Signalman Action Analysis for Container Crane Controlling

Bae, Suk-Tae[†]

ABSTRACT

Human action tracking plays an important place in human-computer-interaction, human action track-ing is a challenging task because of the exponentially increased computational complexity in terms of the degrees of freedom of the object and the severe image ambiguities incurred by frequent self-occlusions. In this paper, we will propose a novel method to track human action, in our technique, a dynamic background estimation algorithm will be applied firstly. Based on the estimated background, we then extract the human object from the video sequence, and the skeletonization method and Hough transform method will be used to detect the main structure of human body and each part rotation angle. The calculated rotation angles will be used to control a crane in the port, thus we can just control the container crane by using signalman body. And the experimental results can show that our proposed method can get a preferable result than the conventional methods such as: MIT, JPF or MFMC.

Key words: Signalman, Human Action, dynamic background estimation

1. INTRODUCTION

Human action tracking and analysis from videos has received a significant amount of attention in recent years driven by its wide applications such as human-computer-interaction, patient re-habilitation, biomechanics, human activity analy-sis, computer animation, etc. Human action track-ing is a challenging task because of the ex-ponentially increased computational complexity in terms of the degrees of freedom of the object and the severe image ambiguities incurred by frequent self-occlusions.

Many approaches have been studied to circum-vent the problems inherent in articulated object tracking. Most earlier efforts of articulated action tracking took advantage of 2-D or 3-D object models [1]. Drummond and Cipolla [2] presented

an algorithm which propagated statistics of proba-bility distributions through a kinematic chain to obtain maximum a posteriori estimates. Erol et al. reviewed the existing hand action estimation methods in [3]. A unified spatio-temporal articu-lated model was proposed by Lan and Huttenlocher [4]. Kalman filters have been employed by many researchers to combat occlusions in articulated ob-ject tracking [5,6]. In the context of multi-view body tracking, Bregler and Malk [7] proposed to track people with twists and exponential maps. Kehl et al. [8] presented an approach based on a volumetric reconstruction and a stochastic meta descent optimization.

In this paper, we present a novel approach for signalman action tracking from web-cam. A sig-nalman performs the works that a Crane driver loads or unloads containers in a ship correctly. The new approach will extract the object based on the estimated dynamic background. Then distinct from the existing approaches, we will extract the skel-etal shape which can be as a new describer of the object. The skeletal structure of the object can avoid the self-occlusions problem. And based on the Hough transform, we derive a novel stretched

※ Corresponding Author : Bae, Suk-Tae, Address :
(608-711) Department of Port & Logistics System,
Tongmyong University, Busan, Korea, TEL : +82-51-
629-3503, FAX : +82-51-629-3529, E-mail : stbae@tu.ac.kr
Receipt date : Sep. 23, 2009, Revision date : Nov. 3, 2009
Approval date : Nov. 17, 2009
[†] Department of Port & Logistics System, Tongmyong University, Korea

body detectors, the detector will detect the region that isn't devious but straight. The region such as forearm, post-brachium and legs can be detected. Signalman action is composed by motions of these parts. The rotated angle can reflect the motion of each part. Thus we will analyze the rotation of each part of skeletal shape and get the rotation angles which will be applied to control a virtual character motion.

The paper is organized as follows: Section 2 presents the dynamic background estimation method; Section 3 describes the objection detection and skeletonization part Section 4 will explain that how to build the stretched body detector and its' analysis and finally we provide the experimental results.

2. BACKGROUND ESTIMATION

In basic terms, we define the background as the stationary portion of a scene. Many applications simply require that there be introductory frames in the sequence which contain only background elements. If pure background frames are available, pixel-wise statistics can be computed directly. The more difficult case is computing the background model in sequences which always contain foreground elements.

