Filtering Methods of Partially Incorrect Geodata

A. A. Golovkov[1], G. S. Ivanova[2]
Bauman Moscow State Technical University
Moscow, Russia
[1]alexander.golovkov1@gmail.com, [2]gsivanova@gmail.com

***Abstract*— At the present time Geoinformation Systems (GIS) are actively using mobile applications to collect and process geodata of field staff in the areas of maintenance and repair (MRO). Such opportunity allows to monitor staff movements and coordinate their actions in the process of work. Processing and filtering geodata on mobile devices is a big challenge especially in in low precision geopositioning, for example, when navigating inside buildings. In the paper we propose the methods of geodata processing oriented to reduce the density of track points, identify and exclude the coordinates with a high error while standing indoors or moving near high-rise buildings, where the accuracy of satellite geopositioning decreases significantly due to reflection, refraction and attenuation of radio signals. The advantages and disadvantages of methods are revealed, their efficiency is estimated.**

Keywords*— *geodata; accuracy; filtering; mobile device; track; GIS

I. Introduction

One of the promising areas in the field of automation of field staff activities is the development of Geoinformation Systems (GIS), which allow monitoring the movement of mobile employees and operatively coordinate their actions in the process of work. Typically, GIS includes mobile applications that provide the collection and preprocessing of geoinformation data with subsequent transfer to the server part. Due to heterogeneity of geodata sources, hardware limitations of mobile devices and high requirements for operation speed, processing of geoinformation is a challenging task. Filtering methods should work efficiently, particularly when navigating indoors and near high-rise buildings, where due to the reflection, refraction and attenuation of radio signals, the error of determination of the coordinates becomes unreliable.

II. Existing Methods for Processing Geodata

The approaches and methods of filtering geocoordinates in mobile devices have been studied in detail in [1–4, 6–14]. Let's consider some of them.

Khrul et al. [3] propose a flow-filtering algorithm based on the analysis of the accuracy of neighboring coordinates. As a filter criterion, the distance between two points is used, which should be not less than the sum of the threshold values that are depend on the type of vehicles being examined. The approach is aimed at excluding coordinate deviations while standing, increasing reliability and reducing the amount of input data. One of the main drawbacks of the method is the lack of dependence of the filter behavior on velocities in neighboring coordinates, which in some cases gives visually poor-quality tracks, especially in cornering movements [2].

The application of the Kalman filter [4, 5, 12, 21] for smoothing geocoordinates is well known. Considering limitations and assumptions [4], the filter effectively smoothes out single deviations of coordinates, but it is not able to deal with long deviations of the trajectory of motion in conditions of unreliable accuracy and/or speed.

Considering high complexity of the task and the large amount of heterogeneous data, many researchers began to apply machine learning models and methods [15] to solve navigation problems in different environments [16–20]. A key feature of these approaches and their significant advantage over other methods is that for their use it is not necessary to develop an accurate mathematical model of the process of user movement or the scatter of coordinates while standing. In the process of learning, the model is adjusted to the input data, modeling the navigation process [22–26]. An essential disadvantage of using such models is that in the inference stage it is impossible to determine the exact criteria on the basis of which the trained model has made a decision regarding the validity of the coordinate.

The Fused Location Provider should be noted – the technology of Android geolocation, designed to generalize all sources of geodata and solve the problem of geolocation indoors [27], but the quality of filtering by this method leaves much to be desired: the deviation of the traveled distance from the real can reach tens of kilometers.

III. Filtering Methods of Partially Incorrect Geodata

We propose filtering methods based on [2, 3], which allow to reduce the density of the points of the resulting track, identify and exclude coordinates with a large error of its determination in the case of partially unreliable geodata parameters.

