

# Where Would You Go on Your Next Vacation? A Framework for Visual Exploration of Attractive Places

Slava Kisilevich\*, Florian Mansmann†, Peter Bak‡, Daniel Keim§

Department of Computer Sciences and Information

University of Konstanz, Germany

{slaks\*,mansmann†,bak‡,keim§}@dbvis.inf.uni-konstanz.de

Alexander Tchaikin

Moscow, Russia

tchaikin@projekts.ru

**Abstract**—Tourists face a great challenge when they gather information about places they want to visit. Geographically tagged information in the form of Wikipedia pages, local tourist information pages, dedicated web sites and the massive amount of information provided by Google Earth is publicly available and commonly used. But the processing of this information involves a time consuming activity. Our goal is to make search for attractive places simpler for the common user and provide researchers with methods for exploration and analysis of attractive areas. We assume that an attractive place is characterized by large amounts of photos taken by many people. This paper presents a framework in which we demonstrate a systematic approach for visualization and exploration of attractive places as a zoomable information layer. The presented technique utilizes density-based clustering of image coordinates and smart color scaling to produce an interactive visualizations using Google Earth Mashup<sup>1</sup>. We show that our approach can be used as a basis for detailed analysis of attractive areas. In order to demonstrate our method, we use real-world geo-tagged photo data obtained from Flickr<sup>2</sup> and Panoramio<sup>3</sup> to construct interactive visualizations of virtually every region of interest in the world.

**Keywords**—Density maps; heat maps; density-based clustering; geovisualization

## I. INTRODUCTION

Travelers often face a great challenge when searching for places of their interests. Information about locations and attractions are widely available. Traditional sources like travel agents, tourist information centers, tour books and trip descriptions are often expensive and involve a time consuming activity. Electronic sources provide an overwhelming amount of information in an inexpensive manner, but their processing is still time consuming and requires a great effort to filter out irrelevant information. Mobile tourist guide systems can recommend a point-of-interest based on user's real time location [1]. But, what if we want to plan a weekend trip or a vacation? We have a lot of questions where to go, what to see, and where are the important places that we should visit first. If we plan a short trip on a weekend, we consider interesting

places in a city or a small region. If we plan a long vacation, usually we know which country we would like to visit and consider interesting places around the area where we stay. In both cases, our real interest is to find attractive places. In our work we assume that an attractive place is characterized by the concentration of a) the number of photos and b) the number of people who take photos there.

Several solutions recently proposed in [2][3][4], use geo-tagged images to reveal popular places in a city or world scale. However, fundamental methods are still missing to deal with interactive analysis and visualization of attractive places.

The principle question – how to make search for and exploration of attractive places easier – is the main motivation for the proposed method. To achieve this aim, we propose a framework for building interactive visualization of attractive places, which we integrate in Google Earth for easy access and distribution by a wide group of users.

One of the most popular and interactive solutions that can provide touristic information is Google Earth. In this environment users have quick access to the geographic data using specialized information layers, such as Panoramio Images, Wikipedia sites, POI databases (dining, commerce, etc.) and other relevant sources with geographic locations shared by the Google Earth user community. In fact, Google Earth has become a popular choice among researchers for (geo)visualization [5][6][7][8]. While a lot of relevant data is available in the layers of Google Earth, it is still very hard to find (and focus on) a attractive location, since there is no visual clue that would guide the user. To simplify the search for attractive places, we propose a method for building visualization maps such that the attractive place can be easily distinguished from other less attractive places. The visualization is provided in KML file format, so the user can overlay other geographic information available in Google Earth.

