



Visualizing Point Density on Geometry Objects: Application in an Urban Area Using Social Media VGI

Moris Zahtila¹  · Martin Knura² 

Received: 16 January 2022 / Accepted: 11 May 2022 / Published online: 28 June 2022
© The Author(s) 2022

Abstract

Point datasets that relate to highly populated places, such as ones retrieved from social media or volunteered geographic information in general, can often result in dense point clusters when presented on maps. Therefore, it can be useful to visualize the relevant point density information directly on the urban geometry to tackle the problem of point counting and density range identification in highly cluttered areas. One solution is to relate each point to the nearest geometry object. While this is a straightforward approach, its major drawback is that local point clusters could disappear by assigning them to larger objects, e.g., long roads. To address this issue, we introduce two new point density visualization approaches by which points are related to the underlying geometry objects. In this process, we use grid cells and heatmap contour lines to divide roads, squares, and pedestrian zones into subgeometry units. Comparison of our visualization approaches with conventional density visualization methods shows that our approaches provide a more comprehensive insight into the point distribution over space, i.e., over existing urban geometry.

Keywords Point density visualization · Subdivision of geometry features · Heatmap · Volunteered geographic information

Visualisierung der Punktdichte auf Geometrieobjekten: Eine Anwendung von Social Media VGI im urbanen Raum

Zusammenfassung

Wenn Punktdatensätze, die sich auf dicht bevölkerte Räume beziehen – beispielsweise räumliche Daten aus Sozialen Medien oder von VGI-Plattformen – auf Karten dargestellt werden, kommt es häufig zu dichten Punktclustern, was Aussagen über die Anzahl der Punkte oder die Intensität der Punktdichte an bestimmten Orten schwierig bis unmöglich macht. Daher kann es nützlich sein, relevante Informationen über die Punktdichte direkt mit Bezug zu urbanen Geometrien zu visualisieren. Eine Lösung besteht darin, jeden Punkt dem nächstgelegenen Geometrieobjekt zuzuordnen. Ein großer Nachteil dieses Ansatzes ist jedoch, dass lokale Punktcluster verschwinden könnten, indem sie größeren Objekten, z. B. langen Straßen, zugewiesen werden. Um dieses Problem zu lösen, werden zwei neue Ansätze zur Visualisierung der Punktdichte eingeführt, bei denen die Punkte mit den urbanen Geometrieobjekten in Beziehung gesetzt werden, lokale räumliche Eigenschaften jedoch erhalten bleiben. Dafür werden Straßen, Plätze und Fußgängerzonen mithilfe von Rasterzellen und Konturlinien von Kerndichteschätzungen in Teilgeometrieinheiten unterteilt. Der Vergleich dieser Visualisierungsansätze mit herkömmlichen Dichteverteilungsmethoden zeigt, dass die vorgestellten Ansätze einen detaillierteren Einblick in die räumliche Punktverteilung mit Bezug zur bestehenden urbanen Geometrie liefern können.

✉ Moris Zahtila
moris.zahtila@tu-dresden.de

Martin Knura
martin.knura@hcu-hamburg.de

¹ Institute of Cartography, Dresden University of Technology, Helmholtzstr. 10, 01062 Dresden, Saxony, Germany

² Lab for Geoinformatics and Geovisualization (g2lab), HafenCity University Hamburg, Henning-Voscherau-Platz 1, 20457 Hamburg, Hamburg, Germany

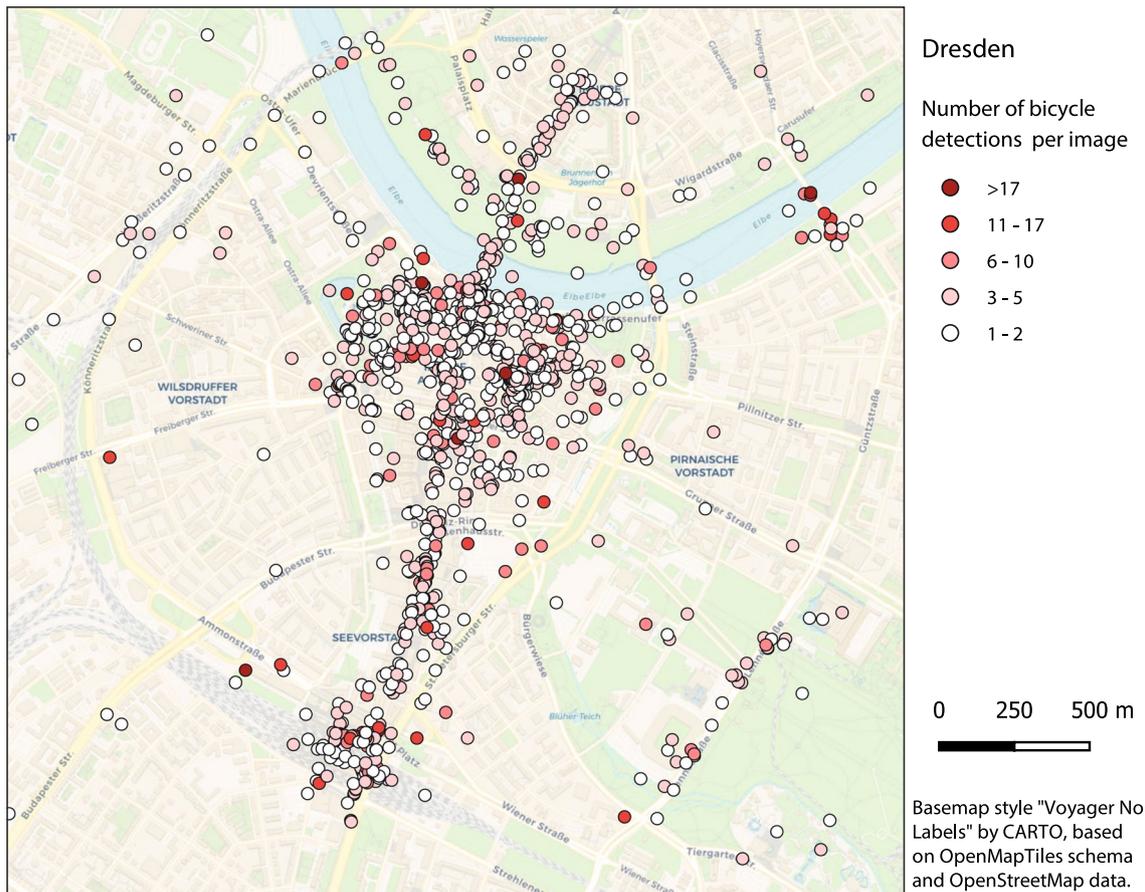


Fig. 1 Location of Flickr images on which bicycles were detected (quantitative aggregation is included through color; overlapping circles in the city center significantly affect data exploration and readability of the map)

1 Introduction

Applications for user-generated spatial web content have steadily grown ever since they were introduced as Volunteered Geographic Information (VGI) by Goodchild (2007). They support a wide range of services like environmental monitoring (Young et al. 2019), movement analysis (Boss et al. 2018), and disaster management (Crooks et al. 2013). Focusing on urban environments, examples of VGI applications are traffic monitoring (Das and Purves 2020), route choice behavior (Scott et al. 2021), decision-making (Giuffrida et al. 2019), and city perception (Bahrehdar et al. 2020).

VGI data are often generated as point data and represent a sample of the underlying population of interest. The combination of these two characteristic features of VGI often causes problems for data exploration and interpretation when *urban* data are presented on maps. While the representativeness, and therefore the fitness for use of a VGI data sample in a particular application has to be assessed based on the

data itself (Senaratne et al. 2017; Zhang and Zhu 2018), even representative VGI data could lead to overcrowded point clouds in very populated places when a large amount of points is georeferenced to a small area. For example, in the recent work of Knura et al. (2021), the information on moving and parked bicycles in the city of Dresden was retrieved utilizing an object detection algorithm on social media images. While outside of the central city areas, one can easily identify each point feature that represents the location of the photo on which bicycles were detected, this is not possible in the central city areas, where point symbolization significantly overlaps (Fig. 1). This effect, known as display clutter (Rosenholtz et al. 2007), makes it hard to read and impossible to visually count the number of points on a map.

