Algorithm to Estimate a Living Area Based on Connectivity of Places with Home

Yuji Matsuo, Sunao Hara, and Masanobu Abe[(✉)]

Graduate School of Natural Science and Technology,
Okayama University, Okayama, Japan
`matsuo@a.cs.okayama-u.ac.jp`,
`{hara,abe}@cs.okayama-u.ac.jp`

Abstract. We propose an algorithm to estimate a person's living area using his/her collected Global Positioning System (GPS) data. The most important feature of the algorithm is the connectivity of places with a home, i.e., a living area must consist of a home, important places, and routes that connect them. This definition is logical because people usually go to a place from home, and there can be several routes to that place. Experimental results show that the proposed algorithm can estimate living area with a precision of 0.82 and recall of 0.86 compared with the grand truth established by users. It is also confirmed that the connectivity of places with a home is necessary to estimate a reasonable living area.

Keywords: Living area · Dementia · Watch-out · GPS

1 Introduction

Thanks to the rapid spread of mobile phones, smartphones, and devices with Global Positioning System (GPS), location data is easily obtained. This is attracting increasing attention of new services and applications. In fact, numerous location-aware applications have been developed. In this study, we are attempting to develop a watch-over system for elderly people and children for two reasons. First, wandering is one of the most problematic behaviors of elderly people with dementia [1]. Surprisingly, it is reported that more than 10,000 people with dementia went missing in 2013 in Japan [2]. Family members or nurses bear a large burden of having to watch the people with dementia to prevent them from going missing. Second, parents also have a large responsibility of preventing their children from being kidnapped, having accidents, or being victims of crimes. For example, many parents have to take their children to and from school. To address this problem, some commercial products are available [3, 4]. Unfortunately, these products are not user friendly and effective; they have complicated procedures for setting the watch-out area, a poor precision of the watch-out area, and so on.

To solve this problem, we propose an algorithm that automatically determines a person's living area using his/her collected location data. Here, we consider a living area to be one of the most important areas to provide care for children and people with dementia. The living area is defined as a set that includes a person's home, important places that he/she frequently visits, and routes that connect them. The criteria for

determining the living area are as follows: (1) home and important places are connected by routes, (2) several routes between them should be available, and (3) the importance of a route is evaluated by its frequency of use. We believe the definition and its criteria are quite reasonable because home plays a central role in everyday life, and users have several routes to a particular place depending on contexts such as shopping lists, time constraints, accompanying people, and so on. Although some research has been performed on finding routes [5–7], the proposed algorithm is unique because it uses GeoHex [8] code and considers a route as a set of GeoHex codes; this results in implicitly expressing a route. Another advantage is that the precision of routes can be manipulated with this code.

This paper is organized as follows. In Sect. 2, we describe details of the proposed algorithm. In Sect. 3, we conduct evaluation experiments for three users. Finally, we conclude the paper and suggest future work.

2   Algorithm to Estimate Living Area

Figure 1 shows an outline of the proposed algorithm consisting of two steps. The first is the preprocessing of the collected location data and the second is the estimation of a living area.
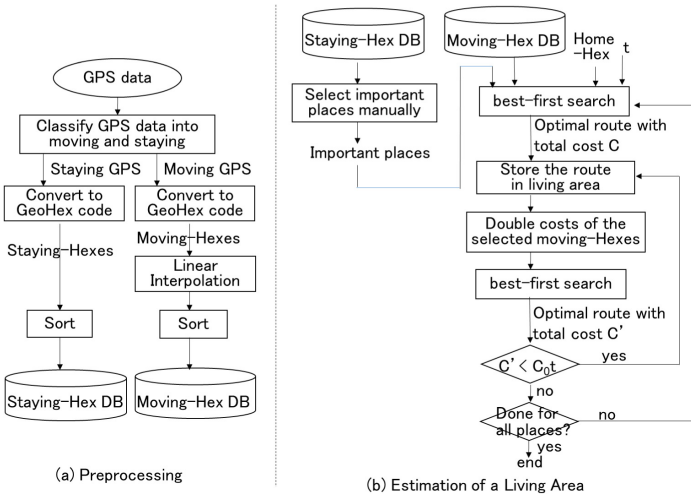

