

The circle object detection with the use of Msplit estimation

Artur Janowski^{1,*}

¹ University of Warmia and Mazury, Faculty of Geodesy, Geospatial and Civil Engineering, Institute of Geodesy, Prawochenskiego 15, 10-720 Olsztyn, Poland

Abstract. The paper presents the use of $M_{split(q)}$ - estimation in the filtration and aggregation of point clouds containing a known number of elliptical shapes with preliminary unknown - locations and dimensions. These theoretical solutions may have practical relevance especially in the modelling of terrestrial laser scanning data of objects that have similar shape to circles. Mentioned shapes can be scanned of tree trunks, columns, gutters, etc., that are elliptical in the horizontal plane. The results are satisfied and encourage furthermore detailed research, particularly with the extension of 3D applications.

1 Introduction

Laser scanning measurement technology is popular and well-known technology. This technology is improving and has new applications. It can be specified as:

- SLS (Satellite Laser Scanning) [1],
- TLS (Terrestrial Laser Scanning) [2-4],
- MLS Mobile Laser Scanning [5],
- ALS (Airborne Laser Scanning) [6].

In the case of TLS, the measurement of spatial point locations must be performed from several measurement stations to get complete information, such as the envelope of spatial objects. Despite the efforts, the obtained data set will not fully represent the observed space. It will also contains a lot of surplus data (redundancy) that hinders the assessment of the measured space as well as its numerical analysis.

Therefore, the raw measurement data is usually taken to modelling, which aims to present space as homogeneous sets of qualitatively [7], semantically and functionally objects described in the simplest way.

This is usually done by specification the location of the object and its dimensions and orientation in space.

The spatial and topological relations obtained in this way can be used to gain knowledge about the study space as well as to perform analyses based on them such as [8-11].

The modelling process compared to the measurement process is more time consuming and demanding from operator's point of view.

It seems that it can never be fully automated, but many of the tasks can be accelerated enough and performed by the application of a supervised mode.

Nowadays spatial data modelling applications based on TLS measurements include many efficient tools for automatic object recognition. They are based on the nature of the mutual distribution of measurement points, their density and reflectance values [12-16].

However, it is always required to indicate the analysis (modelling) area and the initial parameters determining the operation of the selection, filtering or aggregation tools and the selection of geometric primitives best matched to the modelled element.

The main goal of this article is to present the algorithm for automatic detection, on a given points cloud, assumed number of elliptical shapes (horizontal cross-section).

2 M-estimation

$M_{split(q)}$ [17,18] assumes existence set of n observations l (l_1, l_2, \dots, l_n) that contain q mixed random variables with expected values (1):

$$\begin{aligned} E\{l_i\}_1 &= \mathbf{a}_{(1)}\mathbf{X}_{(1)}, \\ E\{l_i\}_2 &= \mathbf{a}_{(2)}\mathbf{X}_{(2)}, \\ &\dots \\ E\{l_i\}_q &= \mathbf{a}_{(q)}\mathbf{X}_{(q)} \end{aligned} \quad (1)$$

where:

\mathbf{a} – vectors of known coefficient values,

\mathbf{X} – vectors of unknown parameters q random variables.

According to the assumptions of the $M_{split(q)}$, for each observation, the division of potential is defined as the possibility of its belonging to the one of defined functional models. Therefore, each observation l_i are described by the functional model (2):

$$l_i = E\{l_i\} + v_i = \mathbf{a}\mathbf{X} + v_i \quad (2)$$

corresponds to q competing functional models (3):

$$l_i = E\{l_i\}_{(q)} + v_{i(q)} = \mathbf{a}_{(q)}\mathbf{X}_{(q)} + v_{i(q)} \quad (3)$$

* Corresponding author: artur.janowski@geodezja.pl

This assumption allows presents the traditional functional model (2) as a conglomerate of functional models as follow (4):

$$l_i = \mathbf{a}_i \mathbf{X} + v_i \rightarrow \text{split} \rightarrow \begin{cases} l_i = \mathbf{a}_{(1)} \mathbf{X}_{(1)} + v_{i(1)} \\ l_i = \mathbf{a}_{(2)} \mathbf{X}_{(2)} + v_{i(2)} \\ \vdots \\ l_i = \mathbf{a}_{(q)} \mathbf{X}_{(q)} + v_{i(q)} \end{cases} \quad (4)$$

