A Precise Proximity-Weight Formulation for Map Matching Algorithms

Ali Oran
Future Urban Mobility IRG
Singapore MIT Alliance for Research Technology (SMART)
Singapore, 138602
Email: aoran@smart.mit.edu

Patrick Jaillet
Department of Electrical Eng. & Computer Science (EECS)
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
Email: jaillet@mit.edu

***Abstract*—With the increased use of satellite-based navigation devices in civilian vehicles, map matching (MM) studies have increased considerably in the past decade. Frequency of the data, and denseness of the underlying road network still dictate the accuracy limits of current MM algorithms. One practical way that can improve the accuracy of most MM approaches is to use more precise weights for the candidate road segments. Because of the geometric nature of the MM problem, proximity-weights have been considered in almost every MM study. However, being formulated through the shortest distance measure, these weights are prone to inaccuracies. We propose a new, more precise, proximity-weight formulation based on a cumulative proximity function which only assumes that the positioning data displays Gaussian distribution errors. Proposed formulations are developed independent of any MM approach, and for this reason they can be used easily under any future MM algorithm.**

I. Introduction

The Map Matching (MM) problem for a vehicle can be defined as the identification of the road segments on which the vehicle is traveling, and also determining the vehicle's location on those links by integrating positioning data of the vehicle with the spatial road network data [1]. It is a problem that needs addressing as a result of the noisy nature of vehicle positioning device measurements coming from a satellite-based navigation device or another sensor, and the inaccuracies in the road map. Currently, it is of primary interest for the intelligent transportation systems (ITS), where determination of the physical location of a vehicle on the underlying road network is essential [2].

Various methods have been proposed to solve the MM problem over the past two decades; and since the beginning of 2000s, following the increased usage of GPS devices on civilian vehicles, interest in MM studies have also increased considerably. Early MM studies were mostly based on the intuitional but highly unreliable "point-to-point", and the slightly more reliable "point-to-curve" and "curve-to-curve" matching methods, all of which are now placed under the "*geometric approaches*" category. In time, these early methods have been mostly replaced by more accurate methods, which constitute "*topological*", "*probabilistic*" and "*advanced*" MM approaches (see survey [1]). Further categorization is also possible depending on how MM methods harness available data, e.g. *offline* vs. *online* methods; or on the data used in MM methods, e.g. methods using solely positioning data vs. methods using other data in addition to positioning data.

Independent of the approach category they belong to, almost all recent MM methods start their analysis by identifying possible road segments that could be the original segment on which the vehicle was traveling [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]. This initial analysis, which can be termed as the "*candidate identification stage*", have become common in recent MM methods following insufficient performances of early methods that had identified the nearest road segments to observed vehicle positions as the original segments. Overall, along with succeeding analyses, this stage improves the accuracy of an MM algorithm by introducing candidate segments for each vehicle position.

Once the candidates are identified, MM methods proceed to identify their relative weights. Generally, a composite weight score is defined as a function of individual weights representing different aspects of the candidates, such as weights based on geometry, or weights based on topology. Because of the geometric nature of the MM problem, geometric weights are still used widely; even though the geometric methods, that had introduced them, have been mostly replaced. The most common geometric weight is *proximity-weight*, which quantifies the spatial closeness between a vehicle position and its candidate segment. This weight is essential in MM algorithms, following the fact that a segment close to a given location point is more likely to be the true segment of the vehicle compared to another segment further away [11], [15]. However, being directly inherited from early geometric methods, in general proximity-weights have been defined through simple proximity measures, the shortest distance between the positioning data and the road segments has been the most widely used measure so far [3], [4], [6], [7], [10], [11], [12], [13], [14]. For clarity, utilization of this measure is illustrated in Fig. 1, where the proximities between point p and segments S_1, and S_2 are defined by shortest distances Δ_1, and Δ_2. This measure is indeed one valid choice for quantifying the proximity relationship between the point, and the segments. Meanwhile, notice that for doing that, it solely relies on the

This research was supported by the National Research Foundation Singapore through the Singapore MIT Alliance for Research and Technology's FM IRG research program.