A potential solution to eliminate display clutter is to replace the individual point symbols with a representation for the point density at each location. For VGI applications in urban environments, cumulative and statistical values for distinguished areas of interest are often more important than the individual elements of the dataset (Hasan

and Ukkusuri 2014). Therefore, it can be useful to visualize point datasets by directly linking the point density information to urban geometry objects, such as roads or pedestrian zones, instead of visualizing them as an *overlaying* layer. The most straightforward method would be to relate each point to the nearest geometry object and count all points per object. A major drawback of this approach is that local clusters could disappear when assigned to larger objects, e.g., long roads or large squares. In this paper, we introduce new visualization approaches based on relating the density information of a point dataset to the underlying geometry objects by dividing roads and pedestrian zones into smaller units using point density information. Our intention is to:

- resolve the problem of overlay between points and geometry layers to identify underlying geometry features more easily
- resolve the problem of point counting and density range identification in highly cluttered areas

We start with an overview of point density visualization methods in Sect. 2. In Sect. 3, we introduce our new methods of relating density information to urban geometry. In Sect. 4, we apply these methods to the bicycle detection dataset, discuss the results in Sect. 5, and make a conclusion in Sect. 6.

2 Density Visualization

Considering that VGI often contains large numbers of point data, there is a high likelihood of visualizing overcrowded or dense point clouds, which bears certain drawbacks for interpretation (Li et al. 2014). The exploration of the dataset can be demanding, even when varying sizes or colors of points are used as a way of aggregation and abstraction (Opach et al. 2019). The cartographic solution to overcome display clutter is point generalization, where different operations like selection and displacement can be performed to reduce clutter and the overall number of points on the map (Slocum et al. 2009). If the point density of a certain area or the whole map is of interest, the generalization process should be guided by respective constraints that preserve local and global density characteristics (Knura and Schiewe 2022).

While the generalization operation of *aggregation* also reduces the number of points on the map Burigat and Chittaro (2008), it could be used in the same way to represent quantitative data—and therefore also point density—through point symbolization (Brewer and Campbell 1998). An approach that tackled the drawback of overlapping VGI points in this way was by arranging micro-diagrams in a regular grid on a map (Gröbe and Burghardt 2020). This approach achieved high diagram readability and a high-level

overview of the data spread because of avoided overlaps. However, the interpretation of the data in relation to the basemap still stayed highly dependent on the current zoom level and estimating where the exact grid boundaries are. The same can be said for word clouds, which were recognized in visualizing spatial information as a promising approach that contains text georeferenced to a point location (Bertone and Burghardt 2017). This can be used in urban planning to better understand the common usage of urban areas, e.g., by visualizing the locations of keywords contained in georeferenced social media data (Dunkel 2015) or diaries of activities in urban areas (Li and Zhou 2017). Here, the keyword's font size communicates the occurrence of that word on a specific location or urban area. However, it can still bring ambiguity in interpretation because it is not intuitively clear how large the buffer zones are to which the keywords relate. Additionally, the basemap is again visually covered by the layer that contains keywords visualizations.

A common, but completely different example of density visualization are heatmaps as a type of isarithmic maps that use color to represent zones of same values. In their most simple form, point density is measured by a form of zoning and counting the number of points in each zone, while kernel density procedures are the most common way to create heatmaps (De Smith et al. 2007). On standard visualizations which use heatmaps, the layer displaying the densities is usually an overlay above the basemap, e.g., in Fisher (2007). These overlays are semi- or non-transparent layers which can easily decrease the overall map readability and make the process of relating the density information to urban geometry more difficult and an exploratory process. A juxtaposed visualization of basemap and heatmap makes it harder to precisely identify units of urban geometry, because it is first necessary to identify a density area of interest, and then visually relate the information conveyed by it to a street, square, or park in the adjacent map (Lobo et al. 2015). Manually relating the density information to urban geometry can be done, e.g., by turning on and off the density visualization layer in an interactive map (especially in the case of non-transparent density visualization), or by trying to identify the geometry features on the basemap through a semi-transparent density visualization layer. Generally, these tasks can easily become time-consuming and require an increased cognitive effort while suffering from problems of split attention.

Some solutions for visual exploration of densities related to urban areas are designed as interactive dashboards. Already Hotta and Hagiwara (2009) recognized the importance of the efforts to provide interactive maps. In their work, the authors focused on developing a method for a faster generation of heatmaps from big data. Recently, Zhu et al. (2019) designed a dashboard that

displays additional information in side views upon selecting an area or a point on the map. Similar tasks allow also Li et al. (2018) in their dashboard with the difference that it uses semi-transparent points that are classified by colors.

3 Relating Points to Geometry Features

Considering the drawbacks of existing approaches—such as overlapping of layers, dense point or word clutters, or spatial imprecision—we have identified the need to develop a density visualization approach that would ease and allow more precise identification of basemap features in relation to the density visualization. In this section, we present density visualization approaches that we developed with the intention of minimizing these drawbacks.

Because VGI data are often obtained within urban environments, our aim is to relate point density information to the existing geometry in urban areas—to which we further refer as *urban geometry*. For convenience and simplification purposes, the urban geometry on which we focus in this paper are roads, squares, and other pedestrian zones. We therefore only use line objects of roads and polygon objects of squares and pedestrian zones. We are aware that for many VGI applications, this subset of urban geometry is not sufficient, but the approaches shown in this section can be expanded to more heterogeneous datasets. As a second simplification, we removed all road parts that spatially overlap with polygons of squares and pedestrian zones to guarantee a consistent visualization.

We start this section with a short introduction into kernel density estimation, which provides an overlaying estimation of point density at each location. We then present a contrary concept of density visualization by relating each point to its nearest urban geometry object. In the last

part of this section, we introduce two new approaches that extend the first two density visualizations.

3.1 Deriving Point Density Surfaces Using Kernel Density Estimation

Kernel density estimation (KDE) is an important method to visualize the spatial pattern of point events over a given location and is applied to numerous fields, such as criminology (Hart and Zandbergen 2014), epidemiology (Shi et al. 2021), or ecology (Péron 2019). The idea of KDE is to estimate densities of specified features at a location using a window (i.e., *kernel*) of its surrounding area with a specified bandwidth or radius and can be determined as (adapted from Silverman (1986)):

$$\hat{f}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{d_{i,(x,y)}}{h}\right), \quad (1)$$

where \hat{f} is the estimated density at a location with spatial coordinates (x, y) , n is the total number of points under concern (e.g., images with bicycle detections), h is the kernel bandwidth (i.e., the search radius from (x, y)), $d_{i,(x,y)}$ is the distance between location i (where i are the number of points under concern) and location (x, y) , and K is the kernel shape, a function used to measure the distance decay effect of the points. While there are several other kernel shape functions available—such as Gaussian, uniform, triangle, and Epanechnikov—we used a weighted quartic function to create heatmap zones in this paper. This function can be represented as:

$$K\left(\frac{d}{h}\right) = \begin{cases} W \frac{15}{16} (1 - d^2)^2 & \text{if } 0 < \left(\frac{d}{h}\right) \leq 1, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

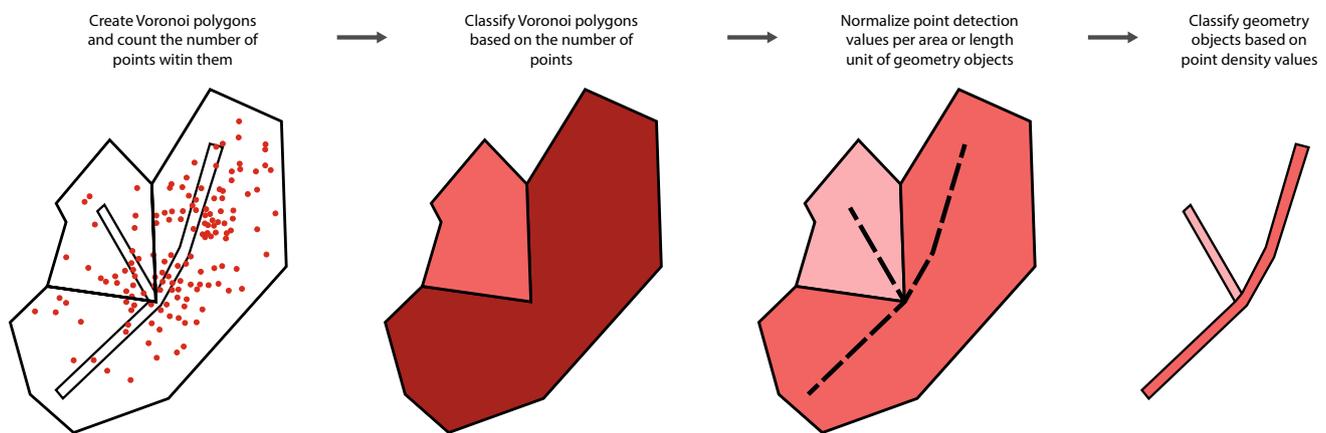


Fig. 2 Workflow for classification of geometry features by assigning points to the nearest geometry object

where W is the weight, e.g., the number of detected bicycles on the image. It is generally accepted that the selection of the kernel shape is not critical, while the bandwidth has a major impact on the KDE's outcome (Shi 2010). A small bandwidth maintains local patterns better, while a larger bandwidth provides a smoother estimate.

