PERGAMON

Pattern Recognition 34 (2001) 1951–1962

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Segmentation based on fusion of range and intensity images using robust trimmed methods

In Su Chang, Rae-Hong Park*

Department of Electronic Engineering, Sogang University, C.P.O. Box 1142, Seoul 100-611, South Korea

Received 3 August 1999; accepted 24 August 2000

Abstract

This paper proposes a segmentation algorithm based on fusion of range and intensity images using robust trimmed methods. Based on the Bayesian theory, a priori knowledge is represented using the Markov random field (MRF). A maximum a posteriori (MAP) estimator is constructed using the edge features extracted from both range and intensity images. Objects are represented by a number of local planar surfaces in range images, and the parametric space for surface representation is constructed with the surface parameters estimated pixel-by-pixel based on the least trimmed squares (LTS) method. Whereas in intensity images, the α-trimmed variance is adopted as the feature for edge extraction. A final edge map is obtained by the MAP estimator that is constructed using the likelihood functions based on the edge information obtained from range and intensity images. Finally, an image is segmented using the fused edge map. Computer simulation results show that our new segmentation algorithm effectively segments test images, independent of shadow, noise, and lighting environment. © 2001 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Image segmentation; Edge detection; Surface parameter estimation; Robust estimation; Markov random field; MAP estimation; α-trimmed method; Least trimmed squares (LTS) method

1. Introduction

The computer vision system provides the visual ability for automatic understanding of the three-dimensional (3D) world. The conventional computer vision systems with intensity images have found difficulties in extracting reliable features for effective image analysis, because of varying statistical characteristics such as sensitivity to noise, reflectance property of object surfaces, lighting environment, and so on. In range images, the pixel value represents the shortest distance from the sensed focal plane to the point on the surface of a 3D object. The range image directly provides explicit 3D geometric information about the objects and is independent of the intensity of a light source, illuminating conditions of the environment, the reflectance of object surfaces, and the shadow of the objects. Thus, 3D modeling and object recognition using range images can be achieved with less high-level processing steps [1].

Noise in range image acquisition is modeled as a Gaussian and impulse noise mixture. Since range and intensity images are represented in different coordinates, coordinates transformation is needed. Because of coordinates transformation, range data in occluded regions are sometimes absent and the missing range data are regarded as impulse noise, which leads to undesirable error in 3D modeling or object recognition.

The edges, important information in image analysis and computer vision, can be classified as step edges, roof edges, shadow edges, albedo edges, etc. The shadow edges and albedo edges in intensity images should not be detected because detection of them yields poor segmentation and performance degradation in related applications. The roof edge in range images detected by second-order differentiation is very sensitive to Gaussian

* Corresponding author. Tel.: +82-2-705-8463; fax: +82-2-705-8629.
E-mail address: rhpark@ccs.sogang.ac.kr (R.-H. Park).

0031-3203/01/$20.00 © 2001 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.
PII: S0031-3203(00)00124-2

or impulse noise. The range image segmentation methods [2,3] based on robust statistics were proposed, but they required a high computational load. For reliable and effective image analysis, fusion of range and intensity images is required [4–7].

Fusion of information obtained from different sensors can be achieved heuristically or by the Bayesian approach. In earlier fusion approaches, the heuristic methods used a mechanism as employed in the Gil et al. method [5] which was based on the pixel-by-pixel logical AND operations applied to the information obtained from different measurement data. Heuristic fusion algorithms lack robustness and generality because of the difficulty in deducing general rules to incorporate sensor uncertainties, whereas the Bayesian approaches use the Bayes rules to incorporate sensor uncertainties and the a priori information. A Markov random field (MRF) was used to model the a priori information of a digital image [6,7]. Zhang and Wallace [6] and Nadabar and Jain [7] proposed the Bayesian approaches, in which a priori modeling of sensor uncertainties and the a priori information were combined.

The rest of the paper is organized as follows. Section 2 reviews the MRF that models the a priori information employed in the Bayesian estimator. Section 3 presents our new segmentation algorithm based on fusion of range and intensity images. Section 4 shows the experimental results to illustrate the effectiveness of our new segmentation algorithm. Finally, conclusions are given in Section 5.

2. MRF

The MRF models have been successfully applied to image modeling in various applications [4–10]. This section briefly describes the background related to the MRF model.

2.1. MRF and the Gibbs distribution

The label image is denoted as $F = \{f_i | i \in S\}$, where $S = \{i | 1 \leqslant i \leqslant m\}$ is the set of sites, with i representing the site (pixel) in the image lattice and m denoting the total number of pixels. The image F can be a labeled map or a binary edge map. Each pixel f_i takes a discrete value from the label set $L = \{l | 1 \leqslant l \leqslant M\}$, where M denotes the total number of labels. The spatial relationship of the sites, each of which is indexed by a single number in S, is specified by a neighborhood system $N = \{N_i | i \in S\}$, where N_i signifies the set consisting of neighboring sites of pixel i. A single site or a set of neighboring sites forms a clique denoted by c. The random field which satisfies the following conditions with the neighborhood system N and the set S is defined as an MRF:

$$\begin{aligned} &(1)\,\text{Positivity} \quad P(f) > 0, \forall f \in F, \\ &(2)\,\text{Markovianity} \quad P(f_i | f_{S-\{i\}}) = P(f_i | f_{N_i}), \end{aligned} \tag{1}$$

where $P(\cdot)$ denotes the probability function, $S - \{i\}$ represents the set formulated by sites except for site i, and f_{N_i} signifies the label set of the neighborhood system at site i. Thus local interaction between neighboring pixels is described by the neighborhood system.

According to the Hammersley–Clifford theorem, the MRF follows the Gibbs probability distribution [10]. The joint probability of the MRF can be obtained from the energy function of the Gibbs distribution.

The Gibbs distribution $P(f)$ is defined by the energy function $U(f)$:

$$P(f) = \frac{1}{Z} \exp\left\{ -\frac{1}{T} U(f) \right\}, \tag{2}$$

where T represents the temperature constant, Z is a normalization constant given by

$$Z = \sum_{f \in F} \exp\left\{ -\frac{1}{T} U(f) \right\} \tag{3}$$

with F denoting the label set, and $U(f)$ is the energy function defined by the sum of the potential function $V_c(f)$ of the clique c:

$$U(f) = \sum_{c \in C} V_c(f) \tag{4}$$

with C representing the clique set consisting of all possible cliques.

