

## A Novel Ship-Bridge Collision Avoidance System Based on Monocular Computer Vision

<sup>1</sup>Yuanzhou Zheng, <sup>1</sup>Weiguo Wu, <sup>1</sup>Haixiang Xu, <sup>2</sup>Langxiong Gan and <sup>3</sup>Jingang Yu

<sup>1</sup>School of Transportation,

<sup>2</sup>School of Navigation, Wuhan University of Technology, Wuhan 430063, China

<sup>3</sup>National Key Laboratory of Science and Technology on Multi-Spectral Information Processing,  
Huazhong University of Science and Technology, Wuhan 430074, China

**Abstract:** The study aims to investigate the ship-bridge collision avoidance. A novel system for ship-bridge collision avoidance based on monocular computer vision is proposed in this study. In the new system, the moving ships are firstly captured by the video sequences. Then the detection and tracking of the moving objects have been done to identify the regions in the scene that correspond to the video sequences. Secondly, the quantity description of the dynamic states of the moving objects in the geographical coordinate system, including the location, velocity, orientation, etc, has been calculated based on the monocular vision geometry. Finally, the collision risk is evaluated and consequently the ship manipulation commands are suggested, aiming to avoid the potential collision. Both computer simulation and field experiments have been implemented to validate the proposed system. The analysis results have shown the effectiveness of the proposed system.

**Keywords:** Monocular vision, moving object detection and tracking, ship-bridge collision

### INTRODUCTION

The frequent occurrence of ship-bridge collision has drawn more and more attention world-widely due to the heavy damage it may cause. During the past decades, a large amount of research work has been done attempting to conquer this problem (Stauffer and Grimson, 1999; Elgammal *et al.*, 2000; Sheikh and Shah, 2005). Despite the significant achievements already made, almost all of these works so far focus on how to remedy the damage after the event, rather than prevent the damage from occurrence beforehand. Therefore, the effect is usually quite limited. On the other hand, a mechanism for active collision avoidance is much more desirable.

Recent years have witnessed the fast development of smart visual surveillance techniques (Zheng, 2010). Thanks to the computer vision and image processing technologies, the smart visual surveillance systems can analyze, without any human interference, the video sequences captured by sensors from a specific sconce of site and furthermore identify, locate and track the changes in the scene and consequently understand the behaviors of objects, which provides strong cues for identifying the abnormal events and raising the alarms actively. Because of its activeness, low cost and immediacy, the smart visual surveillance systems are being more and more widely applied in various fields (Heikkilä and Pietikäinen, 2006).

Motivated by the two aspects above, in this study we propose to utilize the computer vision for active ship-bridge collision avoidance, resulting in a Smart Collision Alarm System (SCAS). From the video sequences captured in field, SCAS first performs the moving object detection and tracking so as to identify the regions in the scene that correspond to the moving ship objects and then calculates the quantities for describing the motion status under world coordinate, including the location, velocity, orientation, etc, which is calculated for each object based on the monocular vision geometry and finally evaluates the ship-bridge collision risk and gives out ship manipulation commands. Extensive testing results, obtained via both computer simulation and field experiments, have shown the effectiveness of the proposed system.

### METHODOLOGY

**The SCAS system:** Figure 1 depicts the systematic structure of SCAS. As shown, the system consists of two main parts, the CCD sensor and the software processing platform. The sensor is installed on top of the bridge, which captures visual signal of the scene and sends to the software platform located in the surveillance center for further processing.

The software processing platform runs core algorithms of the system, of which the overall structure is shown in Fig. 2. The components are detailed in later sections.

