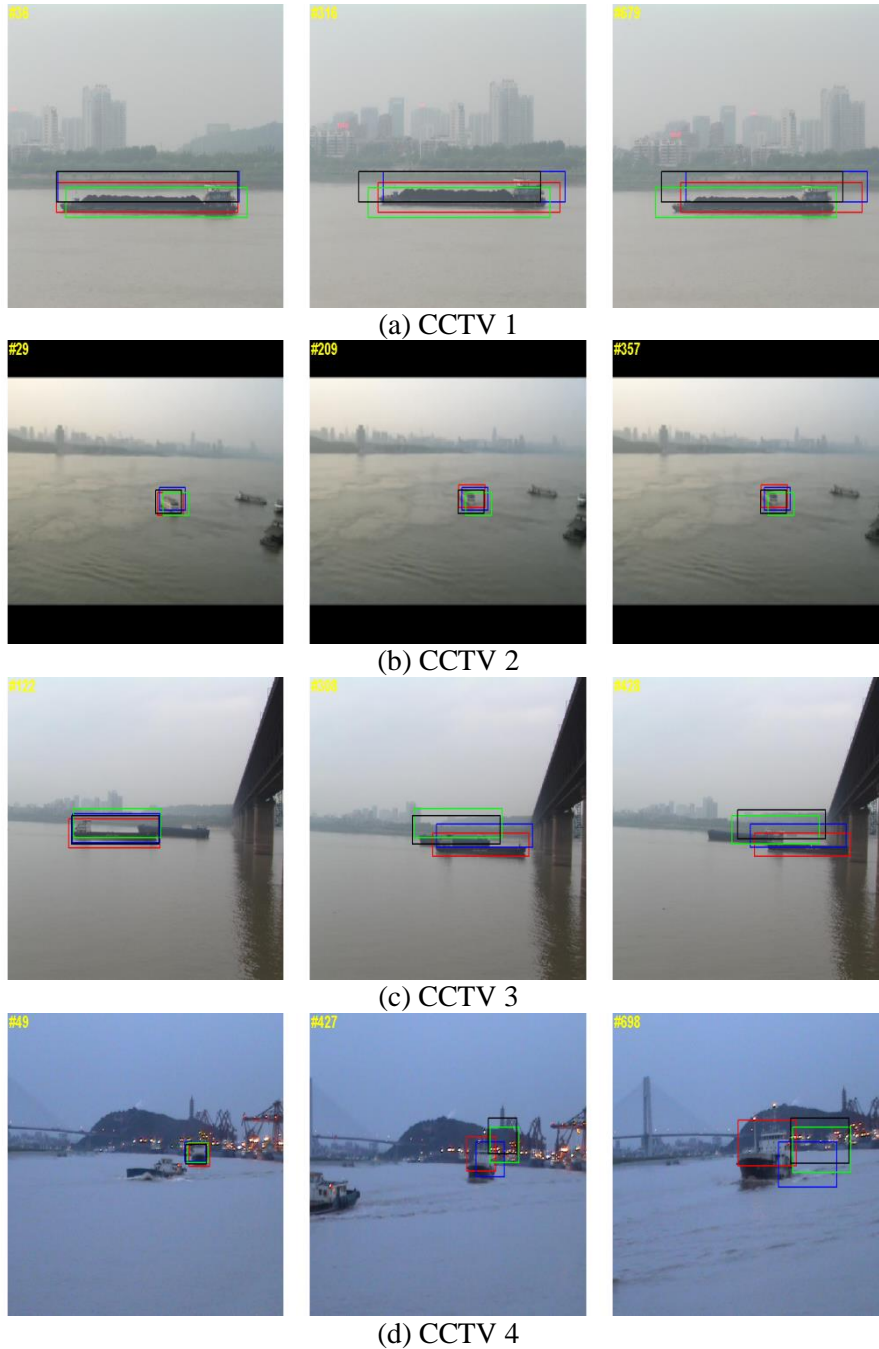


Figure 3. Screenshots of Sampled Tracking Results

(Red—our tracker Blue—Predator Green—IVT tracker Black—MIL tracker)

4.2. Qualitative Evaluations

We use two metrics to evaluate the proposed tracker and the reference trackers. One of the metrics is center location error (CLE) (in pixels). Simultaneously, since one of our main goals is to demonstrate the improved efficiency, we further employ average frames per second (FPS) as the other criteria to evaluate efficiency. The green fonts in Table 3 indicate the best performance and the blue fonts indicate the second-best. We can obviously see that our tracker achieves the best or second-best performance in all sequences, in terms of both CLE and FPS.