The MAP estimator with the Gibbs distribution is optimized by estimating the conditional probability function based on the MRF modeling of the image.

2.2. Optimization of the MAP estimator

An estimator is required to be efficient (deterministic and preferably should make the maximum improvement at each iteration step), predictable (results depending on the inputs and the a priori distribution but not on other parameters), and robust. The highest confidence first (HCF) method [10] satisfies these requirements. It is not a stochastic method like simulated annealing, but a deterministic one such as iterative conditional modes (ICM) [9] estimation. Both HCF and ICM approaches require some initial configuration. At each iteration through the image sites, the state of each site is either changed to the state that yields the maximal decrease of the energy function, or left unchanged otherwise.

In the HCF algorithm, all sites are initially labeled as uncommitted, instead of starting with some specific or arbitrary labels as in most conventional optimization methods. Uncommitted sites do not participate in computation of the energy of the field, nor, therefore, do they actively influence the commitments of their neighbors. However, an uncommitted site always takes into account the states of the active neighbors when making a commitment. For each site, a stability measure is computed. The less stable the state of the site, the more confidence we have in changing its label. In each iteration, the site with the minimum stability measure is selected, and this label is changed to the one that yields the lowest energy, which in turn makes the stability of neighboring sites change. A commitment to assume a label is not a commitment to a fixed label: the label of a committed site will be altered if its neighbors yield a large variance. The process is repeated until no changes in the labeling process result in decrease of the energy function, at which point the energy is the minimum and the algorithm terminates.

The a posteriori knowledge for the labeling is represented by a Gibbs distribution [10,11]. The a posteriori probability of a labeling f in the clique set C is expressed as

$$P(f|O) = \frac{1}{Z}\exp\left\{-\frac{1}{T}\left(\sum_{c\in C} V_c(f) - T\sum_{i\in S}\ln[P(O_i|f_i)]\right)\right\}, \tag{5}$$

where O denotes the observation which represents the pixel value of each image and $V_c(f)$ is the potential function which guarantees the continuity of the neighborhood system configuration. The subscript i denotes the quantity defined at site i.

Initially, the HCF sets the labels at all sites to the null (uncommitted) label ("l_0"). The basic idea of the HCF is to construct the configuration that yields a local minimum energy measure $U(f)$.

The a posteriori energy function $E_i(f)$ at site i is defined as

$$E_i(f) = \sum_{c\in C} V'_c(l) - T\ln[P(O_i|f_i)], \tag{6}$$

where $V'_c(l)$ is given by

$$V'_c(l) = \begin{cases} 0 & \text{for } l = f \text{ or } l = l_0, \\ V_c(l) & \text{otherwise.} \end{cases} \tag{7}$$

The a posteriori energy corresponds to the confidence measure of the label that yields the current configuration of the neighborhood system. The null-labeled site does not have active influence on the current configuration of the neighborhood system, however it is considered in labeling of its neighboring sites.

To construct the minimum energy configuration, the label of the least stable site is first updated. The stability measure $G_i(f)$ at site i is defined as follows:

$$G_i(f) = \begin{cases} \min\limits_{k\in L, k\neq f_i} \Delta E_i(k,f_i) & \text{if } f_i \in L, \\ -\min\limits_{k\in L, k\neq j} \Delta E_i(k,j) & \text{if } f_i = l_0, \end{cases} \tag{8}$$

where $\Delta E_i(a,b) = E_i(a) - E_i(b)$ and $E_i(j) = \min_{k\in L} E_i(k)$ with $j \in L$.

This stability measure is a combined measure of the observations and the a priori knowledge about the current state. A negative value of G indicates that a more stable configuration can be obtained by a label update. In Eq. (8), the null-labeled site has the negative G. The magnitude of G corresponds to the incremental gain or loss of the energy due to a label change: the large negative value of G represents the high possibility of a label change. The HCF can be implemented serially with a heap, maintaining the visiting order of the construction according to the values of G, the top of which denotes the site with the smallest G value. The HCF updates the least stable (the minimum G-valued) site i, adjusts the heap structure, and updates the stability of the neighborhood system of site i. The iterative procedure is terminated when G values at all sites are positive (all sites are stabilized). The label update is to change the label such that

$$l = \begin{cases} \min\limits_{l} \Delta E_i(l,f_i) & \text{if } f_i \in L, \\ \min\limits_{l} E_i(l) & \text{if } f_i = l_0. \end{cases} \tag{9}$$

The stability measure of the updated site becomes positive, and the label update has influence on the configuration of the neighborhood system.

3. Segmentation based on fusion of range and intensity images

This section describes extraction of the edge information based on the α-trimmed variance and surface parameters extracted from intensity and range images, respectively. Then the formulation of likelihood functions based on the edge information and MAP estimation for image segmentation are presented.

3.1. Feature extraction

3.1.1. Feature extraction in intensity images

The intensity image provides the edge information of 3D objects, such as the discontinuities of the objects (step edges) and surfaces (roof edges). Also the shadow and reflectance discontinuities of the objects are represented by the shadow and albedo edges. The step and roof edges are important features in 3D object recognition, whereas the shadow and albedo edges may impair the recognition performance. This paper adopts the binary label set L consisting of two labels ("edge" and "non-edge").

To detect the edge features, the gradient method has been commonly used [12]. But it is difficult to detect a reliable edge map by the gradient method in noisy images because of the sensitivity of the gradient operation to noise. In this paper, the intensity variance at each pixel is used to detect discontinuous regions. For Gaussian noise cases, the intensity variance is large near discontinuous regions whereas small in homogeneous regions if the image is corrupted by Gaussian noise. But for impulse noise cases, the variance is large, independent of the discontinuity. The α-trimmed variance based on the α trimming method [13] is robust to impulse noise, in which outliers are excluded in variance computation:

$$\sigma_\alpha^2 = \frac{1}{(1-2\alpha)n} \sum_{i=\alpha n}^{(1-\alpha)n} (I - \bar{I})^2_{i:n}, \tag{10}$$

where n represents the total number of pixels considered and α denotes a constant value between 0 and 0.5, specifying the ratio of the number of data used in variance computation to the total number of data n. The α-trimmed mean $\bar{I}$ is defined as

$$\bar{I} = \frac{1}{(1-2\alpha)n} \sum_{i=\alpha n}^{(1-\alpha)n} I_{i:n}, \tag{11}$$

where $I_{1:n} \leqslant I_{2:n} \leqslant \cdots \leqslant I_{i:n}$ are the ordered intensities and $(I-\bar{I})^2_{1:n} \leqslant (I-\bar{I})^2_{2:n} \leqslant \cdots \leqslant (I-\bar{I})^2_{i:n}$ are the ordered squared errors. The small α yields the α-trimmed variance sensitive to noise, whereas the large α gives an inaccurate variance estimate due to the small number of data involved in variance computation. To extract the edge information from intensity images, the α-trimmed variance is used in this paper.

