Uncovering Locally Characterizing Regions within Geotagged Data

Bart Thomee
Yahoo! Research
Avinguda Diagonal 177
08018 Barcelona, Spain
bthomee@yahoo-inc.com

Adam Rae
brandcrumb
Carrer de València 352
08009 Barcelona, Spain
adam.rae@brandcrumb.com

ABSTRACT

We propose a novel algorithm for uncovering the colloquial boundaries of locally characterizing regions present in collections of labeled geospatial data. We address the problem by first modeling the data using scale-space theory, allowing us to represent it simultaneously across different scales as a family of increasingly smoothed density distributions. We then derive region boundaries by applying localized label weighting and image processing techniques to the scale-space representation of each label. Important insights into the data can be acquired by visualizing the shape and size of the resulting boundaries for each label at multiple scales. We demonstrate our technique operating at scale by discovering the boundaries of the most geospatially salient tags associated with a large collection of georeferenced photos from Flickr and compare our characterizing regions that emerge from the data with those produced by a recent technique from the research literature.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Spatial databases and GIS*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*

General Terms

Algorithms, Theory, Measurement

Keywords

Geotagged data, spatial analysis, region discovery, scale-space theory, Flickr

1. INTRODUCTION

When considering characteristic geographic areas around the world, there are those that are canonically described in gazetteers (such as cities and states) and there are those that are not, such as the Sahara desert in Africa, the Bordeaux wine region in France and the red light district in Amsterdam. Nonetheless, geographic regions such as the latter are colloquially well-known and there are many others for which official boundaries do not exist or differ from their colloquial boundaries as determined by the consensus of the general populace.

In this paper we describe a novel algorithm that automatically uncovers the colloquial boundaries of locally characterizing regions present in collections of labeled geo-referenced data. Our technique uses scale-space theory [47, 27] to detect those geographic areas for which the occurrence of a label is particularly prominent with respect to the other labels occurring in the same area, at a particular scale. Naturally, the set of scales at which characteristic regions can be found for a certain label depends on the nature of the underlying data and the spatial distributions of the labels. The scale-space paradigm is therefore especially suitable, because it represents the data simultaneously across different scales as a family of increasingly smoothed density distributions, allowing us to automatically derive spatially coherent regions by applying image analysis techniques to the scale-space representations of the data for each label and at each scale in a consistent manner, where the regions are generated by applying the exact same sequence of steps at each scale.

In the past, geo-referenced collections have been predominantly analyzed for the extraction of spatial knowledge for a limited number of pre-defined labels or queries, as well as for specific geographic areas. For instance, in the context of online tagged photos – the scenario we primarily focus on in this paper – regions are often specifically restricted to points of interest such as landmarks [7, 34] or they are defined as dense areas of photos from which the most salient tags are discovered [1, 9]. The analyses are further typically performed at a single fixed scale, ignoring the effects of scale on shape and size of the regions, even though intuitively the scale at which any region discovery is performed has a large influence on the final outcome, i.e. coherent regions at a particular scale may not necessarily be coherent at other scales.

To summarize, the main contributions of our work are the following:

- An effective framework for automatically and efficiently discovering the colloquial boundaries of locally characterizing regions present within large collections of labeled geotagged data at multiple scales using scale-space theory and image processing techniques.

- Insights into the effects of scale on the shape, size and quantity of the derived regions and a comparison of the generated regions with those produced by a recent technique from the research literature using a multi-million sample of geo-referenced photos.

Our technique can be applied to any collection of labeled geographic data and is thus not restricted to any particular domain. The regions our technique uncovers can be of

benefit to many tasks, for instance monitoring the spread of diseases or discovering trending news topics across the world over time at global and local scales. In the context of multimedia our technique is useful for topics such as tag recommendation and disambiguation [46], content tagging and placing [15, 26], and region summarization [39, 48].

The remainder of this paper is organized as follows. We first discuss related work in Section 2 before presenting our algorithm for boundary discovery using scale-space theory in Section 3. In Section 4 we propose optimizations to handle large-scale data and illustrate our technique by discovering and visualizing the detected tag regions within Flickr photos. We follow with an analysis of the properties of our algorithm in Section 5. We conclude the paper with final remarks and future outlooks in Section 6.

2. RELATED WORK

Discovering geographic regions within georeferenced data has been an active research topic for many years. A variety of methods can be applied to uncover the hidden spatial structures, most notably techniques such as clustering [10], density estimation [17, 21] and neural networks [16]. For example, Hargrove et al. [13] used multivariate clustering to uncover distinct ecoregions based on environmental conditions. In recent years, georeferenced multimedia collections have considerably grown in size, in particular due to the availability of digital cameras with built-in GPS receivers that can automatically attach the geographic location to every photo that is taken. We can therefore find a considerable number of methods in the research literature that try to represent or summarize geographical regions in terms of either the tags associated with the photos in that region [1, 9] or by the photos themselves [39, 7, 23, 19, 48, 8]. For example, language modeling approaches [40, 44, 14] characterize cells that partition the world into discrete units for automatic photo and video geo-localization.

However, the notion of scale is typically not naturally integrated into these techniques, requiring tuning of parameters for each different level of detail at which the data is analyzed. The analyses in the literature are furthermore commonly restricted to a limited set of geographic areas, while the visualizations are restricted to modifying objects of known shape and size by increasing details at small scales and reducing details at large scales. Yet, especially in a geographic context, scale is inherently part of the labeled data and thus should form an integral part of the data modeling, analysis and visualization process. Scale-space theory [47, 27] provides a framework for analyzing the structure of multi-dimensional signals at multiple scales, which has been successfully used within many different fields of research. In the field of computer vision performing analysis at multiple scales benefits many tasks, such as feature detection [32, 5] and object detection [30], while in a geographic context the scale-space framework has been used for analyzing [28, 45] and for visualizing [4, 2] data collections.

In our work, we apply the concept of scale-space theory to labeled data in order to automatically infer regions that are locally characterizing across any area at any scale. While the notion of scale was used in [36, 24] for the detection of photo tags that exhibit significant invariant spatial patterns, we in contrast present a more general approach that focuses on defining the boundaries of the characteristic regions for any label, whether or not it demonstrates spatial

invariance. We furthermore do not assume beforehand that labels are related, unlike the authors of [9], who assume that when they are used in close proximity to each other they are therefore semantically related. The most closely related techniques to ours assign localized weights to labels to derive the boundaries of their associated geospatial regions. Nonetheless, these have as drawbacks that the discovered regions can only be elliptical of shape [3], that they have only been applied at very limited scales and geographic areas [22], or that they generate an overwhelming amount of regions [18].

3. ALGORITHM

In our approach we do not focus on the items in a data collection, but instead only focus on the individual instances of their metadata labels (e.g. tags in the case of online photos). We aim to directly discover coherent regions of each label by analyzing their spatial distributions, one at a time. Our technique is thus unlike many of the existing techniques, such as those discussed in the related work section, which first try to discover coherent regions by analyzing how the data is geographically distributed and only afterwards attempt to find the label(s) that most accurately describe each of these regions. Our region discovery approach consists of four main steps, namely:

1. **Data labeling** – Each instance in the collection of geographic data is assigned one or more labels for which ultimately the region boundaries are to be discovered.

