Characterization of Geographic Regions Based on Georeferenced Data from the Social Web

Eduardo Cunha (eduardo.cunha@ist.utl.pt)

Instituto Superior Técnico and INESC-ID

Abstract

The characterization of specific places or more general geographic regions is essential to a variety of decision-making processes, particularly in the context of problems related with urbanism or demographic studies. In the context of my MSc thesis, I propose new ways of characterizing geographic regions, through the usage of georeferenced information extracted from location-based social networks and from popular Web 2.0 services, such as Twitter, FourSquare or Flickr. The specific methods that I propose in my dissertation characterize geographic regions with basis on information extracted from publicly available georeferenced photos, shared by the users of Flickr, together with auxiliary information available from raster datasets containing geographic information (e.g., elevation or population density) about the desired locations. Data classification techniques are used to estimate the boundaries of vague regions, or to infer geographic characteristics like the land coverage. The classification methods are based on Support Vector Machines, leveraging on multiple Gaussian kernels to increase the estimation accuracy. An extensive set of experiments attests to the effectiveness of the proposed methods.

1 Introduction

The characterization of geographic regions is an important aspect for urban planning, also having important applications in marketing (e.g., in choosing business locations, advertisement placing, etc.) or in urbanism and demographic studies, among others. In the context of my MSc dissertation, I am proposing new ways to characterize geographic areas with information extracted from georeferenced data published on on-line services and location-based social networks, like Twitter[1], FourSquare[2] or Flickr[3]. The increasing usage of these services has turned them into sources that are rich in georeferenced data that can be exploited to extract patterns relevant to geographic characterization. These patterns can be extracted from characteristics like the number of entries associated to different places, timestamps associated to visits to the different entries/regions, multimedia contents related with the different regions, the kinds of users that visit specific places, etc. To characterize geographic regions, I proposed to leverage on data classification techniques, based on Support Vector Machines (SVM) and Multiple Kernel Learning (MKL). I have specifically made experiments on two real-world problems related to geographic characterization, namely estimating the boundaries of vague regions, and estimating the land coverage classification of geographic regions. I proposed to use one-class SVMs for the estimation of vague regions, and multi-class SVMs for the estimation of land coverage classes. The remaining contents of this article are organized as follows. Section 2 surveys previous work related with the two experiments that I have made. Section 3 details the two experiments, showing how data collected from Flickr can be used to define vague regions and to classify various zones of cities in terms of their land coverage. Section 4 presents the results that I obtained in the context of both experiments. Finally, Section 5 summarizes the main conclusions of this work.

[1] http://twitter.com/
[2] http://foursquare.com/
[3] http://www.flickr.com/

2 Related Work

This section is composed of two different subsections, where Section 2.1 presents previous work concerned with defining the boundaries of vague regions, and Section 2.2 overviews previous work related to land coverage classification and to the usage of georeferenced multimedia resources for terrain classification.

2.1 Defining the Boundaries of Vague Regions

Territorial subdivisions and geographic borders in general are essential for the analysis of many different types of phenomena, not just in the geographical information sciences but also in areas such as sociology, political science, history, and economics (Grady et al., 2012; Newman, 2006). Since these vague regions are frequently used in natural language discourse, the development of methods for their assessment and cartographic representation assumes a particular importance.

The delineation of vague geographic concepts has indeed been widely studied in the geographical information sciences (Schockaert, 2011; Vasardani et al., 2013). Previous work has explored different methods for delimiting imprecise regions (Schockaert, 2011; Vasardani et al., 2013), most of them based on knowledge about points which are known to be inside or outside of the regions to be defined, often assuming that names for vague regions co-occur frequently with other place names. Different sources of information have been experimented with, including user questionnaires (CIESIN and CIAT, 2005; Clough and Pasley, 2010), information about points and geospatial relations described in maps and gazetteers (Alani et al., 2001), textual information published on the Web (Arampatzis et al., 2006; Jones et al., 2008; Goldberg et al., 2009), or georeferenced photos published on sites like Flickr (Grothe and Schaab, 2009).

In what I consider to be the most relevant previous work, Grothe and Schaab (2009) described two methods for the automatic delineation of vague regions based on Flickr data, namely a method based on Kernel Density Estimation (Brunsdon, 1995) and another based on One-Class Support Vector Machines, i.e. a variant of the classical Support Vector Machines (SVM) classification approach requiring only positive examples (Munoz and Moguerza, 2006). The same authors have also described techniques for the optimization of the parameters required by both algorithms, when used for delimiting vague regions.

2.2 Land Classification with Georeferenced Multimedia

Previous work has also argued that georeferenced photo collections can enable a new form of observational inquiry, which has been termed *proximate sensing* (Jacobs et al., 2007). While the traditional field of remote sensing is mostly based on the usage of overhead images from distant scenes (e.g., satellite imagery) to derive geographic information (Hu and Wang, 2013; Vatsavai et al., 2011), proximate sensing is instead concerned with the usage of ground-level images of close-by objects and scenes, also with the objective of deriving geographic information of relevance to a variety of problems.

The idea of creating maps with basis on ground-based image sensors was perhaps first introduced as an application of the Archive of Many Outdoor Scenes (Jacobs et al., 2007, 2009b,a), which is a dataset of images from publicly available webcams collected from approximately 20,000 outdoor webcams located all around the world, with images recorded at every 30 minutes and with many of the cameras associated with geolocations that were either provided by the maintainer or inferred automatically (e.g., from the IP address). Relying on this, or on similar datasets, different machine learning methods have been used to predict environmental properties from Web imagery. For instance, image segmentation for the detection of tree regions was used in applications related to plant phenology (Jacobs et al., 2009a; Riordan et al., 2010), while canonical correlations analysis was used to predict wind velocity in scenes with visible trees or flags (Jacobs et al., 2009a). Water vapor pressure was inferred in scenes with large depth of field, using an image feature based on contrast (Jacobs et al., 2009b). Cloud maps have been estimated with basis on regression analysis (Murdock et al., 2013), and semi-supervised methods have been proposed to estimate atmospheric visibility from webcam data (Xie et al., 2010). However, these previous approaches were mostly based on the availability of properly placed image sensors (i.e., cameras), and they relied on very specific features that are only effective for certain types of environmental applications and certain types of scenes.

