# A comparative study of spatial, temporal and content-based patterns emerging in YouTube and Flickr

Milan Mirkovic*, Dubravko Culibrk*, Symeon Papadopoulos†, Christos Zigkolis†,
Yiannis Kompatsiaris†, Gavin McArdle‡, Vladimir Crnojevic*
*Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
Email: mmirkov@uns.ac.rs
†Informatics and Telematics Institute - Centre for Research and Technology Hellas, Thessaloniki, Greece
Email: papadop@iti.gr
‡National Centre for Geocomputation, National University of Ireland Maynooth, Dublin, Ireland
Email: gavin.mcardle@nuim.ie

***Abstract*—Due to the recent advances and wide adoption of Web 2.0 technologies, there is an abundance of publicly available user generated content, which can be a valuable resource for researchers, enabling them to apply sophisticated analysis methods on data of unprecedented scale.**

**This paper focuses on the detection and comparison of spatial, temporal and content-based patterns from user-generated content available through two major content-sharing services – YouTube and Flickr – aiming to explain some of the observed differences. In particular, we use different clustering approaches and visualization techniques to discover interesting spots in a city using the publicly available content (images, videos and associated metadata), we then classify them as either landmarks or events, and compare their ranking for each of the respective service.**

**Based on that ranking, we confirm the expected pattern that people tend to take pictures of still objects (monuments, sights) and make videos of events. We also consider the temporal aspect of the data, and extract movement trajectories of users. Lastly, we discover that even single users can generate noticable patterns, and that people find diverse uses of these content-sharing services.**

***Keywords*-clustering; youtube; flickr; spatio-temporal mining**

## I. Introduction

Technological advances frequently affect the habits of people in profound ways; this is especially true in the case of mobile devices. The ubiquity and versatility of mobile components (cameras, wireless and GPS modules, accelerometers, gyroscopes, etc.) has made them pervasive in the users' daily lives.

Such abundance of devices being capable of creating all sorts of content (from plain text documents, over images and audio recordings to videos and more) has inevitably led to an enormous increase in the amount of data generated in any given time interval. Telecommunication advances have made Internet access ubiquitous and cheap, easing the online sharing of content. This, coupled with plummeting prices of storage, has led to creation of services that make it possible to share most of the content created by the end-users. Some of those services have become known world-wide, and are used every day by millions of people, who share the content they make with the online community.

With so much content at hand, researchers have devised new data analysis and exploration methodologies in order to understand real-world phenomena as they are reflected in the online data. Detecting movement trajectories [1], performing temporal analysis related to habits of usage of a particular service and uncovering social patterns and relations among the users [2], [3] are just some of the research topics that have been actively pursued in the past. Building on top of them, this paper aims to identify spatial, temporal and content-based patterns that emerge among YouTube and Flickr users, and to compare them in order to detect differences and try to explain them. Thus, the main contribution of the paper stems from the comparative analysis between two different content sharing communities (both in terms of content and in terms of practices) that cover the same real-world place.

The paper is organized as follows: Section II presents several related efforts. Section III describes the data used. Section IV discusses methods and techniques used for data visualization and analysis, and Section V presents the results obtained and interprets them. Section VI gives a brief discussion on further research possibilities and concludes the paper.

## II. Related work

There has recently been significant research interest in the large-scale analysis of multimedia content created and shared online through services such as Flickr and YouTube. Our work makes use of methods and tools stemming from three research topics: (a) urban data mining, (b) landmark and event detection, and (c) motion pattern analysis and itinerary construction. Below, we briefly review each of these topics. Compared to related research efforts, the main novelty of our work is the comparative study of two different online content communities, both in terms of content type (photos vs. videos) and in terms of content creation and sharing practices, for the detection of interesting patterns.

