

Методы фильтрации частично недостоверных геоданных

А. А. Головков¹, Г. С. Иванова²

Московский государственный технический университет им. Н. Э. Баумана

¹alexander.golovkov1@gmail.com, ²gsivanova@gmail.com

Аннотация. В настоящее время геоинформационные системы (ГИС) активно используют мобильные приложения для сбора и обработки геоданных полевых сотрудников в областях технического обслуживания и ремонта (ТОиР). Такая возможность позволяет осуществлять оперативный мониторинг перемещений сотрудников и координировать их действия в процессе работы. Обработка и фильтрация геоданных на мобильных устройствах представляет собой сложную задачу особенно в условиях низкой точности геопозиционирования, например, при навигации внутри зданий. В работе предложены методы обработки геоданных, ориентированные на уменьшение плотности точек трека, выявление и исключение координат с высокой погрешностью в условиях продолжительных стоянок внутри помещений или движения вблизи высотных зданий, где за счет переотражения, преломления и затухания радиосигналов точность спутникового позиционирования существенно снижается. Выявлены преимущества и недостатки методов, оценена их эффективность.

Ключевые слова: геоданные; точность; фильтрация; мобильное устройство; трек; ГИС

I. ВВЕДЕНИЕ

Одним из перспективных направлений в области автоматизации деятельности полевых сотрудников является разработка геоинформационных систем (ГИС), позволяющих осуществлять мониторинг перемещений мобильных сотрудников и оперативно координировать их действия в процессе работы. Как правило ГИС включают в себя мобильные приложения, которые обеспечивают сбор и предобработку геоинформационных данных с последующей передачей в серверную часть. По причине разнородности источников геоданных, аппаратных ограничений мобильных устройств и высоких требований к быстродействию, обработка геоинформации является сложной задачей. Методы фильтрации должны работать эффективно в том числе при навигации внутри помещений и вблизи высотных зданий, где за счет переотражения, преломления и затухания радиосигналов параметры погрешности определения координат становятся недостоверными.

II. СУЩЕСТВУЮЩИЕ МЕТОДЫ ОБРАБОТКИ ГЕОДАНЫХ

Способы и методы фильтрации геокоординат в мобильных устройствах были детально исследованы в работах [1–4, 6–14]. Рассмотрим некоторые из них.

Хруль и др. в работе [3] предлагают алгоритм потоковой фильтрации, основанный на анализе точности смежных координат. В качестве критерия фильтрации используется расстояние между двумя точками, которое должно быть не меньше суммы пороговых значений, зависящих от типа исследуемых транспортных средств. Подход ориентирован на исключение отклонений координат в местах стоянок, повышение достоверности и уменьшение объема входных данных. Одним из главных недостатков метода является отсутствие зависимости поведения фильтра от скоростей в смежных координатах, что в ряде случаев дает визуально некачественные треки, особенно в движении на поворотах [2].

Известно применение фильтра Калмана [4, 5, 12, 21] для сглаживания геокоординат. С учетом ограничений и допущений [4] фильтр эффективно сглаживает единичные выбросы координат, однако не способен справиться с продолжительными отклонениями траектории движения в условиях недостоверных значений точности и/или скорости.

Ввиду высокой сложности задачи и большого количества разнородных данных, многие исследователи стали применять модели и методы машинного обучения [15] для решения задач навигации в различном окружении [16–20]. Ключевая особенность этих подходов и их существенное преимущество над другими методами заключается в том, что для их использования не требуется разрабатывать точную математическую модель процесса движения пользователя или разброса координат в точках стоянок. В процессе обучения модель подстраивается под входные данные, моделируя процесс навигации [22–26]. Существенным недостатком применения таких моделей является то, что в процессе работы невозможно определить точные критерии, на основе которых обученная модель приняла решение относительно валидности координаты.

Стоит отметить Fused Location Provider – технологию геолокации Android, призванную обобщить все источники геоданных и решить проблему геолокации в помещениях [27], однако качество фильтрации данным методом оставляет желать лучшего: отклонение пройденного расстояния от реального может достигать десятков километров.