Murdock et al. (2013) have for instance tackled the specific problem of estimating satellite cloud maps from a collection of ground-based photos, through a random forest regression technique. Their approach uses historical georeferenced satellite imagery to learn a regression model that maps the ground imagery to the satellite cloud map, considering infrared imagery obtained from Geostationary Operational Environmental Satellites[4] as the ground truth cloud measurements. The authors explored eight different representational choices for inferring the cloud status based on the ground-level imagery (i.e., low-dimensional projections based on Principal Component Analysis or Partial Least Squares, computed on either the full image, the sky pixels only, or the ground pixels only, and the histogram of hues for the sky pixels only or for the top 20 rows in the image). The authors also considered several different alternatives for spatially interpolating the sparse measurements obtained from the individual images (i.e., spline interpolation, nearest neighbors interpolation, and kriging), in order to produce a complete cloud map.

3 Overview on the Experiments

This section describes the general methodology for the two experiments that constitute the focus of my MSc thesis. Section 3.1 introduces the usage of one-class SVMs within a multiple kernel learning setting, which was the method used in the experiments related to the delineation of vague regions. Section 3.2 introduces the usage of multi-class SVMs, which supported the experiments related to land coverage classification.

3.1 Characterizing Vague Regions with Flickr Data and One-Class SVMs

In my work, I proposed to use one-class supervised learning to find the boundaries of vague regions, mostly leveraging on data collected from photo sharing services like Flickr. Let R be the actual footprint of a region (i.e., the set of all points that belong to the region). My objective is to infer an approximation R' for the footprint, that is as similar as possible to R, derived from a sample dataset S. The dataset S consists of N individual observations x_i corresponding to geospatial coordinates that represent points on the surface of the Earth, with $1 \leq i \leq N$. Each of the observations $x_i \in S$ can also be associated to a set of descriptive features (e.g., the estimated population, the terrain elevation, or the land coverage classification type at point x_i). Only samples of a single class are considered for the set S (i.e., I only consider positive examples of points belonging to the vague region), as I only have access to cases where a given user uploaded a photo into Flickr, containing an association to some specific geospatial coordinates, and containing a tag with the name of the vague region. Hence, the problem is to identify the complete set of points that most probably lie in R, through the supervised classification procedure. A model is learned from the dataset S, and I then apply it to all possible points, classifying them as either belonging or not to the region. This results in an estimated footprint R'. Points located in $R \cap R'$ are correctly identified as belonging to R (true positives), while those located in $R' - R$ are falsely identified as belonging to R (false positives). The goal is to maximize the number of true positives, while at the same time minimizing the number of false positives.

As for the supervised learning model, I considered the usage of one-class Support Vector Machines (Munoz and Moguerza, 2006). Support Vector Machines (SVMs) are widely applied in data classification tasks. The general method is based on the statistical learning theory, and is characterized by having good efficiency and generalization capabilities, compared to other data classification methods. The goal in traditional SVM classification is to decide to which one of two classes a given observation belongs to, but if the dataset consists only of specimen of a single class, the formulation of the SVM problem changes slightly. In the one-class setting, the goal is to decide whether a new observation has the same properties as the samples included in the training dataset or not. In the latter case, the point is considered an outlier.

The result of an SVM learner is an α-weighted linear combination of kernel values plus a bias term b, corresponding to the following equation where the x_i, with $i = 1, \ldots, N$, are the training examples, labeled with $y_i \in \{\pm 1\}$ in the case of binary classification tasks, and always with $y_i = 1$ for one-class SVMs.

[4] `http://www.oso.noaa.gov/goes/`

$$\mathrm{f}(x) = \mathrm{sign}\left(\sum_{i=1}^{N} y_i \times (\alpha_i \times \mathrm{k}(x_i, x) + b)\right)$$

The learning method for one-class SVMs considers a free hyper-parameter ν that expresses the maximum fraction of outliers, obtained after solving an optimization problem, with $0 \leq \nu \leq 1$. If information on the uncertainty included in the training dataset is available (e.g., by an analysis of the training data), the expected fraction of errors can be expressed in ν. In a previous related work, Grothe and Schaab (2009) reported on good results for delimiting vague regions with one-class SVMs and Flickr data, when setting the ν parameter to 0.14. In my experiments, I also used this value.

Recent developments, as reported in the literature on SVMs and on other kernel methods, have shown the benefits of considering multiple kernels (Gönen and Alpaydin, 2011). Thus, in this set of experiments, I also relied on a modern SVM-based method that uses multiple kernels.

Kernels have typically to be chosen a-priori (i.e., in my experiments, I used a combination of multiple Gaussian kernels, with different features and/or kernel widths). The parameters of the one-class SVM model are determined by solving the following optimization problem:

$$\begin{aligned}
&\max \theta \\
&\text{w.r.t. } \theta \in \mathbb{R}, \beta \in \mathbb{R}^M \text{ and } \alpha \in \mathbb{R}^N \\
&\text{s.t. } 0 \leq \beta, \sum_{k=1}^{M} \beta_k = 1, \sum_{i=1}^{N} \alpha_i = 1, \text{ and } \theta \leq \sum_{k=1}^{M} \beta_k \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \mathrm{k}_k(x_i, x_j) \; \forall \; 0 \leq \alpha \leq \frac{1}{\nu N}
\end{aligned}$$

In the previous equation, the parameter ν is the pre-specified regularization parameter, N is the number of training examples, the parameters α are the weights assigned to each training example, and the parameters β are the weights assigned to each of the M sub-kernels $\mathrm{k}_k(x, x')$. I used the implementation for multiple kernel learning of one-class SVM classifiers that is provided in the `shogun`[5] machine learning toolkit. Within `shogun`, the above optimization problem is solved using semi-infinite programming (Hettich and Kortanek, 1993). The reader is referred to the paper by Sonnenburg et al. (2006) for more details about `shogun` and about multiple kernel learning in general.

Besides the geospatial coordinates of latitude and longitude for each point $x_i \in S$, I also considered additional features for each of the observations, namely (i) features relative to population statistics, (ii) features relative to land coverage types, (iii) features derived from elevation data, and (iv) features obtained from textual tags associated to the photos in Flickr. In the case of features (i), (ii) and (iii), the `GDAL-convert`[6] tool was used to generate raster datasets in a common representation format, from publicly available datasets encoding these types of information (i.e., from the Gridded Population of the World (GPW[7]) dataset encoding the distribution of human population across the globe, the Global Land Cover Facility (GLCF[8]) dataset encoding land coverage features, and the NASA Shuttle Radar Topographic Mission (SRTM[9]) dataset encoding terrain elevation features). The features derived from textual tags were computed from information directly available from Flickr (i.e., the top most relevant tags associated to the training photos).