Based on the definition of background, we consider a sequence of video frames as Fig 1 shows that. In the sequence, each pixel can be considered as an independent statistical process. We record the gray observations at each pixel over a sequence of frame.
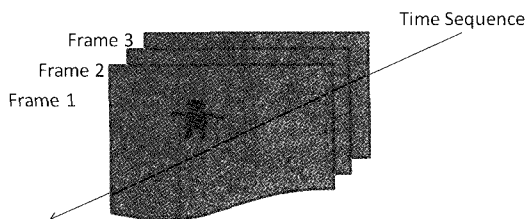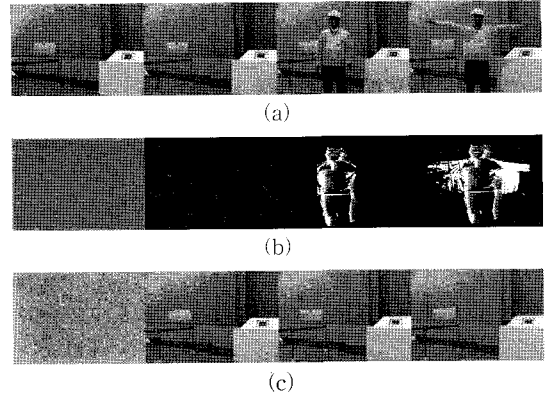

(a)

(b)

(c)

Fig. 2. Background Estimation Result: (a) Frame Sequence (b) S_h Result of the Sequence and (c) Background.

In the video, each frame will be recorded to sequence as a history frame as following equation shows that.

$$S_h = \sum_{i < TimeOut} | S_{h-1} - F_i |$$

(1)

$$B_h = \frac{B_{h-1} + F_i}{2}$$

(2)

Here, S_h is the h_{th} history frames in the video sequence, F_i is i_{th} the frame, and B_h is the background frame in h time. When the video with no object, S_h will be closed to 0 with time passed, in the meantime the background B_h will be more stationary. Thus, we can set a threshold value to judge the background. If S_h is smaller than the threshold value we can consider the B_h as background. As long as there is sufficient data representing the back-ground at any given pixel over the sequence, the background can be estimated from the foreground elements. The process result is as shown in Fig. 2.

3. SIGNALMAN BODY DETECTION AND SKELETONIZATION

In the following, with the estimated background, we will extract the object from video sequence by using the minus operation between video sequence and the background. As the following equation

Fig. 1. Graphical Model for a Video Frames

shows that:

$$O = | F_t - B_h |\qquad(3)$$

Here, O symbol for the Object part, F_t is the current frame, and B_h is the background frame at current case.

Then Gaussians filter will be used to remove the noise from the result. The average value can be considered as a threshold value to make the result binaryzation. Result is as shown is Fig. 3.

The skeleton is a very useful structure in image analysis. In the system, we will apply it to describe the structure of object. And for object skeletonization, the chamfer distance transform(CDT) is easily computed by a two-pass sequential algorithm [9]. Chamfer distances are a good approximation of the Euclidean distance. They are characterized by a local mask. Coefficients can be optimized according to a maximal error criterion [10,11]. For the case of rectangular grids, the pixel size is characterized by its Width W and its Height H which generally differ. As graphic figure shown in Fig. 4.
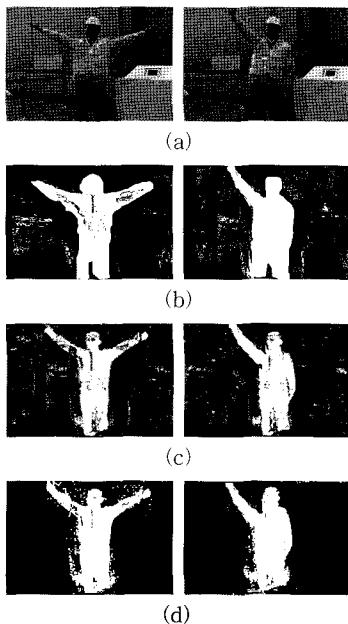

(a)

(b)

(c)

(d)

Fig. 3. Object Detection Result: (a) Test Frame; (b) Minus Operation Result; (c) Gaussians Filtering Result and (d) Object Binaryzation.