Fig. 1. An outline of proposed algorithm

2.1   Preprocessing for Collected Location Data

Location data of longitude and latitude is collected by GPS every 30 s. The preprocessing step has five procedures as follows. (1) On a daily basis, location data is classified into two states: staying and moving. The classification is performed using velocity and distance from the previous location. In the rest of this paper, based on these classifications, we refer to the GeoHex code of the staying state as staying-Hex

and that of the moving state as moving-Hex. (2) Location data is converted into the GeoHex code that corresponds to a small area, a circle with a diameter of approximately 12 m. An aim of the procedure is to ignore small differences among location data because it is rare to obtain the exact same sequence of the data (longitude and latitude) even when people go to the same place taking the same route. This is necessary to accumulate location data over different days. (3) When consecutively sampled moving-Hexes are not adjacent, they are added to fill a gap by linear interpolation. Because location data is sampled every 30 s, a gap can occur between consecutively sampled moving-Hexes when a user moves fast. This procedure makes it possible to easily determine a route between places X and Y by checking if the adjacent moving-Hexes exist from place X to Y. (4) An occurrence frequency of each staying-Hex is counted, and after sorting all staying-Hexes by the frequencies, a staying-Hex database is generated. In this procedure, staying-Hexes located within a circle of 30 m diameter are merged because large buildings occupy several staying-Hexes. (5) An occurrence frequency of each moving-Hex is counted, and after sorting all moving-Hexes by their frequencies, a moving-Hex database is generated.

2.2   Estimating a Living Area

In the current implementation, important places are manually determined by each user. The n-th most frequent staying-Hexes are selected from the staying-Hex database, and are shown on a map in a PC display. Using graphical user interface (GUI), each user selects those that he/she accepts as important places. The living area is estimated using the important places and moving-Hex database. As previously explained, a living area is defined by the set comprised of one's home, important places, and the routes that connect them. Because people usually leave home and return, the proposed algorithm searches routes that connect home and important places by best-first search. The search aims to find a route of minimum cost. A cost is assigned to each moving-Hex; it is the inverse of the occurrence frequency of the moving-Hex. Moreover, to search alternative routes for important places, best-first search is iteratively applied by changing the cost of a moving-Hex that has previously been established. The above procedures are performed between home and each important place one by one. As a result, a living area of a person is represented by a set of staying-Hexes and moving-Hexes.

The proposed algorithm is explained in detail in Fig. 2. In the figure, OpenList contains staying-Hexes or moving-Hexes that have not yet been expanded and CloseList contains staying-Hexes or moving-Hexes that have previously been expanded. The example shown in Fig. 2 attempts to determine a route between Home-Hex and PlaceA-Hex. Here, a letter and a number shown in a Hex are a node ID and a cost, respectively. Hexes painted black indicate that they do not exist in either the staying-Hex or moving-Hex database. Further, an alphabet sequence in OpenList indicates a part of a route. For example, AE means that a user moves from Hex-A to Hex-E. The search is performed as follows. First, moving-Hexes that are adjacent to Home-Hex are expanded. Then, the expanded Hexes are sorted according to cost in increasing order, and the Hex with the lowest cost is expanded. As shown in Fig. 2, Hex A is expanded and added to CloseList because it has the lowest cost. In the same

manner, Hex E is expanded and added to CloseList because the Hex sequence of AE has the lowest cost. In this manner, the Hex expansion and the updating of the open and close lists are repeated. Finally, in this example, an optimal route is obtained as a Hex sequence AEIJ. To find alternative routes, the search is continued after doubling the costs of the moving-Hexes used in the previously established route. The search for alternative routes is terminated when $C' > C_0 t$ is satisfied. Here, C_0 is the total cost of the first route, and C' is the total cost obtained by the repeated search, where t is a control parameter. Figure 3 shows an example of the alternative route search. The costs of moving-Hexes A, E, I, and J have doubled, and a new route CFLK is found.
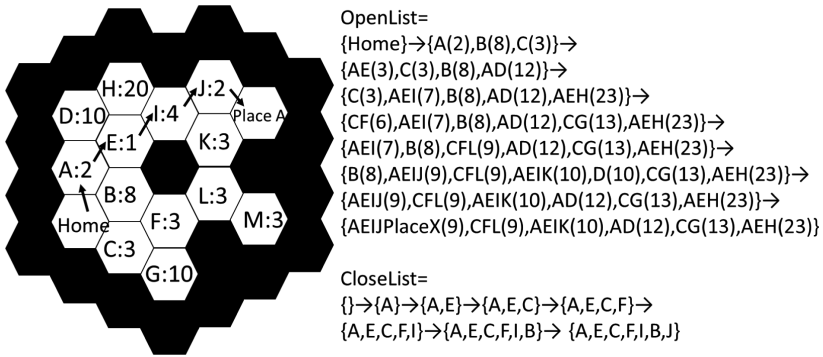