3.1.2. Feature extraction in range images

Roof edges detected using the Laplacian operation are not accurate for noisy range images. In this paper, objects are assumed to consist of a number of planar patches, where three parameters a, b, and c are required to represent each planar patch $z(x, y)$:

$$z = ax + by + c. \tag{12}$$

For faithful representation of a curved surface, more than one planar patch is used, where the accuracy of representation and the number of planar surfaces used depend on the tolerable approximation error. The surface parameters approximating the surface patch provide the geometric information of the 3D object considered. Detection of the discontinuity in surface parameter values yields the roof edges [14]. In this paper, the least trimmed squares (LTS) method [15] based on the robust regression is used to estimate the surface parameters.

The LTS method is more efficient than the least median squares (LMedS) method. In the former, the parameters are estimated by minimizing the non-linear expression:

$$\min \sum_{i=1}^{h} (r^2)_{i:n}, \tag{13}$$

where n denotes the number of data, h represents the number of data used in parameter estimation, and $(r)^2_{1:n} \leqslant (r)^2_{2:n} \leqslant \cdots \leqslant (r)^2_{n:n}$ are the ordered squared errors. Note that the residual error r is defined by the difference of the original data and the one reconstructed by the estimated parameters. In statistics, data that distort the estimate are referred to as outliers. Robust statistics techniques have been used for parameter estimation because of their robustness to outliers. As a performance measure of an estimator, the breakdown point is used, which is defined by the percentage of tolerable outliers. The desirable robustness property is achieved when h is approximately equal to $n/2$, in which the breakdown point attains 0.5.

The planar surface parameters, with which the objective function in Eq. (13) is minimized, are estimated pixel by pixel in an estimation window. Because the LTS method with a large estimation window regards the corner points of the object as noise spikes, it is likely to lose them. In surface parameter estimation, we experimentally set the window size to 9×9, which yields the proper estimation performance and edge-preservation. Estimation window size is selected experimentally.

Our new algorithm extracts edge features from the range image, based on the estimated surface parameters. If the window contains a complex region consisting of a number of planar surfaces, the parameters of the largest surface are obtained. The estimation error computed by the estimated parameters is large for other small surfaces, thus the approximation error is large for such estimation windows containing complex structures. Edge detection in range images is performed by selecting the pixels yielding large estimation error, where the squared estimation error at pixel (x, y) is defined as

$$\varepsilon^2(x, y) = \sum_{i=-4}^{4} \sum_{j=-4}^{4} \{z(x+i, y+j) - \hat{z}(x+i, y+j)\}^2 \tag{14}$$

with z and $\hat{z}$ denoting the depth and the reconstructed depth, respectively, and a 9×9 window is employed.

3.2. Our new segmentation algorithm

3.2.1. Likelihood functions

It is important to define the effective likelihood function for the MAP estimation using the MRF modeling. In this paper, the likelihood function is defined based on edge features extracted from intensity and range images. The label site set consists of "edge" and "non-edge" labels, where "0" represents the "non-edge" label and "1" signifies the "edge" label.

The likelihood functions of the intensity image are defined using the α-trimmed variance. The presence (absence) of the edges in the intensity image yields a large (small) α-trimmed variance, so the likelihood functions are defined as

$$P(O_I|f=0) = \frac{1}{Z_1}\exp(-\sigma_{O_I}^2), \quad (15)$$

$$P(O_I|f=1) = \frac{1}{Z_2}\exp\left(-\frac{1}{\sigma_{O_I}^2}\right), \quad (16)$$

where Z_1 and Z_2 denote the normalization constants, and O_I signifies the observation of the intensity image.

The likelihood functions of the range image are defined based on the surface parameter values estimated in the local window. Since our segmentation method is based on the edge likelihood, the images are segmented even if the largest surface is not extracted from the range images. The estimation error is large in edge regions between different surface patches, whereas it is small in homogeneous regions, so the likelihood functions are defined as

$$P(O_R|f=0) = \frac{1}{Z_3}\exp(-\varepsilon^2), \quad (17)$$

$$P(O_R|f=1) = \frac{1}{Z_4}\exp\left(-\frac{1}{\varepsilon^2}\right), \quad (18)$$

where Z_3 and Z_4 represent the normalization constants, and O_R signifies the observation of the range image.

In this paper, shadows and albedo edges in intensity images are neglected because they do not directly give geometric information. If so, the intensity and range data are assumed to be statistically independent, which makes the fused likelihood function $P(O|f)$ be expressed as the product of likelihood functions for intensity and range images:

$$P(O|f) = P(O_I|f)\cdot P(O_R|f), \quad (19)$$

where O denotes both observations from range and intensity images. By applying the HCF MAP estimation in Eqs. (6)–(19), the final edge map obtained by fusing range and intensity images is optimized.

3.2.2. Clique potential function

The potential function guarantees the continuity of the spatial relationship of edge segments. The optimum potential function is experimentally selected in most conventional methods, whereas in this paper the clique potential function defined by

$$V_c = -\lambda_0(1-l_{ij})\|\mathbf{a}_i - \mathbf{a}_j\|^2 - \lambda_1 l_{ij} \quad (20)$$

is employed, where λ_0 and λ_1 are weighting constants, $l_{ij}=1$ for the edge clique defined between pixels i and j, 0 for the non-edge clique, c represents the second-order clique set for $i,j \in c$, and $\mathbf{a}_i$ and $\mathbf{a}_j$ signify the surface parameter vectors extracted at sites i and j, respectively. The potential function is decreased by the amount of $\lambda_0\|\mathbf{a}_i - \mathbf{a}_j\|^2$ for the non-edge clique and by the amount