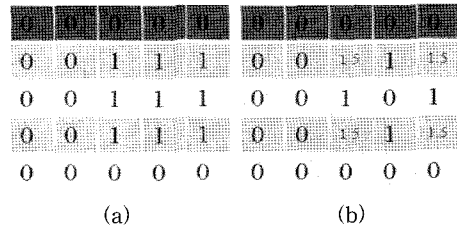|									
0	0	1	1	1	0	0	1	1	
0	0	1	1	1	0	0	1	0	1
0	0	1	1	1	0	0	1	1	
0	0	0	0	0	0	0	0	0	0

(a)   (b)

Fig. 4. A 3×3 Distance Operators on Rectangular Region: (a) Test Region and (b) Distances of Test Region

In the system, for detecting the structure of human body, we use a 3×3 local distance operators. And consider the input object with rectangular pixels. Then apply the distance transformation to the object region. CDT is computed by means of the sequential algorithm given in [12].

The Centre of Maximal Distance(CMD) are detected by a simple inspection of the neighboring pixels. We search for the redundant Moving Average(MA), since the objective is to determine the skeleton.

Let P be the current pixel, N(P) the 3×3 neighborhood of P and n_i a neighbor of P in N(P). Fig. 5 presents the configuration of the neighborhood N(P).

P is a CMD if it is maximal in $N(P)$. We call a local maximum, each point P which does not propagate the distance information to its neighbors, i.e if the distance disc of centre P is not completely contained in an other disc. This means:

$$\forall i \in \{1,...,8\}\qquad P + D_i > n_i\qquad(4)$$

With: $D_i = D_{10}$ if n_i is an horizontal neighbor of P
$D_i = D_{01}$ if n_i is a vertical neighbor of P
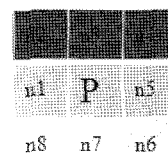$D_i = D_{11}$ if n_i is a diagonal neighbor of P

Fig. 5. The Neighborhood N(P) of Rectangular Pixel P
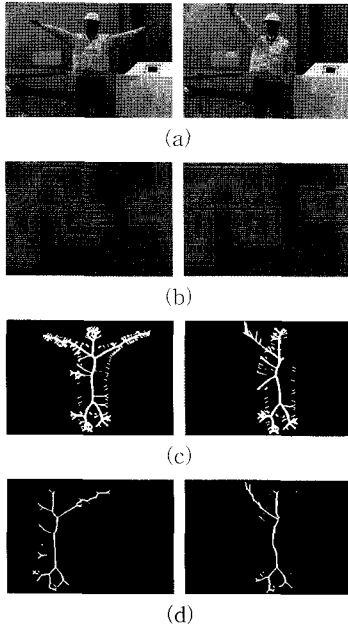
(a)

(b)

(c)

(d)

Fig. 6. Result of Skeletonization: (a) Test Object Frame; (b) Distance Result of Object (Each Value has be magnified 10 times); (c) Medial Axis Searching Result and (d) Skeletonization Result of Object

Local distances D_i are presented in Fig. 4.

Hence, P is a CMD. Each CMD pixel is marked in a list, noticed LSP (Label Symbol Path), which will contain all skeletal pixels. The set of CMD is centered in the pattern and suffices for its reconstruction. But it is generally disconnected and then does not preserve the pattern topology. A connection step is the necessary. For this purpose we search for the connection paths between the different components of the medial axis.