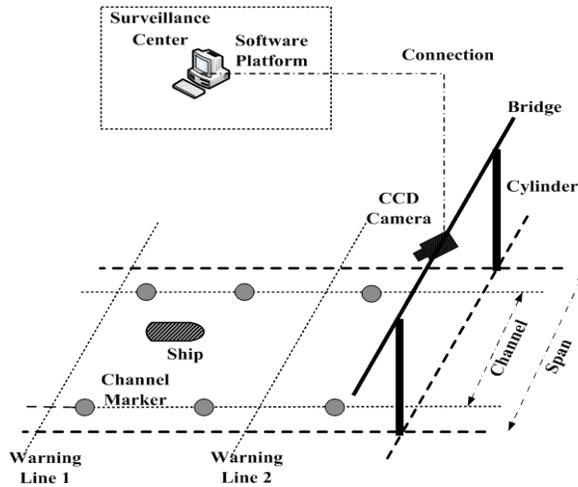


Fig. 1: The systematic structure of SCAS

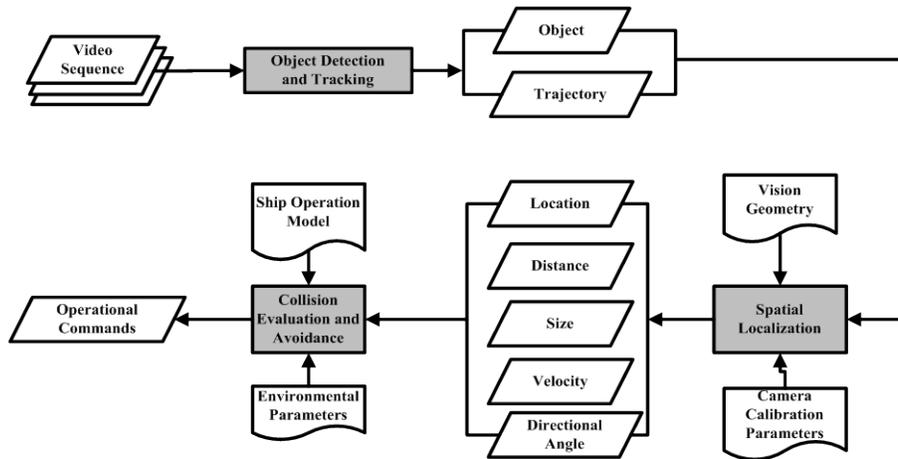


Fig. 2: Algorithmic solution of SCAS

### MOVING OBJECT DETECTION AND TRACKING

Moving object detection and tracking is a preliminary step towards visual surveillance applications. Detection is to identify the regions in images that correspond to the moving objects, while tracking is to associate the objects throughout the frames so as to obtain the trajectory of each over time.

**Moving object detection:** Moving object detection is an intensively addressed topic in literature. In the case of fixed camera, stationary background, or minimal background variation, the problem can be well solved using traditional methods like background differencing, Gaussian or Mixture of Gaussian background modeling, non-parametric background modeling (Stauffer and Grimson, 1999; Elgammal *et al.*, 2000; Sheikh and Shah, 2005), etc. However, when the background contains more complex dynamics, noise, luminance changes, etc, it is usually more difficult to distinguish

the background movements from the object motion and thus more effective methods are desirable (Toyama *et al.*, 1999 ; Li *et al.*, 2004; Ma and Zhang, 2004; Kim *et al.*, 2005; Patwardhan *et al.*, 2008).

In our application, the factors which add the complexity of moving object detection mainly include: weather condition, movement of water in the background, camera vibration, luminance change. An ideal method is expected to adapt to these complex situations as much as possible while satisfy the real-time requirements.

Per our investigation and comparison, in this study we adopt the LBP-based background subtraction method proposed by Heikkilä and Pietikäinen (2006) for motion detection. Considering the computational complexity, we make some modifications to the original algorithm and use the modified version.

The motion detection method used in the study is described in Fig. 3.

