University of Heidelberg, Institute of Geography

ANALYSIS OF TOURIST ACTIVITY BASED ON THE TRACKING DATA COLLECTED BY FLICKR

Yeran Sun, Mohamed Bakillah

Abstract: Volunteered Geographic Information (VGI) provides valuable information to analyze human activities in space and time. In this paper, we develop an integral framework for the study of tourist activity. Flickr photo is chosen as an example to explore the potential of using VGI to analyze spatiotemporal patterns of tourist activity. Munich, Germany is used as study site to conduct the experiments. After a state-of-the-art of tourism activity research, this article employs the Kernel density estimations and spatial scan statistics to explore the distribution of photos. Seasonality of tourism is considered as well to make a temporal analysis. Besides, the influences of both important locations and events are considered in spatial cluster generation. Experimental results show that spatial pattern of tourist activities varies in different months. In addition to famous tourism places, some kinds of events can result in spatial clusters of tourist activities in some months.

Keywords: Volunteered Geographic Information (VGI), Flickr, tourist activity, seasonality, event

ANALYSE TOURISTISCHER AKTIVITÄTEN BASIEREND AUF VON FLICKR GESAMMELTEN TRACKINGDATEN

Zusammenfassung: Freiwillig bereitgestellte geographische Informationen (Volunteered Geographic Information – VGI) sind eine wertvolle Datenquelle, um menschliche Aktivitäten in Raum und Zeit zu analysieren. Dieser Beitrag stellt einen integralen Rahmen zur Untersuchung touristischer Aktivitäten vor, der mithilfe von Flickr-Fotos beispielhaft die Möglichkeiten der VGI zur Analyse von raumzeitlichen Mustern in München als Studienort untersucht. Nach einem Überblick zum Stand der Forschung touristischer Aktivitäten werden die Kerndichteschätzung und räumliche Scan-Statistiken behandelt, die verwendet wurden, um die Verteilung der Fotos zu sondieren, wobei zusätzlich saisonale Aspekte betrachtet wurden. Dieser Beitrag führt zu einem Ansatz, der Standort- und ereignisbasierte Analysen einschließt.

Schlüsselwörter: Volunteered Geographic Information (VGI), Flickr, touristische Aktivitäten, Saisonalität, Ereignis

Authors

Yeran Sun
Dr. Mohamed Bakillah
University of Heidelberg
Institute of Geography
Berliner Straße 48
D-69120 Heidelberg
E: yeran.sun@geog.uni-heidelberg.de
mohamed.bakillah@geog.uni-heidelberg.de

1 INTRODUCTION

Volunteered Geographic Information (VGI) describes geographic information that has been voluntarily gathered, provided, and shared by community members in social media (Goodchild 2007a). Popular examples of VGI include Wikimapia, Flickr, OpenStreetMap. Because it is regularly updated and easily available via application programming interfaces, VGI has gained popularity for spatial scientists, such as geographers, to understand the surface of the Earth and human behavior (Goodchild, 2007b). VGI also contains temporal information of objects (e. g., photos, messages, videos), which can be used to analyze user behaviors in a space-time context. For instance, Flickr and YouTube provide tools for users to edit photos and videos by means of geotags[1]. Geotagged photos and videos are georeferenced by means of the coordinates generated by the camera, smartphone, etc. These georeferenced photos and videos can serve as data source to explore human (user) behavior.

Aiming to analyze tourist activities, statistics data and survey data were the most prevailing. However, for tourist activities sometimes the availability of tourism data is still a major research obstacle. It is always time-consuming and expensive to gather and preprocess statistics data and survey data. In contrast, VGI such as Flickr photos can be a new data source to analyze tourist activities. The reasons are manifold: low price, high convenience, and easy preprocessing, among others. Thus, to investigate the possibilities of VGI on human behavior analysis, the main objective of this paper is to explore whether and how we can use geotagged photos of Flickr, to examine spatiotemporal patterns of tourist activities. Additionally, we analyze the possible effect of seasonality by means of temporal analysis. Furthermore, influences of both place and event are considered in spatial cluster detection.