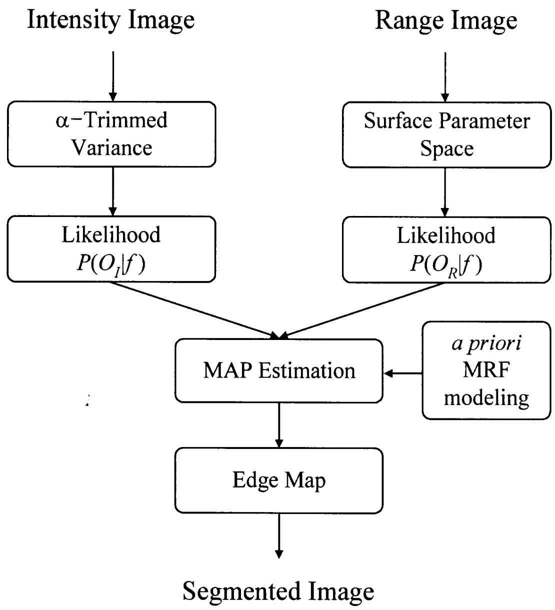

Fig. 1. Flowchart of our new segmentation algorithm based on fusion of range and intensity images.

of λ_1 for the edge clique. The weight constant λ_0 controls the smoothness. For example, large λ_0 smoothes adjacent regions and merges them into a single region, whereas small λ_0 splits a single homogeneous region into multiple regions. λ_1 is related to the a priori information such as the ratio of the numbers of edge and non-edge cliques.

Fig. 1 shows the flowchart of our new algorithm. The fused likelihood function $P(O|f)$ is defined by the likelihood functions $P(O_I|f)$ and $P(O_R|f)$ obtained from intensity and range images, respectively. The final edge map is constructed from the fused likelihood function using the MAP estimation with MRF modeling. Image segmentation is performed with the final edge map.

4. Experimental results

4.1. Test image sets

Fig. 2 shows two sets of test images to show the performance of our new segmentation algorithm. These images are 240 × 240 registered images captured in the Pattern Recognition and Image Processing (PRIP) Lab. at Michigan State University. By a camera calibration algorithm, the range images from a triangulation sensor were registered with the intensity image obtained from a separate CCD camera [7].

The original range and intensity images are dense, except for the shadow regions and for small holes due to quantization and registration error. Because the

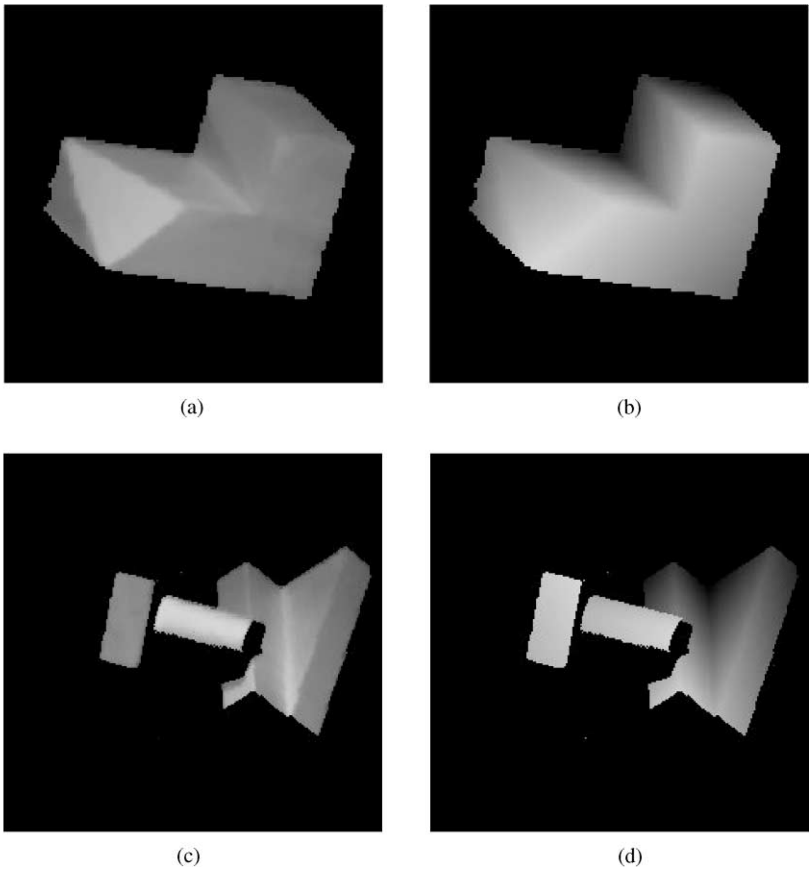

Fig. 2. Intensity and range images used in experiments: (a) intensity Image A, (b) range Image A, (c) intensity Image B, (d) range Image B.

conventional methods are not robust to these holes (impulsive noises), the holes are eliminated by preprocessing: 3×3 median filtering for the intensity image and by the reconstruction method using the estimated surface parameters for the range image.

4.2. Edge detection

To compare our new algorithm with the conventional ones (Gil et al.'s method [5], Zhang and Wallace's method [6], and Nadabar and Jain's method [7]), Gaussian noise ($\sigma = 3.0$) and impulse noise (5%) are added to the test image sets. Three-by-three median filtering is applied as a preprocessing to noisy images because the conventional methods are very sensitive to impulse noise.

Fig. 3 shows the likelihood functions for Image A magnified by a factor of 200, where edges are represented as the dark regions (Figs. 3(a), (c), and (e)) and bright regions (Figs. 3(b), (d), and (f)), respectively. The α-trimmed variance is calculated in the intensity image using five data values in a 3×3 window ($\alpha = 0.22$), whereas the surface parameters are estimated in the range image with a 9×9 window. Note that the fused likelihood function in Fig. 3(e) (Fig. 3(f)) shows a desirable combination of Figs. 3(a) and (c) (Figs. 3(b) and (d)), in the sense that the edge segments not shown in Fig. 3(a) (Fig. 3(b)) are detected in Fig. 3(c) (Fig. 3(d)), resulting in final edge segments in Fig. 3(e) (Fig. 3(f)).