3.2 Classifying Geometry Features Based on the Number of Associated Points

One solution to overcome the issues of overlying point and density information is to directly classify the urban geometry features according to the number of points in their immediate neighborhood. Therefore, we first created a model that assigns each georeferenced point to the nearest urban geometry feature as shown in Fig. 2. If not already located within a polygon feature, each point is assigned to the local Voronoi area, which we created for each line using a create-points-on-line operation and joining the respective areas for each polyline. Then, each feature was classified based on the number of points assigned to it. This approach resolves the issue of overlaps and overlying layers and allows to easily identify the numerical range of assigned points for the different urban geometry features.

As it can be noted on the schematic representation of this approach (Fig. 2), the model classifies geometry features in their full geometry, no matter how large the polygon or how long the polyline is.

3.3 Dividing Geometry Features Into Units Based on Point Density Information

While the aforementioned approach uses the whole geometry of an object, more elaborate approaches should account for the actual spatial spread of point density within the objects. Therefore, we created two GIS models which first relate point detections to density polygons and then use these polygons to divide urban geometry features into smaller units. With this approach, the models provide a better overview of the exact point density spread over longer roads or larger squares and pedestrian zones.

3.3.1 Intersecting Point Clouds with Urban Geometry Using a Grid

In the first model, the whole urban area is first divided into smaller units using a grid, and then the number of points

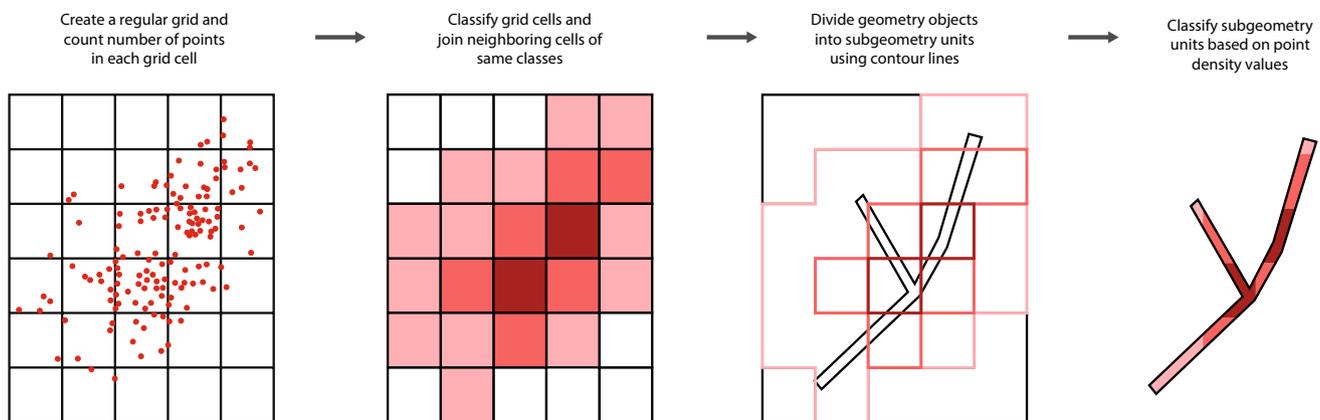


Fig. 3 Workflow for subdivision and classification of geometry features using grid cells

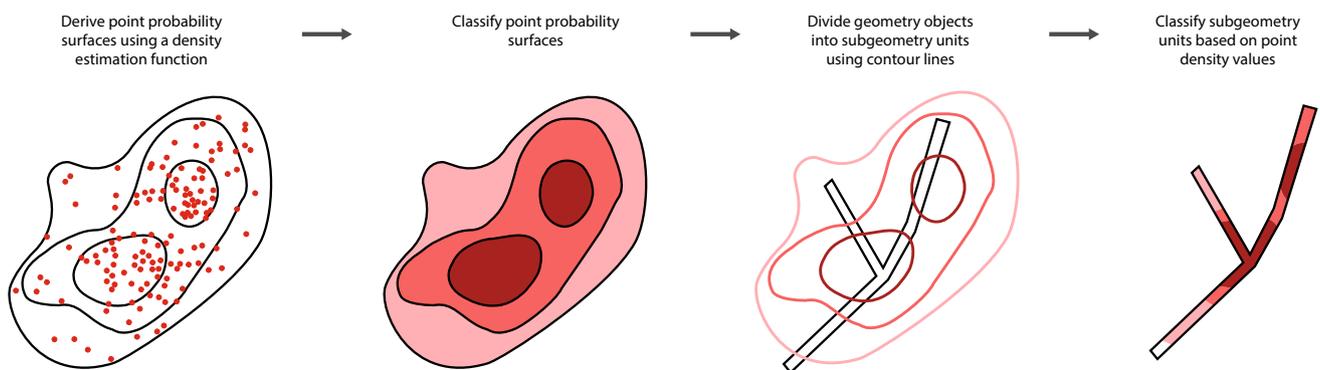


Fig. 4 Workflow for subdivision and classification of geometry features using contour lines derived from a heatmap

in each grid cell is counted. In the next step, the grid cells are classified according to the number of points within them, following a manually defined classification. Then, adjacent cells within the same classes are merged. After this, exterior boundaries of the newly created areas are used to divide the geometry features into smaller units and, finally, to visualize the classified urban geometry units. Figure 3 schematically visualizes this workflow. Compared to the approach in Sect. 3.2, geometry features are divided into subgeometry units of lower and higher point density and can be classified according to the underlying grid cells classes. The intermediate classification of the grid cells has an influence on the actual shape of the joined grid cells, and therefore the result of the process. For that reason, it is meaningful to choose a classification that is appropriate for used dataset, user task, and intended visualization purpose.

3.3.2 Intersecting Point Clouds with Basemap Geometry Using Contour Lines of a Heatmap

In the second model, estimated point density information is derived from a heatmap to be assigned to the urban geometry features. First, a heatmap is created from the input point data as described in Sect. 3.1. Based on the heatmap, respective contour lines that follow the edges of zones with different density values are generated. These contour lines are then intersected with the urban geometry features in the same way as in the grid-based approach. The result is, again, subdivided urban geometry units, which are then classified according to the corresponding heatmap area—i.e., based on the values which are separated through the contour lines. The workflow is shown in Fig. 4. The defined distance between the contour lines influences the actual shape of the resulting contour lines—and, therefore, the derived geometry objects—and should be chosen according to the used dataset, user task, and intended visualization purpose.

4 Application of Visualization Methods

In this section, we apply the density visualization approaches described in Sect. 3 to compare them with the common heatmap and pointmap approaches. We implemented the approaches as QGIS models and created the respective maps using VGI data from the social media platform Flickr and geometry data from the City of Dresden and Open Street Map (OSM). In the following subsections, we first give a short introduction to the data and parameter settings used for creating the maps before we present the respective visualizations that resulted from the developed GIS models.

Table 1 Parameter values used for implementing visualization approaches

Approach	Parameters
Heatmap	Heatmap radius: 50 m; Resolution: 0.1 m × 0.1 m
Nearest neighbor:	Normalization per area of 2500 m ²
Grid cells	Cell size: 50 m × 50 m
Contour lines	Heatmap radius: 50 m; Equidistance: 25 m

4.1 Data and Visualization Settings

VGI data As reference VGI data, we use the dataset from Knura et al. (2021), where images containing bicycle detections in the area of Dresden, Germany, were retrieved from the Flickr YFCC100M dataset. The latter dataset was published under Creative Commons and contains 100 million user posts geolocated all over the world as photos or videos and related textual descriptions, covering the time span between 2004 and 2014. From this dataset, Knura et al. (2021) selected all posts georeferenced in Dresden and containing images and ran on them an object detection algorithm to identify bicycles. The detected bicycles were then classified as either stationary or moving and the case study showed that the geographical spread of the retrieved dataset over the city is typical for VGI data—as previously shown by (Jiang et al. 2016)—and can be seen as *in-the-wild sensing* of the city (Zhu et al. 2013). However, for this article, we do not differentiate between moving and stationary bicycles but use all images with bicycle detections as shown in Fig. 1.