In the flow chart in Fig. 3:

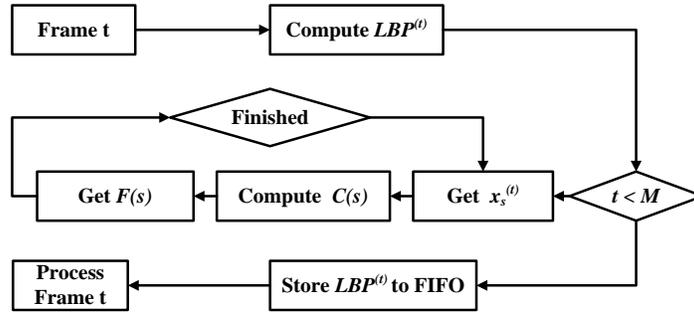


Fig. 3: LBP-based foreground detection

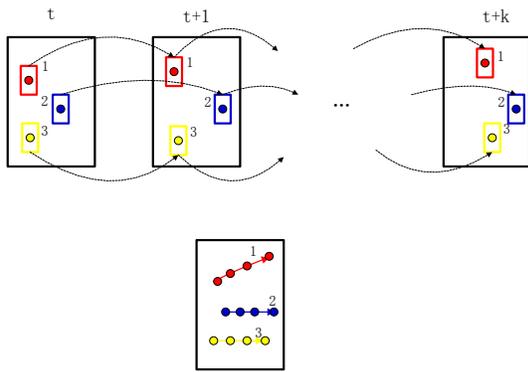


Fig. 4: Demonstration of object tracking

$$C(s) = \sum_{k=1}^M (x_s^{(t)} \otimes x_s^{(t-k)}) \quad (1)$$

where,  $x_s(t)$  denotes the LBP value located at  $s$  in frame  $t$  and  $\otimes$  represents the XOR operation.

$$F(s) = \begin{cases} 1, & C(s) \geq \delta \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

For more details about the algorithm, the readers can refer to Heikkilä and Pietikäinen (2006).

**Object tracking:** Ship object can be reasonably represented as a rectangle with centroid. Object tracking is to associate the multiple objects across frames and obtain the temporal trajectory of each object. In our application, since the relative motion of each object between two consecutive frames is always far smaller than the distance between any two objects, it is reasonable to simply use the nearest corresponding policy for tracking (Fig. 4). More precisely, for the object located at  $s_i$  in frame  $t$ , we calculate the distance from it to each of the objects in the next frame  $t + 1$ , the object with the minimal distance is considered as the association.

**SPATIAL LOCALIZATION**

The location and trajectory obtained by detection and tracking described above are expressed under the

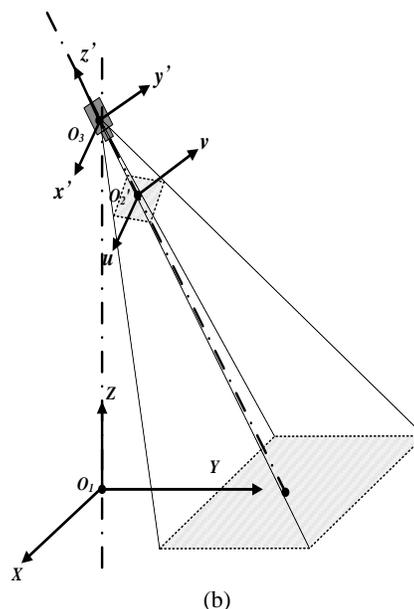
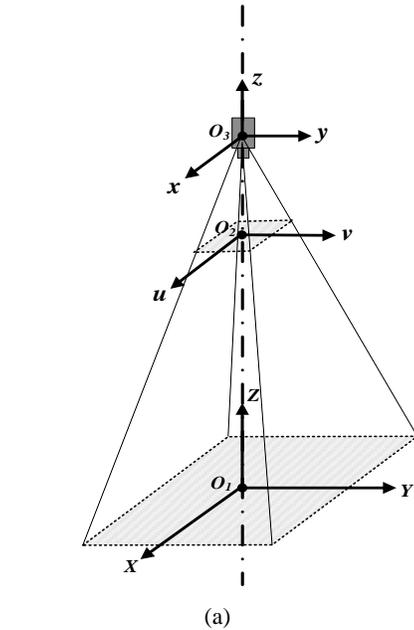


