

DESIGN AND IMPLEMENTATION OF THE VISUAL DETECTION SYSTEM FOR AMPHIBIOUS ROBOTS

Yanlin He,^{*,**} Xu Zhang,^{*,**} Lianqing Zhu,^{*,**} Guangkai Sun,^{*,**} and Junfei Qiao^{***}

Abstract

Aiming at application requirements of our spherical amphibious robot, which has one type of propulsion system to be used both on land and in water, an improved tracking learning detection (TLD) algorithm is designed and implemented, so as to improve the robustness and environmental adaptability of its visual detection system. First, binary normalized gradient method is implemented to reduce the number of image candidate samples. Second, the compressive sensing algorithm is introduced to speed up the image feature calculation. Then, some common videos are applied to test the performance of the improved TLD algorithm, and the results show that the average frame rate of the improved TLD algorithm is increased. Finally, to evaluate the feasibility and effectiveness of the improved TLD algorithm in the visual system of our amphibious spherical robot, a series of experiments are carried out with illumination changes and in underwater environment. Comparison experimental results demonstrated that the improved TLD algorithm has better effectiveness, robustness and real-time performance than the traditional TLD algorithm.

Key Words

Amphibious spherical robot, visual detection system, tracking learning detection, real time

1. Introduction

In an effort to build robots which can autonomously implement some tasks in ever-changing environments, it is critical for a robot to have effectiveness of visual detection system. As a kind of feasible sensor with advantages of low power consumption, strong adaptability and low cost

[1], digital cameras have been widely applied in robotics to realize target detection and tracking. In a machine vision-based robot, the visual detection system plays a very important role in realizing diverse robotic functions, such as target detection and tracking [2], [3], autonomous navigation [4], [5], mutual positioning [6], path planning [7]–[9], visual servoing [10], robot–human interaction [11], *etc.*

In general, some existing detection algorithms can be categorized as template matching-based and feature extraction-based algorithms. Template matching-based algorithm has simple principle and poor manoeuvrability; considering the variability of target, the algorithm is difficult to determine the threshold, and its selection standard is not unified [12]. Examples of this type of algorithm, which are mainly driven by innovations in appearance models, include MeanShift [13], Kalman filter-based algorithms [14] and optical flow-based algorithms [15]. Feature extraction-based algorithms are applying the method of feature function instead of grey value differential accumulation, and it is usually built upon pattern recognition algorithms, such as support vector machine (SVM) [16], Bayes classifier [17], *K*-means [18], *etc.* In 2008, Bibby *et al.* [19] proposed a tracking algorithm based on pixel posteriori probability pascal-web-page (PWP). In 2010, Hare *et al.* [20] proposed a struck algorithm based on multi features, and this algorithm is related to the SVM algorithm and HAAR feature extraction. In 2010, Kalal *et al.* [21] proposed the tracking learning detection (TLD) algorithm, which provided a novel framework for solving the problem of long time, multi-scale change, deformation and other problems in the process of target detection. In 2012, Zhang *et al.* [22] proposed the compressive tracking (CT) algorithm, which provided a novel framework for developing real-time tracking algorithms for mobile robots. In 2015, Yahyaa *et al.* [23] proposed a visual detection and tracking system to guide an autonomous underwater vehicle (AUV) toward its docking station. A tracker using colour thresholding and morphological operations was designed to track artificial objects, but the robotic vision system could only recognize and track specific light sources. In 2016, Zhang *et al.* [24] presented a multi-target tracking system using multiple underwater cameras, and the cameras used were static and the visual system had

* Beijing Engineering Research Center of Optoelectronic Information and Instruments, Beijing information Science and Technology University, Beijing, China; e-mail: heyanlin@bit.edu.cn, zhangxuBISTU@outlook.com, zlq_2018@sina.com

** Bionic and Intelligent Aerospace Vehicles Lab, Beijing Information Science and Technology University, Beijing, China; e-mail: guangkaisun@buaa.edu.cn

*** Beijing University of Technology, Beijing, China; e-mail: junfeiq@bjut.edu.cn

Corresponding author: Lianqing Zhu

Recommended by Prof. Howard Li
(DOI: 10.2316/J.2019.206-0052)

poor real-time performance, which made it unsuitable for robotic applications. In 2017, Chuang *et al.* [25] proposed a novel tracking algorithm on the basis of the deformable multiple kernels to track live fish in an open aquatic environment; the algorithm outperformed the recent tracking-by-detection algorithms for tracking one or multiple live fish in challenging underwater videos, but its low frame rate limited its applications in mobile robotic platforms.

By focusing on the tasks of amphibious environment observations and intelligent surveillance in littoral regions, a moving target detection and tracking system was proposed for our amphibious spherical robot in this paper. Considering that the visual detection system of our amphibious spherical robot is difficult to detect some targets for changes in their appearances that are caused by pose or state changes, random motion and occlusion, and the environment can be easily disturbed by variations with illumination, camera vibration and outside interference, which may cause drift or target loss [26]. What is more, it is difficult to guarantee the detection accuracy, real-time performance and adaptability of the visual detection system of a robot in changeable underwater environment [27], [28].

To realize flexible and economical exploration, our team have designed an amphibious spherical robot [29]–[35], and a three-dimensional printing technology-based amphibious spherical robot with transformable composite propulsion mechanisms was created [36]–[39]. For the sake of realizing the visual detection system of our robot, a camera module was added to the control system of a robot. Given the special working environments and application, an improved TLD algorithm was proposed, binary normalized gradient (BING) method was utilized to reduce the sample numbers, and compressive projection matrix was introduced to reduce the complexity of feature computing down. To evaluate the effectiveness of the improved algorithm, a series of experiments were carried out in various scenarios with illumination changes and in underwater environment, which may cause target partly

occlusion or losses. Experimental results have demonstrated the effectiveness, robustness and real-time performance of the improved algorithm. Furthermore, the accuracy and frame rate of optimized algorithm are largely increased.

