An Adaptive Sampling Approach for Trajectories Based on the Concept of Error Ellipses

Peter RANACHER[1] and Adam ROUSELL[2]

[1]Z_GIS, University of Salzburg/Austria · peter.ranacher@sbg.ac.at
[2]Nottingham Geospatial Institute, University of Nottingham/UK

This contribution was double-blind reviewed as full paper.

Abstract

A trajectory represents an object's path through space and time. These trajectories are often recorded as a series of spatial positions at discrete time intervals. Depending on the sampling rate employed, there can be a large amount of uncertainty relating to how well the trajectory represents the actual path taken. By using the concept of error ellipses, an adaptive sampling strategy has been developed that allows for the determination of an optimal sampling rate whilst ensuring that a pre-defined uncertainty threshold is not surpassed. This decreasing of uncertainty means that any trajectory derived using the algorithm presented would closely match the actual route taken by the object, thus allowing for more accurate spatial interpretation of the data.

1 Introduction

A trajectory – a path through space and time – is a mathematical representation of an object's movement. Conceptually, a trajectory is a continuous function that assigns time instances to spatial positions (ANDRIENKO et al. 2008). In practise, however, we record the spatial positions of a moving object at discrete time intervals and connect these positions by means of interpolation (MACEDO et al. 2008). Obviously, the temporal sampling rate at which we record the movement of an object has a fundamental impact on the representation of movement. In figure 1, the s-shaped path (solid line) shows the moving object's real behaviour in space. The lower the sampling rate $r = 1/\Delta t_i, t_{i+1}$, where $\Delta t_i, t_{i+1}$ is the time interval between two consecutive measurements, the less the two trajectories (dashed lines) resemble the real movement of the object.

When using data that is geospatial in nature, it is important to consider that they will often be subject to spatial uncertainty. Spatial uncertainty (or spatial vagueness) comprises of positional uncertainty (a lack of knowledge regarding the position and shape of an object) and measurement uncertainty (the inability to measure an object precisely) (PAULY & SCHNEIDER 2008). This study is concerned with the concept of positional uncertainty in that it deals with a lack of knowledge about the position of a moving object in an environment.
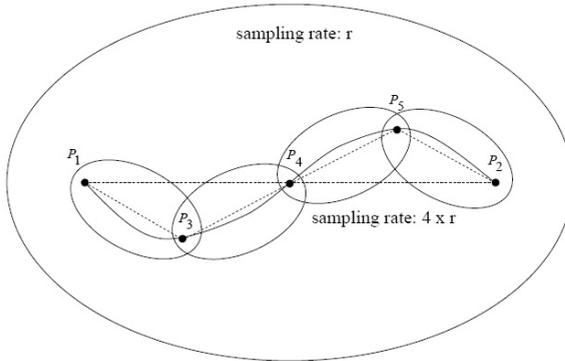
Fig. 1:   Real movement of an object and its trajectory at different sampling rates (from
PFOSER & JENSEN 1999)

There are several methods for describing spatial uncertainty including fuzzy (SCHNEIDER 1999) and vague (PAULY & SCHNEIDER 2008) spatial data types. Rather than classical crisp descriptions where the location of features is assumed to be exact, these methods assign levels of 'knowing' to the location of spatial features. For the vague spatial data types, features consist of kernel and conjuncture parts (PAULY & SCHNEIDER 2008). The kernel part of a feature is the part that is known (it is certain that the feature can be found in that location). The conjuncture part on the other hand represents part of the feature that is assumed (part of the feature may or may not be found in that location). In our case the measured positions of the trajectory represent the kernel part of a moving object in space and time and the interpolation between two consecutive positions is the conjuncture and, hence, affected by error. This also means that our study does not deal with positioning errors (such as GPS errors) associated to a trajectory.

PFOSER & JENSEN (1999) express the spatial uncertainty associated to a trajectory with the help of error ellipses. Error ellipses are closely related to the concept of space-time prisms whereby all possible locations that a person can be in is depicted based on starting location, travel velocity and time taken (KUIJPERS et al. 2010). Although space-time prisms are widely used, KOBAYASHI et al. (2011) note that they classically assume the starting location and time factor are precisely known, which is not always the case. This assumption is shared by the error ellipse concept, but in the scope of this study the effects of this assumption will not be addressed.

