Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-3-2003

MDPE: A Very Robust Estimator for Model Fitting and Range Image Segmentation

H. Wang and D. Suter

MONASH
UNIVERSITY

MDPE: A Very Robust Estimator for Model Fitting and Range Image Segmentation

Hanzi Wang and David Suter
Department of Electrical and Computer Systems Engineering
Monash University, Clayton Vic. 3800, Australia.
{hanzi.wang ; d.suter}@eng.monash.edu.au

Abstract

In this paper, we propose a novel and highly robust estimator, called MDPE (Maximum Density Power Estimator). This estimator applies nonparametric density estimation and density gradient estimation techniques in parametric estimation. MPDE optimizes an objective function that measures more than just the residuals. Both the density distribution of data points and the size of the residual corresponding to the local maximum of the density distribution, are considered as important characteristics in our objective function. MDPE can tolerate more than 85% outliers. Compared with several other recently proposed similar estimators, MDPE has a higher breakdown point and less error variance.

We also present a new range image segmentation algorithm, based on a modified version of the MDPE (Quick-MDPE), and its performance is compared to several other segmentation methods. Segmentation requires more than a simple minded application of an estimator, no matter how good that estimator is: our segmentation algorithm overcomes several difficulties faced with applying a statistical estimator to this task.

Keywords: robust estimation, breakdown point, model fitting, range image segmentation, least median of squares, residual consensus, adaptive least kth order squares, mean shift.

1. Introduction

An important goal of computer vision algorithms is to extract geometric information from an image, or from image sequences. Parametric models play a vital role in this and other activities in computer vision research. When engaged in parametric fitting in a computer vision context, it is important to recognise that data obtained from the image or image sequences may be inaccurate. It is almost unavoidable that data are contaminated (due to faulty feature extraction, sensor noise, segmentation errors, etc) and it is also likely that the data will include multiple structures. We can say that the first set of contaminations will introduce outliers into the data and that the second form (multiple structures) will introduce

pseudo-outliers into the data. Pseudo-outliers are distinguished from outliers in that they represent valid data – merely that they are extraneous to a given single parametric model fit.

Thus it is widely acknowledged that computer vision algorithms should be robust (to outliers and pseudo-outliers). This rules out a simple-minded application of the least squares (LS) method. Although the least squares estimator is highly efficient when data are corrupted by Gaussian noise, it is extremely sensitive to outliers. As a result, this method may break down when the data contain outliers. The breakdown point of an estimator may be roughly defined as the smallest percentage of outlier contamination that can cause the estimator to produce arbitrarily large values (Rousseeuw and Leroy, 1987) The LS estimator has a breakdown point of 0%, because only one single extreme outlier is sufficient to force the LS estimator to produce arbitrarily large values.

Great efforts have been made in the search for high breakdown point estimators in recent decades. Although several robust estimators were developed during the past three decades, most of them (such as the LMedS (Rousseeuw, 1984), the LTS (Rousseeuw, 1984), the RM (Siegel 1982), etc.) can only tolerate 50% gross errors. In computer vision tasks, it frequently happens that outliers and pseudo-outliers occupy the absolute majority of the data. Therefore, their requirement that outliers occupy less than 50% of all the data points is far from being satisfied for the real tasks faced in computer vision. A good robust estimator should be able to correctly find the fit when outliers occupy a higher percentage of the data (more than 50%). Also, ideally, the estimator should be able to resist the influence of all types of outliers (e.g., uniformly distributed outliers, clustered outliers and pseudo-outliers).

This paper presents a novel robust estimator (MDPE). The goals in designing the MDPE are: it should be able to fit signals corresponding to less than 50% of the data points and be able to fit data with multi-structures. In developing the MPDE, we make the common assumption that the residuals of the inliers are contaminated by Gaussian noise (although the precise nature of the noise distribution is not that essential, depending only upon zero mean and unimodality). We also assume that the signal (we seek to fit) occupies a relative majority of the data – that is, there are no other populations, belonging to valid structures, that singly has a larger population. In other words, if there are multiple structures, we seek to fit the largest structure (in terms of population of data – which is often related to but not necessarily identical to geometric size). Of course, in a complete application of the MPDE, such as the

range segmentation algorithm presented later, one can apply the estimator serially to identify the largest structured population, remove it and then seek the largest in the remaining population etc.

Key components of MPDE are: Probability Density estimation in conjunction with Mean Shift techniques (Fukunaga and Hostetler, 1975). The mean shift vector always points towards the direction of the maximum increase in the probability density function (see section 3). Through the mean shift iterations, the local maximum density, corresponding to the mode (or the center of the regions of high concentration) of data, can be found.
The MDPE depends upon an objective function, which consists of two factors:

- The density distribution of the data points estimated by the density estimation technique.
- The size of the residual corresponding to the local maximum of probability density.

If the signal is correctly fitted, the densities of inliers should be as large as possible; at the same time, the center of the high concentration of data should be as close to zero as possible in the residual space.

MDPE can tolerate a large percentage of outliers and pseudo-outliers (more than 85%) and it can achieve better performance than other similar robust estimators. To demonstrate the performance of MDPE we first present some simple tests based upon both synthetic and real images. However, a good estimator is generally only one component of a complete scheme to successfully tackle meaningful computer vision tasks. Thus, in the latter part of this paper, we develop a complete algorithm for the challenging task of range segmentation, using MPDE at its core. We also modify the MDPE to produce a quicker version—QMDPE, with higher computing speed but a little lower breakdown point (still higher than other compared estimators). Experimental comparisons of the proposed approach and several other state-of-the-art methods support the claim that the proposed method has a higher breakdown point and is more robust to outliers.

In the latter part of construction of this paper, the authors became aware of (Cheng and Meer, 2002). This work has some similar ideas to our work in that both methods employ kernel density estimation technique. However, their work places emphasis on the projection pursuit paradigm and on data fusion. Moreover, they use an M-estimator paradigm (see

section 2). Though there are nice theoretical links between M-estimator versions of robust estimators and random sampling methods, as referred to in that paper, the crucial fact remains that LMedS and RANSAC type methods have a higher breakdown point. Moreover, only synthetic examples containing relatively few surfaces are given, and no comparisons to previous approaches are given in that paper. Thus, though their work contains a similar idea that is also a key to our own approach, the differences are significant and, at this point in time, it is difficult to compare the performance of the two approaches.

The contributions of this paper can be summarized as follows:

- We provide a novel estimator that can tolerate more than 85% outliers although it is simple and easy to implement.
- We modify the MDPE to produce a quicker version—QMDPE.
- We propose an algorithm, using the QMDPE, for range image segmentation. The comparative experiments illustrate that the proposed algorithm can achieve good results even when the range images are contaminated by a large number of (impulse) noisy data points.

The organization of this paper is as follows: we review several previous methods and their limits in section 2. The density gradient estimation and mean shift method are introduced in section 3. Section 4 describes the MDPE method. Comparative experimental results of the MDPE and several other robust estimators are contained in section 5. In section 6, a quick version of the MDPE—QMDPE is presented. Section 7 provides a realisitic application of the QMDPE to range image segmentation: indeed, a complete segmentation algorithm is given and is shown to be very effective for this task. Finally, we conclude with a summary and a discussion of further possible work.

2. Previous Robust Estimators.

The maximum-likelihood-type estimators (M-estimators) (Huber, 1973; Huber, 1981) are well known among the robust estimators. The theory of M-estimators was firstly developed by Huber. They minimize the sum of a symmetric, positive-definite function of the residuals with a unique minimum at zero. Although M-estimators can reduce the influence of outliers, they have breakdown points less than 1/(p+1), where p is the number of the parameters to

estimate. This means that the breakdown point of M-estimators will diminish when the dimension of parameter vector increases.

In 1982, Siegel proposed the repeated median (RM) estimator with the high breakdown point of 50% (Siegel, 1982). However, the time complexity of the repeated median estimator is $O(n^p \log^p n)$, which prevents the method being useful in applications where p is even moderately large.

Rousseeuw proposed the least median of squares (LMedS) method in 1984 (Rousseeuw, 1984). The LMedS finds the parameters to be estimated by minimizing the median of residuals corresponding to the data points. In practice, only an approximate LMedS, based upon random sampling, can be implemented for any problem of a reasonable size – we generally refer to this approximate version when we use the term LMedS (a convention adopted by most other authors as well). The LMedS method was based on certain assumptions as follows:

- The signal to estimate should occupy the majority of all the data points, that is, more than 50% data points should belong to the signal to estimate.
- The correct fit will correspond to the one with the least median of residuals. This criterion is not always true when the data includes multiple structures and clustered outliers, and when the variance of inliers is large.
- We already know what kind of primitive model to fit, e.g., to fit a circle, an ellipse, a plane, etc.
- There always is at least one signal in the data. If there is no signal in the data, the LMedS will "hallucinate" and still find one fit involving 50% of the data points.

Although the LMedS has been very successfully applied to a single signal corrupted with uniformly distributed outliers, it is less effective when presented with multiple structures and clustered outliers. Even more, when the gross errors comprise more than 50% of the data, the LMedS method will fail completely.

Obviously, the requirement for 50% or more data belonging to inliers may not be always satisfied, e.g., when the data contain multiple surfaces, when data from multiple views are merged, or when there are more than 50% noise data points existing in the data, the inliers

will contain no more than 50% data points. For these cases, we need to find a more robust estimator that can tolerate more than 50% outliers.

Although Rousseeuw and Leroy (1987) have pointed out that 50% breakdown point is the best that we can expect, they require the estimator to have a unique solution. If we relax this requirement, and permit the case of multiple solutions to exist, we can break though this limit and propose a method that can tolerate more than 50% outliers. RESC is one successful example of these methods (Yu, et al., 1994). The RESC method uses a compressed histogram method to infer residual consensus. Instead of using residual as its criteria, the RESC method uses the histogram power (Yu, et al., 1994) as its criteria. The RESC method finds the parameters by choosing the p-subset corresponding to the maximum histogram power. The authors in (Yu, et al., 1994) claim that the RESC has a higher than 80% breakdown point. They applied RESC to range image segmentation. But no comparative experiments were given.

MINPRAN is another kind of estimator that has a higher than 50% breakdown point (Stewart, 1995). It can find correct the model in the data involving more than 50% outliers without prior knowledge about error bounds and it does not " hallucinate" fits when there are no real structures in the data. However, MINPRAN assumes that the outliers are randomly distributed within a certain range. This makes MINPRAN less effective in extracting multiple structures.

The authors of MUSE (Miller and Stewart, 1996) and those of ALKS (Lee et al., 1998) consider robust scale estimate and they both can obtain a more than 50% breakdown point. MUSE and ALKS can perform better than LMedS and M-estimators at small scale discontinuities. However, MUSE needs a lookup table for the scale estimator correction; ALKS is limited in its ability to handle extreme outliers. Another problem we found in ALKS is its lack of stability under a small percentage of outliers (which will be illustrated in section 4).