The rest of this paper is organized as follows. An overview on our amphibious spherical robot and its vision application requirements is introduced in Section 2. The structure of the traditional TLD algorithm and implementation of the improved algorithm is presented in Section 3. Experimental results under various scenarios are reported in Section 4. Section 5 presents the conclusion of this paper and future research directions.

2. Previous Works of Robots and Traditional Algorithm

2.1 The Structure of Amphibious Spherical Robots

Figure 1 shows the amphibious spherical robot, which consisted of an enclosed hemisphere hull (diameter: 250 mm) and two openable quarter-sphere shells (diameter: 256 mm). Electronic devices and batteries were installed inside the hemispherical hull, which was waterproof and provided protection from collisions. Four legs, each of which was equipped with two servo motors and a water-jet motor, were installed symmetrically in the lower hemisphere of the robot. Driven by the two servo motors, the hip joint and knee joint of a leg were able to rotate around a vertical axis and a horizontal axis, respectively. The water-jet motor was fixed in the knee joint and could generate a vectored thrust in a specific direction in water. We attempted to integrate the design and ensure seamless connections between every part to increase the stability of underwater spherical robot, and the hemispherical hull is adequate to allow the robot to operate up to 10 m underwater. In underwater mode, the openable shells closed to form a ball shape, and the robot was propelled by vectored thrusts from four water jets. In land mode, the openable shells opened and the robot walked using the four legs [26]–[31].

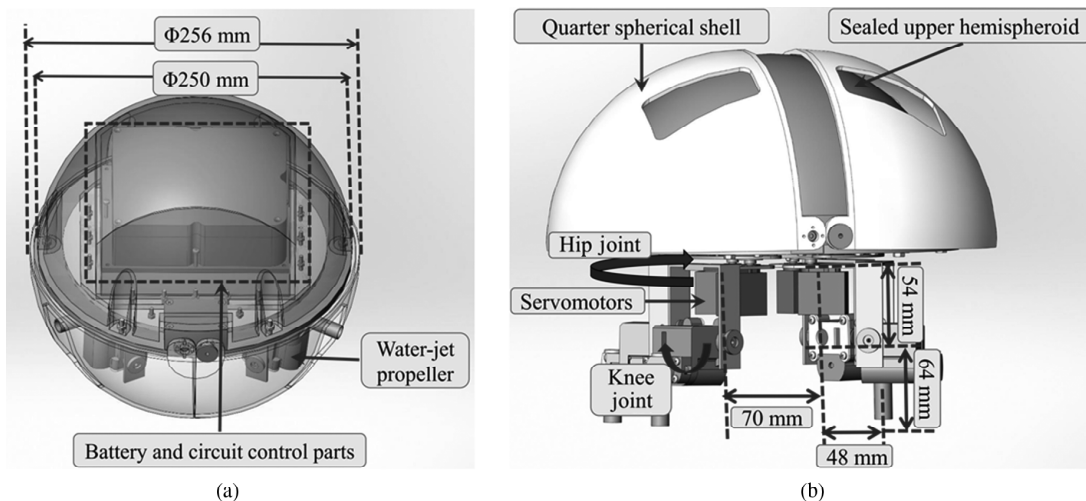


Figure 1. The physical details of amphibious spherical robot: (a) spherical shell closing mode and (b) spherical shell opening mode.

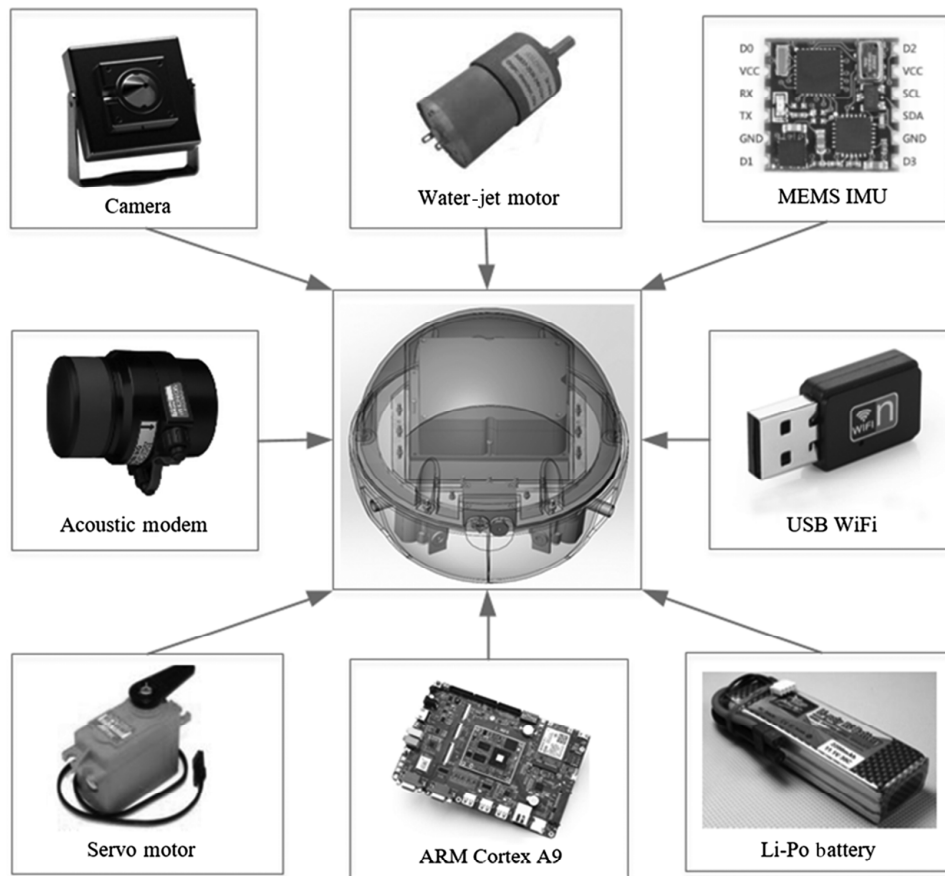


Figure 2. Major components of the amphibious spherical robot.