An error ellipse represents the union of all possible paths that allow the moving object to start at the measured position p_i at time t_i and arrive at the consecutive measured position p_{i+1} no later than t_{i+1} given v_max as the maximum speed of the object (see also figure 1). The two measured positions p_i and p_{i+1} are the two foci of the error ellipse. Hence, the linear eccentricity $e = \Delta s_i, s_{i+1}/2$ is half the distance between p_i and p_{i+1}. The semi-major axis of the ellipse $\alpha = v\_max \cdot \Delta t_i, t_{i+1}/2$ equals half the possible distance the object can cover with maximum speed during the time interval between the two position measurements. From e and a we can calculate the semi-minor axis

$$b = \sqrt{\left(\frac{v\_max \cdot \Delta t_i, t_{i+1}}{2}\right)^2 - \left(\frac{\Delta s_i, s_{i+1}}{2}\right)^2} = S.I.E. \tag{1}$$

Considering the case of linear interpolation, where p_i and p_{i+1} are connected by a straight line, b represents the one point of the uncertainty ellipse that is farthest away from the interpolation line. Hence, b is the maximum *spatial interpolation error* (S.I.E.) between the interpolated path and all other possible paths. If we assume v_max to be constant and defined by inner or outer constraints – such as the maximum allowed speed on a road network, the magnitude of b depends on two variables: the time interval between two consecutive position measurements t_i, t_{i+1} and the distance between these two positions $\Delta s_i, s_{i+1}$.

From the first term in equation (1), it follows that less-frequent sampling causes the time interval $\Delta t_i, t_{i+1}$ between two measurements to increase. This in turn elongates the semi-major axis of the error ellipse and causes a larger S.I.E. However, intuitively, S.I.E. also depends on the 'shape' of the ellipse and therefore, on linear eccentricity. By increasing the distance between two points whilst maintaining the temporal sampling rate and a constant speed there are less possible routes the trajectory could take to get between the two points in the allotted time. This means that the overall area of the error ellipse decreases as the trajectory cannot deviate as far from the 'as-the-crow-flies' route as doing so would take a longer amount of time resulting in not arriving at the end point in the allotted time. Hence, for a constant $\Delta t_i, t_{i+1}$ if $\Delta s_i, s_{i+1}$ approaches $v_{max} \cdot \Delta t_i, t_{i+1}$, real speed is close to maximum possible speed and the ellipse approaches a line with minimum S.I.E; if $\Delta s_i, s_{i+1}$ approaches zero and the two foci lie together very closely, the ellipse resembles a circle and has maximum S.I.E. (b equals a).

2   An adaptive sampling strategy for trajectory data

By maintaining b as constant and expressing the distance between two measurements with the measured speed of the moving object $s_i, s_{i+1} = v_{meas} \cdot \Delta t_i, t_{i+1}$, we can solve equation (1) for $\Delta t_i, t_{i+1}$ and show the functional relationship between sampling rate, maximum speed and measured speed of the object.

$$\Delta t_i, t_{i+1} = \sqrt{\left(\frac{4b^2}{v\_max^2 - v\_meas^2}\right)} \tag{2}$$