Table 3. Quantitative Evaluation Results of Ten Challenging CCTV Videos

Sequence	Ours	Predator	IVT	MIL
CCTV 1	7.3	7.3	15.1	14.2
CCTV 2	4.2	4.8	5.8	12.5
CCTV 3	8.9	8.5	14.2	24.2
CCTV 4	11.1	13.3	19.5	21.5
Average CLE	7.9	8.5	13.6	18.1
Average FPS	42	16	12.6	18.1

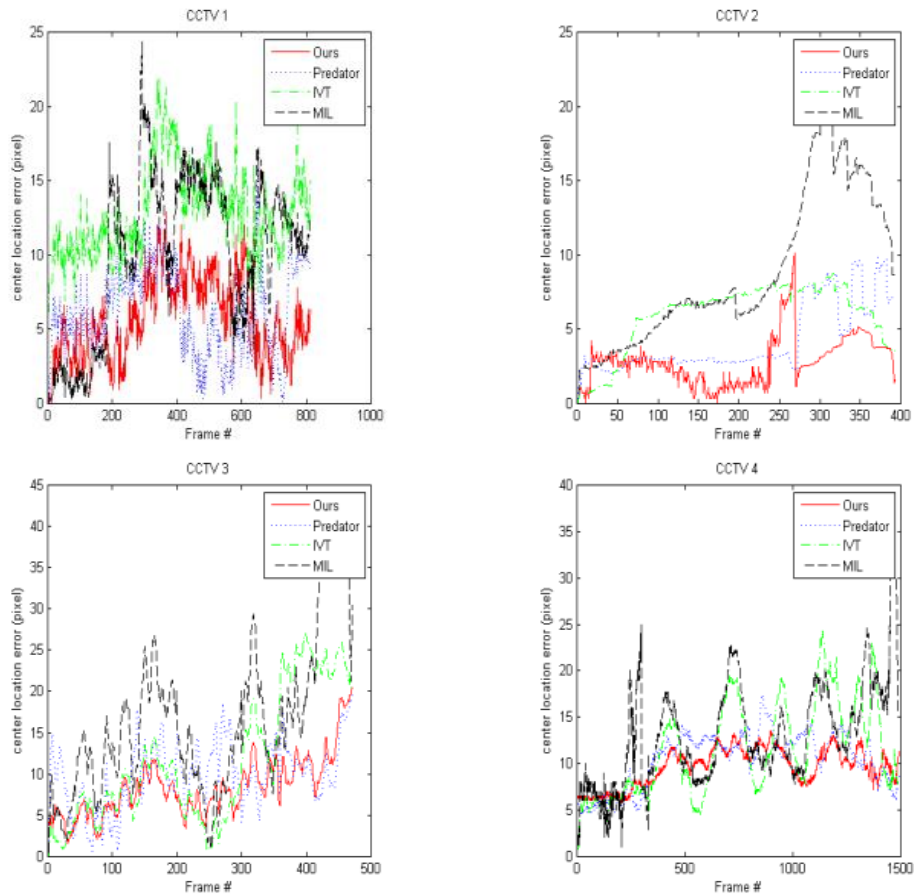


Figure 4. Center Location Error for Each Sequence

CCTV 1 is a very long sequence which suffers cluttered background, making it very challenging to track the ship of interest. Both IVT tracker and MIL tracker drift to the background because the texture of the background is similar to the ship (see frame 300). Our tracker and Predator perform well on this sequence because the appearance model based on P-N learning strategy makes the classifier better differentiate the target from the cluttered background. CCTV 2 comprises motion blur and in-plane rotation. MIL tracker performs not well in this sequence because the feature selection method used in the MIL tracker has potential overfitting problem and this method is easy to select the less discriminative features. This will lead to error accumulation over time. Thus MIL tracker will eventually drift, as shown in frame 310. The ship undergoes partial and even full occlusion in CCTV 3. Both IVT tracker and MIL tracker yield some unstable results when two ships occlude each other severely (see frame 325 and 400). Though the short-term median flow tracker in our system and Predator is also sensitive to appearance change caused by occlusion, these two trackers achieve much preferable results. This is partly due to the cascade of detectors, which helps to locate the object of interest and corrects the error of short-term median flow tracker. Meanwhile, the discriminative model with local features has demonstrated to handle occlusion better than simple feature extraction procedure such as MIL strategy. CCTV 4 is a long sequence