

Maritime Object Segmentation and Tracking by using Radar and Visual Camera Integration

Jae-Jeong Hwang, Sang-Gyu Cho, Jung-Sik Lee and Sang-Hyon Pak, *Member, KIMICS*

Abstract— We have proposed a method to detect and track moving ships using position from Radar and image processor. Real-time segmentation of moving regions in image sequences is a fundamental step in the radar-camera integrated system. Algorithms for segmentation of objects are implemented by composing of background subtraction, morphologic operation, connected components labeling, region growing, and minimum enclosing rectangle. Once the moving objects are detected, tracking is only performed upon pixels labeled as foreground with reduced additional computational burdens.

Index Terms— Radar-camera integrated system, Foreground modeling unit, Background modeling unit, Region growing segmentation.

I. INTRODUCTION

Ship Radar has been widely utilized, since it was developed in 1940 by the U.S. Navy as an acronym for Radio Detection And Ranging. It is an objection detection system by using electromagnetic waves to detect direction, speed and altitude of the objects. The basic principle is that parts of transmitted radio waves from the sender are reflected on the objects. The signal which is weak reflected can be amplified in the receiver to enable detection and tracking. The information of radar includes the bearing and range (position) of the object on the monitor. Thus, it can be used on land, at sea and in the air for military purpose, surveillance, voyage, ranging, etc.

Marine radars are used to measure the bearing and location of ships to navigate through safe way, to prevent collision with unwanted objects and to watch for weather change. Results are displayed on black-white monitor as clustering of the reflected waves from various sources such as ships, islands, clouds, and buoys. Automatic target detection/tracking is a challenging task to be done by radar system at sea that involves the extraction of critical information about a target from a sequence of complex objects [1]. It is very difficult to detect objects if the radar images present a many similar objects and a low signal-to-noise ratio. Even though the image is achieved with

high resolution, recognition of object contents, such as shape, size, color and type of target ships is another problem to be identified, since the radar image does not provide gray levels.

Recognition of image contents is possible by using visual camera with high resolution gray levels. It is possible to display real target images by overlaying on a monitor or on a separate window to identify properties of targets. Since the distance of visibility is limited to a few kilometers, radar and visual camera should be cooperated to provide position and to provide real scene information. Range of radar is usually more than 10 km but that of camera is less than 5 km. It is useful for detecting fast moving targets in severe weather condition.

Only a few trials have been performed for radar and visual camera integration for the purpose of harbor surveillance system. Radar sensors detect dangerous targets in user-defined alarm zones. Alarms are generated to alert security personnel and PTZ cameras are directed to the targets [2]. Performance of airborne earth observing system increases with the integration of two different systems [3]. It can be equipped with helicopters [4],[5] or dump trucks [6]. Short distance targets from vehicles are exactly detected by radar and camera integration on highway and road [7].

Object detection and tracking from visual data are so sensitive due to the noisy nature of marine environment. Fast and robust segmentation is an important process to abstract valuable information. Real-time segmentation of moving regions in image sequences is a fundamental step in many vision systems including automated visual surveillance, human-machine interface, and very low-bandwidth telecommunications [1]. A typical method is the background subtraction. Many background models have been introduced to deal with different problems. One of the successful solutions to these problems is to use a multi-color background model per pixel proposed [8]. However, the method suffers from slow learning at the beginning, especially in busy environments. Fast operation is a crucial factor for marine application. We suggested in this paper a fast algorithm and implementation of foreground and background separation. Performance increases with the help of position information from ship radar.

This paper is structured as follows. In Section 2, we describe overall system of radar and visual camera integration system. Object tracking algorithm is discussed in Section 3 and we present an algorithm based on radar

Manuscript received July 1, 2010; accepted July 1, 2010.

Jae-Jeong Hwang, Sang-Gyu Cho, and Jung-Sik Lee are with the Department of Electronic & Information Eng., Kunsan National University, Kunsan, 573-701, Korea (Email: {hwang, sgcho, leesj}@kunsan.ac.kr)

Sang-Hyon Pak is with DICS Vision Co., Ltd., Korea (Email: {iam-hyon@gmail.com})

information. Simulation and results are shown in Section 4 and conclusions are followed in Section 5.

