

A Study of AI-based Harbor Surveillance System

Dongkoo Shon, Jeongsik Kim, Tae Hyun Yoon, Woo-Sung Jung, Dae Seung Yoo

Intelligent Robotics Research Division

Artificial Intelligence Research Laboratory

Electronics and Telecommunications Research Institute

Daejeon, Republic of Korea

Email: {sdk, j.s.kim, thyo0820, woosung, ooseyds} @etri.re.kr

Abstract—When ships of various types and sizes enter and depart from ports, the risk of collision between ships is high. In particular, collisions between large ships can lead to major accidents. In order to avoid crashes, it is important to quickly identify hazardous situations in advance using harbor surveillance system. But, the harbor surveillance system has limitations in that it continuously requires manpower and the criteria for judgment are different for each person. To solve this problem, we propose an AI based harbor surveillance system including image-based object recognition and abnormal situation recognition. Its superiority was proven through simulation. In the object recognition, YOLOv5, which had the best performance with an mAP of 0.953, was used. The abnormal situation was determined through the distance between the recognized objects, and the comparison of the determination results for 21.5 seconds (66 samples) showed 100%.

Index Terms—Harbor surveillance, Artificial intelligent, Object detection, Anomaly situation determination

I. INTRODUCTION

When ships of various types and sizes enter and depart from ports, the risk of collision between ships is high. In particular, collisions between large ships can lead to major accidents [1]. In order to avoid crashes, it is important to quickly identify hazardous situations in advance. For this purpose, video surveillance, radar, automatic identification system (AIS), CCTV, etc. are being used. However, there is a limit to seeing the entire port as CCTV is fixedly installed. Monitoring using drones is also difficult due to limited batteries. As a way to overcome this, the use of An unmanned aerial vehicle (UAV) such as helikite [2] has been proposed.

A UAV has been developed and commercialized in various forms. In the types of UAVs, there are rotary-wing and fixed-wing, and recently, there is also a form of an unmanned airship. Quadcopters are the most widely studied and used rotary wing type because they can take off and land vertically and can be manufactured in a compact size. The unmanned airship type has the advantage of being able to mount heavier weights and extend operating time. Helikite is a hybrid system of airship and balloons. Helikite does not have a power source like an airship, but and has additional lift and weather stability due to the shape of a balloon attached to a kite wing. So the helikite can be operated for a long time and stable image acquisition is possible, but manpower is continuously required to monitor the image. In addition, there is a difficulty in that the criteria for determining abnormality may vary from person to person. To

solve this problem, we propose a system capable of continuous monitoring without human intervention by automating object recognition and judging abnormal situations based on the recognized object. We compared the object detection performance using 3-various machine learning algorithms(SSD, YOLOv3, YOLOv5) for performance analysis. The location information of the identified object reflecting the camera characteristics was acquired and applied to the collision notification system.

In a recent study, The field of video-based surveillance is also continuing to introduce artificial intelligence. In the [3] study, a study was conducted to graft real-time AI to overcome the limitations of existing methods such as CCTV and patrol. A study on intelligent CCTV was conducted at substations among power facilities [4]. This is a study that implemented AI-based functions for the security and safety of substations while using the existing CCTV system as it is. In the field of defense, research has been done on automated surveillance systems [5]. It is a system that adds the face information of allies to the database and raises an alarm when an intruder not in the database occurs. [6] proposed a real-time object detection edge cloud system for the operational monitoring of airport aprons.

Research is being conducted with artificial intelligence in the maritime and port fields. Because, the performance of computer vision technologies has grow rapidly through integration with deep learning. In [7], AI was applied to automate the detection of objects on the sea surface. [8] used GoogleNet CNN to classify ships in ship traffic monitoring. In particular, the study of Yassir et al. [9] mentioned AIS (Automatic Identification System) and radar data that can be obtained from existing systems as ship tracking data for detecting abnormal conditions. AIS is not mandatory for all ships and has the disadvantage of being able to turn off the power, and the radar has the disadvantage that RCS(Radar Cross Section) is different depending on the size, material, and shape of the ship. The authors proposed that surveillance could be improved through the conduct with a camera vision system.