III. МЕТОДЫ ФИЛЬТРАЦИИ ЧАСТИЧНО НЕДОСТОВЕРНЫХ ГЕОДАНЫХ

Ниже предложены методы обработки, основанные на [2, 3], которые позволят уменьшить плотность точек результирующего трека, выявить и исключить координаты с большой погрешностью определения в случае частично недостоверных параметров геоданных.

A. Фильтрация по направлению движения

Цель фильтрации – уменьшение плотности точек трека, основанное на анализе угла между направлениями движения. В первую очередь необходимо сравнить интервал времени Δt между предыдущей и текущей координатами с интервалом ΔT_{\max} :

$$\Delta t < \Delta T_{\max}(\alpha) = \begin{cases} 1, \text{если } \alpha > P_{\max A}^1 \\ \frac{1 - P_{\max T}^1}{P_{\max A}^1} \alpha + P_{\max T}^1, \text{иначе} \end{cases},$$

где α – разница углов между направлениями движения по трем координатам в последовательности, $P_{\max A}^1$ – максимальный угол в градусах, при котором координата будет отфильтрована, если интервал времени меньше 1 с.

Если $\Delta t < \Delta T_{\max}(\alpha)$, координата предположительно имеет большую погрешность, и ее следует исключить. Параметры $P_{\max T}^1$ и $P_{\max A}^1$ задают максимальный интервал времени в секундах, за который координаты будут отфильтрованы при угле $\alpha = 0$, и максимальный угол, в пределах которого координаты будут исключены, если интервал времени будет соответствовать рассчитанному диапазону. Иными словами, порог интервала времени динамически зависит от угла между направлениями движения. Если $\alpha > P_{\max A}^1$ и интервал времени меньше 1 с, координаты исключаются, в предположении, что совершается поворот, и следует пропустить максимум координат для формирования гладкого трека. Если $\alpha \leq P_{\max A}^1$, то поворот либо не значительный, либо пользователь движется по прямой, и в этом случае целесообразно отфильтровывать большее число координат.

На втором этапе обработки следует анализировать угол α совместно со средней скоростью пользователя \bar{g} по нескольким точкам в последовательности. Если средняя скорость больше определенного значения, порог угла следует увеличить, так как с увеличением скорости

максимально возможный угол поворота за определенное время снижается. Верно и обратное: чем ниже скорость, тем на больший угол можно совершить поворот за определенное время. В соответствии с этим предположением угол α необходимо сравнить с пороговым значением, которое зависит от средней скорости пользователя с учетом максимально возможного интервала времени между координатами:

$$\Delta t < 3P_{\max T}^1 \text{ и } \alpha > \Delta A_{\min}(\bar{g}) = \begin{cases} P_{\max A}^1, \text{если } \bar{g} > P_{\max S}^1 \\ \frac{P_{\max A}^1 - 180}{P_{\max S}^1} \bar{g} + 180, \text{иначе} \end{cases},$$

Где $P_{\max S}^1$ – максимальная скорость в метрах в секунду, при котором координата будет исключена, если $\alpha > P_{\max A}^1$.

Рассмотрим распределение угла α до и после фильтрации на выборке из 368528 координат (см. рис. 1) при значениях параметров:

$$P_{\max T}^1 = 15 \text{ с}, P_{\max A}^1 = 10^\circ \text{ и } P_{\max S}^1 = 25 \frac{\text{м}}{\text{с}}.$$

Фильтр сместил среднее значение угла в центр распределения, однако дисперсия поменялась незначительно, при этом количество координат после фильтрации уменьшилось на 78.976 %.

B. Адаптивная диагностическая фильтрация

Предлагаемый метод адаптивной диагностической фильтрации основан на подходе, описанном в [2, 3]. Цель фильтрации – исключить недостоверные координаты в местах стоянок пользователя за счет анализа скорости в смежных точках. Напомним, что в соответствии с [2], критерий фильтра – $D_{1-2} > D l_1 + D l_2$, где D_{1-2} – расстояние в метрах между точками 1 и 2 в метрах; $D l_1$ и $D l_2$ – характеристики точек в метрах. Если неравенство выполняется, точка 2 считается валидной и проходит фильтрацию, иначе точка отбрасывается.