II. RADAR-CAMERA INTEGRATED SYSTEM

The proposed system acquires image sequences formed by a radar and visual camera as shown in Fig. 1. In radar images, ships appear as bright regions, while the sea is a darker area, if it is clear enough. The grey-level variance over ships and sea is caused by a multiplicative correlated noise, called speckle, which is generated by the coherent processing method needed to synthesize the radar images. The speckled image, however, provides rough information of ships and the radar provides position and bearing information.

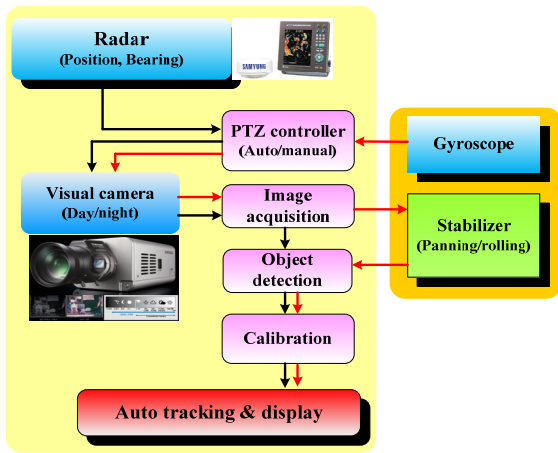


Fig. 1. Radar-camera integrated target tracking system overview.

Position and bearing information activate PTZ controller of camera in automatic or manual mode. Actual status of the PTZ is derived by the gyroscope value that informs orientation, when the camera mounted on a ship is stationary or moving. Target images are obtained by camera with the finest resolution as possible. Final target is defined in an image by the proposed object tracking algorithm. Detailed discussions will be made in following sections.

III. OBJECT TRACKING BASED ON FOREGROUND AND BACKGROUND SEGMENTATION

Fig. 2 shows the ship tracking algorithm block diagram which consists of ship radar, visual camera and data processing units by using foreground and background modeling units. The ship radar provides real-time position information in the NMEA (National Marine Electronic Association) format. This data includes the position, velocity and time information computed by the GPS receiver, which will be used to control the PTZ of camera to take

an image sequence.

Each image of the camera sequence is first segmented based on background and foreground modeling. Then noise removal is performed by binary and morphology operation. Resulting image contains shadow area that causes misleading segmentation. Shadow areas are removed by color and brightness similarity algorithm. Candidate objects from pre-processed results are analyzed by the connected component labeling process, and the central point (coordinate) of object is determined in an image frame. The point of object which is a seed point of region-based image segmentation is used for finding a clear object on the next procedure. At the last step, MER (minimum enclosing rectangle) procedure is applied to draw a rectangle over detected object.

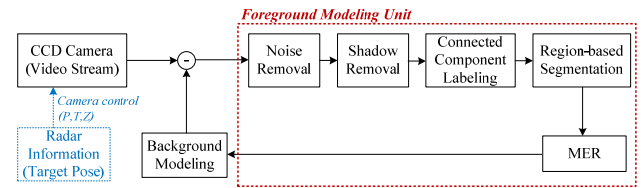


Fig. 2. Structure of marine target tracking algorithm.

A. Foreground Modeling Unit

To detect foreground objects (target) from the background in video stream is a critical first step in many computer vision applications. Traditionally, background subtraction (BS) is a commonly used technique for achieving this subject. Despite of numerous BS algorithms has been proposed in applications such as a human computer interaction, video surveillance, and traffic monitoring to meet real-time goals, the popularity of BS comes from its computation efficiency [9].

To detect foreground object, we use AGMM (Adaptive Gaussian Mixture Model) [10],[11] algorithm with $K(=3)$ components, and its results passed some post-processing. Each pixel in the scene of the video is modeled by a mixture of K Gaussians distributions. The probability that a certain pixel has a value of X_t at time t can be written as

$$P(X_t) = \sum_{i=1}^K w_{i,t} \cdot \eta(X_t, \mu_{i,t}, \sum_{i,t}) \quad (1)$$

where $w_{i,t}$ is the weight parameter of the i^{th} Gaussian component. η is the i^{th} Gaussian probability density function represented by

$$\eta(X_{i,t}, \mu_{i,t}, \sum_{i,t}) = \frac{1}{(2\pi)^{n/2} |\sum_{i,t}|^{1/2}} \cdot e^{-1/2(X_{i,t} - \mu_{i,t})^T \sum_{i,t}^{-1} (X_{i,t} - \mu_{i,t})} \quad (2)$$

where μ_i is the mean and $\sum_{i,t}$ is the covariance matrix of the i^{th} component ($\sum_{i,t} = \sigma_{i,t}^2 I$).