Figs. 4 and 5 show the performance comparison of the conventional and our new segmentation methods. Figs. 4(a) and 5(a) show the edge maps obtained by Gil et al.'s method. The edge in the intensity image is detected by the Kirsch operator, whereas the edge in the range image is extracted by detecting the discontinuities of the normal vector's direction, where the normal vector represents the direction normal to the surface reconstructed by the estimated surface parameters. Since the range edge is very sensitive to noise, some roof edges may not be detected and the noisy false edges may be detected. Figs. 4(b) and 5(b) show the edge maps detected by the Zhang and Wallace method, whereas Figs. 4(c) and 5(c) show the edge maps generated by the Nadabar and Jain method. The performance of the conventional methods is some-

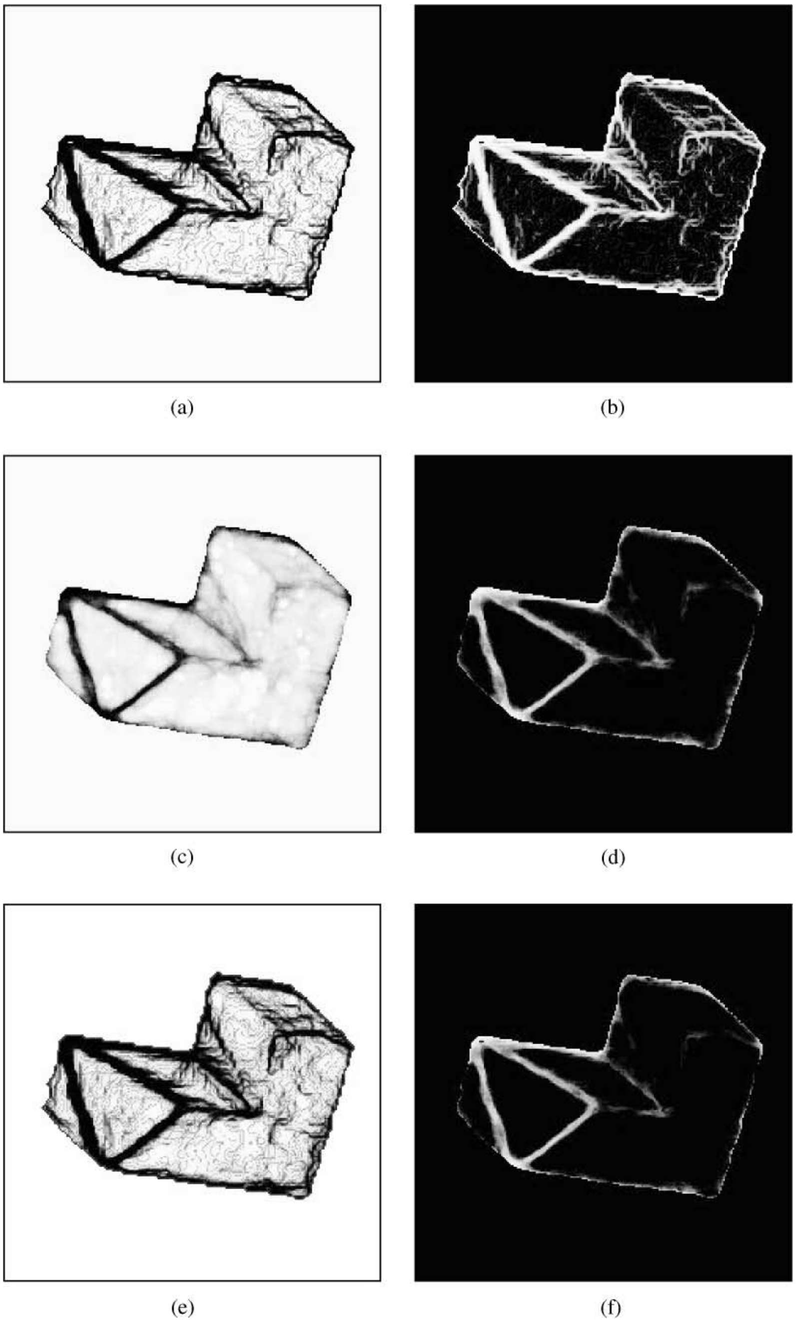

Fig. 3. Likelihood functions for Image A with Gaussian noise ($\sigma = 3.0$) and impulse noise (5%): (a) $P(O_I|f=0)$ (intensity image), (b) $P(O_I|f=1)$ (intensity image), (c) $P(O_R|f=0)$ (range image), (d) $P(O_R|f=1)$ (range image), (e) $P(O|f=0)$ (fused), (f) $P(O|f=1)$ (fused).

what poor for the non-Gaussian noise case, because these methods are not robust to impulse noise. Figs. 4(d) and 5(d) show the edge maps obtained by our new algorithm. In experiments, $\lambda_0 = 15.0$ and $\lambda_1 = 20.0$ are used, where these values are experimentally selected. As shown in Figs. 4(d) and 5(d), the false edges are eliminated by our new method, however the corner points are somewhat smoothed because of the inherent limitation of the LTS estimation. The corner points can be preserved if the surface parameters are estimated based on the multi-resolution method [16] which requires high computational complexity.

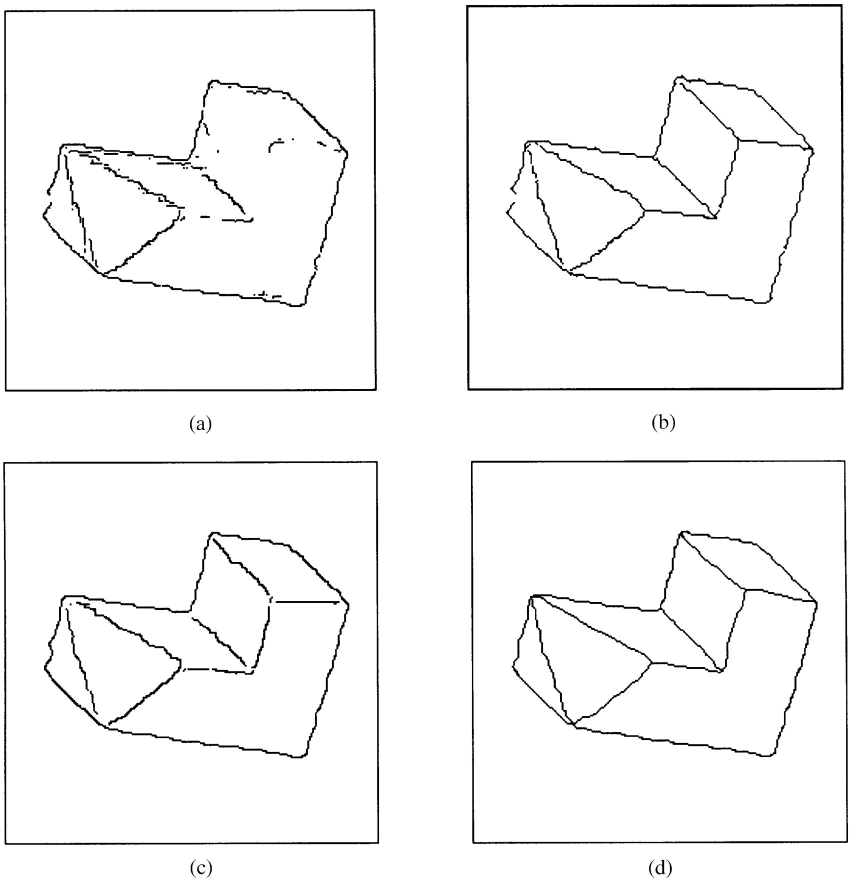

Fig. 4. Performance comparison of edge detection methods for Image A with Gaussian noise ($\sigma = 3.0$) and impulse noise (5%). (a) the Gil et al. method, (b) the Zhang and Wallace method, (c) the Nadabar and Jain method, (d) our new method.

