LAURA PÉREZ VERA

SPATIO-TEMPORAL DYNAMICS OF THE REGIONS OF INTEREST OBTAINED FROM GEOTAGGED PHOTOS

Dissertation presented to the *Universidade Federal de Viçosa*, as part of the requirements of the Graduate Program in Computer Science, to obtain the title of *Magister Scientiae*.

Advisor: Jugurta Lisboa Filho

Co-advisor: Giovanni Comarela

VIÇOSA - MINAS GERAIS

2020

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
P438s
2020

Pérez Vera, Laura, 1992-
Spatio-temporal dynamics of the regions of interest obatined from geotagged photos / Laura Pérez Vera. – Viçosa, MG, 2020.
46 f. : il. (algumas color.) ; 29 cm.

Orientador: Jugurta Lisboa Filho.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Sistemas de informação geográfica. 2. Dados geoespaciais. 3. Turismo. 4. Geografia urbana. I. Universidade Federal de Viçosa. Departamento de Informática. Programa de Pós-Graduação em Ciência da Computação. II. Título.

CDD 22. ed. 003.3

LAURA PÉREZ VERA

SPATIO-TEMPORAL DYNAMICS OF THE TOURISTIC REGIONS OF INTEREST OBTAINED FROM GEOTAGGED PHOTOS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae.*

APROVADA: 18 de dezembro de 2020.

Assentimento:

Laura Pérez Vera
Autora

Jugurta Lisboa Filho
Orientador

ACKNOWLEDGMENTS

First of all, I thank my family, who is responsible for who I am and where I am today, my father Luis M. Pérez Pérez, my mother Beatriz Vera Gómez and my sister Beatriz Pérez Vera.

To my boyfriend Davi Hagap Emmanuel da Silva, for always being there for me and giving me strength when I needed it.

To my friends who from anywhere in the world were always there to help me out.

To the *Universidade Federal de Viçosa* that gave me the opportunity to meet great people and help me be the person I am today.

To the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior in Brasil* (CAPES) for giving me the opportunity and the space to grow professionally.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To all my teachers from primary school to today, mainly my advisors Jugurta Lisboa-Filho and Giovanni Comarela, for their support at all times and their guidance on this path in the academic area.

To all the colleagues that the UFV gave me, and that the laboratory and endless nights united, Juan Niño for his constant encouragement and Luiz for friendship.

.

LIST OF ABBREVIATIONS AND ACRONYMS

VGI	Volunteered Geographic Information
ROI	Region of Interest
API	Application Programming Interface
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
ST-DBSCAN	Spatiotemporal DBSCAN
HDBSCAN	Hierarchical Density-Based Spatial Clustering
AOI	Area of Interest

RESUMO

VERA, Laura Pérez, M.Sc., Universidade Federal de Viçosa, dezembro de 2020. **Dinâmica espaço-temporal das regiões de interesse obtidas a partir de fotos com geo-tags.** Orientador: Jugurta Lisboa Filho. Coorientador: Giovanni Comarela.

Regiões de Interesse são tipos de dados geográficos que destacam áreas com algum tipo de interesse dentro de uma cidade. Eles podem ser usados para apoiar o planejamento de viagens do usuário, bem como para melhorar a distribuição de recursos nessa área de planejamento. É do interesse das agências de turismo do governo não só obter as Regiões de Interesse onde os visitantes focalizam a sua atenção, mas também compreender o seu comportamento para melhorar a experiência turística na área geográfica. Os dados de mídia social como fonte de informações geográficas registram as interações entre os usuários e o ambiente circundante e têm o potencial de descobrir informações valiosas. Métodos e técnicas de mineração de dados espaciais foram usados e aprimorados para ajudar a compreender esses comportamentos. O turismo como uma das indústrias economicamente mais importantes em Cuba e sendo vulnerável a diferentes eventos, como desastres naturais, relações políticas ou a passagem do tempo, tem recebido muita atenção e métodos foram desenvolvidos para obter, monitorar e avaliar a recuperação e a situação de as regiões de interesse dentro de uma área geográfica. Neste trabalho, os metadados das fotos geomarcadas do Flickr são usados como fonte de dados para obter as regiões de interesse e para entender a dinâmica espacial e temporal. Havana, a cidade turística mais importante de Cuba, é usada como área geográfica.