which suffers extreme scale change. Both IVT tracker and MIL tracker eventually fail to track the ship of interest due to its intrinsic weakness with regard to appearance change caused by scale change. It can also be seen that our tracker performs comparatively better overall than Predator in terms of accuracy. This is due to the power of particle optimization and re-combination with orthogonal design, which leads to a more effective and efficient motion model. The motion model better relates the location of the object over time, which facilitates the search procedure for finding the most likely location in the current frame. Moreover, since our main purpose is to design a more effective and efficient motion model, we notice that Predator runs about 16 FPS on our 2.2 Ghz Intel i3 processor on a 320*240 CCTV video. However, our tracker runs about 42 FPS on average, meaning that our proposed tracker runs almost 3 times as fast as the Predator.

5. Conclusions

In this paper, we deeply analyze Predator in a perspective of control system and indicate the greedy search strategy in cascade object detection procedure is the most time-consuming part. In order to solve this problem, we propose a novel motion model based on extended particle filter with orthogonal design. We significantly improve the efficiency in terms of both time and space complexity by employing particle optimization and re-combination with orthogonal design. In the experiments, we demonstrate that our enhanced tracker outperforms some state-of-the-art trackers in terms of both accuracy and efficiency.

Acknowledgements

We would like to thank the editor and reviewers for their careful read and thoughtful comments on our previous manuscript. This work is supported by the National Science Foundation of China "Intelligent Analysis on Inland Waterway Video for Electronic Cruise System" (NSFC 51279152).