The first Gaussian component will be updated by

$$\begin{aligned} w_{k,t} &= (1 - \alpha) \cdot w_{k,t-1} + \alpha \cdot (M_{k,t}), \\ \alpha &: \text{learning rate}, \\ M_{k,t} &: \begin{cases} 1, & \text{model matched} \\ 0, & \text{remaining model matched} \end{cases} \end{aligned} \quad (3)$$

The BS is performed by marking a foreground pixel which is greater than some standard deviation away from any of the B distributions and first B distributions are used as a model of the background of the scene where B is estimated as

$$B = \arg_b \min \left(\sum_{j=1}^b w_j > T \right) \quad (4)$$

where T is threshold for a minimum fraction of background model.

1) Noise Removal

Morphological operation includes four basic operations: erosion, dilation, opening and closing. The opening of P and Q is obtained by the erosion of P and Q , followed by dilation, while the closing of P and Q is obtained by the dilation followed by the erosion. With some relevant operators, noise in an image is reduced. Parameters to be determined are window size and number of operations. We set the window size 3 and execute opening/closing operations.

2) Shadow Removal

The results of BS method are often confused by shadow regions and it caused misclassification of the objects and shadow regions disturb further processes perceived the shapes or exact positions of moving objects. In this paper, we use RGB color model for shadow elimination which is based on similarities between components [12] and it is able to correspond adaptively to illumination changes. Normalized RGB is given by

$$r = \frac{R}{T}, \quad g = \frac{G}{T}, \quad b = \frac{B}{T}, \quad T = R + G + B \quad (5)$$

Color similarity $P_{color}(i)$ to the background at each pixel is calculated by

$$P_{color}(i) = e^{-\left(\frac{(i_r - m_r)^2}{2s_r^2} + \frac{(i_g - m_g)^2}{2s_g^2} \right)} \quad (6)$$

where i_k denotes normalized RGB component, m_k and

s_k are mean and standard deviation of background, respectively. The color similarity becomes one, if the pixel color is close or equal to the background color.

We also define *brightness similarity* $P_{bright}(i)$ to the background as given by

$$P_{bright}(i) = e^{-\left(\frac{(i_d - s_m)^2}{2s_s^2} \right)} \quad (7)$$

where i_d denotes difference between input image and background model and s_m and s_s are mean and standard deviation of difference between shadow image and background model, respectively. If the value i_d is close or equal to the mean value s_m , the brightness similarity becomes one, meaning maximum similarity.

3) Connected Components Labeling

Objects are expressed by means of the connected components labeling operation for binary images, in which assigns a unique label to each maximal connected region of foreground pixels [13]. It relies on two subsequent raster scans of the image. In the first pass, a temporary label is assigned to each foreground pixel based on the values of its neighbors already visited by the scan. In the second pass the labels associated with the pixels in the neighborhood are registered as being equivalent. Thus, all pixels in a connected component share similar pixel intensity values and are in some way connected with each other.

Connected component is performed for binary input images with 8-connectivity. The input image is scanned along a row until it comes to a point p for which the binary value $V=1$, belonging to foreground pixel. Then, it examines the four neighbors of the pixel which have already been encountered in the scan. Based on this information, the labeling occurs as follows:

In the first pass:

- Scan each pixel by column(row), then by row(column)
- If the pixel is the foreground,
- If there are no neighbors, uniquely label the pixel.
- Otherwise, find the neighbor with the smallest label and assign it to the current pixel.

In the second pass:

- If the pixel is foreground, relabel the pixel with the lowest equivalent label.

Connected component labeling is useful to detect unconnected regions and to obtain the shape of objects. However, the efficiency is limited to the small 3x3 regions. Object connection which is greater than the small window is still unconnected.

4) Region-growing segmentation

The goal of segmentation is to partition an image into objects, which is one of necessary step for tracking and identification. Region-based image segmentation is a simple method involving of initial seed point. This is achieved by looking for the boundaries between regions based on discontinuities such as thresholding in gray level or color properties. Correct seed point and relevant thresholding are two important factors in the region growing method. Pixels or sub-regions are grouped into larger regions, which are grown by appending to each seed those neighboring pixels that are similar to the seed [14].