978-1-4673-6033-3/13/$31.00 ©2013 IEEE

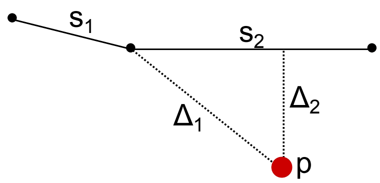

Fig. 1. Shortest distances, Δ_1, and Δ_2, from point p to segments S_1, and S_2, defining the proximity of S_1, and S_2 to p, respectively.

proximity of the segments' closest points to p. With this reliance on a single point, it can not be expected to quantify the proximity of these segments to p completely. Thus, any proximity-weight defined by the shortest distance measure can be prone to inaccuracy, and can undoubtedly lower the accuracy of any MM method using it. Possible issues related to this measure will be discussed in greater detail in the next section.

Unfortunately, inaccurate proximity-weight formulations due to the shortest distance measure have been overlooked in the literature. The primary reason of this neglect could be attributed to the characteristics of the problems that found interest in MM community, and the promising results associated with these. Most studies have analyzed the MM problem under the presence of frequently sampled position data, which makes the uncertainties regarding the travel of a vehicle from one data point to another low (see [6], [16] for location uncertainty arising from low-sampling of moving objects). When uncertainties are low, the influence of inaccurate proximity-weights could be easily lessened when combined with other weights. Consequently, MM algorithms using proximity-weights based on the shortest distance measure would have still resulted in highly accurate mappings, while the proximity-weights were not accurate. Meanwhile, when data frequency is low, matching results would deteriorate, as some of the other weights also become less reliable with increasing uncertainty. Therefore, in the last few years there has been a growing interest in addressing this type of MM problems with various approaches using complex algorithms (e.g. hidden Markov models of [11], [12], [13], [14]). When analyzing low-sampled position data, precise proximity-weights could be one easy to be used, yet substantially beneficial addition for improving the performances of MM algorithms.

Note that, precise proximity-weights would not just benefit the special problem above, and it could be used to tackle MM problems in general to achieve highly accurate matchings. For instance, the denseness of the underlying road network (or the closeness of road segments) is another factor that affects the accuracy of MM methods. Most methods perform well when the underlying road network is sparse, or when the matching is done for vehicles moving along major roads. However, accuracy could deteriorate when matching is done for vehicles moving on a dense road network, e.g. city centers where shorter and closely spaced roads are present. The major reason for this deterioration is the increased number of candidate road segments, which makes it more difficult to identify the true road segment among others. Under these circumstances, precise proximity-weights could distinguish more likely candidates from unlikely ones with better accuracy, thus improving the MM performance. Deterioration of accuracy is also an important consideration when MM is done in an online way, since initial estimation errors might grow quickly in succeeding iterations. Precise proximity-weights can benefit the overall performance of online methods by reducing estimation error at each iteration. Finally, precise proximity-weights would benefit MM problems the most when positioning data is the sole data available. When MM algorithms cannot harness other data, the positioning data should be used with precise formulations to yield the best possible estimates.

In this paper, we propose a new, more precise, proximity-weight for the candidate segments, which is developed independent of any underlying MM approach. It is based on a cumulative proximity function, as opposed to the shortest distance based weights, which are dependent on a single point. The only assumption we make when developing this weight is that the positioning data displays Gaussian distribution errors. The final weight formula is in a compact form, and can yield segment weights without demanding calculations. We believe, this precise proximity-weight could be a practical, yet substantially promising improvement that could be utilized in various MM algorithms.

The rest of the paper is organized as follows. First, in II-A, we will discuss the shortest distance based proximity-weights, and will point out the sources of inaccuracies in them. Later in II-B, considering these inaccuracies, we will develop a more precise proximity-weight formulation. In III, we will compare the results of proposed formulation with the shortest distance based weight formulation.