One of the main visualization techniques we use is density maps. The creation of density maps consists of several steps: 1) We measure the attractiveness of a place using density (concentration) of photos and the number of people who

<sup>1</sup><http://earth.google.com>

<sup>2</sup><http://www.flickr.com>

<sup>3</sup><http://panoramio.com>

took photos in that place. 2) We separate dense and sparse regions, by applying density-based clustering. Density-based clustering methods have many advantages over other clustering techniques. In particular, dense clusters are built using density measures and require only two parameters: the neighborhood radius and a minimum number of points in the neighborhood; clusters can have different shapes, which may reflect the natural bounds of an attractive place or a region of attraction; there is no need to define the number of clusters and the method's ability to handle outliers. The above mentioned advantages make density-based clustering especially applicable to the problem of finding concentrated areas, while areas with low concentration will be treated as noise and removed from further analysis. 3) Having a density cluster allows us to estimate the importance of every photo taken in it by applying the *Influence Weight* heuristic on every photo. The influence weight is based on *Kernel Density Estimation* and used as an indication of importance of an image with respect to other nearby images. The weight is then mapped to a color scale to build density maps. 4) For each cluster we select one or more images with high influence weights as representative for that area.

We would like to emphasize the key aspects of our approach and outline differences to existing solutions.

#### **Modeling:**

- Density-based clustering is used to outline areas of high photo concentration instead of grid-based approaches [2][3][9] or equally-sized radial clusters [4].
- Density-based clustering allows finding outliers (sparse areas), which can greatly reduce the amount of data, calculation time and allow the user to concentrate only on the attractive areas.
- We use the *Influence Weights* notion and calculate it for every image in a cluster. It allows us to estimate the importance value of every photo without complex context-based analysis.

#### **Visualization technique:**

- Density maps are built using true point visualization as opposed to interpolation techniques commonly used in heat map visualization [10][2][3][9][11].
- The visual size of the true point can be changed and used in interaction as a zoomable layer - larger point size can be used at country resolution, while smaller size can be used on city or street resolution, thus giving the user or analyst an opportunity not only to assess the level of importance of a place, but also to see where the photos were taken. The latter is not possible with interpolation techniques.

#### **Integrated system:**

- The proposed technique is integrated into a dedicated application with a fully controlled process ranging from a selection of the desired area of interest, clustering parameters to color mapping, scaling and visualization.

The contribution of our paper is a framework, which incorporates a new analysis model, a visual representation of

its outcome and an interface for detailed analysis of attractive areas.

The rest of this paper is structured as follows: Section II outlines related work. Section III presents our approach, which is then demonstrated in the case study in Section IV. The last section concludes the paper and suggests directions for future work.

## II. RELATED WORK

The Mean Shift algorithm [4] was used to find highly photographed areas. The authors applied the algorithm on photo data covering the whole world and presented the most photographed cities on Earth and most photographed landmarks in a city in a tabular form.

In [2] heat maps were used to visualize the concentration of tourists using coordinates of geo-tagged images. According to the authors, the studied area was divided into rectangular cells counting the number of photos and photo owners in each cell as a measure of concentration. The heat maps were produced using interpolation over every cell. This approach is also mentioned in consecutive articles [3][9]. Since the technique for building heat maps was not the purpose of their research, no explanations were given on the rationale of the selected technique and its parameters. The authors also state that heat maps succeeded in providing an overview of the tourist concentration but lacked on explanation of the quantitative meaning of the colors.

Fisher [12] proposes *Hotmap*, a mash-up system that visualizes the usage of a Microsoft's Live Search Maps by building heat maps over the number of downloaded tile images. The author also discusses how Hotmap can be used to reflect the *prominent* points as touristic places by analyzing the places the user was looking at while working with Live Search Maps. *Logarithmic color scaling* is proposed by the author to increase the variance in color and to differentiate between non-popular and popular places, that most people are looking at.

Kernel density estimation hotspot maps are used among other techniques by authors of [11] to map crime cases and to help police in crime investigation and analysis.

The heat maps techniques described above are based on interpolation of data between points of known values and commonly used when data represent smooth continuous phenomena [13]. Common interpolation methods are *triangulation*, *inverse-distance*, *kriging* [14] or kernel density estimation [11]. Gaussian filters [15] and Alpha Blending [16] are additional methods that are adopted from the image processing area. Gaussian filters are originally used for image or mesh smoothing, which make them appropriate for approximation of continuous phenomena. Many researchers suggested *alpha blending* or *alpha transparency* in the domain of visualizing large datasets. The method uses the alpha-transparency of the color system to represent data points. As a result, highly overplotted areas have high opacity and sparse areas have higher transparency.