The remainder of this paper is organized as follows. The next section reviews previous work dealing with tourism analysis using georeferenced images. Subsequently, the methodology, study area, and data processing are introduced and important results are presented as well. Finally, this paper makes conclusions and presents the future works.

2 RELATED WORK

Recent studies illustrate the high potential of VGI to examine tourist activities. Regardless of spatial and temporal characteristics of users, Julio et al. (2007) as well as Lindgren & Lundstrom (2011) focus on the statistical analysis of social relationships and interactions, while others only take the temporal changes into account. Examples are the monitoring of users' lifestyle made by Lee et al. (2011) who extract user' behavioral patterns in urban areas from Twitter. Using Flickr data, Popescu & Grefenstette (2009) examine temporal attributes of tourist sites, such as the temporal length of visits as well as the minimum, maximum, and average visiting times. In terms of statistical analysis Zwol (2007) investigates the temporal, social and spatial behaviors of Flickr's users. Results show that a) users discover new photos within hours after being uploaded and b) the spatial distribution of views on infrequently viewed photo is more concentrated around a geographic location.

Subsequently, more researchers paid attention on the spatiotemporal pattern of tourist activities. Using Flickr Girardin et al. (2007) visualized the point density of tourists, subsequently detecting tourist concentrations. After that, Girardin et al. (2008) uncovered the spatial and temporal presence of tourists during their visit to Rome. Subsequently, Girardin et al. (2009) measured the attractiveness of POIs based on the presence of photographers. To explore the representative trajectory patterns, Yin et al. (2011) generated many trajectories and ranked the trajectory patterns quantitatively by means of an exemplar-based algorithm. Using Panoramio geotagged photos, Andrienko et al. (2009) built a flow map aiming to show aggregated moves (collective movement patterns) of photographers between different places. Besides, Hollenstein & Purves (2010) leveraged geotagged photos of tourists to describe city core.

Considering the spatial and temporal aspects of VGI, event detection becomes a popular research field. Rattenbury et al. (2007) present an approach to extract event semantics from Flickr image tag. Model comparisons (scale-structure identification (SCI) vs. standard burst detection methods) underpin that SCI outperforms the naive scan method as well as the spatial scan method. Chen & Roy (2009) employ a wavelet transform to suppress noise by identifying the tags related with events. Afterwards, event-related tags of Flickr are clustered such that each cluster, representing an event, consists of tags with similar spatial and temporal patterns. Andrienko et al. (2009) investigate visual analytics to combine computational techniques with interactive visual displays. Kisilevich et al. (2010) extend this work by applying systematic visual analytics to accomplish different tasks in the event-based analyses of image data. Additionally, geomessages of Twitter and geotags of YouTube videos are leveraged as well. Lee et al. (2011) present a geosocial event detection approach to detect crowded places based on the conception of social networking sites. Tahayna et al. (2011) propose a technique (denoted as GAoptSVM) aiming at an optimal SVM-based video event classification via an evolutionary optimization technique. Thom et al. (2012) present an approach that allows for a real-time interactive analysis of geotagged microblog messages. A cluster analysis approach is used to detect spatiotemporal anomalies automatically.

3 METHODOLOGY

The Kernel Density Estimation (KDE) can present the spatial pattern of tourist activities visually by means of generating local clustering regions. After that, we apply spatial scan statistic method to detect significant spatial clusters. KDE is able to produce local estimates of densities, resulting in a smooth density surface of tourist activities. KDE helps to identify locations of possible "cluster" or at least sub-regions worth for further examination (Gatrell et al. 1996). Normally the Gaussian kernel function is employed.

