International Cartographic Association
Association Cartographique Internationale

Automatic cleaning of outliers in multibeam bathymetric data with clustering algorithm

Yuan Wei, Shujun Li, Shaohua Jin, Hongchao Ji

Department of Military Oceanography and Hydrography & Cartography, Dalian Naval Academy, Yuan Wei, 383073307@qq.com, Shujun Li, lishujunhao@163.com, Shaohua Jin, jsh_1978@163.com, Hongchao Ji, Jihongchao@163.com

Keywords: multi-beam bathymetry, clustering algorithm, outliers, automatic cleaning

Abstract:

Eliminating the invalid data and locating the doubtful data are realized by using the algorithm of density clustering by referring to the image features of outliers in the rearview from the perspective of manual interactive processing. The MultiBeam EchoSounder (MBES) bathymetric datasets are automatically divided into three types: credible data, doubtful data, and invalid data by the clustering algorithm. The reliable data are retained, invalid data are automatically eliminated, and the doubtful data are judged manually. This kind of automatic outlier cleaning algorithm with partial manual intervention can solve the contradiction between the low reliability of automatic processing algorithms and the low efficiency of manual interactive processing. An example shows that the algorithm improves the reliability of the results obtained from the automatic processing algorithm to a certain extent, and also, the algorithm is of great significance for the realization of high reliability and high-efficiency cleaning of outliers in MBES bathymetric data.

As the world continues to explore the ocean, the demand for efficient access to high precision underwater topography is increasing, the accuracy and efficiency of underwater terrain acquisition nowadays will not be able to meet the needs of technology development. As one of the most efficient and precise equipment, the MBES system is widely used worldwide in underwater topography surveys. In recent years, the original MBES data proliferating. However, for a long time, the cleaning of outliers in MBES data with high reliability mainly relies on manual operation, which has an extended processing cycle and dramatically limits data processing efficiency. The automatic processing algorithms that have been proposed improve processing efficiency, but the credibility can not reach that of human processing; it is challenging to apply it in multibeam surveying, especially with nautical charting. Therefore, there is a crucial problem in MBES data processing:how to improve the credibility of the results processed with the automatic algorithm.After years of research, many automatic data cleaning algorithms have been proposed both at home and abroad. The cleaning mechanism can be summarized as follows:Setting up several simple thresholds; Filtering by conditions based on the statistical results;establishing local topography model.

The boundary between the recognition and validation of outliers is blurred by most of the existing automatic algorithms, and the MBES soundings are directly divided into two categories according to the specific conditions: credible and ineffective. But in surveying practice, there are usually a few data belonging to the third category, which we call doubtful data. The third category of data has to be verified by professionals, who need to determine with the experience combined with the investigation data of the survey area and instrument characteristics. For example, the causes of a small cluster of points isolated from the central cluster may be suspended solids in water or angular objects, such as stones; the former should be marked as invalid data and eliminated, while the latter should be marked as trusted data to be retained. Most of the existing automatic cleaning algorithms mark the doubtful data as credible data or invalid data across the board, which is the key to the low reliability of the current automatically processing algorithms. Although dubious data accounted for a small percentage, it is crucial because it often represents the characteristics area where the underwater topography changes dramatically or targets unnatural topographic relief. Once been misjudged, it would have a significant influence on the measuring result. Currently, there is no appropriate unartificial solution to validate the doubtful data. Although the doubtful data that need manual validation is rare, manually Screening for doubtful data ping by ping would also take much time. Therefore, it is a feasible program that could improve the credibility and efficiency of data processing, that the doubtful data are located automatically and validated manually.

With the matures of the machine learning algorithm, more and more scholars apply it to the cleaning of multibeam outliers， due to its unique advantages in target recognition. Yang Fanlin et al. introduced an erosion and expansion clustering algorithm to realize the automatic cleaning of MBES outliers in the deepwater area(Yang et al., 2007). Xiaolong Chen et al. used the K-Means clustering algorithm to detect outliers based on single pulse-sounding data. In this algorithm, a ping was divided into multiple clustering windows according to the topographic change. With beams in

Abstracts of the International Cartographic Association, 3, 2021.
30th International Cartographic Conference (ICC 2021), 14–18 December 2021, Florence, Italy.
https://doi.org/10.5194/ica-abs-3-314-2021

the center of a ping as initial clustering samples, a cluster on behalf of the actual underwater topography was clustered from the central beam to the edge; and the outliers are weed out automatically. This method solves the problem of recognition accuracy, but the K-Means algorithm is sensitive to the noise. As a result, the algorithm may fail when there are large areas of continuous abnormal or missing data. Although the methods above can classify MBES soundings appropriately, they still directly divide the data into two categories: credible and invalid, and the validity of the doubtful data is not isolated; the processing results have the problem of credibility.