OpenList=
{Home}→{A(2),B(8),C(3)}→
{AE(3),C(3),B(8),AD(12)}→
{C(3),AEI(7),B(8),AD(12),AEH(23)}→
{CF(6),AEI(7),B(8),AD(12),CG(13),AEH(23)}→
{AEI(7),B(8),CFL(9),AD(12),CG(13),AEH(23)}→
{B(8),AEIJ(9),CFL(9),AEIK(10),D(10),CG(13),AEH(23)}→
{AEIJ(9),CFL(9),AEIK(10),AD(12),CG(13),AEH(23)}→
{AEIJPlaceX(9),CFL(9),AEIK(10),AD(12),CG(13),AEH(23)}

CloseList=
{}→{A}→{A,E}→{A,E,C}→{A,E,C,F}→
{A,E,C,F,I}→{A,E,C,F,I,B}→ {A,E,C,F,I,B,J}

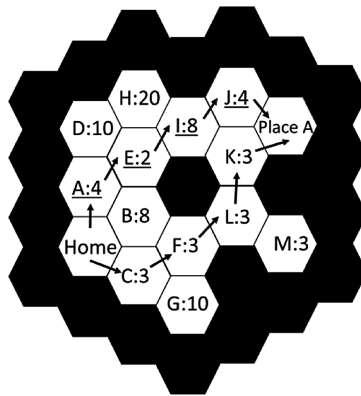**Fig. 2.** An example of the best-first-search

Fig. 3. Searching an alternative route

3 Evaluation of the Proposed Algorithm

The proposed algorithm is evaluated using location data collected from three users. Moreover, to observe the advantages of the proposed algorithm, we compared its performance with that of a conventional method that simply selects all GeoHexes counted more than 3 times as a living area.

3.1   Data for Evaluations

GPS data was collected every 30 s for 12 months for three graduate students of Okayama University. They live alone in Okayama city and travel to the university by walking or bicycling. Using the data, grand truth of a living area is created as follows. (1) The important places selected by each user were shown on a map in a PC display. Each user was asked to freely draw routes on the map using a GUI between home and the important places based on memory. The drawn routes were converted into GeoHex code and stored as the living area. (2) The routes generated by the proposed algorithm were shown on a map in a PC display over the GeoHexes selected in (1). Each user was asked to subjectively judge whether the generated routes were sufficient in defining his/her living area. For routes that were difficult to judge, they physically went to the site before making their decision. The routes accepted as the living area are added to the output of (1). As a result, a grand truth of a living area is represented by a set of the selected Hexes.

3.2   Experimental Results of Estimating Living Area

The living area is estimated for each student. The performance of the proposed algorithm is evaluated by precision, recall, and F-measure, as defined in Eqs. (1)–(3).

$$\text{precision} = \frac{\{\text{Grand truth area}\} \cap \{\text{Estimated area}\}}{\{\text{Estimated area}\}} \tag{1}$$

$$\text{recall} = \frac{\{\text{Grand truth area}\} \cap \{\text{Estimated area}\}}{\{\text{Grand truth area}\}} \tag{2}$$

$$\text{F - measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

Because the experimental results are nearly identical among the users, the average values of precision, recall, and F-measure are shown in Fig. 4. The x axis indicates the parameter t. As shown in the figure, recall increases with t. This means that there are several routes between home and each important place. When t is equal to 15, the recall is at its highest value of 0.95, whereas the precision is at its lowest value of 0.77. The lower the precision, the greater the inadequacy of the Hexes selected as the living area. According to the F-measure, the performance of the proposed algorithm is saturated with a precision of 0.82 and recall of 0.86 when parameter t is equal to 10.

Table 1 shows the precision, recall, and F-measure of the conventional method and proposed algorithm. We can see that the proposed algorithm outperforms the conventional method. Figure 5 shows examples of the estimated living area. As shown in Fig. 5(a), the conventional method generates numerous "Hex-islands" that have no connection to other Hexes. This is not acceptable for estimating a living area.