2. **Scale-space representation** – The labeled data is represented in a scale-space as a family of increasingly smoothed two-dimensional density distributions.

3. **Region generation** – Two-dimensional regions are generated at multiple scales for each label separately.

4. **Region selection** – The generated regions are validated in order to discard those that are not appropriate at particular scales or do not meet criteria suited to the envisioned use case.

In the following sections we detail each of the aforementioned steps in the algorithm.

3.1 Data labeling

The labels for which we aim to derive their characteristic regions can be simple, e.g. words present in a georeferenced document, or more complex, e.g. visual concepts detected in a georeferenced image. If the data collection is unlabeled, each instance in the collection first needs to be associated with one or more labels before our algorithm can be applied, where the set of labels may be manually specified by a user beforehand or automatically derived from the data. The labeling of an instance is typically done by extracting suitable features and then performing feature analysis and classification to obtain its labels. Whenever an instance or label does not meet pre-defined quality criteria it can be removed from the dataset. To illustrate, a textual document could be considered as an instance and the words it contains as labels, where those labels that are stop words may be discarded.

Given a set of labels Λ, we can define the data collection as $\mathbb{D}_\Lambda = \biguplus \mathbb{D}_\lambda | \lambda \in \Lambda$, where the $\biguplus$ operator performs a union of the data items associated with each label $\mathbb{D}_\lambda$. A data item that has multiple labels will therefore be included as many times in $\mathbb{D}_\Lambda$ as it has labels. We further define $\mathbb{D}_\lambda \triangleq \{d\}$,

where d has label λ and is represented by a tuple (l_x, l_y), which contains a geographic location that is expressed by longitude l_x and latitude l_y.

3.2 Scale-space representation

Scale-space theory is predominantly used in image analysis, where an image is represented as a family of images obtained by convolving it using a smoothing kernel parameterized with a scaling factor. Scale-spaces are particularly useful for isolating regions of pixels that exhibit segmentation at certain scales by comparing subsequent increments of smoothing, a notion that techniques such as SIFT [32] and SURF [5] have successfully used for the purpose of extracting scale-invariant features from images. We exploit this property of the scale-space representation in order to uncover regions of sizes appropriate at a given scale. In this work we define a *region* as a closed area on our planet that is bounded by a non-self-intersection polygon, which we refer to as its *contour*.

In order to apply scale-space theory to our data we consider a discretized histogram of label counts to be analogous to a grayscale image. We first represent the data for each label $\mathbb{D}_\lambda$ as a two-dimensional density histogram $f_\lambda(x, y)$, such that

$$|\mathbb{D}_\lambda| = \sum_{x=1}^{w} \sum_{y=1}^{h} f_\lambda(x, y) \tag{1}$$

where w and h are the number of bins along the width and height of the histogram. Even though geographic locations can be represented as a bounded function of continuous longitude and latitude values, they are typically expressed in the discrete domain. In particular, one of the most precise ways of obtaining a geographic location is through the use of digital global positioning systems, which are discrete by nature and yield locations with an accuracy usually in the order of meters, depending on equipment and environmental conditions. We thus represent our data as a two-dimensional histogram along longitude and latitude using finite addressability, where the exact resolution of the histogram depends on the precision of the underlying data as well as on the envisioned use case. For example, for geographic coordinates expressed with up to two decimal places of precision in longitude and latitude, over the whole world the histogram would be of dimensions $w = 36,000$ and $h = 18,000$, ensuring that no cell would be larger than the equivalent of $0.01°$ longitude at the equator, which approximately measures 1km.

Using this data definition, we then define the linear scale-space representation of a family of increasingly smoothed histograms $L_\lambda(x, y; t)$ for a given label λ as a convolution of $f_\lambda(x, y)$ with Gaussian kernels $G(x, y; t)$, as expressed by

$$L_\lambda(x, y; t) = G(x, y; t) * f_\lambda(x, y) \tag{2}$$

where t denotes the variance of the kernel and thus $\sigma = \sqrt{t}$ its standard deviation. At $t = 0$ we obtain the original density distribution of the spatial data for a certain label, whereas for increasing values of t the data is convolved with a kernel of increasing width, so that the density distribution is increasingly smoothed and as a result more and more high frequency details are removed. We exploit this property in the next step to discover the boundaries of the various regions where a label is used across the world at individual scales.

3.3 Region generation

For a given label λ, its scale-space representation L_λ, and a certain scale factor t, we aim to extract a set of locally characterizing regions. These characteristic regions are geographic areas where the occurrence of a particular label is particularly prominent with respect to the other labels occurring in the same area, at the specified scale.

Label prominence

To account for the prominence of a label for a particular point in space at a particular scale, we weight $L_\lambda(x, y; t)$ in proportion to how much it contributes to the overall data density $L_\Lambda(x, y; t)$, such that

$$L'_\lambda(x, y; t) = \frac{(L_\lambda(x, y; t))^2}{L_\Lambda(x, y; t)} \tag{3}$$

Spatial activity isolation

From the resulting histogram we then produce a binary representation that isolates the high frequency spatial activity found in $L_\lambda(x, y; t)$ by removing the lower frequency activity represented in a histogram smoothed at a higher value of t, according to

$$b(x, y; t) = \begin{cases} 1 & \text{if } (L'_\lambda(x, y; t) - L'_\lambda(x, y; t + a) > \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where a is a constant and ϵ is a small threshold applied to the magnitude of the response to suppress noise. When a approaches zero the Difference of Gaussians operation above closely approximates the Laplacian of Gaussian, effectively detecting edges between two areas of uniform but different intensities in the weighted histogram.

By considering t as a continuous variable, each label generates closed contours across one or more ranges of scales [31], provided it is not considered as noise and filtered out. Across scales contours can be formed, cease to exist, merge with other contours or split into multiple contours. By fixing the scale factor t contours are obtained that are appropriate for that scale. We illustrate this in Figure 1.
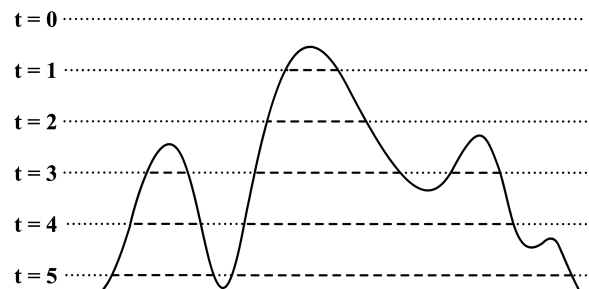

Figure 1: The contours for a certain label across a continuous set of scales undulate, forming peaks and valleys. For visualization purposes we show a two-dimensional view of the three-dimensional plot $b(x, y; t)$ by fixing y. Here, the scale dimension t runs from top to bottom, the spatial dimension x runs from left to right and the spatial dimension y would run from front to back.