Table 1
The Electrical Specifications of Robot

Electrical Equipment	Specific Type	Parameter Specifications
Main control unit	ARM Cortex A9	Processor speed is 20 fps
Coprocessor	AVR Atmeg2560	
Power supply	Li-Po batteries	Maximum voltage supply is 14.8 V
Camera	KC352	Maximum resolution is 640 × 480
USB WiFi	FW150US	Transmission rate is 150 Mbps
Servo motor	HS-5086WP	The maximum torch is 3.6 kg cm

Restricted by the narrow load space and limited power resources of the small-scale amphibious spherical robot, the robotic electronic system was fabricated around an advanced RISC machine (ARM), as shown in Fig. 2. The robot was powered by a set of Li Po batteries, with a total capacity of 12,000 mA h. Sensors including a micro electromechanical system (MENS) inertial sensor were used to achieve attitude angle of the robot, an KC352 camera model was adopted in this paper, and a complementary metal-oxide-semiconductor (CMOS) sensor was employed to ensure the quality of image data, which has a maximum resolution of 640 × 480. FW150US has been selected as wireless network card, and a micron modem was selected as underwater sonar module. Some electrical specifications and physical properties of the spherical robot are presented in Tables 1 and 2.

Table 2
The Physical Properties of Spherical Robot

Components	Size Specifications
Upper shell diameter	250 mm
Lower shell diameter	256 mm
Weight of robot	2.26 kg
Actuating system height	118 mm (standing state)
Servomotor	31 × 15.2 × 31 mm
Water-jet propeller	15 × 21 × 42 mm

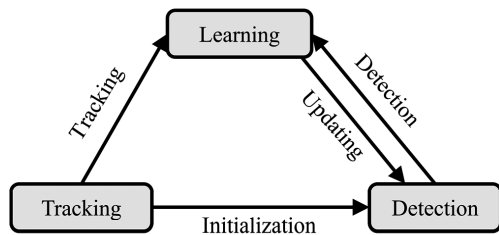


Figure 3. The block diagram of the TLD framework.

2.2 Traditional Tracking Learning Detection Algorithm

In 2010, Kalal *et al.* [21] proposed the TLD algorithm, which provided a novel framework for solving the problem of long time, multi-scale change, deformation and other problems in the process of target detection, and received wide recognition in the field of research scholars. As an effective and efficient discriminative algorithm, traditional TLD algorithm consists of three stages: tracking, learning and detection. As shown in Fig. 3, in the stage of tracking and detection, the location of target is finished and then output to the learning stage, when a new target pattern is identified, the detection stage is updated, and the probability of detection is further increased. When the location of target obtained by the tracking module is not accurate, the location of target obtained by the detection module and the status of the tracking module are updated in time, so as to better accomplish the target detection tasks.

Accordingly, the detection module is a paramount one of TLD algorithm, and its accuracy will largely effect the algorithm and the realization of detection module as shown in Fig. 4. The detection of target is accomplished by designing three stages of classifiers: the first stage is the variance judgement of image fragments, some fragments

with low variance is considered as background. The second stage is the ensemble classifier, and the classification results were obtained by using multiple similar weighted Bayes classifier. The third stage is the k nearest neighbour classifier, because of the existence of target space, the change of the target state can be recorded, and the validity of its long-time detection could be guaranteed.

The TLD algorithm succeeds in solving the problem of long time, multi-scale changes and deformation in the process of target detection. However, the accuracy and frame rate are relatively low when the environment is complex and changeable, and its efficiency will restrict its further promotion. To compensate for the deficiencies of TLD algorithm, Zhou *et al.* [40] proposed an improved TLD algorithm employing a Kalman filter in the detector of TLD and adding Markov model to increase the discretion for targets with similar appearance. To solve tracking drift problem of tracking algorithm, Yang *et al.* [41] proposed a gesture tracking method by combining TLD algorithm with Kalman filtering. However, due to the special working environment, our amphibious spherical robot has higher requirements for the robotic vision system in terms of robustness and efficiency, so it is still problematic for applications of these algorithms with our amphibious spherical robot.

3. Improved Tracking Learning Detection Algorithm

The efficiency of the TLD algorithm is associated with the number of candidate samples. Therefore, BING features were adopted to reduce the number of candidate samples [42], and the integral graph strategy was adopted to reduce the complexity of feature calculation in the view of compressive sensing theory. Figure 5 illustrates the details of the optimization and acceleration of the improved TLD algorithm.

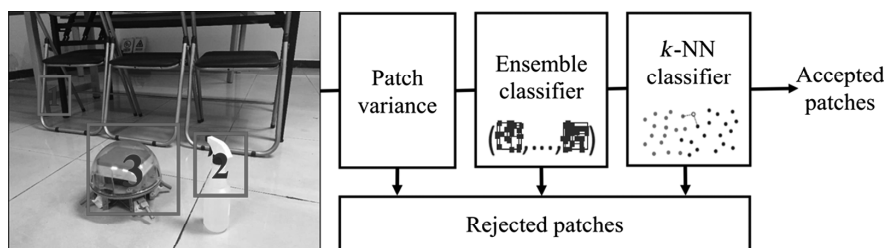


Figure 4. Block diagram of the target detector [18].

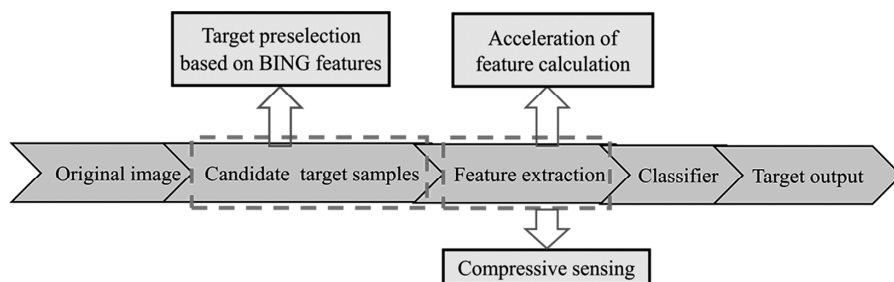


Figure 5. The optimization and acceleration of algorithm.