Palavras-chave: Regiões de Interesse. Pegadas espaciais. Dinâmica espaço-temporal. Turismo.

ABSTRACT

VERA, Laura Pérez, M.Sc., Universidade Federal de Viçosa, December, 2020. **Spatio-temporal dynamics of regions of interest obtained from geotagged photos.** Advisor: Jugurta Lisboa Filho. Co-advisor: Giovanni Comarela.

Regions of Interest are types of geographic data that highlight areas with some type of interest within a city, they can be used to support user's travel planning, as well as to improve the distribution of resources in that planning area. It is in the interest of the tourist agencies of the government not only to obtain the Regions of Interest where the visitors focus their attention but understand their behavior in order to improve the tourist experience in the geographic area. Social media data as a source of geographic information record the interactions between users and their surrounding environment and have the potential to discover valuable information. Methods and techniques of spatial data mining had been used and improved in order to help understand these behaviors. Tourism as one of the most economically important industries in Cuba and being vulnerable to different events such as natural disasters, political relations or the passage of time has received much attention and methods have been developed to obtain, monitor, and evaluate the recovery and status of the Regions of Interest within a geographical area. In this work, the metadata of the geotagged photos from Flickr is used as a data source to obtain the Regions of Interest and to understand the spatial and temporal dynamics. Havana, the most important touristic city of Cuba is used as the geographic area.

Keywords: Regions of Interest. Spatial footprints. Spatio-temporal dynamics. Tourism.

SUMMARY

1 INTRODUCTION

There has been given a very advanced growth of the different Web technologies. The emergence of Web 2.0 allows users to make contributions, share information and interact between them, among the various types of information contributed and shared by users, the ones that have geographic properties are called Volunteered Geographic Information c(VGI) (GOODCHILD, 2007).

VGI brings a new notion of infrastructure to collect, synthesize, verify, and redistribute geographic data through geolocation technology, mobile devices, and geographic databases. In recent years, VGI has led to an explosive growth in the availability of geographic information generated by the user. The most common providers of VGI are Flickr, Instagram, OpenStreetMap, Twitter, Facebook, YouTube, Wikimapia, Foursquare, among others (SUN, FAN, et al., 2015).

Spatial footprints are a special type of VGI on social media platforms. By posting stories or sharing information on social media, people can georeferenced their posts by attaching their locations. Compared to traditional VGI platforms (e.g.: OpenStreetMap) in which people directly contribute geographic information, the spatial footprints can be considered indirectly as VGI, that is, the main objective of the user is to share information instead of contributing data (CHENG, CAVERLEE, et al., 2011). This is a kind of spatial information provides a rapid and cost-efficient alternative to the traditional surveys (HEIKINHEIMO, MININ, et al., 2017).

People record and note their spatial footprints when they travel and experience the world. Large volumes of this type of geographic information provide a great opportunity for the study of place concepts. These spatial footprints directly show specific phenomena, scenes or, status of reality, which is a great opportunity to discover valuable geographic information and notice changes in reality over time, such as Points of Interest (POI) that refer to specific locations or real-world entities in geographic space (DE TRÉ, VAN BRITSOM, et al., 2013), or Region of Interest (ROI) that may contain several co-located geographic features (ELIAS, 2003), also the ROI

can also include areas that do not have important landmarks, but simply provide panoramic views (e.g.: areas in Paris that provide a good view of the Eiffel Tower) (HU, GAO, et al., 2015).

Spatio-temporal data is defined as an object that has spatial and temporal properties, the spatial properties are the location and geometry of the object, the temporal property is timestamp or time interval for which the object is valid. Spatio-temporal data sets essentially capture changing values of spatial and thematic attributes over a period of time. The analysis of spatio-temporal data is growing with the increasing availability and awareness of huge amount of geographic and spatio-temporal datasets (RAO, GOVARDHAN e RAO, 2012),. this analysis is useful for sectors of the economy such as tourism, being is one of the most economically important industries in the world (YAN, ECKLE, et al., 2017), knowing the state of the touristic areas is a major priority for countries whose economy depends almost entirely on the tourism industry. This industry can be affected by several events such as natural disasters, changes in political relations, and the pass of the time in the tourism infrastructure.