**Urban data mining:** Girardin et al. [4] construct heat maps from geo-referenced data, which are then used as overlays in Google Earth in order to identify concentration of tourists in certain areas. They also identify inbound and outbound trajectories of tourists for the wider region of city of Florence (and later the province of Florence). In their subsequent work [5], the authors combine geospatial data obtained from Flickr with data obtained from a local telecom company to separate tourists from residents in the city of Rome, and visually inspect their activity (movements and concentration around points of interest). Mirkovic et al. [6] perform (visual) density analysis of YouTube data labelled as recorded in the central Serbia region, in order to discover attractive places.

**Landmark and event detection:** Quack et al. [7] identify landmarks and events by means of a two-level clustering approach, relying on spatial features for the first level of clustering and on visual similarities for the second level. They also map the extracted clusters to Wikipedia articles. Similarly, Papadopoulos et al. [8] conduct a hybrid clustering on the photos captured within a particular place (city) with the goal of identifying and mining landmarks and events. The hybrid clustering method they use takes into account both visual and tag-based similarities between the photos of the collection.

**Motion pattern analysis and itinerary detection:** Popescu et al. [9] mine trip-related information, including visit times, interior vs. exterior and panoramic views, from a corpus of more than 70 million Flickr photos and by making use of geo-located Wikipedia entries. Andrienko et al. [10], in addition to spatial analysis of Flickr and Panoramio geo-referenced content, use temporal information to build flow maps that do not only show the characteristic routes, but also the movement direction. Mirkovic et al. [11] use YouTube geo-referenced data to discover user behavior patterns in a case study to compare these patterns between the countries of Serbia and Japan.

## III. Dataset

In order to enable the comparative analysis between YouTube and Flickr data, several common data attributes have been identified. Each of the services exposes an API that enables programmatic data and content retrieval, but due to the differences in the content types, not all of the available attributes were included in the unified analysis. For example, YouTube forces its users to classify each video into one of 15 categories, and stores the length of the uploaded video. Flickr on the other hand stores make and model of the camera used to capture the picture, focal length of the lens used, information on whether the flash went off or not, etc. The content used in this study comprises images retrieved from Flickr, thumbnails (user-selected or auto-generated representative frames of respective videos) retrieved from YouTube and their associated metadata. In particular, for every object (photo or video), the following metadata was collected:

- ID of the object
- Username of the uploader
- Time and date the object was published (uploaded)
- Coordinates (latitude and longitude)
- Title of the object
- Tags associated with the object

Data were collected for the city of Dublin, Ireland, since it is a well-known tourist destination and the independent domain expert who interpreted the results was a resident of the city. Data were split into two subsets and meta-data was obtained for both of them. Some basic dataset statistics are provided in Table I.

Table I: Basic dataset properties

| Service | Period | # Photos / # Videos | # Users |
|---|---|---|---|
| Youtube | 2006–2011 | 5,592 | 2,939 |
| Flickr | 2005–2010 | 90,978 | 3,504 |

## IV. Analysis Framework

To be able to carry out visual exploration of the results by a domain expert, several visualization techniques were applied to the data, and a specialized tool was used to help the process (CommonGIS – Visual Analytics Toolkit [10]). To better understand the visual patterns that emerge from the data and to be able to detect differences between them, each data subset (Flickr and YouTube respectively) was processed using the same techniques and methods. For a simple overview of spatial distribution of objects, they were plotted on a map using their geographic coordinates, in such a way that each object was presented with a distinct point (circle). Then a density-based clustering was performed, using the OPTICS (Ordering Points To Identify the Clustering Structure) algorithm [12]. This algorithm creates an ordering of a database, additionally storing the core-distance and a suitable reachability-distance for each object. In contrast to another popular clustering algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [13], OPTICS responds well to datasets with varying densities. Density clusters are then built using only two parameters: the neighborhood radius and a minimum number of points in the neighborhood. Resulting clusters can have different shapes and there is no need to predefine their number (which is of essence when working with large datasets and area-dependent parameters).

To further detect the differences between the subsets, several series of heat maps were constructed. Heat maps are a good way of visually representing grouping of data points, since they make it easy to distinguish between hotspots and areas scarce with objects – by displaying continuous data in such a fashion that lower values are presented using one pattern or color and higher values are presented using a different color, hue or lightness. In this context, when used as an overlay in another software (e.g. Google Earth geo-browser), heat maps can be used to visually identify points of interest and interactively explore the area around those places. Also, they were constructed for different time periods, in order to detect temporal differences.