Urban geometry data The urban geometry features that we work with in this paper are roads and pedestrian zones. The vector data of the road network in Dresden have been made openly available by the City of Dresden (State Capital Dresden 2021). In the dataset, roads are modeled as single lines that represent road lanes of both driving directions, as well as sidewalks. The dataset also includes several road categories that are not applicable for bicycles, such as railroad lines, so we removed these entries before running the models. For pedestrian zones in Dresden that include squares and other pedestrian areas, we used OSM polygons because these were not included in the former Dresden dataset.

Visualization settings To ensure comparability of results between different approaches, we used comparable parameter settings for the operational steps within our models. Table 1 provides an overview of the parameters used. Before deciding on these specific values, we analyzed the effects of various parameter combinations and decided on the latter because they provided a suitable

visual representation of our VGI dataset considering the map scales in this paper. We based the parameters in each approach on the value of 50 m because lower values resulted in patterns that were too small to be well visible on maps showing wider Altstadt—the city center of Dresden—and because too many points would not be assigned to any urban feature (e.g., a significant number of grid cells smaller than 50 m × 50 m did not intersect any road or square object in the grid-based approach). Greater values resulted in visualizations where the point hotspots were spreading over large areas (e.g., over whole squares), providing an overly generalized result. As the roads are modeled as polylines, to make them comparable with pedestrian zones that are polygons, we used average road width of 10 m in the process of normalizing density values.

To make the visualizations reflect the results of the models presented in this paper in a meaningful way, we implemented a classification that considers the characteristics of the social media dataset used. In the dataset, more

than half of images with bicycle detections contain only one bicycle, while merely 6.7 percent contain more than five (Knura et al. 2021). Considering also that in the latter paper most of the 100 m × 100 m grid cells contained up to 35 parked bicycle detections on photos posted within the cells, we expected a similar pattern for bicycle detections in general. Therefore, we decided to use an incremental classification that starts with five bicycle detections and doubles the value of the previous class in each next class. That way, we defined seven classes of bicycle detections from posted photos: 0, 1–5, 6–15, 16–35, 36–75, 76–155, and more than 155 bicycles; which resulted in a good distinction between urban geometry units in different classes. This classification visually distinguishes well the classes with a smaller, middle, and higher number of points. A common classification on all maps helped to compare the results of different approaches throughout the maps. All visualizations were created using QGIS Desktop, version 3.16.1 Hannover.

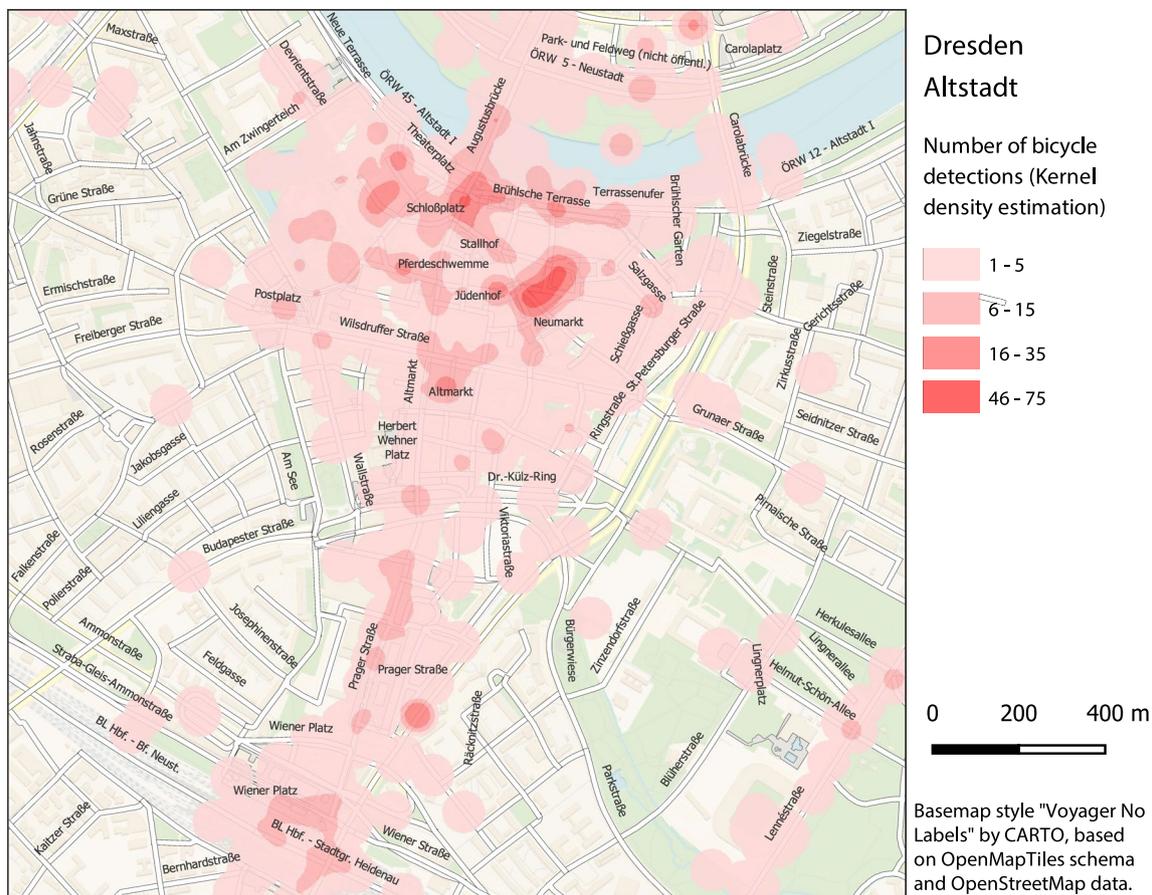


Fig. 5 Heatmap showing interpolated bicycle density information based on bicycle detections on photos posted on Flickr

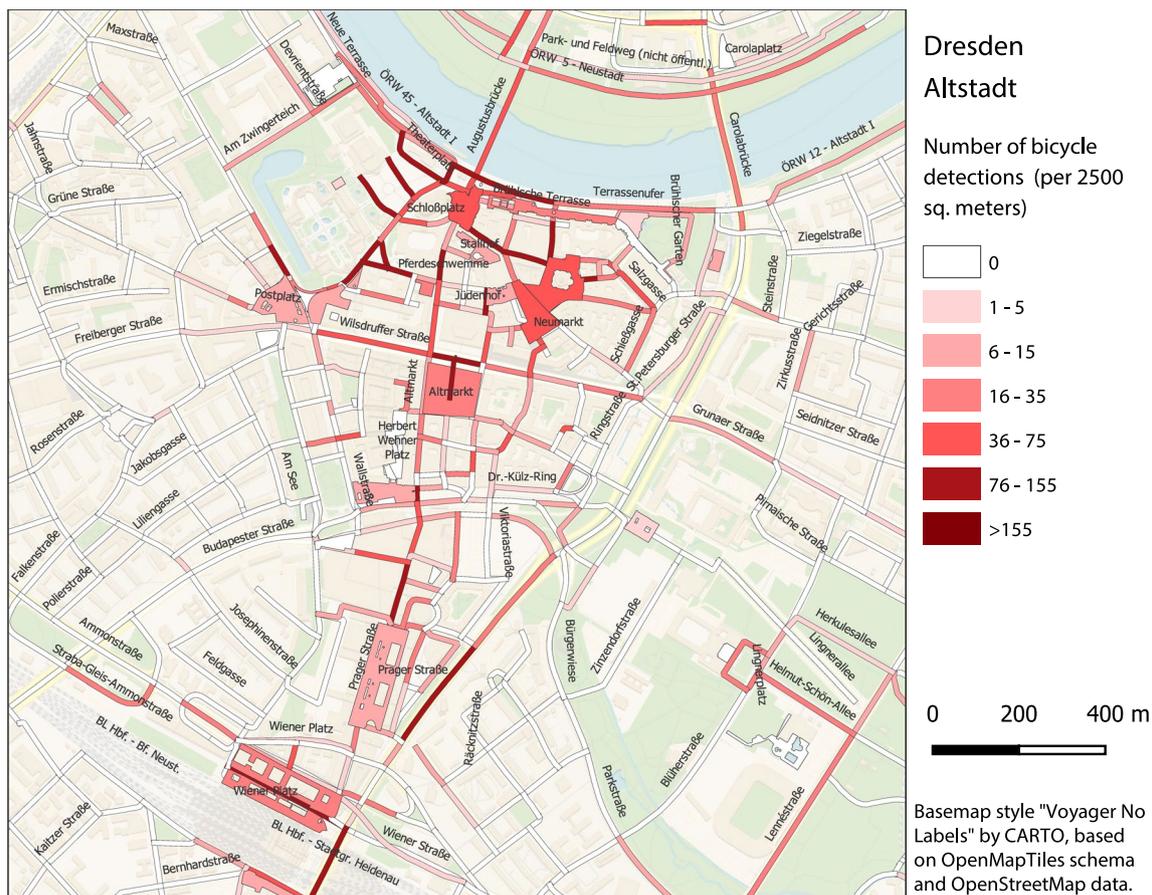


Fig. 6 Points of photos with bicycle detections related to original urban geometry (without further subdivision to smaller units)