II. Improving Proximity-Weights for Map Matching Algorithms

A. Current Candidate Measures in MM Studies

Almost all recent MM works start their analysis with the candidate identification stage, identifying possible road segments that could be the original segment on which the vehicle was traveling. The most common way of doing this is considering a geometric proximity criteria to given position points, and identifying the segments satisfying this criteria as the candidate segments. Generally, this criteria is defined through an error region, either ellipsoidal, or circular, and its geometric parameters are chosen depending on the accuracy of the positioning device. Segments falling under this error region are then identified as candidate road segments for that point. We introduce Fig. 2 as a motivational example to illustrate this stage, and also to aid our development of ideas. In this figure, point p_1 represents the GPS point of a vehicle, and solid lines $S_1, S_2, ..., S_5$ are road segments in the vicinity of p_1. The green error circle is drawn around p_1 to identify the candidate links associated with it. Segments S_1, S_3, S_4 fall under this region partially, while segment S_2 falls completely; and they all belong to the set of candidate segments for p_1, which we

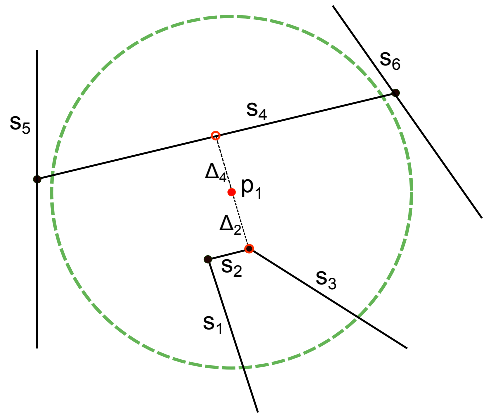

Fig. 2. Circular error region defined around GPS point, p_1. Segments S_1, S_2, S_3, S_4 are the candidate road segments for p_1. Δ_2, and Δ_4 are the shortest distances from p_1 to segments S_2, and S_4, respectively. $\Delta_2 \approx \Delta_4$.

denote by $\mathcal{C}(p_1)$. Other nearby segments S_5 and S_6, being out of the circle, do not belong this set. In this example, we also assumed that p_1 was observed almost equidistant away from segments S_2 and S_4, i.e. $\Delta_2 \approx \Delta_4$.

After identifying candidate segments, MM methods determine proximity-weights of these segments, which are assessed by the closeness of segments to the GPS point. As could be expected, every MM study has proposed its own proximity-weight, through some function of distance. In deterministic MM studies, e.g. [4], [10], monotonically decreasing functions of distance are the most common. Similarly, in probabilistic methods, e.g. [11], likelihood functions of distance defined through the Gaussian distribution is common. Meanwhile, all these works, deterministic and probabilistic, define proximity-weight functions with the shortest distance from the GPS point to the segment as the argument. For a general mathematical discussion, let $d(p, s)$ define the distance between points p and s, $d_m(p, S)$ the shortest distance between point p and road segment S, and $\mathcal{W}_m(p, S)$ the weight of S with respect to p defined through the shortest distance measure. Also, let $f : \mathbb{R}^+ \to \mathbb{R}^+$, be the function that defines the proximity-weight of segments in MM methods. Then, in general form, weights of candidate segments with the shortest distance measure can be formulated as,

$$\mathcal{W}_m(p, S) \doteq f(d_m(p, S)), \quad d_m(p, S) \doteq \min_{s \in S}(d(p, s)). \quad (1)$$

In (1), independent of the function f chosen in an MM method, shortest distance measure is the sole factor that defines the proximity-weights. Meanwhile, as we discussed in the previous section, the shortest distance measure may not fully characterize the proximity relationship between a point and a segment. For this reason, (1) will be liable for yielding incorrect weights. Notice this situation in Fig. 2, where independent of chosen f, weighting the candidate segments based on the shortest distance measure would yield segments S_2 and S_4 be given almost the same weights following the fact that p_1 was observed almost equidistant away from both segments. Meanwhile, notice that segment S_4 has a considerably longer portion remaining inside the error region compared to S_2. Under this situation, assigning the same weights to S_2 and S_4 would be arguable. One would rather expect S_4 to get a higher weight, as it has a longer portion remaining inside the error region while being almost the same distance away from p_1 as S_2.