4. SIGNALMAN BODY PARTS SEGMENTATION BASED ON HOUGH TRANSFORM AND ACTION ANALYSIS

The Hough transform [13] is a technique which is most commonly used to extract features of a particular shape within an image. The classical transform is used as a method to detect lines and extended to recognize regular curves such as lines, circles, ellipses and arbitrary shapes that are repre-

sented in some parametric form by Richard Duda and peter Hart [14]. The purpose of the transform is to determine which lines fit most closely to the data in the image. In order to determine the line, it is necessary to convert the points on a *(x, y)* plane to a *(r, Θ)* plane that consists of the length and the angle from the origin to the line to be made up two points in image like shown in Fig. 7 and Fig. 8. For all data points, a number of potential lines described here as solid lines are plotted going through them. For all potential lines, a line described here as solid lines are plotted going through them. For all potential lines, a line described as a dashed line is plotted which is perpendicular to them and which is crossed the origin. The length and angle of each dashed line corresponds to a sinusoidal curve in the *(r, Θ)* plane. The tables in Figure 7 show the results for each diagram. A graph of length and angle, sometimes referred as a Hough space, is created as shown in Fig 8. The

Angle(°)	Dist
0	40
30	68.5
60	81.6
90	70.5
120	41.2
150	0.5

Angle(°)	Dist
1	57.2
30	79.4
60	80.6
90	60
120	23.6
150	-18.9

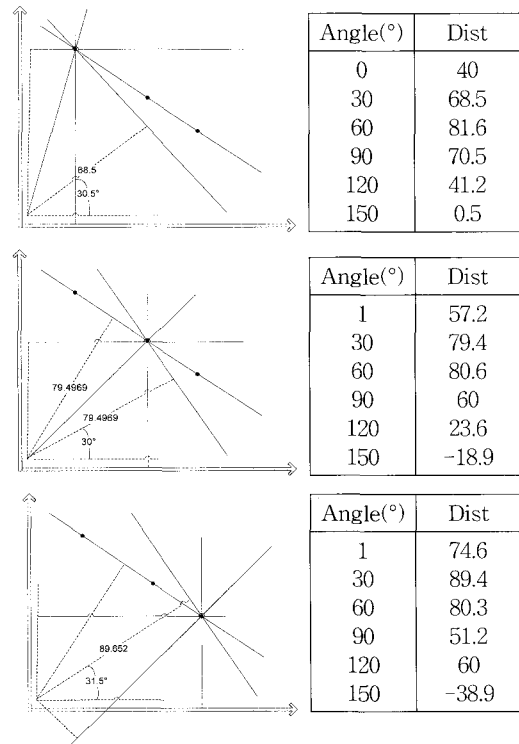| Angle(°)	Dist
1	74.6
30	89.4
60	80.3
90	51.2
120	60
150	-38.9

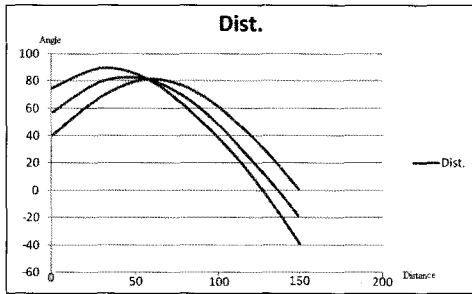Fig. 7. Potential Line Diagrams for Three Points

Fig. 8. Sinusoidal Curves on the Hough Space

point where the sinusoidal curves intersect gives a distance and angle. This distance and angle indicate the most fitted line which bisects the three points. The major advantage of the Hough transform technique is that it is tolerant of gaps in feature boundary descriptions and is relatively robust against image noise and reduces the dimension of candidate feature points.

As seen from the result of object skeletonization, parts of signalman body such as: arms, legs or head. the skeletonization of arms (or leg) can be composed of two lines, one is forearm (lead leg) and the other is postbrachium (draw leg). So the Hough transform can be adapted to detect these lines information from the skeletonization result. The connection point of two lines is the body joint. Rotation angle of two connected lines can describe the body action information.