3.2 Multi-Class SVMs and Multiple-Kernel Learning

Large collections of georeferenced photos can also perhaps be used to derive maps depicting what-is-where on the surface of the Earth. I specifically focused on the task of using georeferenced image collections to perform land-coverage classification, a problem for which one can easily access ground-truth data for performing the evaluation. Thus, I investigated whether the classification of feature vectors derived from georeferenced images can be used to assign land-coverage labels to the individual cells of a raster, in order to create maps.

[5] `http://shogun-toolbox.org`
[6] `http://www.gdal.org/`
[7] `http://sedac.ciesin.columbia.edu/data/collection/gpw-v3/`
[8] `http://www.landcover.org/`
[9] `http://srtm.csi.cgiar.org/`

Let R be a raster-based representation for a given study region, composed of a set of individual cells $r_i \in R$ that are organized as a rectangular grid. Each cell $r_i \in R$ is associated to a land-cover class c_{r_i} from a finite set of classes C. My objective was to infer an approximation R' for the land-coverage classes of the cells in the study-region, that is as similar as possible to R (i.e., that has as many cells as possible assigned to the correct class), derived from a sample dataset S. The dataset S consists of N individual observations x_i corresponding to geospatial coordinates that represent points within the study region, with $1 \leq i \leq N$. Each cell from R denotes a region that can contain zero, one, or many observations x_i (i.e., the N observations are sparse and unevenly distributed). Each of the observations $x_i \in S$ is associated to a set of descriptive features (e.g., the geospatial coordinates, visual content descriptors, or textual tags annotating a photo taken at point x_i). The feature vectors from a subset of observations $S_{train} \subset S$, in my case corresponding to observations annotated with tags that are highly related to the names of the land-coverage classes in C, are used to learn a classification model. The complete set of observations S is used to derive feature vectors for all of the cells $r_i \in R'$, by averaging the feature vectors of all observations contained in each given cell r_i, or by averaging the feature vectors of the five closest photos in the case of cells that do not contain any observations (i.e., cells without any observation are represented through a weighted average of their five closest observation, where the weights correspond to the inverse of the geometrical distance towards the center of the cell, as computed through Vincenty's geodetic formulae (Vincenty, 1975)). The learned classification model is then applied to the feature vectors of all cells $r_i \in R'$, this way producing the land-coverage map.

To efficiently find the five closest photos to a cell, I relied on a KD-Tree data structure to index the positions of the photos (Bentley, 1975). A KD-Tree is essentially a binary tree in which every node is a k dimensional point (i.e., two dimensional, in my case). Every non-leaf node corresponds to a splitting hyperplane, which can guide nearest neighbor searches through half-spaces.

In what regards the classification approach, I again considered the usage of Support Vector Machine classifiers (Munoz and Moguerza, 2006). The goal in traditional SVM classification is to decide to which one of two classes a given observation belongs, although multi-class problems can also be handled through SVMs, for instance through the heuristic one-vs-one or one-vs-all strategies, or through slight changes in the formulation of the SVM optimization problem. In the set of experiments repeated on this section, I relied on a modern SVM-based method that uses multiple kernels and that directly handles multi-class problems.

$$\mathrm{f}(x) = \arg\max_{y \in Y} \left(\sum_{i=1}^{N} y_i \times \left(\alpha_i \times \sum_{j=1}^{M} \beta_j \times \mathrm{k}_j(\Phi_j(x_i; y_i) - \Phi_j(x_i; y), \Phi_j(x; y_i) - \Phi_j(x; y)) + b_{y_i} \right) \right)$$

In the previous equation, the parameters α correspond to the weights assigned to each training example, β corresponds to the kernel weights, and $\Phi_{\mathrm{K}}(x; y)$ are joint feature maps given by $\lambda(y) \otimes x$, where the parameter $\lambda(y)$ is a class attribute vector, and x is a training instance. As in the experiment described in Section 3.1, kernels have to be chosen a-priori, and the parameters of the multi-class model are determined by solving the following optimization problem:

$$\begin{aligned} \min_{\alpha} \ & \gamma - \sum_i \alpha_{i_{y_i}} \\ \text{s.t.} \ & \forall i : 0 \leq \alpha_{i_{y_i}} \leq C, \ \forall i : \forall u \neq y_i : \alpha_{iu} \leq 0, \ \forall i : \sum_{u \in Y} \alpha_{iu} = 0, \ \forall u : \sum_i \alpha_{iu} = 0 \text{ and} \\ & \forall K : \gamma \geq \frac{1}{2} \sum_{i,j,u,v} \alpha_{iu} \alpha_{jv} \langle \Phi_{\mathrm{K}}(\mathrm{x}_i, u), \Phi_{\mathrm{K}}(\mathrm{x}_j, v) \rangle \end{aligned}$$

In the previous equation, the parameter C is a pre-specified regularization parameter. The expression $\langle \Phi_{\mathrm{K}}(\mathrm{x}_i, u), \Phi_{\mathrm{K}}(\mathrm{x}_j, v) \rangle$ is the dot product between two feature maps. In my experiments, I used the implementation for multiple kernel learning of multi-class SVM classifiers that is provided in the `shogun` machine learning toolkit. Within `shogun`, and similarly to the case of one-class SVMs, the above optimization problem is solved using semi-infinite programming (Hettich and Kortanek, 1993). The reader is again referred to the paper by Sonnenburg et al. (2006) for more details about `shogun` and multiple kernel learning in general.

Besides the geospatial coordinates of latitude and longitude for each observation x_i, and similarly to the case of the experiments concerned with delimiting vague regions, I also considered additional features for each

of the observations, namely (i) external features relative to population statistics gathered from the GPW dataset, (ii) external features derived from elevation data taken from the NASA SRTM dataset, (iii) visual features extracted from the photos, and (iv) features obtained from textual tags associated to the photos in Flickr. In what concerns the visual content features, most of them correspond to MPEG-7 descriptors extracted from the photos, although I also used some other popular content descriptors from the literature on image retrieval, that are available in the dataset from the 2013 MediaEval Placing Task (Hauff et al., 2013). The features related to the textual tags correspond to 300 binary features indicating the presence or absence of each tag. The textual tags were chosen based on 3 heuristics that determine each tag's relevance to land-coverage classes, and that are detailed in the dissertation.

4 Experimental Results

This section presents the experimental methodology and the obtained results for the two tasks addressed in the context of my MSc research, with Section 4.1 addressing the delineation of vague regions, and Section 4.2 addressing the task of land coverage classification.