We see that if the moving object moves far slower than assumed, we need a higher sampling rate for b to remain constant. In figure 2, we portray this relationship for different S.I.E. and for the case that the maximum possible speed of the moving object is 60 km/h.
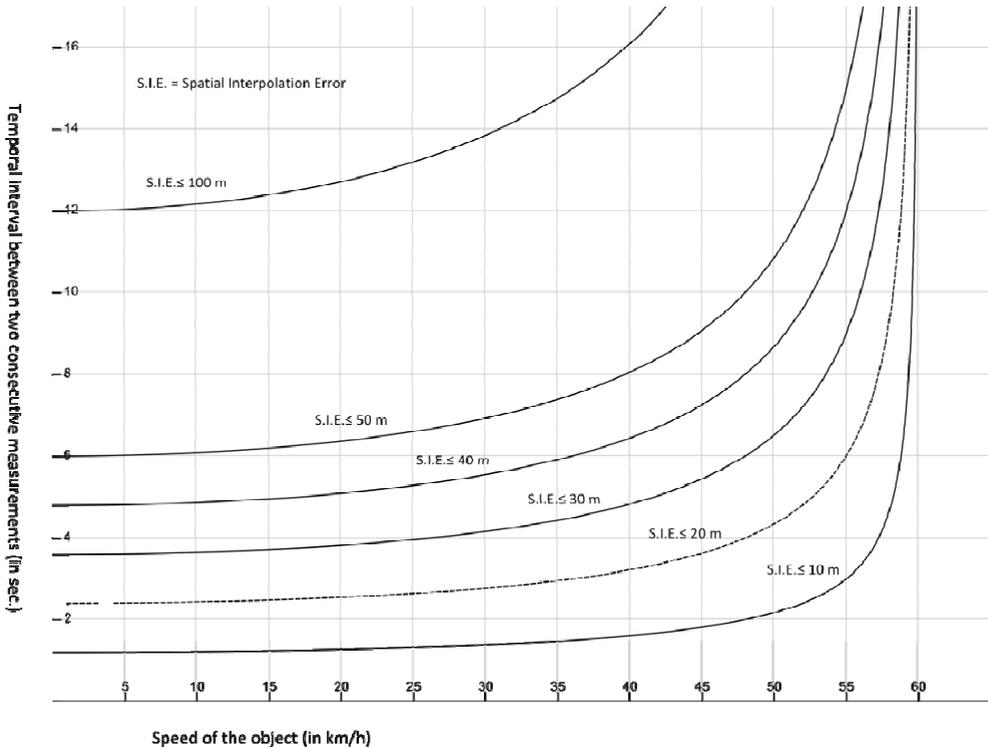
Fig. 2:   Relationship between the sampling temporal interval and object speed on constant S.I.E. values

The curves in figure 2 represent different levels of maximum spatial interpolation error. If an object moves, for example, with a speed of 50 km/h a temporal interval of 4 seconds between two consecutive positions in the trajectory is required for the interpolation error to be less than 20 meters (dashed line). It must be mentioned that the speed of the object in this case does not refer to the momentaneously measured speed, but the average speed between the current position and the last point of the trajectory. Hence, if the floating car changes its momentaneous speed *or* its direction the resulting average speed will decrease and require that the position of the moving object is stored in the trajectory.

Based on the assumptions above we propose an adaptive sampling strategy for trajectories in general and floating car trajectories in particular. This sampling strategy records the position of a moving object at varying temporal intervals. The value of the sampling rate is calculated according to the object's average speed as proposed in equation (2). This guarantees that the trajectory of the movement will be within a certain spatial uncertainty corridor. Intuitively, a time interval between two measurements other than an integer will not be used for sampling trajectories. On the one hand, this implies that a sampling rate of one measurement per second represents the upper boundary of sampling. On the other hand, we must map any results from equation (2) to the largest previous integer. In that case the curves in figure 2 resemble a step function.

In the algorithm 1 we present an adaptive sampling strategy for trajectories.

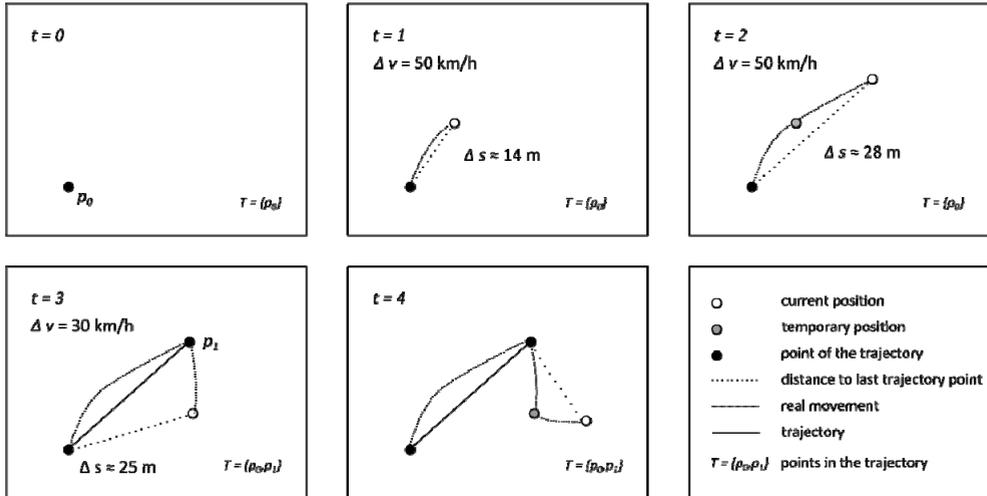**Algorithm 1:** Adaptive sampling for trajectories with guaranteed S.I.E.

