The 9th International Conference "ENVIRONMENTAL ENGINEERING"
22–23 May 2014, Vilnius, Lithuania
SELECTED PAPERS
eISSN 2029-7092 / eISBN 978-609-457-640-9
Available online at *http://enviro.vgtu.lt*

Section: Technologies of Geodesy and Cadastre

Modified RANSAC transformation of coordinates burdened with outliers

Joanna Janicka, Jacek Rapinski

Uniwersity of Warmia and Mazury,Oczapowskiego 2, Olsztyn 10-719, Poland

Abstract

One of the method that can be used in the process of coordinate transformation, where the coordinate of some points may be affected by gross errors, is a method called RANSAC (Random Sample and Consensus). The stage hypothesis of this method is related to the concept of the minimal sample set (MSS) which is also called a hypothetical model. First minimal sample set is randomly selected from the input dataset and the model parameters are computed using only the elements of the MSS. Next, in testing step, RANSAC iteratively checks which observations of the entire dataset are consistent with the hypothetical model.
The authors presented the possibility of using the RANSAC method in the process of transformation parameters estimation with some modification. The authors propose the use of some functions during the selection the MSS from the input data set so that they were not chosen randomly.

Keywords: RANSAC; transformation of coordinates; parameter estimation.

1. Introduction

The transformation of coordinates allows to convert coordinates from one geodetic system to another. Usually the determination of transformation parameters is performed by the means of the least-squares method. Unfortunately, the least squares method is not immune to outliers. In geodetic practice, each observation is measured with an unknown true error. So in coordinate transformation there are also situations when the coordinates of some reference points are burdened with outliers. This outliers should be identified and eliminated from the process of transformation parameter estimation. One can also decrease their influence on the estimated parameters by applying special estimation methods e.g.: robust M-estimation [3], [5] R-estimation [1] or the Msplit estimation [6]. The application of robust estimation methods in the process of coordinate transformation tested and confirmed the effectiveness of these methods. The biggest problem in the use of these methods is the total number of points which are burdened with gross errors. If the number of outliers is greater than 40–45%, these methods do not generate the correct results [4]. In this paper, the general RANSAC algorithm and modification of the RANSAC transformation are described.

2. RANSAC algorithm

Fischler and Bolles [2] proposed the RANSAC algorithm in 1981 as a method to estimate the parameters of a certain model, starting from a set of data contaminated by large amounts of outliers. It is an iterative algorithm which uses least-squares method to estimate model parameters. The basic premise of RANSAC is the presence in the data set of both observations that fit the model (inliers) and those which differ from the values (outliers). The source of data that do not fit into the model are gross errors (measurement errors), noise or other disturbances. A set of data and a model that will be matched to the data set are the input data of the algorithm. The advantage of this method is that the percentage of outliers which can be handed by RANSAC can be larger than 50% of the entire data set. Such a percentage, know also as the "breakdown points", is commonly assumed to be the practical limit for many other commonly-used techniques for parameter estimation such as a robust estimation method for example, for the M-estimaton methods, R-estimation). Information, which is used in the process of RANSAC estimation:

Corresponding author: Joanna Janicka. E-mail address: joannasuwm@gmail.coml
http://dx.doi.org/10.3846/enviro.2014.210

- the minimum number of points (observations) required to determine the model parameters,
- the minimum number of iterations,
- parameters determining the extent to which data are correct (inliers),
- the size of the data set, which completes the process of the iteration.

The RANSAC algorithm is essentially composed of two steps: hypothesis and tests that are repeated in an iterative process.

- **Hypothesis**. The first minimal sample set (MSS) is randomly selected from the input dataset and the model parameters are computed using only the elements of the MSS. The MSS is the amount of data (observations) required to clearly define the model (the minimum number of observations that is required to describe the model is equal to the number of model parameters). The minimum number is determined by the selected function describing the model. In the Helmert transformation, the minimum number of total points is 2 (to set four parameters, transformation is necessary to know the x and y coordinates of the two points). Thus, the first phase starts with selecting a necessary and minimum number of observations of the data set. Based on these selected observations, the output model (hypothetical) is estimated. All of the remaining data are tested in terms of fit to the hypothetical model.
- **Testing.** In the second step, RANSAC iteratively checks which observations of the entire dataset are consistent with the hypothetical model. This requires determining the value of the parameter m_d specifying the maximum distance from a test point to a hypothetical model. If it fits the criterion of m_d, the point is treated as just another hypothetical inliers. The minimum percentage of observations that must be the correct data in the whole data set is also defined (for example, the model can be regarded as properly defined if 80% of the observations are those that are not burdened with outliers). The estimated model is correct if it has a sufficient number of points that have been classified as correct observations (inliers). The best set of observations which is selected from the entire dataset is called the consensus set (CS). Defining an iteration as a single process of random selection of MSS and fit testing, the number of iterations is determined by the following formula