After noticing that similar weights could be given for a very short segment and a long one, one should question the validity of defining proximity-weights by the shortest distance measure. This approach rules out one essential information, the length of segments remaining inside the error region. In a simple way, one might think about introducing length information to weight calculations by considering lengths as factors to the proximity-weights. However, this approach would also yield questionable results, as it would favor longer segments to shorter ones, and thus could distort the proximity analysis. Notice this again in Fig. 2, where with this proposed approach S_2 and S_4 would get their correct weights, but S_1 and S_3 could be given higher weights than S_2, even though most points on these two segments are father away to p_1 than the points on S_2. Considering the fact that, nearby segments are more likely to be the real location of a GPS point rather than the farther away segments, using segment lengths as factors to improve the accuracy of the shortest distance based weights would not be sufficient. For accurately determining the proximity-weight of a candidate segment, its proximity to the GPS point should be analyzed through the whole segment, rather than through a single point as done in the shortest distance based analysis.

B. Improving Candidate Weights Through New Measures:

One way of considering the whole segment proximity is to introduce a point-wise proximity-weight on the segment, and then to sum the weights of points along the portion of this segment that stays inside the error region. This way, one could avoid the inaccuracies associated with the shortest distance measure, and also would not omit length considerations from the analysis. A subsequent normalization of segment weights could also be done to add consistency to the final results. In some of the recent MM works, similar proposals to shift from the shortest distance measure has been noted [17], but to the best of our knowledge a total compact formula for proximity-weights has not been proposed in the literature yet. From a probabilistic perspective, this proposed approach would be equivalent to the consideration of the cumulative distribution function (CDF) over an entire segment rather than the consideration of the probability density function (PDF) at a segment's closest point to the GPS point (as done in [11], [12], [13], [14]), for characterizing the entire segment. Consistent with our previously defined notations, for a GPS point p, and its candidate segment $S \in \mathcal{C}(p)$, consider the distance from a point s on S to p, $d(p, s)$, and the proximity-weight of this particular point defined by some function f, as $f(d(p, s))$. The overall weight of candidate segment S with respect to p could be formulated as,

$$\mathcal{W}(p,S) \doteq \int_{\bar{S}} f(d(p,s))\, dl, \tag{2}$$

where the integral is the line integral along line segment $\bar{S}$, the portion of S that remains inside the error region. Once the weights of all candidate segments of a GPS point, p, are calculated, normalization could be done to obtain relative weights,

$$w(p,S) \doteq \frac{\mathcal{W}(p,S)}{\sum_{S\in\mathcal{C}(p)} \mathcal{W}(p,S)}. \tag{3}$$

Note that, (2) is formulated with a generic distance-based weighting function f, and thus it would be still valid under various MM approaches, either deterministic or probabilistic. Considering the uncertain nature of the MM problem, we favor progressing with a probabilistic approach, in where we will formulate (2) with a likelihood function of distance. Following the normalization in (3), proximity-based likelihoods can be utilized as proximity-weights. For this reason, the upcoming weights we introduce in this paper, through the likelihood function, can also be used in deterministic algorithms as well. Let $pr(p\,|\,d(p,s))$ denote the likelihood of the vehicle being away from point p by distance $d(p,s)$. Following (2), the overall likelihood of being on S when observing the vehicle at p, $Pr(p\,|\,S)$, would be,

$$\mathcal{W}(p,S) = Pr(p\,|\,S) = \int_{\bar{S}} pr(p\,|\,d(p,s))\, dl \tag{4}$$

where again the integral is the line integral along the line segment $\bar{S}$. The choice for the likelihood function, $pr(p\,|\,d(p,s))$, has not been standard either. So far, Gaussian distribution has been one of the most used distributions in MM works describing the spread of the GPS data [6], [8], [9], [11], [12], [13], [14]. Also, since the year 2000, after the removal of Selective Avability (SA), the GPS data has shown more clear pattern of a Gaussian distribution [15]. Considering this pattern of GPS data, we have also chosen to develop proximity-weights based on a Gaussian distribution, centered around the GPS point with zero mean. With this assumption, we have,