The main assumption of $M_{\text{split}(q)}$ is to assign the greatest influence (during estimation process of the particular parameters (5)) those observations that best fit to the particular functional models:

$$\mathbf{X}_{(1)}, \mathbf{X}_{(2)}, \dots, \mathbf{X}_{(q)} \quad (5)$$

and can be used to determine estimators (6):

$$\hat{\mathbf{X}}_{(1)}, \hat{\mathbf{X}}_{(2)}, \dots, \hat{\mathbf{X}}_{(q)} \quad (6)$$

The competition of functional models for every observation adjust them into the whole set of observations according to the conglomerate's assumptions.

Graphically, the idea of $M_{\text{split}(q)}$ estimation of the set points, that lays on the plane, of three circles - three functional models ($q = 3$) is shown in Fig. 1.

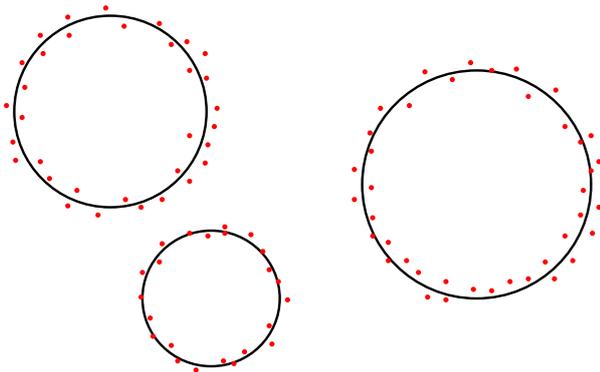


Fig. 1. $M_{\text{split}(q)}$ for sets of points belonging to three circles.

The obvious problem in the iterative estimation of the $M_{\text{split}(q)}$ is the proper choice of the number of q models in (4). Incorrect assumption of this magnitude will result in the uncertainty of the observation set modelling (the residuals will be too big) and the result will be far from expectation. The number of approximating models can be from 1 to n where n is the number of all observations in the measurement set.

In the presented paper, the choice of the value of the number of models was assumed as a additional task for further considerations. Therefore, in this analysis value of q is the result of the application the statistical method of the cluster analysis with the use of single-neighborhood agglomeration (nearest neighbourhood) [19,20]. This method consists in determining the distance between any clusters as the distance between the two nearest objects (closest neighbours) belonging to distinct clusters. Elements of one cluster form one chain. This method fails with overlapping clusters belonging to

different models, then in practical implementations, it may require operator supervision to correct (increase) number of clusters .

After determining the number of clusters and the types of functional models (4) for the all of the $M_{\text{split}(q)}$ observations, it can be performed. The result will depend on the correctness of the selected models, their number q , and the relation of the number of observations belonging to each model [21], as well as the way of weighting the observations in the estimation process.

3 Functional model

The use of $M_{\text{split}(q)}$ estimation in the elaboration of the 2D or 3D set of points projected to the plane, where they can be treated as q sets of circles requires the use of a suitable functional model. The alignment of a set of 2D points is reduced to determination of the circle centre coordinates and its radius and thus the parameters fulfilling the classical equation:

$$(x_i - x_c)^2 + (y_i - y_c)^2 = R^2 \quad (7)$$

where:

x_c, y_c – centre coordinates of the circle,

x_i, y_i – coordinates of circle points,

R – radius of the circle.