The spatio-temporal data used for this work is collected from Flickr. Flickr is a powerful platform created to share professional photos and non-professional photos. Flickr also makes available a free API that allows to obtain detailed information of the shared photos.

In order to better understand the changes in the tourism industry of a geographic area and how these changes are related to the different kinds of events in time, it is necessary to make a spatio-temporal analysis of the data. There are traditional surveys that are carried out by the tourism agencies in order to obtain that kind of information from the travelers, but these methods have proven to be time consuming and do not reflect reality because most of the time failed to obtain all the information needed.

Methods have been developed to obtain the places where people focus their attention using geotagged photos, one of the most recent paperwork developed an efficient method for POI/ROI discovery from Flickr (KUO, CHAN, et al., 2018). Most of these articles use known techniques of data mining clustering (such as DBSCAN or ST-DBSCAN methods) for obtaining the POI/ROI in the different geographic areas

(HU, GAO, et al., 2015) and they do not perform an analysis upon the result to study the relations of the ROIs obtained in the geographic area.

Cuba as one of the countries with the tourism industry being one of the most important economic polo (FERRER, 2018) and being subjected to some important political changes in terms of economic relations serves this work as a provider for the VGI data needed. Based on different sources of VGI this study identifies the different touristic POI/ROI in a geographic area and establishes how the dynamics between the POI/ROI are related to different events, to make a better contribution of information in the tourism scope of each region.

The objective of this work is to analyze the spatial distribution and temporal dynamics of the ROIs obtained from geotagged photos. This work is based on the hypothesis that in countries where tourism is the main economic polo all political relations had an effect on the tourism dynamic.

1.1 Objectives

The main goal of this research is to analyze the relations between the ROIs of a geographical area and the political events in a country, taking as a data source the spatial footprints of the users that visited that area. This analysis has the purpose to make an information contribution to the tourism in the area, providing insights about the effect of the political events in the tourism and how the dynamic of the ROIs were affected by these events.

Specific objectives are:

1. Obtaining spatial footprints of the geographical area of Havana;
2. Obtention of ROIs in the geographic area Havana by clustering methods and algorithms to define the spatial shape of the clusters;
3. Identify, through analysis of the dynamics of ROIs over the years, events that had an impact on tourism.

1.2 Organization of the Dissertation

This dissertation was created as a collection of articles produced as a result of the research. About the articles, the first was presented in *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems 2020* and the second will still be submitted to *the* ***geographical analysis*** *journal* (https://onlinelibrary.wiley.com/journal/15384632).

This dissertation is organized as follows:

Chapter 2 is composed of the two articles resulting from the research with their respective bibliographic references, Article I: “What about Cuba? Political tourism or tourism policies?” is reproduced in Section 2.1 and Article II: “Analysis of the international tourism in Havana through their ROIs dynamics.” is presented in Section 2.2.

In Chapter 3, general conclusions are made, discussions about the results and objectives achieved. Some suggestions for future work are also presented.

2 SCIENTIFIC ARTICLES

This chapter contains the two articles resulting from the research. The first article with the title “What about Cuba? Political tourism or tourism policies?” was presented in the *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems 2020,* conducted an online version with a base in Seattle, United States. This article shows the analysis made with the ROIs in Havana obtained as a result of applying HDBSCAN clustering method on the spatial footprints obtained from Flickr. The analysis is based on the political event of the visit of Obama to Cuba in 2017 and how this event affects the flow of American tourists to Havana.

The second article entitled “Analysis of the international tourism in Havana through their ROIs dynamics.” performs an analysis of the ROIs in Havana, taking into account the countries of origin of the users, who mostly visit the island in the last twenty years. This article, unlike the first, uses a spatial overlapping clustering method, obtaining spatio-temporal ROIs, and with this a spatio-temporal analysis is made in order to show the relation between the events of the countries and the dynamic of the tourism.

