The sensitivity of RANSAC iterations to measurement noise and facet shape

Joost Toben (4018656) and Eva Linssen (3902749)

Abstract

The theoretical estimate on the number of iterations that RANSAC needs to find a plane with a certain probability depends on the number of inliers, the number of outliers, and this probability. It assumes that the inliers lie exactly on the plane, and hence were measured with no measurement noise. In this short paper we will examine by experiments on artificial data how the number of iterations is affected when there is measurement noise. In urban reconstruction the model to be found is not a full plane but some facet on a plane. With measurement noise, the shape of this facet with may also affect the number of iterations, which we also examine experimentally.

Introduction

To determine parameters of an unknown model, a technique called RANdom Sample Consensus (RANSAC) can sometimes be used. It takes observed data and assumes that the model parameters may already be determined from just a few observations. RANSAC is an iterative procedure that tries to find the best-fitting model by trying many, and determining the support for each model from all observations. The model (parameters) with most support is returned. RANSAC was introduced by Fischler and Bolles [1]. The particular use of RANSAC for building reconstruction was studied by Schnabel et al. [2], among others.

When RANSAC is used to determine the plane in a 3D point set P that contains the maximum number of points of P, we have to set a number of iterations that makes us sufficiently certain that the best plane we have found so far is in fact the plane with the most points. If P has n points, and the number of points inside the plane with the most points is k, we call k/n the inlier ratio. To have a probability p that we found the plane with the most points after r iterations, we have the expression $p = 1 - (1 - (k/n)^3)^r$. This expression can be rewritten to know the value of r needed in the algorithm. However, it is valid only if there is no measurement noise, but data acquired with LiDAR will typically have 5 cm measurement noise. Hence the expression is incorrect in practice. We will examine how r depends on the measurement noise and also on the plane shape. To this end, we will run experiments on artificial data.

We phrase the research questions precisely:

1. Does the number of iterations required to find the best plane increase when noise is present, and how does it increase in the amount of noise?
2. When noise is present, does the facet shape influence the number of iterations that is needed, and in what way?

For both questions we assume that the probability with which we want to find the best plane is fixed to a constant. We aim to know r for p=0.95.

Experiment set-up

To make our artificial data correspond to a possible scenario in the real World, we choose our point set P to lie in a box of 50x50x20 meters, where 20 meters is the height. We will place n-k points uniformly at random in the box, and another k points on the model. Then we perturb every one of the k points on the model by making a ball with radius b around it, and randomly replacing the point inside its ball, with uniform distribution.

The set-up has four aspects that can be varied to generate the test data: the values k and n, the shape of the model, and the radius b of the ball. We will use the following settings:

- n = 8000
- k = 500 or k = 1000
- The shapes are rectangles of different aspect ratios but the same area: 6x6, 9x4, 12x3, 15x2.4. We place them as rectangles horizontally in the middle of the box.
- The radius b of the ball is 0, 1, 2, 3, 4 cm.

This leads to 40 different setting combinations, called settings from now on, that all need to be repeated sufficiently often to get an estimate on r.

To run RANSAC, we need one more parameter, namely the distance to the plane (model) that is allowed for a point to be counted as support for that plane. We will choose this distance to be 4 cm. If we would choose it smaller than 4 cm, the setting with b = 4 cm might not even contain any plane that gets support from all k points originally placed on the facet.

Notice that we may find a plane with support higher than k. The 8000 randomly placed points could lie such that several of them support the plane as well. We will say that we found the plane when we get a support of at least 90% of k, so either 450 or 900, depending on the choice of k. We can then stop this test and record the number of iterations needed.

For every setting we do the following. We run RANSAC and count how many iterations are needed until we find a plane with the desired support. We do this 100 times and then choose as the estimate of r, the smallest number of iterations so that 95 of the runs found a plane with sufficient support within r iterations (and 5 of the runs needed more than r iterations).