Based on the empirical results presented by Helbich & Leitner (2012), the Spatial Scan Statistic (SSS) (Kulldorf 1997) is chosen as cluster detection method. Compared to other clustering methods, the SSS method detects local clusters and tests if such clusters occur by chance (Kulldorf 1999). Following Helbich (2012), this paper chooses the Bernoulli model (Kulldorff 1997) to detect significant local spatial cluster compared to a heterogeneously distributed control population.

Besides, in this paper we take event effect into consideration, with the aim to

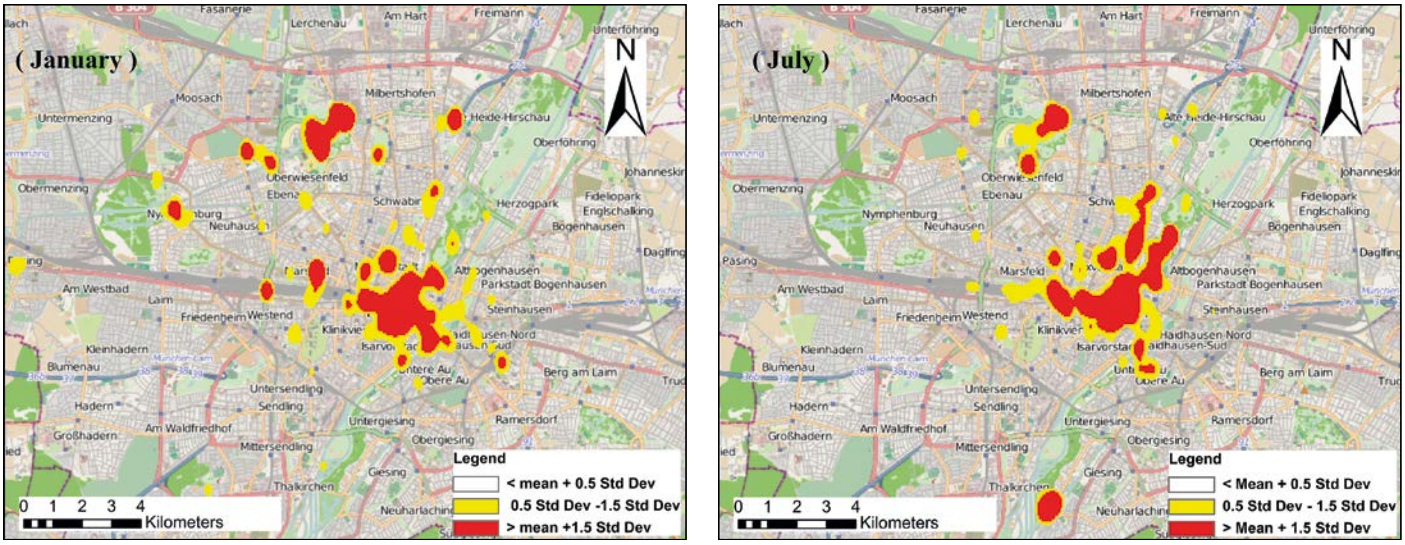

Figure 1: KDE results of tourist activities in January and July (bandwidth = 0.3 km) (Note: "0.5 Std Dev – 1.5 Std Dev" represents "Mean + Std Dev – Mean + 1.5 Std Dev")

examine the influences of social events on tourist activities in terms of density estimation and cluster detection.

4 EXPERIMENTAL RESULTS

4.1 DATA

In this paper, the study case is Munich, Germany. Using a bounding box, we acquired the georeferenced images taken within Munich from 2010 and 2011 via Flickr API. There are 21,404 and 24,547 geo-referenced images taken in 2010 and 2011 respectively. Considering the user bias and errors, we preprocessed the image set using the method of Grothe & Schaab (2009). After that, we distinguished the images taken by tourists with the one taken by residents using a method proposed by Girardin et al. (2007). At last, the image set of tourist activities was transformed into point set. Here, we focus on the data in 2011. In addition, we use the point set of 2010 as the control population in the cluster detection.