4.2 Visualizations

4.2.1 Pointmap and Heatmap

Compared to the pointmap, a heatmap could provide more clarity in areas where numerous points overlap. For example, by eyeing a point clutter on a pointmap, only points at the top can be interpreted unambiguously, while the quality and quantity of underlying points are largely occluded (Fig. 1)—which negatively affects the overall readability of the map. In contrast, a heatmap like on Fig. 5, which was created using the Heatmap (Kernel Density Estimation) tool in QGIS, provides the cumulative point density estimation for each map pixel, so there is no cluttering and overlapping of points. Still, the density visualization is here as well a separate layer that overlays geometry features, significantly occluding parts of the layers below it. Because of the overlay, identifying urban geometry features below the areas that would display probability of very high density would, due to much darker shades, require a comparison with a map that is not occluded and, therefore, would require also a higher cognitive effort. As it can be seen in Fig. 5, it is easy to find

hotspots where images with bicycle detections are expected, but the readability and contrast of the underlying layers are significantly reduced.

4.2.2 Density-Geometry Map Using Nearest Neighbor Relation

One of the possible solutions to relate points to urban geometry—a straightforward approach where each point is assigned to the nearest or underlying urban geometry object—was described in Sect. 3.2. The implementation of this approach is presented in Fig. 6. The map shows urban geometry objects classified based on the number of detected bicycles located on or in close proximity of the object. A map created using this approach allows the user to easily identify roads and pedestrian zones where a high number of point features were located (in this case, image posts with bicycle detections). For example, there are more than ten road segments with more than 75 bicycle detections from images. Among squares, three with the highest values are classified into the middle class of 36–75 bicycle detections. However, in the case of larger pedestrian zones (like the one

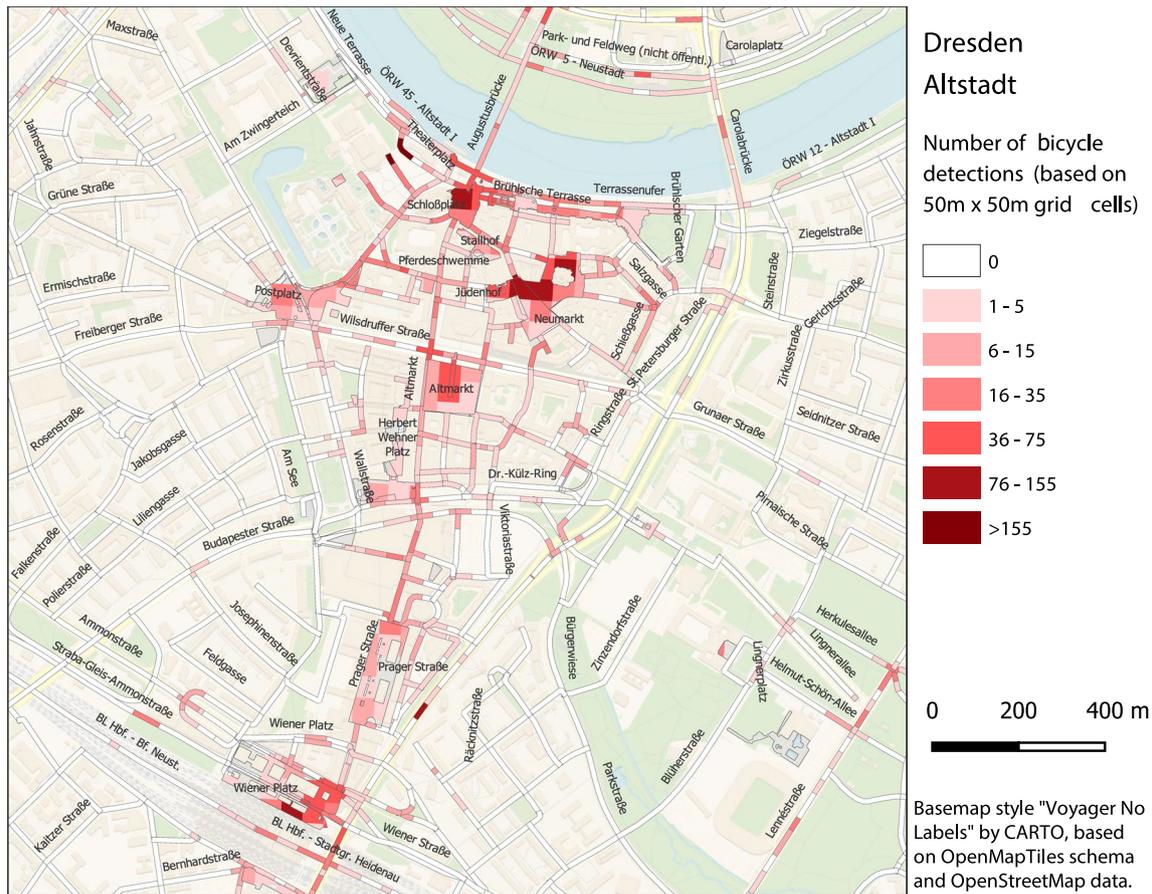


Fig. 7 Subdivision of streets, roads, and pedestrian zones based on their intersection with classified grid cells

that is a combination of the Neumarkt and An der Frauenkirche squares), or elongated ones (like Brühlsche Terrasse that is around 300 m long), classifying the whole object into one single class seems too general for specific purposes, providing too little details about the spatial pattern spread.

4.2.3 Density-Geometry Map Using a Grid

Figure 7 visualizes the result of implementing the grid-based approach (Sect. 3.3.1). Large polygons and long roads are here divided into smaller, classified subgeometry units. The underlying grid structure is visible on most of the polygons. This differentiation reveals a unique pattern for each of the three largest squares in the center of the map. None of them has an evenly spread point distribution. In the North, both Schlossplatz and Neumarkt and An der Frauenkirche (considered together as one larger pedestrian zone) have most of the photos with bicycle detections in their north-western part, while Altmarkt in the central part of the map has the majority in its center. All the more, Brühlsche Terrasse is here divided into six zones, while in Fig. 6, it is one. In the same fashion, more details can also be identified on roads

because smaller road sections at which the photos with bicycle detections occurred are now highlighted instead of having the whole original road segment classified in one class. This effect is nicely visible on most of the roads, e.g., on Lennéstrasse in the lower right corner.