$$T_{iter} = \frac{\log \varepsilon}{\log(1-q)} \tag{1}$$

where:

ε – probability of incorrect identification of the model,

q – is calculated based on the following equation:

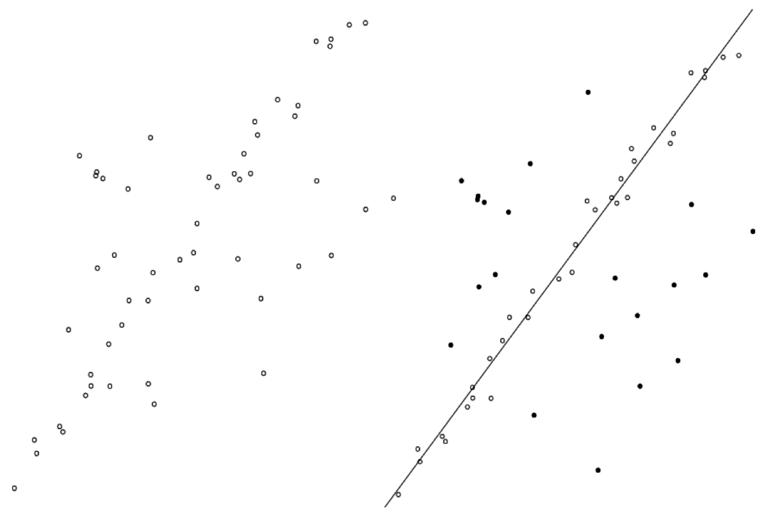
$$q = \left(\frac{N_i}{N}\right)^k \tag{2}$$

in which: N_i – the number of points that belong to the consensus set,

N – the total number of points,

k – minimal number of data that are necessary to clearly define the model.

Figure 1 presents an example of the application of RANSAC to estimate line parameters. The data set containing correct observations (inliers) and observations burdened with gross errors (outliers) is shown in Figure 1. Figure 1(a) presents the entire input data set, from which the correct observations describing a straight line are selected. Figure 1(b) shows the line that is based on the correct observation (inliers) selected by the RANSAC algorithm and the solid dots are the outliers that do not fit the model.



(a) (b)

Fig. 1. (a) A data set with many outliers for which a line has to be fitted (b) Line fitted with RANSAC

3. Modfified RANSAC application in transformation parameter estimation

3.1. RANSAC transformation

The RANSAC algorithm can be used in the process of coordinate transformations, especially if outliers are present in the entire set of reference points. To confirm the effectiveness of the proposed method, the Helmert transformation was adopted. The determination of transformation parameters is commonly performed by the least-squares method using all available points. But the least squares method is a neutral method, so is not immune to outliers. It means that if, for any reason, one or more of the reference point coordinates is not correct, the transformation parameters will be estimated with this error. Such observations must be identified and eliminated from the data set before performing coordinate transformations. In this paper, the RANSAC algorithm for coordinate transformation is proposed to complete this task.