4.2 RESULTS OF KDE

The KDEs of tourist activities were calculated with different bandwidths ranging from 1 to 500 m. The generated clustering is distributed discretely when the bandwidth is too small (e. g. 10 m). On the contrary, local clustering is easily missed when the bandwidth is too large (e. g. 500 m). Here we showed the results at a bandwidth of 300 m which is a medium level. Figure 1 displays the KDE results of tourist activities in January and July. We can see that there is clustering of tourist activity in both months, and in addition the distribution of tourist activity is slightly different between these two months despite that the cores of both are in the city center.

Figure 2 shows the change of the mean as well as the standard deviation value of KDE in 2011. If the mean value of KDE is higher, the point intensity of tourist activity will be stronger. And if the standard deviation value of KDE is higher, the point clustering of tourist activity will be stronger. The shifty curve indicates the seasonality of tourist activities. Additionally, we also present the change of the image amount (see Figure 3). Globally, the curve of point number has a similar tendency with that of the mean of the KDE as well as the standard of the KDE. Specifically, both of the curves are low in January and subsequently become high in April. At last, they come to the maximum in September and October. However, in July, the number of images is relatively high, while the mean and standard deviation of KDE are relatively low. It indicates that the intensity of clustering in July is relatively low. In other words, in July the clustering of tourist activities is distributed more homogenously.

To explore the possible influence of events on the spatial pattern of tourist activities, we focus some important events in each month. To be simple, such important events here are the ones owning the most geotags. Figure 4 shows the tag numbers of the two most important events in January, April, July and October. We can see that the tag numbers of the events in July and

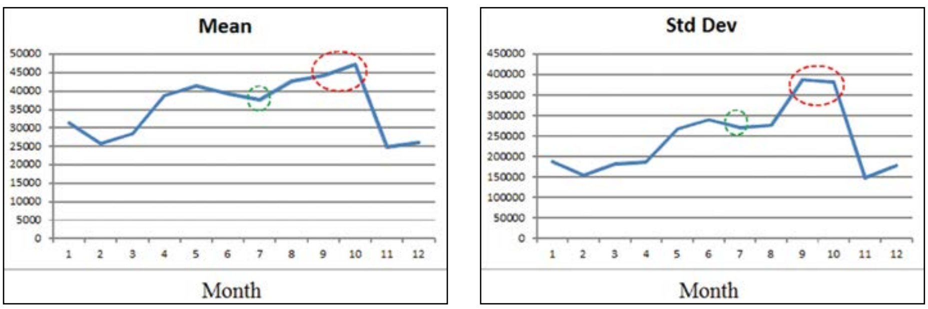

Figure 2: Mean and Standard Deviation of the KDE in each month of 2011 (bandwidth = 0.3 km)

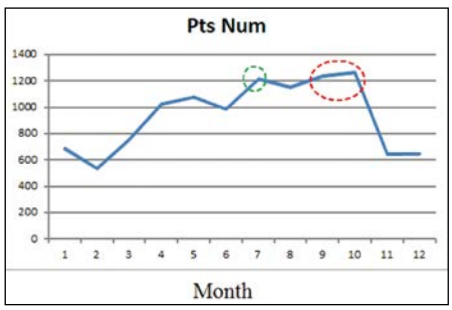

Figure 3: Change of the image amount

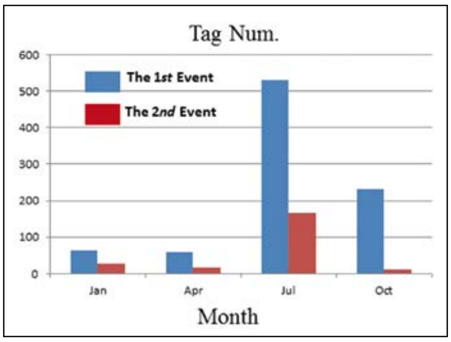

Figure 4: Number of the event-related tags within Munich in 2011

October are higher than that in January and April. Considering the curve of KDE, the changes in tag number confirm that some events can trigger the increase of the images and influence the spatiotemporal pattern of tourist activities. These events are shown in Table 1.