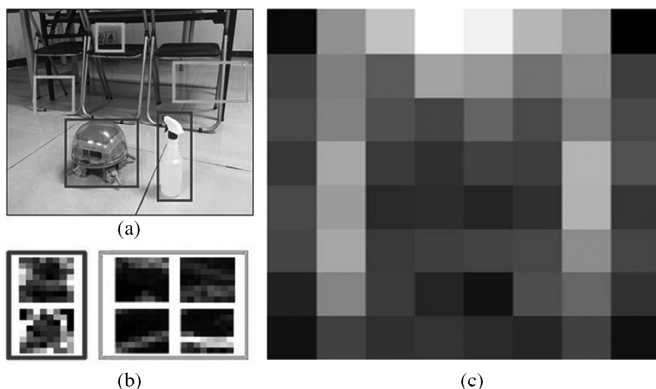


Figure 6. The feature of normalized gradient: (a) original image; (b) 8×8 NG feature; and (c) the model of leaning.

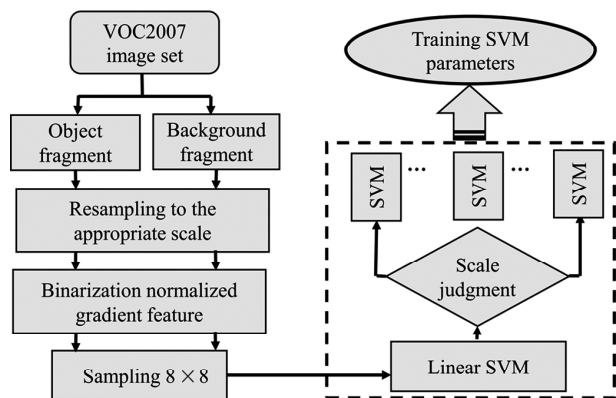


Figure 7. Target detection training based on BING feature.

3.1 BING Feature-based Target Preselection

The advantage of BING feature is that the image processing and “target” searching can be completed by simple operations, such as adding, subtracting and shifting. The most evident characteristics of “target” in the image are that it has a closed boundary, and the background information is disorganized, and the application of BING feature is to find target in the image. In the gradient space map, the general target has obvious edge gradient while it is normalized to the same scale, and it has excellent contour commonality and could make the target forming a closed contour. In the fixed window size, the gradient patterns of target and background information are different [43], [44], and SVM classifier can be invoked to distinguish the target.

Figure 6(a) manifests the original image including the target and background information, the target and background information is marked with black frame. The image is normalized by gradient operator, the fragments of target and background are extracted respectively, before being adjusted to the size of fixed 8×8 , so that the differences between the target and background can be found. As shown in Fig. 6(b), the gradient distribution of the target is messy, and the gradient distribution of the background is single, then the training of SVM classifier can

be completed by preparing enough sample size. Figure 6(c) illustrates the target model established after studying some certain amount of samples.

There are two steps to implement the original localization of targets by using a BING feature: the first step is to extract the image fragment of target and non-target from the data sets, the size of target is defined in this paper, and the length and width are setting as a discrete combination of 8, 16, 24, ..., 256. For extracting the image fragment, resampling its size to the nearest neighbour’s scale combination, the Sobel operator is used to obtain the normalized gradient feature of image fragment. The second step is to resample the gradient image to the size of 8×8 , and the feature calculations of samples are completed, with category labels matching; it can be transported to the SVM classifier to complete training tests. The overall detection training process is based on the BING feature as shown in Fig. 7.

Here we selected the Pascal VOC 2007 as data set; it provides the standard evaluation system for detection algorithm and learning performance of the standard image annotation. This data set contains 20 categories of images, including human, animal, vehicle, *etc.*; these objects are most common in our daily life, the total number of the data sets is 9963, and the location and category information are provided in images for the purpose of research.

Following that, the trained classifier is employed to detect the actual target. With regard to the image for detection, the image was scanned in turns by using the method of sliding window, and its scales are traversed from 8×8 to 256×256 , with a step size of 8. After that, gradient calculation is normalized for the image fragment of each scale obtained by traversal, the gradient value ranges from 0 to 256, and this value is saved as a variable of single byte, its high position served as a result of binarization. Then the scale of 8×8 is resampled and sent them to the SVM classifier for weighted judgement, the specific flow chart of real-time target detection based on BING feature as Fig. 8 shows.

To validate the practicability of the proposed BING feature-based target preselection, 2,501 training samples are used for testing, the first stage of SVM parameters training takes 14.457 s, and the second stage takes 52.271 s. The average consuming time of target detection for each image is 10^{-6} s due to there are 3,000 samples for testing.

3.2 Compressive Sensing-based Acceleration

In addition to the use of the BING feature, some method for improving the efficiency and optimizing the algorithm is also necessary, and this section depicts the compressive sensing algorithm to speed up the feature calculation. The compressive sensing algorithm succeeded in real-time performance, and it can provide a new probability for image sampling, namely if the sparsity of the signal is good enough, the signal could be sampled through the frequency which is lower than the sampling frequency of Shannon, and the signal could be reconstructed with high probability and high precision.

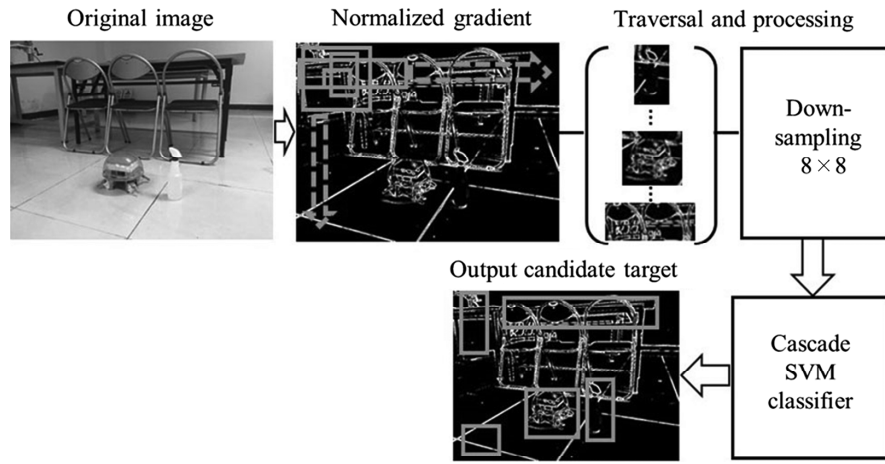


Figure 8. Real-time target detection based on BING feature.