In addition, a two-level cluster analysis was carried out based on the framework of [14]. It uses spatio-temporal analysis and photo clustering for ranking areas, landmarks and events in a city. As a first step, the geo-tagged images are clustered solely based on their spatial proximity in order to derive "city areas". To this end, the BIRCH algorithm [15] is used. Subsequently, in each of the extracted areas, the hybrid clustering scheme of [8] is applied that relies on both visual and tag-based similarities among images in order to derive clusters corresponding to landmarks and events. In addition, a time slice analysis is carried out both at the level of the whole city and at the area level in order to derive a popularity ranking for the different temporal intervals, such as seasons, months and times of the day.

We opted for the parallel use of two different clustering frameworks, namely OPTICS and the two-level ClustTour framework [14], in order to enable the exploration of different aspects of the dataset. The use of OPTICS targets at a coarse-grained study of the spatial density distribution of content items. In complement to this, the ClustTour analysis results in a popularity-based ranking of different fine-grained areas of the city and a fine-grained clustering of content items around landmarks and events.

Finally, based on temporal data (i.e. time and date of picture taken / video made), for users who have taken several images/videos, paths of movement (flows between places) were extrapolated – which were then grouped to identify common routes taken [10]. In addition, it was possible to determine the direction of travel as well – indicating whether inbound or outbound flow is greater. Also, for further visual inspection, common routes were exported to Google Earth so their emergence could be dynamically presented using a player and a custom time-scale (i.e. the dynamics of their appearance could be explored).

## V. Results and Discussion

Using the methods and techniques described in Section IV, numerous patterns were identified in Flickr and YouTube subsets. The overview comparative plot of content objects can be seen in Figure 1. According to it, the majority of the pictures and videos were taken in the city centre. There are a lot more pictures present (blue circles) than videos (red circles), and they seem to span a wider area around the city centre (in fact, they are so densely packed along the coastline that it can be traced rather accurately just by observing circles representing images).

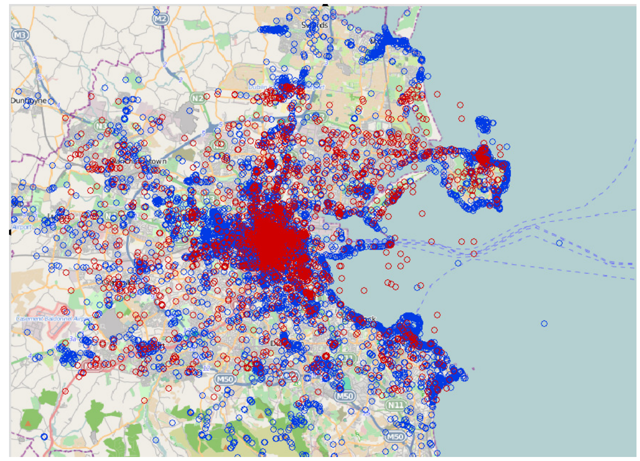


Figure 1: Spatial data distribution of Flickr and YouTube subsets (Flickr - blue circles, YouTube - red circles)

The next analysis step was based on the OPTICS algorithm. As subsets varied greatly in the number of records available, different values for distance ($d$) and minimum number of points ($p$) (used by the OPTICS algorithm) were applied for YouTube ($d = 100m$, $p = 5$) and Flickr ($d = 50m$, $p = 15$) subset respectively. Results (Figure 3) suggest that the city centre contains many interesting places for both photographers (Figure 3a) and video-makers (Figure 3b), but the former tend to generate much more content that spreads out more uniformly from the city centre. It can also be noted that the Flickr subset yielded more clusters than the YouTube subset, which was interpreted by the domain expert as a result of many famous sights that are often photographed as opposed to fewer locations where events take place (which are usually recorded and hence posted on YouTube).