4.3 RESULTS OF SSS

Via the KDE analysis, we find out some interesting local regions. Subsequently, using the SSS method, we conduct the cluster detection in these four months. Some significant clusters (*p-value* < 0.05) are detected and the primary clusters (Most likely clusters) in the four months are shown in Table 2. Significance values (*p-values*) are based on 999 Monte Carlo simulations. Not surprisingly, these primary clusters are located in the downtown. However, their sizes are clearly different. The primary cluster in October is the largest one with the most observed points (830 cases) and second largest radius (2.712 km) of scan window. And the one in July is the smallest one bearing the fewest number of observed points (60 cases) and smallest radii of scan window (0.398 km). On the other hand, there are only 2 significant clusters detected in October and 11 clusters detected in July globally. By means of comparing the spatial coverage of the cluster, we can see that tourist activities clustered within two large regions in October while they clustered within many small regions in July.

To investigate the possible reasons resulting in the clustering of tourist activities, we analyze the images of the detected significant clusters. In terms of the text (tag and title) as well as sometimes imagery content, we can find out two main reasons. The first one is that some places (e. g. tourism attractions) attract a vast amount of tourists. The second one is that some events (e. g. parades) result in the gathering of tourists within a certain time period. Table 3 shows the number of significant clusters as well as the main reasons of these clusters. We can see that some events (e. g. Improv

Month	The 1st event	Tags	The 2nd event	Tags
January	Ski World Cup	64	Red Bull Crashed Ice	26
April	Workshop TUI Cruises	58	City Bike Marathon	16
July	Election of the 2018 Host City Winter Olympics	530	Christopher Street Day	167
October	Oktoberfest	231	one academic lecture	12

Table 1: The tag amount of the important events in the four months

Month	Observed cases	Expected cases	Observed expected ratio	Log likelihood ratio	*p*-value	Circle radius (km)
January	123	32.557	3.778	89.564	<0.001	2.071
April	106	39.933	2.654	46.039	<0.001	2.712
July	60	9.547	6.285	83.448	<0.001	0.398
October	830	626.345	1.325	75.148	<0.001	2.483

Table 2: Statistics of the detected primary cluster (most likely cluster)

Month	Significant clusters	Place	Event
January	3	Highlight Towers	–
April	7	New Botanical Garden; BMW World	Improv Festival
July	11	Bavarian National Museum; Prince Regent's Theatre; Bavarian State Chancellery	Christopher Street Day; Tollwood Summer Festival
October	2	Nymphenburg Palace	–

Table 3: Statistics of the detected significant clusters (p-value < 0.05)

Festival) have significant influences on tourist activities and trigger clustering in April and July.

4.4 DISCUSSION

In terms of analysis based on the tracking data of Flickr, we can explore the spatial and temporal patterns of tourist activities. KDE is dedicated to describe the spatial pattern and SSS is utilized to detect the cluster. The temporal analysis indicates that the KDE changes seasonally. It is demonstrated that the spatial distribution of tourist activities is heterogeneous and some hotspot regions exist. The local spatial clustering in September and October is obviously higher than that in November and February. On the other hand, in July the clustering of tourist activities is distributed more homogenously. Subsequently, by means of SSS, some significant clusters are detected. There exists clearly seasonal difference on the size as well as the number of cluster. These detected clusters are prevailingly located in some important tourism attractions (e. g. museum and palace).

Additionally, the event is considered in analyzing the possible reasons of clustering of tourist activities. Several great social events have influences on the spatiotemporal pattern of tourist activities. However, compared to the hotspot places, events have relatively low effects. Besides, the events, which can result in clustering of tourist, are the ones owing lots of participators and a relatively long time period (e. g. Tollwood Summer Festival is more than 3 weeks). At the same time, these events always contain parades, demonstrations, etc. which took place at important streets. Obviously, some of events owning most tags (e. g. Ski World Cup, Workshop TUI Cruises, etc. in Table 1) could not result in significant spatial clusters. These events always have relatively large or small spatial coverage. For instance, *City Bike Marathon* has a large spatial coverage while *Workshop TUI Cruises* only takes place within one building.