In this algorithm, the transformation parameters are still estimated by the least squares method but the RANSAC algorithm is also used during the parameter estimation process. RANSAC algorithm select from entire set of observation only those points that are not outliers. The greatest benefit of this approach is that the percentage of outliers can be larger than 50% of the entire set of reference points. The RANSAC transformation is an iterative process and is described as follows:

1. Helmert transformation in two-dimensional space is a four-parameter transformation with the minimum number of reference points. So in the first step, two reference points are randomly selected from the entire set of reference points and the transformation parameters are calculated by solving a set of four linear equations. This is the step of creation of the hypothetical model.
2. The next step is the transformation of reference points coordinate from one coordinate system to another with the hypothetical transformation parameters (from the first step, on the basis of the hypothetical model).
3. Then the parameter m_d is defined, the minimum number of iterations and the minimum number of points (observations) required to fit the model. This is an arbitrary parameter which corresponds to the maximum allowable residual after adjustment, so the value of this parameter depends on the required accuracy of transformation.
4. The distances to each point from the points transformed are calculated using a hypothetical model and testing conditions: if $d < md$ then the point is added to the minimal set of reference points described in the first step, then if the number of inliers is sufficient (e.g. min 80% of the entire set of total points).
5. If all conditions from step 4 are satisfied, the iteration process is finished. The Consensus Set (CS) is then defined and based on it, the transformation parameters are re-estimated by the least-squares method. If conditions are not fulfilled, then RANSAC algorithm once again selects the minimum number of points to define a hypothetical model and the procedure 1–4 is repeated.
6. The final step is the transformation of coordinates from one geodetic system to another using the CS model parameters.

3.2. Modified RANSAC transformation

In the RANSAC transformation described in this section two reference points are randomly selected from the entire set of reference points to create the hypothetical model MSS. Then on the basis of MSS the estimation parameters are calculated. But there is a risk that the algorithm will not select the proper two points for the best solution (RANSAC chooses randomly so there is no control which two points will be selected). During the transformation of coordinates, the location of the reference points is very important. Reference points should cover evenly transformed object and its surroundings. Therefore, it is necessary to introduce a modification in the algorithm of RANSAC transformation. The authors propose a modification of the RANSAC algorithm during the selection the MSS from the input data set so that they were not chosen randomly. So the first point is selected randomly, but the second point is selected from the other reference points on the basis of the distance. The distance between those two points should be the largest. This change in the algorithm would avoid the incorrect location of the reference points.

4. Example

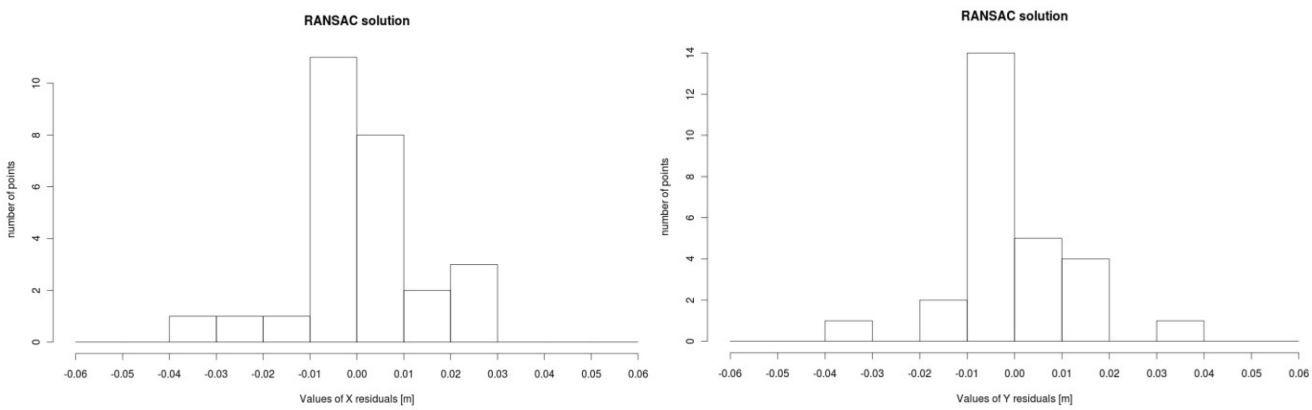
To confirm the effectiveness of the proposed algorithm performed two types of the calculations. In the section 4.1 the effectiveness of the proposed RANSAC transformation is presented. In the section 4.2 it is shown that despite the advantages of this method, it has some flows. It is not an efficient method from a computational point of view. It requires many iterations and many operations which sometimes (especially in the case of large sets) take a longer time than the standard procedure. There is also a risk that the algorithm will not select the proper two points for the best solution (which depends on the *a priori* selection of parameters).

An analysis performed using a horizontal geodetic network. The object includes 150 points. The coordinates of these points are defined in the local coordinate system and in the Polish Coordinate System “2000” where 54 of them are reference points.

