Density calculation for moving points

Stefan Peters[1] and Jukka M. Krisp[1]

[1] Technische Universität München (TUM)
Bauingenieur- und Vermessungswesen, Kartographie (LFK)
Arcisstraße 21, 80333 München, Germany
stefan.peters@bv.tum.de, jukka.krisp@bv.tum.de

ABSTRACT

To detect clusters and hotspots in moving point datasets is a difficult task. It depends on the data analysis and on the visualisation or rather simulation. This paper presents the investigation of the calculation and representation of temporally and spatially highly dynamic point datasets. Within the paper we provide a flexible approach that aims to explore point patterns and densities of points with temporal dimension. Two new approaches are presented, based on kernel density estimation (KDE); both of them consider the direction and the speed of movement for the individual points. One approach is based on a linear directed cone and the other one considers a "trigonometric bell". The core benefit of this approach is that the resulting directed kernel density maps represent the point density at the initial time of the moving points as well as the point density change towards the end time of the recorded point movement in one single density map.

INTRODUCTION

Large amounts of movement data are generated daily for example through technological devices such as mobile phone equipped with a GPS or through analysis of satellite images. The exploration of moving point datasets to identify movement patterns has already become a research focus in GIScience (Dykes and Mountain, 2003). To generate knowledge from patterns and benefit from human cognitive abilities, meaningful visualization of patterns is crucial. Generally clustering and more specifically kernel density estimation (KDE) as a visual clustering method, is a data mining technique that aims at grouping data to patterns based on mutual (dis-)similarity (Assent et al., 2007). Simply plotting data (points) on a suitable background map or image will not necessarily be effective. The goal is to effectively present abstract information that reveals unseen patterns (or movements) that are not directly visible (Andrienko et al., 2008). In our specific case we investigate traditional density maps and develop a density map with dynamic density information where trends and directions of the density change can be interpreted from the visualisation.

BACKGROUND OF POINT DENSITIES

From a theoretical perspective our new approach relates to the provision of "theory, methods and tools for visual exploration, analysis, synthesis, and presentation of geospatial data" as defined as a research field within the context of geovisualization (MacEachren and Kraak, 2001). An important issue in Geovisualization and Visual analytics, published in the latest ICA research agenda, is that Geovisualization techniques have extended the map medium to embrace dynamic, three- and four-

dimensional data representation (Virrantaus et al., 2009). Effective visual exploration may offer opportunities for analysing these data sets in a timely fashion, but current techniques may not be effective when applied to the visual analysis of the kinds of large and complex data sets that are increasingly common. The displays may become illegible due to visual clutter and massive overplotting associated with large numbers of cases, users may have difficulty perceiving, tracking and comprehending numerous visual elements that change simultaneously (Andrienko et al., 2008). With the evolving concept of geovisualization to visual analytics (VisMaster, 2009), the need to analyze and make sense of overwhelming amounts of complex, disparate, conflicting and dynamic data and information is recognized. Core parts of data analysis require human judgment. Concerning data representations and transformations, the question is how to transform data into a representation that is appropriate to the analytical task and effectively conveys the important content?

One way to deal with dynmic point patterns is the calculation and representation of trajectories. Trajectories are defined as the path of a moving object that follows through space and have been investigated in several studies, with a conclusion that trajectories of moving objects are complex spatio-temporal constructs, and analysis of multiple trajectories is a challenging task (Rinzivillo et al., 2008).

Generally point density is a continuous function and in order to present an effective and accurate impression of its distribution, a scheme that recognizes this continuity is needed (Langford and Unwin, 1994). The point density function, known as kernel, is a well-defined smooth, optionally unbounded and three dimensional function. The kernel shape, its bandwidth as well as the output-size of the resulting raster-cells determine this function. The color-intensity of each output-cell depends on the z-value of the overlaying three dimensional kernel function above each sample point.

Kernel density estimation (KDE) may also provide an informative (exploratory) tool for hot-spot and cool-spot identification and analysis (Smith et al., 2008). Smith et al. (2008) point out that the kind of kernel function (e.g. linear, Gaussian function) does not have a big influence on the densities. Rather the bandwidth, also called search radio or spread parameter, tend to have a major impact on the resulting estimated density surface (Krisp et al., 2009).