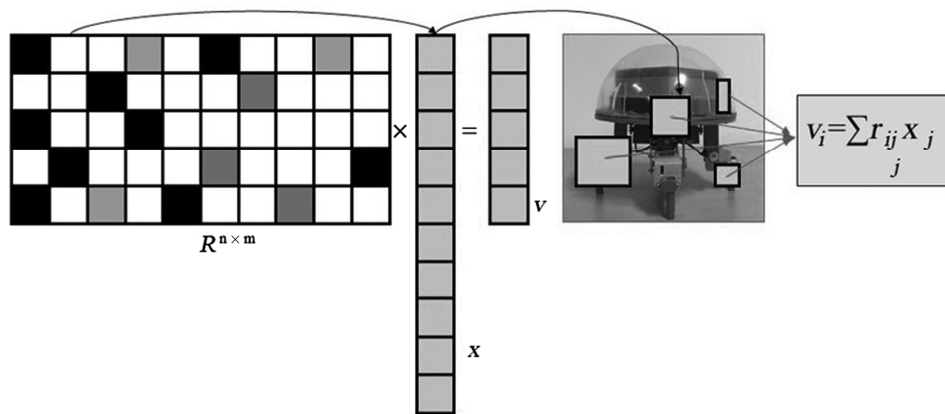


Figure 9. Image feature calculation of compressive sensing.

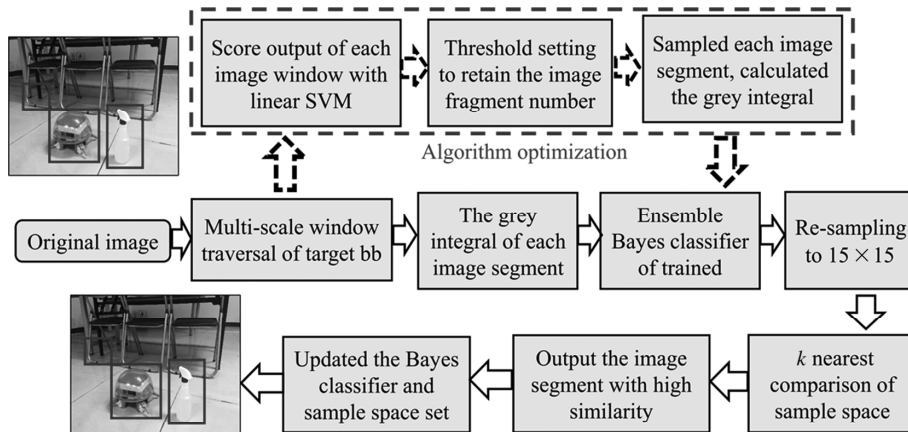


Figure 10. The process of improved TLD algorithm.

The traditional sampling method is time consuming, and according to the theory of compressive sensing, the recovery of the signal with good sparsity could be accomplished by low-frequency sampling, so the number of reconstruction signal is reduced, and the time of reconstruction signal is also greatly reduced. In summary, compressive sensing is the projection of original high-dimensional signal

in the low dimension; thus, the vectors can be projected to the dimension of vector through the selected dimension projection matrix. Figure 9 illustrates the process of feature calculation of image [45], [46].

The realization of improved TLD target detection process with the fusion of BING feature selection and compressive sensing is shown in Fig. 10; as shown in the figure,



Figure 11. Display of target detection demonstration system of MFC.

the initial selection of the target and the feature calculation process are indicated by dotted arrow. First, the normalized gradient of sample BB was calculated by each sliding window and then binarizing them simultaneously. Next, their size was scaled to 8×8 and input them into the trained linear SVM classifier, the score of probability of the existed target could be get-through classification, the number of the candidate target reduced, and the velocity of traversal could be achieved at 300 fps. Above all, the improved algorithm could effectively reduce the feature calculation for unqualified candidate target, the

amount of feature calculation, and improve the efficiency and performance of the target detection.

4. Experimental Evaluation on Land and Underwater

To evaluate the effectiveness of the improved algorithm, this section fabricated a demonstration interface system, as shown in Fig. 11; this demo software can be used to complete TLD algorithm and the validation of the improved algorithm. The demo system can be used to read a single frame picture, continuous picture stream, video stream and camera, and to activate buttons such as open, play, save and others, according to different options. The selection of target could be completed through the rectangular frame, and the feature of which could be represented by x , y , width, height; x and y are the horizontal and vertical coordinates of the rectangular frame; height and width respectively are the ones corresponding to the rectangular frame. By selecting the corresponding algorithm, the button “play” could reveal the detection operation. Real-time output target rectangle frame will be displayed in the following text box, while the “save” button can save the current frame to start the output video stream, including the size of picture and initial rectangle frame, also with its detection trajectory.

To guarantee the objectivity, authenticity and accuracy of test results, herein 20 common videos were applied to test that the environmental conditions and target models have deteriorated (partly occlusion, illumination variance or pose change), which are used to verify the performance of the algorithm. The rectangle data of the



Figure 12. Display of the test scenario like partly occlusion, illumination variance or pose changes.