The remainder of this paper is organized as follows: Section 2 describes the object detection and anomaly situation determination for AI based harbor surveillance system. Finally, Section 3 concludes this paper with discussion and present future work.

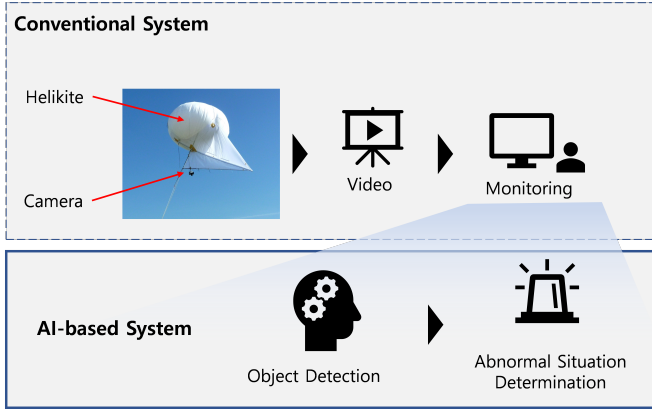


Fig. 1. AI based harbor surveillance system using helikite

II. AI BASED HARBOR SURVEILLANCE SYSTEM

Figure 1 shows an overview of a video-based harbor surveillance system. The conventional system that could continuously acquire images using helikite was configured to automate monitoring of the port using AI. Harbor surveillance automation consists of two main steps. In the first step, it recognises objects such as ships, vehicles, and people from helikite image data. In the second step, the recognized object is used as input data to determinate an abnormal situation that detects the risk of collision between ships.

A. Harbor Object Detection

In order to learn the object recognition algorithm, we created an image dataset for training. Ships, vehicles, and people were found in the harbor images acquired through helikite and labeled using Yolo_mark [10]. The number of images used is 11,318, and the number of objects in the images is 64,151. Table I shows the number of objects by object type. Also, a sample of acquired images is shown in Figure 2.

TABLE I
NUMBER OF TARGETS PER CLASS

Calss (Target Object Type)	Number of Targets
Merchant	10,608
Boat	10,690
Warship	642
Special	115
Uncertain	191
Person	16,541
Car	21,684
Truck	2,100
Bus	512
Bike	693
Etc. vehicle	375

TABLE II
PERFORMANCE COMPARISON BY RECOGNITION MODEL

Object Detection Models	Precision	Recall	mAP
SSD	0.652	0.766	0.844
YOLOv3	0.690	0.725	0.715
YOLOv5	0.932	0.929	0.953