$$pr(p\,|\,d(p,s)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(d(p,s))^2/2\sigma^2} \tag{5}$$

where σ is the assumed standard deviation of the GPS measurements. Then, the correct proximity-weight of a candidate segment could be found by (4) and (5). For better clarity, let's see how this approach works practically for segment S_4 in our motivational example in Fig. 3, an update of Fig. 2. To evaluate the line integral of (2) for segment S_4, we first parameterize the portion of S_4 remaining inside the error region, $\bar{S_4}$. We defined the end points of $\bar{S_4}$ as $S_4^A = [x_A,\ y_A]^T$, and $S_4^B = [x_B,\ y_B]^T$. Then,

$$\bar{S_4}(t) \doteq S_4^A + (S_4^B - S_4^A)t = \begin{bmatrix} x_A + (x_B - x_A)\,t \\ y_A + (y_B - y_A)\,t \end{bmatrix},\ t\in[0,1]. \tag{6}$$

Note that, independent of the underlying road network, the parametrization of road segments will be standard following

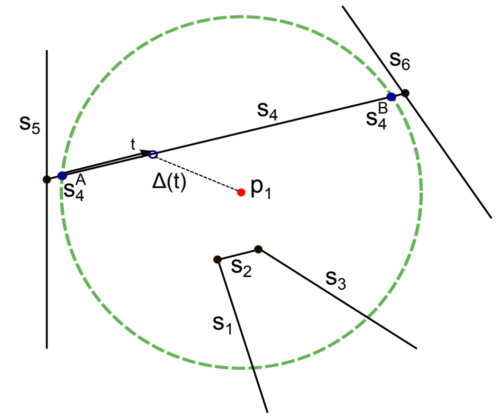

Fig. 3. S_4^A and S_4^B are the end points of the portion of S_4 that falls inside the error region, which we name as $\bar{S_4}$. This line segment is parametrized with parameter t, and the distance from a point on $\bar{S_4}$ to p_1 is formulated by $\Delta(t)$.

the fact that in spatial maps road segments are defined as straight lines connecting nodes or shape points. For this reason the above formula holds in general for any candidate segment whose end points inside the error region are defined by coordinates $[x_A,\ y_A]^T$, and $[x_B,\ y_B]^T$. We use $\Delta(t)$ to denote the parametrized distance from a candidate segment $\bar{S}$ to GPS point $p_1 = [x_1,\ y_1]^T$, and it can be formulated as,

$$\Delta(t) \doteq \sqrt{[(x_A - x_1) + (x_B - x_A)\,t]^2 + [(y_A - y_1) + (y_B - y_A)\,t]^2} \tag{7}$$

With the Gaussian-based likelihood function in (5), (4) can be written in the parametrized form,

$$\mathcal{W}(p,S) = Pr(p\,|\,S) = \int_0^1 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\Delta(t))^2/2\sigma^2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt. \tag{8}$$

where, $x(t)$, and $y(t)$ define parametrized segment's coordinates. Notice from (6) that,

$$dx/dt = (x_B - x_A), \quad \text{and} \quad dy/dt = (y_B - y_A), \tag{9}$$

which yields,

$$Pr(p\,|\,S) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^1 e^{-(\Delta(t))^2/2\sigma^2} \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}\, dt. \tag{10}$$

With expanding the squared terms above, and regrouping them, one can write (8) in the following compact form,

$$Pr(p\,|\,S) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^1 e^{-(at^2+bt+c)/2\sigma^2}\ \sqrt{a}\ dt \tag{11}$$

where,

$$
\begin{aligned}
a &= (x_B - x_A)^2 + (y_B - y_A)^2, \\
b &= 2\left[(x_A - x_1)(x_B - x_A) + (y_A - y_1)(y_B - y_A)\right], \quad (12) \\
c &= (x_A - x_1)^2 + (y_A - y_1)^2.
\end{aligned}
$$