4.1 Delimiting Vague Region Boundaries

I evaluated the proposed approach for delimiting vague regions using a large collection of Flickr photos originally gathered in the context of the SAPIR[10] (Search on Audio-visual content using Peer-to-peer Information Retrieval) European project, namely the Content-based Photo Image Retrieval (CoPhIR) collection, which has recently been made available, and that is described in a paper by Bolettieri et al. (2009). Several previous studies have collected their own subsets of Flickr photos, but I argue that using a common dataset facilitates the reproducibility of the experiments reported here.

As precise regions for my evaluation, I started with the 11 European countries that were used by Grothe and Schaab (2009), and that were originally selected for their diversity in size, shape, and the availability of geotagged photos, this way allowing for testing the proposed approaches under varying conditions. The ground-truth geographic boundaries were obtained from the shapefiles available from the GADM[11] database of global administrative areas. Having unambiguous footprints makes it possible to evaluate the estimations with quantitative measures, for which I used the traditional accuracy, recall, precision, and F_1 metrics.

In a first set of experiments, I focused on models that only used geospatial coordinates associated to Flickr photos, comparing the usage of a traditional one-class SVM classifier, as reported by Grothe and Schaab (2009), against the usage of one-class SVM models combining three different Gaussian kernels, with widths γ corresponding to the values of $\sqrt{2}/2$, $\sqrt{2}$, and $2 \times \sqrt{2}$. The hyper-parameter ν of the SVM classifiers was set to 0.14 on all experiments (i.e., the same value that was used by Grothe and Schaab (2009)), after an initial set of tests in which I have also tried the values of 0.01, 0.1 and 1.

Table 1 presents the obtained results for the 11 European countries that were considered, showing that the usage of multiple kernel learning outperforms traditional one-class SVMs. Approximate randomization testing (Edgington, 1969) was used to compare the outputs of these two methods over all the considered countries, in terms of the 4 different evaluation metrics. The results showed that the classifiers are indeed significantly different at the level of 0.05. Notice that even though I am using the same 11 countries that were considered by Grothe and Schaab (2009), my results cannot be directly compared, given that I used a different set of Flickr photos, and a different source of ground truth information for the boundaries. However, I did re-implement the approach originally presented by these authors (i.e., one-class SVMs using a single kernel and the geospatial coordinates), afterwards testing it in my dataset – see Table 1.

Table 2 presents the obtained results over the same 11 European countries that were considered in my first set of experiments, comparing models that only used the geospatial coordinates, against models that used the additional features. In this case I used one-class SVM models combining six different Gaussian kernels, three of them using only information from the geospatial coordinates (i.e., similarly to the case of

[10] `http://www.sapir.eu/`
[11] `http://www.gadm.org/`

Table 1: Comparing models that use a single kernel, or a combination of Gaussian kernels.

One kernel using geospatial coordinates					Three kernels using geospatial coordinates				
	Pre	Rec	F1	Acc		Pre	Rec	F1	Acc
Albania	0.84	0.80	0.82	0.84	Albania	0.85	0.80	0.82	0.85
Belarus	0.93	0.30	0.45	0.61	Belarus	0.95	0.37	0.53	0.65
Croatia	0.63	0.48	0.54	0.79	Croatia	0.63	0.54	0.58	0.79
France	0.95	0.58	0.72	0.81	France	0.96	0.67	0.79	0.85
Germany	**0.99**	0.58	0.73	0.74	Germany	**0.99**	0.66	0.79	0.79
Greece	0.39	0.58	0.47	0.72	Greece	0.37	0.60	0.46	0.70
Ireland	0.72	0.80	0.76	0.77	Ireland	0.73	0.87	0.79	0.79
Italy	0.87	0.75	0.81	0.91	Italy	0.86	0.77	0.81	0.91
Luxembourg	0.92	0.63	0.74	0.83	Luxembourg	0.92	0.63	0.75	0.83
Switzerland	0.85	0.87	0.86	**0.86**	Switzerland	0.85	0.88	**0.87**	**0.86**
Ukraine	0.87	0.13	0.22	0.57	Ukraine	0.88	0.16	0.27	0.59

the experiments reported over Table 1), and the other three using the remaining features. For each group of three kernels, the kernel widths were again selected as $\sqrt{D_m}/2$, $\sqrt{D_m}$, and $2 \times \sqrt{D_m}$, where D_m is the dimensionality of the corresponding feature representation. The values in bold that are shown both in Tables 1 and 2 correspond to the best results that were obtained for each of the 11 countries, in terms of the four different evaluation metrics that were considered in my experiments.

The results on Table 2 show that the additional features often lead to improved results, particularly the features derived from population counts (i.e., in 5 of the 11 countries that were used in my experiments, the best results in terms of precision were obtained with models that combined geospatial coordinates with the population feature) and from the textual tags associated to the photos (i.e., in 8 of the 11 countries, the best results in terms of recall were obtained with models that combined geospatial coordinates with the features derived from the textual tags). Using photos from the CoPhIR dataset, the method that corresponds to the usage the complete set of features that were proposed, through multiple kernel learning of SVMs, corresponds to an average improvement of approximately 5.5% in terms of the F_1 metric, over the one-class SVM approach that corresponds to the method reported by Grothe and Schaab (2009).

In Figure 1, and for illustration purposes, we can see the results for six of the eleven countries that were considered in my formal evaluation, namely for *Italy*, *Switzerland*, *Greece*, *Germany*, *Ireland* and *Luxembourg*. All these results were obtained with the models that combined all the features that were proposed. In the different maps that are shown in Figure 1, the areas painted in red represent the real footprints of each country, while the areas in blue represent the estimated footprints. The green dots correspond to the locations of the Flickr photos that were used to generate the results, and they show that while the data acquired from Flickr naturally contains some errors (i.e., I am occasionally using some points that are not located at the region being defined, because either the photo was assigned wrong coordinates, or because the placename tag does not denote the place where the photo was taken), the assumption that these points will be predominantly located within the vague region seems reasonable.

Besides the formal evaluation with quantitative measures of classification quality, involving regions whose boundaries are well known, I have also analyzed the obtained results for a small set of vague regions. Formally evaluating the performance of the proposed approach on vague regions is much harder, given that I cannot easily access ground-truth information. However, I attempted to gather several illustrative examples of the results obtained for vague regions, showing that they are indeed meaningful. For instance, for illustration purposes, Figure 2 shows results for two vague concepts corresponding to large geographic regions, namely the *Alps* and *Anatolia*. These results were obtained with models involving the complete set of features.