At this point we have obtained disjoint contours representing the area where a label is particularly prominent, although depending on the underlying data distribution these may be incoherent. In order to turn this into a set of dis-

Figure 2: Region generation given a binary representation of the label `spain`. We first fill all outer contours (left) and then smooth and connect disjoint areas by applying a morphological closing operation that first dilates (middle-left) and then erodes (middle-right) the binary representation, after which the contours are traced (right) to extract the characteristic regions. Here, the label was also prominently used on the Balearic islands (an archipelago of Spain), Melilla (an exclave of Spain) and in a small area in France, causing regions to emerge there as well.

tinct, contiguous regions, we apply three image processing techniques to the binary representation of a label.

Contour filling

We observe that the Difference of Gaussians operation may yield more than one concentric contour for a label for each crossing of the threshold ϵ. To emphasize the full spatial extent of a label rather than its internal density, we apply contour tracing [6] to the binary representation to extract the outermost contours that mark the areas where the label's influence drops below the threshold for the last time. We fill the entire area enclosed by the outer contours, implicitly discarding any inner contours.

Morphological closing

To smooth the contours and to connect disjoint regions that are separated from each other by a small gap, we apply morphological closing to the binary representation after it has been filled. Morphological closing first performs a dilation step followed by an erosion step, where in both steps the same disk-shaped structuring element S is applied to the binary representation, such that

$$b(x, y; t) \bullet S = (b(x, y; t) \oplus S) \ominus S \tag{5}$$

where the radius ρ of the structuring element can be adjusted to enforce more or less smoothing of a region's contour. The value to use for ρ depends on the envisioned use case, where the parameter can be manually explored or automatically optimized using a particular objective function, e.g. see Section 5.1.

Contour tracing

We again apply contour tracing to the binary representation of each label to extract the outer contours of the locally characterizing regions that have emerged.

We illustrate the image processing operations in Figure 2. Note that these techniques form just one of many ways of generating regions from the binary representation. The contour filling operation could for instance be substituted by applying flood filling that would skip any inner (nested) contours, while the morphological closing operation could be replaced by applying dilate and erode distance transforms [11] to achieve a similar effect. If necessary, the resulting polygon description could even be further simplified, e.g. through iterative end-point fit algorithms [35]. The main motivation for using the image processing techniques we described earlier is because they are mature, well-understood and because efficient implementations exist.

3.4 Region selection

Each binary representation for a label λ at scale t may result in multiple regions emerging at a particular scale. These regions can still vary substantially in their possible sizes, ranging from very small to very large, where their size is primarily determined by the spatial prominence of the label and the magnitude of the signal that remains after the Difference of Gaussians bandpass filter has been applied. Furthermore, the total number of regions generated for all of the labels combined may be too large for effective use. Depending on the envisioned use case further region selection may therefore be necessary. For instance, it may be appropriate to apply a ranking function to only pick a subset of the regions to be shown; the function could be parameterized with characteristics of the regions, such as area size or data density, or alternatively could include constraints to limit the maximal permissible amount of overlap between regions in order to select a disjoint or only partially overlapping set of regions to reduce visual clutter.

3.5 Parameterization

Our technique requires the specification of several parameter values that are, to a certain extent, dependent on the envisioned use case as well as the data itself. For analysis and visualization purposes, one usually focuses on only a limited number of scale levels $s \in \mathbb{N}$ at which the contours are obtained. The distance between consecutive scale levels is not necessarily linear in terms of the scale factor t but in online map-based interfaces, such as Yahoo! Maps, it is often roughly quadrupled so that $t = 4^s$. In addition, the distance between the scale-space representations to which the Difference of Gaussians operation is applied is controlled by the parameter a, which does not necessarily cover the entire range between consecutive scale levels. We illustrate these notions in Figure 3. By adjusting how t is obtained from s, as well as suitably setting a, the amount of smoothing between consecutive scales can be optimized to emphasize certain frequency ranges within the underlying data at certain scales. Finally, the noise threshold ϵ and the default radius ρ of the disk-shaped structuring element used by the morphological closing operation both affect the number, shape and size of the regions that emerge from the data.

If ground truth regions are available it is straightforward to derive the (near-)optimal parameter values by exhaustively comparing the obtained regions with the ground truth across a range of parameters. Multi-parameter learning algorithms [38, 42] can be applied to more efficiently direct the search and thus speed up convergence towards optimality.
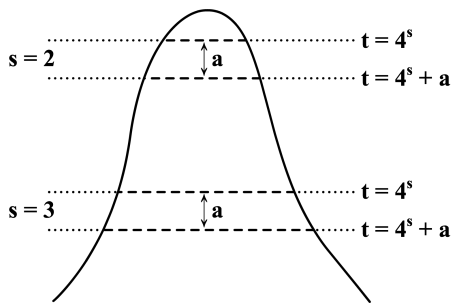
Figure 3: Illustration of the relationship between scale level s, scale factor t and Difference of Gaussians distance a. At scale level $s = 2$ the associated scale factor is $t = 16$, whereas at scale level $s = 3$ the scale factor is $t = 64$. At both locations the underlying density histogram is smoothed using a kernel parameterized by t and a kernel parameterized by $t+a$ after which the difference between them is taken.

3.6 Complexity

In contrast with techniques that process individual labeled instances, such as clustering-based approaches, our technique processes aggregate label counts, which for large quantities of data is much more computationally efficient, since the computational complexity does not increase with the size of the collection once the data has been represented as a histogram. Our approach is furthermore inherently parallelizable. Nonetheless, examining millions of labeled instances in order to assign them to the correct cells within the histograms can still be a time-consuming process – an issue of importance for highly dynamic online datasets – although in preliminary experimentation we observed that for our particular labeled collection the amount of computation could be reduced by subsampling the data without substantially compromising the shape and size of the final regions; however this may not necessarily be the case for other datasets.

4. IMPLEMENTATION

To demonstrate our region discovery technique we envisage a scenario prompted by users wanting to explore and browse large-scale geographically annotated media collections. This kind of scenario is particularly important to address, considering exploration has been shown to be the predominant mode of interaction for users of media systems [20] and exploration and browsing systems that allow the user to gain insight into and support the exploration of media collections are therefore needed [29]. In this section we present details on the implementation of our technique in the context of such a system, where we uncover the regions present within a large collection of labeled georeferenced photographs taken from the online image sharing website Flickr[1]. A preliminary version of our photo exploration and browsing system was presented in earlier work [43]. We emphasize, however, that our region discovery technique can be applied to any kind of labeled dataset of similar nature and not just to visual media; whether (meaningful) regions emerge entirely depends on the characteristics of the dataset. The optimizations we outline below for representing large-scale datasets are generalizable and not specific to our particular dataset.