In traditional way of calculating kernel densities a normally distribution function is used and the points are visited by the following “three-dimensional-moving-function”, based on (Scott, 1992)

$$\hat{f}_h(x) = \frac{1}{N \cdot h} \sum_{i=1}^{N} K(u) \quad \text{with} \quad u = \frac{X - X_i}{h} \quad \text{and} \quad K_G(u) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}u^2\right) \qquad \text{(Eq.1)}$$

with:
- $\hat{f}_h(x)$ = general Kernel density function
- K = Kernel function
- K_G = standard Gaussian function
- h = smoothing parameter (bandwidth)
- X = points (x,y) for which the density will be estimated
- $X_1, X_2, ..., X_N$ = sample points, placed within the kernel radius h

Silverman (1986) worked on the issue of determining the optimal bandwidth for a certain case, before making the actual density computation. Silverman’s optimal bandwidth calculation (equation 2) has two outcomes, the optimal radius for the x-coordinates and the optimal one for the y-coordinates. Depending on which one is smaller, it takes either the variance (var) or the interquartile range (IQR) of the coordinates to the basis.

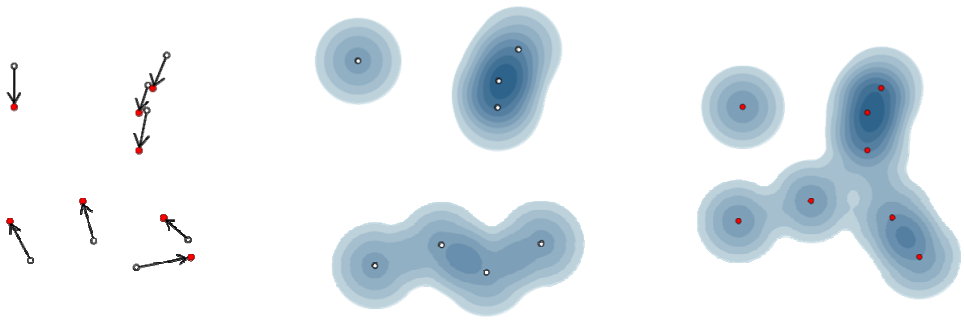
$$bw\_optimal = 1.06 * \min\left(\sqrt{\text{var}(P)}, \frac{\text{IQR}(P)}{1.34}\right) * n^{-\frac{1}{5}} \quad \text{(Eq.2)}$$

with: P = dataset: points coordinates
bw_optimal = optimal bandwidth
IQR (P) = interquartile range (distance between the 1st and 3rd quartile)
var(P) = variance of P
n = number of points (length of P)

In Krisp et al. (2009) a computational tool - using Silverman's formula (equation 2) for calculating an optimal bandwidth – is described. The objective thereby is to visually select the bandwidth parameter using an interactive graphical interface. In this interface the kernel density maps are drawn near real-time while the user can change the bandwidth using a slider tool. This helps to determine the appropriate bandwidth setting and provide a suitable visualization of a particular point dataset.

MOTIVATION

In classical Kernel density estimation the point dataset normally is linked to only one specific time. In case of different time series of the same point dataset different density maps can be produced and compared visually. This visual comparison is difficult because the specific changes of the point densities (directions, distances) are hardly to recognize. The idea of just one resulting density map representing the time and space change of the point densities between two moments of times would improve this comparison.



(a) (b) (c)

Figure 1: Schematic point dataset with blank points at time t1 and filled points at time t2 (a); corresponding kernel density estimations: Density for t1 (b) and Density for t2 (c)

Figure 1 illustrates the initial situation. In figure 1 (a) two time stages, t1 and t2, of a potential point dataset are shown. In figure 1 (b) the corresponding kernel density estimations for time t1 is drawn and in (c) the corresponding density map for time t2. The point dataset is highly dynamic; each point may move to a different direction with a different speed or remain at the same position. Consequently the direction of movement of each point can be assigned as a vector to each point, like shown in figure 1 (a). How can we consider the direction and the speed of movement for each point when calculating a density surface? Therefore the following method will describe our solution.