4.2.4 Density-Geometry Map Using a Heatmap

As a result of applying the approach that subdivides the geometry objects based on heatmap contour lines (Sect. 3.3.2), there are much fewer subgeometry units classified into the class of zero-values than in the grid-based approach, which can be observed in Fig. 8. While this is the consequence of assigning the heatmap density estimation values to the respective subgeometry units, small patterns of local extreme values are less visible and highlighted here. Besides that, within squares, very much the same patterns as in the grid-based approach appear, but here have rounded shapes. For example, at Altmarkt, in the center of the map, two hot spots within the square are visible, and it can be identified that the majority of image posts with bicycle detections should appear at the upper center of the square. In comparison to the grid-based approach, on linear

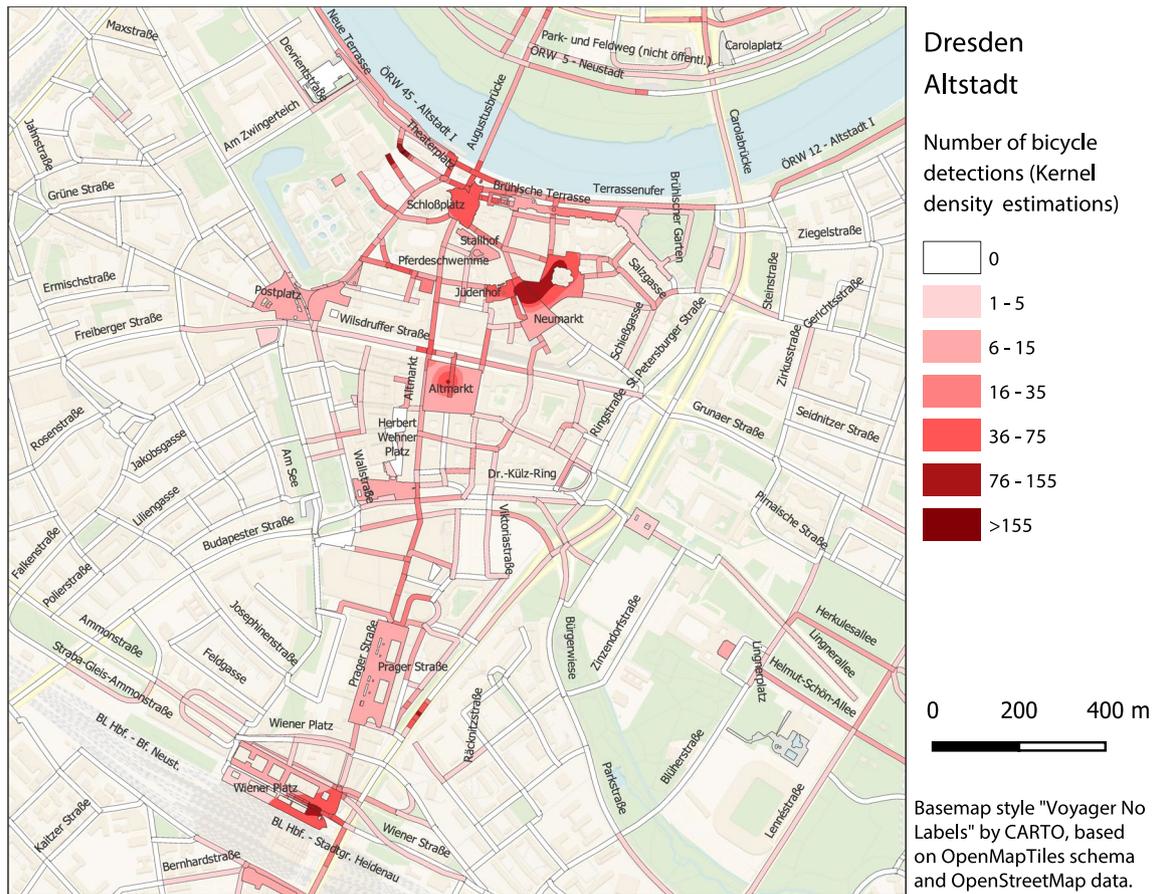


Fig. 8 Subdivision of streets, roads, and pedestrian zones based on intersection with contour lines derived from the heatmap

objects, this visualization possibly provides fewer details by differentiating a lower number of segments. The difference becomes even more obvious on the shorter streets southern of Altmarkt. These geometry units largely kept their original length here but were classified differently than in the baseline-based approach. In general, this approach provides a very smooth visual appearance of density zones and provides a more accurate precision representation in respect to the uncertainties of geocoding and localization in VGI, but with the cost of eliminating potentially critical micro-locations.

5 Discussion

5.1 Resolving the Problem of Density Information as an Overlying Layer

Compared to the map with point clouds in Fig. 1, maps in Figs. 5, 6, 7, and 8 all provide a density overview of points with less display clutter and therefore much higher readability of the map. Still, there is another major difference between the heatmap in Fig. 5 and the following three maps:

The surfaces of the density estimation are on the heatmap noticeably overlaying the basemap and urban geometry, making it rather inconvenient to relate spatial patterns of point density to exact local infrastructure. A possible method to solve this problem is to lower the opacity of the density estimation layer.

While this allows establishing a relation between the density pattern and the underlying layers, significantly lowering the opacity also results in a lower contrast between the point density estimation surfaces and underlying layers, which again decreases the readability of the density values. The maps in Figs. 6, 7, and 8 address this problem by relating the point density information directly to the urban geometry features. These features can, as a result, then be clearly and quickly identified as they directly provide the information about the (estimated) number of points—i.e., georeferenced images with bicycles detections—related to each urban geometry object. Visualizing an overlying layer with point density information above the urban geometry layer and relating the point density information between layers is, in this case, no longer needed.

Furthermore, heatmaps still require a visual exploration, while our methods allow also other ways of handling the data

because the density-related information becomes an attribute appended to the geometry features dataset. This information can then be filtered, searched, further processed, or presented in additional ways (e.g., as a tabular list of subgeometry objects ordered by density values). Additionally, storing the density-related information in such a manner allows the visualizations to more easily be recreated; which provides a better time efficiency knowing that not all of the heatmap production methods can in GIS software be stored as permanent layers. In other words, the heatmap-creating tool would usually need to be re-run to re-create visualizations, which makes it more time-consuming and requires more effort.

5.2 Resolving the Problem of Point Density Identification

The approach presented in Sect. 3.2 has a major advantage: maintaining the original geometries. Technically, the numbers of geometry objects and object edges stay the same compared to the input dataset, which can, in some cases, be limiting in the exploration of spatial patterns. Despite this advantage, the approach does not offer the possibility to separately detect important pattern-related micro-locations within objects. This is, for small or short objects, not necessarily a significant issue, but can be limiting for larger or longer objects whose parts may be differently classified according to the number of corresponding points features. As a result, no meaningful spatial patterns can be identified on those objects other than the general conclusion: the closer to the city center (or the more popular place), the higher the number of points. This can be the case even for a dataset like a road network segmented using road intersection points.

Even with segmented roads, critical to visualize are segments within the city center where there is a large fluctuation of people and, thus, multiple micro-locations exist where important points clusters (e.g., georeferenced posts) could occur. Not even normalizing the number of point detections over the polygon's area or length results in more helpful information: It only shows how frequented the urban geometry object in general is and does not allow a precise detection of all the important micro-locations.

In contrast, the approaches presented in Sect. 3.3 *subdivide* these geometry features, leading to a higher number of map objects and respective edges. By implementing subdivision, they generally provide a higher level of details regarding the identification of spatial patterns *within* geometries, which is especially meaningful for investigating large or elongated geometry objects. Using the models described in Sect. 3.3, we were able to differentiate density hotspots on all geometry objects. The major advantage of these approaches is that the point density patterns are well visible on polygonal objects like squares and pedestrian zones—an effect that can be well observed in Fig. 9. The grid-based approach appears to be useful for the subdivision and classification of roads and other linear objects, providing a comprehensive subgeometry classification. It also allows visualizations of micro-locations for areas like squares or pedestrian zones (Fig. 9b). The advantage of the approach based on contour lines is that the point density patterns are very well visible on polygonal objects. The patterns receive a smoother shape on larger areas, which may be more intuitive to interpret than the grid cells because of high similarity to the zones of a heatmap (Fig. 9c). A drawback of the contour-line-based model in comparison to the grid-based is that on shorter

