Faculty of Engineering

Department of Computer Science

Trajectory Features

PhD student and author:

Avgoustinos Vouros

Supervisor: Prof. Eleni Vasilaki

Contents

1 Key concepts

1.1 Minimum enclosing ellipsoid

Given a set of points in the plane the minimum enclosing ellipsoid is defined as the unique closed ellipse of smallest volume which enclose these points [1, 2].

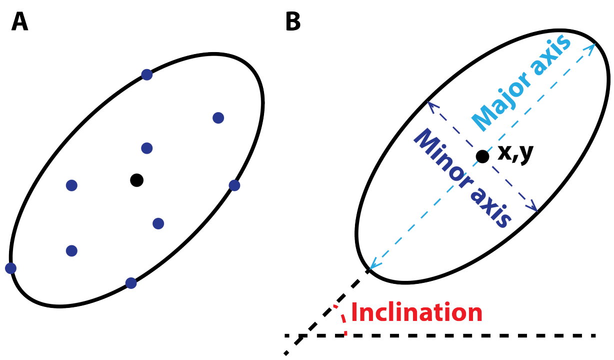

Figure 1: *Minimum enclosing ellipse.* **A.** The minimum enclosing ellipsoid over a set of points; blue dots are the points and the black dot corresponds to the center of the ellipse. **B.** Ellipsoid enclosing metrics: center of the ellipse (x,y); major and minor axes of the ellipse; inclination of the ellipse.

1.2 Circumcircle

Let a triangle formed by using any three points as vertices; the circumcircle is then defined as the unique circle that passes through all the vertices of the triangle [3].

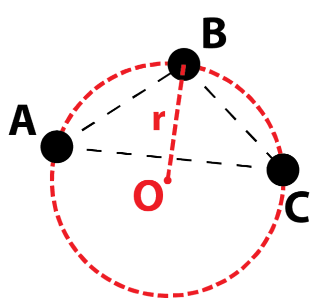

Figure 2: *Circumcircle.* Let A, B and C be three points in space which form the triangle $\triangle ABC$ if are used as vertices. The unique circle O with radius r which passes through the vertices A, B and C is then called circumcircle.

1.3 Statistics

A number of path features were the result of measurements which were performed point-by-point across all the successive coordinates of a path. To assign a single value for these features the following statistics were considered over the measurements:

- **Median:** median was selected to measure the central measurement of the measured values over the mean because of its robustness to outliers.
- **Inter-quartile range (IQR):** IQR is a robust measurement of the spread or dispersion of the measured values similar to the standard deviation. It is equal to the first quartile subtracted from the third quantile, $IQR = Q_3 - Q_1$
- **Coefficient of variation (CV):** CV is another measurement of variability and is defined as the ratio of the standard deviation to the mean ($CV = std/mean$). It is particularly useful when we want to compare the spread from two different populations. However, here we define the CV as the ratio of the IQR to the median, $CV' = IQR/median$. The logic behind this choice is that IQR and median are robust statistics while standard deviation and mean are not.

2 Geometric features

2.1 Path eccentricity

The eccentricity measures how elongated are the paths. It is defined as $\epsilon \equiv \sqrt{1 - \frac{b^2}{a^2}}$, where a and b are the semi-major and semi-minor axis of the enclosing ellipse of the path [4].

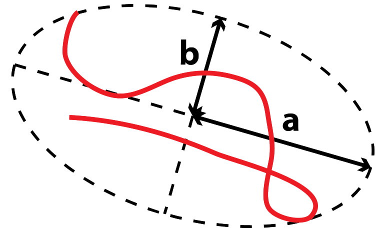

Figure 3: *Path eccentricity.* Let a path consisted of a set of points being enclosed inside the minimum ellipsoid with semi-major axis a and semi-minor axis b. The eccentricity is then defined as $\epsilon \equiv \sqrt{1 - \frac{b^2}{a^2}}$.

2.2 Path focus

The focus is a measurement of how concentrated a path is on specific locations. It is defined as $f \equiv 1 - \frac{4A}{\pi d^2}$, where A is the area of the minimum enclosing ellipse if the path and d is the total length of the path. With this definition a focus of 0 means that the path is perfectly circular; larger focus values give an indication of increasingly closed paths [4].