On MIPS-5000 workstation, the Gil et al. method, Zhang and Wallace method, The Nadabar and Jain method, and our new method take about 20 s, 6, 8 and 15 min, respectively. It takes about 10 min to estimate surface parameters by the LTS method, thus further research will focus on the development of the fast surface parameter estimation algorithm.

4.3. Segmentation results

Figs. 6(a) and (b) show the segmentation results based on the final edge maps shown in Figs. 4(d) and 5(d), respectively. The test images are corrupted by Gaussian ($\sigma = 3.0$)/impulse (5%) noise mixture. As shown in Fig. 6, the image is segmented into a number of planar surfaces and curved surfaces. Although some fitting error appears large in the curved surface which is approximated by the first-order planar surface model, fusion of the range and intensity edge information reduces the error effect in the curved surface.

Figs. 7 and 8 show the segmentation performance, in terms of the number of incorrectly segmented pixels, as a function of Gaussian and impulse noise level for Images A and B, respectively. Gaussian and impulse noise independently corrupt the range and intensity images. Amount of Gaussian noise is represented by the standard deviation σ of the Gaussian function and that of impulse noise is expressed by the noise ratio percentage (%), in which the noise ratio is defined by the ratio of the number of contaminated pixels to the total number of pixels in an image. The segmentation performance is represented by the number of incorrectly segmented pixels, where total numbers of pixels considered are 15,984 and 9381 for Images A and B, respectively. The edge pixels are not considered in computing the segmentation performance because they are merged to neighboring regions. In Figs. 7 and 8, 'miss' represents the number of pixels in undetected regions, 'under-segmentation' signifies the number of pixels in under-segmented regions, and 'over-segmentation' denotes the number of pixels in over-segmented

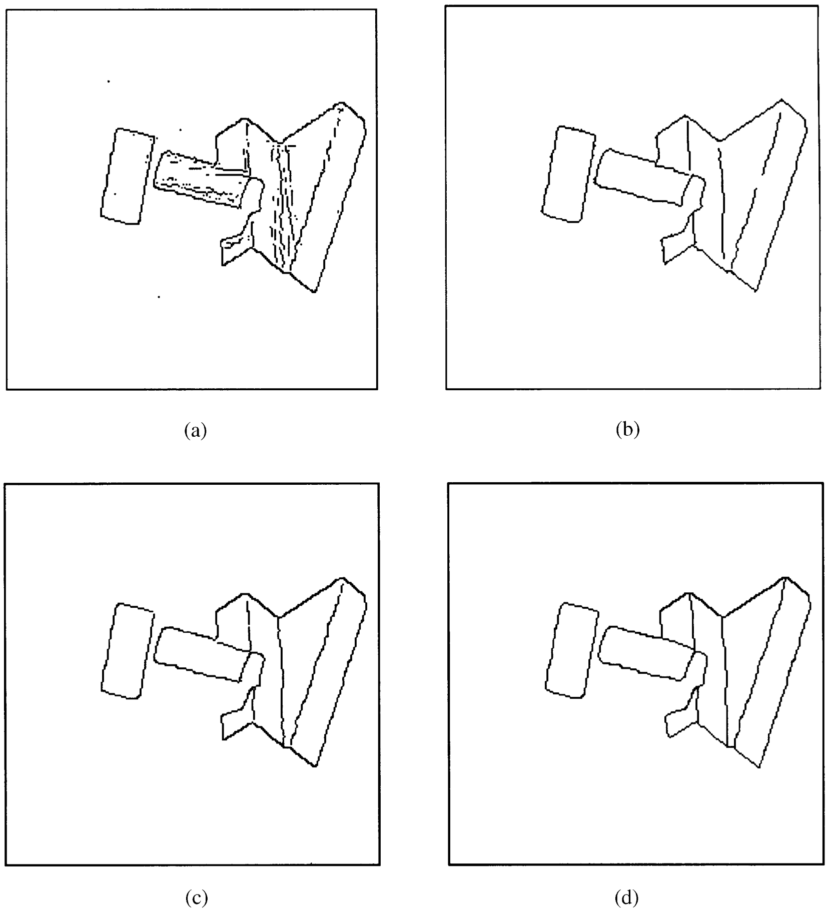

Fig. 5. Performance comparison of edge detection methods for Image B with Gaussian noise ($\sigma = 3.0$) and impulse noise (5%): (a) the Gil et al. method, (b) the Zhang and Wallace method, (c) the Nadabar and Jain method, (d) our new method.

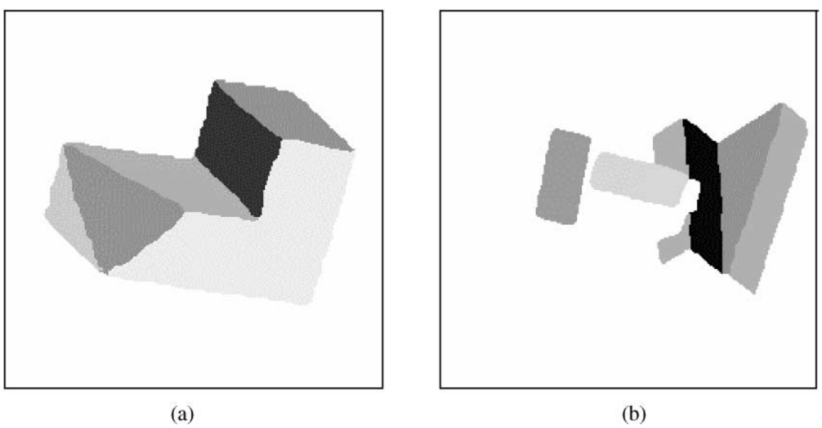

Fig. 6. Segmentation results of our new method (with Gaussian noise ($\sigma = 3.0$) and impulse noise (5%)): (a) Image A, (b) Image B.