---

[1] http://www.flickr.com/

4.1 Dataset

Our Flickr dataset contains a sample of over 56 million geo-referenced images uploaded to Flickr before the end of 2010. Each photo is represented by a geographic location, indicated by longitude and latitude, and one or more tags assigned to the photo by the uploading user. We consider each instance of a tag associated with a photo as a label and annotate each label with the location information of its source photo. A single photo may thus generate multiple instances of annotated labels. We perform sanitization on the data by removing all non-latin characters, reducing all remaining latin characters to their lowercase representation and removing all diacritics, so that tags like `España` and `Gaudí` become `espana` and `gaudi` respectively. We remove tags that refer to years, such as `2006` and `2007`, as well as camera manufacturer names, such as `canon` and `nikon`, because these at times get automatically added by capture devices or photo applications and thus are not representative for the tagging behavior of users. We additionally remove infrequently used tags by ensuring we have at least 50 instances per tag occurring around the world; while infrequently used tags could certainly yield a locally characterizing region, it was shown in [41] that the tag frequency distribution in photos uploaded to Flickr follows a power law with the very long tail containing the infrequent tags that typically were categorized as incidentally occurring words, making it unlikely such tags will actually generate regions.

4.2 Representation

We aim to compute the regions for all remaining tags in the dataset at several different scales, ranging from street level to world level. At the lowest scale level (i.e. zoomed in) the finest details are captured from the highest frequency components present within the normalized data, while at higher scale levels (i.e. zoomed out) the coarser structure is uncovered from the lower frequency components. This also corresponds well with the paradigm used by typical online systems that show interactive map-based views of the world, where the act of zooming in increases the spatial resolution to reveal more detail, whereas zooming out decreases the spatial resolution and reduces the level of detail.

In order to capture the finest details at the lowest scale the underlying scale-space representation of each tag needs to be of sufficiently high spatial resolution, whereas at higher scales such high resolution is not as necessary; provided that the resolution of the representation is at least high enough to capture the details at each scale level and the Gaussian kernel is correctly adjusted to retain the correct amount of smoothing, our scale-space region derivation technique generates equivalent regions at a reduced resolution at lower computational cost.

We therefore express the difference between two consecutive scales by changing the spatial resolution of the scale-space representation by a factor of two along each dimension, where incrementing the scale level s quarters the spatial resolution u, while decrementing s quadruples it, which also approximates the effects of zooming in and out on the spatial resolution in typical online map-based systems. At the same time we apply the reverse to the scale factor t, where incrementing the scale level s quadruples the scale factor t, while decrementing s quarters it. Since convolving a histogram with a kernel of width σ is analogous to convolving a histogram of quadruple the resolution with a kernel of width

2σ, the inverse relationship between u and t effectively allows us to apply the same Gaussian kernel at each scale level. We illustrate our representation in Figure 4.
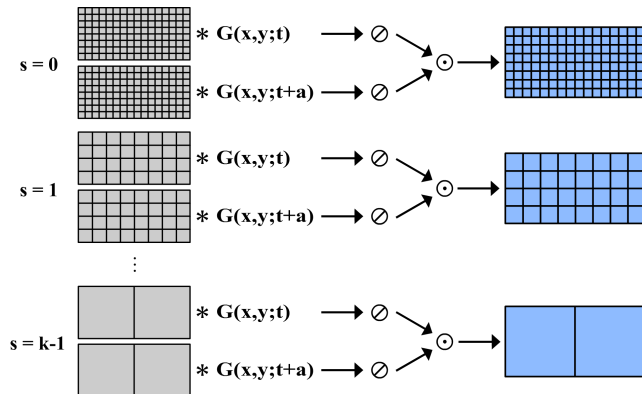

Figure 4: Scale-space representation using variable spatial resolution, where the weighting of the scale-space representation of a label in proportion to how much it contributes to the overall data density is represented by the symbol $\oslash$ and the difference operation between two scales is represented by the symbol $\odot$. The scale-space representation is divided into a varying number of blocks that can be processed in parallel.

As the earth is commonly represented using a cylindrical projection that results in longitude values ranging twice as far as latitude values ($\pm 180°$ versus $\pm 90°$), we found for varying numbers of s that scale-space representations of height $h = r2^{k-s-1}$ and width $w = 2r2^{k-s-1}$ gave results equivalent to those obtained when using representations of the highest resolution at each scale, where k refers to the total number of scale levels being used and r is a constant. In our case we set $k = 12$, which is comparable to the number of scales found in online mapping systems such as Yahoo! Maps, and $r = 1,440$, which for convenience is a power-of-two multiple of the total number of degrees along the longitudinal and latitudinal axes. At street level, where $s = 0$, this effectively results in a scale-space representation of approximately 6 million by 3 million cells, whereas at world level, where $s = k - 1$, the representation measures 2,880 by 1,440 cells; at street level the largest cell is approximately the equivalent of $0.0001°$ longitude at the equator, measuring roughly 13m, while at world level it is the equivalent of $0.25°$ longitude at the equator, or about 28km.

At lower levels of s the representations are typically too large to fit into the memory of a 'standard' personal computer, e.g. at $s = 0$ the scale-space representation for a single tag at a single scale level using 4-byte floating-point values would require 63TB of memory (although admittedly storing the representation in a sparse format would substantially reduce this number). We therefore further optimize the computations by partitioning each histogram into overlapping blocks measuring $o + r + o$ by $o + r + o$ cells and process each of them separately in parallel. Here, r is defined as before and refers to the core area of the block, while o is a sufficiently large border area to avoid the boundary effects otherwise caused by applying the Gaussian kernels or morphological operations to a block containing solely the core area. A labeled data point will thus appear only once in a core area, but may also appear in one or more bor-

der areas of its neighboring core areas. Once each block has been processed we then discard the borders, after which we can trivially connect regions spanning across adjacent blocks along their seams.

4.3 Visualization

The final step in our implementation is to integrate all aforementioned aspects into a single system that supports zoomable browsing and exploration of the photos taken across the globe through their tag regions. Our interface, presented in earlier work [43] and shown in Figure 5, initially displays a geographic map of the user's estimated location along with any tag regions for their current map zoom level. The user can explore the regions detected in the world using standard map-based navigation operations such as zooming and panning. A region can be visually browsed by clicking on it, after which our system issues an API call to Flickr to retrieve photos taken within that region that are associated with the region's tag. Our system also supports querying for regions associated with one or more specific places or tags, the latter of which is particularly useful when one wants to visually compare the photos belonging to different regions having the same tag. To provide additional insight into the discovered regions we used the Geonames[2] gazetteer to label tags that refer to place names by matching tags with entries in the gazetteer in order to give the user the ability to show or hide regions based on this property.
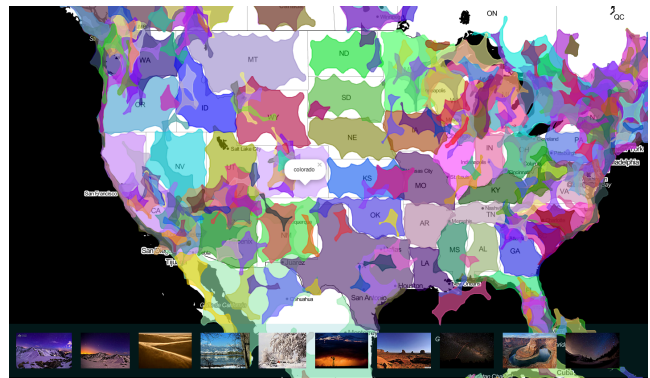

Figure 5: A screenshot of our interface showing all tag regions detected in the United States at scale level $s = 10$ and a selection of Flickr photos associated with the tag region `colorado`. The region colors are derived from the tag using a hash function, thus when a tag generates multiple regions they are assigned the same color.