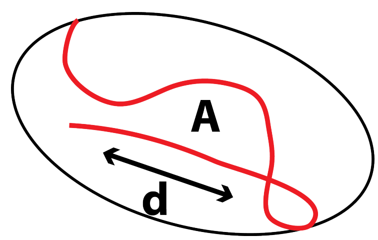

Figure 4: *Path focus.* The path is enclosed inside the minimum volume ellipsoid with semi-major axis a and semi-minor axis b. The focus is then defined as $f \equiv 1 - \frac{4A}{\pi d^2}$, where d is the total length of the path and A is the area of the ellipse.

2.3 Path density

A method to determine if a path favors exploration or not, is to quantify the circularity of the enclosing ellipse. A more circular ellipse, i.e. ellipse where the difference between the semi-major and the semi-minor axes is minimized, can be linked to less exploration, thus indicate paths more focused on specific areas. Mathematically we can define the following equation of the density of a path around a certain area: $density \equiv (a-b)/l$, where a and b are the semi-major and semi-minor axes of the path enclosing ellipse. Given this equation, a perfect circle will equal to a value of zero.

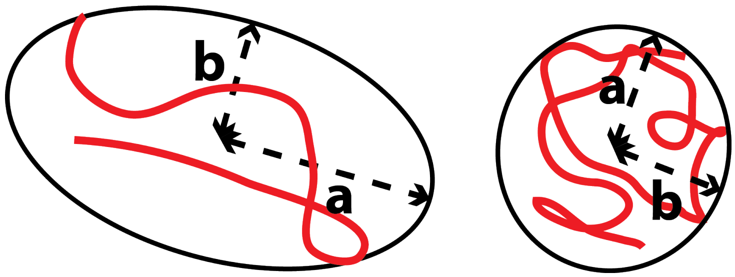

Figure 5: *Path density.* Given two paths and their enclosing ellipses then the more circular the ellipses are the more exploration the paths represent. Since the circularity of an ellipse is proportional to the equality of the semi-major a and semi-minor a axis we can measure how focused is a path on a specific area by calculating the difference between a and b. For a perfect circle this difference will be equal to 0.

2.4 Path inner radius variation

The inner radius is defined as $CV \equiv (Q_{ri_3} - Qri_1)/\widetilde{ri_2}$, where $\widetilde{ri_2}$ is the median distance between each point of the path to the center of the minimum enclosing ellipsoid of the path, and $Q_{ri_3} - Qri_1$ is the inter-quartile range (IQR) of this distance. The defined equation is the coefficient of variation which measures the relative dispersion of points relative to a circle. A perfect circle has a coefficient of variation equal to zero. [4].

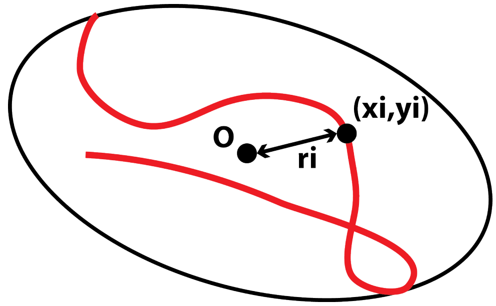

Figure 6: *Path inner radius variation.* The path is enclosed inside the minimum volume ellipsoid with center O. The inner radius variation is then defined as the ration of the inter-quartile rate of the distances between each point to the center of the ellipse to the median of the distances between each point to the center of the ellipse.

2.5 Path longest loop

This feature measures the length of the longest loop, or self-intersecting sub-segment of the path. To compute this value all pairs of lines defined by two consecutive path points are tested for intersection. If no intersection was present a value of zero was assigned to the feature. The path longest self-intersecting is divided by the total length of the path for normalisation purposes [4].