Using the above equation (7) in the LSM (least squares method) estimation requires transformation into a form (8)[22]:

$$P_1 x_i + P_2 y_i + P_3 = -(x_i^2 + y_i^2) \quad (8)$$

where:

$$P_1 = 2x_c \quad (9)$$

$$P_2 = 2y_c \quad (10)$$

$$P_3 = x_c^2 + y_c^2 - R^2 \quad (11)$$

Such a functional model (8) was chosen for the estimation of q circles in the $M_{\text{split}(q)}$ process. Hence:

$$\mathbf{L} = \mathbf{A}\mathbf{X}_{(q)} + \mathbf{v}_{(q)} \quad (12)$$

where:

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & 1 \\ x_n & y_n & 1 \end{bmatrix} \quad (13)$$

$$\mathbf{X} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & 1 \\ x_n & y_n & 1 \end{bmatrix} \quad (14)$$

$$\mathbf{L} = \begin{bmatrix} -(x_1^2 + y_1^2) \\ -(x_2^2 + y_2^2) \\ \vdots \\ -(x_n^2 + y_n^2) \end{bmatrix} \quad (15)$$

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad (16)$$

4 Iterative block of algorithm

An integral part of the algorithm using the $M_{split(q)}$ estimation for segregation of measurement data that represents sets q circles is the "iterative block of algorithm" shown in Fig. 2. The input to this block is the data set as the matrices A , L , X (13-15) and parameter q . On such set of data, the LSM estimation is iterated q times. Weights for each estimation are calculated based on the residual values for observations from the previous estimate. The rule here is to assign higher weight values for observations that have higher residual values in the preceding estimation. This rule is implemented by the weight function module (Fig.2).

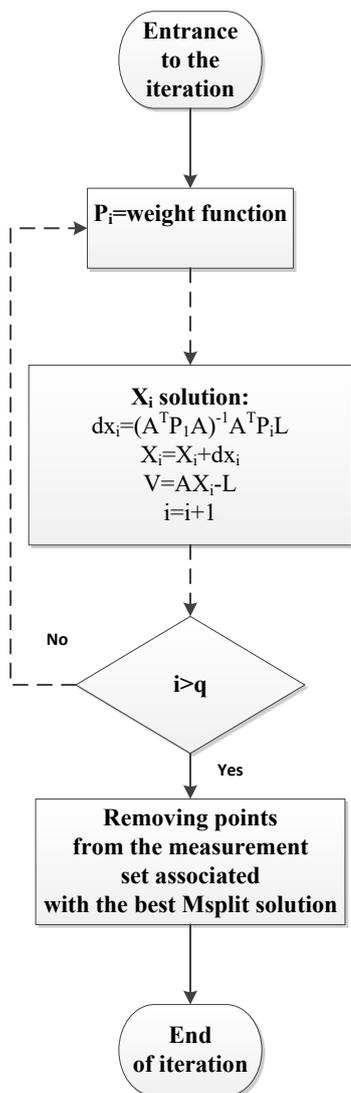


Fig. 2. Iterative block of algorithm.

The method of weighing should be chosen in relation to the optimality of the given task, theoretically and empirically tested. The activity sub-block marked with a dashed line can also be executed iteratively (more than once). Then the weights of the first estimation are obtained by analyzing the residuals of the q th estimate of the previous iteration. Such an approach may optimize the execution time of task (based on presented algorithm) for some systems of functional models.

After completing the iterative estimation and obtaining the parameters for the assumed q functional models, it must be chosen the least defective, it means the lowest error fitting into a hypothetical subset of data. A hypothetical subset of data is understood here as a set of observations that Euclidean distance from the found solution of the functional model is less than the assumed value of the parameter τ . The final step of this iterative block is removing, from the measured dataset, observations assigned to the optimal model just found (based on the assumed value of the parameter τ). It means that the remaining observations have a connection with others $q-1$ solutions.

The iterative block usage process is repeated q times, and each time one optimal solution is selecting and its observations are reducing from all data set. The obtained results of the parameters of all q functional models constitute the approximate parameters of the next iteration that are presented as a dotted line (Fig. 2). The gradient values of these parameters is the basis for the decision to end the algorithm. The schema of main algorithm was depicted on Fig. 3.