5. ANALYSIS

In our exploration and browsing scenario, we note that the tags assigned to Flickr photos are user-generated for which no valid ground truth exists against which we can compare the locally characterizing tag regions that emerge from the data. However, we observe that the tags that generate regions often are toponyms, i.e. referring to place names, and a visual inspection of the regions generated for tags that match the name of a country revealed that their shapes and sizes often closely matched the territorial boundaries, which was particularly evident when there was sufficient spatial

---

[2] http://www.geonames.org/

coverage of photos taken in that country. Since countries do have verifiable and canonical boundaries that can be used for evaluation we will therefore first present an exploration of parameters appropriate at country level and then discuss to what degree our parameters are suitably tuned to describe any type of region in general, in particular given that toponym and non-toponym boundaries may not behave in the same manner and that a single set of parameters may not sufficiently account of differences between regions that emerge at different scale levels.

5.1 Parameter exploration

We extracted the polygonal boundaries from OpenStreetMap[3] for all 192 countries in the world (according to the U.N.) and computed the regions of each tag matching the name of one of these countries for an extensive range of parameter values and scales. Our aim was to determine the optimal parameters that would achieve the highest overlap between the Flickr tag regions and the OpenStreetMap regions by only varying the kernel size σ and the morphological disk radius ρ. Since the scale level s is coupled with a scale factor t, which in turn is coupled with the kernel width σ, through our parameter exploration we thus empirically determine the best σ and its associated scale factor t and scale level s.

To get a snapshot of the region contours at a certain scale level s the spacing distance a ideally goes to zero, although this would leave gaps in the scale-space from where we would not obtain any contours. To this end we set our parameter a to cover the entire distance between consecutive scale levels to maximize the signal difference and to capture all contours. At each scale level s we thus applied the Difference of Gaussians operation to the scale-space representations that were spaced a distance of $a = 4^{s+1} - 4^s$ apart. This also had as benefit that the kernel width σ conveniently differed a factor of two between both representations. We suppressed low signal responses resulting from the Difference of Gaussians operation by removing those with magnitudes less than $\epsilon = 0.0005$ by setting them to zero in the binary representation, while all signals above the threshold were set to one.

In order to compare the tag regions with the polygonal boundaries of the countries, we first converted the longitude and latitude coordinates of all points along their exterior paths to an equal-area coordinate system using the Gall-Peters projection [12], allowing us to measure the extent their geographic areas overlap according to

$$overlap(A_{ssr}, A_{osm}; s, \lambda, \sigma, \rho) = \frac{A_{ssr} \bigcap A_{osm}}{A_{ssr} \bigcup A_{osm}} \qquad (6)$$

where A_{ssr} refers to the projected area of the scale-space region with tag λ at scale level s generated using a Gaussian kernel of width σ and a morphological disk of radius ρ, whereas A_{osm} refers to the projected area of the OpenStreetMap region of the country with name λ.

We explored the parameter space where σ ranged from 0.2 to 10.0 in steps of 0.2 and ρ ranged from 5 to 75 in steps of 5, where for each parameter combination we retained the highest overlap for each country across all scale levels after which we averaged the overlaps over all countries. The results are shown in Figure 6 in which the lines indicate changes in overlap in steps of 0.02. Our exploration yielded the overall best results with a highest average overlap of 54%

[3] http://www.openstreetmap.org/

at $\sigma = 4.8$ and $\rho = 35$. From a practical perspective, given our variable scale-space representation implementation, we can simply smooth the overlapping histogram blocks at each scale level once with a kernel of size $\sigma = 4.8$ and once with a kernel of size $\sigma = 9.6$.
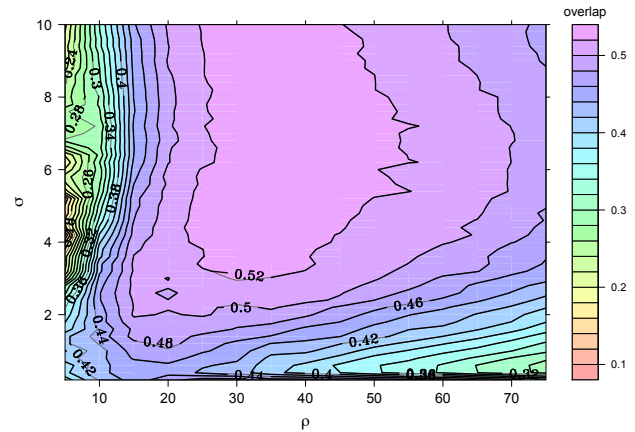

Figure 6: Highest average overlaps across all scales between the areas of Flickr tag regions referring to country names and their matching OpenStreetMap regions shown in a contour plot. Note that the colors do not vary between contour lines and thus do not exactly reflect the actual overlap values.

In Figure 7 we show the countries that emerged from our dataset using the optimal parameter values we established. As can be seen, the generated regions often follow the outline of the countries except in areas where there was little photographic coverage in our Flickr sample, such as in Africa, Russia and the Middle East. Interestingly, we observe several instances where the country regions overlap more than just along the border, e.g. Mexico and the state of Texas in the United States; India that covers Nepal, Bangladesh and Bhutan. Furthermore, we noticed – as expected – that the size of a country did appear to matter at which scale level its best overlap was attained, where larger countries tended to be found at a more global level and smaller ones at a more local level, see Figure 8.
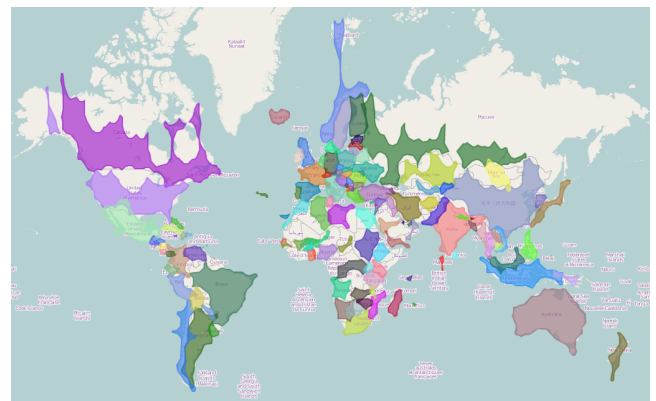

Figure 7: Flickr tag regions referring to country names generated at parameters $\sigma = 4.8$ and $\rho = 35$ extracted at the scales where they had the best overlap.
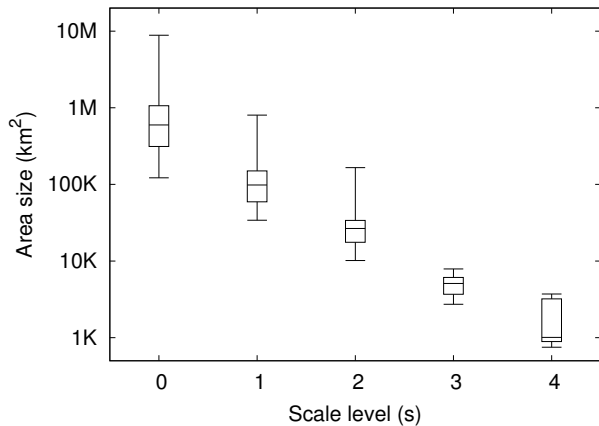
Figure 8: Box and whisker plot showing scale levels 0 to 4 versus area size at which Flickr tag regions referring to country names had the best overlap.