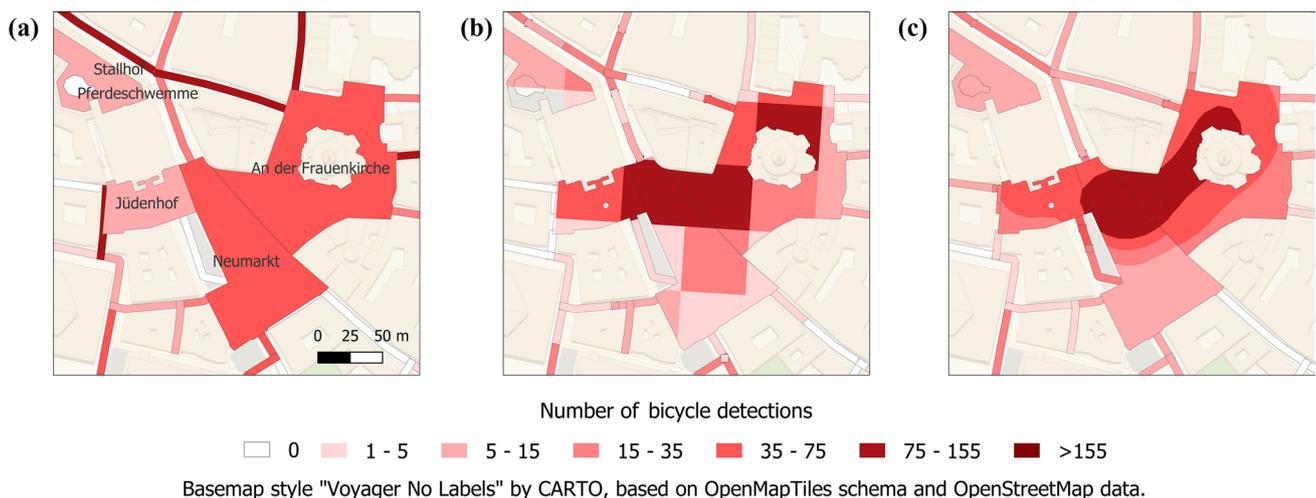


Fig. 9 Comparison of data visualization approaches on an urban micro-location in Dresden: **a** classification with Voronoi areas (per 2500 m²); **b** grid-based subdivision (50 m × 50 m grid cells); **c** contour-lines-based subdivision (heatmap radius 50 m)

linear objects, it tends to provide fewer details, meaning it results in a smaller number of subgeometry units.

5.3 Influence of Input Data and Parameter Values on Visualizations

As a matter of course, the level of resulting details for each of the visualization approaches is influenced by the input data and respective parameter values. Using a dataset where the urban geometry is modeled differently than in the dataset used in this paper could have a major impact on spatial precision and granularity of visualized patterns: e.g., using road data from OSM, where driving lanes of opposite directions on wider roads are modeled separately, would be especially meaningful when there is a larger distance between road lanes, like a green area or a tram line between them. Thereby, increasing the level of granularity of the input data—i.e., having more objects and a higher cumulative object area—will lower the density values for some of the urban geometries when using the nearest neighbor relation. By contrast, an advantage of both approaches presented in Sect. 3.3 is that the local density is calculated independently from the urban geometry, and therefore the visualized density pattern is not affected when changing the input data.

For all but the nearest neighbor approach, the parameter value selection has a crucial effect on the resulting map and has to be a task-related decision. When using the grid-based approach, the grid cell size, as well as the classification on which the cells get aggregated, influence the final result. The smaller the grid cell is, the more detailed the spatial pattern within the urban geometries will be. However, if the cell size is too small, a significant number of the grid cells could not contain any points, which could lead to the point spread pattern getting underrepresented on the urban geometry features because intersections of empty cells with urban geometry units will be more common. The result of the second step in Fig. 3 would then be comparable to a heat map created using the uniform kernel function with a very small bandwidth. As described in Sect. 3.1, the bandwidth selection is crucial for KDE. For the approach that uses contour lines, the equidistance between the contour lines is the second important parameter. The larger the distance between the contour lines is, the less will the spatial pattern be visible in the resulting visualization.

5.4 Introduction of Uncertainties

In this paper, we aggregated point data to improve the usability of the map, but this generalization step also introduced uncertainty to our approaches. For the grid-based approach, we discussed the influence of the grid cell size on the result in the section before. Besides the size of the cells, their initial positioning could also have a major impact on the

resulting spatial patterns displayed on the map. For example, it could be interesting to research the effects of shifting the cells for 50% of their size into different directions. Our assumption is that, in some locations, the resulting density pattern could change significantly. Another approach could also be to design an algorithm that automatically places the cells in a fashion that takes into account the constellation of the densest point clusters. Then, the cells would not be distributed within a bounding box that surrounds the dataset strictly from border to border starting from one side but would be distributed independently on a perfect fit within the bounding box in a way that best covers most populated locations.

In general, placing grid cells without taking into account the positions of dense point clusters potentially highlights density borders as much more certain than they actually are—especially because edges are an important factor in the cognitive map reading process. When intersected with urban geometry objects, these uncertainties could increase even more. To reduce this edge-related bias, possible solutions (other than the one mentioned in the paragraph before) would be to rotate the grid cells according to the major orientation of the road network in a city, or to use other shapes for grid cells, such as diamond or hexagon shapes, to see if they suit better to the urban geometry. Another solution for the edge-related bias—implementing softer border shapes—is already implemented in the approach based on the contour lines of a heat map because the kernel shapes prevent hard edges through the distance decay effect.

5.5 Influence of Scale and Data Privacy

When searching for the best parameter values, the targeted map scale and its impact on the visualization also have to be considered. First, as we use point density to *divide* urban geometries, the size of the input geometries must be suitable for the targeted map scale, even after being divided into several smaller units. At this point, legibility constraints from map generalization could help to decide if the input geometries are suitable (Stigmar and Harrie 2011). Second, the cell size of the grid-based approach, as well as the equidistance between the contour lines in the approach based on heatmap zones, has to be selected accordingly to the map scale and dependently on the task.

In this paper, we used an input dataset where each point represents the location of a social media post, which could possibly corrupt its contributor's privacy due to georeferencing. Using privacy-sensitive data like social media data—and VGI in general—needs to guarantee contributors' privacy even when the data are not directly linked to the original dataset that contains information on the contributors. In this regard, the resulting dataset of our approaches provides data privacy because the information on single

points is lost by being aggregated within density surfaces that are assigned to the geometry objects. We also foresee the possibility of adapting the models of our approaches for processing privacy-aware data format proposed by Dunkel et al. (2020) as long as the spatial precision is applicable to the used geometry.

6 Conclusion

The presentation of VGI point data on maps tends to generate dense point clutters in highly populated places when a large amount of data are georeferenced to a small area. A common way to visualize the point density information of overlapping objects is to create a heatmap using kernel density estimation. While this solves the problem of cluttered point symbols, heatmaps have significant shortages for user tasks where density information has to be related to underlying geometry features. In this article, we tackled this problem and introduced two new approaches of visualizing point density using grid cells and heatmap contour lines, respectively. We showed that both approaches provide a way to visually differentiate micro-patterns within smaller and larger geometry objects, and by this in a detailed way and without overlaying and occluding the geometry features reveal characteristics of the dataset such as a specific spatial spread. Our approach based on a grid exactly highlights areas of high point density within squares and pedestrian zones and points out the parts of roads where points occur. In contrast to it, the approach based on contour lines generates a dataset representation in rounded shapes due to the edge-smoothing effect of the heatmap's contour lines.

An aspect to which we want to focus on in future work is the optimization of the underlying grid regarding cell positioning, shape, and orientation. The other main focus of our future work will be to integrate our geometry subdivision and classification approaches into a dashboard for urban planning. In it, the used dataset of bicycle detections on social media images can be combined with other relevant data sources, e.g., existing bicycle parking stations, which could help to identify locations in a city that have a high need of bicycle-related infrastructure.

To further evaluate the usability of our approaches in a working environment, a user study needs to be conducted with experts facing similar density visualization and exploration tasks, such as urban planners. This will provide us better insights in the usability of our approaches when solving practical tasks, and reveal other datasets and applications that could benefit from approaches of visualizing point density as an urban geometry attribute, as presented in this paper. Possible applications we can anticipate are analyzing the return areas of e-scooters, bicycle or car-sharing systems,

identifying locations of frequent traffic accidents, or deducing the best bird-observing spots based on historical bird-observation data.

Acknowledgements We would like to thank our research group colleague Florian Kluger who provided valuable help in this research. His comments were an important factor during the ideation phase, and his proofreading of the manuscript allowed us to improve it meaningfully. We would also like to thank professor Dirk Burghardt for his support during all research phases and for feedback that helped to improve the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This collaboration was realized within the DFG Priority Programme (SPP 1894/2) and supported by grants EVA-VGI 2 (BU 2605/8-2) and TOVIP (SCHI 1008/11-1).