2.1 ARTICLE I: What about Cuba? Political tourism or tourism policies?

Laura Pérez Vera, Giovanni Comarela, Jugurta Lisboa-Filho

In: ACM SIGSPATIAL International Workshop on Geospatial Humanities (GeoHumanities'20), 4., 2020, Seattle, USA. Proceedings. Seattle: ACM, 2020.

ABSTRACT

In this paper, we use the metadata of the spatial footprints obtained from Flickr of the geographical region of Havana to find the Regions Interest within the city. We conduct a spatio-temporal analysis of the changes in the Regions of Interest visited by Americans tourists in order to understand the impact of Obama's visit to Havana. This analysis is guided by the following research question: Given the political changes, are there changes in tourism? HDBSCAN and alpha-shape algorithms are used to obtain the Regions of Interest. Our results show that there exists a connection between the political event and changes in the tourism dynamics in the period of Obama's visit to Cuba. Furthermore, our analyses show that the regions where the American tourists have more activity are the ROIs near the Havana port, which is consistent with the major event in the area being the entrance of the cruises in the country.

2.1.1 Introduction

Cuba was for a long time a country closed to the world, characterized by its revolutionary policies in the last sixty years, its economic and political history has always been linked to one important economic and political partner, first Spain, then the United States and finally the former Soviet Union. Today, Cuba has changed substantially, tourism became the country's solution to get foreign exchange income, assuming international tourism as one of the basic axes in the redefinition of its economy. This economic opening led to the establishment of relations with countries such as Spain, Mexico, Canada, Japan, France, and Jamaica [6]. These established economic relations also constitute, in the case of Cuba, in new political relations. The relationship between tourism and politics has been analyzed in tourism studies,

showing that the political impact on a region can carry a significant cultural and psychic/symbolic impact in the tourism of that region [11]. There exists a much deeper relationship between tourism and politics, since the moment we understand that tourism is much more than ``a place to stay during holidays''. This relation can be shown when studying the dynamics in tourism related to a political event.

To show the tourism dynamics, this paper uses the spatial footprints left by the tourists on social media platforms when they travel and experience the world.
Compared to traditional Volunteered Geographic Information (VGI) platforms (e.g., Open Street Map), in which people directly contribute geographic information, the spatial footprints can be considered indirectly as VGI, that is, the main objective of the user is to share information instead of contributing data [3]. Spatial footprints directly show specific phenomena, scenes or status of reality, which is a great opportunity to discover valuable geographic information and notice changes in reality over time such as Region of Interest (ROI) that may contain several co-located geographic features [5], also the ROI can include areas that do not have important landmarks [9]. The ROIs are obtained from geotagged photos, representing the real interests of the tourists who visit the island. Millions of photos are uploaded to the site on a daily basis from all around the world and of every subject. For millions of budding photographers, it's a place to freely share their photos with friends, family and categorized communities of other camera wielders around the world.

This paper makes an approach to the study of the ROIs of Havana in order understand the connection between the political events and tourism. With the emergence of the Web 2.0 and users being able to make content contributions, share information, and interact with each other. Hence, data-driven approaches may be used to observe, through the tourists ``eyes'', changes in Cuba.

The study of the ROI became a major priority when trying to understand the changes in the countries or cities whose economy depends almost entirely on this industry. Events such as natural disasters, change in the political relations and the pass of the time in the tourism infrastructure can characterize those geographic areas. The objective of this work is to bring a better understanding and provide an approach to the discussion on the changes of the tourism in Cuba, when trying to analyze the changes of the tourism through the political changes. The research question that

guides this paper is: Given the political changes, are there changes in tourism? Using data clustering to obtain the ROI from the spatial footprints collected from the city of Havana, and analyzing the data inside each ROI, this paper responds to such a research question.