Table 2: Comparison between SVM models using multiple kernels and different sets of features.

Coordinates and terrain elevation

	Pre	Rec	F1	Acc
Albania	**0.88**	0.79	**0.84**	0.86
Belarus	0.96	**0.43**	**0.59**	**0.69**
Croatia	0.62	0.56	0.59	0.79
France	0.95	0.63	0.76	0.83
Germany	0.98	0.60	0.75	0.75
Greece	0.38	0.59	0.46	0.71
Ireland	0.74	0.84	0.78	0.79
Italy	0.84	0.71	0.77	0.89
Luxembourg	0.87	0.63	0.73	0.81
Switzerland	0.85	0.86	0.85	0.85
Ukraine	0.87	0.18	0.30	0.59

Coordinates and land coverage

	Pre	Rec	F1	Acc
Albania	0.84	0.78	0.81	0.83
Belarus	0.97	0.34	0.51	0.64
Croatia	0.47	0.67	0.55	0.71
France	0.93	0.74	0.82	0.86
Germany	**0.99**	0.64	0.77	0.78
Greece	0.33	0.65	0.43	0.64
Ireland	0.77	0.87	0.78	0.77
Italy	0.66	0.68	0.67	0.84
Luxembourg	**0.93**	0.62	0.74	0.83
Switzerland	**0.88**	0.85	0.86	**0.86**
Ukraine	0.83	0.17	0.28	0.59

Coordinates and population

	Pre	Rec	F1	Acc
Albania	0.87	0.82	**0.84**	**0.87**
Belarus	0.96	0.33	0.49	0.63
Croatia	**0.69**	0.54	0.61	**0.81**
France	**0.98**	0.65	0.78	0.85
Germany	**0.99**	0.65	0.79	0.79
Greece	0.57	0.66	0.61	0.82
Ireland	0.78	0.91	0.84	0.84
Italy	**0.98**	**0.78**	**0.87**	**0.94**
Luxembourg	0.88	0.62	0.73	0.81
Switzerland	0.87	0.86	0.86	**0.86**
Ukraine	**0.98**	0.14	0.25	0.59

Coordinates and data from all rasters

	Pre	Rec	F1	Acc
Albania	0.84	**0.83**	**0.84**	0.85
Belarus	0.99	0.37	0.54	0.66
Croatia	0.66	0.57	0.61	**0.81**
France	**0.98**	0.69	0.81	0.86
Germany	**0.99**	0.64	0.78	0.78
Greece	0.52	0.67	0.59	0.80
Ireland	0.78	0.90	0.84	0.84
Italy	0.97	0.73	0.84	0.93
Luxembourg	0.89	0.62	0.73	0.81
Switzerland	0.87	0.83	0.85	0.85
Ukraine	0.97	0.13	0.23	0.58

Coordinates and tags

	Pre	Rec	F1	Acc
Albania	0.83	0.82	0.83	0.84
Belarus	0.97	0.15	0.26	0.55
Croatia	0.42	**0.73**	0.54	0.66
France	0.95	**0.87**	**0.91**	**0.92**
Germany	**0.99**	**0.74**	**0.85**	**0.84**
Greece	0.36	**0.72**	0.48	0.66
Ireland	0.72	**0.92**	0.81	0.80
Italy	0.69	0.74	0.72	0.85
Luxembourg	0.83	**0.76**	**0.80**	**0.84**
Switzerland	0.80	**0.91**	0.85	0.84
Ukraine	0.90	**0.37**	**0.52**	**0.68**

Complete set of features

	Pre	Rec	F1	Acc
Albania	0.86	0.82	**0.84**	0.86
Belarus	**1.00**	0.27	0.42	0.61
Croatia	0.65	0.62	**0.64**	**0.81**
France	0.96	0.70	0.81	0.86
Germany	0.98	0.67	0.80	0.79
Greece	**0.59**	0.71	**0.64**	**0.83**
Ireland	**0.80**	0.90	**0.85**	**0.85**
Italy	0.97	0.73	0.84	0.93
Luxembourg	0.86	0.68	0.76	0.82
Switzerland	0.84	0.82	0.83	0.82
Ukraine	**0.98**	0.17	0.29	0.60

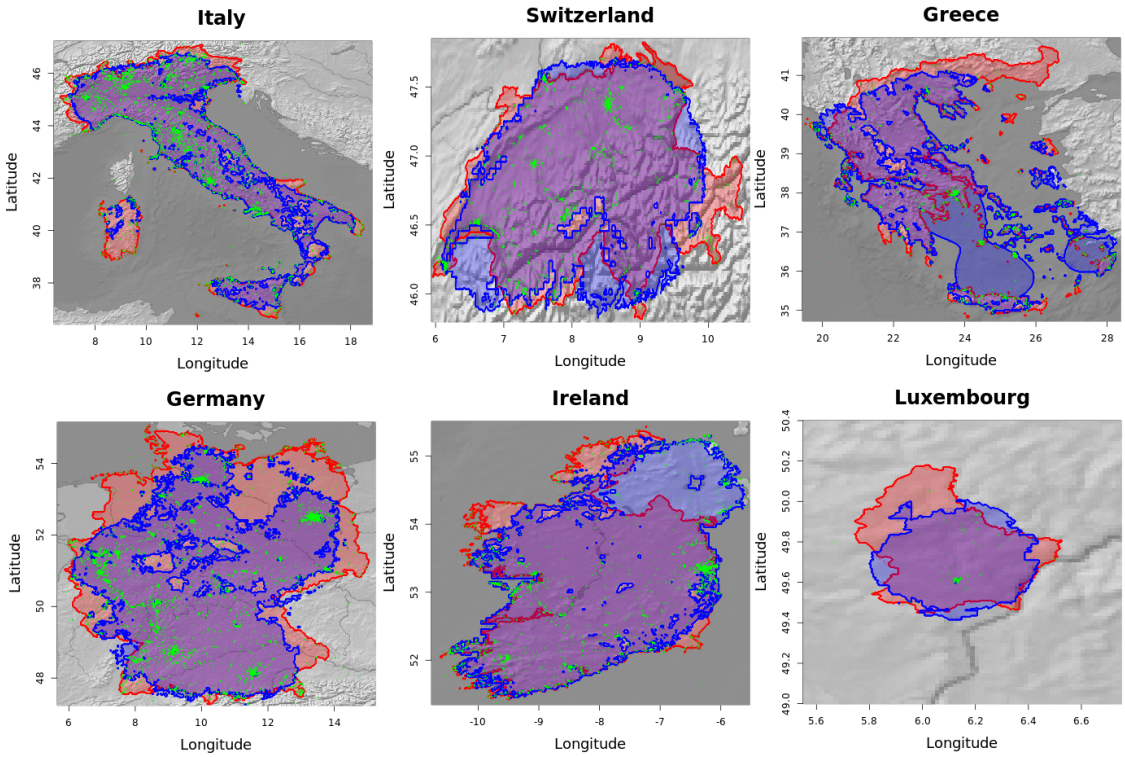

Figure 1: Results for six different countries using the complete set of features.