For this reason, an automatic data cleaning algorithm with partially manual intervention was proposed, Based on the Density Clustering algorithm DBSCAN (Density-based Spatial Clustering of Applications with Noise).DBCSAN algorithm is used to divide the bathymetric data into three categories: credible, doubtful, and invalid, retains the credible data and eliminates the invalid data, and the dubious data submitted to manual validation. The method could effectively improve the credibility of the automatic data cleaning algorithm. Simultaneously, compared with manual processing, the method in this paper still dramatically enhances efficiency. A contradictory problem has been solved, that the automatic processing method has high efficiency but low credibility, while manual processing method has high reliability but low efficiency of this contradiction

The main steps of the algorithm include two parts: identification of outliers based on DBSCAN and validation with manual intervention. The first step is to divide the original data into three categories: credible data, invalid data, and doubtful data through the clustering algorithm. The trusted data is retained, the invalid data is eliminated, and the doubtful data is submitted to the manual validation. Finally, the doubtful data will be classified as credible data or invalid data.

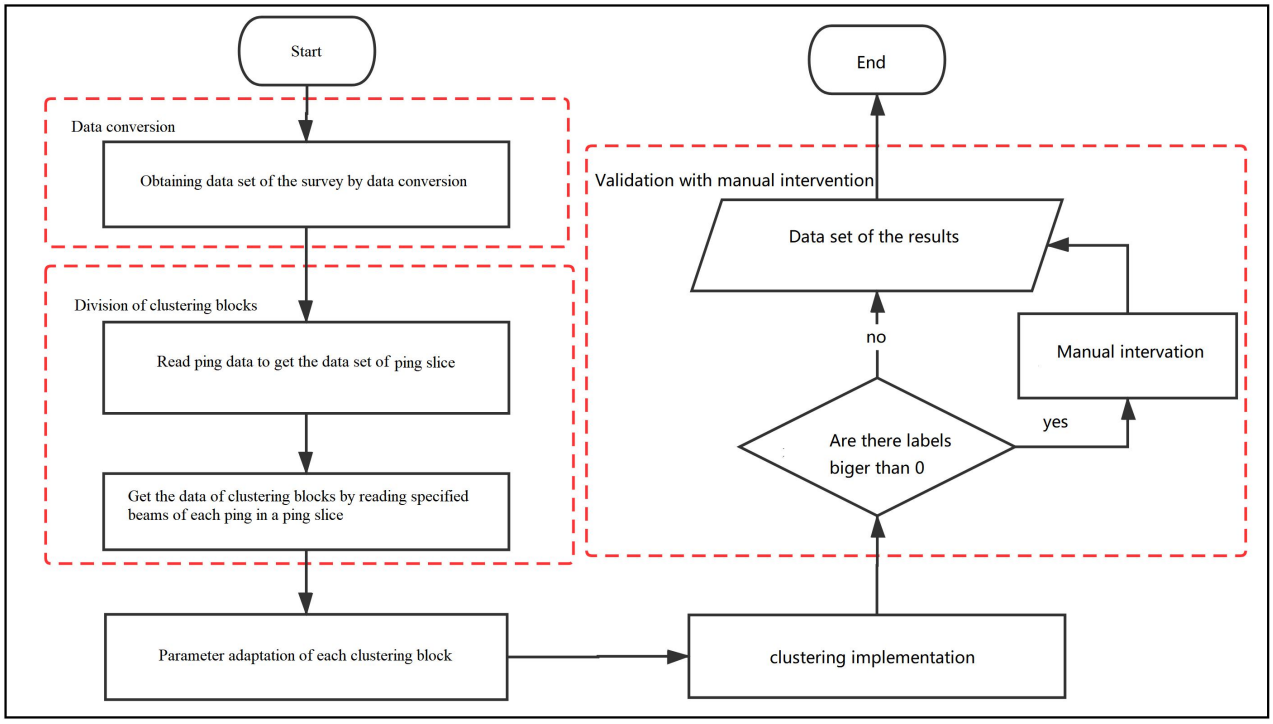

Figure 1. Flow chart of the algorithm

The concept of ‘doubtful data’ is introduced into the automatic cleaning of outliers in multibeam bathymetric data, which provides a new idea for the automatic cleaning of outliers. A new automatic cleaning method in multibeam bathymetric data is established, which uses the DBSCAN algorithm to classify bathymetric data and manually validate the doubtful data. This automatic processing algorithm effectively divides raw data into three types: credible data, doubtful data, and invalid data. Retain the credible data and eliminate invalid data. The validation of the dubious data is carried with manual intervention. To a certain extent, this method improves the credibility of the automatic outlier cleaning algorithm and has practical application value.

Abstracts of the International Cartographic Association, 3, 2021.
30th International Cartographic Conference (ICC 2021), 14–18 December 2021, Florence, Italy.
https://doi.org/10.5194/ica-abs-3-314-2021 | © Author(s) 2021. CC BY 4.0 License.