Bab-Hadiashar and Suter (1999) have used least K-th order (rather than median) methods and a heuristic way of estimating scale to perform range segmentation. However, though their method can handle large percentages of outliers and pseudo-outliers, it does not seem as successful in tolerating extreme cases, situations where the method presented here, still does well.

3. Density Gradient Estimation and Mean shift Method

There are several nonparametric methods available for probability density estimation: histogram, naive method, the nearest neighbor method, and kernel estimation (Silverman, 1986). The kernel estimation method is one of the most popular techniques used in estimating density. Given a set of n data points $\{X_i\}_{I=1,\ldots,n}$ in a d-dimensional Euclidian space R^d, the multivariate kernel density estimator with kernel K and window radius (band-width) h is defined as follows (Silverman, 1986, p.76)

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{1}$$

The kernel function K(x) should satisfy some conditions (Wand and Jones, 1995, p.95).

There are several different kinds of kernels. The Epanechnikov kernel (Silverman, 1986, p.76) is one optimum kernel which yields minimum mean integrated square error (MISE):

$$K_e(\mathrm{X}) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-\mathrm{X}^T\mathrm{X}) & if\ \mathrm{X}^T\mathrm{X} < 1 \\ 0 & otherwise \end{cases} \tag{2}$$

where c_d is the volume of the unit d-dimensional sphere, e.g., $c_1=2$, $c_2=\pi$, $c_3=4\pi/3$.

The estimate of the density gradient can be defined as the gradient of the kernel density estimate (1)

$$\hat{\nabla} f(x) \equiv \nabla \hat{f}(x) = \frac{1}{nh_d}\sum_{i=1}^{n} \nabla K(\frac{x - X_i}{h}) \tag{3}$$

According to (3), the density gradient estimate of the Epanechnikov kernel can be written as

$$\hat{\nabla} f(x) = \frac{n_x}{n(h^d c_d)}\frac{d+2}{h^2}\left(\frac{1}{n_x}\sum_{X_i \in S_h(x)}[X_i - x]\right) \tag{4}$$

where the region $S_h(x)$ is a hypersphere of the radius h, having the volume $h^d c_d$, centered at x, and containing n_x data points.

The mean shift vector $M_h(x)$ is defined as

$$\mathrm{M}_\mathrm{h}(\mathrm{x}) \equiv \frac{1}{n_x}\sum_{X_i \in S_h(x)}[X_i - x] = \frac{1}{n_x}\sum_{X_i \in S_h(x)} X_i - x \tag{5}$$

Equation (4) can be rewritten as

$$\mathrm{M}_\mathrm{h}(\mathrm{x}) \equiv \frac{h^2}{d+2}\frac{\hat{\nabla} f(x)}{\hat{f}(x)} \tag{6}$$

Equation (6) firstly appeared in (Fukunaga and Hostetler, 1975). Equation (5) shows that the mean shift vector is the difference between the local mean and the center of the window. Equation (6) shows the mean shift vector is an estimate of the normalized density gradient. The mean shift is an unsupervised nonparametric estimator of density gradient. One characteristic of the mean shift vector is it always points towards the direction of the maximum increase in the density.

The Mean Shift algorithm can be described as follows:

1. Choose the radius of the search window
2. Initialize the location of the window.
3. Compute the mean shift vector $M_h(x)$.
4. Translate the search window by $M_h(x)$.
5. Step 3 and step 4 are repeated until convergence.

The converged centers (or windows) correspond to modes (or centers of the regions of high concentration) of data represented as arbitrary-dimensional vectors. The proof of the convergence of the mean shift algorithm can be found in (Comaniciu and Meer, 1999, 2002). Since its introduction by Fukunaga and Hostetler (1975), the mean shift method has been extensively exploited and applied in low levels computer vision tasks (Cheng, 1995; Comaniciu and Meer, 1997, 1999, 2002) for its ease and efficiency.

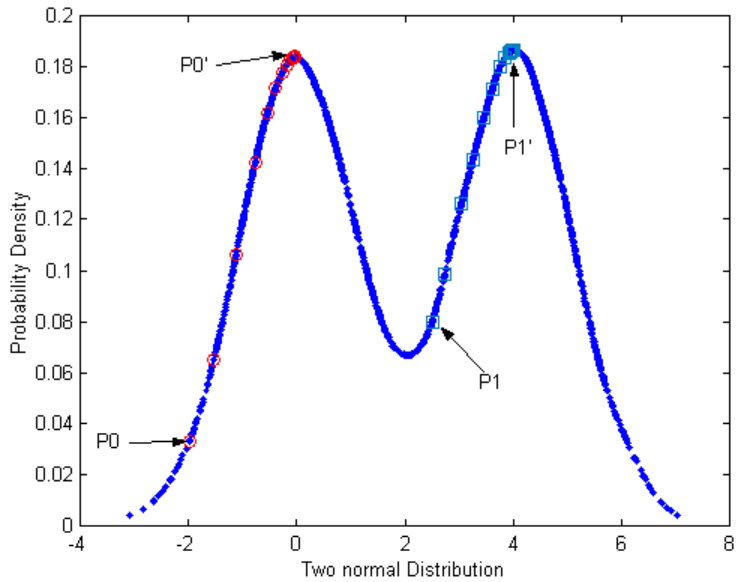


Figure 1. One example where the mean shift estimator found the local maximum of the probability densities.

To illustrate the mean shift method, two normal distributions are generated, each having 1000 data points and with unit variance. One has a distribution with zero mean, and the other has a mean of 4.0 (see figure 1). We selected two initial points: P0 (-2.0) and P1 (2.5). The search window radius was chosen as 1.0. After applying mean shift algorithm, the mean shift estimator automatically found the local maximum densities (converged points). Precisely, P0' located at -0.0305, and P1' with 4.0056. The centers (P0' and P1') of the converged windows correspond to the local maximum probability densities, that is, the two modes.

4. Maximum Density Power Estimator—MDPE

4.1 The density power (DP)

Random sampling techniques have been widely used in a lot of methods, for example, LMedS, RESC, ALKS, etc. Each uses the random sampling techniques to choose p points, called a p-subset, determines the parameters of a model for that p-subset (p equals 2 for a line, 3 for a circle or plane, 6 for a quadratic curve), and finally outputs the parameters determined by the p-subset with the minimum or maximum of the respective objective function. Here we derive a new objective function.

When a model is correctly fitted, there are two criteria that should be satisfied:

(1) Data points on or near the model (inliers) should be as many as possible;

(2) The residuals of inliers should be as small as possible.

Most objective functions of existing random sampling methods consider either one of the criteria or both. RANSAC (Fischler and Rolles, 1981) applies criterion (1) into its optimization process and outputs the results with the highest number of data points within an error band; The Least squares method uses criterion (2) as its objective function, but minimizes the residuals of all data points without the ability to differentiate the inliers from the outliers; MUSE, instead of minimizing the residuals of inliers, minimizes the scale estimate provide by the kth ordered absolute residual. RESC combines both criteria into its objective function, i.e., the histogram power. Among all these methods, RESC obtains a highest breakdown point. It seems that it is preferable to consider both criteria in the objective function.

The new estimator we introduce here, MDPE, also considers these two criteria in its objective function. We assume the residuals of the inliers (good data points) satisfy a zero mean, smooth and unimodal distribution: e.g., a Gaussian distribution. If the model to fit is correctly estimated, the data points on or near the fitted structure should have a higher probability density; and at the same time, the center of the converged window by the mean shift procedure (corresponding to the highest local probability density) should be as close to zero as possible in residual space. According to the above assumptions, our objective function ψ_{DP} considers two factors: (1) the densities $\hat{f}(X_i)$ of all data points within the converged window W_c and (2) the center C of the converged window. Thus $\psi_{DP} \propto \sum_{X_i \in Wc} \hat{f}(X_i)$ and $\psi_{DP} \propto \frac{1}{|C|}$.

We define the probability density power function as follows:

$$\psi_{DP} = \frac{\left(\sum_{X_i \in W_c} \hat{f}(X_i) \right)^{\alpha}}{\exp(|C|)^{\beta}} \tag{8}$$

where C is the center of the converged window W_c obtained by applying the mean shift procedure; α, β are the parameters that adjust the relative influence of the probability density and the residual of the point corresponding to the center of the converged window. They are empirically set to 1.0.

The density power is proportional to the sum of densities of all data points within the converged window and inversely proportional to the center of the converged window obtained by the mean shift. If a model is found, $|C|$ is very small, and the densities within the converged window are very high. Thus our objective function will produce a high score. Experiments, presented next, show the MDPE is a very powerful method for data with a large percentage of outliers.

4.2 The MDPE algorithm

As Lee stated, any one-step robust estimator cannot have a breakdown point exceeding 50% (Lee et al., 1998). The MDPE adopts a multistep procedure. The procedure of the MDPE can be described as follows:

(1) Choose a search window radius h, and a repetition count m. The value m, for the probability P that at least one "clean" p-subset being chosen from m p-subsets, is determined by $m = \frac{\log(1-P)}{\log[1-(1-\varepsilon)^p]}$, where ε the fraction of outliers (possibly including pseudo-outliers, clustered and uniformly distributed outliers) contained in the whole set of points.

(2) Randomly choose one p-subset, estimate the model parameters by the p-subset, and calculate the signed residuals of all data points.

(3) Apply the mean shift steps in the residual space with initial window center zero. Notice that the mean shift is employed in one dimensional space –signed residual space. The converged window center C can be obtained by the procedure above.

(4) Calculate the densities (using equation (1)) corresponding to the positions of all data points within the converged window with radius h in the residual-density space.

(5) Calculate the density power according equation (7).

(6) Repeat step (2) to step (5) m times. Finally, output the parameters with maximum density power.

The results are from one p-subset, corresponding to the maximum density power. In order to improve the statistical efficiency, a weighted least square procedure (Rousseeuw, 1987, p.202) can be carried out after the initial the MDPE fit.

Instead of estimating the fit involving the absolute majority in the data set, the MDPE finds the relative majority of the data points. This makes it possible for the MDPE to obtain a greater than 50% breakdown point.

5. Experiments and Analysis

In this section, we will compare the abilities of several estimators (MDPE, RESC, ALKS, and LMedS) to deal with data with a large percentage of outliers. We investigate the characteristics of the methods under clustered outliers and different percentages of outliers, and test the influence of the choice of window radius on the MDPE. Unless we specify the window radius h, it is set at 2.0.

Experiment 1.

We generated four kinds of data (step, three-step, roof, and six-line), each with a total of 500 data points. The signals were corrupted by Gaussian noise with zero mean and standard

variance σ. Among the 500 data points, α data points were randomly distributed in the range of (0, 100). The ith structure has γ_i data points.

(a) Step: x:(0-55), y=30, γ_1=65; x:(55-100), y=40, γ_2=30; α=405; σ=1.

(b) Three-step: x:(0-30), y=20, γ_1=45; x:(30-55), y=40, γ_2=30; x:(55-80), y=60, γ_3=30; x:(80-100), y=80, γ_4=30; α=365; σ=1.