METHOD FOR DENSITY CALCULATION CONSIDERING DYNAMIC POINTS

Kernel functions used in traditional density estimations are non-directed. For in position and time highly dynamic datasets a new approach is needed which considers the time and position changes of points. A solution is to implicate the movement direction of every point and modify the kernel towards a directed one.

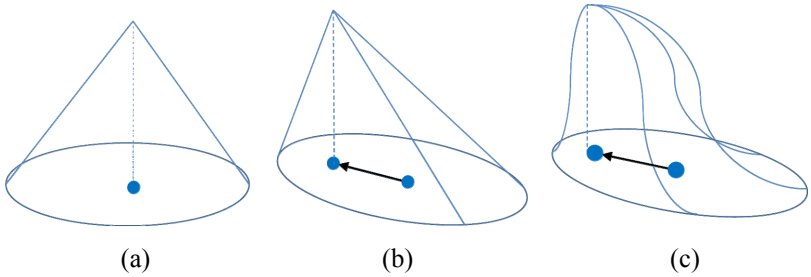


Figure 2: Normal linear Kernel (a); directed kernel considering the movement direction and the speed of each point: linear cone (b) as suggested by Krisp and Peters (Krisp and Peters, 2010) and a *trigonometric bell* (c).

Figure 2 (a) illustrates a normal linear kernel over a sample point. We assign to each point, which changed its position between two different moments of time, a movement vector. Next we apply a directed kernel to the start and to the end point of the vector. So a directed kernel is defined as a non-symmetrical function which is tilted towards the point movement direction. A tilted kernel can be designed for example as a tilted cone or as a trigonometric (or Gaussian-) bell. In Figure 2 the assigned directed kernel is shown. Figure 2(b) displays a linear tilted kernel and figure 2(c) a tilted trigonometric kernel. In this work both forms were used. Krisp and Peters (2010) already introduced the idea of a linear tilted kernel on a conceptual level. For these points without position change in time, there will be no vector and the kernel consequently has a non-directed shape.

Within a directed kernel the shifted is defined by the perpendicular distance from the sample point at time t2 (point X_{i2}) to its apex (figure 3).

For the tilted cone (figure 3a) the calculation of the pixel kernel density values (h_i) relies on the theorem on intersecting lines (equation 2). Therefore the intersection point *S* has to be determined before (equation 4). The point *S* results from the intersection of the line ($X_{i2,}$ *X)* and the circular arc around point X_{i1} with the radio *h* (Bandwidth).

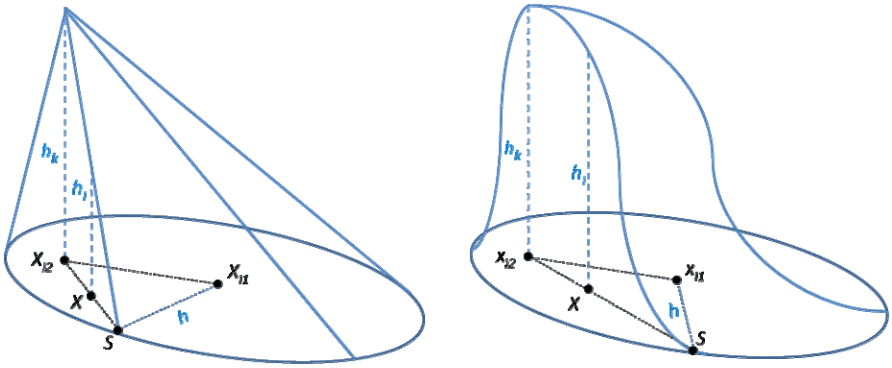


Figure 3: tilted cone (a) and tilted trigonometric bell (b)

Description of kernel elements in figure 3:

X = any point (x,y) inside the bandwidth-circle around X_{i1}
h_k = perpendicular distance from the base to the apex
h_i = perpendicular distance from the base to the surface of a cone / bell over any point X
X_{i1} = sample point in time t1; center-point of the cone ground area / bell ground area
X_{i2} = sample point in time t2; point foot of perpendicular of the tilted cone / tilted bell
S = intersection point of the line (X_{i2}, X) and the circular arc X_{i1} with radio h
h = search radio (bandwidth)