Идея адаптивной фильтрации заключается в 2 основных гипотезах:

1. уменьшая значения $D l_1$ и $D l_2$ при «входе» и «выходе» из стоянки, возможно сохранить координаты движения близ стоянок, что увеличит качество трека;
2. увеличивая значения $D l_1$ и $D l_2$ прямо пропорционально точности и обратно-пропорционально скорости, возможно уменьшить количество выбросов в местах стоянок, так как скорость в этом случае существенно ниже, чем в движении.

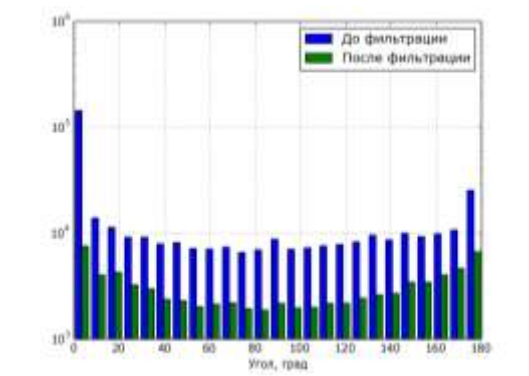


Рис. 1. Распределение угла α до и после обработки фильтром по направлению

На основе описанных гипотез критерий фильтра соответствует:

$$D_{i-1,i} > \begin{cases} f_a(a_{i-1}, p_{i-1}) + f_a(a_i, p_i), \\ \text{если } s_i = 1 \text{ и } \mathcal{G}_{i-1} > P_S^2 \\ \text{или } s_{i-1} = 1 \text{ и } \mathcal{G}_i > P_S^2 \\ f_a(a_{i-1}, p_{i-1})f_K(\mathcal{G}_{i-1}) + f_a(a_i, p_i)f_K(\mathcal{G}_i), \\ \text{иначе} \end{cases}$$

Где a – радиус круга в метрах, в котором находятся координаты с вероятностью $\sim 68\%$; p – тип источника геоданных: 0 – GPS источники (NMEA), 1 – мобильное геопозиционирование (API), 2 – Wi-Fi и сотовые вышки [1]; $s_i = \begin{cases} 1, \text{если } \mathcal{G}_i < P_{max}^2 - \text{признак стоянки;} \\ 0, \text{иначе} \end{cases}$ P_{max}^2 –

минимальная скорость пользователя в метрах в секунду, при которой точку местоположения можно считать стоянкой; $f(a_i, p_i) = \begin{cases} \frac{a_i}{P_K^2}, \text{если } p_i = 2 \\ a_i, \text{иначе} \end{cases}$ – функция,

учитывающая низкие точности координат от Wi-Fi и вышек сотовой связи; P_S^2 – минимальная скорость в метрах в секунду, которая предположительно является скоростью пользователя после выхода из стоянки;

$$f_K(\mathcal{G}_i) = \begin{cases} 1, \text{если } \mathcal{G}_i > P_S^2 \\ \frac{1 - P_K^2}{P_S^2} \mathcal{G}_i + P_K^2, \text{иначе} \end{cases} \quad \text{в соответствии с [2].}$$

Первое выражение в системе соответствует первой гипотезе выше, значение порога здесь прямо пропорционально точности двух смежных точек без какого-либо множителя. Фактически это выражение является критерием выхода из стоянки. Второе выражение отражает вторую описанную гипотезу: параметры точности в каждой из точек домножаются на соответствующие коэффициенты, которые зависят от скоростей пользователя в этих точках.

Если критерий фильтра выполняется, и предыдущая координата – стоянка, метод не исключает координату, а модифицирует ее параметры в предположении, что текущая координата является продолжением стоянки.