(c) Roof: x:(0-55), y=x+30, γ_1=35; x:(55-100), y=140-x, γ_1=30; α=435; σ=1.

(d) Six-line: x:(0-25), y=3x, γ_1=30; x:(25-50), y=150-3x, γ_2=20; x:(25-50), y=3x-75, γ_3=20; x:(50-75), y=3x-150, γ_4=20; x:(50-75), y=225-3x, γ_5=20; x:(75-100), y=300-3x, γ_6=20; α=370; σ=0.1.

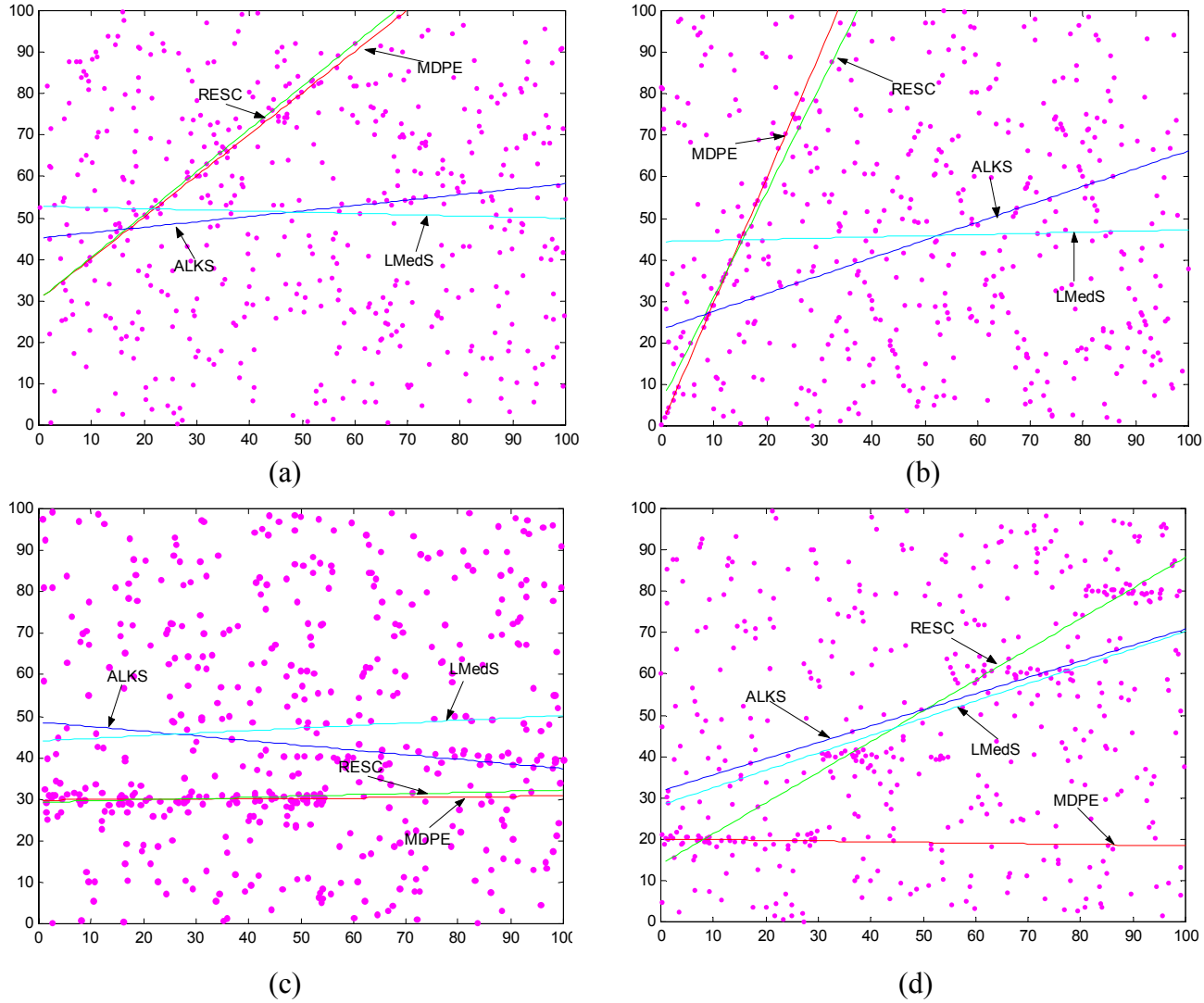


Figure 2. Comparing the performance of four methods: (a) fitting a step with a total of 87% outliers; (b) fitting three steps with a total of 91% outliers; (c) fitting a roof with a total of 93% outliers; (d) fitting six lines with a total of 94% outliers.

From figure 2, we can see that because LMedS has only a 0.5 breakdown point, it cannot resist more than 50% outliers. Thus, LMedS failed to fit all the four signals; The ALKS, RESC and MDPE approaches all have a more than 50% breakdown point. But the results show that ALKS is not applicable for the signals with such large percentages of outliers because it failed in all four cases. In contrast, RESC successfully fitted three models, but failed one. Only the MDPE method correctly fitted all the four signals. The MDPE didn't breakdown even with 94% outliers.

The MDPE is a general method which can be easily extended to fit other kinds of models, such as circles, ellipsis, planes, etc. Figure 3 shows the ability of the MDPE to fit circles under 95% outliers. Five circles were generated, each with 100 data points and σ=0.1. 1500 random outliers were distributed at range (-75 - 75). H was set 7.0. Thus, for each circle, it has 1900 outliers (400 pseudo-outliers plus 1500 random outliers). The MDPE method gave the most accurate results of the four methods.

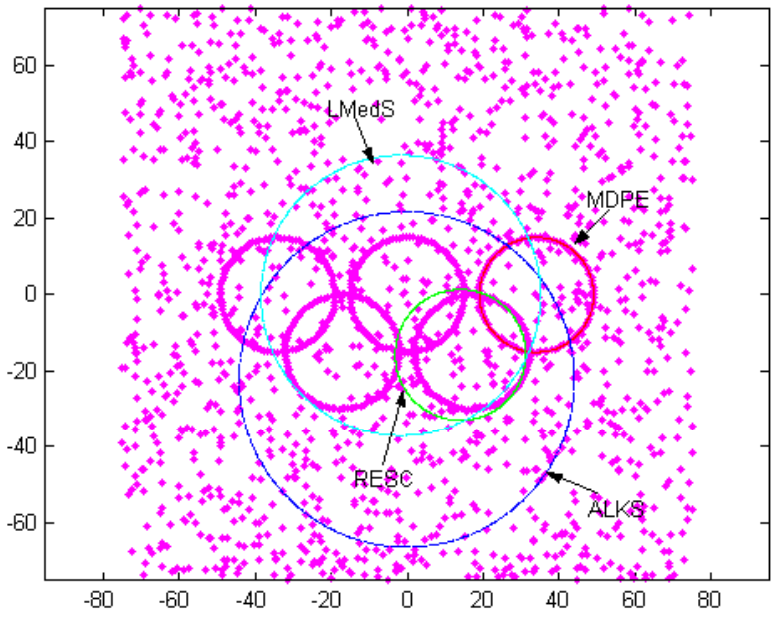


Figure 3. One example of fitting circles by the four methods. The data had 95% outliers.

Experiment 2.

In the previous experiment, we investigated the characteristics of the four methods to fit data with multiple structures. Here, we will explore the abilities of the four methods to fit data

with clustered outliers. We generated a line (y=x-1) corrupted by Gaussian noise with zero mean and standard variance σ_1. The line had γ data points. Among the total 500 data points, α data points were randomly distributed in the range of (0, 100.0), and β clustered outliers were added to the signals, possessing a spherical bivariate normal distribution with standard variance σ_2 and mean (80.0, 30.0).

(a) γ=100, σ_1=1.0; α=200; β=200; σ_2=5.0.

(b) γ=100, σ_1=1.0; α=200; β=200; σ_2=2.0.

(c) γ=275, σ_1=1.0; α=0; β=225; σ_2=1.0.

(d) γ=275, σ_1=5.0; α=0; β=225; σ_2=1.0.