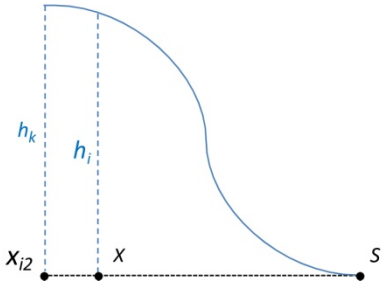


Figure 4: Calculation of h_i recording to Trigonometric function

For the tilted trigonometric bell (figure 3b) the pixel kernel density value (h_i) is calculated with a trigonometric approach: the connection between the apex in direction to the 'to be interpolated' point X to the intersection point S is thereby defined as a cosines function to achieve a "Gaussian-like" bell (figure 4, equation 4).

Tilted cone case (figure 3a): Calculation of h_i recording to the theorem on intersecting lines:

$$h_i = \frac{h_k \cdot \overline{X,S}}{\overline{X,S}} \quad \text{(Eq.3)}$$

Tilted trigonometric bell case (figure 3b): Calculation of h_i recording to trigonometric functions:

$$h_i = \left(1 + \cos\frac{\overline{X,X_{i2}}}{\overline{X_{i2},S}}\right) \cdot h_k \quad \text{(Eq.4)}$$

whereas for both cases the intersection point S is calculated as following (t = azimuth):

$$\begin{aligned} x_s &= x_{i1} + \overline{X_{i1},S} \cdot \sin t_{x_{i2},x} \\ y_s &= y_{i1} + \overline{X_{i1},S} \cdot \cos t_{x_{i2},x} \end{aligned} \quad \text{(Eq.5)}$$

$$\overline{X_{i1},S} = \sqrt{\overline{X_{i2},X_{i1}}^2 - \left(\overline{X_{i2},X_{i1}} \cdot \sin(\left|t_{X_{i2},x} - t_{X_{i2},X_{i1}}\right|)\right)^2} + \sqrt{h^2 - \left(\overline{X_{i2},X_{i1}} \cdot \sin(\left|t_{X_{i2},x} - t_{X_{i2},X_{i1}}\right|)\right)^2}$$

For those points whose position change is bigger than the Kernel bandwidth h > (X_{i1}-X_{i2}), the position change will be set as the bandwidth h. X_{i2} therewith will be the intersection of the change direction and the bandwidth-circle around X_{i1}.

Consequently the vertical intersection of each grid point with the envelope of the tilted cone defines the density of any grid point. In case of several overlapping shifted cones, all resulting density values will be added, like it is done in traditional kernel density estimation.

CASE APPLICATION: DENSITY OF AIRPLANES OVER GERMANY

Test data and results

Our sample datasets have been provided by the Deutsche Flugsicherung GmbH and the EUROCONTROL Information Centre. The dataset contains points representing airplane positions over the area of Germany. As test data we used airplane positions at two different moments of time. Alltogether 405 airplane coordinates were used, 15.05.2009 at 9:26:00 o'clock (t1) and at 9:31:00 o'clock (t2).

In our case we applied the directed linear and the directed trigonometric kernel to the test data of 405 airplanes. The positions at time t1 and time t2 determine the direction in which the kernel is tilted. The distance between these two positions defines the speed of the airplane and thus the skewness of the directed kernel.

Figure 5 demonstrate the directed kernel approach. The "straight" standard kernel, either linear or trigonometric (a,c) is changed while considering two different time stages into either a linear directed kernel (b) respectively into a trigonometric directed kernel (d).