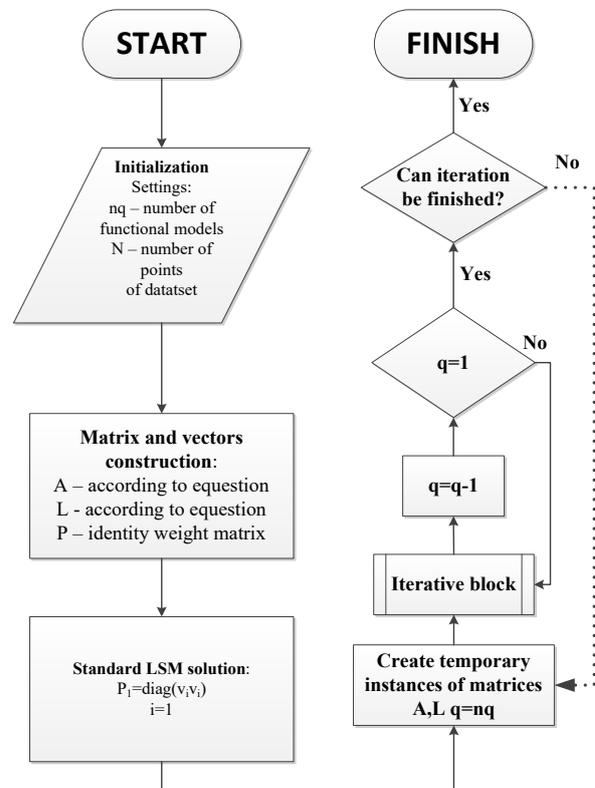


Fig. 3. The schema of main algorithm.

5 Practical implementation

To test the algorithm (Fig. 3), $i = 1050$ points data was assumed. Points (x_i, y_i) can be interpreted as the creation of 7 functional models in the form of circle equations (8). Three of these circles were overlaid on each, in order to test the reliability of the algorithm to the penetrate results of functional models (Fig. 4).

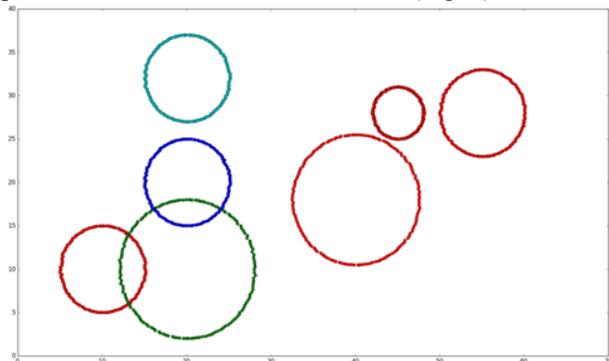


Fig. 4. Visualization of dataset created for validation algorithm purposes.

Parameters of the theoretically defined circles (the theoretical coordinates of the centres of circles x_c, y_c and theoretical radiuses R) with coordinates of points contaminated with Gaussian noise with the maximum value 0.20 m, are presented in Table 1. Coordinates of circle centres and their radiuses, are estimated from data with Gaussian noise, are given respectively as $x_{cest}, y_{cest}, R_{est}$. Assumed value $\tau = 1.50$ m.

The weighing function was the fiftieth power of residual for observations from the previous estimation.

Table 1. Theoretical and estimated parameters of test circles.

Circle no.	x_c	y_c	R	x_{cest}	y_{cest}	R_{est}
1	10.0	10.0	5.0	10.09	10.00	5.00
2	20.0	10.0	8.0	20.10	9.99	8.00
3	20.0	20.0	5.0	20.11	20.00	5.01
4	20.0	32.0	5.0	20.10	32.00	5.00
5	55.0	28.0	5.0	55.10	28.00	5.00
6	45.0	28.0	3.0	45.10	28.00	3.00
7	40.0	18.0	7.5	40.10	18.0	7.50

As a result of the algorithm use in its 10 main iterations (dotted line in Fig. 3) and increased to 2 iterations in the "iterative block" (dashed line in Fig. 2), one obtains the solution of $q = 7$ functional models - circles depicted as continuous line in Fig. 5.