A. *Filtering by the Direction of Moving*

The purpose of the filter is to reduce the density of track points, based on the analysis of the angle between the directions of motion. First of all, it is necessary to compare the time interval Δt between the previous and the current coordinate with the interval $\Delta T_{\max}$:

$$\Delta t < \Delta T_{\max}(\alpha) = \begin{cases} 1, if \ \ \alpha > P^1_{\max A} \\ \dfrac{1 - P^1_{\max T}}{P^1_{\max A}}\alpha + P^1_{\max T}, otherwise \end{cases},$$

where α is the difference in degrees between the directions of motion along three coordinates in the sequence, $P_{\max A}$ is the maximum angle in degrees at which the coordinate will be filtered if the time interval is less than 1 second.

If $\Delta t < \Delta T_{\max}(\alpha)$, the coordinate is assumed to have a large error, and it should be rejected. Parameters $P_{\max T}$ and $P_{\max A}$ specify the maximum time interval in seconds for which the coordinates will be filtered at the angle $\alpha = 0$ and the maximum angle within which the coordinates will be excluded if the time interval corresponds to the calculated range. In other words, the threshold of the time interval dynamically depends on the angle between the directions of motion. If $\alpha > P^1_{\max A}$ and the time interval is less than 1 s, the coordinates are excluded, assuming that a turn is made, and the maximum number of coordinates should be passed to form a smooth track. If $\alpha \le P^1_{\max A}$, then the turn is either not significant, or the user moves along a straight line, and in this case, it is preferable to filter out a larger number of coordinates.

In the second stage of processing, it is necessary to analyze the angle α together with the average user speed $\bar{\vartheta}$ at several points in the sequence. If the average speed is greater than a certain value, the threshold angle should be increased, because the maximum possible steering angle decreases with increasing speed. The converse case is also true: the lower the speed, the greater the angle you can make a turn at in a certain time. In accordance with this assumption, the angle α has to be compared with the threshold value, which depends on the average user speed with considering the maximum possible time interval between the coordinates:

$$\Delta t < 3P^1_{\max T} \ and \ \alpha > \Delta A_{\min}(\bar{\vartheta}) = \begin{cases} P^1_{\max A}, if \ \ \bar{\vartheta} > P^1_{\max S} \\ \dfrac{P^1_{\max A} - 180}{P^1_{\max S}}\bar{\vartheta} + 180, otherwise \end{cases},$$

where $P^1_{\max S}$ is the maximum speed in meters per second at which the coordinate will be filtered if $\alpha > P^1_{\max A}$.

Consider the distribution of the angle α before and after filtering on a sample of 368528 coordinates (Fig. 1) with the parameters: $P_{\max T} = 15s$, $P_{\max A} = 10^o$ and $P_{\max S} = 25\dfrac{m}{s}$.

The filter shifted the average value of the angle to the center of the distribution, but the dispersion changed not so great, while the number of coordinates after filtering decreased by 78.976%.

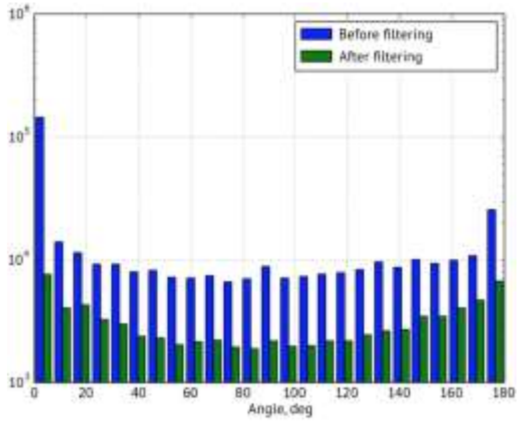

Fig. 1. Distribution of the angle α before and after filtering by the direction of moving

B. Adaptive Diagnostic Filtering

The proposed method of adaptive diagnostic filtering is based on the approach [2, 3]. The purpose of the filtration is to reject incorrect coordinates in the standing areas of the user by analyzing the speed at neighboring points. It will be remembered that, according to [2], the filter criterion is $D_{1-2} > Dl_1 + Dl_2$, where D_{1-2} is the distance in meters between points 1 and 2 in meters; Dl_1 and Dl_2 are the characteristics of the points in meters. If the inequality holds, point 2 is considered valid and passes the filtering, otherwise the point is discarded.