4.1. RANSAC transformation calculation

The transformations were performed in two scenarios with different numbers of reference points burdened with gross errors and different value of these errors. The artificial errors were added to both X and Y coordinates.

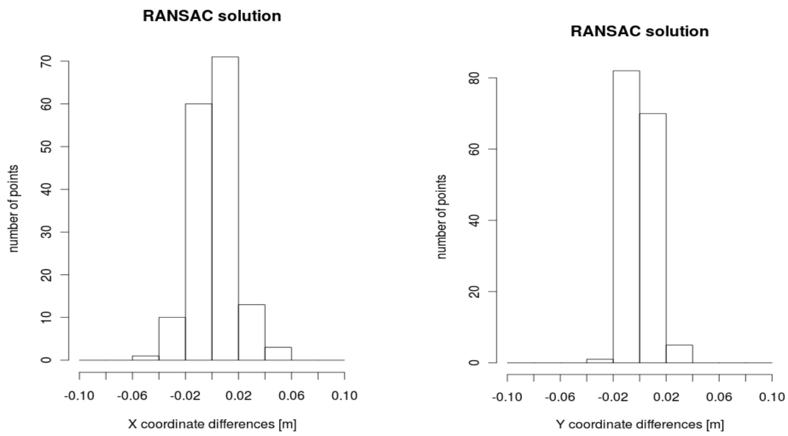
The first scenario assumed that 50% of the reference points coordinate include outliers of magnitude of 0.15 to 0.30 m. The gross errors were added to the X and Y coordinates of 27 out of 54 reference points. The values of the residuals obtained in this scenario are shown in the Figure 2.



(a) (b)

Fig. 2. (a) Histogram of X residuals (b) Histogram of Y residuals

The Figure 2 presents the histogram of X and Y residuals obtained in the first scenario. In this scenario, the values of the residuals are between –0.04 and 0.04 m. The histograms of the differences between coordinates after transformation and catalogue values are shown in the Figure 3.

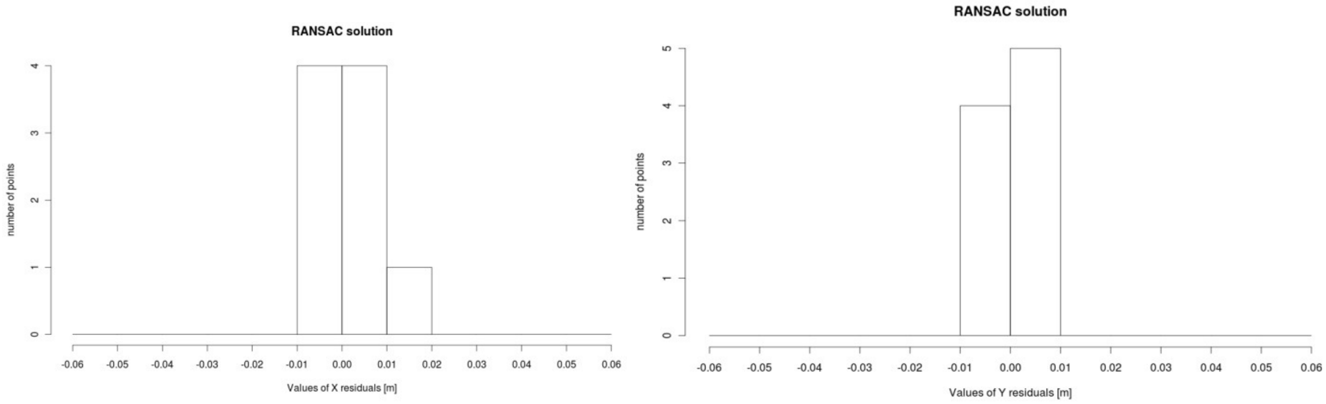


(a) (b)

Fig. 3. (a) The histogram of the X coordinate differences (b) The histogram of the Y coordinate differences

The Figure 3 presents the histogram of the X and Y coordinate differences obtained in the first scenario. The values of these differences are between –0.06 and 0.06 m. However, almost 80% of the differences are less than ± 0.02 m.