Code Availability The QGIS models are available upon request.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bahrehdar AR, Adams B, Purves RS (2020) Streets of London: using flickr and openstreetmap to build an interactive image of the city. *Comput Environ Urban Syst*. <https://doi.org/10.1016/j.compe-nvurbsys.2020.101524>
- Bertone A, Burghardt D (2017) A survey on visual analytics for the spatio-temporal exploration of microblogging content. *J Geovisualiz Spatial Anal* 1:2. <https://doi.org/10.1007/s41651-017-0002-6>
- Boss D, Nelson T, Winters M, Ferster CJ (2018) Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. *J Transp Health* 9:226–233. <https://doi.org/10.1016/j.jth.2018.02.008>
- Brewer C, Campbell AJ (1998) Beyond graduated circles: varied point symbols for representing quantitative data on maps. *Cartogr Perspect* 29:6–25. <https://doi.org/10.14714/CP29.672>
- Burigat S, Chittaro L (2008) Decluttering of icons based on aggregation in mobile maps. In: Meng L, Zipf A, Winter S (eds), *Map-based mobile services: design, interaction and usability*. Springer, Berlin, pp 13–32. https://doi.org/10.1007/978-3-540-37110-6_2
- Crooks A, Croitoru A, Stefanidis A, Radzikowski J (2013) #earthquake: Twitter as a distributed sensor system. *Trans GIS* 17(1):124–147. <https://doi.org/10.1111/j.1467-9671.2012.01359.x>
- Das RD, Purves RS (2020) Exploring the potential of twitter to understand traffic events and their locations in greater Mumbai, India.

- IEEE Trans Intell Transp Syst 21(12):5213–5222. <https://doi.org/10.1109/TITS.2019.2950782>
- De Smith M, Goodchild M, Longley P (2007) Geospatial analysis: a comprehensive guide to principles, techniques and software tools. Troubador Publishing Limited, Harborough
- Dunkel A (2015) Visualizing the perceived environment using crowd-sourced photo geodata. *Landsc Urban Plan* 142:173–186. <https://doi.org/10.1016/j.landurbplan.2015.02.022> (Special Issue: Critical Approaches to Landscape Visualization)
- Dunkel A, Löchner M, Burghardt D (2020) Privacy-aware visualization of volunteered geographic information (vgi) to analyze spatial activity: a benchmark implementation. *ISPRS Int J Geo Inf* 9:10. <https://doi.org/10.3390/ijgi9100607>
- Fisher D (2007) Hotmap: looking at geographic attention. *IEEE Trans Vis Comput Gr* 13(6):1184–1191. <https://doi.org/10.1109/TVCG.2007.70561>
- Giuffrida N, Le Pira M, Inturri G, Ignaccolo M (2019) Mapping with stakeholders: an overview of public participatory GIS and VGI in transport decision-making. *ISPRS Int J Geo Inf* 8:4. <https://doi.org/10.3390/ijgi8040198>
- Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. *GeoJournal* 69(4):211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- Gröbe M, Burghardt D (2020) Micro diagrams: visualization of categorical point data from location-based social media. *Cartogr Geogr Inf Sci* 47(4):305–320. <https://doi.org/10.1080/15230406.2020.1733438>
- Hart T, Zandbergen P (2014) Kernel density estimation and hotspot mapping. *Policing* 37(2):305–323. <https://doi.org/10.1108/PIJPSM-04-2013-0039>
- Hasan S, Ukkusuri SV (2014) Urban activity pattern classification using topic models from online geo-location data. *Transp Res Part C Emerg Technol* 44:363–381. <https://doi.org/10.1016/j.trc.2014.04.003>
- Hotta H, Hagiwara M (2009) Online geovisualization with fast kernel density estimator. In: 2009 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology, vol 1, pp 622–625. <https://doi.org/10.1109/WI-IAT.2009.105>
- Jiang B, Ma D, Yin J, Sandberg M (2016) Spatial distribution of city tweets and their densities. *Geogr Anal* 48(3):337–351. <https://doi.org/10.1111/gean.12096>
- Knura M, Schiewe J (2022) Analysis of user behaviour while interpreting spatial patterns in point data sets. *KN J Cartogr Geogr Inf*. <https://doi.org/10.1007/s42489-022-00111-9>
- Knura M, Kluger F, Zahtila M, Schiewe J, Rosenhahn B, Burghardt D (2021) Using object detection on social media images for urban bicycle infrastructure planning: a case study of Dresden. *ISPRS Int J Geo Inf* 10:11. <https://doi.org/10.3390/ijgi10110733>
- Li D, Zhou X (2017) “leave your footprints in my words”—a georeferenced word-cloud approach. *Environ Plan A Econ Sp* 49(3):489–492. <https://doi.org/10.1177/0308518X16662273>
- Li C, Baciu G, Han Y (2014) Interactive visualization of high density streaming points with heat-map. In: 2014 international conference on smart computing, pp 145–149. <https://doi.org/10.1109/SMARTCOMP.2014.7043852>
- Li C, Dong X, Yuan X (2018) Metro-Wordle: an interactive visualization for urban text distributions based on Wordle. *Vis Informat* 2(1):50–59. <https://doi.org/10.1016/j.visinf.2018.04.006> (Proceedings of PacificVAST 2018)
- Lobo MJ, Pietriga E, Appert C (2015) An evaluation of interactive map comparison techniques. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. Association for Computing Machinery, New York, pp 3573–3582. <https://doi.org/10.1145/2702123.2702130>
- Opach T, Korycka-Skorupa J, Karsznia I, Nowacki T, Gołabiowska I, Rød JK (2019) Visual clutter reduction in zoomable proportional point symbol maps. *Cartogr Geogr Inf Sci* 46(4):347–367. <https://doi.org/10.1080/15230406.2018.1490202>
- Péron G (2019) Modified home range kernel density estimators that take environmental interactions into account. *Mov Ecol* 7(1):16. <https://doi.org/10.1186/s40462-019-0161-9>
- Rosenholtz R, Li Y, Nakano L (2007) Measuring visual clutter. *J Vis* 7(2):17–17. <https://doi.org/10.1167/7.2.17>
- Scott DM, Lu W, Brown MJ (2021) Route choice of bike share users: leveraging GPS data to derive choice sets. *J Transp Geogr*. <https://doi.org/10.1016/j.jtrangeo.2020.102903>
- Senaratne H, Mobasheri A, Ali AL, Capineri C, Haklay MM (2017) A review of volunteered geographic information quality assessment methods. *Int J Geogr Inf Sci* 31(1):139–167. <https://doi.org/10.1080/13658816.2016.1189556>
- Shi X (2010) Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds. *Int J Geogr Inf Sci* 24(5):643–660. <https://doi.org/10.1080/13658810902950625>
- Shi W, Tong C, Zhang A, Wang B, Shi Z, Yao Y, Jia P (2021) An extended weight kernel density estimation model forecasts Covid-19 onset risk and identifies spatiotemporal variations of lockdown effects in china. *Commun Biol* 4(1):126. <https://doi.org/10.1038/s42003-021-01677-2>
- Silverman BW (1986) Density estimation for statistics and data analysis. Routledge <https://doi.org/10.1201/9781315140919>
- Slocum T, McMaster R, Kessler F, Howard H (2009) Thematic cartography and geovisualization. Pearson Prentice Hall, Hoboken
- State Capital Dresden (2021) Straßenknotennetz-Kanten [data file]. Office for Geodata and Cadastre. <https://opendata.dresden.de/informationsportal/#app/mainpage//Straßenknotennetz-Kanten>. Accessed 10 Mar 2022
- Stigmar H, Harrie L (2011) Evaluation of analytical measures of map legibility. *Cartogr J* 48(1):41–53. <https://doi.org/10.1179/1743277410Y.0000000002>
- Young BE, Dodge N, Hunt PD, Ormes M, Schlesinger MD, Shaw HY (2019) Using citizen science data to support conservation in environmental regulatory contexts. *Biol Cons* 237:57–62. <https://doi.org/10.1016/j.biocon.2019.06.016>
- Zhang G, Zhu AX (2018) The representativeness and spatial bias of volunteered geographic information: a review. *Ann GIS* 24(3):151–162. <https://doi.org/10.1080/19475683.2018.1501607>
- Zhu Z, Blanke U, Calatroni A, Tröster G (2013) Human activity recognition using social media data. In: Proceedings of the 12th international conference on mobile and ubiquitous multimedia. Association for Computing Machinery, New York. <https://doi.org/10.1145/2541831.2541852>
- Zhu M, Chen W, Xia J, Ma Y, Zhang Y, Luo Y, Liu L (2019) Location2vec: a situation-aware representation for visual exploration of urban locations. *IEEE Trans Intell Transp Syst* 20(10):3981–3990. <https://doi.org/10.1109/TITS.2019.2901117>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.