The idea of adaptive filtering consists of two main hypotheses:

1) by decreasing the values Dl_1 and Dl_2 at "entrance" and "exit" from the standing point, it is possible to save the coordinates of the movement near the stands, which will increase the quality of the track;

2) increasing the values Dl_1 and Dl_2 directly proportional to the accuracy and inversely proportional to the speed, it is possible to reduce the amount of deviations while standing, since the speed in this case is significantly lower than in movement.

Based on the described hypotheses, the filter criterion corresponds to:

$$D_{i-1,i} > \begin{cases} f_a(a_{i-1}, p_{i-1}) + f_a(a_i, p_i), \\ \quad if \ \ s_i = 1 \ and \ \vartheta_{i-1} > P^2_S \\ \qquad or \ s_{i-1} = 1 \ and \ \vartheta_i > P^2_S \\ f_a(a_{i-1}, p_{i-1}) f_K(\vartheta_{i-1}) + f_a(a_i, p_i) f_K(\vartheta_i), \\ \quad otherwise \end{cases}$$

where a is the radius of the circle in meters, in which the coordinates are located with a probability of ~68%; p is type of the geodata source: 0 – GPS navigation (NMEA), 1 – mobile geopositioning (API), 2 – Wi-Fi and mobile cell sources [1]; $s_i = \begin{cases} 1, if \ \vartheta_i < P_{maxS}^2 \\ 0, otherwise \end{cases}$ is a flag of standing, P_{maxS}^2 is the minimum user speed in meters per second, at which the location point can be considered as a stand point; $f(a_i, p_i) = \begin{cases} \dfrac{a_i}{P_K^2}, if \ \ p_i = 2 \\ a_i, otherwise \end{cases}$ is a function that considers the low accuracy of coordinates from Wi-Fi and cell sources; P_S^2 is the minimum speed in meters per second, which is assumed to be the user's speed after leaving the stand point; $f_K(\vartheta_i) = \begin{cases} 1, если \ \vartheta_i > P_S^2 \\ \dfrac{1 - P_K^2}{P_S^2} \vartheta_i + P_K^2, иначе \end{cases}$ in accordance with [2].

The first formula in the system corresponds to the first hypothesis above, the threshold value here is directly proportional to the accuracy of two neighboring points without any multiplier. In fact, this formula is the criterion for leaving the stand point. The second formula reflects the second described hypothesis: the accuracy parameters at each point are multiplied by the corresponding coefficients, which depend on the user's velocities at these points.

If the filter criterion is satisfied and the previous coordinate is stand, the method does not reject the coordinate, but modifies its parameters under the assumption that the current coordinate is a continuation of the stand.

Consider the operation of the filter. In the Fig. 2 are presented respectively:

- the track of a four-hour stand indoors with a small moving part;
- the result of filtering by the diagnostic method [3] without considering the user's speed and the entrances/exits from the stand point;
- the result of filtering by the proposed diagnostic method after filtering by the direction of moving.

Red markers are the points with a zero speed, blue ones correspond to the positive speed.

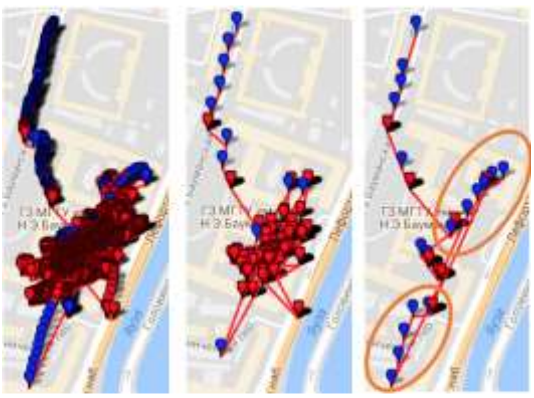

Fig. 2. Comparison of diagnostic filtering methods

The filter dropped out more points directly in the stand point, but the movement part of the track in the top-left corner did not change essentially, as can be seen from the histogram (Fig. 3) too.