To evaluate the the Gaussian integral in (11), the polynomial exponent is first transformed into the following sum of a squared term and a constant:

$$(at^2 + bt + c) = a(t + b/2a)^2 - b^2/4a + c.$$

Then,

$$\int_0^1 e^{-(at^2+bt+c)/2\sigma^2} dt = e^{(b^2/4a-c)/2\sigma^2} \int_0^1 e^{-a(t+b/2a)^2/2\sigma^2}\, dt. \quad (13)$$

Now, with another transformation, $x = \sqrt{a}(t + b/2a)/\sigma$, the integral in (13) can be simplified to,

$$\int_0^1 e^{-a(t+b/2a)^2/2\sigma^2} dt = \int_{\frac{b}{2\sigma\sqrt{a}}}^{\frac{2a+b}{2\sigma\sqrt{a}}} \frac{\sigma}{\sqrt{a}} e^{-x^2/2}\, dx. \quad (14)$$

Combining (11), (13), and (14), one arrives at,

$$Pr(p_1 \,|\, S) = e^{(b^2/4a-c)/2\sigma^2} \int_{\frac{b}{2\sigma\sqrt{a}}}^{\frac{2a+b}{2\sigma\sqrt{a}}} \frac{1}{\sqrt{2\,\pi}} e^{-x^2/2}\, dx$$

which leads us to the final compact form of the proximity-weight of a candidate segment S with respect to point p:

$$\mathcal{W}(p, S) = e^{(b^2/4a-c)/2\sigma^2} \left[\Phi\left(\frac{2a+b}{2\sigma\sqrt{a}}\right) - \Phi\left(\frac{b}{2\sigma\sqrt{a}}\right)\right] \quad (15)$$

where Φ is the standard cumulative distribution function for the Gaussian distribution with mean as zero and standard deviation as one.

(15) provides MM developers a precise, closed form, proximity-weight formulation, under the sole assumption that GPS data displays a Gaussian distribution. Once a candidate segment's end points inside the error region are found, i.e. $[x_A,\ y_A]^T$, and $[x_B,\ y_B]^T$, then its proximity-weight could be calculated directly with (12), and (15).

C. Defining Parameters

Two important factors, the geometric parameters of the error region, and the standard deviation σ has not been discussed in detail up to this point yet. Notice that, in candidate preparation stage one implicitly assumes that the segments remaining inside the error region are the only possible segments corresponding to the given GPS point. Consequently, one should choose the error region parameters to make sure that the likelihood of being on a segment out of the error region is negligible. One way of satisfying this constraint is to choose the parameters of the error region according to σ, and so that a high confidence level is attained. Geospatial positioning statistics given in US Army Corps Engineers handbook [18] provides a good outline of the relationship between σ, and the expected confidence levels. A trade off will most likely be made about the desired confidence interval, and computational speed, as high confidence intervals comes with larger error regions that contain more candidates compared to lower confidence intervals. There have been different approaches for determining the standard deviation in MM studies. Some studies calculate standard deviation through an analysis of actual GPS data, while others choose it through assumptions. In this study, the guidelines in [18] were followed for determining the geometric parameters of the error region, and also σ.

III. Comparison of Weight Measures

Remember that, we have developed the precise proximity-weight formulation with the goal that it could be later used by any MM algorithm. For this reason, in order to have a clear and algorithm-independent comparison, we will compare the proposed weight measure of (15) with the classical shortest distance-based weight of (1), for calculating the proximity-weights of the candidate segments in our motivational example Fig. 2. To have comparable results, we will use the Gaussian function of (5), also as the weighting function f in (1), i.e.,

$$\mathcal{W}_m(p, S) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(d_m(p,S))^2/2\sigma^2}. \quad (16)$$

The road network in Fig. 2 was defined with p_1 at the origin, and the following nodes are the end points of road segments.

$$
\begin{aligned}
&p_1 = [0, 0]^T, \\
&n_1 = [-34.481, 0.095]^T,\ n_4 = [28.734, 19.060]^T, \\
&n_2 = [-5.336, -12.041]^T,\ n_5 = [2.873, -9.578]^T, \\
&n_3 = [28.854, -24.578]^T,\ n_6 = [2.079, -34.866]^T.
\end{aligned}
$$

As mentioned, we considered guidelines of [18] to determine the error circle radius, r_c, and the standard deviation, σ. To make sure that the segments remaining under the error circle represent all possible candidate segments with a very high probability, we have chosen to utilize a 99.9% probability criteria. To achieve this, a circular error radius of 30 meters, which is more than 3 times the deviation root mean square error (3DRMS), was chosen. After choosing the circular radius as 30m, in order to preserve the 3DRMS measure, standard deviation was chosen as $7m \approx 30m/4.24$, following the relation suggested in [18].