Nevertheless, the difference between users' preferences as well as their photographing behaviors might have impacts on the result of KDE. For instance, some tourists like to take pictures everywhere or take series of pictures in one place, and other might take only one picture on each place or even never take any picture in some places. Also, some users uploaded lots of images while other users only shared few images. Thus, it will be better to consider the photographing behavioral pattern of Flickr users when analyzing the spatiotemporal pattern of tourist activities.

5 CONCLUSION AND FUTURE WORK

In this paper, we proposed an integral framework to analyze tourist activities in space and time. Flickr geo-referenced images were utilized as the tracking data of tourist activities. The results demonstrate good potential of the proposed approach to analyse tourist activities. In addition, seasonality of tourism as well as the event influence are taken into consideration. And, the experimental results demonstrate that spatial pattern of tourist activities varies in different months. In addition to famous tourism places, some kinds of events (e. g. carnival, demonstration, exhibition, etc.) might result in spatial clusters of tourist activities in some months. After the discussion above, the characteristics of the events as well as the photographing behavior pattern of the users should be considered in the future work. Furthermore, to support result of the analysis, other VGI data (e. g. Twitter, Facebook and OpenStreetMap) might be integrated into the analysis as well.

[1] http://geotag.sourceforge.net

Hrsg.: Strobl, Josef / Blaschke, Thomas / Griesebner, Gerald / Zagel, Bernhard

Wichmann

Angewandte Geoinformatik 2013

Beiträge zum 25. AGIT-Symposium Salzburg

2013, XV, 650 Seiten

ISBN 978-3-87907-533-1

82,– € Subskriptionspreis bis zum 15.07.2013: 69,– €

Preisänderungen und Irrtümer vorbehalten. Es gelten die Liefer- und Zahlungsbedingungen des VDE VERLAGs.

Jetzt gleich hier bestellen: **www.vde-verlag.de/130701**

www.wichmann-verlag.de

VDE VERLAG GMBH

Bismarckstr. 33 · 10625 Berlin · Tel.: (030) 34 80 01-222

Fax: (030) 34 80 01-9088 · kundenservice@vde-verlag.de

VDE VERLAG

References

Andrienko, G.; Andrienko, N.; Bak, P.; Kisilevich, S.; Keim, D. (2009): Analysis of Community-contributed Space- and Timereferenced Data by Example of Panoramio Photos. In: Vision, Modeling, and Visualization Workshop, Braunschweig, 2009, pp. 323-330.

Chen, L.; Roy, A. (2009): Event detection from flickr data through wavelet-based spatial analysis. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, Hong Kong, China, pp. 523-532.

Gatrell, A.; Bailey, T.; Diggle, P.; Rowlingson, B. (1996): Spatial point pattern analysis and its application in geographical epidemiology. In: Transactions of the Institute of British Geographers, 21, pp. 256-274.

Girardin, F.; Calabrese, F.; Dal Fiore, F.; Ratti, C.; Blat, J. (2008): Digital footprinting: Uncovering tourists with user-generated content. In: IEEE Pervasive Computing, 7 (4), pp. 36-43.

Girardin, F.; Dal Fiore, F.; Ratti, C.; Blat, J. (2007): Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. In: Journal of Location Based Services, 2 (1), pp. 41-56.

Girardin, F.; Vaccari, A.; Gerber, A.; Ratti, C. (2009): Quantifying urban attractiveness from the distribution and density of digital footprints. In: Journal of Spatial Data Infrastructure Research, 4, pp. 175-200.

Goodchild, M. F. (2007a): Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. In: International Journal of Spatial Data Infrastructures Research, 2, pp. 24-32.