Each of these methods has its own disadvantages and advantages and the selection of any of them is guided by

many considerations such as correctness of the estimated data, sensitivity to model specification, uncertainty handling, complexity of interpolation parameters, speed of execution, ease of understanding, etc. However, the common characteristic of these methods is interpolation, which is based on estimation of unknown values from known values to generate a continuous surface from discrete points. We, on the other hand, propose a model for finding attractive places using discrete point visualization. This approach makes it possible not only to give a visual clue of attractive places, but also to gain insight about exact hotspot places where the majority of people like to take photos. Additionally, it provides a quantitative measure of the attractiveness of a place.

### III. METHOD

The method for creating visualization maps is based on a series of steps, as shown in Figure 1.

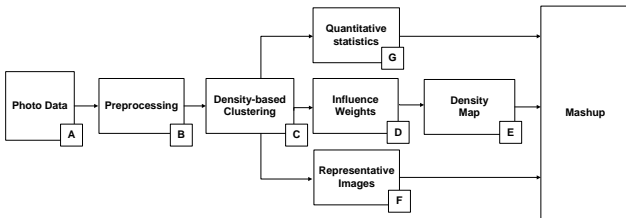


Fig. 1. Steps in creating visualization maps

#### A. Photo Data

We use geo-tagged photo information collected using APIs provided by Flickr and Panoramio. We downloaded the Flickr photo data in a similar way described in [4]. Panoramio API allows us to download photo data by using a precise geographic boundary. A PostgreSQL database was used to store the collected data because of its support for spatial queries.

#### B. Preprocessing

The first preprocessing step is to convert coordinates expressed in degrees into Universal Transverse Mercator (UTM) coordinate system in order to work with Euclidean distances. This step was performed during data collection, so it reduced the overhead of applying orthodromic distances during distance calculation. The second preprocessing step was to reduce all photos taken by the same user having the same coordinates, leaving only a single photo. This was done to reduce the computation time during weight calculation, since more than one photo taken by the same user having the same coordinates do not contribute to the overall attractiveness of the place.

#### C. Density-based Clustering

As it was mentioned in Section I, the features of density-based clustering are as following:

- It usually requires only two input parameters: the minimum radius of the neighborhood  $\epsilon$  and the minimum number of points inside a neighborhood  $MinPts$ . Regions with high density are connected into density clusters. The intuition is that for any point in a cluster, the local point density around that point has to exceed some threshold.
- There is no need to preset the number of clusters. The number of clusters is determined by using only two parameters as described above.
- The clusters can be of arbitrary shapes. Unlike grid-based clustering, where one needs to define rectangular grids, density clusters can be of any shape and this only depends on the density of the regions.
- It can handle noise. Points not connected into clusters are considered outliers.

The above mentioned features make the use of density-based clustering in our problem very intuitive: we want to find only dense areas and disregard photos which were taken in sparse areas. There is one question that has to be answered: how do we determine initial parameters  $\epsilon$  and  $MinPts$ . Several suggestions can be found in the literature [17][18]. However, our experiments indicate that  $\epsilon$  of 100 meters and  $MinPts$  between 5 to 20 points is an adequate choice for finding dense regions on a city level.

#### D. Influence Weights

In the previous step we obtained clusters which represent dense regions. However, density is different between clusters and different parts of a cluster also have different densities. In this step we obtain weights of every photo point in a cluster contributed by points in a neighborhood. The intuition behind the influence weight is the following: the weight of a point will be high if a lot of people took photos nearby that point, otherwise it will be low. The influence weights can be used to draw conclusions about attractiveness of a region by mapping the values to colors and building density maps as described in step. III-E or to select representative images from a cluster (see step. III-F).