Given:

- b: maximum spatial interpolation error;
- v_max: maximum possible speed of the moving object;
- r: minimum possible sampling rate (must guarantee that the trajectory is *always* within b, no matter the current speed of the object)

Description of the adaptive sampling approach:

1. **let** $T = \emptyset$; $i = 0$;
   /* At the beginning the trajectory T is empty */
2. **measure** p_0;
   /* Measure the starting position of the object */;
3. **let** $T[i] = p_0$;
   /* Assign p_0 to be the first point of the trajectory*/;
4. **wait** (r);
5. **measure** p_{curr} ;
   /* measure the current position of the moving object after a time interval r */
6. **let** $p_{last} = T[i]$; $i{+}{+}$;
   /* assign the current point of the trajectory to be the last known position*/
7. **calculate** $v\_meas = \Delta s_{last}, s_{curr}/\Delta t_{last}, t_{curr}$;
   /* calculate the average speed between the current and the last position in the trajectory based on time difference and distance */
8. **calculate** $\Delta t_i, t_{i+1} = floor\,(sqrt\,(4b^2/(v\_max^2 - v\_meas^2\,)))$;
   /* Enter the average speed into equation 2 and calculate the temporal interval that guarantees that the S.I.E. is within a certain threshold; map this interval to the largest previous integer */
9. **if** $\Delta t_i, t_{i+1} > /\Delta t_{last}, t_{curr}$ **then**

   > **let** $p_{temp} = p_{curr}$;
   >
   > /* if the temporal interval between current and last position is smaller than the interval required for sampling, the current position is stored in a temporary variable*/

   else
   > **let** $T[i] = p_{temp}$;
   > **let** $p_{temp} = p_{curr}$;
   > /* else assign the last point, which fulfils this condition, to be the next point of the trajectory */

10. **go to** 4

We want to illustrate the functioning of algorithm 1 with an example (see also, figure 3). Let us assume to record the trajectory of a floating car in an urban environment. We set *v_max* – the maximum possible speed of the floating car – to be *60 km/h*. This results from a speed limit in the urban area (*50 km/h*) plus a threshold of *10 km/h*. Moreover, we want the spatial interpolation error of the trajectory to be below 20 meters. The minimum possible sampling rate equals *r =1//1s*. The trajectory of the object is empty.

Fig. 3:   Adaptive sampling for trajectories with guaranteed maximum S.I.E.

In scene *t=0* we start recording the movement of the floating car. We measure the car's first position p_0 and store it as the first point of the trajectory. After one second we measure the car's next position (scene *t =1*). We calculate the average speed between p_0 and the current position, which equals 50 km/h. We enter the 50km/h into equation (2), calculate the minimum time interval needed for sampling and map the result to the largest previous. This results in a temporal interval of 4 seconds. We compare this to the interval that has passed since p_0 (1 second) to the current position and find that the former is smaller. However, we may not be sure whether we have to include the current position in the trajectory at a later stage. Therefore, we store it as a temporary position. After two seconds we again measure the position of the car (scene *t=2*). We calculate the average speed from p_0 to the current position. As this speed still equals 50 km/h we again calculate 4 seconds as the minimum time interval needed for sampling. This is still lower than the two seconds that have passed since p_0. Hence we store the current position as the new temporal position. In scene *t=3* we see that the moving object has taken a sharp turn. Therefore our calculated average speed between p_0 and the current position decreases to 30 km/h. Hence, also the minimum time interval needed for sampling drops to 2 seconds. This value exceeds the time that has passed since p_0. Consequently, we add the temporary position stored in scene *t=2* as a new point to the trajectory and store the current position as the new temporary position (scene *t=4*). As the minimum sampling rate is one second and the function requires a minimum sampling rate of at least 2 seconds for the object to be within the proposed S.I.E., we

guarantee that the same temporary position will not be included into the trajectory more than once.