It is worth noting that the coordinate deviations, highlighted in orange, are better filtered by the diagnostic method [3].

Quantitative results confirm the efficiency of the method (Table 1): the rejected coordinates percentage in stand points increased, in the movement decreased, while the ratio of distance traveled to the real decreased by approximately 3 times.

TABLE I. COMPARISON OF DIAGNOSTIC FILTERING METHODS

Characteristic		Method [3]	Proposed method
Filtered coordinates, %	all	98.872	**99.559**
	with a zero speed	99.006	**99.863**
	with a positive speed	96.085	**93.238**
Distance traveled (real = 357.260 m), m		3545.094	**1195.716**

Based on the analysis of the speed distribution and comparison results of the methods given in Table 1, it can be argued that the proposed method of adaptive diagnostic filtering is more efficient than the method [3]. Particular attention should be paid to the distance traveled. In many GIS this is one of the most important characteristics, and almost all filtering methods are aimed at increasing its reliability.

IV. CONCLUSION

In the paper we propose two methods of filtering geodata, focused on reducing the density of track points, as well as identifying and excluding coordinates with a large determination error. Both methods use different sources of geolocation data and can be used in mobile devices in real time processing. The experimental results confirm the efficiency of the proposed approaches: the reliability of the ratio of the distance traveled to the real has increased by 19.8% in comparison with the existing methods considered.

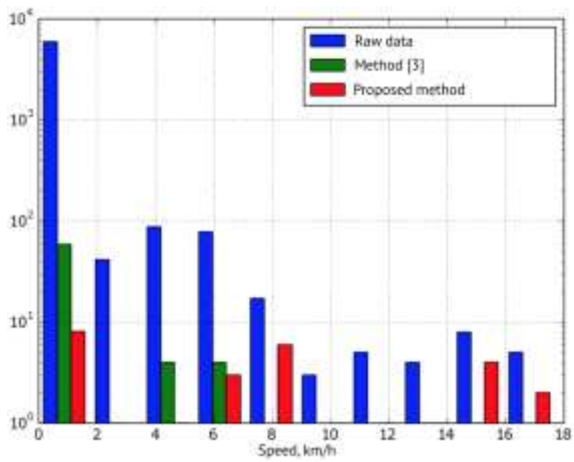

Fig. 3. Distribution of velocity before and after filtering by diagnostic methods

REFERENCES

[1] Golovkov A.A. Geodata sources in mobile devices. Dinamika slozhnykh sistem – XXI vek [Dynamics of Complex Systems – XXI century], 2017, vol. 11, no. 4, pp. 94-101. Available at http://www.radiotec.ru/media/article_files/14_0948-Головков_А.А..pdf (accessed 6 April 2018). (in Russian)

[2] Golovkov A.A., Ivanova G.S. Real time adaptive stream-oriented geo-data filtering. Nauka i obrazovanie. MGTU im. N.E. Baumana [Science & Education of the Bauman MSTU], 2016, no. 4, pp. 156-169. Available at http://technomag.edu.ru/jour/article/download/10/12 (accessed 6 April 2018). (in Russian). DOI: 10.7463/0416.0835995

[3] Khrul' S.A., Son'kin D. M. Adaptive algorithm for processing the flow of navigational data based on the diagnostic filtering method. Izvestiya Tomskogo politekhnicheskogo universiteta [Proceedings of Tomsk Polytechnic University], 2012, vol. 321, no. 5, pp. 217-222. Available at https://cyberleninka.ru/article/v/adaptivnyy-algoritm-obrabotki-potoka-navigatsionnyh-dannyh-na-osnove-metoda-diagnosticheskoy-filtratsii (accessed 6 April 2018). (in Russian)