Рассмотрим работу фильтра. На рис. 2 представлены соответственно:

- трек четырехчасовой стоянки внутри здания с небольшим участком движения;
- результат фильтрации диагностическим методом [3] без учета скорости пользователя и

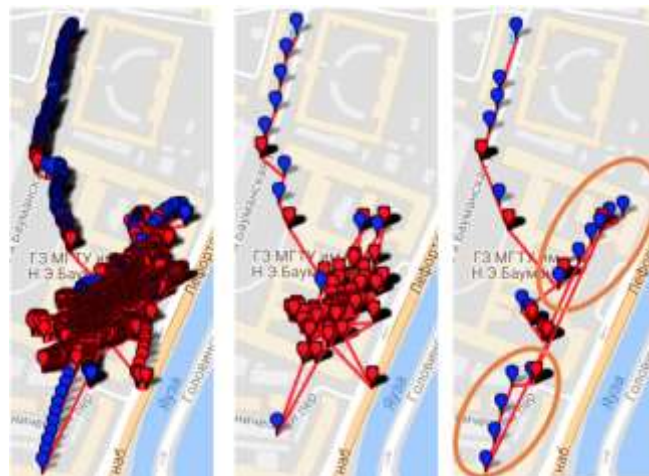


Рис. 2. Сравнение диагностических методов фильтрации

входов/выходов из стоянки, при этом $D_{i-1,i} > 3(a_{i-1} + a_i)$;

- результат фильтрации предложенным диагностическим методом после фильтрации по направлению при $P_K^2 = 20$ и $P_S^2 = 2 \frac{M}{c}$.

Красными маркерами показаны точки, где скорость равна нулю, синим, где скорость положительна.

Фильтр отбросил большее количество точек непосредственно в стоянке, однако часть трека в движении слева сверху практически не изменилась, как можно заметить и по гистограмме (см. рис. 3).

Стоит отметить, что девиации координат, выделенные оранжевым, лучше отфильтровал метод диагностической фильтрации [3].

Количественные результаты подтверждают эффективность метода (см. табл. 1): процент исключенных координат в стоянках увеличился, в движении – уменьшился, при этом отношение пройденного расстояния к реальному уменьшилось примерно в 3 раза.

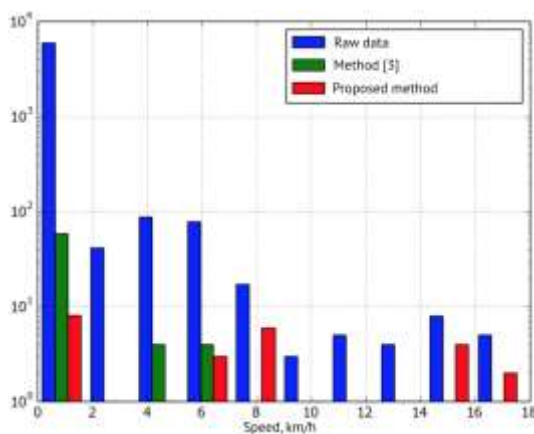


Рис. 3. Распределение скорости до и после фильтрации диагностическими методами

ТАБЛИЦА I СРАВНЕНИЕ ДИАГНОСТИЧЕСКИХ МЕТОДОВ ФИЛЬТРАЦИИ

Характеристика		Метод [3]	Предложенный метод
Отфильтрованных координат, %	всех	98.872	99.559
	с нулевой скоростью	99.006	99.863
	с положительной скоростью	96.085	93.238
Пройденное расстояние (реальное расстояние 357.260 м), м		3545.094	1195.716

Исходя из анализа распределения скорости и результатов сравнения методов, приведенных в таблице 1, можно утверждать, что предложенный метод адаптивной диагностической фильтрации является более эффективным, чем метод [3]. Особое внимание стоит обратить на пройденное расстояние. Во многих ГИС это один из наиболее важных параметров, и практически все методы фильтрации направлены на увеличение его достоверности.

IV. ЗАКЛЮЧЕНИЕ

В работе предложены два метода фильтрации геоданных, ориентированные на уменьшение плотности точек трека, а также выявление и исключение координат с большой погрешностью определения. Оба метода используют различные источники геолокационной информации и могут быть использованы в мобильных устройствах в режиме реального времени. Экспериментальные результаты подтверждают эффективность предложенных подходов: достоверность отношения пройденного расстояния к реальному увеличилась на 19.8 % в сравнении с существующими рассмотренными методами.

СПИСОК ЛИТЕРАТУРЫ

[1] Головков А.А. Источники геоданных в мобильных устройствах // Динамика сложных систем – XXI век. № 4. 2017. С. 94-101.