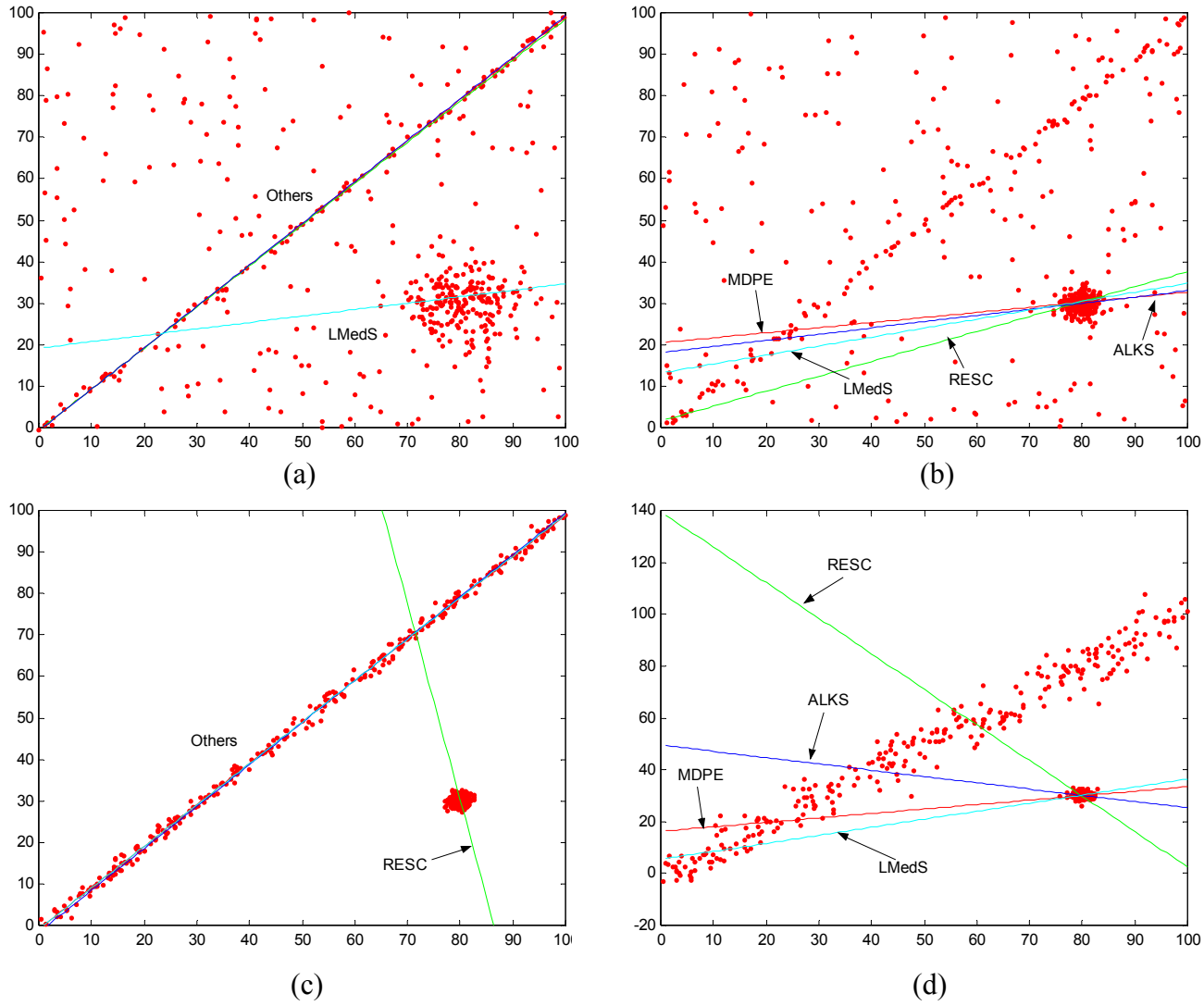


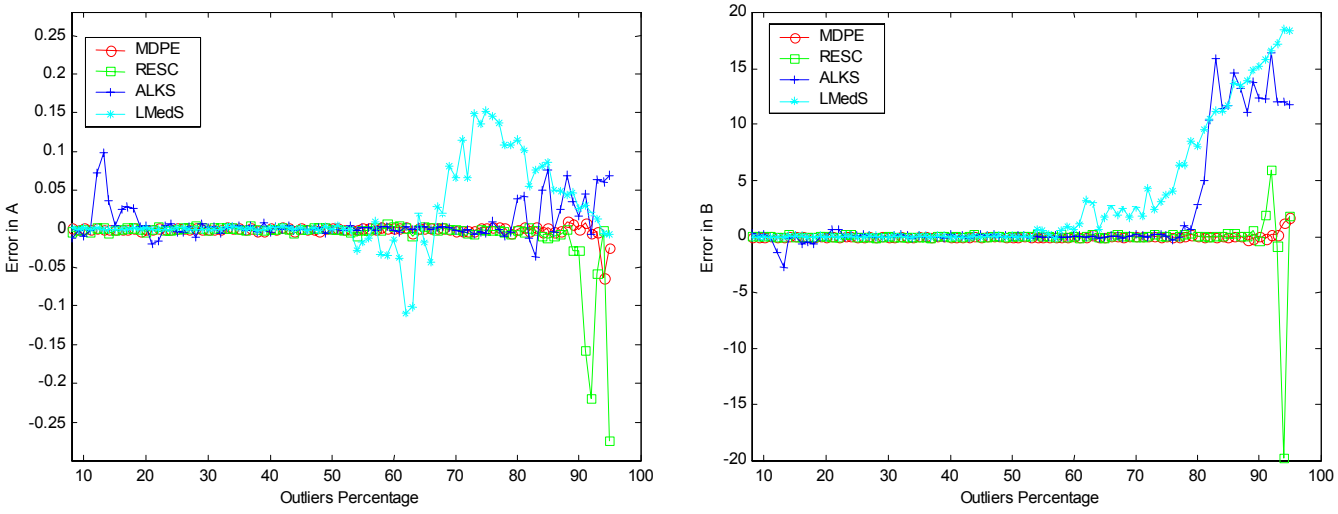
Figure 4. Experiments where the four methods are fitting a line with clustered outliers. The standard variance of both clustered outliers and inliers will affect the results of the four methods.

Figure 4 shows that both the standard variance of clustered outliers σ_2 and the standard variance of inliers to the line σ_1 will decide the accuracy of the results estimated by the four methods. When σ_1 is small and σ_2 is large, the MDPE, the RSEC, and the ALKS all can correctly fit the line although a large number of clustered outliers existed in the data (see figure 4 (a)). The LMedS failed because it cannot tolerate more than 50% outliers. When the standard variance of clustered outliers is small, i.e., the outliers are densely clustered within a small range; the ability of all the four methods to resist the influence of clustered outliers will be greatly reduced (see figure 4 (b)). As shown in figure 4 (c) and figure 4 (d), the standard variance of inliers to the line will also affect the accuracy of the results by the four methods. When σ_1 was increased from 1.0 to 5.0, all the four methods failed to fit the line even with only 45% clustered outliers.

Experiment 3.

It is important to know the characteristics of the various methods when the signals were contaminated by different percentages of outliers. In this experiment, we will draw the "breakdown plot" and compare the abilities of the four methods to resist different percentages of outliers. We generated step signals (y=Ax+B) as follows:

Signals: line 1: x:(0-55), A=0, B=30, γ_1 will be decreased with the increase of uniformly distributed outliers α; line 2: x:(55-100), A=0, B=60, γ_2=25; for both lines: σ=1.



(a) (b)

Figure 5. Breakdown plot for the four methods: (a) error in A vs. outlier percentage; (b) error in B vs. outlier percentage.

In total 500 points. 15 clustered outliers centred at (80,10) with unit variance were added to the signals. At the beginning, $\gamma_1 = 460$, $\alpha=0$, so the first signal had an initial 8% outliers; then every repeat of the experiment 5 points was moved from γ_1 to uniform outliers (α) ranging over (0-100) until $\gamma_1=25$. Thus the percentage of outliers in the data points changed from 8% to 95%. The whole procedure above was repeated 20 times.

As figure 5 illustrated, the LMedS had a least breakdown point (about 50%) among all these four estimators. In contrast, the MDPE has a highest breakdown. RESC began to breakdown when outliers comprised more than 88% of the total data; ALKS broke down even when outliers comprised less than 80%. The MDPE began to breakdown only at 94% outliers. However, even at 94% and 95% outliers, the MDPE had still about 75% correct estimation rate out of the 20 times; while the other three methods all wrongly estimated the models. Another thing we noticed is that the ALKS has some obvious fluctuations in the results when the outliers are less than 30%, while the other three have not this undesirable characteristic. This may be because the robust estimate of the noise variance is not valid for small or large k values (k is the optimum value to be determined by the data). Among all the four methods (the MDPE, the RESC, the LMedS and the ALKS), the MDPE has the least error variance.

Experiment 4.

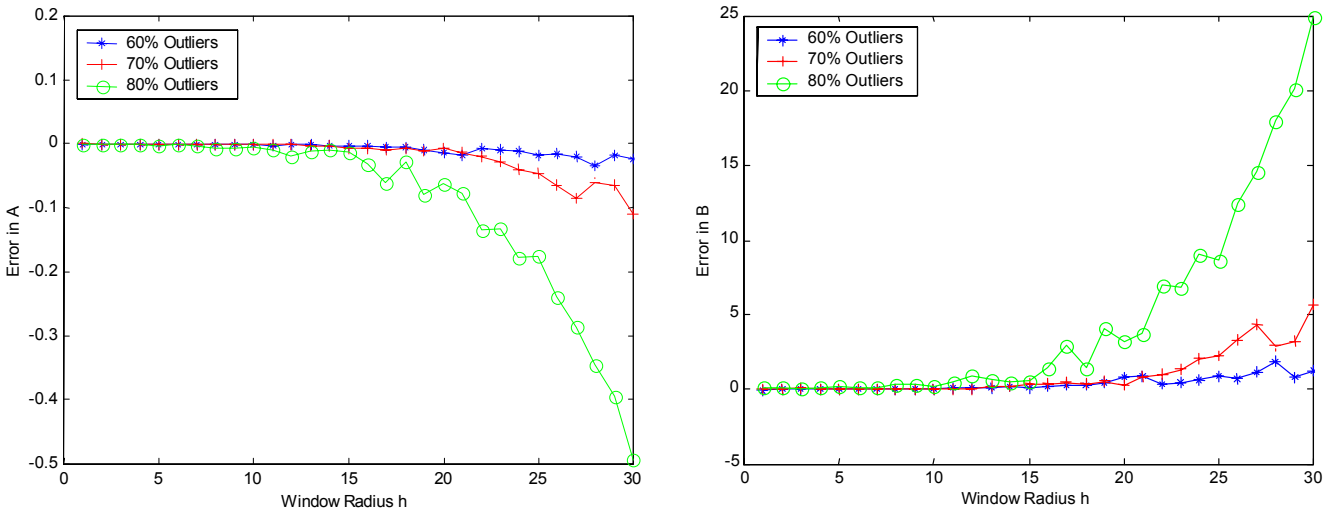


Figure 6. The influence of window radius and percentage of outliers on the results of the MDPE.

Although the MDPE has showed its powerful ability to tolerate large percentage of outliers (including pseudo-outliers), its success is decided by the correct choice of window radius h. If h is chosen too small, it is possible that the densities of data points in the residual space may not be corrected estimated (the density function is a noisy function with many local peaks and valleys), and some inliers may possibly be neglected; on the other hand, if h is set too large, the window will include all the data points including inliers and outliers; all peaks and valleys of the density function will also be smoothed out. In order to investigate the influence of the choice of window radius h and percentage of outliers on the estimated results, we generated a line $y=Ax+B$, where $A= 1$ and $B= -1$; $x \varepsilon (0\text{-}100)$. The line was corrupted by Gaussian noise with a unit variance. In total, 500 data points were generated. 60%, 70% and 80% percentage of uniformly distributed outliers in the range (0-100) were added to the signal. The window radius h was set from 1 to 30 with increasing step by 1 each time. The results were repeated 20 times. Figure 6. shows that the errors in A and B increase with the window radius h because when the radius becomes larger, it is possible that more outliers were included within the converged window. But the influence of different window radii on the results is small for different percentages of outliers when h is within a certain range: for example, for this case, when h is set within the range at (1-15), the results will not be affected greatly. At the same time, the percentage of outliers also has influence on the sensitivity of the results to the choice of window radius. The higher the percentage of outliers, the more the window radii influence the results.

Experiment 5.

In this experiment, we will give two real images to show the ability of MDPE to tolerate large percentage of outliers. The window radius was set 2.0 (for line fitting) and 7.0 (for circle fitting).

The first example is to fit a line in the pavement shown in figure 7. The edge image was obtained by using Canny operator with threshold 0.15 and included 2213 data points (shown in figure 7 (b)). There were about 85% outliers (most belonging to pseudo-outliers which had structures and belonged to other lines) in the data. Four methods (MDPE, RESC, ALKS, and LMedS) were applied to fit a line in the pavement. As shown in figure 7(c), both MDPE and RESC correctly found a line in the pavement. However, ALKS and LMedS failed to correctly fit the line.