	Tag	Overlap
1	germany	84.1%
2	belarus	83.1%
3	poland	83.0%
4	sweden	82.6%
5	cambodia	81.8%
6	south africa	81.6%
7	czech republic	81.3%
8	turkey	80.8%
9	finland	80.0%
10	greece	79.6%

With the best parameter combination for the tag region generated for an individual country we ranked them by their overlap scores, shown in Table 1, demonstrating that with correctly tuned parameters we could achieve overlaps as high as 84.1%. This must be considered with respect to the available coverage of geotagged Flickr photos in each country, as those with poorer coverage across the whole of the territory are likely to give poorer overlap values.

While we acknowledge that we cannot draw any definite conclusions based on our parameter exploration in light of the absence of ground truth beyond country regions, the results suggest that the best parameter values we have obtained for tag regions relating to countries may transfer to other types of regions as well, in particular because the parameters remained stable for regions of varying sizes, while at the same time they were typically obtained at scales appropriate for the region given its size. However, upon closer inspection of local regions, we discovered that the tag regions that were already detected at more global levels were still present at local levels. While the regions emerging at global scales have a large footprint that in principle should be removed at local scales – since the Difference of Gaussians operation acts as a band-pass filter – at street level the coverage of the world in terms of Flickr photos is patchy and they thus still emerge in the form of disjoint smaller regions. One possible solution to this would be to hierarchically connect tag regions by determining superior, inferior and substitution relations between regions [22] and avoid displaying regions at scales substantially lower than the highest scale they have already been found at. This technique would at the same time resolve an issue we observed with near-identical regions emerging that only differ in the language of their associated tags.

5.2 Comparison

As already described in the related work section, alternative techniques for deriving colloquial regions from labeled data can be found in the research literature. In this section we present a comparison between the characteristic regions our technique generates and those obtained using a recently proposed approach that we briefly describe below; our main criteria for choosing this technique was its ability

to analyze data at different scales, its applicability to noisy user-generated data on a per-label basis without requiring too many parameters to be tuned and its ability to generate regions of arbitrary shapes. These constraints therefore directly ruled out many other approaches that focus on deriving regions by inspecting the distribution of the georeferenced objects rather than the labels [1, 7, 9, 33], those that focus on the stability of regions across scales rather than the relevance of a region at a particular scale [37], those that can only model regions as elliptical shapes [3], as well as those that do not derive any kind of statistics that measure the influence of a label with respect to all other labels [25], as this would result in regions being generated that can impossibly be locally characterizing. To elaborate on the latter issue, in our preliminary experiments we observed the tag `beach` to be present all over the city center of Barcelona, likely due to users having bulk tagged the photos they took on their holiday trip; with the prominence weighting these tags were voted down at those locations by our algorithm, whereas without such weighting these tags formed a large region that clearly was not locally characterizing.

Crowd-based noise filtering

The objective of the technique of Intagorn and Lerman [18] is the same as ours, namely to derive regions for labels at several scales that are locally special with respect to the use of other labels. However, their approach is different in that it is based on clustering the data relevant for a particular label at a particular scale, after which the polygonal boundaries of the clusters are derived. In addition, the method essentially treats each scale independently from all other scales, unlike our work. Note that the technique is very similar to the approach proposed by Jung et al. [22], although the latter focused on four well-known cities only, whereas the crowd-based noise filtering technique analyzed tags at various scales by varying the parameters and thus fits the context of our comparison better.

To briefly describe the crowd-based noise filtering technique, the underlying algorithm first partitions the world into cells, where for each cell it computes a statistic involving the number of distinct people that used each of the distinct tags in that cell. For each distinct tag the method then assesses whether its occurrence at a certain cell is particularly significant. All cells in which a tag is not used significantly enough are discarded, whereas the remaining cells are clustered by treating them as connected nodes in a graph, in which edges are formed between nodes when they are close enough according to some threshold and disjoint clusters are

then identified using connected component analysis. Finally, a polygonal boundary is derived for each resulting cluster by finding a Delauney triangulation that completely subsumes the data within each cluster, where the smoothness of the boundary can be controlled by setting a parameter.

Parameterization

For the evaluation of the colloquial regions generated by both techniques, we focused on nested bounding boxes within the world at different scales corresponding to intuitive geographic areas, where the scales zoom in from the World, to Europe, to Spain, to Catalonia and finally ending in Barcelona, allowing us to investigate the kind of regions that emerge and the nature of their differences.

For our scale-space region derivation technique we used the optimal parameters as determined through our parameter exploration, $\sigma = 4.8$, $\rho = 35$ and $\epsilon = 0.0005$, and as scale levels for the World $s = 0$, Europe $s = 2$, Spain $s = 3$, Catalonia $s = 6$ and Barcelona $s = 10$. As mentioned earlier, while appropriate for a particular scale, our technique can produce regions that may vary substantially in size; considering that a typical computer screen has a resolution on the order of 1440x1440 pixels (hence we had set our block size r to these values), we required that the regions occupied a minimum surface area of 1000 cells and a maximum of 400,000 cells to ensure they are not too large or too small for viewing purposes.

For the crowd-based noise filtering technique we in principle used the parameter values mentioned in the paper at the scale levels that were presented, unless we observed that an adjustment resulted in better results; we used parameters $\alpha = 10$, $k = 1$ and as scale levels for the World $Eps = 500$, Europe $Eps = 100$, Spain $Eps = 50$ and Catalonia $Eps = 10$. Due to the minimum cell size of the world within which the statistics were computed, as specified by α, we could not compute the regions for Barcelona, because this required a finer grid size.

Results

From the quantity of regions obtained at the various scales, shown in Table 2, we can instantly see that the crowd-based noise filtering technique discovers a large number of regions, even after we already limited it by setting a minimum surface area. In contrast, our scale-space region derivation technique finds a small number of regions at world level and larger numbers as the scale level becomes more local. The notion of finding more regions at local scales makes intuitive sense, since at global scales only the regions with large footprints will emerge, of which there are few, while at local scales only the regions with small footprints will be discovered, of which there are many.