Heat maps enabled us to gain additional insights into temporal and spatial patterns present in the data. The two subsets were partitioned into different time slices, namely seasons (spring, summer, autumn and winter), months (January through December) and time of day (day and night). Then for each of the groups several heat maps were constructed, one for every interval within the group.[1] They were then imported to Google Earth as discrete layers, to allow for visual inspection of changes in the density of photos and videos taken at popular locations in the respective time slice. Flickr subset analysis shows that there were more pictures taken during the day (8am – 8pm / 59% of images) than at night (8pm – 8am / 41% of images), and that their geographical distribution is more disperse throughout the city. Monthly analysis revealed some further intriguing patterns: March is the most popular month for taking pictures around Morton stadium, and June appears

[1]Due to the large amount of heat maps created, only a few screenshots were included in the paper. The rest, with accompanying .KML files, can be freely downloaded from: http://www.iim.ftn.uns.ac.rs/media/cason2011.zip

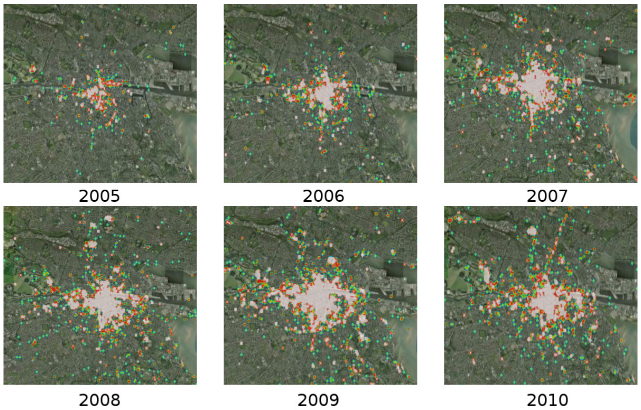
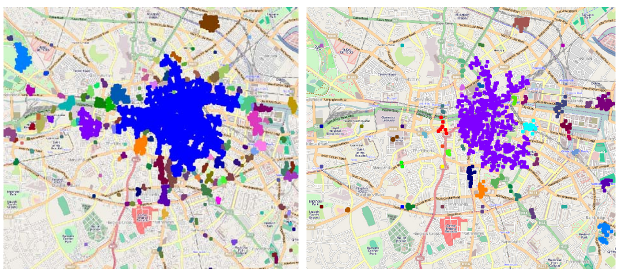



Figure 2: Flickr subset heat maps, data split by years



(a) Flickr (b) YouTube

Figure 3: OPTICS Clustering results

to be the preferred time for heading out to the docks and visiting the southern part of the city. When sliced into yearly intervals, data reveals an expanding pattern: images marked as taken in 2005 are closely clustered around the city centre. Afterwards, a growing trend can be observed where each following year seems to have photos covering a larger area expanding from the city centre (Figure 2).

One interesting pattern was captured in the Flickr subset, that could be observed only in January 2010 (Figure 4) – a high concentration of images were tagged as taken along a road that connects the city with the airport in the north. At first, the domain expert thought it could have been a wedding, a funeral or a parade, but could not find any such event taking place in the observed month and year. After further investigation, it turned out that the photos were generated by a single person, who is a member of a Flickr group documenting the condition of the Irish bicycle lanes. It is interesting to note that such an impact on the heat maps can be made by a single person, and it also shows the diverse uses of the Flickr service.

The heat map analysis was repeated for YouTube subset. Due to the smaller size of the set, a less detailed city view was often obtained. YouTube day / night heat maps do not show a large deviation of videos with respect to the time of the day they were taken at; most of them are concentrated