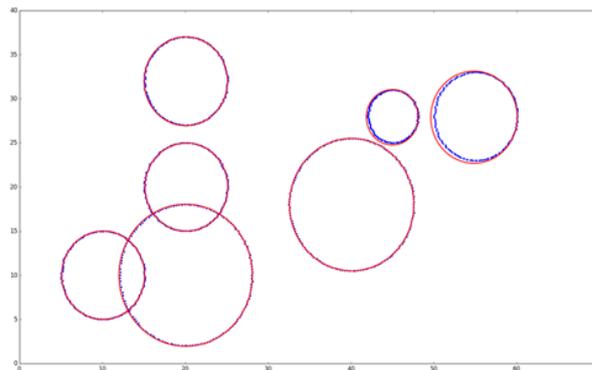


Fig. 5 The solution $M_{split(q)}$ for 7 circles (continuous line).

The obtained values of functional models parameters (x_{mc}, y_{mc}, R_m) and their absolute differences, with the values estimated from the Table 1, are presented in Table 2.

Table 2. Result values for $M_{split(q)}$ with 7 circles.

Circle no.	x_{mc}	y_{mc}	R_m	$ x_{mc} - x_{cest} $	$ y_{mc} - y_{cest} $	$ R - R_{est} $
1	10.10	10.00	5.00	0.01	0.00	0.00
2	20.06	9.99	8.05	0.04	0.00	0.05
3	20.14	19.99	4.99	0.03	0.01	0.02
4	20.07	31.97	5.03	0.03	0.03	0.03
5	54.85	27.90	5.22	0.25	0.10	0.22
6	44.99	28.90	3.13	0.11	0.90	0.13
7	40.09	18.00	7.51	0.01	0.00	0.01

The largest differences from the assumed theoretical values can be seen in solutions 5th and 6th of circle. The numerical justification for such a result is noticed in the order of finding solutions of functional models during using the algorithm. 5th and 6th circles in the iteration process were found earliest then others (there was the greatest negative impact of observation recognized in the further process that belongs to the other solutions).

6 Conclusion

It is worth to notice that among the existing applications of M_{split} [23] there is lack of practical implementation of using this application for more than 2 spited sets ($q > 2$). As the industrial application of the presented algorithm, the measurements of space with a highly structured form, but unknown objects parameters with circle shapes can be indicated. Developing an algorithm for 3D applications would allow the detection of linear infrastructures with circle cross-sections.

Mass acquisition of observation data, which is the domain of most geodetic survey nowadays, forces the search for numerical solutions capable of meeting the expected data processing time and the accuracy of the analysis results.

The proposed modification of algorithm, the extension of the standardized approach of $M_{split(q)}$, works in theoretical considerations. Its practical utility depends on

the quality of the data, the number of outliers, the number of observations associated with each of the functional models.

The evident problem of the algorithm needed to solved in the future is the automation of the definition of the value q - the number of functional models. This article assumes its knowing, which greatly influences the versatility of applications. One of potential solution is to place the entire algorithm into another external iteration responsible for the iterative changing value of parameter q . Store of solution results for the accepted value range $q = 1 \dots \max q$ and choosing this solution where the residual errors will be the smallest. The work involved in the optimal selection of parameter q is the basis for the next study and is closely related to its implementation and commercial applications.