[4] Listerenko R.R. Applying a Kalman filter to process the GPS coordinate sequence. Molodezhnyy nauchno-tekhnicheskiy vestnik [Youth scientific and technical bulletin], 2015, no. 9. Available at https://readera.ru/primenenie-filtra-kalmana-dlja-obrabotki-posledovatelnosti-gps-koordinat-14413076 (accessed 6 April 2018). (in Russian)

[5] Zabegaev A.N., Pavlovskiy V.E. Adaptation of Kalman filter for use with local and global navigation system. Preprinty IPM im. M.V. Keldysha [Preprints of M.V. Keldysh], 2010, no. 82. Available at http://keldysh.ru/papers/2010/source/prep2010_82.pdf (accessed 6 April 2018). (in Russian)

[6] Lei Gonga, Takayuki Morikawab, Toshiyuki Yamamotoc, Hitomi Satob. Deriving personal trip data from GPS data: A Literature Review on the Existing Methodologies. Procedia-Social and Behavioral Sciences, 2014, no. 138, pp. 557-565. Available at https://www.sciencedirect.com/science/article/pii/S1877042814041597 (accessed 6 April 2018). DOI:10.1016/j.sbspro.2014.07.239

[7] Changqing Zhou, Nupur Bhatnagar, Shashi Shekhar, Loren Terveen. Mining personally important places from GPS tracks. Data Engineering Workshop, IEEE 23rd International Conference on. 17-20 April 2007, pp. 517-526. Available at http://ieeexplore.ieee.org/document/4401037/ (accessed 6 April 2018). DOI: 10.1109/ICDEW.2007.4401037

[8] Velikanova E.P., Voroshilin E.P. Adaptive estimation of maneuvering object position in changeable radio channel transmission gain. Doklady TUSUR [Proceedings of TUSUR], 2012, no. 2-1 (26), pp. 29-35. Available at https://cyberleninka.ru/article/n/adaptivnaya-filtratsiya-koordinat-manevriruyuschego-obekta-pri-izmeneniyah-usloviy-peredachi-v-radiolokatsionnom-kanale.pdf (accessed 6 April 2018). (in Russian)

[9] Makarenko G.K., Aleshechkin A.M. Study on filtering algorithms to determine the coordinates of an object by the signals of satellite navigation systems. Doklady TUSUR [Proceedings of TUSUR], 2012, no. 2-2 (26), pp. 15-18. Available at http://old.tusur.ru/filearchive/reports-magazine/2012-26-2/015.pdf (accessed 6 April 2018). (in Russian)

[10] Prokhortsov A. V., Savel'ev V. V. Ways of determination a vehicles coordinates and velocities by means of a satellite navigation systems. Izvestiya TulGU [Proceedings of the TSU], 2011, no. 2, pp. 264-274. Available at https://cyberleninka.ru/article/v/metody-opredeleniya-koordinat-i-skorosti-podvizhnyh-obektov-s-pomoschyu-sputnikovyh-radionavigatsionnyh-sistem (accessed 6 April 2018). (in Russian)

[11] Pudlovskiy V.B. Metody i algoritmy navigatsionnykh opredeleniy s ispol'zovaniem retranslirovannykh signalov sputnikovykh radionavigatsionnykh sistem. Kand. Diss. [Methods and algorithms for navigational determinations using retransmitted signals from satellite radio navigation systems. PhD Thesis]. Moscow, 2009. 263 p.

[12] Jaime Gomez-Gil, Ruben Ruiz-Gonzalez, Sergio Alonso-Garcia, Francisco Javier Gomez-Gil. A Kalman filter implementation for precision improvement in low-cost gps positioning of tractors. Sensors, 2013, no. 13, pp. 15307-15323. Available at https://www.ncbi.nlm.nih.gov/pubmed/24217355 (accessed 6 April 2018). DOI: 10.3390/s131115307