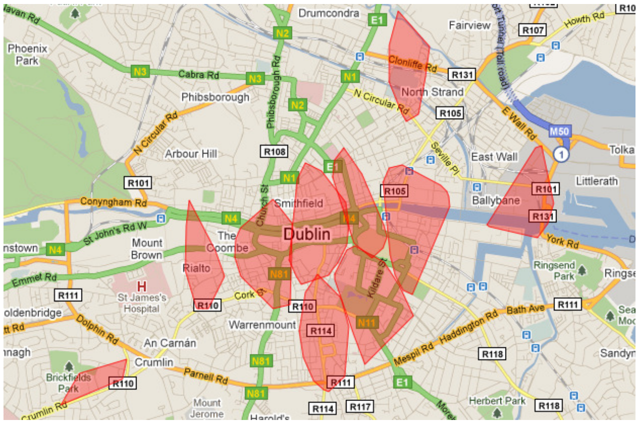
Figure 4: Anomaly in the Flickr subset – a high concentration of photos taken along the road connecting the city centre with the airport in the north

near the city centre, with many scattered seemingly randomly throughout the rest of the city. Partitioned to months, data shows that the least popular month for taking videos in Dublin appears to be June, closely followed by September. In general, it can also be observed that there are not many videos taken in the south-western part of the city, which might make sense since a closer inspection of that area revealed that it is a residential area dotted with parks (with probably not too many events going on there).

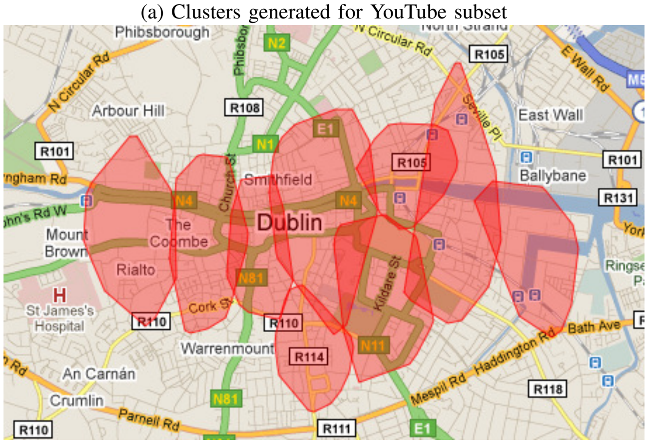
After clustering the YouTube and Flickr content using the ClustTour framework (Figure 5), a visual analysis of the cluster distribution was carried out by a domain expert. In general, the “areas” formed from the YouTube data[2] are the result of locations where events take place, while the ones resulting from Flickr[3] represent famous sightseeing locations in Dublin. More specifically, examination of the YouTube areas (Figure 5a) reveals that much interest is focused around the main stadiums of Dublin: Croke Park and The Aviva Stadium. These are used for national and international soccer and rugby matches as well as concerts. Additionally, Dublin’s main concert venue, The O2, also known as The Point, which is located in the docklands area of the city forms another popular area. When examining the content-based clusters, it is seen that acts performing in smaller venues such as Vicar Street and Whelans generate large quantities of YouTube data. Additionally, tourist attractions such as The Guinness Store House have substantial amounts of YouTube content associated with them.

It is also clear that the areas formed by spatial clustering of the Flickr photos correspond to locations which are famous for their scenic beauty or architecture (Figure 5b). For example, The Ha’penny Bridge, which spans the River Liffey, is an iconic symbol of Dublin and forms the most important cluster along with O’Connell Street and The Spire.

[2]Can be seen online at: http://www.clusttour.gr/youtube/?content=placeid=1
[3]Can be seen online at: http://www.clusttour.gr/?content=placeid=24




(a) Clusters generated for YouTube subset




(b) Clusters generated for Flickr subset

Figure 5: ClustTour generated clusters

This landmark was included in the recent NBC list of the top 25 most photographed locations in the world ([16]). Additional clusters form around other visitor attractions such as St. Stephen's Green, a park in the city centre, and Trinity College. Both are areas of beauty and located close to monuments and statues such as Molly Malone. Architecture also plays a key role in the forming of clusters with Christchurch Cathedral and The Custom House both contributing to the formation of distinct clusters in different parts of the city. Similarly, The Guinness Store House forms a cluster of its own indicating its importance as an attraction of Dublin.