regions. As shown in Figs. 7 and 8, the segmentation performance for Gaussian noise cases is lower than that for impulse noise cases because the trimmed methods (α-trimmed method and LTS method) are robust to impulse noise. Since the step edge pixels contaminated by Gaussian or impulse noise may not be detected, there are

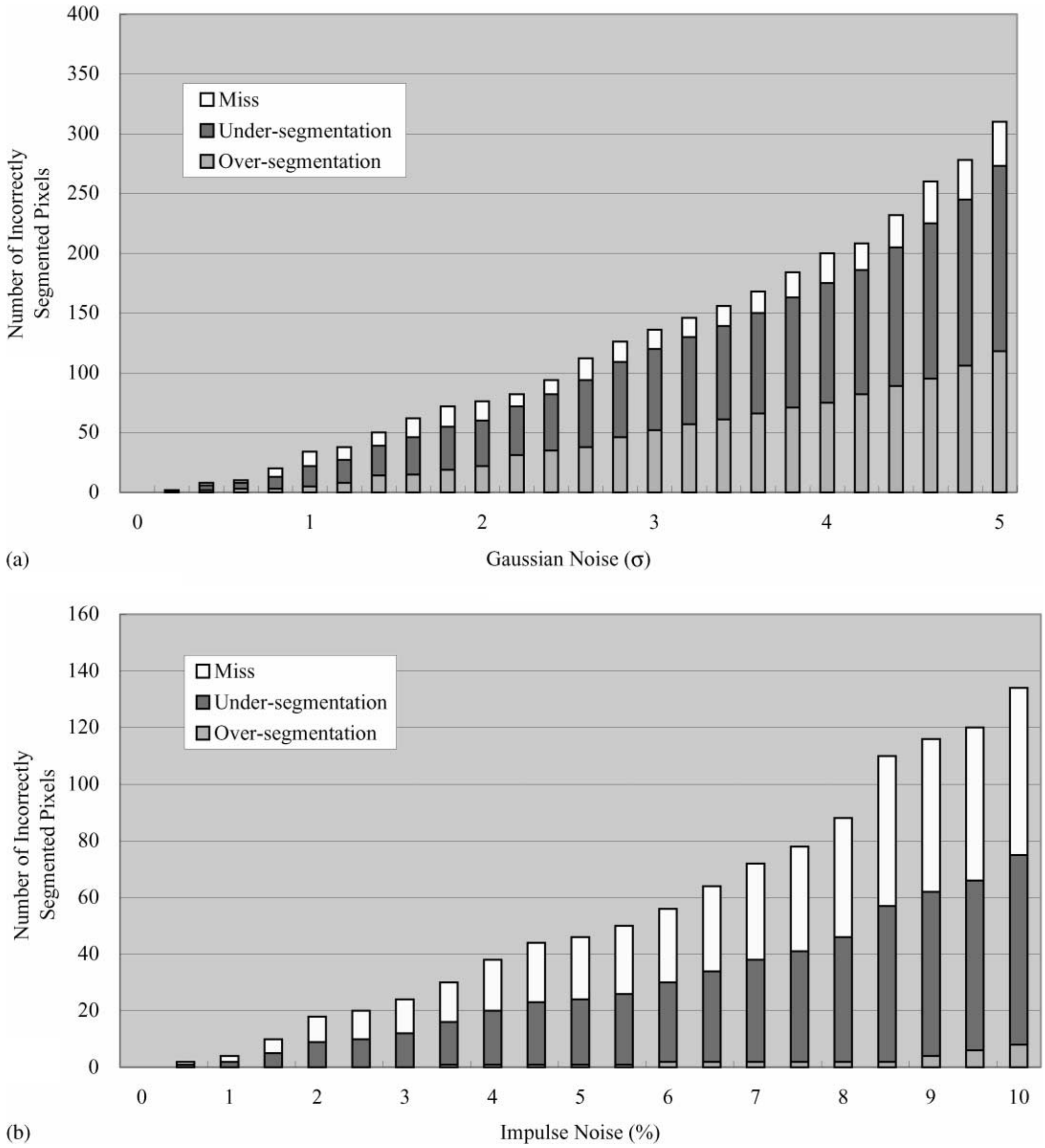

Fig. 7. Segmentation performance of our new method for Image A (total number of pixels considered: 15984): (a) Gaussian noise, (b) impulse noise.

more 'under-segmentation' pixels than 'over-segmentation' pixels in Images A and B. Also Figs. 7 and 8 show that the large number of step edge pixels yields the 'under-segmentation': 567 and 774 in Images A and B, respectively.

5. Conclusions

This paper presents the segmentation algorithm based on fusion of range and intensity images using robust trimmed methods. Fusing the edge information extracted from both images, the MAP estimator is implemented based on the MRF modeling. The likelihood functions of the intensity image are based on the statistical α-trimmed variance, whereas those of the range image are based on the surface parameters estimated based on the LTS method.

Simulation with various test image sets shows that segmentation results of our new algorithm are robust to Gaussian/impulse mixture noise. Future research will focus on the development of the fast parameter estimation algorithm and on the extension of the algorithm to higher-order surface cases.