The definition of the influence weight function is described in Eq. 1. Let  $F$  be the set of images where every image  $p \in F$  is described by its coordinates  $\mathbf{x}_p \in \mathbb{R}^2$  and its owner  $O_p$ . Then, the influence weight function for a photo point  $p$  in cluster  $C$  is defined as follows:

$$f(p) = \sum_{\substack{l \in C \\ O_l \neq O_p}} K(\mathbf{x}_p, \mathbf{x}_l) \quad (1)$$

where  $K$  is a kernel function. Square wave, parabolic or Gaussian kernels are examples of candidate kernels. The influence weight for a photo point  $p$  is calculated as a sum of kernel functions between point  $p$  and all other points in a cluster whose owners are not equal to the owner of the point  $p$ .

#### E. Density Map

For creating the density map, we draw one dot for every photo at its geographic coordinates using its weight for color-

coding. According to the zoom level of the map, different sizes of the points are used. For example, when displaying information at country level, we need significantly larger points in order to still see the respective attractive areas, whereas a detailed street level view needs finer grained points.

An important aspect of building the density maps is color mapping. The design of effective color schemes that respond to the needs of the user is a non-trivial task [19]. Many aspects have to be taken into account such as: color-blindness, print reproduction, task dependency. Some suggestions can be found in various literature [20][13][19][21]. We developed our own color mapping tool [22] (see Figure 2) that extends some existing tools like [23][24] with the following features:

- RGB and HSL model with complete adjustability.
- Arbitrary number of color pivots.
- Color ramping is performed between every color pivot.
- Every color pivot can offset manually to the left or right of the color scale, which allows to decrease or increase the range of weights mapped to a specific color (see Figure 2).
- Flexibility to adapt to the data distribution.

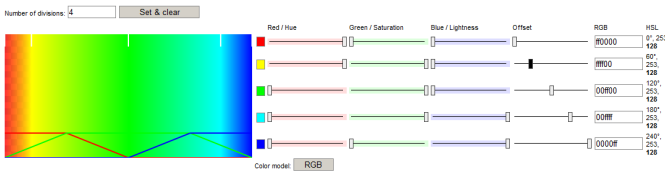


Fig. 2. Example of color distribution using the Gradient Tool. Red and Blue occupies 10%, Yellow and Cyan - 30% and Green - 20% of the color scale.

### F. Representative Images

Photos with high influence weights are not necessarily those that have attractive content, but were taken in a dense and thus, attractive area. We call such photos *representative photos*. We can use such photos to illustrate what was photographed in a specific area. The clusters produced by the density-based clustering algorithms and described in step III-C are used for finding dense areas. Such areas, as described above, can have different sizes. The sizes depends on the attractiveness of an area and on the parameters that are chosen for building clusters. It is possible that one density cluster includes many different points of interest. Thus, it is evident that there can be a lot of *representatives* in a cluster. In such case we split a cluster into a number of *local* clusters, where the representative image is retrieved from a local cluster. We propose two ways of achieving this:

(a) Application of Mean Shift [25], a non-parametric feature space clustering algorithm. This algorithm is usually used in image processing, but was recently used in [4] to find highly-photographed places. Mean Shift is a non-parametric technique for estimating the modes of an underlying probability distribution. Dense regions in the feature space correspond to local maximum of the density probability function [26]. The

advantages of the Mean Shift algorithm are: (1) no assumption about the cluster shapes, (2) arbitrary feature spaces and (3) one control parameter: radius of the cluster. Once local clusters are obtained,  $n$  representative photos can be selected from every cluster in a decreasing order of their influence weights. (b) Influence weights can be considered as local maxima of the data distribution so we can apply a naive approach to obtain local clusters in the following way: (1) Sort the photo points according to their influence weights in descending order. (2) Select the first point  $p_1$ , which will be the center point of the first cluster (3) Apply spatial query and find all the images whose distance from  $p_1$  is less than radius  $R$ . (4) Ignore points that were already processed (5) Continue with the first unprocessed point with the highest influence weight.