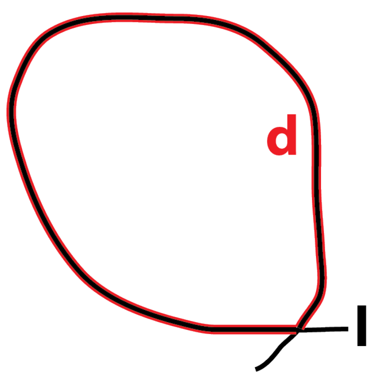

Figure 7: *Path longest loop.* Given that d is the longest loop and l the total length of the segments then this feature is computed as $LL = \frac{d}{l}$

.

2.6 Path sinuosity

Sinuosity is a measurement of the tortuosity of a path. It can be simply defined as the distance between the first and the last point of the path d_{pe} divided by the total length of the path l, $sinu = \frac{d_{pe}}{l}$ [5]. However, such simplistic measurement is not enough for paths that are completely random or exhibit both random and targeted behaviours and a more sophisticated index of sinuosity can take into account the distribution on the turning angles [5]. In the study of [5] there are two such indexes described by the equations 1 and 2

$$S = 2 * \sqrt{p * \frac{1+c}{1-c} + b^2} \qquad (1) \qquad\qquad S = 2 * \sqrt{p * \frac{1+c^2-s^2}{(1-c)^2+s^2} + b^2} \qquad (2)$$

where p is an expected step length and b the coefficient of variation of p and s and c are the mean sine and cosine of the turning angles. Given these formulas equation 2 is more generic while equation 1 assumes a Gaussian distribution, thus $s = 0$. In order to adjust the given equations for path coordinates, it is considered that p is the average of the lengths between the successive points forming the path and b is the ratio of the standard deviation of the lengths and p. Angles where computed based on the absolute angle formula given in 2.8.1.

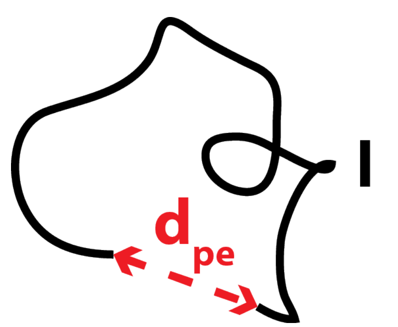

Figure 8: *Path sinuosity.* A simplistic formula of sinuosity where $sinu = \frac{d_{pe}}{l}$ [5].

2.7 Path curvature

The curvature is computed between any three points and it is equal to the radius of their circumcircle, which is also known as curve radius. When we have a sequence of points then a sequence of curve radii can be computed for each segment formed by two successive points. The curve radius of each segment is then equal to the average of the two curve radii which specify the segment. Finally the curvature of a path as a whole is normally computed by adding the length of the segments whose radius is less than a certain threshold [6]. However, as a curvature feature the median and IQR of the curve radius of the segments is considered.

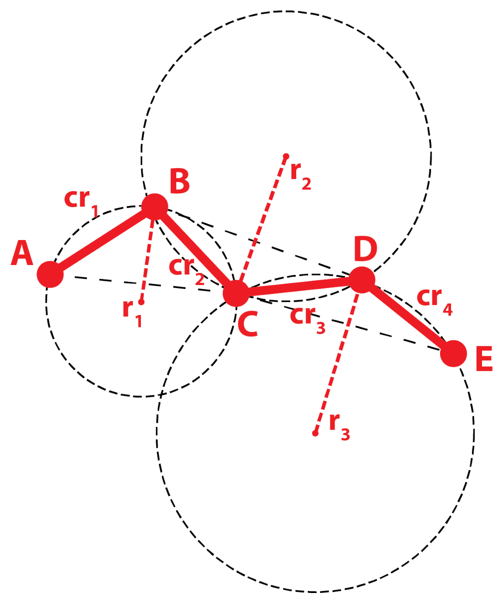

Figure 9: *Path curvature.* Let the points A-E be the coordinates of a path. For each sequence of three points there is a intersecting circle with radius r and a triangle formed by connecting the first to the last of the three points. The segment formed between each two points is then having a curve radius equal to the average radius of the radii for the triangles it is part of. For example The curve radius cr_2 of segment BC is equal to $\frac{r_1+r_2}{2}$ and the curve radius cr_3 of segment CD is equal to $\frac{r_2+r_3}{2}$. In this example the first segment AB and the last segment DE are having curve radii cr_1 and cr_4 which are equal to r_1 and r_3 [6].