Table 2: Number of regions detected by our scale-space region derivation technique (SSR) and by the crowd-based noise filtering technique (CNF) for different areas of the world.

Area	SSR	CNF
World	488	492197
Europe	1274	633193
Spain	3285	663528
Catalonia	6968	617047
Barcelona	9963	—

We illustrate the resulting regions detected within the four geographic areas in Figure 9. As can be seen, both approaches have roughly equivalent coverage of the world, where differences show up at different scale levels. For instance, the crowd-based noise filtering technique is able to cover Africa almost completely and also a larger part of Russia than our scale-space regions approach, and furthermore produces more regions in the Spanish state of Catalonia.

Qualitatively, the scale-space tag regions appear to more closely fit the country regions, whereas the regions produced by the crowd-based noise filtering technique seem to be rather crude; when measuring this quantitatively, by comparing both approaches in terms of overlap with the OpenStreetMap country regions, the scale-space regions approach achieved an average overlap of 56.1%, while the crowd-based noise filtering approach reached an average overlap of 41.9% using the clustering distance threshold of $Eps = 100$ as specified in the paper. Furthermore, due to the scale-space technique generating substantially less regions the interface appears less cluttered than when it is filled with the regions produced by the crowd-based noise filtering approach.

When analyzing the underlying technique of the crowd-based noise filtering approach, it becomes clear why it produces such a large quantity of regions. Namely, once a single labeled data point exceeds the threshold set by the statistic associated with the grid cell the data point is assigned to a region will be generated; the region initially is very small and may get clustered together with other regions, but if not it will remain to exist since the algorithm does not perform any filtering on the produced regions. We can clearly see this in the figures, where even after filtering out small regions and showing a small subset of regions the interface still shows many regions.

6. CONCLUSIONS

In this paper we presented a novel, generalized framework that automatically uncovers locally characterizing regions present within geotagged data using scale-space theory. The regions are generated by applying image analysis techniques to the scale-space representations of each data label. To handle large labeled collections across a wide range of scales we presented optimized implementation techniques to perform the region derivation in an efficient parallel manner. We analyzed the obtained tag regions in terms of the parameterization of the scale-space model at different scales and compared them against the regions produced by a recently proposed technique, showing that our tag regions can closely resemble the polygonal boundaries of the places they refer to. We envisage our technique being used to help users explore labeled geographic data in new ways, allowing them to draw insights in a manner that current interfaces to such data do not yet support. In future work we would like to investigate the implications of a dynamic dataset on the regions produced by our scale-space model and extend our technique to incorporate time as an additional dimension in our scale-space model for the purpose of detecting regions for labeled spatiotemporal data. Furthermore, we would like to use the discovered regions for purposes beyond media exploration, such as supporting spatially-aware and temporally-aware tagging, for instance learning when to suggest certain tags based on the geographic location and the timestamp of a given photo.
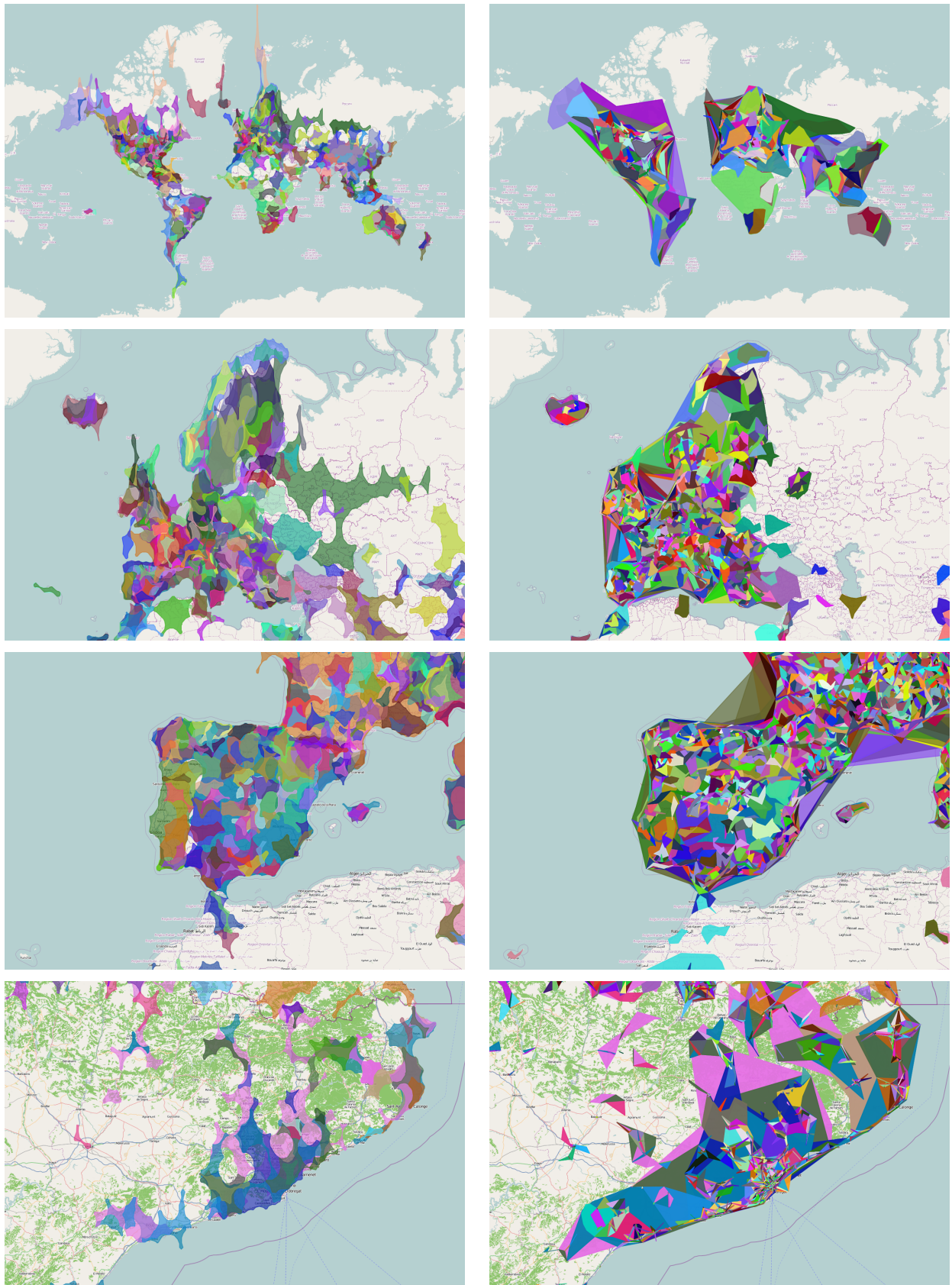
Figure 9: The regions detected by our scale-space region derivation technique (left column) and those by the crowd-based noise filtering technique (right column) for the geographic areas of the World (top), Europe (middle-top), Spain (middle-bottom) and Catalonia (bottom). To avoid excessive numbers of regions to be returned we limited the crowd-based noise filtering to return at most 10,000 regions.