### G. Quantitative Statistics

To enrich the reasoning about attractive places we can provide additional information to the user such as the shape of the clusters, number of photos taken in the cluster and number of different users who took photos. The shape of the cluster as a convex hull can be easily produced using PostgreSQL spatial queries. In addition to this, the database provides functionalities to inquire the area of a cluster and other geographic statistics.

## IV. CASE STUDY

As a case study, we present Saint Martin, a tropical island in the northeast Caribbean. In a series of density maps<sup>4</sup> presented in Figure 3 we show the feasibility and tasks of analysis and exploration of attractive places with our methods; namely we can (a) observe all attractive places as a whole, (b, c) zoom into a smaller region and investigate the place of interest in more details, (d) investigate boundaries of attractive places, (e) locate what is typically being photographed there using representative photos, and (f) obtain some statistical information about the area.

Density clusters were obtained by the GDBScan algorithm [27] with a neighborhood radius  $\epsilon$  of 100 meter and a minimum number of 10 photos in a neighborhood. 2380 photos were left in 40 clusters after identifying 2009 photos as outliers. The Gaussian density function was used for calculating influence weights.

$$K(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2R^2}}$$

where  $R$  is a radius of a neighborhood and  $\|\cdot\|$  is the standard euclidean norm. The radius of local clusters was preset to 100 meters. Local clusters with less than 6 users were not visualized. One representative photo was obtained in each local cluster.

Figure 3(a) presents an overview of attractive places of Saint Martin island at high resolution, such that the whole island is seen on a map. All the subfigures use the same colormap created using the tool described in Section III-E. Let

<sup>4</sup>The Google Earth map can be downloaded from <http://infovis.uni-konstanz.de/kml/StMaarten.kmz>

us concentrate on Mahot beach which is very popular among island visitors because of its vicinity to the landing runway. Figure 3(b) shows the area of Mahot beach. Figure 3(c) shows the detailed view of Mahot beach. It can be seen that the points are smaller in close zooming than on Figure 3(b), which allows the user to see exact places where people took photos. The yellow area is highly concentrated, so it is covered completely even by small points. Boundaries of the concentrated areas are shown in Figure 3(d) as convex hull. The yellow lines outline the cluster obtained by density-based clustering as described in step III-C, while the green lines outline several local clusters that were produced as described in step III-F. Figure 3(e) shows a representative image of one of the highly photographed local clusters. Quantitative statistics (the number of photos and owners in a cluster) are presented in Figure 3(f).

## V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a framework to visualize and explore attractive places using density maps. The underlying model is based on a visualization of every geo-tagged image coordinate. This approach is opposed to techniques commonly used in GIS where a visualized region is created by interpolation. We introduced the notion of attractive places and showed how we can find such regions using density-based spatial clustering. We adopted the definition of influence and density function from the domain of non-parametric clustering and show how they can be reused for implicitly obtaining weights of photo locations. This allowed us to estimate the characteristic of the attractive places by retrieving photos with high influence weights without content-based analysis. We believe that our approach can be of great importance to travelers by reducing search time of attractive places, to providers of tourist services or researchers who analyze spatial events.

In future research, we will consider additional issues, such as (a) analysis of attractive places as a function of time and (b) embedding additional data sources, like Wikipedia geo-tagged web pages, in a process of density map creation in a way similar to creation of digital touristic maps.