2.8 Path angles

Angular aspects of a path can indicate different navigation patters and behaviours. Angles can be defined considering the joint of two path segments (relative angles) or considering the segment between two points and the right horizontal of the distal end (absolute angles). Since for every three (relative) or two (absolute) points we will have a different angular measurement as feature values we can consider the median, $median(angles)$ the inter-quartile range, $IQR(angle)$ and the coefficient of variation defined as $f_{angle} = \frac{IQR(angle)}{median(angles)}$.

2.8.1 Absolute angle

The absolute angle of a path formed between two points is computed using the coordinates of the two points and the inverse tangent function. It is given by the formula $\theta = arctan(\frac{opp}{adj})$, where opp equals the opposite side and adj the adjacent side of the right-angled triangle formed by using these two points as vertices.

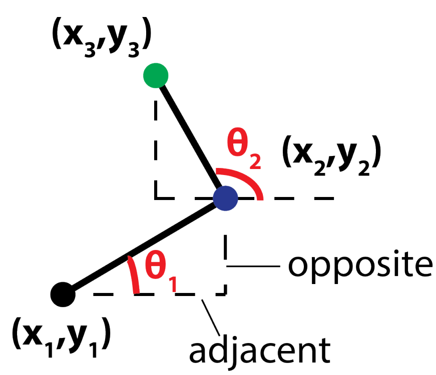

Figure 10: *Absolute angles.* Given a set of successive three points (black, blue, green) the two absolute angles between them can be computed as follows: for the first two points (black, blue) the adjacent will be equal to $adj_1 = x_2 - x_1$ while the opposite will be equal to $opp_1 = y_2 - y_1$; finally the first absolute angle is computed using the formula $\theta_1 = arctan(\frac{opp}{adj})$. The same process is repeated for the next angle (points blue, green).

2.8.2 Relative angle

The relative angle of a path between three points is computed using the law of cosines. Forming the triangle between these three points then the relative angle θ is given by the formula: $c^2 = a^2 + b^2 - 2ab(cos\theta) \Rightarrow \theta = arccos(\frac{a^2+b^2+c^2}{2ab})$, where a, b, c are the sides of the triangle (c is the hypothetical side formed by connecting the first with the last point of the path).

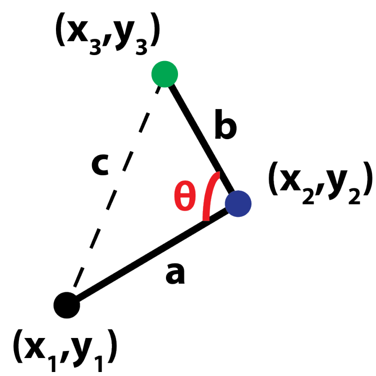

Figure 11: *Relative angles.* Given a set of successive three points (black, blue, green) the absolute angle θ between them can be computed as follows: we form a triangle by connected the first and the last point (side c) and then we compute each side $a = \sqrt{(x_2 - x_1)^2 - (y_2 - y_1)^2}$ and $b = \sqrt{(x_3 - x_2)^2 - (y_3 - y_2)^2}$; finally we solve the equation $c^2 = a^2 + b^2 - 2ab(cos\theta)$ for θ, thus $\theta = arccos(\frac{a^2+b^2+c^2}{2ab})$.

3 Spatial features

3.1 Path distance to center

The distance to the center of the arena is a useful measurement that indicates if the animal spends most of its time next to the walls or to the more central parts of the arena.