As shown in Fig. 9, $\theta_1 \sim \theta_4$ can describe the motion of arms and $\theta_5 \sim \theta_7$ can describe the motion of legs. Thus a human motion can be composed of these angles. We can set a vector as shows following:

$$V = \{\theta_1, \theta_2, \theta_3, ... \theta_7\}^T \qquad (5)$$

Here, we named the V as the Motion Vector, and range of each element can be ranged as following:

$$\theta_i (\forall i = 1,...7) > 0$$
$$\theta_i (\forall i = 1,...7) \leq 180 \qquad (6)$$

For calculating these angles, we can set a coordinates, each line will intersect either Y axis or X axis with a or β as shown on Fig. 10. On the basis
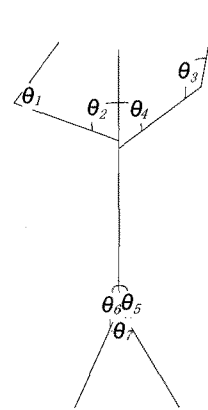

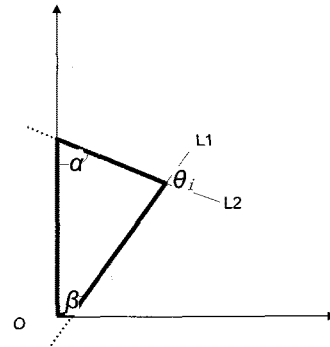Fig. 9. Graphic Diagram of Human Body Gesture

Fig. 10. Graphic Diagram of Human Body Gesture

of Hough transform the angle a can be calculated as following:

$$\alpha = \pi / 2 - \arctan(Y_i - b) / X_i \qquad (7)$$

$$\beta = \pi / 2 - \arctan(Y_j - b) / X_j \qquad (8)$$

$$\begin{aligned}\theta_i &= 2\pi - \frac{\pi}{2} - \alpha - \beta \\ &= \frac{3}{2}\pi - (\frac{1}{2}\pi - \arctan(Y_i - b) / X_i) - (\frac{1}{2}\pi - \arctan(Y_j - b) / X_j) \\ &= \frac{1}{2}\pi + \arctan(Y_i - b) / X_i + \arctan(Y_j - b) / X_j\end{aligned} \qquad (9)$$

Here X_i, Y_i are points on the first line and X_j, Y_j are on the other line which intersects with the current line.

We applied the calculated angles to control the crane in port. If the port worker moves his arm up, that means to control the crane up or if the worker moves the arm down, the crane can move down. So by applying the system to the port crane
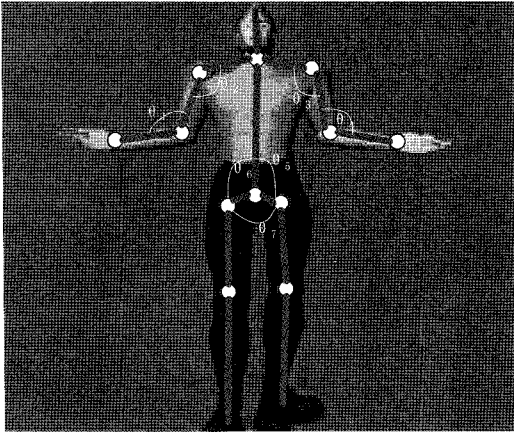
Fig. 11. A Model of Virtual Character and its' Joints
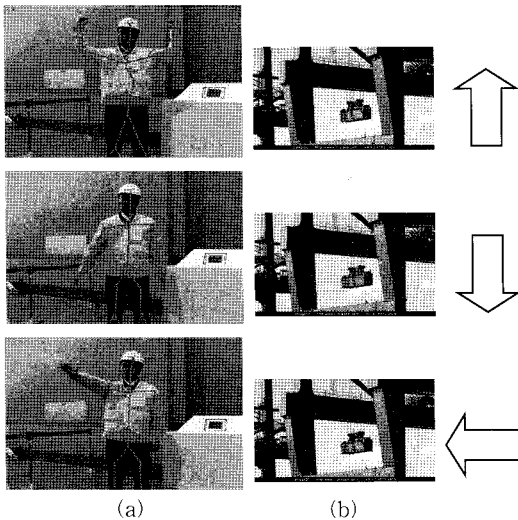

(a)   (b)

Fig. 12. Virtual Character Control: (a) Real Motion
and (b) Crane Control Direction

controlling, human can just control the crane using body gesture. And for recognizing human action, we separate human body as the Fig. 11 parts, the rotation angles of each joints are as shown following.