Fig. 5: Vision Geometry: (a) the down-look case (b) arbitrary perspectives

image coordinate, which can not accurately describe the relationship among objects in realistic space. Instead, quantities expressed under the world coordinate system are required. So, a problem is brought forwards: how to convert *the image coordinate* (u,v) into *the geographical coordinate* (X,Y) in the water plane. We show in this section that the problem can be solved based on the geometry of monocular computer vision (Ma and Zhang, 2004).

The conversion between (u,v) and (X,Y) involves three coordinate systems, namely the world coordinate system, the camera coordinate system and the image coordinate system. Let us first look into the geometric relationship in the down-look perspective, while in other arbitrary perspectives, the geometric relationship can be derived based on the down-look case.

In the down-look case, the relationship among the three coordinate systems is shown in Fig. 5a, where  $O_1 - XYZ$  is the world coordinate system,  $O_2 - uv$  the image coordinate system,  $O_3 - xyz$  the camera coordinate system and  $O_1O_2O_3$  the optical axis. The relationship between  $O_3 - xyz$  and  $O_2 - uv$  is

$$\begin{cases} x = u \Delta_x \\ y = v \Delta_y \end{cases} \quad (3)$$

where,

$$\begin{cases} \Delta_x = 2H \cdot \text{tg}(\tau_x / 2) / M \\ \Delta_y = 2H \cdot \text{tg}(\tau_y / 2) / N \end{cases} \quad (4)$$

where, H is height of the camera above the water plane,  $\tau_x, \tau_y$  are view range angles along the X, Y directions respectively, M, N the pixel resolutions of the image along the u, v directions, respectively. The relationship between  $O_1 - XYZ$  and  $O_3 - xyz$  is:

$$\begin{cases} X = x \\ Y = y \\ Z = z + H \end{cases} \quad (5)$$

In other arbitrary perspectives, the pose of the camera can be treated as to be generated based on the down-look case as shown in Fig. 5b, where  $\psi, \theta$  and  $\varphi$  are the rotation angles along the three axes respectively. The relationship between  $O_3 - x'y'z'$  and  $O_3 - xyz$  is:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (6)$$

where, the rotation matrix is:

$$R = R_\psi \cdot R_\theta \cdot R_\varphi \quad (7)$$

$$R_\psi = \begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

$$R_\varphi = \begin{bmatrix} \cos \varphi & 0 & \sin \varphi \\ 0 & 1 & 0 \\ -\sin \varphi & 0 & \cos \varphi \end{bmatrix} \quad (9)$$

$$R_\theta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \quad (10)$$

Note the relationship between  $O_3 - xyz$  and  $O_2 - uv$  is determined by the intrinsic parameters of the camera, which is identical to Eq. (5):

$$\begin{cases} x' = u \Delta_x \\ y' = v \Delta_y \end{cases} \quad (11)$$

Combining Eq. (5-11), we have:

$$\begin{bmatrix} \Delta 0 & 0 \\ 0 \Delta 0 & y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ -H \end{bmatrix} \quad (12)$$

By solving the Eq. (12), the conversion between (X, Y) and (u, v) can be obtained.

### COLLISION EVALUATION AND AVOIDANCE

The main purpose of developing SCAS is to predict the risk of ship-bridge collision and obtain the operation command suggestions to avoid the potential collision. By video analysis stated previously, we can extract quantities to measure the level of collision risk. For example, if the distance between ship and bridge is very small and the velocity of the ship is very large, one can claim that collision is quite likely to occur.