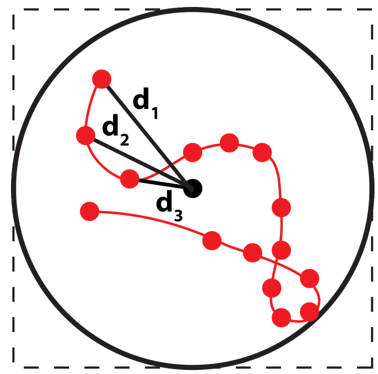

Figure 12: *Path distance to center.* The distance (d_i, where $i = 1, 2, 3...$) of every point forming the path to the center of the arena is computed in order to approximate the amount of time the animal spent next to the walls. This feature can be computed for both circular and square arenas.

3.1.1 Median distance

This value is computed by measuring the distance of each datapoint forming the path to the arena center and dividing it with the arena radius. Afterwards the median of the distances is considered instead of the mean because it is more robust to outliers, $median(d_1, ..., d_n)$.

3.1.2 Inter-quartile range distance

This value is computed by measuring the distance of each datapoint forming the path to the arena center and dividing it with the arena radius. Afterwards the IQR value of the distances is considered instead of the standard deviation because it is more robust to outliers, $IQR(d_1, ..., d_n)$. IQR provide information about the spread of the sample.

3.2 Path angle to center

Using the center of the arena as a reference point the angle between each point and the arena center can be calculated. Finally as features the median and IQR can be considered over all the measurements. Such features can indicate if the animal is focused on a certain quadrant of the arena.

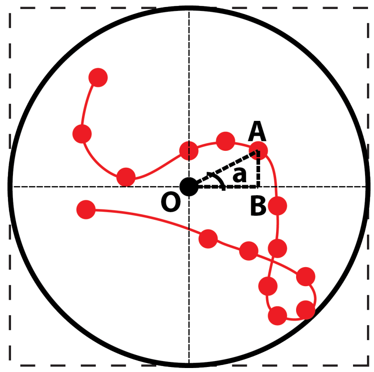

Figure 13: *Path angle to center.* Let the triangle OAB where vertex O is the center of the arena and the vertex A is a point of a path. The angle a can be then computed using the formula $\theta = arctan(\frac{OA}{OB})$. This feature can be computed for both circular and square arenas.

3.3 Central displacement

The central displacement is the Euclidean distance of the centre of the minimum enclosing ellipsoid to the centre of the arena, dividing with the arena radius. It is an important measurement to identify concentric paths with the arena [4].

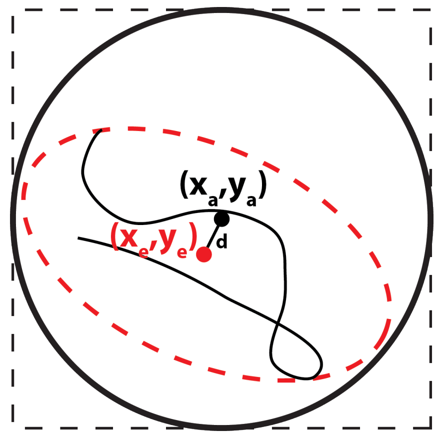

Figure 14: *Central displacement.* It is the distance d form the center of the arena (x_a, y_a) to the center of the minimum enclosing ellipsoid of the path (x_e, y_e). This feature is applicable to both circular and square arenas.

References

[1] B. Gärtner and S. Schönherr, "Exact primitives for smallest enclosing ellipses," *Information Processing Letters*, vol. 68, no. 1, pp. 33–38, 1998.

[2] M. J. Todd and E. A. Yıldırım, "On khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids," *Discrete Applied Mathematics*, vol. 155, no. 13, pp. 1731–1744, 2007.

[3] E. W. Weisstein, "Circumcircle," http://mathworld.wolfram.com/Circumcircle.html.

[4] T. V. Gehring, G. Luksys, C. Sandi, and E. Vasilaki, "Detailed classification of swimming paths in the morris water maze: multiple strategies within one trial," *Scientific reports*, vol. 5, p. 14562, 2015.

[5] S. Benhamou, "How to reliably estimate the tortuosity of an animal's path:: straightness, sinuosity, or fractal dimension?" *Journal of theoretical biology*, vol. 229, no. 2, pp. 209–220, 2004.

[6] A. Franco, "Curvature," https://github.com/adamfranco/curvature/wiki, 2016.