Goodchild, M. F. (2007b): Citizens as sensors: the world of volunteered geography. In: GeoJournal, 69 (4), pp. 211-221.

Grothe, C.; Schaab, J. (2009): Automated Footprint Generation from Geotags with Kernel Density Estimation and Support Vector Machines. In: Spatial Cognition & Computation, 9 (3), pp. 195-211.

Helbich, M. (2012): Beyond Postsuburbia? Multifunctional Service Agglomeration in Vienna's Urban Fringe. In: Journal of Economic & Social Geography, 103 (1), pp. 39-52.

Helbich, M.; Leitner, M. (2012): Evaluation of Spatial Cluster Detection Algorithms for Crime Locations. In: Gaul, W.; Geyer-Schulz, A.; Schmidt-Thieme, L.; Kunze, J.: Challenges at the Interface of Data Analysis, Computer Science, and Optimization. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, pp. 193-201 (DOI: 10.1007/978-3-642-24466-7_20).

Hollenstein, L.; Purves, R. S. (2010): Exploring place through User Generated Content: Using Flickr tags to describe city cores. In: Journal of Spatial Information Science, 1, pp. 21-48.

Julio, G.; Jussi, K.; Paul, C. (2007): iCLEF 2006 Overview: Searching the Flickr WWW photo-sharing repository. In: Evaluation of Multilingual and Multi-modal Information Retrieval, 7th Workshop of the Cross-Language Evaluation Forum (CLEF 2006), Alicante, Spain.

Kisilevich, S.; Krstajić, M.; Keim, D.; Andrienko, N.; Andrienko, G. (2010): Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. In: Proceedings of 14th International Conference on Information Visualization, London, pp. 289-296.

Kulldorff, M. (1997): A spatial scan statistic. In: Communications in Statistics: Theory and Methods, 26, pp. 1481-1496.

Kulldorff, M. (1999): Spatial scan statistics: Models, calculations and applications. In: Balakrishnan, N.; Glaz, J.-B. (Eds): Recent Advances on Scan Statistics and Applications. Birkhäuser, Boston, USA.

Lee, R.; Wakamiya, S.; Sumiya, K. (2011): Discovery of unusual regional social activities using geotagged microblogs. In: World Wide Web-Internet and Web Information Systems, 14 (4), pp. 321-349.

Lindgren, S.; Lundstrom R. (2011): Pirate culture and hacktivist mobilization: The cultural and social protocols of #WikiLeaks on Twitter. In: New Media & Society, 13 (6), pp. 999-1018.

Popescu, A.; Grefenstette, G. (2009): Deducing trip related information from flickr. In: Proceedings of the 18th international conference on World wide web, Madrid, Spain, ACM, pp. 1183-1184.

Rattenbury, T.; Good, N.; Maaman M. (2007): Towards automatic extraction of event and place semantics from flickr tags. In: Proceedings of the 30th Annual International Conference on Research and Development in Information Retrieval. Amsterdam, the Netherlands, pp. 103-110.

Tahayna, B.; Belkhatir, M.; Alhashmi, S. M.; O'Daniel, T. (2011): Evolutionary optimization of video event classification. In: International Journal of Computer Mathematics, 88 (18), pp. 3784-3802.

Thom, D.; Bosch, H.; Koch, S.; Worner, M.; Ertl, T. (2012): Spatiotemporal Anomaly Detection through Visual Analysis of Geolocated Twitter Messages. In: 2012 IEEE Pacific Visualization Symposium (PacificVis), pp. 41-48.

Yin Z.; Cao L.; Han J.; Luo J.; Huang T. (2011): Diversified Trajectory Pattern Ranking in Geotagged Social Media. Proceedings of the Eleventh SIAM International Conference on Data Mining, SDM 2011, April 28-30, 2011, Mesa, Arizona, USA, pp. 980-991.

Zwol, R. (2007): Flickr: Who is Looking? In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society, pp. 184-190.