[2] Головков А.А., Иванова Г.С. Адаптивная фильтрация потока геолокационных данных в реальном времени // Наука и Образование. МГТУ им. Н.Э. Баумана. № 4. 2016. С. 156-169.

[3] Хруль С.А., Сонькин Д. М. Адаптивный алгоритм обработки потока навигационных данных на основе метода диагностической фильтрации // Известия Томского политехнического университета. № 5. 2012. С. 217-222.

[4] Листеренко Р.Р. Применение фильтра Калмана для обработки последовательности GPS-координат // Молодежный научно-технический вестник. № 09. 2015.

[5] Забегаев А.Н., Павловский В.Е. Адаптация фильтра Калмана для использования с локальной и глобальной системой навигации // Препринты ИПМ им. М.В.Келдыша. № 82. 2010. 24 с.

[6] Lei Gong, Takayuki Morikawa, Toshiyuki Yamamoto, Hitomi Satoh. Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies // Procedia-Social and Behavioral Sciences. № 138. 2014. P. 557-565.

[7] Changqing Zhou, Nupur Bhatnagar, Shashi Shekhar, Loren Terveen. Mining Personally Important Places from GPS Tracks // Data Engineering Workshop, 2007 IEEE 23rd International Conference on. 17-20 April 2007. P. 517-526.

[8] Великанова Е.П., Ворошилин Е.П. Адаптивная фильтрация координат маневрирующего объекта при изменениях условий передачи в радиолокационном канале // Доклады Томского государственного университета систем управления и радиоэлектроники. № 2-1 (26). 2012. С. 29-35.

[9] Макаренко Г.К., Алешечкин А.М. Исследование алгоритма фильтрации при определении координат объекта по сигналам спутниковых радионавигационных систем // Доклады Томского государственного университета систем управления и радиоэлектроники. № 2-2 (26). 2012. С. 15-18.

[10] Прохорцов А. В., Савельев В. В. Методы определения координат и скорости подвижных объектов с помощью спутниковых радионавигационных систем // Известия Тульского государственного университета. Технические науки. № 2. 2011. С. 264-274.

[11] Пудловский В.Б. Методы и алгоритмы навигационных определений с использованием ретранслированных сигналов спутниковых радионавигационных систем. Дисс. М. МГТУ им. Н.Э. Баумана. 2009. 263с.

[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. № 13. 2013. P. 15307-15323.

[13] Садковский Б.П., Садковская Н.Е., Трошкин Д.О. Взаимодействие систем глобального позиционирования ГЛОНАСС/GPS и наземных вычислительных центров // Инженерный журнал: наука и инновации. № 6(30). 2014.

[14] Салычев О.С. MEMS/GPS — малогабаритная интегрированная навигационная система // Геопрофи. № 3. 2013. С. 16-17.

[15] Christopher M. Bishop, 2006. Pattern Recognition and Machine Learning. Springer, pp: 738.

[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.

[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. P. 140-149.

[18] Christian Manasseh, Raja Sengupta. Predicting driver destination using machine learning techniques. 16th International IEEE Conference on Intelligent Transportation Systems. 2013. P. 142-147.

[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.

[20] Berk Kapicioglu. Applications of Machine Learning to Location Data. PhD Thesis. Princeton University. P. 80. 2013.

- [21] Гаврилов А.В. Использование фильтра Калмана для решения задач уточнения координат БПЛА // Современные проблемы науки и образования. № 1. 2015.
- [22] Jeffrey L. Elman. Finding Structure in Time // Cognitive Science № 14. 1990. P. 179-211.
- [23] Hochreiter Sepp, Jürgen Schmidhuber. Long short-term memory // Neural Computation № 9(8). 1997. P. 1737-1780.
- [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.
- [25] Herbert Jaeger, Harald Haas. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication // Science. № 304. 2004. P. 78-80.
- [26] Alex Graves, Greg Wayne, Ivo Danihelka. Neural Turing Machines // arXiv preprint arXiv: 1410.5401. 2014.
- [27] Сетевой ресурс
<https://developers.google.com/android/reference/com/google/android/gms/location/FusedLocationProviderApi>