We solve the two critical issues of correct seed point and thresholds by using Radar information in this research. Position information in latitude and longitude from the Radar has higher priority than that calculated in the camera image processing. Larger threshold results in larger size of regions, while lower threshold provides smaller size of segmented regions. This issue is solved by using binary image which has small variation.

5) Minimum Enclosing Rectangle

Final stage of the object segmentation is the minimum enclosing rectangle (MER). The MER for an object in the foreground is defined as its bounding box aligned such that it encloses all the points in the object with the area minimized. To determine the MER, the object boundary is rotated through 90° in steps of $\sim 3^\circ$. Following each stepwise rotation, a horizontally aligned bounding box is fit to the object boundary, and the minimum and maximum horizontal and vertical values of the rotated boundary points are recorded. At a particular angle, the area of the bounding box is minimized, and this defines the MER. The rectangularity of an object can be measured with the rectangle fit factor R .

$$R = \frac{A}{A_r} \quad (8)$$

This is simply the ratio of the object's area A to the area of its MER, A_r . The rectangle fit factor represents how well an object fills its minimum enclosing rectangle. This parameter takes on a maximum value of 1 for rectangular objects. It is bounded between 0 and 1, taking the value $\pi/4$ for circular objects and smaller values for slender, curved objects.

Another related shape measure is the aspect ratio computed as:

$$A_{MER} = \frac{W}{L} \quad (9)$$

It is the ratio of the width W to the length L of the minimum enclosing rectangle, and it is used to distinguish slender objects from roughly square or circular object.

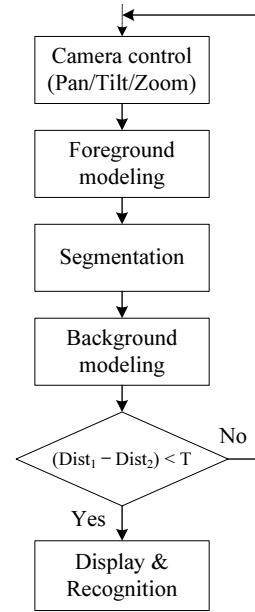


Fig. 3. Proposed ship object segmentation algorithm.

B. Background Modeling

Background modeling is integrated with the foreground modeling to find deviations from the ship objects at sea as indicated in the flow diagram in Fig. 3. Any significant change in an image from the background model signifies a moving object. Frame differencing of temporally adjacent frames is called background subtraction, which is simple and well represent moving information. It is combined with the foreground modeling including connected component algorithm, since the residual information does not represent complete information of the object. One limitation of the background modeling is that it requires a static background. Since application of this work includes ship detection and tracking, the system should be able to deal with time-varying background (e.g., the waves on the water, moving clouds) [15]. To solve the problem, we utilize the position information from the ship Radar and further image processing algorithm such as region-growing and the MER.

Ship object is centered in an image buffer to be processed for segmentation. If we assume static circumstance, centered image is not so moving from time to time, assuming that the Radar object tracking works perfectly. However, we have to take into account two position data: one from Radar and one from image processing system. Location of object is likely to move outward image frame in bad condition. Thus, we assign half the horizontal frame size as the maximum displacement, $Dist_2$, which is allowed moving in an image frame. This is the stability algorithm in the system.

Detection range of maritime objects is an important factor in this system. It is related to CCD size, focal length, and object size as illustrated in Fig. 4. It also depends on the size of object in the screen, which should

not be too small and too big. For example, assuming the vertical size of an object (10 m height) is a third of the screen (30 m height), focal length 99 mm, and 1/4 inch CCD, the maximum detection range is about 1100 m as given by

$$WD = \frac{f \cdot V}{v} = \frac{99\text{mm} \cdot 30\text{m}}{2.7\text{mm}} = 1100\text{m} \cdot \quad (10)$$

Thus, the ship operator is able to detect unidentified object 3.6 min earlier before collision, assuming the ship speed of 10 knots, which is enough to activate the proper counter-action.

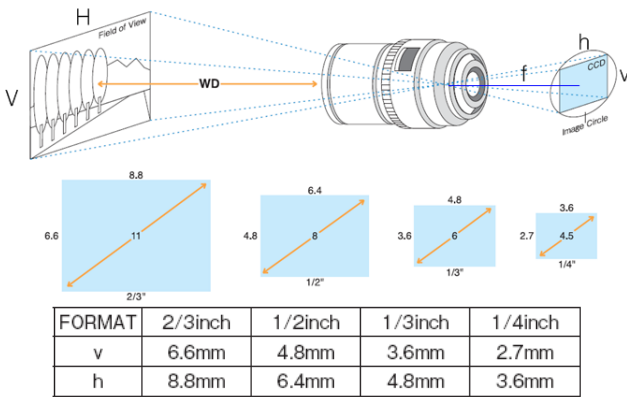


Fig. 4. Parameters for detection distance which is related to CCD size, focal length, and object size.