We first look into weights based on the shortest distance formula, (16). Even before calculation, one would expect S_1, S_2, and S_4 to have almost the same weights, and S_1 to have a slightly lower weight, as it's shortest distance is larger than others. Using (16), the likelihoods can be found as, $Pr(p_1|\, S_2) = Pr(p_1|\, S_3) = Pr(p_1|\, S_4) = 7.9772 * 10^{-4}$, and $Pr(p_1|\, S_1) = 7.9761 * 10^{-4}$. When normalized,

$$w(p_1, S_2) = w(p_1, S_3) = w(p_1, S_4) \approx w(p_1, S_1) \approx 1/4.$$

Using (15), we expect to improve these uniform weights. As a first step, the intersection points between the error circle and the segments, $S_1^A, S_3^A, S_4^A, S_5^5$, are found by standard algebraic methods. Then by (12) and, (15) one would arrive at likelihood values of, $Pr(p_1|\,S_1) = 0.0643, Pr(p_1|\,S_2) = 0.025, Pr(p_1|\,S_3) = 0.1023$, and $Pr(p_1|\,S_4) = 0.1572$. When normalized,

$$w(p_1, S_1) = 0.1842, \quad w(p_1, S_2) = 0.0717,$$
$$w(p_1, S_3) = 0.2934, \quad w(p_1, S_4) = 0.4507.$$

These results show that the proposed formulation also considers the segment lengths besides proximity, with the longest segment S_4 getting the highest weight, and the shortest segment S_2 the lowest weight, as opposed to the almost uniform weights of the other approach. On the other hand, while the length of segments were taken into consideration, these lengths did not distort the proximity analysis as opposed to what would have happened if the shortest distance based weights were simply combined with the lengths of segments.

We also want to use a practical mathematical argument to show that (15) yields more consistent weights compared to (16). For this argument, we analyze the weight results given a simpler underlying distribution than the Gaussian, such as the uniform distribution. If a vehicle's GPS data was to display a uniform distribution, then one would merely expect the relative proximity weights of segments to be the same as the relative lengths of segments remaining inside the error circle. With this idea, we observed the normalized weight results when we increased the standard deviation of the Gaussian from its initial value of 7 to a large number of 1000. Note that, for large standard deviation values the Gaussian PDF becomes very flat almost equal to the PDF of a uniform distribution, at least around the mean. This gives us the opportunity to test consistency of (15), without changing the underlying assumption of the Gaussian distribution. For $\sigma = 1000$, the weights are observed to be,

$$w(p_1, S_1) = 0.1782, \quad w(p_1, S_2) = 0.0809,$$
$$w(p_1, S_3) = 0.207, \quad w(p_1, S_4) = 0.5339,$$

which are equal to the normalized lengths of the segments remaining inside the error circle. The original lengths are,

$$|\bar{S_1}| = 18.879\,m, \quad |\bar{S_2}| = 8.571\,m$$
$$|\bar{S_3}| = 21.928\,m, \quad |\bar{S_4}| = 56.569\,m.$$

The matching results of normalized lengths and the normalized weights under very large standard deviation values show that the lengths of candidate segments have indeed become the sole factor, overruling proximity affects. This result in return shows that the new proposed weight, (15), is consistent. On the contrary, under this large value of σ, the shortest distance based weights would still yield normalized weights of approximately $1/4$ for all candidate segments.