Interestingly, the top two areas formed from each content source represent equivalent areas of the city (the top 10 clusters from each subset are presented in Table II, along with the season the most data was generated in). However, when the content-based clusters are examined, it is noteworthy that the data contributing to these areas are different. The area around the Ha'penny Bridge is indicated as the top area in both content sources, although its prominence in the YouTube set is not the result of the bridge itself but of its co-location to The Academy, an intimate concert

Table II: Top clusters detected by ClustTour framework

| | YouTube | | Flickr | |
|---|---|---|---|---|
| Rank | Cluster name | Season | Cluster name | Season |
| 1 | Academy | Autumn | Ha'penny Bridge | Spring |
| 2 | Grafton | Autumn | Campanile | Spring |
| 3 | O'Connell | Autumn | Cathedral | Summer |
| 4 | Whelans | Autumn | Storehouse | Spring |
| 5 | Docklands | Autumn | Famine Memorial | Spring |
| 6 | Vicar | Autumn | Merrion | Spring |
| 7 | Croke | Summer | Whiskey | Summer |
| 8 | Crumlin | Autumn | Whelans | Spring |
| 9 | Ballybane | Summer | Hanover | Spring |
| 10 | Guiness | Autumn | Howth Harbour | Summer |

location in the city generating a lot of YouTube content. This also highlights the fact that the size of the venue does not necessarily contribute to the larger amount of videos produced. For example, Croke Park has a capacity of over 40,000 for a concert while The Academy had a capacity of less than 1000 and yet The Academy has a greater impact on online content generation.

Overall, comparing the clusters returned, it is clear that the clusters formed from the Flickr photo sets reflect the main tourist attractions of Dublin, while at the same time capturing the main event locations. On the other hand, the YouTube clusters capture events and the location of events, but miss many of the main tourist attractions.

Finally, an estimate of the movement flow of people around the city was derived based on the timestamps of content items (Figure 6; YouTube data are presented in red and Flickr data in blue). Both sets of movement statistics are similar and show a uniform flow of movement in all directions within the city centre. An axis along the River Liffey can be observed. The river forms a natural boarder between the north and south of the city. The flow of people along it suggest that people travel along one side of it before crossing over and travelling on the other side, rather than crossing back and forth. When areas outside of the city centre are considered, there is a greater flow of people out of the city centre towards the south of the county, particularly towards the coast which is an area of scenic beauty and popular with tourists and those interested in hiking.

## VI. Conclusion

The work presented in this paper focuses on examining and explaining differences between patterns detected in publicly available multimedia data gathered from two major content-sharing services, namely Flickr and YouTube. Most of the differences emerge as a result of habits of people: they tend to take pictures of places, and make videos of events. This can clearly be observed in ranking of top clusters for each of the subset. On the other hand, some of the patterns would have escaped detection had only one approach been used – for example, anomaly detected in the Flickr subset using heat maps would have gone unnoticed had only the

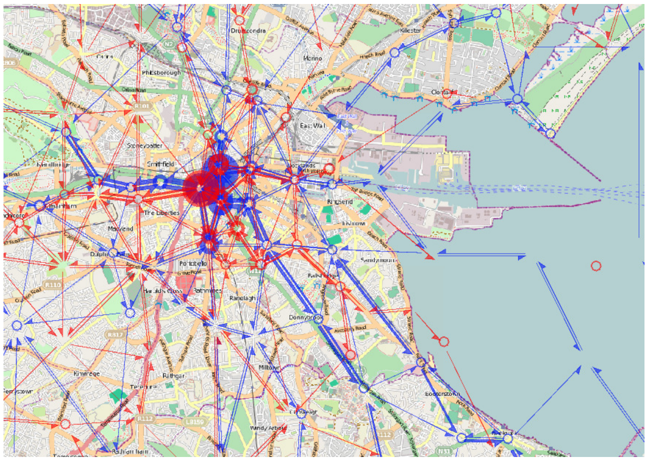


Figure 6: Aggregated trajectories of Flickr and YouTube users (Flickr - blue trajectories, YouTube - red trajectories)

clustering of the data been performed. Hence, we demonstrated the importance of combining different methods for pattern extraction in order to gain a better understanding of phenomena that drive the data generation. Also, it is worth mentioning that our future research efforts will be directed towards determining if the results obtained in this paper are true for other cities as well, and towards quantitatively describing similarities and differences between the detected clusters (and obtained results in general).