IV. TEST RESULTS

This section demonstrates the results of our proposed algorithms on an image sequence. Resolution of the sequences is 720x480, taken by visual camera on the top of ship mount. The CCD Camera has the flowing features:

- 40,000 pixels, CCD camera, 720x480(HxV), 24bits
- PCI capture board with BNC connect, No compression

Fig. 5 shows two input images taken by visual camera. Target images should position on the center of an image frame. However, two images are on off-center to show some displacement for the purpose of demonstration.



Fig. 5. Two displaced target images.

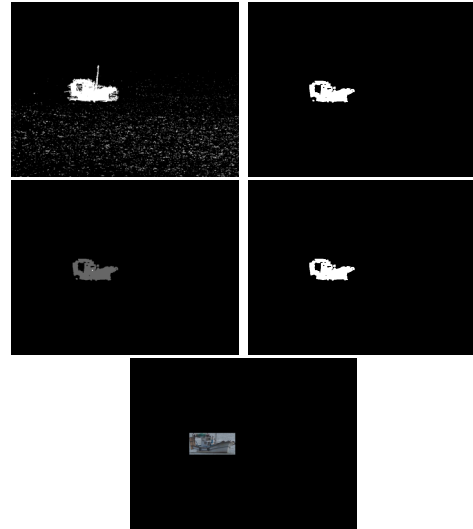


Fig. 6. Resulting images at the starting position: (from top-left) a) differenced image, b) morphologic operation, c) connected components labeling, d) region growing, and e) MER.

We used an adaptive mixture of three Gaussian components. Fig. 6a shows the result of adaptive Gaussian mixture model with noises from waves as expected. Morphology operation is applied to remove noises and shape the candidate object as shown in Fig. 6b, which is followed by labeling algorithm in Fig. 6c. During the process of labeling, a central point of the object is calculated, which will be used for seed point in the region growing. Noises are further removed in this process. Final step is to draw a rectangle surrounding the object and to display the natural scene in Fig. 6e.

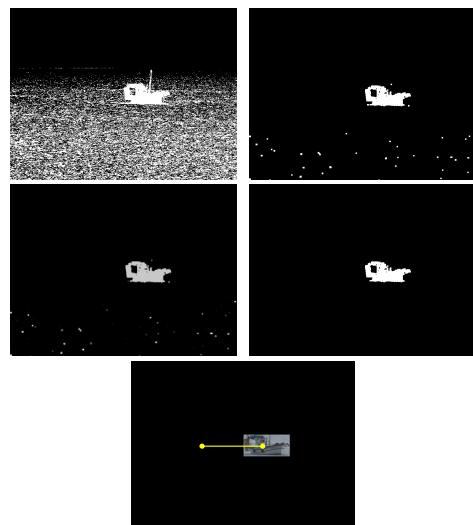


Fig. 7. Resulting images at the displaced position: (from top-left) a) differenced image, b) morphologic operation, c) connected components labeling, d) region growing, and e) MER.

Result of background modeling includes severe noises from moving waves as shown in Fig. 7a, which has been displaced from the starting position. The effect of noise is eliminated by following morphologic operation, connected component labeling, and region growing processes. Displaced position is informed by visual arrow starting from the previous location after the MER step.

V. CONCLUSIONS

We have presented new segmentation algorithms for ship object tracking at sea. The algorithms run under the framework of the real-time ship navigation system, consisting of Radar and visual camera. Although Radar has been used for a long time, integrated scheme with visual camera is an efficient way to enhance marine surveillance system. We have proposed a method to detect and track moving ships using position from Radar and image processor. The two slightly different positioning results are adaptively controlled to fix the object in the center of an image frame. Algorithms for segmentation of objects are implemented by composing of background subtraction, morphologic operation, connected components labeling, region growing, and minimum enclosing rectangle. Once the moving objects are detected, tracking is only performed upon pixels labeled as foreground with reduced additional computational burdens.