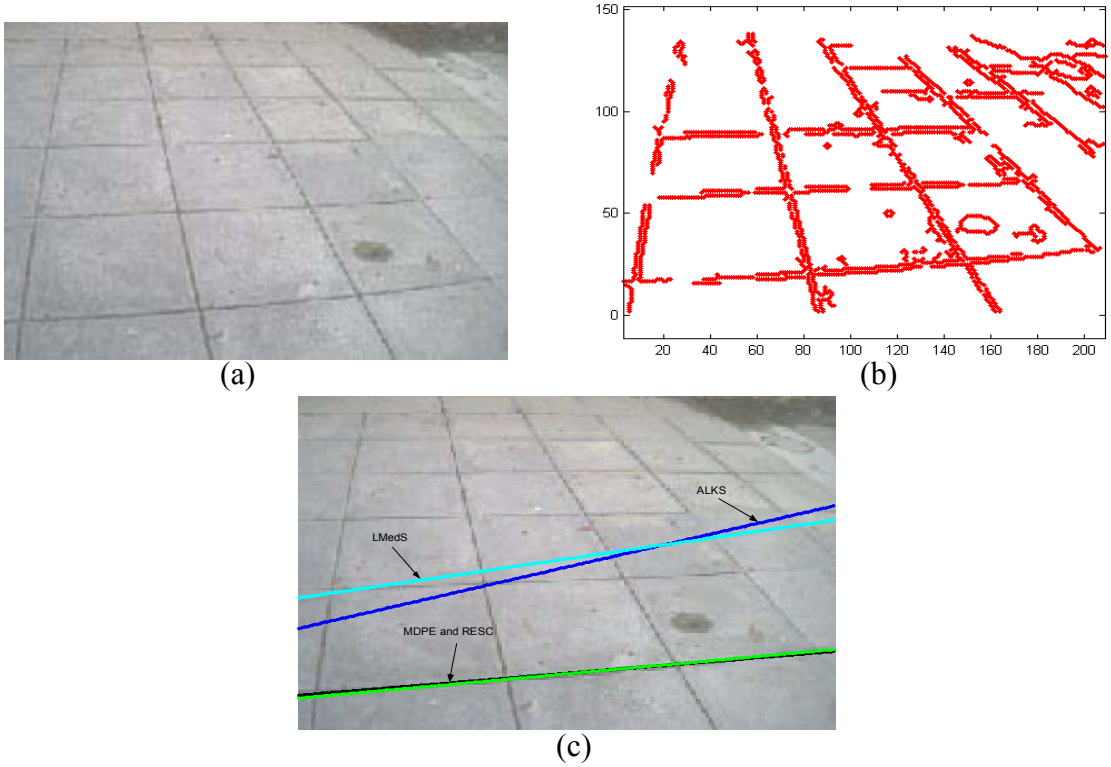
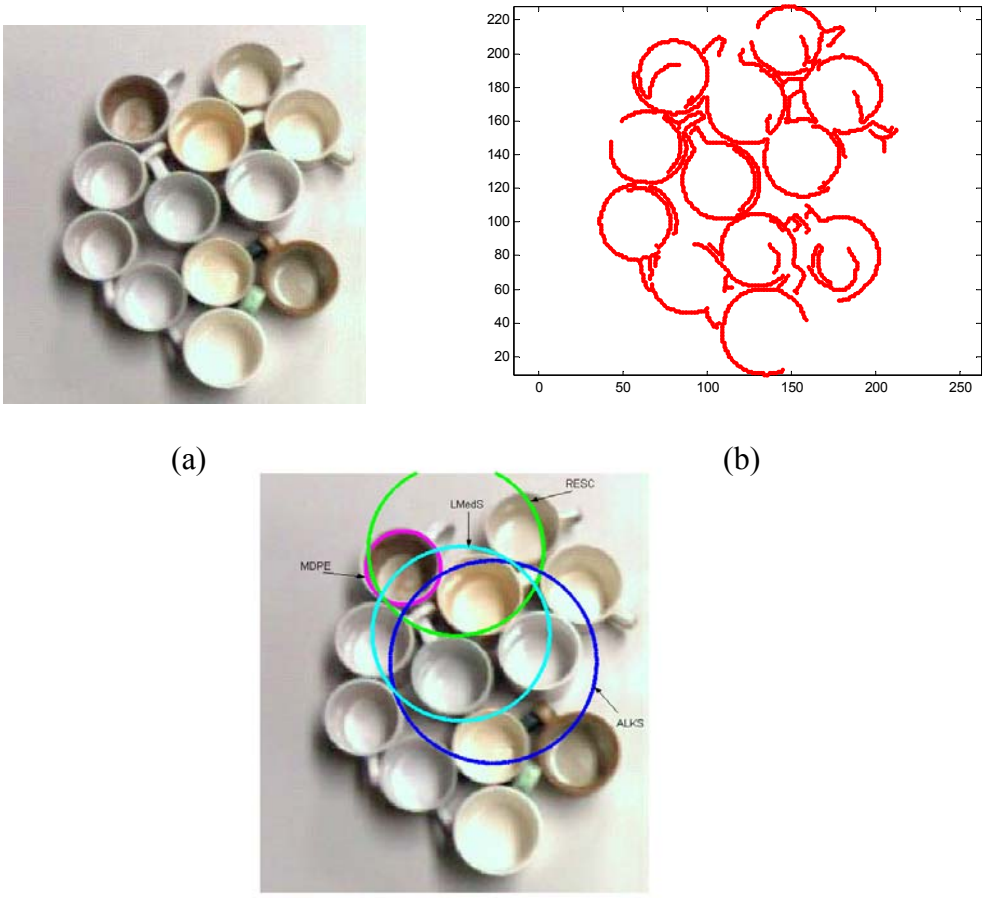


Figure 7. Fitting a line (a) one real pavement; (b) the edge image by using Canny operator; (b) the results of line fitting obtained by four methods.

The second example is to fit a circle edge of one cup out of twelve cups. Among the total 1959 data points, the inliers corresponding to each cup were less than 10% of the total data points. This is a multiple-solution case: the fitted circle can correspond to any cup in the twelve cups. As shown in figure 8, only MDPE correctly found the cup edge. However, all other three methods failed to fit the circle edge of the cup.



(a) (b)

(c)

Figure 8. Fitting a circle edge. (a) twelve cups; (b) the edge image by using Canny operator; (c) the results of circle fitting obtained by four methods.

6. Modifying MDPE to Quick MDPE (QMDPE)

As shown in section 5, the MDPE has very high breakdown point and can tolerate a large percentage of outliers including gross noises and pseudo-outliers. However, the time cost needed to calculate the densities $\hat{f}(X_i)$ of all data points within the converged window W_c is large when the number of the data points is huge. It takes O(n) time to calculate the density $\hat{f}(X_i)$ at one point X_i. If there are nw data points within the converged window W_c, the time complexity of computing the probability density power function ψ_{DP} is O(n*nw). In range image processing, nw may be tens of thousands to hundreds of thousands. For such huge range data points, the MDPE is not computationally efficient. A quicker version of MDPE with similar higher breakdown point to outliers is needed for range image segmentation.

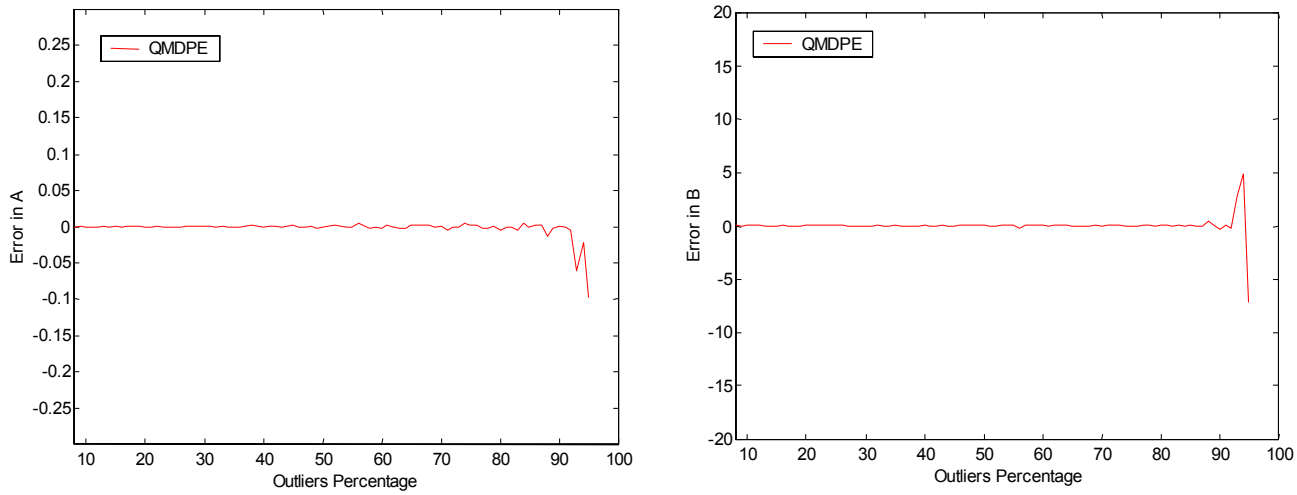
In this section, we will modify our MDPE to a quicker version, called QMDPE.

Unlike the MUSE, the RESC, the MINPRAN, etc, our MDPE does not require the outliers to be uniformly distributed. The MDPE was originally designed to extract specified primitives from the data that may include multiple structures and noises. The outliers could include gross errors and pseudo-outliers. The assumption that MDPE depends on is that the data corresponding to the fit we want to estimate are in the relative majority. These relative majorities can be less than 50%. If this assumption is satisfied, it is obvious that when the model to fit is correctly estimated, the center of the converged window in residual space by the mean shift procedure should be as close to zero as possible. At the same time, the probability density $\hat{f}(\mathrm{X_c})$ of the point Xc (corresponding to the center of the converged window) should be as high as possible. Then we define the probability density power function, which uses only one point's probability density, as follows:

$$\psi_{\mathrm{DP}} = \frac{\left(\hat{f}(Xc)\right)^{\alpha}}{\exp(|C|)^{\beta}} \tag{9}$$

Where α, β are the parameters that adjust the relative influence of the probability density and the residual of the point corresponding to the center of the converged window. They are empirically set 2.0 and 1.0.

Here we notice that only the probability density on the point corresponding to the center of the converged window need to be calculated. Thus, the time cost to compute the probability density power for each p-tuple is greatly reduced.



(a) (b)

Figure 9. Breakdown plot for the QMDPE method: (a) error in A vs. outlier percentage; (b) error in B vs. outlier percentage.

Now, we compare the QMDPE breakdown point with other four estimators (including the LMedS, ALKS, RESC, and the MDPE as shown in Figure 5). From figure 9 (we repeated 20 times), we can see that the QMDPE began to breakdown when outliers involved more than 92% of the data. However, even when outliers occupied more than 92% of the data, the QMDPE has still more than 70% correct estimation rate. The breakdown point of the QMDPE is higher than that of the LMedS (50%), the ALKS (79%), and the RESC (87%) methods. Although its breakdown point is a little lower than the MDPE, the QMPDE is much faster than the MDPE because it saves a lot of time in calculating the probability density power for each randomly sampled p-tuple.