## Acknowledgment

The work of Milan Mirkovic and Dubravko Culibrk was supported by Ministry of Science and Education of Republic of Serbia, under grant III-44002. The work of Symeon Papadopoulos was funded by the GLOCAL project under contract number FP7-248984. The work of Gavin McArdle was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. Research presented in this paper is part of the collaboration under the COST (European Cooperation in Science and Technology) ICT Action IC0903, Knowledge Discovery from Moving Objects (MOVE).

## References

[1] M. C. Gonzlez, C. A. Hidalgo, and A. L. Barabsi, “Understanding individual human mobility patterns,” *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[2] N. Eagle and A. Pentland, “Reality mining: sensing complex social systems,” *Personal and Ubiquitous Computing*, vol. 10, no. 4, p. 268, 2006.

[3] M. Cha, H. Kwak, P. Rodriguez, Y. Y. Ahn, and S. Moon, “I tube, you tube, everybody tubes: analyzing the world’s largest user generated content video system,” in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007, p. 14.

[4] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat, “Digital footprinting: Uncovering tourists with user-generated content,” *Pervasive Computing, IEEE*, vol. 7, no. 4, pp. 36–43, 2008.

[5] F. Girardin, F. D. Fioreb, C. Rattib, and J. Blata, “Leveraging explicitly disclosed location information to understand tourist dynamics: a case study,” *Journal of Location Based Services*, vol. 2, no. 1, pp. 41–56, 2008.

[6] M. Mirkovic, D. Culibrk, S. Milisavljevic, and V. Crnojevic, “Detecting attractive locations using publicly available user-generated video content - central serbia case study,” *18th Telecommunications Forum TELFOR 2010*, 2010.

[7] T. Quack, B. Leibe, and L. V. Gool, “World-scale mining of objects and events from community photo collections,” in *Proceedings of the 2008 international conference on Content-based image and video retrieval*, 2008, pp. 47–56.

[8] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali, “Cluster-based landmark and event detection on tagged photo collections,” *Multimedia, IEEE*, no. 99, pp. 1–1, 2011.

[9] A. Popescu, G. Grefenstette, and P. A. Moellic, “Mining tourist information from user-supplied collections,” in *Proceeding of the 18th ACM conference on Information and knowledge management*, 2009, pp. 1713–1716.

[10] G. Andrienko, N. Andrienko, P. Bak, S. Kisilevich, and D. Keim, “Analysis of community-contributed space-and time-referenced data (example of panoramio photos),” in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2009, pp. 540–541.

[11] M. Mirkovic, D. Culibrk, and V. Crnojevic, “Detecting behavior patterns using publicly available user-generated video content - comparative study of serbia and japan,” in *DOGS 2010*, Novi Sad, 2010. [Online]. Available: http://www.ftn.uns.ac.rs/dogs/en-splash.htm

[12] M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. Sander, “OPTICS: ordering points to identify the clustering structure,” *ACM SIGMOD Record*, vol. 28, no. 2, p. 60, 1999.

[13] S. Kisilevich, F. Mansmann, and D. Keim, “P-dbscan: a density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos,” *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*, p. 38, 2010.

[14] S. Papadopoulos, C. Zigkolis, S. Kapiris, Y. Kompatsiaris, and A. Vakali, “City exploration by use of spatio-temporal analysis and clustering of user contributed photos,” in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, 2011, p. 65.

[15] T. Zhang, R. Ramakrishnan, and M. Livny, “BIRCH: an efficient data clustering method for very large databases,” in *ACM SIGMOD Record*, vol. 25, 1996, pp. 103–114.

[16] S. O’Neill, “The 25 most photographed places on earth,” 2010. [Online]. Available: http://www.msnbc.msn.com/id/43240059/ns/travel-travel$_tips/$