References

- [1] A. Yilmaz, O. Javed and M. Shah, "Object tracking: A survey", *ACM Comput Surv*, vol. 38, no. 4, (2006).
- [2] K. Branson and S. Belongie, "Tracking multiple mouse contours (without too many samples)", *International Conference on Computer Vision and Pattern Recognition*, (2005) June 20-25; San Diego, CA, United states, pp. 1039-1046.
- [3] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms", *International Conference on Computer Vision and Pattern Recognition*, (1998) June 23-25, Santa Barbara, CA, United states, pp. 232-237.
- [4] M. Isard and J. MacCormick, "BraMBLe: A Bayesian multiple-blob tracker", *International Conference on Computer Vision*, (2001) July 9-12; Vancouver, BC, United states, pp. 34-41.
- [5] B. Babenko, M.-H. Yang and S. Belongie, "Robust object tracking with online multiple instance learning", *IEEE Trans Pattern Anal Mach Intell*, vol. 33, no. 8, (2011), pp. 1619-1632.
- [6] F. Coudert, "Towards a new generation of CCTV networks: Erosion of data protection safeguards?", *Comput Law Secur Rev*, vol. 25, no. 2, (2009), pp. 145-154.
- [7] N. Dadashi, A. W. Stedmon and T.P. Pridmore, "Semi-automated CCTV surveillance: The effects of system confidence, system accuracy and task complexity on operator vigilance, reliance and workload", *Appl. Ergon*, vol.44, no. 5, (2013), pp. 730-738.
- [8] F. Teng, Q. Liu, X.Y. Gao and L. Zhu, "Real-time ship tracking via enhanced MIL tracker", *Fourth International Conference on Emerging Trends in Engineering and Technology*, (2013), October 25-27; GIMT, Kurukshetra, India, pp. 399-404.
- [9] F. Teng and Q. Liu, "Multi-scale ship tracking via random projections", *Signal Image Video Process*, vol. 8, no. 3, (2014), pp. 1-8.
- [10] Chaudhuri, S. Sinha, and A. Konar, "Vision Based Target-Tracking Realized with Mobile Robots using Extended Kalman Filter", *Engineering Letters*, vol. 14, no.1, (2007).
- [11] J. Lim, D. A. Ross, R.-S. Lin and M.-H. Yang, "Incremental Learning for Visual Tracking", *Nips*, vol. 17, (2004), pp. 793-800.
- [12] A. Blake and M. Isard, "Active contours: the application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion", *Springer-Verlag New York, Inc.*, (1998).
- [13] Z. Kalal, K. Mikolajczyk and J. Matas, "Face-TLD: Tracking-learning-detection applied to faces", *International Conference on Image Processing*, (2010) September 26-29, Hong Kong, China, pp. 3789-3792.
- [14] Z. Kalal, K. Mikolajczyk and J. Matas, "Forward-backward error: Automatic detection of tracking failures", *International Conference on Pattern Recognition*, (2010), August 23-26; Istanbul, Turkey, pp. 2756-2759.
- [15] Z. Kalal, K. Mikolajczyk and J. Matas, "Tracking-learning-detection", *IEEE Trans Pattern Anal Mach Intell*, vol. 34, no. 7, (2012), pp. 1409-1422.
- [16] Z. Kalal, J. Matas and K. Mikolajczyk, "Online learning of robust object detectors during unstable tracking", *International Conference on Computer Vision*, (2009), September 27-October 4; Kyoto, Japan, pp. 1417-1424.
- [17] Z. Kalal, J. Matas and K. Mikolajczyk, "P-N learning: Bootstrapping binary classifiers by structural constraints", *International Conference on Computer Vision and Pattern Recognition*, (2010) June 13-18; San Francisco, CA, United states, pp. 49-56.
- [18] H. G. Zaini, "A novel approach for designing a feedback controller of linear time invariant networked control systems with delayed-transmission time", *International Conference on Information Networking*, (2013), January 27-30; Bangkok, Thailand.
- [19] H. Kandil, H. Kandil, and A. Atwan, "A Comparative Study between SIFT-Particle and SURF-Particle Video Tracking Algorithms", *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 5, no. 3, (2012), pp. 111-122.
- [20] Y. Yang, X. Wang and C. Hu, "An overview and analysis of particle filter's key technology", *Information Optoelectronics, Nanofabrication and Testing*, (2012) November 1-2; Wuhan, China.
- [21] T. Germa, F. Lerasle and Thierry Simon, "Video-based face recognition and tracking from a robot companion", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 3, (2009), pp. 591-616.
- [22] Y. W. Leung and W. Yuping, "An orthogonal genetic algorithm with quantization for global numerical optimization", *IEEE Trans Evol Comput*, vol. 5, no. 1, (2001), pp. 41-53.

Authors



Fei Teng, he received the BS degree in School of Automation, Taiyuan University of Science and Technology in 2011, the MS degree in School of Automation, Wuhan University of Technology in 2013. He is now a PhD candidate in School of Energy and Power Engineering, Wuhan University of Technology. His research interests include machine learning and computer vision, with a focus on visual tracking techniques, such as compressive sensing, sparse representation, feature extraction, and feature selection.



Qing Liu, she received the BS degree, MS degree and PhD degree in School of Automation, Wuhan University of Technology in 1995, 1998 and 2002 respectively. Her current research interests include machine learning and computer vision.



Lin Zhu, she received the BS degree in School of Automation, Wuhan University of Science and Technology in 2012. She is now a graduate student in School of Automation, Wuhan University of Technology. Her research interests include image processing and pattern recognition.