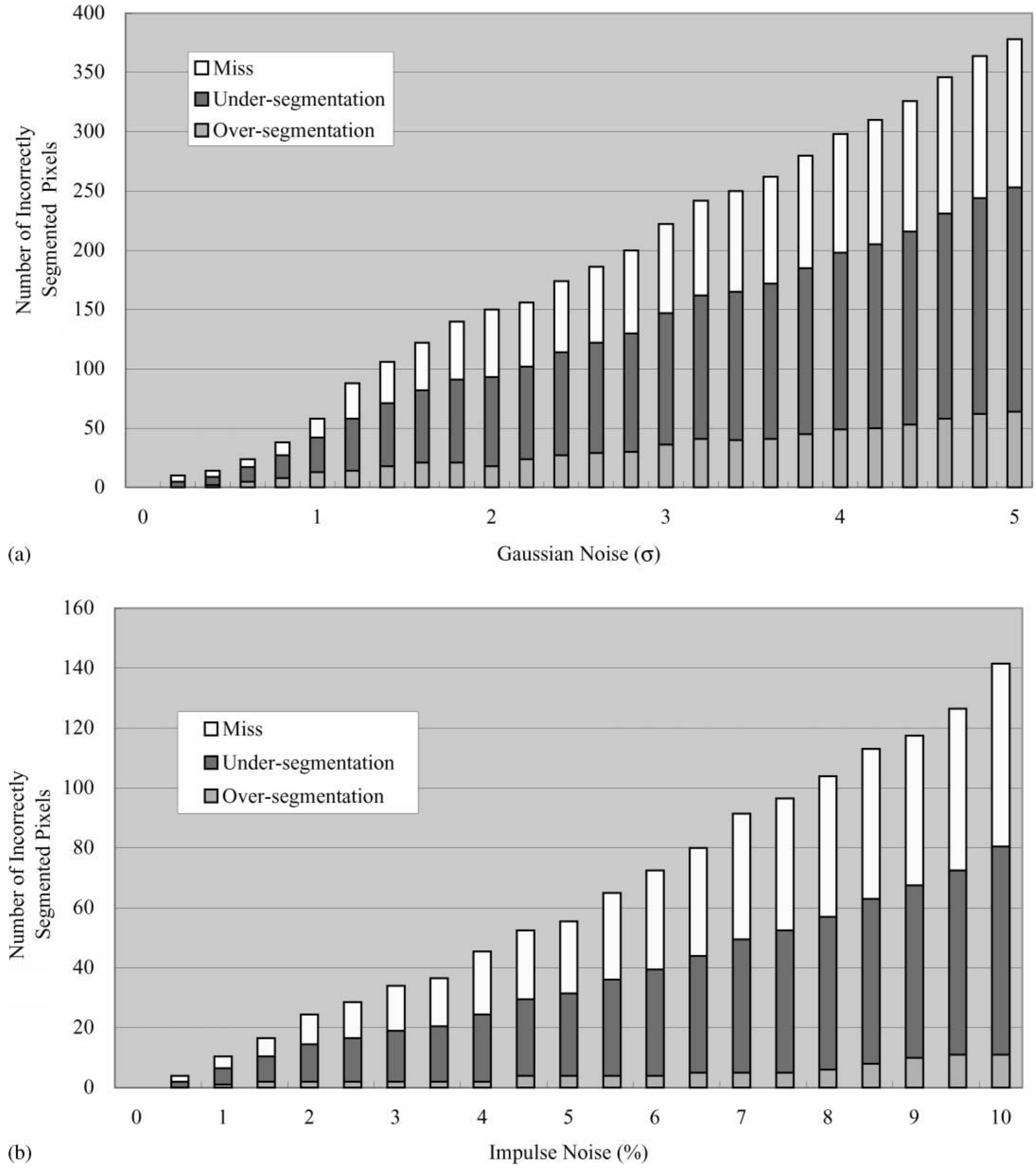

Fig. 8. Segmentation performance of our new method for Image B (total number of pixels considered: 9381): (a) Gaussian noise, (b) impulse noise.

Acknowledgements

This work was supported in part by the Engineering Research Center-Advanced Control and Instrumentation (KOSEF) and the Brain Korea 21 Project.

References

[1] P.J. Besl, Surfaces in Range Image Understanding, Springer, New York, 1988.

[2] J. Clark, A. Yulle, Data Fusion for Sensory Information Processing System, Kluwer, Boston, MA, 1990.

[3] A. Hoover, J.-B. Gillian, X. Jiang, P.J. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, R.B. Fisher, An experimental comparison of range image segmentation algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 673–689.

[4] K.-M. Lee, P. Meer, R.-H. Park, Robust adaptive segmentation of range images, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 200–205.

[5] B. Gil, A. Mitiche, J.K. Aggarwal, Experiments in combining intensity and range edge maps, Comput. Vision Graphics Image Process. 21 (1983) 395–411.

[6] G. Zhang, A. Wallace, Physical modeling and combination of range and intensity edge data, CVGIP: Image Understanding 58 (1993) 191–220.

[7] S.G. Nadabar, A.K. Jain, Fusion of range and intensity images on a connection machine (CM-2), Pattern Recognition 28 (1995) 11–26.

[8] S. Geman, D. Geman, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1984) 721–741.

[9] J. Besag, On the statistical analysis of dirty pictures, J. Roy. Statist. Soc. B 48 (1986) 259–302.

[10] P.B. Chou, C.M. Brown, The theory and practice of Bayesian image labeling, Int. J. Comput. Vision 4 (1990) 185–210.

[11] S.Z. Li, Markov Random Field Modeling in Computer Vision, Springer, New York, 1995.

[12] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. 8 (1986) 679–698.

[13] J.B. Bennar, T.L. Watt, Alpha-trimmed means and their relationship to median filters, IEEE Trans. Acoust., Speech, Signal Process. 32 (1984) 145–153.

[14] A. Leonardis, A. Gupta, R. Bajcsy, Segmentation of range images as the search for geometric parametric models, Int. J. Comput. Vision 14 (1995) 253–277.

[15] P.J. Rousseeuw, A.M. Leory, Robust Regression and Out Detection, Wiley, New York, 1987.

[16] I.S. Chang, D.-G. Sim, R.-H. Park, Multiresolution surface parameter estimation for range images, Proceedings of the IEEE International Conference Image Processing, Lausanne, Switzerland, 1996, pp. 37–40.

About the Author—IN SU CHANG received the B.S. and M.S. degrees in Electronics Engineering from Sogang University, Seoul, Korea, in 1995 and 1997, respectively. He is working toward the Ph.D. degree at the Sogang University. His current research interests are image processing, computer vision, lane detection, and image watermarking.

About the Author—RAE-HONG PARK received the B.S. and M.S. degrees in Electronics Engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively. He received the M.S. and Ph.D. degrees in Electrical Engineering from Stanford University, Stanford, California, in 1981 and 1984, respectively. He joined the faculty of the Department of Electronic Engineering at Sogang University, Seoul, Korea, in 1984, where he is currently a professor. In 1990, he spent his sabbatical year at the Computer Vision Laboratory of the Center for Automation Research, University of Maryland, College Park, Maryland, as a visiting associate professor. He received a scholarship from the Korean Government, Ministry of Education, from 1979 to 1983, and a Post-Doctoral Fellowship from the Korea Science and Engineering Foundation (KOSEF) in 1990. He received the Academic Award in 1987 from the Korea Institute of Telematics and Electronics (KITE). Also he received the First Sogang Academic Award in 1997 and the Professor Achievement Excellence Award in 1998 and 1999, all from Sogang University. He served as Editor for the KITE Journal of Electronics Engineering in 1995–1996. His current research interests are computer vision, pattern recognition, and video communication.