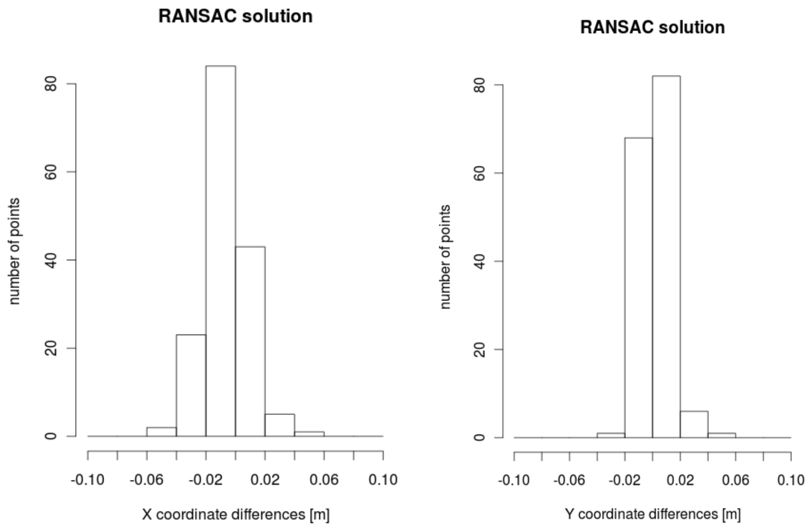
The second RANSAC transformation was performed assuming that 83% (45 out of 54) of the reference point coordinates include different values of gross errors. The residuals, calculated to the nine correct reference point coordinates, are presented in the Figure 4.



(a) (b)

Fig. 4. (a) The histogram of X residuals (b) The histogram of Y residuals

The Figure 4 presents the histogram of X and Y residuals obtained in the second scenario. In this scenario, the values of the residuals are between –0.01 and 0.02 m. The resulting coordinate differences are presented as histograms in the Figure 5.



(a) (b)

Fig. 5. (a) The histogram of the X coordinate differences (b) The histogram of the Y coordinate differences

In the Figure 5, the histogram of the X and Y coordinate differences obtained in the second scenario is presented. The values of these differences are between –0.06 and 0.06 m as in the second scenario. The results of the transformation are almost the same in those two scenarios. In this scenario, the coordinates of 45 points are contaminated by outliers. Thus, only 9 out of 54 points are not burdened with gross errors. In this case, traditionally used robust estimation methods would not give the correct result. These methods would recognize those nine correct points as outliers and 45 bad points as correct points. RANSAC transformation correctly identified 9 points (inliers) and the transformation parameters were estimated based on them.

4.2. Modified RANSAC transformation

To confirm the necessity for modification of RANSAC transformation performed additional calculations. In this case the transformation performed also in two scenarios. In the first scenario to estimate the transformation parameters two points were selected randomly. In the second scenario selected two points furthest from each other. Then performed the transformation of coordinates. The results are presented in the Figure 6 and Figure 7 as a differences between coordinates after transformation and the catalogue values.