7. REFERENCES

[1] S. Ahern, M. Naaman, R. Nair, and J. Yang. World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. *JDCL*, pp. 1–10, 2007.

[2] T. Ai and J. Li. The lifespan model of GIS data representation over scale space. *GEOINFORMATICS*, pp. 1–6, 2009.

[3] Y. Baba, F. Ishikawa, and S. Honiden. Extraction of places related to Flickr tags. *ECAI*, pp. 523–528, 2010.

[4] T. Barkowsky, L. J. Latecki, and K.-F. Richter. Schematizing maps: simplification of geographic shape by discrete curve evolution. *Spatial Cognition II*, pp. 41–53, 2000.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. Speeded-Up Robust Features (SURF). *CVIU*, 110(3):346–359, 2008.

[6] F. Chang, C.-J. Chen, and C.-J. Lu. A linear-time component-labeling algorithm using contour tracing technique. *CVIU*, 93:206–220, 2004.

[7] W.-C. Chen, A. Battestini, N. Gelfand, and V. Setlur. Visual summaries of popular landmarks from community photo collections. *ICMR*, pp. 789–782, 2009.

[8] M. Cristani, A. Perina, U. Castellani, and V. Murino. Content visualization and management of geo-located image databases. *CHI*, pp. 2823–2828, 2008.

[9] D.-P. Deng, T.-R. Chuang, and R. Lemmens. Conceptualization of place via spatial clustering and co-occurrence analysis. *LBSN*, pp. 49–56, 2009.

[10] V. Estivill-Castro and I. Lee. AUTOCLUST: automatic clustering via boundary extraction for mining massive point-data sets. *GeoComputation*, 2001.

[11] P. F. Felzenszwalb and D. P. Huttenlocher. Distance transforms of sampled functions. *Theory of Computing*, 8(1):415–428, 2012.

[12] J. Gall. Use of cylindrical projections for geographical, astronomical, and scientific purposes. *Scottish Geographical Magazine*, 1(4):119–123, 1885.

[13] W. W. Hargrove and F. M. Hoffman. Using multivariate clustering to characterize ecoregion borders. *CiSE*, 1(4):18–25, 1999.

[14] C. Hauff and G.-J. Houben. Geo-location estimation of Flickr images: social web based enrichment. *ECIR*, pp. 85–96, 2012.

[15] J. Hays and A. A. Efros. IM2GPS: estimating geographic information from a single image. *CVPR*, pp. 1–8, 2008.

[16] H. Hotta and M. Hagiwara. A neural-network-based geographic tendency visualization. *WI-IAT*, pp. 817–823, 2008.

[17] H. Hotta and M. Hagiwara. Online geovisualization with fast kernel density estimator. *WI-IAT*, pp. 622–625, 2009.

[18] S. Intagorn and K. Lerman. Learning boundaries of vague places from noisy annotations. *SIGSPATIAL*, pp. 425–428, 2011.

[19] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of geo-referenced photographs. *MIR*, pp. 89–98, 2006.

[20] R. Jain. Experiential computing. *CACM*, 46(7):48–54, 2003.

[21] C. B. Jones, R. S. Purves, P. D. Clough, and H. Joho. Modelling vague places with knowledge from the Web. *IJGIS*, 22(10):1045–1065, 2008.

[22] D. Jung, H. Park, R. Maeng, and S. Han. A geometric pattern-based method to build hierarchies of geo-referenced tags. *SocialCom*, pp. 546–551, 2010.

[23] L. S. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. *WWW*, pp. 297–306, 2008.

[24] L. S. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How Flickr helps us make sense of the world: context and content in community-contributed media collections. *MM*, pp. 631–640, 2007.

[25] C. Kessler, P. Maué, J. T. Heuer, and T. Bartoschek. Bottom-up gazetteers: learning from the implicit semantics of geotags. *GeoS*, pp. 83–102, 2009.

[26] J. Kleban, E. Moxley, J. Xu, and B. S. Manjunath. Global annotation on georeferenced photographs. *CIVR*, 2009.

[27] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[28] I. Laptev, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner. Automatic extraction of roads from aerial images based on scale space and snakes. *Machine Vision and Applications*, 12:23–31, 2000.

[29] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: state of the art and challenges. *TOMCCAP*, 2(1):1–19, 2006.

[30] O. Linde and T. Lindeberg. Composed complex-cue histograms: An investigation of the information content in receptive field based image descriptors for object recognition. *CVIU*, 116(4):538–560, 2012.

[31] T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus of attention. *IJCV*, 11(3):283–318, 1993.

[32] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[33] C.-A. Lu, C.-H. Chen, and P.-J. Cheng. Clustering and visualizing geographic data using geo-tree. *WI-IAT*, pp. 479–482, 2011.

[34] R. Raguram, C. Wu, J.-M. Frahm, and S. Lazebnik. Modeling and recognition of landmark image collections using iconic scene graphs. *IJCV*, 95(3):213–239, 2011.

[35] U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1(3):244–256, 1972.

[36] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from Flickr tags. *SIGIR*, pp. 103–110, 2007.

[37] T. Rattenbury and M. Naaman. Methods for extracting place semantics from Flickr tags. *TWEB*, 3(1):1, 2009.

[38] S. Robertson and H. Zaragoza. The probabilistic relevance framework: BM25 and beyond. *FTIR*, 3(4):333–389, 2009.

[39] S. Rudinac, A. Hanjalic, and M. Larson. Finding representative and diverse community contributed images to create visual summaries of geographic areas. *ICMR*, pp. 1109–1112, 2011.

[40] P. Serdyukov, V. Murdock, and R. van Zwol. Placing Flickr photos on a map. *SIGIR*, pp. 484–491, 2009.

[41] B. Sigurbjornsson and R. v. Zwol. Flickr Tag Recommendation based on Collective Knowledge. *WWW*, pp. 327–336, 2008.

[42] M. Taylor, H. Zaragoza, N. Craswell, S. Robertson, and C. Burges. Optimisation methods for ranking functions with multiple parameters. *CIKM*, pp. 585–593, 2006.

[43] B. Thomee and A. Rae. Exploring and Browsing Photos through Characteristic Geographic Tag Regions. *MM*, pp. 1273–1274, 2012.

[44] O. Van Laere, S. Schockaert, and B. Dhoedt. Finding locations of Flickr resources using language models and similarity search. *ICMR*, article 48, 2011.

[45] F. Wang. Job access and homicide patterns in Chicago: an analysis at multiple geographic levels based on scale-space theory. *JQC*, 21(2):195–217, 2005.

[46] K. Q. Weinberger, M. Slaney, and R. van Zwol. Resolving tag ambiguity. *MM*, pp. 111–120, 2008.

[47] A. Witkin. Scale-space filtering: a new approach to multi-scale description. *ICASSP*, 9:150–153, 1984.

[48] K. Yanai, H. Kawakubo, and B. Qiu. A visual analysis of the relationship between word concepts and geographical locations. *CIVR*, article 13, 2009.