To that end, first, we collect the metadata from the geo-tagged photos inside the geographic area of Havana using the Flickr's public API (Application Programming Interface). Second, the data is clustered using HDBSCAN algorithm and we apply the α-shape method on the clustered data to obtain the ROI for each cluster. Then, we make a temporal analysis based on the origin of the photo's owner that visited the city through the years. Finally, we analyze the ROIs with photos taken by Americans.

We show that the political relation between the United States and Cuba had a significant effect on the tourism dynamics in Havana in the year of 2017, being the ROIs near the port of Havana, the ones with more activities by the Americans tourists, which is consistent with the activities by cruises on the port of Havana.

2.1.2 Related Works

There are several papers that study Regions of Interest from different geographic areas using geotagged photos as data sources.

A three-layer framework for extracting the Areas of Interest from geotagged photos and understanding their spatio-temporal dynamics is presented by Hu et al. [7]. To retrieve data, the authors used Flickr's public API. Then, the data pre-processing is performed on the data layer. The spatio-temporal layer focuses on extracting information about the spatial extents as well as the temporal growth of the ROIs. Finally, the semantic layer serves the purpose of discovering knowledge from the extracted ROIs.

Kuo et al. [8] also proposed a framework to obtain the ROIs using as data source the geotagged photos from Flickr. The attractive footprints discovery is conducted to eliminate noise and to select valid footprints with a local maximum, using a voting value equation. In order to obtain the attractive footprints, a threshold is set. The clustering step is performed by processing those spatio-temporal properties and attributes of those attractive footprints.

The study of Li et al. [10] conducts an exploratory analysis to identify tourist attractions as hot spots in Flickr usage patterns using partial least squares regression. The study exposes the relationships between the tweet and photo densities and the characteristics of people in different counties of California. The authors find that, after removing probable tourist data, well-educated people in management, business, science, and the arts are more likely to be involved in the generation of georeferenced tweets and photos.

Yun et al. [14] find the seasonal distribution of urban walking tourists in the Bukchon Hanok Village using as data source a GPS-based smartphone application combined with an additional questionnaire to track tourists' spatio-temporal movement.

The previous papers conduct a study on how to obtain the ROIs from geotagged photos, and a few of them perform an analysis that takes into account important events in the geographic area.

In the area of tourism in Cuba, there are articles published that perform an analysis on the history of tourism in the country. For instance, Chavez et al. [2] made an economic and geographical approach to analyze the progress of tourism and how the changes on the internal economic policies allowed many Cubans to start a second inflow of money, which had a positive impact on the island’s economy and tourism. Salinas et al. [12] conducted a historical study of the evolution of tourism in Cuba, affirming that the emergence and evolution of tourism in Cuba correspond to the country’s economic and socio-political history.

In contrast to the aforementioned work, here we make a connection between the occurrence of important events in Cuba and the dynamics on its ROIs, obtained from geotagged photos.

2.1.3 Building the Region of Interests

This section has two main goals. First, we show our strategy for obtaining the metadata of the photos from Flickr and how the data is cleaned in order to eliminate the noise and issues that commonly come with this type of data. Then we describe the clustering approach that we used to obtain the Regions of Interest.

2.1.3.1 Data Collection and Pre-processing

For the data extraction process, we relied on Flickr's public API[1] which provides a wide range of information of each photo that has been uploaded. We only collected information from photos that were made publicly available by the users. This information represents the presence of the tourists in the city that used Flickr as a social platform for uploading their photos. We work under the hypothesis that this information is a representative sample of the tourists in the city, despite recognizing the sample's bias of containing only Flickr users.

The **flickr.photos.search** endpoint[2] returns records (i.e., photos) for a given area, which can be delimited by a bounding box described as latitude and longitude intervals. Unfortunately, such an endpoint has a limitation of returning at most the four thousand records per query. In order to overcome this limitation, we started by querying the whole Havana area. Then, every time that we obtained four thousand records, we split the box in half, alternating latitude and longitude, and we recursively applied the same process to both subareas. Whenever a query returned less than four thousand points, we stopped. For each photo, we saved the following metadata: photo ID, photo title, description, tags, upload time, time when a photo was taken, location (in the form of latitude and longitude), and owner ID.