However, only the immediate dynamic state of ship is insufficient for making correct decisions, because the occurrence of collision is not only relevant to ship, but also to bridge (e.g., quality of the bridge) and environment (e.g., speed of wind and water flow). Therefore, we believe an ideal mechanism for risk evaluation and avoidance should simultaneously take into consideration the two aspects of factors, as well as empirical knowledge of ship navigation and operation. The scheme of collision risk evaluation and avoidance in SCAS is shown in Fig. 6. For details on We implemented the SCAS software on Microsoft Visual C++ 6.0 platform. A plenty of video sequences captured on Wuhan Yangtze River Bridge have been used to test and analysis the performance of our system, both qualitatively and quantitatively. Because it is quite difficulty to obtain a dataset of real collision scenario, we have to qualitatively test and evaluate the risk

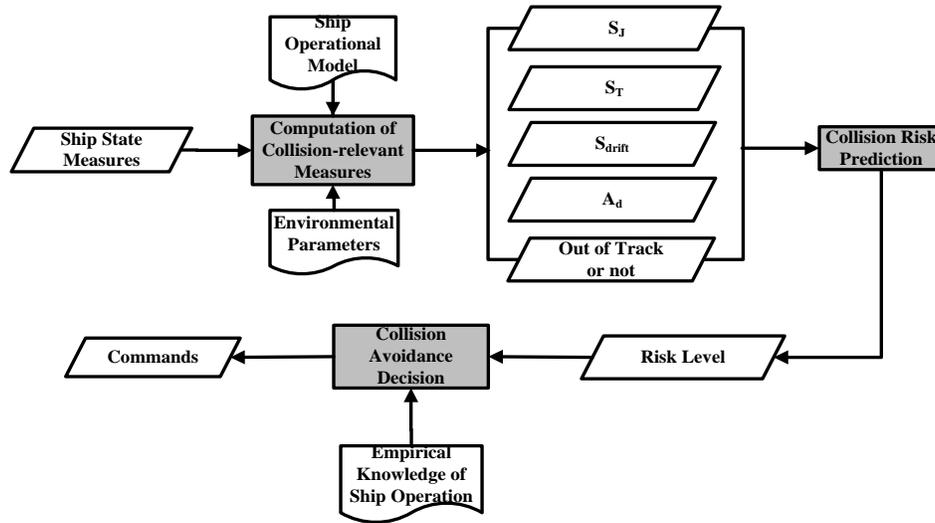


Fig. 6: The scheme of collision risk evaluation and avoidance

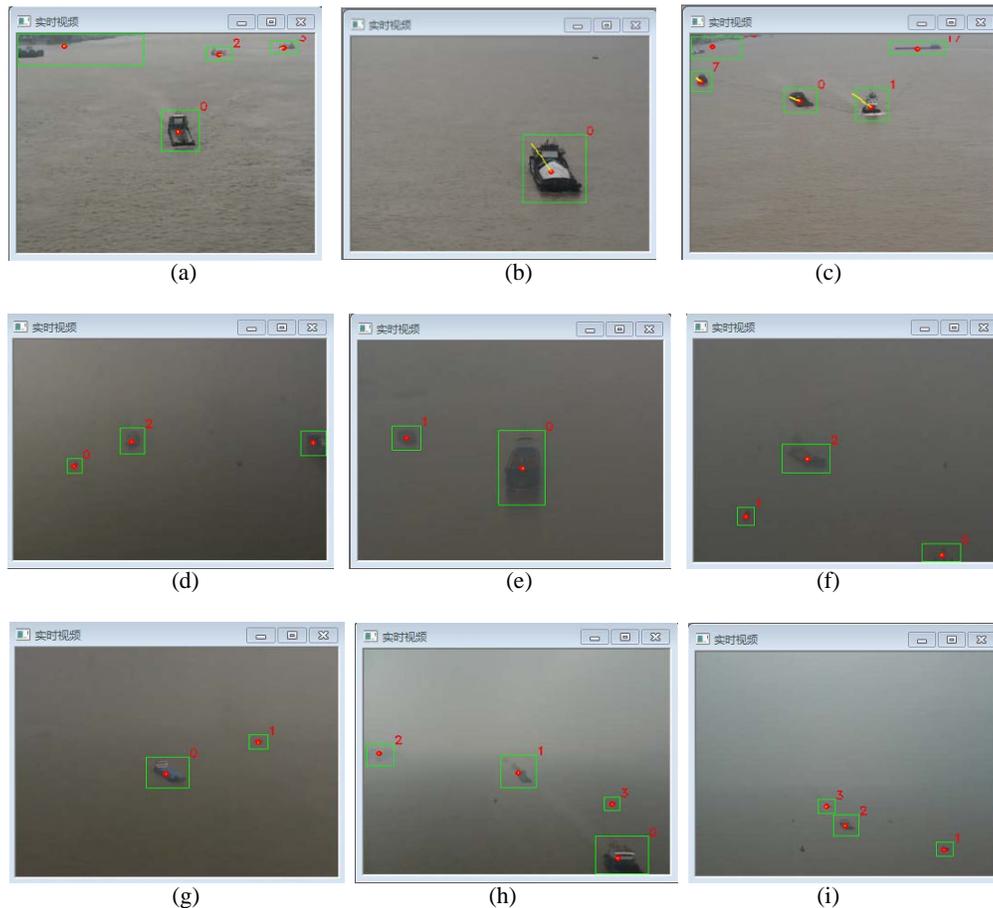


Fig. 7: Detection and tracking results