REFERENCES

- [1] A. Mecocci, et al., "Radar image processing for ship-traffic control," *Image and Vision Computing*, vol. 13, no. 2, pp. 119-128, Mar. 1995
- [2] T. Kitade and I. Yagi, "Harbor surveillance Radar/Camera system," *Osaka and its Technology*, pp. 22-27, Dec. 2002
- [3] H. Guo et al., "A new airborne Earth observing system and its applications," *Int. Symp. on Geoscience and Remote Sensing (IGARSS '01)*, vol. 1, pp. 549-551, 9-13 July 2001
- [4] K. Yamamoto et al., "Millimeter wave radar for the obstacle detection and warning system for helicopters," *IEE Conf. RADAR*, pp. 94-98, 15-17 Oct. 2002
- [5] C. Migliaccio et al., "Millimeter-wave Radar for rescue helicopters," *9th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV '06)*, pp. 1-6, 5-8 Dec. 2006
- [6] T. Ruff, "Evaluation of a radar-based proximity warning system for off-highway dump trucks," *Accident Analysis & Prevention*, vol. 38, no. 1, pp. 92-98, Jan. 2006
- [7] H.-Y. Lin, K.-J. Li, and C.-H. Chang, "Vehicle speed detection from a single motion blurred image," *Image and Vision Computing*, vol. 26, no. 10, pp. 1327-1337, Oct. 2008
- [8] C. Stauffer and W.E.L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. on Pattern Analysis & Machine Intell.*, vol. 22, pp. 747-757, 2000
- [9] D.H. Parks and S.S. Fels, "Evaluation of background subtraction algorithms with post-processing," *IEEE Fifth International Conf. on Advanced Video and Signal Based Surveillance*, pp. 192-199, 2008
- [10] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, 27(7), pp. 773-780, 2006
- [11] P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detec-

tion," *In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems*, Sept. 2001

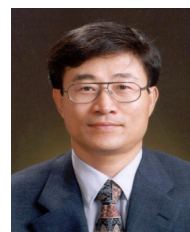
- [12] B.S. Kim et al., "An effective shadow elimination method using adaptive parameters update," *Trans. of Inst. of Electron. Eng. of Korea*, vol. SP-45, no. 3, pp. 11-19, May 2008
- [13] L. di Stefano and A. Bulgarelli, "A Simple and Efficient Connected Components Labeling Algorithm," *Proc. 10th Int. Conf. on Image Analysis and Proc.*, pp. 322, 1999
- [14] R.C. Gonzalez, R.E. Woods, and S.L. Eddins, *Digital image processing using Matlab*, Prentice Hall, 2004
- [15] A. Yilmaz, O. Javed, and M. Shah, "Object Tracking: A Survey," *ACM Computing Surveys*, vol. 38., no. 4, pp. 1-45, Dec. 2006



Jae-Jeong Hwang received the B.S., M.S., and Ph.D. degrees in electronic engineering from the Chonbuk National University in 1983, 1986, and 1992, respectively. He is currently a professor at the Kunsan National University, Korea, and as adjunct professor of RMIT University, Australia. His research interests are digital image/video coding & processing, information theory, object segmentation and tracking, and 2D/3D visual quality assessment and evaluation. He is the coauthor of *Techniques and standards for image, video and audio coding* (Prentice Hall, 1996) and *Fast Fourier transform – Algorithms and applications* (Springer, 2010).



Sang-Gyu Cho received the B.S. and M.S. degrees in electronic engineering from the Kunsan National University in 2002 and 2004, respectively. He is currently a doctoral course student at the Kunsan National University, Korea. His research interests are digital image/video coding & processing, computer vision, object segmentation and tracking, and 2D/3D visual quality assessment and evaluation.



Jung-Sik Lee received the B.S. in Electronic Communications Engineering from the Hanyang University, Korea in 1983, and M.S., and Ph.D. degrees in Electrical Engineering from the Florida Institute of Technology in 1990 and 1996, respectively. He is currently a professor at the Department of Information Telecommunication Engineering of Kunsan National University, Korea. In 2004, he visited the University of South Florida, U.S.A. as a research professor. His specialties are digital communications, signal processing, neural networks applications.



Sang-Hyon Pak received the B.S. degrees in electronic engineering from the Chonbuk National University after the maritime telecommunication of the Kunsan National Fisheries College in 1993. He is currently an administrator in DICS Vision Co., Ltd., Korea. His research interests are an application and devices for a growth industry under digital image/video coding & processing, information theory, object segmentation and tracking.