Result of virtual character control can be seen as Fig. 12 shows that.

5. EXPERIMENTING RESULT

In this section, we proposed some of the experimental results. The action tracking perform-

ance of the proposed methods were compared both qualitatively with the multiple independent trackers (MIT) [15], Joint Particle Filter (JPF) [16] and mean field Monte Carlo (MFMC) [17]. All experiments were performed using VC++ on a 3.2-GHz Pentium IV PC without code optimization.

The video MAN contains a man moving his arms. It has 455 frames and was captured by 25 fps with a resolution of 320 × 240 pixels. The complex background and fast motion make it challenging for articulated motion tracking. The proposed method achieved robust tracking results even when the arms moved rapidly as shown in Fig. 13.

The video WALKING Signalman has a man who is walking. It was captured by 20 fps with a resolution of 160 × 120 pixels and has 456 frames as shown in Fig. 14. We use this video to demonstrate the ability of the proposed algorithm to track legs. The identical color of the torso and arms, and the frequent severe self-occlusions among limbs make it difficult for articulated motion tracking, when somebody move his arm over his body, the arm will obdurate the body part, we consider this case as the self-occlusions.

To compare the proposed method against the state of the art, we independently implemented the MIT, JPF and MFMC. In Fig. 15, we illustrate the RMSE (root-mean-square error) sample result
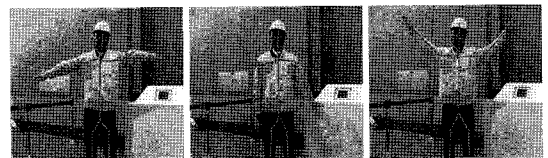

Fig. 13. Tracking Results of Proposed Method on
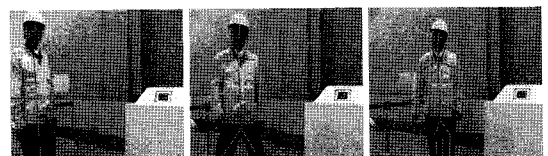the Sequence Signalman

Fig. 14. Tracking Result of the Proposed Method
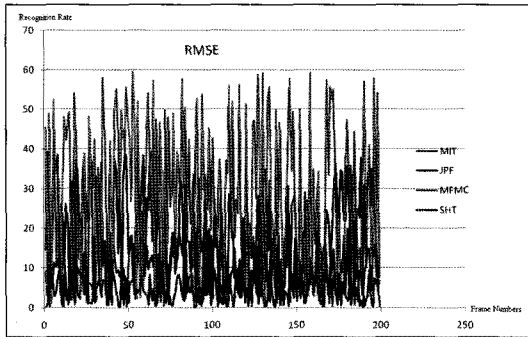on the Sequence WALKING Signalman

Fig. 15. Comparison of RMSE using MIT, JPF and MFMC and the Proposed Method (Skeletonization & Hough Transform Based Method) on the Signalman Action Sequence

frames. MIT kept the connection among parts. However, it could not produce satisfactory results when self-occlusion presented. Compared with the MIT, the proposed method improved the performance in that the connections among parts were preserved well. This is because the proposed method uses the skeletonization model to represent the human body and each part. Thus for each part such as arm, its' skeletonization model cannot cover the torso, so for solving the self-occlusion problem, the proposed method can gave a better results.