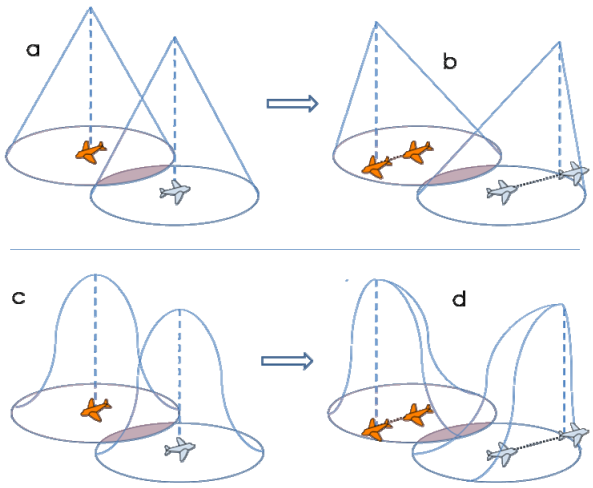


Figure 5: Straight linear cone (a), directed linear cone approach (b), straight gaussian cone (c), directed trigonometric cone approach (d); in each case above the airplane positions

To evaluate the applied directed kernel density methods, we computed first the density map using the traditional kernel density (KDE) approach. Thereby a Gaussian function was used like displayed in figure 5c. The result for the KDE method is shown in figure 6, while a shade illuminated effect was used for the individual kernels.

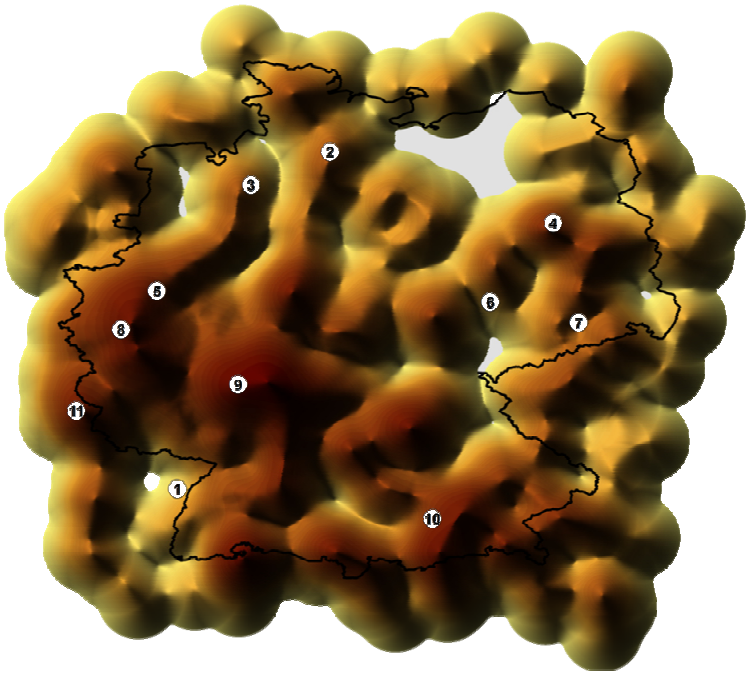


Figure 6. Kernel density estimation (KDE) of 405 airplanes at time t1

The directed linear kernel approach (tilted cone) is applied to the test data of 405 airplanes, which result is illustrated in figure 7. The positions at time t1 and time t2 determine the direction and the dimension of the tilt of the cones. A shade illuminated effect was used.

Furthermore the directed trigonometric kernel approach (tilted bell) is applied to the test data of 405 airplanes, which result is illustrated in figure 8. The positions at time t1 and time t2 determine the direction and the dimension of the tilt of the cone. Again a shade illuminated effect was used. In all three figures 6,7 and 8 the numbers with white circles stand for the biggest airports (1 Straßburg, 2 Hamburg, 3 Bremen, 4 Berlin, 5 Dortmund, 6 Leipzig, 7 Dresden, 8 Köln, 9 Frankfurt a.M., 10 München and 11 Luxemburg).