evaluation and avoidance part, relying on simulation. While for the detection and tracking part, quantitative analysis can be conducted more conveniently. computing  $S_J$ ,  $S_T$ ,  $S_{drift}$  and  $A_d$ , the readers can refer to Zheng (2010).

## EXPERIMENT RESULTS

**Experiment 1: Quantitative analysis:** Fig. 7 demonstrates some typical experimental results. All of the datasets are retrieved on Wuhan Yangtze River

Table 1: Quantitative results

Sequence	Derivation of Images	DR (%)	FR (%)
(a)	48.59	100	11.1
(b)	49.57	50	0
(c)	25.97	91.97	0.74
(d)	3.10	100	30.1
(e)	3.68	100	0
(f)	1.63	100	21
(g)	3.22	100	18.2
(h)	12.59	50	0
(i)	4.527	85.7	0
Average		86.4	8.9

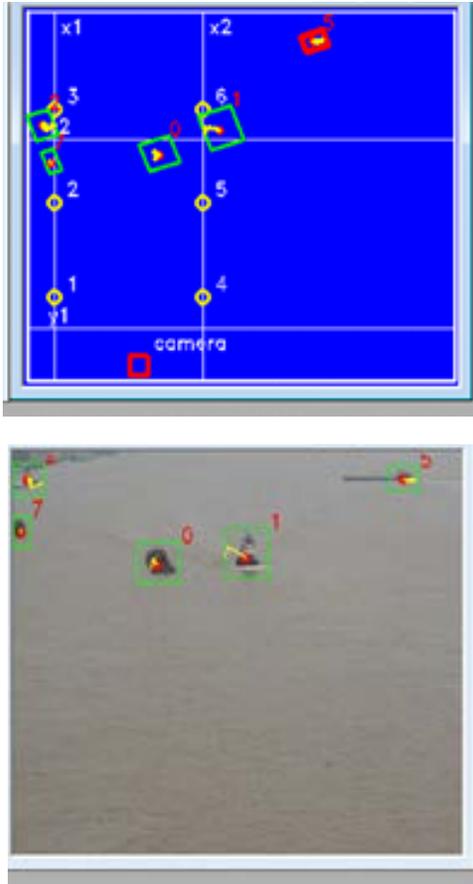


Fig. 8: Results of collision risk evaluation and avoidance

Bridge, in which (a)-(c) are captured in sunny day with short focus length, (d)-(f) in rainy day and (e)-(i) in foggy day and thus the images are in low quality. All the sequences contain camera vibration to some extent. Solid circle in red denotes centroid of the detected ship, rectangle in green denotes the detected region, number in red the identifier of tracking and line in yellow the trajectory. A fixed parameter configuration is used for all the experiments:  $M = 30$ ,  $\delta = 10$ .