Fig. 2. Sample of Acquired Image Obtained from Helikite

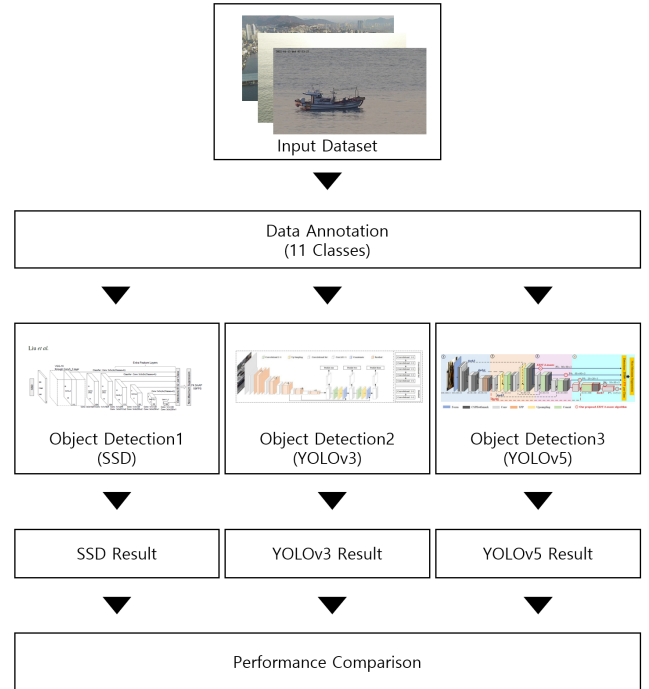


Fig. 3. Overall Process for Harbor Object Detection

For object recognition for harbor monitoring, SSD [11], YOLOv3 [12] and YOLOv5 [13] models were trained and compared performance using datasets, respectively. A overall process shown in Figure 3. Since deep learning requires a large amount of learning data, data augmentation techniques such as color and brightness adjustment randomly and random perspective were used to overcome the small amount of learning data. To reduce the likelihood of a particular result for a particular training data, we use k-fold validation to separate the training and test data for each experimental (i.e., k=5 in this study). The results of calculating Precision, Recall, and mAP for each of the three models are shown in Table II. In the performance comparison, the object recognition rates are 0.844, 0.715, and 0.953 in SSD, YOLOv3, and YOLOv5, respectively. Therefore, in the next section, YOLOv5, which showed the highest performance, was used.

TABLE III
SYMBOLS AND EQUATIONS USED FOR EXTRACTION OF ACTUAL POSITION FROM A PIXEL COORDINATES

Symbols	Description
x_i	a x-axis center of object i, $x_i \in [0, 1]$
y_i	a y-axis center of object i, $y_i \in [0, 1]$
$\rho_{x,i}$	an angular difference horizontally from the center of the image(angular coordinates), $\rho_{x,i} = (x_i - 0.5) \rho_{xMax,i}$
$\rho_{y,i}$	an angular difference vertically from the center of the image(angular coordinates), $\rho_{y,i} = (y_i - 0.5) \rho_{yMax,i}$
$\rho_{xMax,i}$	angular difference from the center of the image to the left based on the camera direction
$\rho_{yMax,i}$	angular difference from the center of the image to the bottom based on the camera direction
ϕ_{cam}	a pitch of a camera
H_{cam}	altitude of a camera
α_i	a converted actual position from the x_i coordinate, $\alpha_i = H_{cam} \cdot \frac{\tan(\rho_{x,i})}{\sin(\phi_{cam} - \rho_{y,i})}$
β_i	a converted actual position from the y_i coordinate $\beta_i = H_{cam} \left\{ \frac{1}{\tan(\phi_{cam} - \rho_{y,i})} - \frac{1}{\tan(\phi_{cam})} \right\}$
w_i	width of the image
h_i	height of the image

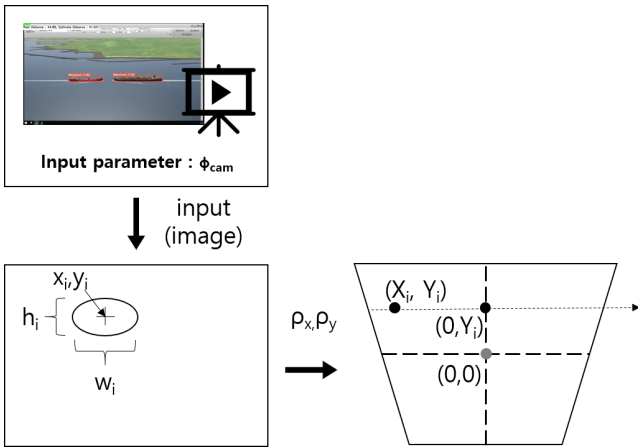


Fig. 4. Process of extracting object coordinates in the image

B. Anomaly situation determination

We consider an approach situation that can cause a ship collision as an abnormal situation and establish a system that can judge it. Since it is quite difficult to obtain the ship collision situation data in the real offshore, the situation was simulated using the ship simulator (K-Sim Navigation, Kongsberg Digital). In the ship simulator, latitude and longitude and simulation image data of two ships were acquired at three samples per second under the condition that the camera's altitude and vertical angle were fixed.