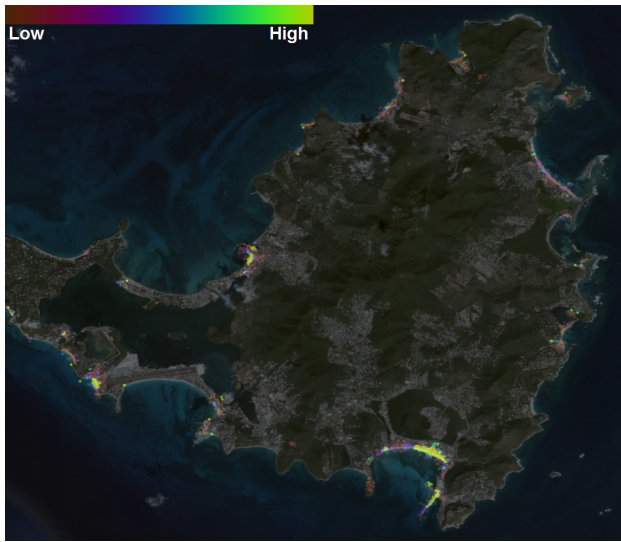
## ACKNOWLEDGEMENTS

This work was partially funded by the German Research Society (DFG) under grant GK-1042 (Research Training Group “Explorative Analysis and Visualization of Large Information Spaces”), and by the Priority Program (SPP) 1335 (“Visual Spatio-temporal Pattern Analysis of Movement and Event Data”).

## REFERENCES

- [1] J. Pavón, J. M. Corchado, J. J. Gómez-Sanz, and L. F. C. Ossa, “Mobile tourist guide services with software agents,” in *MATA*, 2004, pp. 322–330.
- [2] F. Girardin, F. D. Fiore, J. Blat, and C. Ratti, “Understanding of tourist dynamics from explicitly disclosed location information,” in *4th International Symposium on LBS and Telecartography, Hong-Kong, China*, 2007.
- [3] F. Girardin, F. D. Fiore, C. Ratti, and J. Blat, “Leveraging explicitly disclosed location information to understand tourist dynamics: a case study,” *Journal of Location Based Services*, vol. 2, no. 1, pp. 41–56, 2008.
- [4] D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, “Mapping the world’s photos,” in *WWW*, 2009.
- [5] T. Smith and V. Lakshmanan, “Utilizing Google Earth as a GIS platform for weather applications,” in *22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, 2005.
- [6] A. Slingsby, J. Dykes, J. Wood, and K. Clarke, “Interactive tag maps and tag clouds for the multiscale exploration of large spatio-temporal datasets,” in *IV ’07: Proceedings of the 11th International Conference Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 497–504.
- [7] J. Wood, J. Dykes, A. Slingsby, and K. Clarke, “Interactive Visual Exploration of a Large Spatio-temporal Dataset: Reflections on a Geovisualization Mashup,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1176–1183, 2007.
- [8] A. Slingsby, J. Dykes, J. Wood, M. Foote, and M. Blom, “The Visual Exploration of Insurance Data in Google Earth,” *Proceedings of Geographical Information Systems Research UK (GISRUK)(Manchester, UK, 2007)*, pp. 24–32, 2008.
- [9] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat, “Digital footprinting: Uncovering tourists with user-generated content,” *Pervasive Computing, IEEE*, vol. 7, no. 4, pp. 36–43, 2008.
- [10] R. van Liere and W. de Leeuw, “Graphsplatting: Visualizing graphs as continuous fields,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, no. 2, pp. 206–212, 2003.
- [11] S. Chainey and L. Tompson, *Crime Mapping Case Studies. Practice and Research*. Wiley, 2008.
- [12] D. Fisher, “Hotmap: Looking at geographic attention,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1184–1191, 2007.
- [13] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. Howard, *Thematic Cartography and Geovisualization*. Prentice Hall, 2008.
- [14] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics, 1999.
- [15] W. M. Wells, “Efficient synthesis of gaussian filters by cascaded uniform filters,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 2, pp. 234–239, March 1986.
- [16] A. Unwin, M. Theus, and H. Hofmann, *Graphics of Large Datasets: Visualizing a Million Series, In: Statistics and Computing*. Springer, 2006.
- [17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” *Data Mining and Knowledge Discovery*, pp. 226–231, 1996.
- [18] A. Hinneburg and D. A. Keim, “An efficient approach to clustering in large multimedia databases with noise,” *Knowledge Discovery and Data Mining*, vol. 5865, 1998.
- [19] P. B. Adam Light, “The end of the rainbow? color schemes for improved data graphics,” *Eos*, vol. 85, no. 40, 2004.
- [20] M. Harrower and C. Brewer, “Colorbrewer.org: an online tool for selecting colour schemes for maps,” *Cartographic Journal*, vol. 40, no. 1, pp. 27–37, 2003.
- [21] B. E. Rogowitz and L. A. Treinish, “Why should engineers and scientists be worried about color?” <http://www.research.ibm.com/people/l/lloyd/color/color.HTM>.
- [22] A. Tchaikin, “Gradientcreator tool,” 2009, <http://infovis.uni-konstanz.de/tools/gradientcreator/>.
- [23] C. Brewer, “Colorbrewer,” <http://colorbrewer2.org/>.
- [24] “Colormap-tool,” <http://infovis.uni-konstanz.de/tools/colormap/>.
- [25] K. Fukunaga and L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition,” *Information Theory, IEEE Transactions on*, vol. 21, no. 1, pp. 32–40, 1975.
- [26] D. Comaniciu, P. Meer, and S. Member, “Mean shift: A robust approach toward feature space analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603–619, 2002.
- [27] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, “Density-based clustering in spatial databases: The algorithm gdbcscan and its applications,” *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 169–194, 1998.