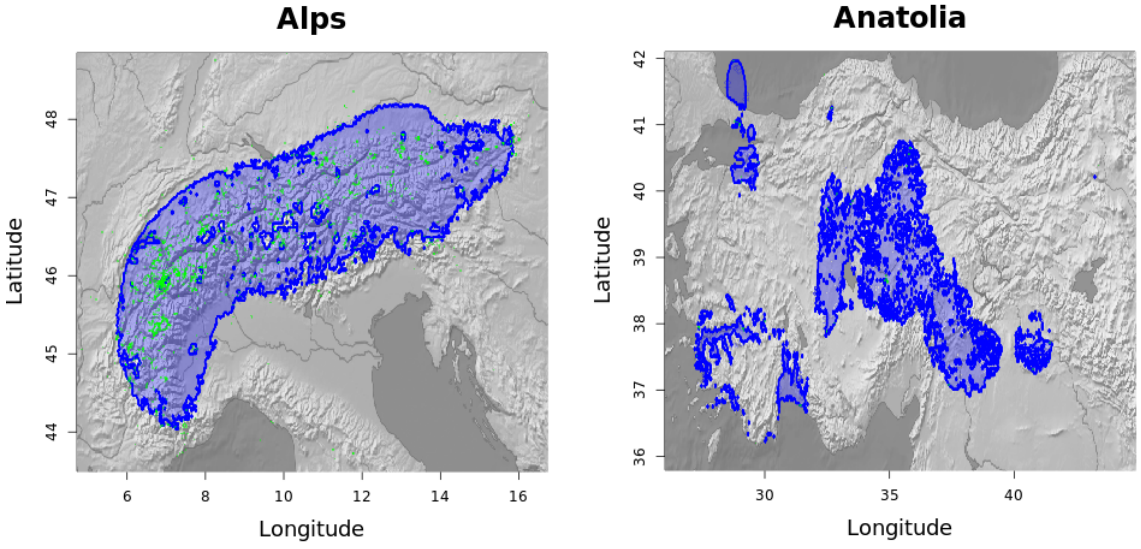

Figure 2: The obtained results for two large vague regions, namely the *Alps* and *Anatolia*.

4.2 Land Coverage Classification

On what concerns the land coverage classification task, I evaluated the proposed approach using a large collection of Flickr photos originally gathered in the context of the 2013 MediaEval Benchmarking Initiative for Multimedia Evaluation[12], a joint evaluation initiative focusing on new algorithms for multimedia access and retrieval. Several previous studies have collected their own subsets of Flickr photos, but I again argue that using a common dataset facilitates the reproducibility of the experiments reported here.

I focused on 4 separate regions that were chosen for their diversity in land-coverage types, and for the availability of many geo-tagged photos taken from within these regions in the MediaEval collection. The four regions correspond to metropolitan areas containing the cities of *London*, *New York*, *Paris* and *Rome*. The ground-truth land-coverage classes for the considered study regions were obtained from the Global Land Cover Facility (GLCF) dataset at a resolution of 1km per pixel. The availability of ground truth data makes it possible to evaluate the estimations with quantitative measures, for which I again used the traditional metric of accuracy, as well as macro-averaged recall, precision, and F_1 scores.

After collecting the initial set of photos for each study region, I have that each region was then divided into 4 quadrants, and I gathered a maximum of 375 photos from each quadrant, and for each land-coverage class. The idea was to produce balanced training datasets, containing 1500 examples for each land-coverage class, geospatially distributed over each region. After gathering photos from the different quadrants, if there are still particular land-coverage classes with less than 1500 example photos, I attempt to gather other examples associated to that particular class, from anywhere within the study region.

The two maps in Figure 3 show, for the regions of *London* and *New York*, the location of the photos in the balanced training set, in blue, and all the remaining photos for these two regions, in red. Notice that the remaining photos are used in the construction of the vectors that are to be classified, for each cell of the resulting land-coverage rasters.

Each of the training instances, selected according to the aforementioned procedure, is associated to a feature vector containing the geospatial coordinates from where the photo was taken, the elevation and population density at that particular geospatial position, the descriptions for the visual contents of the photo, and information regarding the occurrence of 300 specific tags. In my experiments, I then proceeded to using different sub-sets of these particular features.

In a first set of experiments, I compared the usage of a single-kernel multi-class SVM classifier, against the usage of multi-class SVM models combining three different Gaussian kernels, using only information from the geospatial coordinates as the features describing each position. I wanted to see if the proximity towards photos associated to a particular land-coverage class is enough to achieve a high accuracy. When

[12] `http://www.multimediaeval.org/`

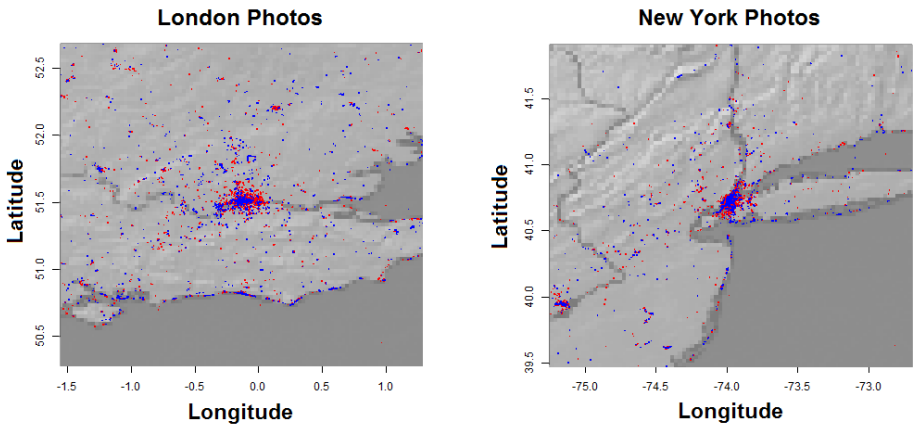

Figure 3: Distribution of training photos, in blue, and for all the remaining available photos, in red, for the metropolitan regions of *London* and *New York*.

combining multiple kernels, the three kernel widths were selected as $\sqrt{D_m}/2$, $\sqrt{D_m}$, and $2 \times \sqrt{D_m}$, where D_m is the dimensionality of the corresponding feature representation (i.e., $m = 2$ in this case).