An image data from the camera is two-dimensional pixel data. It is not easy to find out the distance between ships using only pixel coordinates. However, the actual ship is located at a point in the three-dimensional Cartesian coordinate system. If the three-dimensional position of the ship can be determined by transforming the pixel coordinates, the distance between the ships can be calculated. For obtaining 3D coordinates

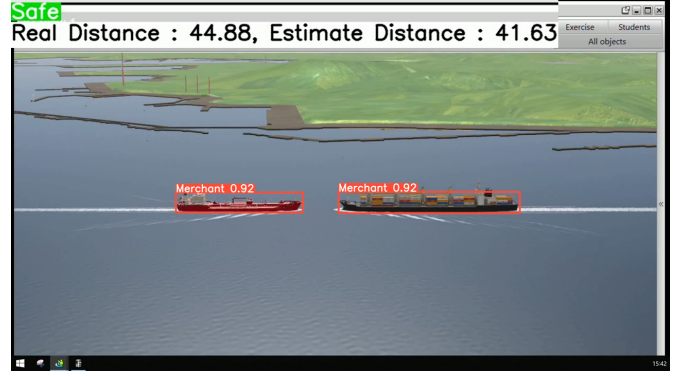


Fig. 5. Video Output using Ship Simulator(Safe)



Fig. 6. Video Output using Ship Simulator (Warning)

using a single camera, we make two assumptions that all objects are at sea(or ground) level and that the angle of one pixel is the same. The symbols and equations used for extraction of actual position from pixel coordinates from an image are summarized in Table III, and the conversion process is schematically illustrated as in Figure 4. [14]

The initial distance between the two ships is about 330 m, and the simulation was conducted so that the distance between the two ships decreased as the experiment time passed. The actual distance between the ships was calculated from the latitude and longitude of the two ships according to the time obtained from the simulator, and the estimated distance between the two ships, a recognized object, was derived based on the image. The criteria for determination an abnormal situation was classified into a dangerous state if the distance between two ships was less than 20 meters, and a safe state otherwise. Table IV shows the results of the abnormal situation judgment experiment from about 3 seconds before the collision. The total length of the experiment is about 21.5 seconds. Although the average distance error between the real distance and the estimated distance is 3.4m, the determination results achieved a high success rate in the experiment environment. Figure 5 and 6 show two examples of video outputs for safe and warning cases.

TABLE IV
TABLE OF ABNORMAL SITUATION DETERMINATION RESULT BY TIME
(FROM 3 SEC. BEFORE THE COLLISION UNTIL THE COLLISION)

Time (Sec.)	Real distance(m)	Estimate distance(m)	Determination result
18.6	44.876	41.634	Safe
18.933	39.8	35.234	Safe
19.25	34.848	32.836	Safe
19.567	29.861	26.402	Safe
19.883	24.873	22.479	Safe
20.2	19.99	17.629	Warning
20.533	14.915	12.043	Warning
20.867	9.839	6.078	Warning
21.2	4.764	0.313	Warning
21.517	-0.057	5.053	Warning

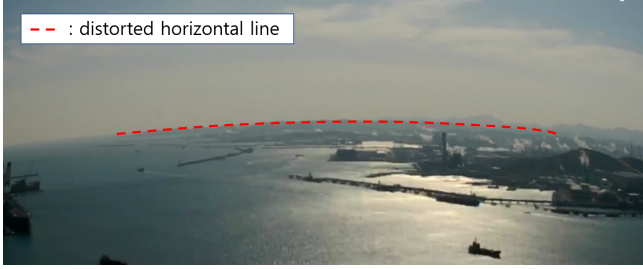


Fig. 7. Example Distortion Occurred Image

III. DISCUSSION AND CONCLUSION

We propose AI based harbor surveillance system consisting of object recognition and anomaly situation determination. In the object recognition, YOLOv5, which had the best performance with an mAP of 0.953, was used. The abnormal situation was determined through the distance between the recognized objects. The comparison of the determination results from real and estimated distance are same.

However, there is a limitation in that the recognition rate of small objects(e.g. person, bike) is somewhat lowered when recognizing objects. It is because that, YOLO model determines objects by dividing a fixed cell grid to match real-time performance. [15] In order to solve the problem, we will research the object detection model for considering small size target scenario, such as determining the access of unauthorized persons and bicycles to the no-port zone is being considered.

Moreover, there is another limitation in the early detection of anomaly situations. We set the criteria for anomaly situations based on only distance that is less than 20 meters. According to the simple criteria, in the experimental results, only 1.5 seconds was given from anomaly detection to collision. In future research, it is necessary to classify the warning stage so that anomalies can be recognized even at a longer distance based on the direction and speed from AIS information.