Results

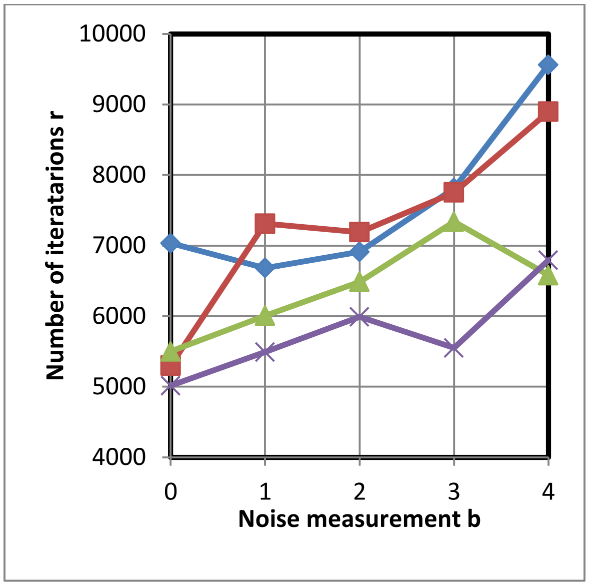

Figure 1: iterations r per noise measurement b with k= 500

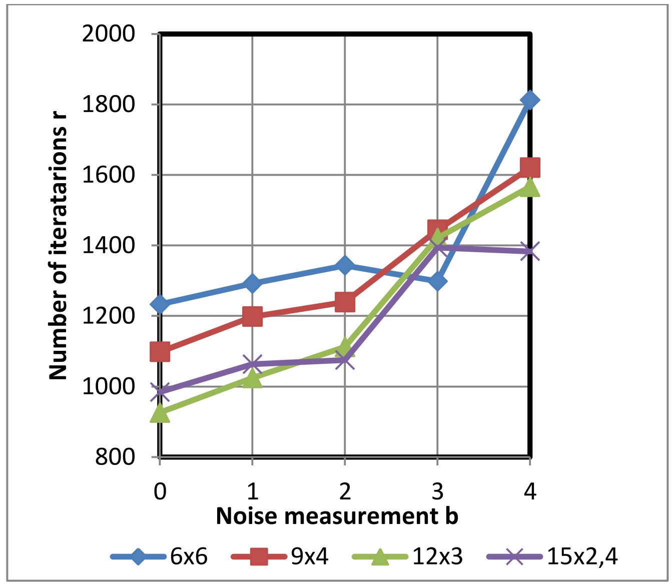

Figure 2: iterations r per noise measurement b with k= 1000

Figure 1 and 2 both show the relationship between the amount of noise and the number of iterations. The differently coloured lines represent the different plane ratios. K is respectively 500 and 1000. The x-axis depicts the noise, ranging from 0-4. The y-axis depicts the number of iterations. The number of runs per setting is 700.

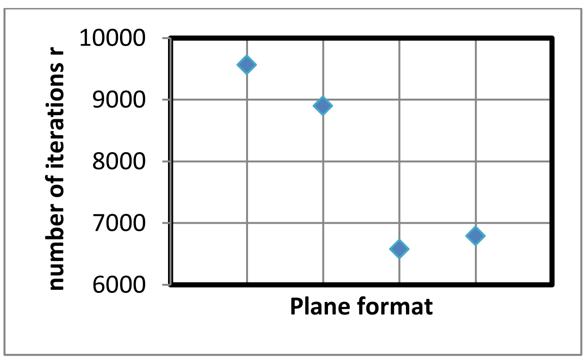

Figure 3 : Iterations per plane format with b=4 and k= 500

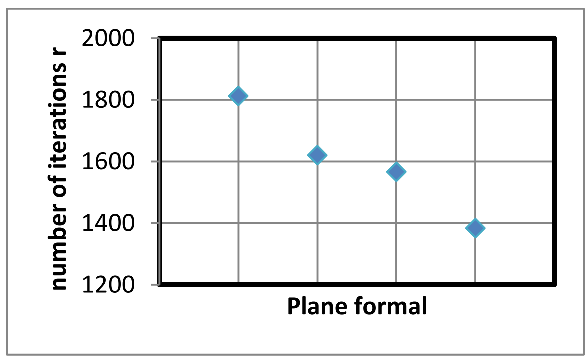

Figure 4: Iterations per plane format with b=4 and k= 1000

Figure 3 and 4 both show the relationship between the shape of the plane and the number of iterations. The x-axis depicts the shape of the plane, from left to right: 6x6, 9x4, 12x3, 15x2.4. The y-axis depicts the number of iterations. Both figures depict the data with noise = 4. K is 500 and 1000 respectively. The number of runs per setting is 700.