3   Discussion

At first glance, our proposed method for sampling trajectories seems counterintuitive. During those moments where the object moves slower we require a higher sampling rate compared to those situations where the object moves faster. This is however related to the comparison between the maximum possible speed of the object and the average speed determined by the distance and time covered between two consecutive points on its path. At situations where the object is fast it moves much closer to how we predict it as the average speed will approach the maximum speed and thus less variation in routes is possible. Moreover, we may recall the aim of our sampling method: we want to guarantee that the entire trajectory of the moving object does not exceed a specific uncertainty threshold. When storing the position of an object less frequently as it moves slowly, two consecutive positions will remain spatially close together yet their temporal distance may be large. This means that the uncertainty ellipse of the moving object will resemble a circle due to there being a large difference between average and maximum possible speeds. As mentioned before, circles cover a larger area compared to more 'elongated' ellipses. Hence, the uncertainty of the moving object grows when it moves slower.

In other words, when analysing a trajectory consisting of spatially close positions that were attended by the moving object consecutively but after a considerable time, we might not be sure whether the object really moved slowly as implied by the trajectory, or whether the object moved very fast and the positions are close by pure chance. Our approach, guarantees that the uncertainty of the moving object does not exceed a certain level.

However, adaptive sampling is only feasible if we have a realistic understanding of the maximum possible speed of a moving object. If an object is able to exceed its proposed maximum speed, equation (2) will result an imaginary number, which for obvious reasons makes a difficult interval to sample. Moreover, we have to be aware that the resulting trajectory is denser when the moving object moves slower and that the sampling rate is irregular. This might not be desirable for applications that rather concentrate on the movement of the object and aim at leaving out those parts where the object is at rest, and applications that require a uniform sampling rate for the whole trajectory.

4   Conclusion and Outlook

In this paper we show how we can use the concept of error ellipses for sampling trajectories of moving objects and guaranteeing that their interpolation error does not exceed a certain threshold. For reasons of simplicity we do not consider measurement uncertainty (such as GPS error) for our proposed sampling strategy. Therefore, a logical next step of our research is to adapt our approach and also consider the error that originates from inaccurate position fixing. Moreover, we plan to test and compare our algorithm to traditional sampling methods. We conjecture that our algorithm stores less superfluous information than traditional sampling approaches and want to show this in an empirical analysis. Every

interpolation between two positions is a guess. With our approach we want to make this guess more predictable.

Acknowledgement

References

ANDRIENKO, N., ANDRIENKO, G., PELEKIS, N. & SPACCAPIETRA, S. (2008), Basic Concepts of Movement Data. In: GIANNOTTI, F. & PEDRESCHI, D. (Eds.), Mobility, data mining, and privacy: Geographic knowledge discovery. Berlin, Springer, 15-38.

KOBAYASHI, T., MILLER, H. J. & OTHMAN, W. (2011), Analytical methods for error propagation in planar space-time prisms. Journal of Geographical Systems, 13 (4), 327-354.

KUIJPERS, B., MILLER, H. J., NEUTENS, T. & OTHMAN, W. (2010), Anchor uncertainty and space-time prisms on road networks. International Journal of Geographical Information Science, 24 (8), 1223-1248.

MACEDO, J., VANGENOT, C., OTHMAN, W., PELEKIS, N., FRENTZOS, E., KUIJPERS, B., NTOUTSI, S., SPACCAPIETRA, S. & THEODORIDIS, Y. (2008), Trajectory Data Models. In: GIANNOTTI, F. & PEDRESCHI, D. (Eds.), Mobility, data mining, and privacy: Geographic knowledge discovery. Berlin, Springer, 123-150.

PAULY, A. & SCHNEIDER, M. (2008), Vague Spatial Data Types. In: Encyclopaedia of GIS. Springer US, 1213-1217.

PFOSER, D. & JENSEN, C. (1999), Capturing the uncertainty of moving-object representations. In: Advances in Spatial Databases. Springer, 111-131.

SCHNEIDER, M. (1999), Uncertainty Management for Spatial Data in Databases: Fuzzy Spatial Data Types. In: Proceedings of the 6th Int. Symp. on Advances in Spatial Databases, LNCS 1651. Springer, 330-351.