Judging from the results, the occurrence frequency of the GeoHexes is not sufficient to estimate a living area, and it is necessary to explicitly use the constraint that a
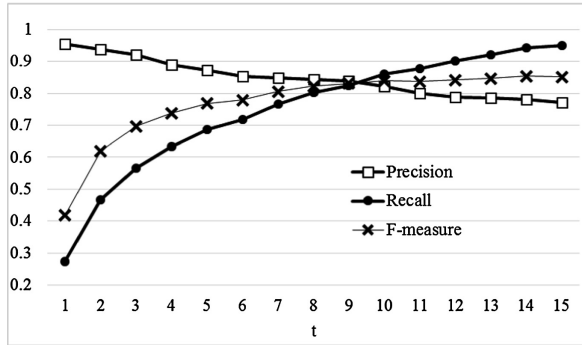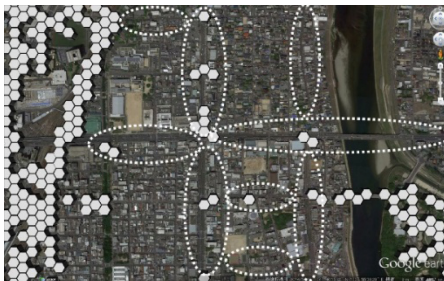
Fig. 4. Average performance for the three users

Table 1. Precision and recall of three users

	Precision		Recall		F-measure	
	Con-ventional	Pro-posed (t=10)	Con-ventional	Pro-posed (t=10)	Con-ventional	Pro-posed (t=10)
User A	0.30	0.84	0.26	0.82	0.28	0.84
User B	0.30	0.82	0.73	0.91	0.43	0.86
User C	0.60	0.81	0.67	0.85	0.63	0.83

Routes are missing in the dotted elliptical areas. This is inadequate for estimating the living area.

(a) Conventional method
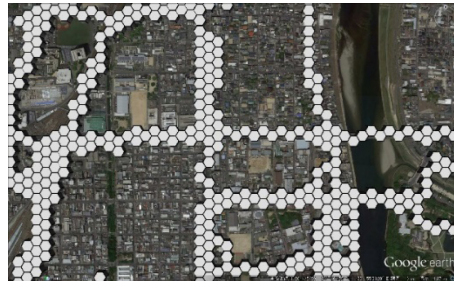
All places are connected with routes. Several paths are clearly observed between staying places.

(b) Proposed algorithm

Fig. 5. An example of estimated living area for the three users

GeoHex must be a connection between a home and an important place. The performance of the proposed algorithm depends less on a user than the conventional method. This may be a direct effect of the constraint.

4   Conclusion

We propose an algorithm to estimate a living area of a person using his/her collected location data. The defining characteristic of a living area is that a home and important places must be connected by several routes. Experimental results for the three users showed that the proposed algorithm works well and illustrated the importance of the requirement. In future, we will develop an algorithm to estimate the location of important places and to determine the amount of data required for estimating a living area. Thereafter, we would apply the algorithm to a watch-out system and evaluate it in the field.

References

1. Lin, Q., Zhang, D., Chen, L., Ni, H., Zhou, X.: Managing elders' wandering behavior using sensors-based solutions: a survey. Int. J. Gerontol. **8**(2), 49–55 (2014)
2. https://www.npa.go.jp/safetylife/seianki/H24yukuehumeisha.pdf (in Japanese)
3. http://www.secom.co.jp/english/cocosecom/index.html
4. http://www.dokoiruka.jp/ (in Japanese)
5. Zhou, C., Shekar, S., Treveen, L.: Discovering personal paths from sparse GPS traces. In: Proceedings of the JCIS 2005 Workshop on Data Mining (2005)
6. Liao, L., Fox, D., Kautz, H.: Extracting places and activities from GPS traces using hierarchical conditional random fields. Int. J. Rob. Res. **26**(1), 119–134 (2007)
7. Yamada, N., Isoda, Y., Minami, M., Morikawa, H.: Incremental route refinement for GPS-enabled cellular phones. In: Proceedings of the Fifth International Conference on Mobile Computing and Ubiquitous Networking (ICMU 2010), pp. 87–93 (2010)
8. GeoHex. http://geogames.net/geohex/v3