If we use a wide-angle lens with a wide field of view of the camera, we can monitor a wide area, but this may cause a problem that the image is relatively severely distorted. In addition to visual problems, this image distortion is especially problematic when physical coordinate numerical calculations such as exact position calculation through image analysis

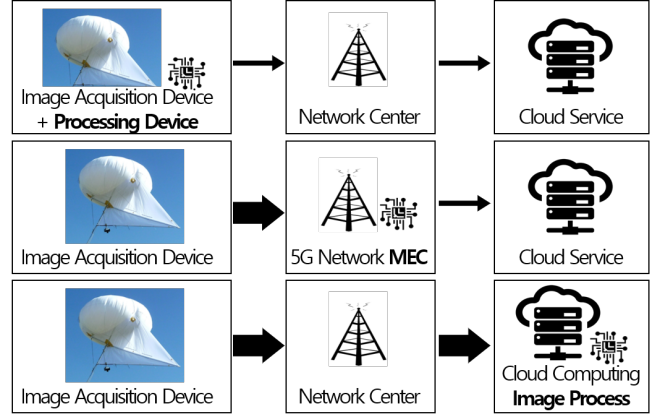


Fig. 8. Various Operation Method of Image Processing Mission Equipment

are required. There are two types of video image distortion: radial distortion and tangential distortion. Radial distortion is caused by the refractive index of the convex lens, and it is relatively easy to correct as the degree of distortion of the image is determined by the distance from the center. On the other hand, tangential distortion is distortion that occurs in the form of an ellipse due to horizontal misalignment between the camera lens and image sensor or lens centering mismatch. Figure 7 is an example of what Figure 2 would look like if distortion occurred. The horizontal line in Figure 2 is straight, whereas the horizontal line in Figure 7 is curved. Obtaining the distortion-corrected image coordinates from the distorted image coordinates is a very difficult problem without a closed-form solution, and the development of technology to correct images in real-time moving unmanned aerial vehicles such as helikites is a problem to be solved.

The operation method of image processing mission equipment can be largely divided into three methods as shown in Figure 8. The first method is to install an image acquisition and processing device on an aerial device such as a helikite, and obtain image processing results directly from the air device. Since the original image does not have to be transmitted through the network, it has the advantage of reducing the burden on the network. Conversely, there is also a restriction that additional mission equipment requiring computing power must be attached to the aerial image acquisition device. The second method is to use a centralized image processing center such as a cloud center. Since only image acquisition equipment and communication equipment need to be operated in the airborne equipment, disadvantages such as increased weight and reduced operating time due to additional equipment are reduced. Instead, large-capacity network resources are required because large-capacity video streaming can be requested for precise video analysis. In addition, due to the remote location of the cloud, there may be limitations in application of application services that require immediate response. The last method is to operate the cloud at the end of the network, similar to the operation in MEC (Multi-access Edge Computing) of the 5G network. It has the advantage of reducing the image

processing burden and speeding up the response time. In the case of port monitoring, since it is difficult to build a network infrastructure, additional research is needed on which method is suitable for port monitoring.

ACKNOWLEDGMENT

This work was supported by Institute of Information & Communication Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2020-0-00869, Development of 5G-based Shipbuilding & Marine Smart Communication Platform and Convergence Service).

A preliminary version of this paper has been presented at KICS [14] [16]. This research provides an enhancement to the existing work by modified analysis results and suggested new intuition.