7. Applying QMDPE to Range Image Segmentation

To test utility of QMDPE, we apply it to range image. However, segmentation is a (surprisingly) complex task and an estimator cannot simply be applied directly without considering:

1. The computational cost. QMDPE is an improved (in speed) MDPE. Its computational cost is much less than MDPE's computational cost. Even so, for a range image with a large number of data points (262,144 data points in our case), employing a hierarchical structure in our algorithm greatly optimise the computational speed.
2. Handle intersections of surface. When two surfaces intersect, the intersection line may be possibly assigned to either surface. When one surface is found first, it is hard to say which surface the intersection line should belong to (see figure 10). In fact, the intersection line is on the both surfaces and the data points are inliers to the both surfaces. Additional information (such as the normal to the surface at each pixel) should be used to handle data near the intersection line.
3. Handle virtual intersection. It is popular in model-based methods to directly estimate parameters of a primitive; and then find a surface belonging to the primitive according to the estimated parameters; then the data points on the surface will be masked and not be processed in later procedure. However, some times two surfaces do not actually intersect, but the extension of one surface is intersected by the other surface. In this case, the connected component algorithm (Lumia et al., 1983) should be employed to avoid that same labels for regions that are not connected. The component with the maximum number of data points will be chosen.

4. Remove the isolated outliers. When all surfaces are estimated, some isolated outliers, due to the noise introduced by range image camera, may remain. At this stage, a post processing procedure should be made to eliminate the isolated outliers.

The originators of other novel estimators (e.g. ALKS, RESC, MUSE, MINPRAN) have also applied their estimators to range image segmentation, but they have not generally tackled all of the above issues. Hence, even those interested in applying ALKS/RESC or any other estimator to range image segmentation may find several of the components of our complete implementation independently useful.

Now, we will present a model-based top-down algorithm, based on the QMDPE, for range image segmentation. Our method directly extracts the required primitives from the raw images. The presented algorithm deals with the whole image as raw image, and it is very robust to noisy or occluded data due to the adoption of robust estimator QMDPE. Because we adopt hierarchical techniques, this makes the algorithm computationally efficient and makes it possible to deal with large size range images with only small extra computational cost.

7.1 The background of Range Image Segmentation methods

Perception of surfaces in the images has played a very important role in image understanding and three-dimensional object recognition. Three are many three-dimensional image segmentation methods published in the literature. Generally speaking, these segmentation methods can be classified into two major classes:

1. Edge-based segmentation techniques (Ghosal and Mehrotra, 1994; Wani and Batchelor 1994).
2. Region-based segmentation techniques or clustering techniques (Fitzgibbon et al., 1995; Hoffman and Jain, 1987; Jiang and Bunke, 1994).

In edge-based segmentation methods, it is important to correctly extract the discontinuities—surface discontinuities (boundaries and jumps) and orientation discontinuities (creases and roofs), which will be used to guide the followed segmentation process. The main difficulties that edge-based segmentation techniques meet are:

- The effectiveness of these methods will be greatly reduced when range images contain noise;

- When the edge operator mask size is increased, the computational time will be greatly increased.
- When the edge pixels detected by edge operator are not continuous (especially in noisy image), it will be difficult to link these discontinuous pixels.
- Also, the reliability of the crease edge detectors makes edge-based methods questionable.

Region-based techniques have wider popularity than edge-based techniques. The essence of region growing techniques is that it segments range images based on the similarities of feature vectors corresponding to pixels in range images. The region-based techniques first estimate the feature vectors at each pixel, and then aggregate the pixels that have similar feature vectors; and at the same time, separate the pixels whose feature vectors are dissimilar, to form a segmented region.

Like other methods, region-based methods also have some problems in applications:

- They need many parameters to control the processing of region growing. Most of these parameters are predetermined.
- The choice of initial region greatly affects the performance of most region-based methods. When the seeds are placed on a boundary or on a noise corrupted part of the image, the results will break down.
- The region boundaries are often distorted because of the noise in the range images.
- In clustering-based methods, to adaptively estimate the actual number of clusters in the range image is difficult.

The model-driven (top-down) methods are appealing because it has proved that these methods have similarities to the human cognitive process (Neisser, 1967; Gregory, 1970). The model-based methods can directly extract the required primitives from the unprocessed raw range images. However, it is difficult to directly extract specified primitives by the edge-based methods and region-based methods. The features that are used in model-driven methods are primitives. So the matching takes place very early in the recognition procession. Primitive geometric features are matched instead of similar features that are used in region-based methods. Then matches are checked for local consistency by putting some geometric constraints, e.g. distance, normal, etc. Model-based methods have been attracted more and

more attention (see Yu et al., 1994; Stewart, 1995; Miller and Stewart, 1996; Lee et al., 1998). These methods, which introduce robust statistics, are very robust to noisy or occluded data.

7.2 The Algorithm, based on QMDPE, for range image segmentation.

The proposed algorithm is a model-based method and can directly extract planar primitives. Because the QMDPE is very robust to noise, the algorithm has the advantage that can resist the influence of a large amount of random noise in the range image. Also, the proposed algorithm is robust to the presence of multiple structures. Because we employed the hierarchical structure in the algorithm, the computing time for range image segmentation is greatly reduced. Because we sequentially removed the detected surfaces one by one, the averaged time to segment range image will be affected by how many surfaces the range image includes, etc. But the computing time for the segmentation of range image will not be greatly affected by the size of the range image for the reason that we uses a sampling hierarchical technique.

The steps of the algorithm can be described as follows:

1. Mark all invalid points. Shadow pixels may occur in a structured light scanner (e.g. ABW) image, these points will not be processed in next steps.
2. Calculate the normal of each range pixel and identify the jump edge pixels.
   Although the QMDPE algorithm was designed to fit the data including noise and multiple structures, it requires that the data points of the model should occupy relative majority of the whole data. This can be satisfied in a lot of range images. However, for some very complicated range images (including many objects and surfaces), this requirement is not always satisfied. Using the information of jump edge will help to coarsely segment the range image to some small regions, and for each region, the segmentation based on the QMDPE is performed.
3. Employ a hierarchal sampling technique.
   The proposed algorithm employs a hierarchal structure based on the fact that when an image is regularly sampled, the main details will remain while some minor details may be lost. The procedure is as follows: For a 512 by 512 range image, we begin with 64 by 64 regularly sampled data. Every step to higher hierarchy, the size in each dimension (x and y) is increased by 2 until to the original size of the image to deal with, i.e., 128

by 128, 256 by 256, and 512 by 512. Using hierarchy in our algorithm greatly improves the computing speed.

4. Apply the QMDPE to obtain the parameters of the estimated primitive.
   For the current hierarchy, we use the whole sampled image as the data to deal with. And we apply the QMDPE to the data. Then, the plane parameters will be estimated. The inliers corresponding to the estimated plane parameters will be identified. At this stage, it is difficult to tell which plane, of any two intersected planes, the data that are on or near the intersection line belong to. Unfortunately, this case is not considered in the popular range image segmentation methods employing robust estimators such as RESC, ALKS and MUSE.
5. Using normals information.
   When the angle between the normal of the data point that has been classified as an inlier, and the estimated plane normal, is less than a threshold value (T-angle, 40 degree in our case), the data point is accepted for step 5. Otherwise, the data point is rejected and will be used as some of the left points for further processing. As shown in figure 10, when we did not consider the normal information, the range image was over segmented because of the intersection of two planes (pointed out by the arrow in figure 10 (b) and (c)). As comparison, we obtain the right result when we considered the normal information.
6. Using the connected component algorithm to extract the maximum connected component and label them.
   The remained unlabeled inliers will be used in next loop for further processing.
7. Select the connected component for processing in the next loop.
   For all unlabeled data points, we use jump edge information and connected component to extract the component with the maximum number of the connected data points for the next loop. When the number of the data points belonging to the maximum connected component is larger than a threshold (T-cc), we repeat step 4-6. Otherwise, we stop this hierarchy and go to the higher hierarchy until the top of the hierarchy (512-by-512).
8. At last, we eliminate the isolated outliers and assign them to the majority of their eight-connected neighbors.

7.3 Experiments in Range Image Segmentation.

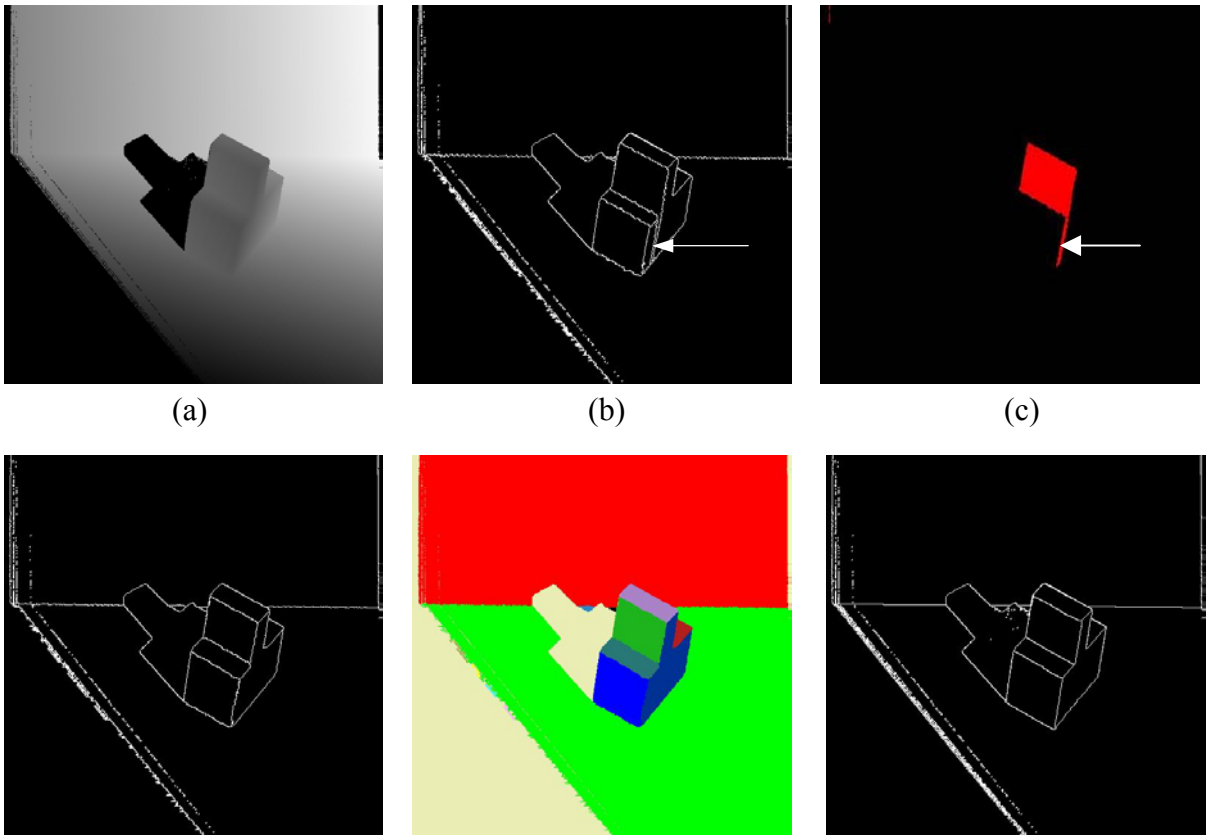


Figure 10. A comparison of using normal information or not using normal information. (a) Range image (ABW test.10 from the USF database); (b) The segmentation result without using normal information; (c) The points near or on the intersection of two planes may be classified to both planes without considering normal information; (d, e) The result using normal information; (f) The ground truth result.