The data was collected from the geographic area of Havana using the coordinates belonging to the bounding box defined by latitude in the interval [23.0499, 23.1470] and longitude in the interval [-82.4551, -82.3025]. In total, 136,970 records within the bounding box of Havana were retrieved, with a total of 7323 users.

After collecting the data, we filtered out some records with inconsistent information. For instance, some photos had the field ***time when a photo was taken*** with values before the year 1900 or after the data collection time. After the cleaning phase, our dataset contained 68101 photos related to 3364 different users.

---

1 www.flickr.com/services/api/

2 www.flickr.com/services/api/explore/flickr.photos.search

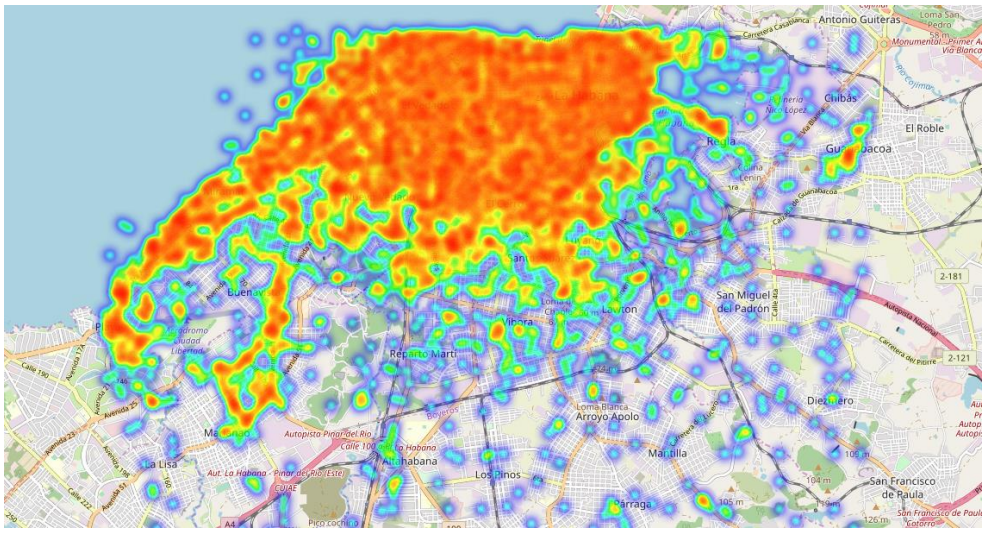

Figure 1. Distribution of photos over the Havana area

Figure *1* shows the distribution of photos in our dataset. The spatial distribution of these photos is consistent with the geographical location of Havana and shows that the higher concentration of points is in the Central Havana area, corresponding to its main touristic area.

We decided to focus on the more touristic regions in Havana for the purpose of this paper. Figure *2* shows the concentration of the photos taken in the area to be studied.