Table 3 shows the obtained results, where we can see that the model that combines three kernels performs better in average that a more traditional SVM classifier. In terms of accuracy, we have that only for the zone of *Paris* the traditional SVM classifier model performed better, also obtaining the same result for the zone of *New York*, while the model with three kernels performed better for the zones of *London* and *Rome*.

Table 4 presents the obtained results over the same 4 regions that were considered in my first set of experiments, but now comparing different types of models that used geospatial coordinates together with

Table 3: Comparison between models using a single kernel or a combination of Gaussian kernels.

One kernel using only the geospatial coordinates

		Macro-Average			Per-Class F1			
Region	Accuracy	Precision	Recall	F1	Water	Forest	Land	Urban
New York	0.77	0.56	0.57	0.56	0.86	0.83	0.25	0.28
London	0.58	0.54	0.57	0.56	0.86	0.49	0.47	0.36
Paris	0.52	0.28	0.36	0.32	0.00	0.20	0.69	0.15
Rome	0.57	0.48	0.59	0.53	0.78	0.49	0.39	0.02

Three kernels using only the geospatial coordinates

		Macro-Average			Per-Class F1			
Region	Accuracy	Precision	Recall	F1	Water	Forest	Land	Urban
New York	0.77	0.58	0.58	0.58	0.87	0.83	0.25	0.29
London	0.61	0.55	0.58	0.57	0.90	0.53	0.47	0.34
Paris	0.49	0.29	0.36	0.32	0.03	0.34	0.63	0.12
Rome	0.60	0.45	0.63	0.52	0.81	0.42	0.48	0.05

Table 4: Comparison between SVM models using multiple kernels and three or more different sets of features.

Coordinates, rasters and tag features

		Macro-Average			Per-Class F1			
Region	Accuracy	Precision	Recall	F1	Water	Forest	Land	Urban
New York	0.78	0.58	0.59	0.58	0.86	0.83	0.14	0.47
London	0.48	0.52	0.56	0.54	0.90	0.27	0.40	0.23
Paris	0.64	0.35	0.40	0.38	0.08	0.18	0.79	0.19
Rome	0.63	0.48	0.61	0.54	0.95	0.36	0.44	0.02

Coordinates, rasters and image content features

		Macro-Average			Per-Class F1			
Region	Accuracy	Precision	Recall	F1	Water	Forest	Land	Urban
New York	0.74	0.56	0.50	0.53	0.77	0.81	0.15	0.34
London	0.58	0.53	0.57	0.55	0.79	0.48	0.47	0.40
Paris	0.56	0.26	0.40	0.31	0.00	0.12	0.71	0.29
Rome	0.57	0.46	0.58	0.51	0.83	0.47	0.21	0.02

Complete set of features

		Macro-Average			Per-Class F1			
Region	Accuracy	Precision	Recall	F1	Water	Forest	Land	Urban
New York	0.74	0.56	0.50	0.53	0.77	0.81	0.15	0.34
London	0.58	0.53	0.57	0.55	0.79	0.48	0.47	0.40
Paris	0.56	0.26	0.40	0.31	0.00	0.12	0.71	0.29
Rome	0.57	0.46	0.58	0.51	0.83	0.47	0.21	0.02

sets of other features (e.g., the textual tags, the visual content descriptors, or the information derived from external raster datasets encoding population or terrain elevation). These models combined the same three kernels from the first set of experiments, with three additional kernels that use the geospatial coordinates together with the different sets of other features. The kernel widths were also selected in the same manner, namely as $\sqrt{D_m}/2$, $\sqrt{D_m}$, and $2 \times \sqrt{D_m}$.

Figure 4 illustrates the obtained results for the geographic region of *New York*, placing side-by-side the ground-truth information and the estimate produced by the best-performing model. The red dots in the map correspond to the location of the training photos, and the four different types of terrain are represented in different colors. The class *water* is represented in blue, the class *forest* is represented in green, *land* is displayed in yellow, and *urban* terrain is pictured in gray. Figure 4 shows that the model has a good performance in the identification of the class *water*, represented in blue. In the zone of *New York*, the best-performing model was also able to identify the main areas of the class *urban*. However, these are still significant differences between the ground-truth maps and the estimates produced by my classification models.

5 Conclusions

The characterization of geographic regions assumes a particular importance in the context of urban planning, demographic studies, and several other topics related to the geographic information sciences. With the popularity of social web services and with their massive utilization, we have that large amounts of useful raw data are nowadays publicly available, which I claim that can be used to extrapolate geographic characteristics. In my dissertation, I presented two case studies, related to the characterization of geographic regions, that leverage on georeferenced photos from Flickr. These case studies provide good examples to my claim that information from social Web 2.0 services can be used to characterize geographic regions.

In a first set of experiments, which is detailed in Section 3.1, I evaluated an automated method, based on multiple kernel learning of one-class SVMs, for delimiting imprecise geographic regions with basis on publicly available data. The method uses one-class SVMs for interpolating from a set of points which are assumed to lie in the region that is to be defined. These points correspond to geospatial coordinates associated to Flickr photos tagged with the name of the vague region. Besides considering the geospatial coordinates for the points, I also considered a rich set of descriptive features obtained from population, elevation and land coverage raster datasets, as well as from textual tags. The overall approach for finding region boundaries was evaluated by means of statistical classification measures, using a set of 11 regions whose boundaries are well defined. The obtained results show that the refined method performs better than the simpler methods described in the literature, based solely on interpolating from geospatial coordinates (i.e., better than the previous study that was reported by Grothe and Schaab (2009), also based on one-class SVMs).