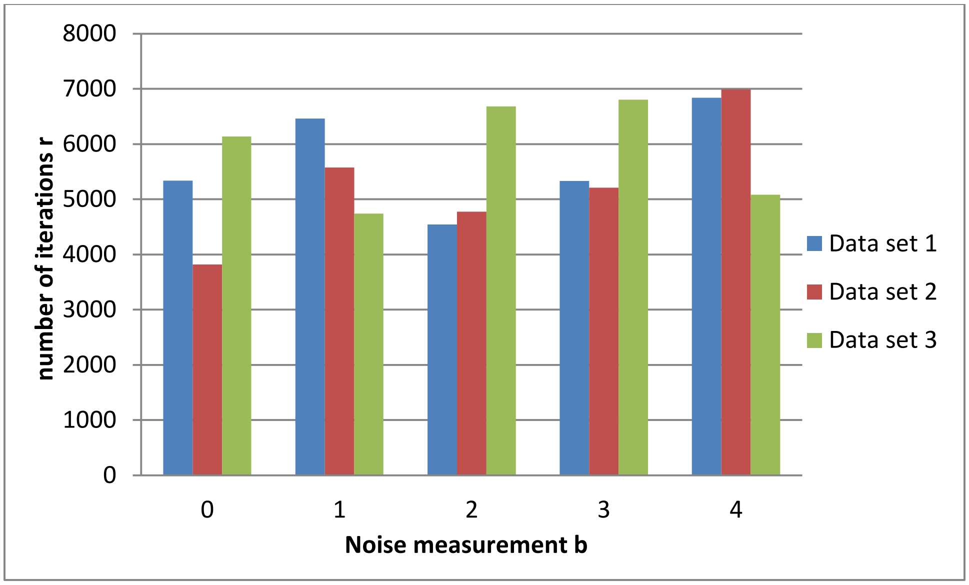

Figure 5: Iterations r per noise measurement b for the plane 15x2.4, k = 500 and several datasets

Figure 5 depicts the differences between dataset 1, 2 and 3. The x-axis depicts the amount of noise and the y-axis depicts the number of iterations. The figure depicts the differences for k = 500 and the shape of the plane is 15x2.4. The number of runs per setting is 100.

Evaluation and discussion

For our first attempt at this experiment, we performed it as described in the experiment set-up. We ran it a total of three times, resulting in three different datasets. As can be seen in figure 5, these sets varied wildly. Some variation was to be expected, but these differences were large enough to consider all data irrelevant. This caused us to reconsider the amount of runs per setting. We redid the experiment, but this time with 700 runs per setting instead of the standard 100. 700 was chosen as a compromise between our available time and the quality of the results. In the following two paragraphs, our research questions will be answered with the data gathered from the experiment with 700 runs.

To answer our first research question, ´does the number of iterations required to find the best plane increase when noise is present, and how does it increase in the amount of noise?´, we but need to look at figure 1 and 2. A clear relationship is shown between the amount of noise and the number of iterations, namely an increase in the one means an increase in the other. Some small deviations do exist, but not enough to decrease/diminish the strength of this claim. Aside from the deviations, the relationship between noise and iterations is a linear one.

As for the second research question, ‘when noise is present, does the facet shape influence the number of iterations that is needed, and in what way?’, we look at the other two figures, 3 and 4. Both show a clear decline in the amount of iterations as the shapes grow more elongated. One outlier in the data is the point corresponding with plane shape 12x3 for k = 500. However, due to this being our only dataset, and taking note of the differences between the other datasets as described above, we can’t draw any meaningful conclusions about the quality of our results, and thus the (ab)normality of this outlier.

Conclusions

For this paper, research was done to find relationships between noise and iterations and plane shape and iterations when using RANSAC. We gathered data from a program that creates its own uniformly distributed random point set , and drew the following conclusions: more noise results in an increase in iterations, and more elongated plane shapes result in a decrease.

However, we ran into some possible fallibilities during the experiment. The first and most obvious is the amount of runs per setting. We found huge differences between our data sets when we used the standard 100 runs. Due to time constraints we only have one data set with this number. A role for future research would be to see if 700 is indeed high enough to limit the amount of variation between datasets.

Another point is the way we utilized our point set . In the experiment set-up in this paper, nothing is said about this utilization. We were therefore forced to create our own implementation, which resulted in the following: For every setting we created a new point set to put the plane in. However, it is possible that the point set influences the resulting data. Future research could determine if this is case.

References

[1] Martin A. Fischler and Robert C. Bolles: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM 24 (6): 381–395 (1981).

[2] Ruwen Schnabel, Roland Wahl, and Reinhard Klein: Efficient RANSAC for Point-Cloud Shape Detection. Comput. Graph. Forum 26(2): 214-226 (2007).