[13] Sadkovskiy B.P., Sadkovskaya N.E., Troshkin D.O. Interaction of global positioning systems Glonass/GPS and ground-based computing centers. Inzhenernyy zhurnal: nauka i innovatsii [Engineering Journal: Science and Innovation], 2014, no. 6(30). Available at http://engjournal.ru/catalog/it/nav/1284.html (accessed 6 April 2018). (in Russian). DOI: 10.18698/2308-6033-2014-6-1284

[14] Salychev O.S. MEMS/GPS — compact integrated navigation system. Geoprofi [GeoProfi], 2013, no. 3, pp. 16-17. Available at http://www.geoprofi.ru/Services/Doc/1776/6fff32044f3f4ff5abeece7614dc7eb2/True (accessed 6 April 2018). (in Russian)

[15] Christopher M. Bishop. Pattern recognition and machine learning. Springer, 2016, 738 p.

[16] Huiqin Li, Gang Wu. Map matching for taxi gps data with extreme learning machine. Advanced Data Mining and Applications: 10th International Conference proceedings, 2014, vol. 8933. Available at https://link.springer.com/chapter/10.1007/978-3-319-14717-8_35 (accessed 6 April 2018). DOI: 10.1007/978-3-319-14717-8_35

[17] Celestino Ordóñez, José R. Rodríguez-Pérez, Juan J. Moreira, J. M. Matías, Enoc Sanz-Ablanedo. Machine learning techniques applied to the assessment of gps accuracy under the forest canopy. Journal of Surveying Engineering, 2011, no. 137, pp. 140-149. DOI: 10.1061/(ASCE)SU.1943-5428.0000049

[18] Christian Manasseh, Raja Sengupta. Predicting driver destination using machine learning techniques. 16th International IEEE Conference on Intelligent Transportation Systems, 2013, pp. 142-147. Available at http://ieeexplore.ieee.org/document/6728224/ (accessed 6 April 2018). DOI: 10.1109/ITSC.2013.6728224

[19] Abhishek Goswami, Luis E. Ortiz, Samir R. Das. WiGEM: a learning-based approach for indoor localization. CoNEXT '11 Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies, 2011. Available at https://dl.acm.org/citation.cfm?id=2079299 (accessed 6 April 2018). DOI: 10.1145/2079296.2079299

[20] Berk Kapicioglu. Applications of machine learning to location data. PhD Thesis. Princeton University. 2013. 80 p.

[21] Gavrilov A.V. Using the Kalman filter to solve the UAV coordinate refinement problems. Sovremennye problemy nauki i obrazovaniya [Modern problems of science and education], 2015, no. 1-1. Available at https://www.science-education.ru/ru/article/view?id=19453 (accessed 6 April 2018). (in Russian)

[22] Jeffrey L. Elman. Finding structure in time. Cognitive Science, 1990, no. 14, pp. 179-211. Available at https://crl.ucsd.edu/~elman/Papers/fsit.pdf (accessed 6 April 2018).

[23] Hochreiter Sepp, Jürgen Schmidhuber. Long short-term memory. Neural Computation, 1997, no. 9(8), pp. 1737-1780. Available at http://www.bioinf.jku.at/publications/older/2604.pdf (accessed 6 April 2018).

[24] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. NIPS 2014 Deep Learning and Representation Learning Workshop, 2014. Available at https://arxiv.org/abs/1412.3555 (accessed 6 April 2018).

[25] Herbert Jaeger, Harald Haas. Harnessing Nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science, 2004, no. 304, pp. 78-80. Available at http://www.columbia.edu/cu/biology/courses/w4070/Reading_List_Yuste/haas_04.pdf (accessed 6 April 2018). DOI: 10.1126/science.1091277

[26] Alex Graves, Greg Wayne, Ivo Danihelka. Neural Turing machines // arXiv preprint arXiv: 1410.5401, 2014. Available at https://arxiv.org/pdf/1410.5401.pdf (accessed 6 April 2018).

[27] Fused location provider API. Available at: https://developers.google.com/android/reference/com/google/android/gms/location/FusedLocationProviderApi (accessed 6 April 2018).