IV. Conclusion

Considering the arguable results of the proximity-weight formulations that were based on the shortest distance measure, we propose a new, more precise, alternative weight formulation based on the sole assumption that GPS data displays Gaussian distribution errors. This weight formulation defines a candidate segment's weight through the line integral of a point-wise weight function over the segment. The final form is free of the integral, and can determine proximity-weights with simple calculations. It was developed independent of any MM approach, and for this reason could be used easily with any future MM algorithm.

References

[1] M. A. Quddus, W. Y. Ochieng, and R. B. Noland, "Current map-matching algorithms for transport applications: State-of-the art and future research directions," *Transportation Research Part C: Emerging Technologies*, vol. 15, no. 5, pp. 312 – 328, 2007.

[2] G. Taylor and G. Blewitt, *Intelligent positioning: GIS-GPS unification*. John Wiley, 2006.

[3] W. Ochieng, M. Quddus, and R. Noland, "Map-matching in complex urban road networks," *Brazilian Journal of Cartography (Revista Brasileira de Cartografia)*, vol. 55, no. 2, pp. 1–18, 2003.

[4] M. Quddus, W. Ochieng, L. Zhao, and R. Noland, "A general map matching algorithm for transport telematics applications," *GPS Solutions*, vol. 7, pp. 157–167, 2003.

[5] J. S. Greenfeld, "The map matching algorithm of gps data with relatively long polling time intervals," in *Proceedings of the 81st annual meeting of the Transportation Research Board (TRB 2002)*, January 2002.

[6] S. Brakatsoulas, D. Pfoser, R. Salas, and C. Wenk, "On map-matching vehicle tracking data," in *31st international conference on Very large data*, ser. VLDB '05. VLDB Endowment, 2005, pp. 853–864.

[7] F. Marchal, J. Hackney, and K. Axhausen, "Efficient map matching of large global positioning system data sets: Tests on speed-monitoring experiment in zürich," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1935, pp. 93–100, 2005.

[8] M. E. E. Najjar and P. Bonnifait, "A road-matching method for precise vehicle localization using belief theory and kalman filtering," *Autonomous Robots*, vol. 19, pp. 173–191, 2005.

[9] C. Smaili, M. E. El Najjar, and F. Charpillet, "A road matching method for precise vehicle localization using hybrid bayesian network," *Journal of Intelligent Transportation Systems*, vol. 12, no. 4, pp. 176–188, 2008.

[10] N. R. Velaga, M. A. Quddus, and A. L. Bristow, "Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 6, pp. 672 – 683, 2009.

[11] P. Newson and J. Krumm, "Hidden markov map matching through noise and sparseness," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '09. ACM, 2009, pp. 336–343.

[12] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang, "Map-matching for low-sampling-rate gps trajectories," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2009, pp. 352–361.

[13] J. Yuan, Y. Zheng, C. Zhang, X. Xie, and G.-Z. Sun, "An interactive-voting based map matching algorithm," in *Proceedings of the 2010 Eleventh International Conference on Mobile Data Management*, ser. MDM '10. IEEE CS, 2010, pp. 43–52.

[14] C. Goh, J. Dauwels, N. Mitrovic, M. T. Asif, A. Oran, and P. Jaillet, "Online map-matching based on hidden markov model for real-time traffic sensing applications," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, sept. 2012, pp. 776 –781.

[15] F. Diggelen, "System design & test-gnss accuracy-lies, damn lies, and statistics," *GPS World*, vol. 18, no. 1, pp. 26–33, 2007.

[16] D. Pfoser and C. S. Jensen, "Capturing the uncertainty of moving-object representations," in *Proceedings of the 6th International Symposium on Advances in Spatial Databases*. Springer-Verlag, 1999, pp. 111–132.

[17] M. Bierlaire, J. Chen, and J. Newman, "A probabilistic map matching method for smartphone gps data," *Transportation Research Part C: Emerging Technologies*, vol. 26, no. 0, pp. 78 – 98, 2013.

[18] *NAVSTAR Global Positioning System Surveying*. US Army Corps of Engineers, February 2011, no. 1110-1-1003, ch. 4, pp. 4–1 – 4–18.