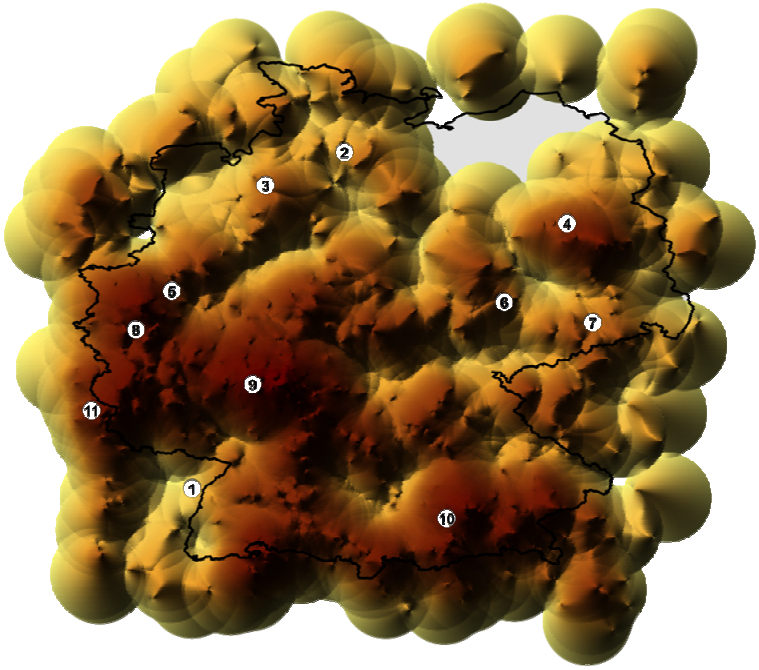


Figure 7. Directed linear kernel density estimation of 405 airplanes; based on Krisp and Peters (2010)

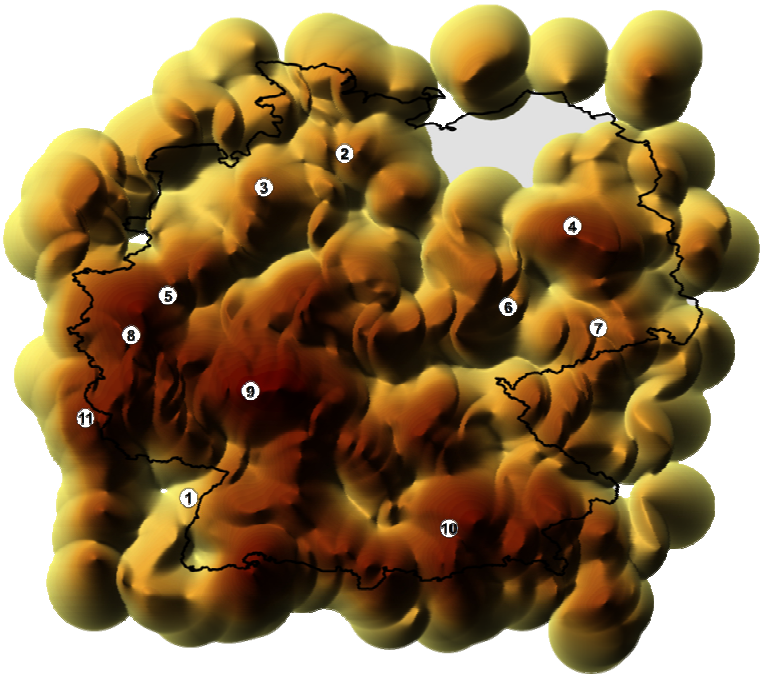


Figure 8. Directed *trigonometric* kernel density estimation of 405 airplanes

Discussion of the results

Let's compare the traditional density map using a non-directed linear kernel like in figure 6 with the directed density map whereby the new directed kernel method was used, either with a directed linear kernel (figure 7) or with a directed trigonometric kernel (figure 8). With the traditional kernel density method two maps have to be produced and compared to analyze the density changes of moving point data. But using the new directed trigonometric kernel method, only one map resulted which reflects the density change in time. Figure 7 and 8 illustrate obviously the density change and in particular the direction in which the density is changing. For areas with sparse airplanes the plane position change can be recognized as followed. The midpoint of the kernel ground area represents the point position at time t1, whereas its position at time t2 can be recognized throughout the apex of the shifted kernel. This can be seen in our data for example in figure 7 upper right corner where point distribution is sparse. The method took the speed of the points into consideration. In the resulting maps (figure 7, 8) the speed is reflected by the moving vector, so by the kernel apex (point position at time t2) and its distance from the ground circle midpoint (point position at time t2). In areas with a denser point distribution in the resulting directed map (centre of figure 7 and 8) "ripple" or "wavelike" effects occur. Thus the density direction change is illustrated.