Table 3
Accuracy Rate and fps of TLD and Improved Algorithm

Test set	Accuracy rate (%)		Frames per second (fps)	
	TLD	Proposed algorithm	TLD	Proposed algorithm
Animal	87	96	12	25
Biker	2	9	10	21
Bolt	0	8	10	30
Cliff bar	63	44	13	26
Chasing	76	85	15	32
Coupon book	31	98	13	32
David indoor	98	33	20	35
Dark car	67	18	15	22
Football	59	62	23	24
Goat	48	59	25	27
Occluded face	87	97	22	25
Panda	34	43	12	28
Pedestrian	0	1	23	33
Skating	37	84	17	25
Shaking 1	15	48	12	27
Shaking 2	56	53	19	29
Sylvester	89	80	12	25
Tiger 1	40	87	20	25
Tiger 2	54	62	18	22
Twinning	91	95	13	20
The average accuracy (fps)	52	58	16	27

target image detection output was recorded as T , which was compared with the real target image detection data (G) of the same rectangle in the initial frame, the feature of the rectangle frame, including its upper left corner coordinates x , y , the width (w) and high (h) of the rectangle frame. The similarity score of the two rectangular frames can be expressed as

$$\text{score} = \frac{\text{area}(\text{ROI}_T \cap \text{ROI}_G)}{\text{area}(\text{ROI}_T \cup \text{ROI}_G)} \quad (1)$$

where ROI indicates that the image was output as a rectangular frame, the $\text{ROI}_T \cap \text{ROI}_G$ refer to the intersecting total number of pixels of rectangle frames T and G , and the $\text{ROI}_T \cup \text{ROI}_G$ refer to the union total number of pixels of rectangle frames T and G . This formula can be used to evaluate the similarity of rectangle frames T and G , and when the score is greater than 75%, the target image detection results are considered acceptable. Figure 12 manifests some of the test scenario like partly occlusion, illumination variance, and the detection target in the scenario have been

mark out with a black rectangle frame, which contains some characteristics of common environment and can be used to evaluate the performance of the algorithm. The frames per second (fps) and the average accuracy, which represent the efficiency of detection algorithm, are some results shown in Table 3.

4.1 Experiments with Normal Illumination

To verify the practicability of the improved algorithm, a visual detection system of our robot was developed, and another amphibious spherical robot was also selected as the detection target. When the robot is in motion state with normal illumination, the detection results of TLD algorithm and the improved algorithm were obtained, and these results also demonstrate the effectiveness of the improved algorithm. In Fig. 13, the frame rates of TLD algorithm with different frequencies are 7.8 fps and 4.7 fps, respectively, and Fig. 14 shows the frame rates of the improved algorithm are 18.2 fps and 16.8 fps, respectively.



Figure 13. Detection results of TLD algorithm: (a) detection image in #28 and (b) detection image in #138.



Figure 14. Detection results of improved algorithm: (a) detection image in #28 and (b) detection image in #138.



Figure 15. Detection results of TLD algorithm: (a) detection image in #25 and (b) detection image in #325.



Figure 16. Detection results of improved algorithm: (a) detection image in #25 and (b) detection image in #325.



Figure 17. Detection results of TLD algorithm: (a) on-land walking; (b) just entering the water; (c) completely into the water and (d) underwater swimming.



Figure 18. Detection results of the improved algorithm: (a) on-land walking; (b) just entering the water; (c) completely into the water and (d) underwater swimming.

4.2 Experiments with Illumination Changes

The detection and recognition results of the algorithm may be influenced by illumination changes and any other factors; the variation of the illumination could cause great influence on the grey feature of the target, and the grey information is often served as detection feature. In this paper, the TLD algorithm and the improved algorithm are both involved in the grey level information in feature extraction; because of the existence of the learning module of TLD algorithm, it can update the feature of target in time and reduce the effect of illumination. In the improved algorithm, the learning module of TLD algorithm and the compressive sensing algorithm was combined together, and experimental results of detection with illumination changes are shown in Figs. 15 and 16. In Fig. 15, the frame rates of TLD algorithm with different frequencies are 2.8 fps and 3.9 fps, respectively. Figure 16 shows the frame rates of the

improved algorithm are 33.7 fps and 18.4 fps respectively, so the improved algorithm could fully meet the real-time requirement of visual detection system for our amphibious spherical robot.

4.3 Experiments in Underwater Environment

To verify the amphibious practicability of the improved algorithm with our robot, the comparison experiment was carried out when the robot changed its motion state from on-land walking to underwater swimming. At this time, the robot's morphological structure continued to change, and the attitude adjustment of robot in underwater also caused the characteristic change of robot, which will also bring some difficulties for the detection, and the algorithm needs to re-complete the detection process as soon as possible. Therefore, the efficiency requirement of the real-time detection is higher. Figures 17 and 18 show the frame

Table 4
Average Frame Rate of Experiment with TLD and Improved Algorithm

Test set	Accuracy (%)		Frame rate (fps)	
	TLD	Proposed algorithm	TLD	Proposed algorithm
Normal illumination	28.2	38.8	6.4	17.6
Illumination changes	24.1	54.2	3.6	26.5
Underwater environment	33.9	43.9	10.2	16.6
The average accuracy/frame rate	31.1	41.4	8.3	17.1

rates of TLD algorithm and the improved algorithm with different frequencies when the robot changed its motion state from on-land walking to underwater swimming; as we can see from experimental results, the phenomenon of occlusion led to temporary loss of the target. However, when the robot completely enters the underwater, the improved algorithm updated in time and finished detection and re-localization. In Figs. 17 and 18, we also see the frame rates of the improved algorithm are largely enhanced.

According to the detection experiment results of robot under different environments, the average frame rate of TLD and improved TLD algorithm is obtained, when the average frame rate of the algorithm is greater than 15 fps, the average accuracy of the algorithm is greater than 30%. We considered that the algorithm could meet the requirements of real-time visual detection system of our amphibious spherical robot, and experimental results of different environments are shown in Table 4.

5. Conclusion and Future Work

In this paper, considering the traditional TLD algorithm could not meet the requirement of the visual detection system of our amphibious spherical robot, a real-time visual detection system of an improved TLD algorithm was designed and implemented in our robot, and BING method and compressive sensing algorithm were introduced to accelerating and optimizing the traditional TLD algorithm. To evaluate the effectiveness of the improved algorithm, a series of experiments have been carried out in some common videos with the impact of illumination changes and various scenarios. Finally, a series of experiments are carried out with the visual system of our robot, and experimental results demonstrated that the improved TLD algorithm has better effectiveness, robustness and real-time performance than the traditional TLD algorithm. The average frame rate of the improved TLD algorithm is lifted from 8.3 fps of TLD to 17.1 fps, and its average accuracy increased from 31.1% to 41.4 %. Consequently, the improved TLD algorithm could fully meet the requirements of real-time visual detection system of our amphibious spherical robot. In the future, we will conduct some

more experiments in complex and variable amphibious environment.