For the sake of quantitative evaluation, we adopt two metrics, namely the *Detection Rate* (DR) and the *False Alarm Rate* (FR), defined as follows:

$$DR = \frac{\text{\#of correctly detected objects}}{\text{\# of true objects}} \quad (13)$$

$$FR = \frac{\text{\#of incorrectly detected objects}}{\text{\# of true objects}} \quad (14)$$

The numbers of various kinds of objects are counted manually. The overall results are listed in Table 1, which show that the average DR and FAR on the nine sequences are 86.4 and 8.9%, respectively. The results show that our method is able to accurately detect and track the moving ship objects and adapt well to various complex situations.

**Experiment 2: Qualitative analysis:** By manually modifying the states of detected objects on-the-fly, we simulate the scenarios that have high risk of collision, which are then used to test the effectiveness of the proposed scheme for collision evaluation and avoidance. Result on sequence (c) is given in Fig. 8 as an example, where the object identified as “5” is correctly predicted to have high risk of collision and alarms and operation command suggestions are given accordingly by the system.

## CONCLUSION

In this study, we propose a novel system for ship-bridge collision avoidance based on monocular computer vision. From the video sequences captured in field, SCAS first performs the moving object detection and tracking so as to identify the regions in the scene that correspond to the moving ship objects and then calculates the quantities for describing the motion status under world coordinate, including the location, velocity, orientation, etc, which are calculated for each object based on the monocular vision geometry and finally evaluates the ship-bridge collision risk and gives out ship manipulation commands. Extensive testing results, obtained via both computer simulation and field experiments, have shown the effectiveness of our proposed system. Moreover, the computational complexity of SCAS is reasonably low, making it appropriate for real-time implementation. SCAS is very promising for further research and employment in the realistic application systems.

## REFERENCES

- Elgammal, A.M., D. Harwood and L.S. Davis, 2000. Non- parametric model for background subtraction. Proceedings of the 6th European Conference on Computer Vision-Part II, Springer-Verlag, London, UK, pp: 751-767.
- Heikkilä, M. and M. Pietikäinen, 2006. A texture-based method for modeling the background and detecting moving objects. IEEE T. Pattern Anal., 28(4): 657-662.
- Kim, K., T. Chalidabhongse, D. Harwood and L. Davis, 2005. Real-time foreground-background segmentation using codebook model. Real-Time Imaging, 11(3): 172-185.

- Li, L., W. Huang, Y. Gu and Q. Tian, 2004. Statistical modeling of complex backgrounds for foreground objects detection. *IEEE T. Image Process.*, 13(11): 1459-1472.
- Ma, S. and Z. Zhang, 2004. *Computer Vision*. Science Press, Beijing, China.
- Patwardhan, K., G. Sapiro and V. Morellas, 2008. Robust foreground segmentation using pixel layers. *IEEE T. Pattern Anal.*, 30(4): 746-751.
- Stauffer, C. and W. Grimson, 1999. Adaptive background mixture models for real-time tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2: 246-252.
- Sheikh, Y. and M. Shah, 2005. Bayesian modeling of dynamic scenes for object detection. *IEEE T. Pattern Anal.*, 27(11): 1778-1792.
- Toyama, K., J. Krumm, B. Brumitt and B. Meyers, 1999. Wallflower: Principles and practice of background maintenance. *Proceeding of IEEE International Conference on Computer Vision*, 1: 255-261.
- Zheng, Y., 2010. *Study on Emergency Maneuvering Technologies on Vessel not under Command and Ship-Bridge Collision Avoidance on Bridge Waters*. Wuhan University of Technology Press, Wuhan, China.