References

1. J. A. B. Rosette, P. R. J. North, J. C. Suárez. Vegetation height estimates for a mixed temperate forest using satellite laser altimetry *International Journal of Remote Sensing*, 29 (5), 1475 (2008)
2. T. Cao, A. Xiao, L. Wu, L. Mao. Automatic fracture detection based on Terrestrial Laser Scanning data: A new method and case study *Computers & Geosciences* 106, 209 (2017)
3. P. Tysi c, A. Wojtowicz, J. Szulwic. Coastal Cliffs Monitoring and Prediction of Displacements Using Terrestrial Laser Scanning 2016 Baltic Geodetic Congress (Geomatics), 61, DOI: 10.1109/BGC.Geomatics.2016.20 (2016)
4. K. Bobkowska, A. Janowski, J. Szulwic. 3D modelling of cylindrical shaped objects from lidar data – an assessment based on theoretical modelling and experimental data. *Metrology and Measurement Systems*, 25(1) (2018)
5. D. Barber, J. Mills, S. Smith-Voysey. Geometric validation of a ground-based mobile laser scanning system *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1), 128 (2008)
6. A. Janowski, J. Szulwic, P. Tysi c, A. Wojtowicz. Airborne and mobile laser scanning in measurements of sea cliffs on the southern Baltic. *SGEM2015 Conference Proceedings* 2, 17, ISBN 978-619-7105-32-2, ISSN 1314-2704 (2015)
7. M. Walacik. Orthodox valuation method application on the example of coworking *SGEM2017 Conference Proceedings* 17, 421 (2017)
8. K. Bobkowska, A. Ingot, M. Mikusova, P. Tysi c. Implementation of spatial information for monitoring and analysis of the area around the port using laser scanning techniques. *Polish Maritime Research* 24 (s1), 10-15, DOI: 10.1515/pomr-2017-0015 (2017)
9. A. Janowski, J. Rapinski. The Analyzes of PDOP factors for a Zigbee ground – based augmentation systems. *Polish Maritime Research*, 24(s1), pp. 108-114, DOI: 10.1515/pomr-2017-0028 (2017)
10. M. Renigier-Bi zor. Modern classification system of real estate markets. *Geodetski Vestnik*. 61(3). DOI: 10.15292/geodetski-vestnik.2017.03.441-460 (2017)
11. M. Bednarczyk. Application possibilities of augmented reality in analog maps *Geographic Information Systems Conference and Exhibition “GIS ODYSSEY 2017”*, 22-27 (2017)
12. C. Wang, P. Hsu. Adaptive building edge detection by combining LiDAR data and aerial images . *Journal of Photogrammetry and Remote Sensing* 12(4), 365 (2007)
13. A. Janowski, J. Rapi ski. M-Split Estimation in Laser Scanning Data Modeling. *Journal of the Indian Society of Remote Sensing* 42, 15 (2013) DOI: 10.1007/s12524-012-0213-8
14. J. Szulwic, P. Tysi c. Searching for road deformations using mobile laser scanning. *MATEC Web of Conferences* 122 (2017) DOI: 10.1051/matecconf/201712204004
15. J. Zheng, J., T. Mccarthy, A. S. Fotheringham, L. Yan. Linear feature extraction of buildings from terrestrial lidar data with morphological techniques. *International Archives of the Photogrammetry Remote* 37(B), 241 (2008)
16. P. Axelsson. DEM generation from Laser Scanner Data Using Adaptive TIN-Models. *The International Archives of Photogrammetry and Remote Sensing* 33/B4/1, 110 (2000)
17. Z. Wi niewski. Estimation of Parameters in a Split Functional Model of Geodetic Observations (Msplit Estimation). *Journal of Geodesy* 83, 105 (2009)
18. Z. Wi niewski. Msplit(Q) Estimation: Estimation of Parameters in a Multisplit Functional Model of Geodetic Observations. *Journal of Geodesy* 84, 355 (2010)
19. M. Renigier-Bi zor, A. Bi zor, R. Wi niewski. Rating engineering of real estate markets as the condition of urban areas assessment. *Land Use Policy* 61, 511 (2017)
20. M. Renigier-Bi zor, R. Wi niewski, A. Bi zor, A. Kaklauskas. Rating methodology for real estate markets – Poland case study. *International Journal of Strategic Property Management* 18, 198 (2014)
21. A. Bi zor, M. Renigier-Bi zor. Optimization and polyoptimization in the management of land. *SGEM2015 Conference Proceedings* 2, 1011 (2015)
22. A. Janowski, J. Szulwic, M.  uk. 3D modelling of liquid fuels base infrastructure for the purpose of visualization and geometrical analysis. *SGEM2015 Conference Proceedings* 1, 753 (2015)
23. J. Janicka, J. Rapinski. Msplit transformation of coordinates. *Survey Review* 45(331), 269 (2013)