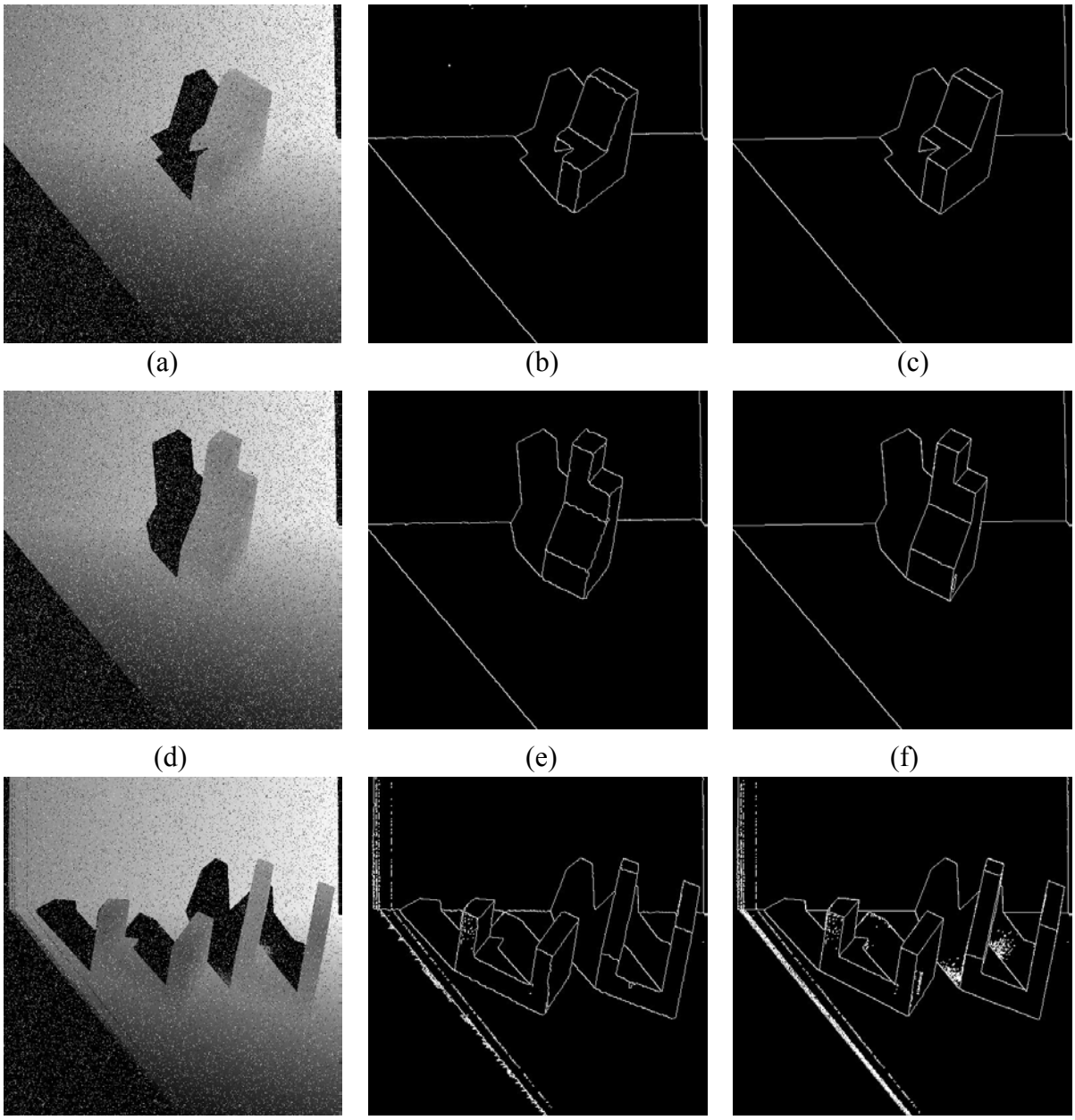


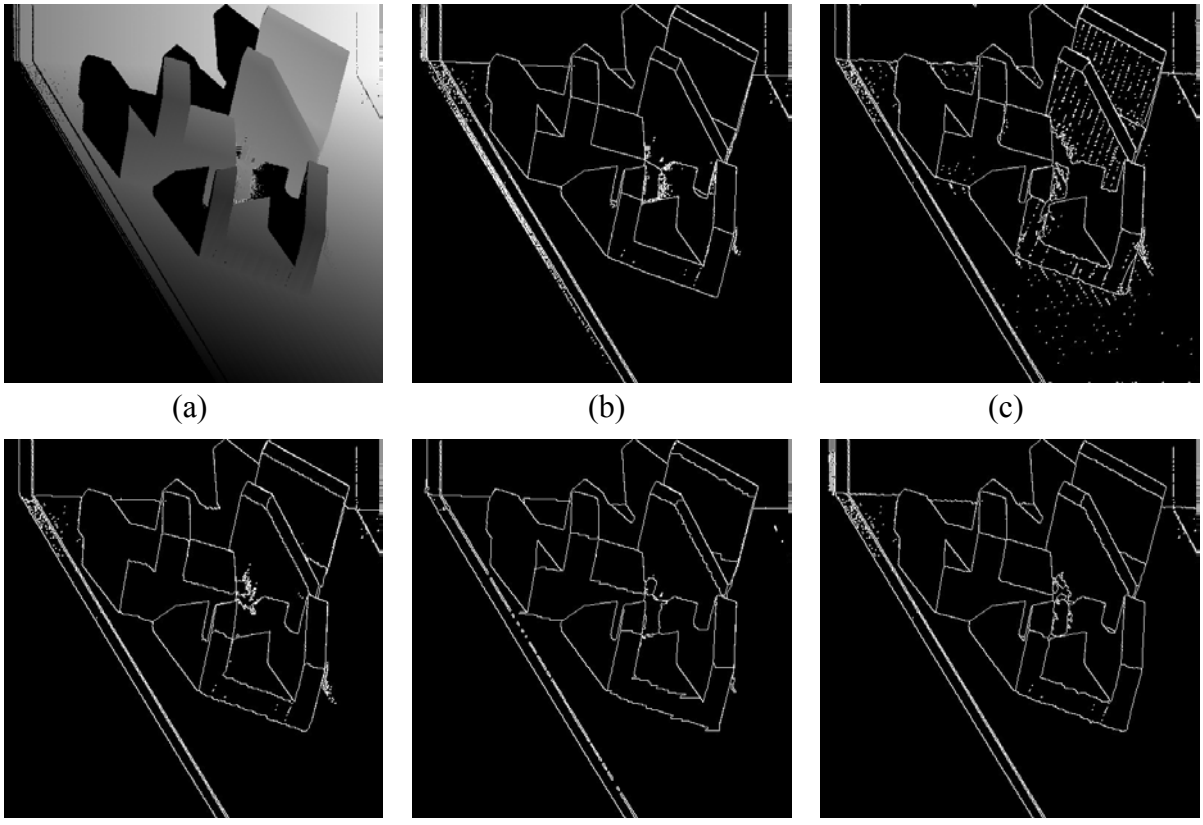
Figure 11. Segmentation of ABW range images from the USF database. (a, d, f) Range image with 15% random noises; (b, e, g) Segmentation result by the proposed algorithm; (c, f, h) The ground truth results.

Next, we will apply the proposed algorithm to real range images from the USF database.
In this subsection, we will provide examples that show how to use the proposed algorithm to segment range images. We will also compare our results with those of several popular approaches.

Since one main advantage of our method, over the traditional methods, is that it can resist the influence of noise, we put some randomly distributed noise into the range images (Note, as

the whole image is dealt with at the beginning of the segmentation, there is also a high percentage of pseudo-outliers existing in the data.)

As shown in figure 11, we add 15% randomly distributed noise, i.e. 39322 noisy points were added to each range image taken from the USF ABW range image database (test28, test27, and test13). The main surfaces were recovered by our method. Only a slight distortion appeared on some boundaries of neighbouring surfaces. In fact, the accuracy of the range data, and the accuracy of normal at each range point, will have an effect on the distortion.



(a) (b) (c)

(d) (e) (f)

Figure 12. Comparison of the segmentation results for ABW range image (test.1) from the USF database. (a) Range image; (b) The result of ground truth; (c) The result by the USF; (d) The result by the WSU; (e) The result by the UB; (f) The result by the proposed method.

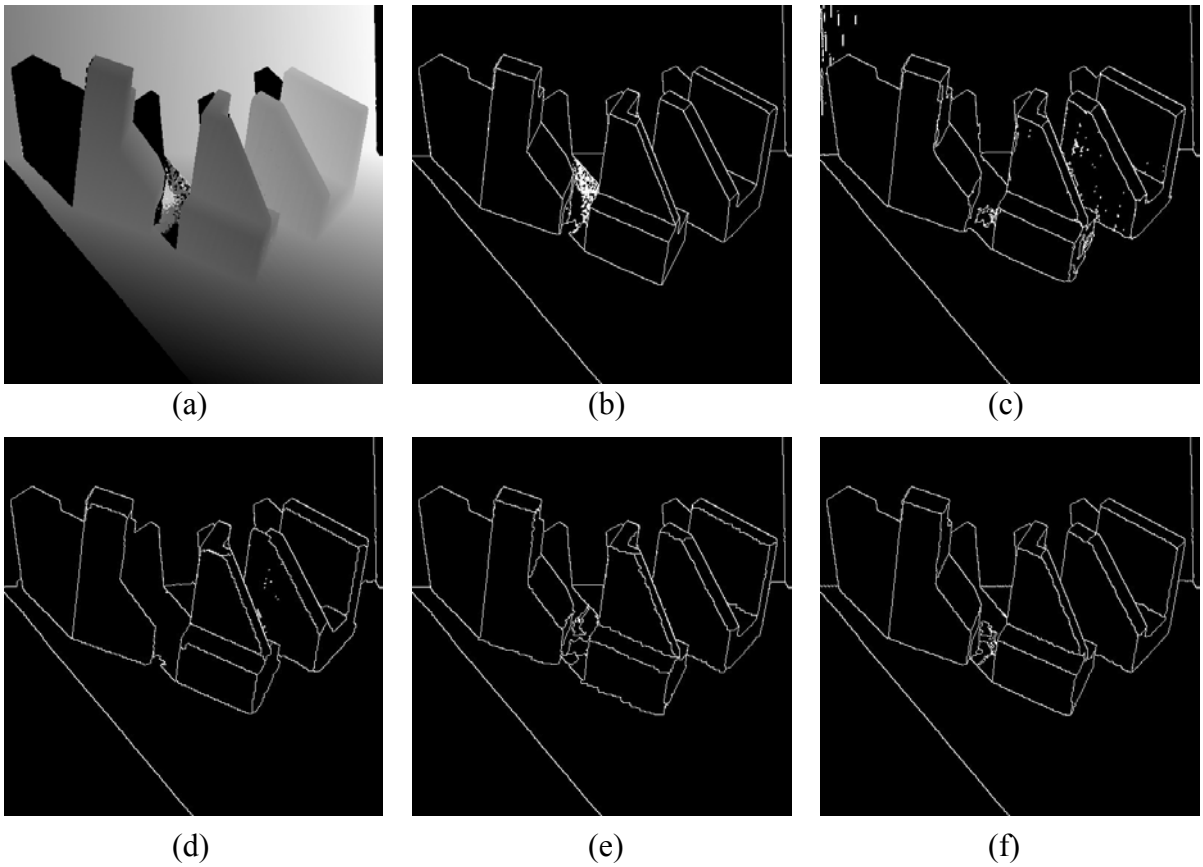


Figure 13. Comparison of the segmentation results for ABW range image (train 6) from the USF database. (a) Range image; (b) The result of ground truth; (c) The result by the USF; (d) The result by the WSU; (e) The result by the UB; (f) The result by the proposed method.