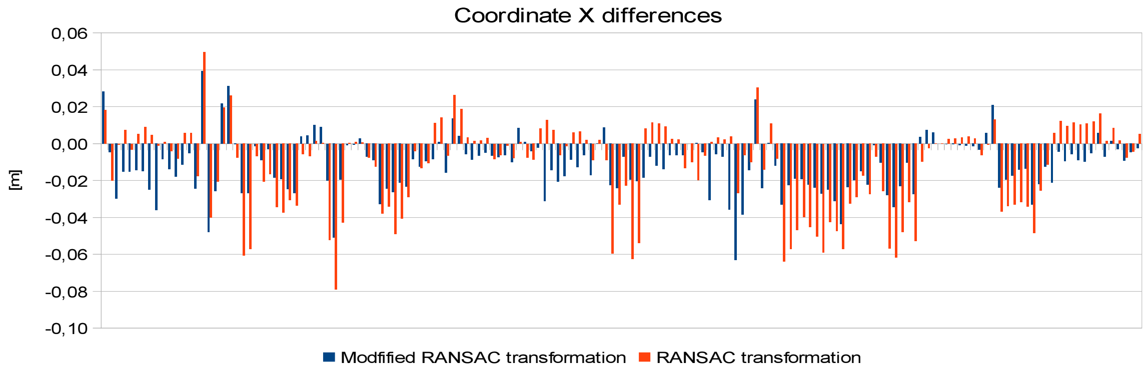


Fig. 6. Differences between coordinates

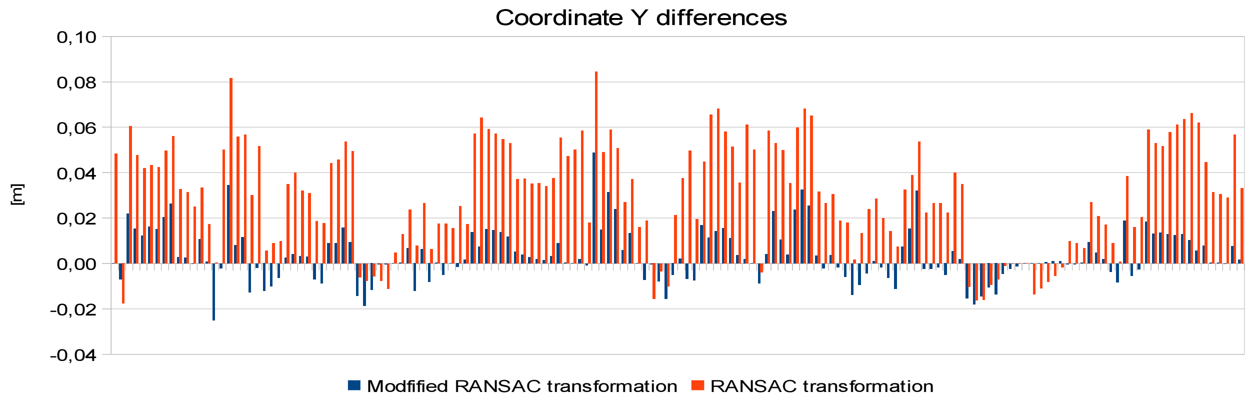


Fig. 7. Coordinate Y differences

Figures 6 and 7 show that the coordinate differences obtained after modified RANSAC transformation with two optimally selected points are less then the differences after RANSAC transformation where two points were chosen randomly. Thus the modification of the algorithm provides better results than the standard RANSAC transformation.

5. Conclusions

In order to show the robustness of this approach, the method was applied at different scenarios, considering different numbers of outliers. It started with a case with 50% of points contaminated with outliers and continued with a very high number of 83% of points contaminated with outliers.

In each variant of calculation, the desired results was achieved, therefore, the effectiveness of RANSAC algorithm in application to coordinate transformation is confirmed. The main goal of the study was to confirm the possibility to properly estimate coordinate transformation parameters, when the total number of points burdened with errors is greater than 50%. Calculations performed in the second and third variant prove the effectiveness of the proposed algorithm, with the number of observations burdened with errors at the level of 83%.

Despite the advantages of this method, it has some flows. It requires many iterations and many operations which sometimes (especially in the case of large sets) take a longer time than the standard procedure. There is also a risk that the algorithm will not select the optimally points for the best solution. Therefore the modification proposed by the authors optimizes the RANSAC transformation. The number of iteration is reduced and the location of the reference points is the most optimally.

References

[1] Duchnowski, R. 2011. Robustness of strategy for testing leveling mark stability based on rank tests, *Survey Review* 43(323): 687–699. http://dx.doi.org/10.1179/003962611X13117748892551
[2] Fischler, M.; Bolles, R. 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM* 6(24): 381–395. http://dx.doi.org/10.1145/358669.358692
[3] Huber, P. J. 1964. Robust estimation of a location parameter, *Annals of Mathematical Statistics* (1): 73–101. http://dx.doi.org/10.1214/aoms/1177703732
[4] Janicka, J.; Rapinski, J. 2013. *M-split transformation of coordinates,* Survey Review, available online on 21 May 2013. http://dx.doi.org/10.1179/1752270612Y.0000000030
[5] Janicka, J. 2011. Transformation of coordinates with robust estimation and modified Hausbrandt correction, *Environmental engineering* (3): 1330–1334.
[6] Wisniewski, Z. 2008. Estimation of parameters in a split functional model of geodetic observations, *Journal of Geodesy* (83): 105–120.