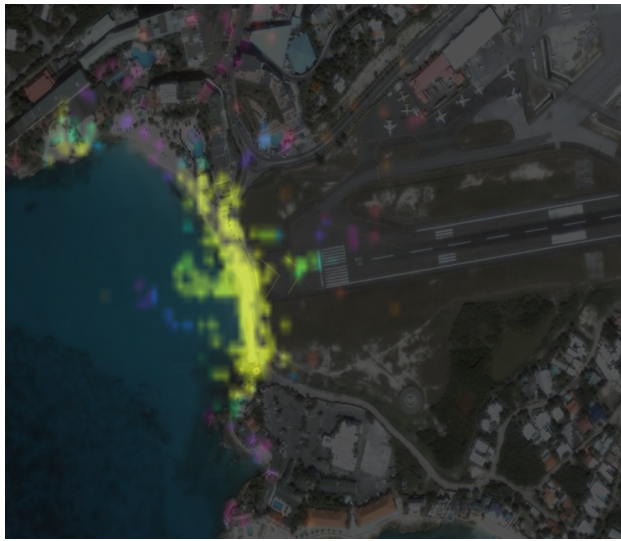




(a) Attractive areas of the Island



(b) Zooming into the Mahot beach area.



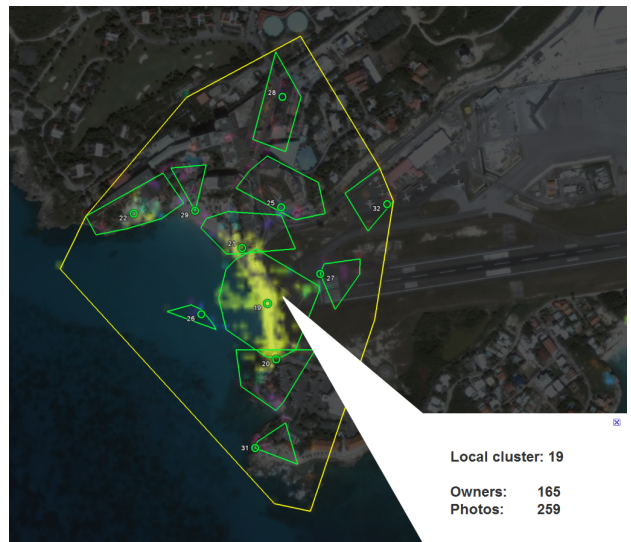
(c) Detailed view of the Mahot beach area.



(d) Boundaries of attractive areas. Yellow: attractive area as one dense cluster. Green: local clusters.



(e) Representative image.



(f) Quantitative Statistics. Number of photos taken in a cluster and number of different people who photographed there.

Fig. 3. Saint Martin island case study.