REFERENCES

- [1] N.-S. Son and S.-Y. Kim, "Validation on the algorithm of estimation of collision risk among ships based on ais data of actual ships' collision accident," *Journal of Navigation and Port Research*, vol. 34, no. 10, pp. 727–733, 2010.
- [2] S. Zakaria, M. R. Mahadi, A. F. Abdullah, and K. Abdan, "Aerial platform reliability for flood monitoring under various weather conditions: A review," in *GeoInformation for disaster management conference*. Springer, 2018, pp. 295–314.
- [3] S.-H. Ko, K.-W. Shin, Y.-S. Jung, and H. Kim, "A study on freeway traffic management through ai based driving image analysis," in *Conference of Korean Society of Transportation*, 2022, pp. 167–168.
- [4] S. Jin, C.-G. Yoon, H.-J. Kim, J.-H. Park, and D. Y. Kim, "Intelligent cctv surveillance system for substation," in *Conference of Korean Institute of Electrical Engineers*, 2022, pp. 2155–2156.
- [5] T. Bhupathi, A. Chittala, and V. Mani, "A video surveillance based security model for military bases," in *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*. IEEE, 2021, pp. 517–522.
- [6] Z. Lyu and J. Luo, "A surveillance video real-time object detection system based on edge-cloud cooperation in airport apron," *Applied Sciences*, vol. 12, no. 19, p. 10128, 2022.
- [7] Y. Guo, Y. Lu, Y. Guo, R. W. Liu, and K. T. Chui, "Intelligent vision-enabled detection of water-surface targets for video surveillance in maritime transportation," *Journal of advanced transportation*, vol. 2021, 2021.
- [8] K. Bobkowska and I. Bodus-Olkowska, "Potential and use of the googlenet ann for the purposes of inland water ships classification," *Polish Maritime Research*, 2020.
- [9] Y. Zardoua, A. Astito, and M. Boulaala, "A comparison of ais, x-band marine radar systems and camera surveillance systems in the collection of tracking data," *arXiv preprint arXiv:2206.12809*, 2022.
- [10] A. Bochkovskiy, T. Hager, I. Ran, R. Moelker, V. Zhang, Lovexyto, and Casper. (2016) Yolo_mark. [Online]. Available: https://github.com/AlexeyAB/Yolo_mark
- [11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.
- [12] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [13] G. Jocher, A. Stoken, J. Borovec, C. L. NanoCode012 C., and H. Laughing. (2020) yolov5 v3.0. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [14] J. Kim, D. Shon, W.-S. Jung, T. H. Yoon, and D. S. Yoo, "A proof-of-concept study on data-driven simulation using vision and an object detection technique based on deep learning," in *Korea Information and Communication Society (KICS) Summer Conference 2022*, 2022, pp. 773–774.
- [15] W.-S. Jung and D. S. Yoo, "Industrial humanless monitoring system over 5g networks in the shipyard environment," in *2022 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2022, pp. 19–24.
- [16] T. H. Yoon, D. Shon, J. Kim, W.-S. Jung, and D. S. Yoo, "Artificial intelligence-based port monitoring system with moored unmanned aerial vehicle," in *Korea Information and Communication Society (KICS) Summer Conference 2022*, 2022, pp. 759–760.



Dongkoo Shon received the B.S. degree in Department of Computer Engineering and Information Technology, and the M.S. and Ph.D. degrees from the Department of Electrical, Electronics and Computer Engineering, University of Ulsan, in 2013, 2015, and 2022, respectively. He is currently a researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. His current research interests are in AI application in industrial field.



Jeongsik Kim received the B.S. degree in system design and control engineering and the Ph.D. degree in mechanical engineering from Ulsan National Institute of Science and Technology, in 2014 and 2021, respectively. He is currently with Electronics and Telecommunications Research Institute. His research primarily focuses on multi-agent systems and digital twins for industrial applications.



Taehyun Yoon received the B.S., M.S., and Ph.D. degrees from the School of Electronics Engineering, Kyungpook National University, in 2005, 2007, and 2015, respectively. Since 2016, he has been a researcher with the Intelligent Robotics Research Division, Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. His current research interests include applied mobile communication, ship-IT convergence, and LoRa networks.



Woo-Sung Jung received the dual B.S. degrees in electrical and computer engineering and in information and computer engineering, and the M.S. and Ph.D. degrees from the School of Computer Engineering, Ajou University, South Korea, in 2007, 2009, and 2015, respectively. He is currently a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. His current research interests are in wireless networking, the Internet of Things, device-to-device communication, and embedded systems.



Daeseung Yoo received the B.S., M.S., and Ph.D. degrees from the Department of Computer Engineering and Information Technology, Ulsan University, in 1998, 2001, and 2011, respectively. Since 2009, he has been a Research Engineer with the Intelligent Robotics Research Division, Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. His current research interests include applied software engineering, mobile communication, ship-IT convergence, and ad hoc networks