It is important to compare the results of our method with the results of other methods. In the next experiment, we will compare our results with those of the three popular range image segmenters (i.e. the USF, WSU and UB, see (Hoover, 1996)).

Consider figure 12 and figure 13: (a) is the range image and (b) is the edge map of the manually made ground truth segmentation result. The results obtained by all methods should be compared with the results of the ground truth. (c) is the results obtained by the USF. From figure 12 (c) and figure 13 (c), we can see the USF's results contained many noisy points. In both figure 12 (d) and figure 13 (d), the WSU segmenter missed surfaces. The WSU segmenter also under segmented the surface in figure 13 (d). From figure 12 (e) and figure 13 (e), we can see the boundaries on the junction of surfaces were distorted relatively seriously. Our results are shown in figure 12 (f) and figure 13 (f). Compared with other methods, the proposed method performed best. Our method directly extracted the planar primitives. In the proposed method, the parameters requiring tuning are less than other traditional methods.

Adopting hierarchy-sampling technique in the proposed method greatly reduces its time cost. The processing time of the method is affected to a relative large extent by the number of surfaces in the range images. The processing time for a range image including simple objects is faster that a range image including complicated objects. Generally speaking, it takes about 40 seconds (on an AMD800MHz personal computer) for segmenting a range image with less surfaces and about 80-100 seconds for a range image including more surfaces. This includes the time for computing normal information at each range pixel (which takes about 12 seconds).

8. Conclusion

The contributions of this paper are twofold: first we introduce a new and highly robust estimator (MPDE), and secondly we apply the new estimator to the computer vision task of segmenting range data. The latter is more than a mere application of the estimator in a straightforward manner. There are a number of issues that need to be addressed when applying an estimator (any estimator) to such a problem. The solutions we have found, to these practical problems that arise in the segmentation task, should be of independent interest. The resulting combination of a highly robust estimator and a very careful application of that estimator, produces a very effective method for range segmentation.

MPDE is similar to many random sampling estimators: we randomly choose several p-subsets, and we calculate the residuals for the fit determined by each p-subset. However, the crux of the method is that we apply the mean shift procedure to find the local maximum density of these residuals. Furthermore, we evaluate a density power measure involving this maximum density. The final estimated parameters are those determined by the one p-subset corresponding the maximum density power over all of the evaluated p-subsets. Our method, and hence our definition of maximum density power, is based on the assumption that when a model is correctly fitted, its inliers in residual space should have a higher probability density, and the residual at the maximum probability density of inliers should have a low absolute value. This captures the dual notions that: the data points having lower residuals should be as many as possible, and that the residuals should be as small as possible. In that sense, our method combines the essence of two popular estimators: Least Median of Squares (low residuals) and RANSAC (maximum number of inliers). However, unlike RANSAC, MDPE

scores the results by *the densities of data points* falling into the converged window *and on the size* of residual of the point corresponding to local maximum density. Contrast this also with the Least Median of Squares, which uses a single statistic (the median).

The result of our innovation is a highly robust estimator. The MDPE can tolerate more than 85% outliers, and has regularly been observed to function well with even more than 90% outliers.

We also developed a quicker version—QMDPE. The advantage of the QMDPE is in that only the probability density corresponding to the center of the converged mean shift window needs to be calculated, therefore the time cost to compute the probability density power is greatly reduced. Although the QMDPE has a lower breakdown point than the MDPE, this difference is very small. The QMDPE still has a higher breakdown point than most available estimators (such as M-estimators, LMedS, LTS, RNSAC, ALKS, and RESC). We recommend that when the data points are small (say less than 5000 points) and the task has a high reliance on the breakdown point of the estimator, then MDPE is an ideal choice. On the other hand, when the task involves a large number of data points (for example, range image segmentation which often involves more than tens of thousands of data), and the speed is a relatively important factor to consider, it is better to choose the QMDPE rather than the MDPE.

The MDPE/QMPDE is an unsupervised method. However a crucial parameter we need to choose is the window radius, h. Normally, the results will be stable for a reasonable range when the window radius is changed from small to large (Fukunaga, 1990, p.541). This range is larger when the percentage of outliers is smaller. According to those authors, the optimal window radius can be decided by the center of the largest operating range that yields the same parameters for a given data. We have yet to investigate this issue.

When the percentage of outliers is very large or there are many structures in the data (pseudo-outliers), one problem in carrying out the methods which use random sampling techniques is: the number of p-subsets to be sampled, m, will be huge. For example, if we require the probability to have at least one “clean” p-subset is 0.95 and the percentage of contaminated data ε is 90%, then the value of m is 2994 for p=3; and 2995700 for p=6! Thus, a more feasible technique for sampling p-subsets is needed for fitting a data with a large number

percentage of outliers and multiple structures. This is true of all methods based upon random sampling techniques to gain robustness and the solution may lie in heuristic approaches such as structured sampling or divide and conquer techniques. This aspect is reserved for future study.

Acknowledgements

This work is supported by the Australia Research Council (ARC), under the grant A10017082. We thank Xiangyi Jiang, Patrick J.Flynn for their code and results. We also thank Raymond Jarvis and Jaesik Min for their kind assistance.

References

Bab-Hadiashar, A. and Suter, D. 1999. Robust segmentation of visual data using ranked unbiased scale estimate. ROBOTICA, International Journal of Information, Education and Research in Robotics and Artificial Intelligence, 17, pp. 649-660

Cheng, H. and Meer, P. May 2002, Robust Computer Vision through Kernel Density Estimation, A. Heyden et al. (Eds.), ECCV 2002, LNCS 2350, Springer-Verlag Berlin, pp. 236-250.

Cheng, Y. 1995. Mean Shift, Mode Seeking, and Clustering. IEEE Trans. Pattern Analysis and Machine Intelligence, **17(8),** pp.790-799.

Comaniciu, D. and Meer, P. June 1997. Robust Analysis of Feature Spaces: Color Image Segmentation. in Proceedings of 1997 IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR, pp.750-755.

Comaniciu, D. and Meer, P. September 1999. Mean Shift Analysis and Applications. in Proceedings 7th International Conference on Computer Vision, Kerkyra, Greece, pp.1197-1203.

Comaniciu, D. and Meer, P. 2002. Mean Shift: A Robust Approach towards Feature Space A

Nalysis. IEEE Trans. Pattern Analysis and Machine Intelligence, **24(5)**, pp. 603-619.

Davis L.S. (Ed), 2001. Foundations of Image Analysis. pp.323-347, Kluwer.

Fischler, M. A. and Rolles, R. C. 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Commun. ACM, **24(6)**, pp.381-395.

Fitzgibbon, A.W., Eggert, D.W., and Fisher, R.B. 1995. High-Level CAD Model Acquisition from Range Images. Technical Report, Dept. of Artificial Intelligence, Univ. of Edinburgh.

Fukunaga, K. and Hostetler, L.D. 1975. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. IEEE Trans. Info. Theory, vol. IT-21, pp. 32-40.

Fukunaga, K. 1990. Introduction to Statistical Pattern Recognition, Boston: Academic Press.

Ghosal, S. and Mehrotra, R., 1994. Detection of Composite Edges. IEEE Trans. PA, **13(1),** pp.14-25.

Gregory, R.L., 1970. The intelligent Eye. New York:McGraw-Hill.

Hoffman R.L. and Jain A.K. 1987. Segmentation and Classification of Range Images. IEEE Trans. Pattern Analysis and Machine Intelligence, **9(5),** pp.608-620.

Hoover, A., Jean-Baptiste, G., Jiang., X., et al. 1996. An Experimental Comparison of Range Image Segmentation Algorithms. IEEE Trans. Pattern Analysis and Machine Intelligence, **18(7),** pp.673-689.

Huber P.J. 1973. Robust Regression: Asymptotics, Conjectures and Monte Carlo. Annals of Statistics, **1**, pp799-821.

Huber P.J. 1981. Robust Statistics, New York, Wiley.

Jiang, X.Y. and Bunke, H. 1994. Fast Segmentation of Range Images into Planar Regions by Scan Line Grouping. Machine Vision and Applications, **7(2)**, pp.115-122.

Lee, K-M., Meer, P., and Park, R-H. 1998. Robust Adaptive Segmentation of Range Images. IEEE Trans. Pattern Analysis and Machine Intelligence, **20(2),** pp.200-205.

Lumia, R., Shapiro, L., and Zungia, O. 1983. A New Connected Components Algorithm for Virtual Memory Computer. Computer Vision, Graphics, and Image Processing, **22(2)**, pp. 287-300.

Miller J.V. and Stewart, C.V. June 1996. MUSE: Robust Surface Fitting Using Unbiased Scale Estimates. Proc. Computer Vision and Pattern Recognition'96, pp.300-306, San Francisco.

Neisser, U., 1967. Cognitive Psychology, New York: Appleton-Century-Crofts.

Rousseeuw, P.J. 1984. Least Median of Squares Regression. J. Amer. Stat. Assoc. **79**, pp. 871-880.

Rousseeuw P.J. and Leroy A., 1987. Robust Regression and outlier detection. John Wiley & Sons, New York.

Siegel, A.F. 1982. Robust Regression using Repeated Medians. Biometrika, **69**, pp.242-244 .

Silverman, B.W. 1986. Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

Stewart, C.V. 1995. MINPRAN: A New Robust Estimator for Computer Vision. IEEE Trans. Pattern Analysis and Machine Intelligence, **17(10),** pp.925-938,

Wand, M. P. and Jones, M. 1995. Kernel Smoothing, Chapman & Hall.

Wani, M.A., and Batchelor, B.G. 1994. Edge Region Based Segmentation of Range images. IEEE Trans. Pattern Analysis and Machine Intelligence, **16(3),** pp.314-319.

Yu, X., Bui, T.D., and Krzyzak, A. 1994. Robust Estimation for Range Image Segmentation and Reconstruction. IEEE Trans. Pattern Analysis and Machine Intelligence, **16(5),** pp.530-538.