On another set of experiments, that was presented in Section 3.2, I have evaluated a method leveraging

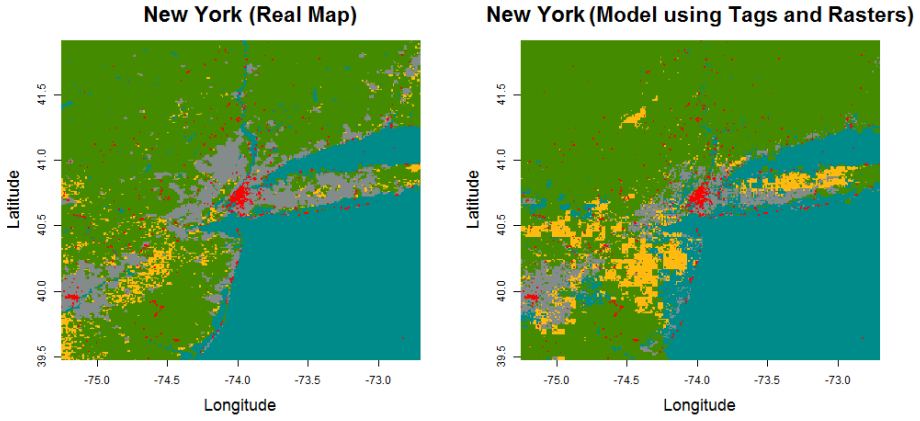

Figure 4: Estimates for the region of *New York paired with the ground-truth information.*

a large collection of georeferenced photos in order to estimate land-coverage maps. I described a collection of features for representing the images that can properly retain information related to land-coverage. With these features, I used multi-class SVMs, combining multiple Gaussian kernels, to predict the land-coverage classes. The evaluation results show that the proposed method was able to obtain an average accuracy of 62%, when considering 4 distinct land-coverage classes. While unlikely to replace satellite imagery in the short term, the results demonstrate the plausibility of leveraging the vast collections of georeferenced photos, existing on repositories such as Flickr, for large-scale monitoring of geospatial properties.

Acknowledgments

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT), under project grants PTDC/EIAEIA/109840/2009 (SInteliGIS), EXCL/EEI-ESS/0257/2012 (DataStorm), EXPL/EEI-ESS/0427/2013 (KD-LBSN), and PEst-OE/EEI/LA0021/2013 (INESC-ID multi-annual funds).

References

Alani, H., Jones, C., and Tudhope, D. (2001). Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science*, 15(4).

Arampatzis, A., van Kreveld, M., Reinbacher, I., Jones, C. B., Vaid, S., Clough, P., Joho, H., and Sanderson, M. (2006). Web-based delineation of imprecise regions. *Computers, Environment and Urban Systems*, 30(4).

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9).

Bolettieri, P., Esuli, A., Falchi, F., Lucchese, C., Perego, R., Piccioli, T., and Rabitti, F. (2009). CoPhIR: a test collection for content-based image retrieval. *CoRR*, abs/0905.4627v2.

Brunsdon, C. (1995). Estimating probability surfaces for geographical point data: An adaptive kernel algorithm. *Computers & Geosciences*, 21(7).

CIESIN and CIAT (2005). Gridded Population of the World, Version 3 (GPWv3) Data Collection.

Clough, P. and Pasley, R. (2010). Images and perceptions of neighbourhood extents. In *Proceedings of the ACM Workshop on Geographic Information Retrieval*.

Edgington, E. S. (1969). Approximate randomization tests. *Journal of Psychology: Interdisciplinary and Applied*, 72(2).

Goldberg, D. W., Wilson, J. P., and Knoblock, C. A. (2009). Extracting geographic features from the internet to automatically build detailed regional gazetteers. *International Journal of Geographical Information Science*, 23(1).

Gönen, M. and Alpaydin, E. (2011). Multiple kernel learning algorithms. *Journal of Machine Learning Research*, 12.

Grady, D., Brune, R., Thiemann, C., Theis, F. J., and Brockmann, D. (2012). Modularity maximization and tree clustering: Novel ways to determine effective geographic borders. In Thai, M. T. and Pardalos, P. M., editors, *Handbook of Optimization in Complex Networks, Theory and Applications*.

Grothe, C. and Schaab, J. (2009). Automated Footprint Generation from Geotags with Kernel Density Estimation and Support Vector Machines. *Spatial Cognition & Computation*, 9(3).

Hauff, C., Thomee, B., and Trevisiol, M. (2013). Working Notes for the Placing Task at MediaEval 2013. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop*.

Hettich, R. and Kortanek, K. O. (1993). Semi-infinite programming: Theory, methods, and applications. *SIAM Review*, 35(3).

Hu, S. and Wang, L. (2013). Automated urban land-use classification with remote sensing. *International Journal of Remote Sensing*, 34(3).

Jacobs, N., Burgin, W., Fridrich, N., Abrams, A., Miskell, K., Braswell, B. H., Richardson, A. D., and Pless, R. (2009a). The global network of outdoor webcams: properties and applications. In *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.

Jacobs, N., Burgin, W., Speyer, R., Ross, D., and Pless, R. (2009b). Adventures in archiving and using three years of webcam images. In *Proceedings of the IEEE CVPR Workshop on Internet Vision*.

Jacobs, N., Roman, N., and Pless, R. (2007). Consistent temporal variations in many outdoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Jones, C. B., Purves, R. S., Clough, P. D., and Joho, H. (2008). Modelling vague places with knowledge from the web. *International Journal of Geographical Information Science*, 22(10).

Munoz, A. and Moguerza, J. M. (2006). Estimation of high-density regions using one-class neighbor machines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3).

Murdock, C., Pless, R., and Jacobs, N. (2013). Webcam2satellite: Estimating cloud maps from webcam imagery. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*.

Newman, D. (2006). The lines that continue to separate us: borders in our 'borderless' world. *Progress on Human Geography*, 30(2).

Riordan, E., Graham, E., Yuen, E., Estrin, D., and Rundel, P. (2010). Utilizing public internet-connected cameras for a cross-continental plant phenology monitoring system. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*.

Schockaert, S. (2011). Vague regions in geographic information retrieval. *SIGSPATIAL Special*, 3(2).

Sonnenburg, S., Rätsch, G., Schäfer, C., and Schölkopf, B. (2006). Large scale multiple kernel learning. *Journal of Machine Learning Research*, 7.

Vasardani, M., Winter, S., and Richter, K.-F. (2013). Locating place names from place descriptions. *International Journal of Geographical Information Science*, 27(12).

Vatsavai, R. R., Bright, E., Varun, C., Budhendra, B., Cheriyadat, A., and Grasser, J. (2011). Machine learning approaches for high-resolution urban land cover classification: a comparative study. In *Proceedings of the International Conference on Computing for Geospatial Research & Applications*.

Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review*, 22(176).

Xie, L., Carreira-Perpinan, M. A., and Newsam, S. (2010). Semi-supervised regression with temporal image sequences. In *Proceedings of the IEEE International Conference on Image Processing*.