Acknowledgement

This work is supported by the Research Project of Beijing Education Committee (KM201911232005). This work is also supported by the National Natural Science Foundation of China (51475047, 61805017), the Pre-research project of General Equipment Department (6140414030101).

References

- [1] J. Fabian, T. Young, J.C. Peyton Jones, and G.M. Clayton, Integrating the Microsoft Kinect with Simulink: Real-time target tracking example, *IEEE/ASME Transactions on Mechatronics*, 19(1), 2014, 249–257.
- [2] O. Araar, N. Aouf, and J.L. Vallejo Dietz, Power pylon detection and monocular depth estimation from detection UAVs, *Industrial Robot—An International Journal*, 42(3), 2015, 200–213.
- [3] Z. Zhang, A. Beck, and N. Magnenat-Thalmann, Human-like behavior generation based on head-arms model for robot tracking external targets and body parts, *Proceedings of IEEE Transactions on Cybernetics*, 45(8), 2015, 1390–1400.
- [4] E. Wirbel, B. Steux, S. Bonnabel, and D.L.F. Arnaud, Humanoid robot navigation: From a visual SLAM to a visual compass, *Proc. IEEE Int. Conf. Networking, Sensing and Control*, Evry, France, 10–12 April 2013, 678–683.
- [5] A. Delibasi, E. Zergeroglu, I.B. Küçükdemiral, and G. Cansever, Adaptive self-tuning control of robot manipulators with periodic disturbance estimation, *International Journal of Robotics & Automation*, 25(1), 2010, 48–56.
- [6] F. Zhang, S. Zheng, H. Yun, and X. Shao, The research on attitude correction method of robot monocular vision positioning system, *IEEE Proc. IEEE Int. Conf. Robotics and Biomimetics*, Macau SAR, China, 5–8 December 2017, 1972–1976.
- [7] X. Huang, Y. Jia, and S. Xu, Path planning of a free-floating space robot based on the degree of controllability, *Science China Technological Sciences*, 60(2), 2017, 1–13.
- [8] L. Li, X. Wang, D. Xu, and T. Min, An accurate path planning algorithm based on triangular meshes in robotic fibre placement, *International Journal of Robotics & Automation*, 32(1), 2017. DOI: 10.2316/Journal.206.2017.1.206-4673.
- [9] C.J. Lin, A GPU-based evolution algorithm for motion planning of a redundant robot, *International Journal of Robotics & Automation*, 2(2), 2017, 00015.
- [10] F. Castelli, S. Michieletto, S. Ghidoni, and E. Pagello, A machine learning-based visual servoing approach for fast robot control in industrial setting, *International Journal of Advanced Robotic Systems*, 16(6), 2017, 172988141773888.