6. CONCLUSIONS

In this paper, we have presented a novel algorithm for tracking human action, which is based on the detecting the skeletonization of signalman body. The proposed algorithm firstly separate the background from the video sequence, a based on the separated background, the method will detect the object by using the difference operation. On the next step, we will extract the skeletonization of the detected object, and apply the Hough transform to detect line information which composes the object structure. The rotation of each part can also be calculated based on Hough theory. Finally, we will apply the calculated angles to control the container crane in port. Through the simulation, we can see

that the proposed Skeletonization & Hough Transform Based Method can get a preferable result than the conventional methods such as: MIT, JPF or MFMC.

REFERENCES

[1] S. Ju, M. J. Black, and Y. Yacoob, "Cardboard people: A parameterized model of articulated image motion," presented at the Int. Conf. Automatic Face and Gesture Recognition, 1996.

[2] T. Drummond and R. Cipolla, "Real-time Tracking of Highly Articulated Structures in the Presence of Noisy Measurements," Presented at the Int. Conf. Computer Vision, Vancouver, BC, Canada, 2001.

[3] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "A Review on Vision-based Full DOF Hand Motion Estimation," Presented at the IEEE Conf. Computer Vision and Pattern Recognition, 2005.

[4] X. Lan and D. Huttenlocher, "A Unified Spatio-temporal Articulated Model fro Tracking," Presented at the IEEE Conf. Computer Vision and Pattern Recognition, 2004.

[5] K. Nickels and S. Hutchinson, "Model-based Tracking of Complex Articulated Objects," IEEE Trans. Robot. Autom., Vol.17, No.1, pp. 28-36, Jan. 2001.

[6] N. Jojic, M. Turk, and T. Huang, "Tracking Self-occluding Articulated Objects in Dense Disparity Maps," Presented at the Int. Conf. Computer Vision, Corfu, Greece, 1999.

[7] C. Bregler and J. Malik, "Tracking People with Twists and Exponential Maps," Presented at the IEEE Conf. computer Vision and Pattern Recognition, Washington, DC, 1998.

[8] R. Kehl, M. Bray, and L. V. Gool, "Full Body Tracking from Multiple Views using

Stochastic Sampling," in IEEE Conf. Computer Vision and Pattern Recognition, San Diego, CA, 2005.

[9] Rosenfeld, A and J. Pflatz. "Sequential Operations in Digital Picture Processing". J. ACM, 13. pp. 471-494, 1966.

[10] D. Copuin, Ph. Bolon. "Discrete Distance Operator on Rectangular Grids," Pattern Recognition Letters, 16, pp. 911-923, 1995.

[11] Borgefors, G. "Distance Transformations in Digital Images, " Computer Vision, Graphics and Image Processing, Vol.34, pp. 344-371, 1986.

[12] Rosenfeld, A and J. Pflatz. "Sequential Operations in Digital Picture Processing," J. ACM, 13, pp. 471-494, 1966.

[13] P.V.C Hough, "A method and means for rec- ognizing complex patterns," U. S Patent 3069564, 1965.

[14] R. O. Duda, P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," Comm. ACM, Vol.15, pp. 11-15, 1972

[15] M. Isard and A. Blake, "Condensation- Conditional Density Propagation for Visual

Tracking," Int. J. Computer, Vol.29, No.1, pp. 5-28, 1998.

[16] M. Isard and J. MacCormick, "BraMBLe: A Bayesian Multiple-blob Tracker," In Proc. Int. Conf. Computer Vision, Vancouver, BC, Canada, Jul. 2001, Vol.2, pp. 34-41.

[17] Y. Wu, G. Hua, and T. Yu, "Tracking Articulated Body by Dynamic Markov Network," Presented at the IEEE Conf. Computer Vision and Pattern Recognition, Washington. DC, 2004.

Bae, Suk-Tae

1988. 2. Mechanical Engineering, Dong-A University (B.S.)
1991. 2. Mechanical Engineering, Dong-A University (M.S.)
1995. 2. Mechanical Engineering, Dong-A University (Ph.D.)
1995. 9~present: Prof. Dept. of Port & Logistics System, Tongmyong University
Research interests : Port Logistics Auto-Equipments, Container Terminal Operation System, e-SCM