Summing up the new directed kernel approach results a density map which displays the density of the point data at a certain time as well as the point density change towards a second moment of time. Thereby the direction but also the speed of movement for the individual points is considered. Between directed linear and trigonometric kernel method there are no major differences. For both methods the exposed benefits are valid. The use of the directed trigonometric kernel method results a smoother map (figure 8) while comparing with the map output (figure 7) using the directed linear kernel method.

CONCLUSIONS & FURTHER RESEARCH

During creating density maps of dynamic point data with airplanes, we advise that analysts can identify spatial relationships and spatial point patterns. Also they are able to connect these with relevant background knowledge of processes, e.g. with possible airplane collisions or upcoming shortages of landing spots.

In this work we considered in point density estimation the direction and the speed of movement for the individual points. Therefore two approaches, a linear and a trigonometric directed kernel were developed and applied to airplane data of two different times. The main advantage of this approach is that instead of comparing two density maps, the density change of a point dataset in time can be obtained from one single "directed density map".

Regarding to the proposed approach, the "distortion" of the kernel function may code some information about the direction of movement and speed, but only for very sparse spatial data. Whenever the "cones" overlap, as over Frankfurt, the information about the density changes on the one hand disappear. But on the other hand the density direction change is illustrated throughout "ripple" or "wavelike" effects that might help the analysis of moving clusters.

In further research we will investigate the directed kernel approach using data sets of a different nature, such as GPS data collected in different situations. We also want to investigate dynamic point densities with point datasets whereby each point at a certain moment of time is not related to a specific point after time and position changed. So the number of points for different time series distinguishes. These data could be for instance points of animal swarms or weather data.

BIBLIOGRAPHY

ANDRIENKO, G., ANDRIENKO, N., DYKES, J., FABRIKANT, S. & WACHOWICZ, M. (2008) Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Information Visualization,* 7**,** 173-180.

ASSENT, I., KRIEGER, R., MUELLER, E. & SEIDL, T. (2007) VISA: Visual Subspace Clustering Analysis. *SIGKDD Explorations,* 9 (2).

DYKES, J. A. & MOUNTAIN, D. M. (2003) Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Computational Statistics and Data Analysis, Data Visualization II Special Edition,* 43**,** 581-603.

KRISP, J. & PETERS, S. (2010) Directed Kernel Density Estimation (DKDE). *submitted to Annals of GIS (in review 03.2010)*.

KRISP, J. M., PETERS, S., MURPHY, C. E. & FAN, H. (2009) Visual Bandwidth Selection for Kernel Density Maps. *PFG - Photogrammetrie Fernerkundung Geoinformation,* 5**,** 441–450.

LANGFORD, M. & UNWIN, D. J. (1994) Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal,* 31**,** 21– 26.

MACEACHREN, A. & KRAAK, M.-J. (2001) Research Challenges in Geovisualization. *Cartography and Geoinformation Science,* Volume 28**,** 3-12.

RINZIVILLO, S., PEDRESCHI, D., NANNI, M., GIANNOTTI, F., ANDRIENKO, N. & ANDRIENKO, G. (2008) Visually–driven analysis of movement data by progressive clustering. *Information Visualization,* 7 (3).

SCOTT, D. W. (1992) *Mulivariate Density Estimation,* London, Wiley.

SILVERMAN, B. W. (1986) *Density estimation for statistics and data analysis,* London, Chapman and Hall.

SMITH, GOODCHILD, M. F. & LONGLEY, P. A. (2008) *Geospatial Analysis - a comprehensive guide,* www.spatialanalysisonline.com Matador (an imprint of Troubador Publishing Ltd).

VIRRANTAUS, K., FAIRBAIRN, D. & KRAAK, J. M. (2009) ICA Research Agenda on Cartography and GI Science. *The Cartographic Journal,* 46**,** 63-75.

VISMASTER (2009) Visual Analytics – Mastering the Information Age.