- [11] S. Lemaignan, M. Warnier, E.A. Sisbot, A. Clodic, and R. Alamiet, Artificial cognition for social human-robot interaction, *Artificial Intelligence*, 247(C), 2017, 45–69.
- [12] C.A. Cifuentes, A. Frizzera, R. Carelli, and T. Bastos, Human-robot interaction based on wearable IMU sensor and laser range finder, *Robotics and Autonomous Systems*, 62(10), 2014, 1425–1439.
- [13] K. Kesorn and S. Poslad, An enhanced bag-of-visual word vector space model to represent visual content in athletics images, *IEEE Transactions on Multimedia*, 14(1), 2012, 211–222.
- [14] Y. Liu, Q. Li, H. Fang, and H. Xu, Research on embedded system with implementation of a moving target tracking algorithm based on improved meanshift on DM6437, *Advanced Materials Research*, 1003, 2014, 207–210.
- [15] J. Liu, K.R. Subramanian, and T.S. Yoo, An optical flow approach to tracking colonoscopy video, *Computerized Medical Imaging and Graphics*, 37(3), 2013, 207–223.
- [16] K. Li, B. Hu, J. Gao, and G. Feng, Nonlinear robust detection Kalman filter algorithm based on M-estimation, *Journal of Computer Applications*, 34(11), 2014, 3214–3217.
- [17] S. Shamshirband, D. Petkovic, H. Javidnia, and A. Gani, Sensor data fusion by support vector regression methodology – A comparative study, *IEEE Sensors Journal*, 15(2), 2015, 850–854.
- [18] Y. Qi, K. Suzuki, H. Wu, and Q. Chen, EK-means tracker: A pixel-wise tracking algorithm using Kinect, *Proc. Third Chinese Conf. Intelligent Visual Surveillance (IVS)*, Beijing, 1–2 December 2011, 77–80.
- [19] C. Bibby and I. Reid, Real-time tracking of multiple occluding targets using level sets, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, 13–18 June 2010, 1307–1314.
- [20] S. Hare, A. Saffari, and H.S. Torr Philip, Struck: Structured output tracking with kernels, *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Barcelona, Spain, 06–13 November 2011, 263–270.
- [21] Z. Kalal, K. Mikolajczyk, and J. Matas, Tracking-learning-detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7), 2012, 1409–1422.
- [22] K. Zhang, L. Zhang, and M.-H. Yang, Real-time compressive tracking, *Proc. European Conf. Computer Vision (ECCV 2012)*, Florence, Italy, 7–13 October 2012, 864–877.
- [23] M. Yahya and M. Arshad, Tracking of multiple light sources using computer vision for underwater docking, *Procedia Computer Science*, 76, 2015, 192–197.
- [24] L. Zhang, B. He, Y. Song, and T. Yan, Consistent target tracking via multiple underwater cameras, *Proc. OCEANS 2016 – Shanghai*, Shanghai, China, 2016, 10–13.
- [25] M. Chuang, J. Hwang, J. Ye, S. Huang, and K. Williams, Underwater fish tracking for moving cameras based on deformable multiple kernels, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(9), 2017, 2467–2477.
- [26] S. Pan, L. Shi, and S. Guo, A Kinect-based real-time compressive tracking prototype system for amphibious spherical robots, *Sensors*, 15(4), 2015, 8232–8252.
- [27] K. Wang, Y. Liu, and L. Li, Visual servoing based trajectory tracking of underactuated water surface robots without direct position measurement, *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Chicago, IL, 14–18 September 2014, 767–772.
- [28] X. Cheng, N. Li, T. Zhou, L. Zhou, and Z. Wu, Object tracking via collaborative multi-task learning and appearance model updating, *Applied Soft Computing*, 31, 2015, 81–89.
- [29] Y. Li, S. Guo, and C. Yue, Preliminary concept of a novel spherical underwater robot, *International Journal of Mechatronics and Automation*, 5(1), 2015, 11–21.
- [30] L. Shi, S. Guo, S. Mao, C. Yue, M. Li, and K. Asaka, Development of an amphibious turtle-inspired spherical mother robot, *Journal of Bionic Engineering*, 10(4), 2013, 446–455.
- [31] L. Shi, Y. He, and S. Guo, Skating motion analysis of the amphibious quadruped mother robot, *Proc. IEEE Int. Conf. Mechatronics and Automation*, Takamatsu, 4–7 August 2013, 1749–1754.
- [32] C. Yue, S. Guo, M. Li, Y. Li, H. Hirata, and H. Ishihara, Mechanronic system and experiments of a spherical underwater robot: SUR-II, *Journal of Intelligent and Robotic Systems*, 2015, DOI: 10.1007/s10846-015-0177-3.
- [33] Q. Fu, S. Guo, Y. Yamauchi, H. Hirata, and H. Ishihara, A novel hybrid microrobot using rotational magnetic field for medical applications, *Biomedical Microdevices*, 17(2), 2015, DOI: 10.1007/s10544-015-9942-0.
- [34] C. Yue, S. Guo, L. Shi, Design and performance evaluation of a biomimetic microrobot for the father-son underwater intervention robotic system, *Microsystem Technologies*, 22(4), 2016, 831–841.
- [35] Y. He, S. Guo, L. Shi, S. Pan, and Z. Wang, 3D printing technology-based an amphibious spherical underwater robot, *Proc. of 2014 IEEE Int. Conf. Mechatronics and Automation*, Tianjin, China, 2014, 1382–1387.
- [36] Y. He, L. Shi, S. Guo, S. Pan, and Z. Wang, Preliminary mechanical analysis of an improved amphibious spherical father robot, *Microsystem Technologies*, 2015, 1–16, DOI: 10.1007/s00542-015-2504-9.
- [37] S. Pan, S. Guo, L. Shi, Y. He, Z. Wang, and Q. Huang, A spherical robot based on all programmable SoC and 3-D printing, *Proc. of 2014 IEEE Int. Conf. Mechatronics and Automation*, Tianjin, China, 2015, 150–155.
- [38] S. Guo, Y. He, L. Shi, S. Pan, K. Tang, R. Xiao, and P. Guo, Modal and fatigue analysis of critical components of an amphibious spherical robot, *Microsystem Technologies*, 2016, 1–15, DOI: 10.1007/s00542-016-3083-0.
- [39] S. Guo, Y. He, L. Shi, S. Pan, R. Xiao, K. Tang, and P. Guo, Modeling and experimental evaluation of an improved amphibious robot with compact structure, *Robotics and Computer-Integrated Manufacturing*, 51, 2018, 37–52.
- [40] X. Zhou, Q. Qian, Y. Ye, and C. Wang, Improved TLD visual target tracking algorithm, *Journal of Image and Graphics*, 18(9), 2013, 1115–1123.
- [41] T. Xu, C. Huang, Q. He, Q.G. Guang, and Y. Zhang, An improved TLD target tracking algorithm, *IEEE International Conf. on Information and Automation*, Macau, China, 2017, 2051–2055.
- [42] M. Cheng, Z. Zhang, W. Lin, and P. Torr, Binarized normed gradients for targetness estimation at 300 fps, *Proc. IEEE Conf. Computer Vision & Pattern Recognition*, Columbus, OH, 23–28 June 2014, 3286–3293.
- [43] B. Alexe, T. Deselaers, and V. Ferrari, Measuring the objectness of image windows, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), 2012, 2189–2202.
- [44] S. Cheng, Y. Cao, J. Sun, G. Liu, and G. Han, Efficient target tracking by TLD based on binary normed gradients, *Optics and Precision Engineering*, 23(8), 2015, 2339–2348.
- [45] K. Zhang, L. Zhang, and M. Yang, Fast compressive tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10), 2014, 2002–2015.

Biographies



Yanlin He received the B.S. degree in communication engineering from Lanzhou Jiaotong University, China, in 2011, and received her Ph.D. degree in 2018 from Beijing Institute of Technology, China. Now she is a lecturer in the School of Instruments Science and Opto-Electronic Engineering at Beijing Information Science and Technology University. Her current research interests include amphibious spherical robot, multi-robot system. She has published about 28 refereed journal and conference papers in the recent years.



Xu Zhang received his B.S. degree in 2007 from Beijing University of Technology, Beijing, China. Now he is a postgraduate in the School of Instruments Science and Opto-Electronic Engineering at Beijing Information Science and Technology University, Beijing, China. His main research interests include image processing technologies and bionic robotics.



Guangkai Sun received his B.S. degree in 2007 and M.S. degree in 2010 from Hebei University of Science and Technology and received his Ph.D. degree in 2015 from Beihang University. Now he is an associate professor in the School of Instruments Science and Opto-Electronic Engineering at Beijing Information Science and Technology University. His main research interests include novel sensing and

testing technologies, and soft robotics.



Lianqing Zhu received his M.S. degree and his Ph.D. in Hefei University of Technology in 1989, and in Harbin Institute of Technology in 2013, respectively. Now he is a professor and Ph.D. supervisor in Beijing Information Science and Technology University. His main research directions include fibre sensing technology and optical measurement.



Junfei Qiao received his M.S. degree and his Ph.D. in Northeastern University in 1998. Now he is a professor and Ph.D. supervisor in Beijing University of Technology. His main research directions include intelligent control theory and its application.