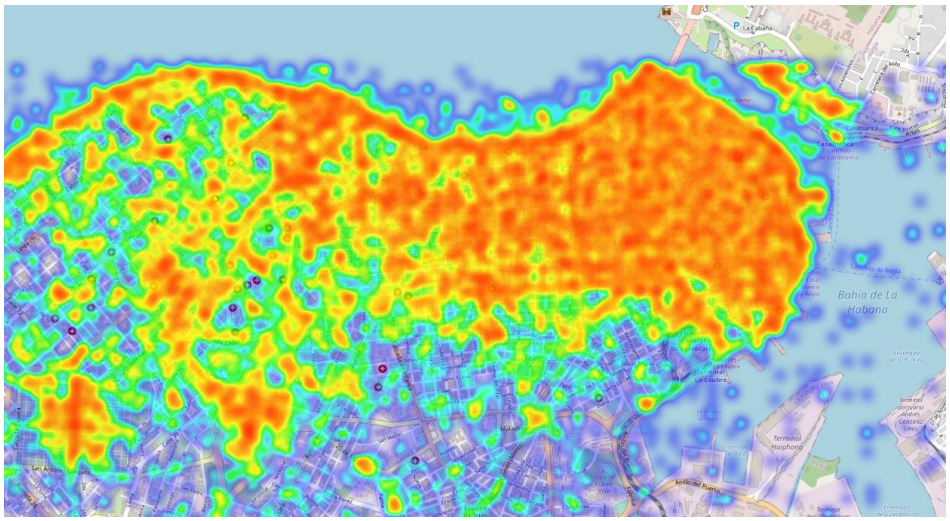

Figure 2. Distribution of photos over the more touristic areas of Havana

2.1.3.2 Clustering

In order to find Regions of Interest in Havana, we used a two-phase strategy. First, we applied the HDBSCAN clustering algorithm [1] to our Flickr dataset. HDBSCAN is a density-based hierarchical clustering method, based on DBSCAN with the ability of handling data with varying density. In this case, each resulting cluster is a representation of a maximal dense region of photos.

The second phase is to transform each cluster into a Region of Interest. To that end, we used the **alpha-shape** algorithm, which is a generalization of the standard convex hull algorithm for a finite set of points [4].

Figure *3* shows the result of applying the two steps above to our Flickr dataset. We obtained a total of 136 ROIs, with different shapes and sizes. Among these ROIs, there are important landmarks of Cuba. For instance, the ROIs 20 and 21 represent two parts of the well-known "*Plaza de la Revolución*" in Havana, most specifically the ROI 20 is the entrance to the monument and museum itself, and the 21 is the area in front of the monument liked by tourists for being an open area and the possibility of taking photos with the view of the monument. The ROIs 116,114, 125, and 128 represent *"Plaza Vieja"*, "*Plaza San Francisco de Asis"*, *"Plaza de Armas"*, and "*Parque Central*" respectively, the most famous squares in Havana.

Figure 3. Regions of Interest in Havana

2.1.4 Temporal Analysis

In this section, we expand our analyses to understand the tourism dynamics in Havana. To this end, we start by looking at the country of origin of the tourists over time and then, we dig deeper into a special case, tourists from the United States.

Figure *4* shows the temporal distribution of tourists visiting Havana, from the year 2004 to 2019, for the 10 countries that most originated visitors to the city during such a period.

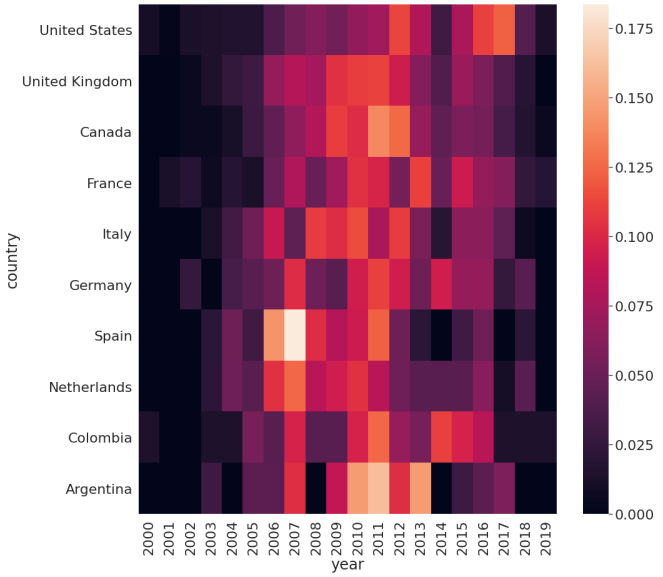

Figure 4. Temporal distribution of visits to Havana per country of origin. Values are normalized according to the sum of each row

.

There are two main takes from the figure. First, it seems that, overall, the number of tourists visiting Havana starts to decrease around 2015 and that this trend is even more significant after 2018. It is important to keep in mind that we are using Flickr photos as proxy for the number of visits, and that Flicker usage dropped significantly in the last years as a result of the emergence of other photo-sharing services (e.g. Instagram) and by an important Flickr's policy change, when it started limiting the number of photos for free accounts [13].

The second take from is Figure *4* that despite the Flicker's cooling down and differently from the other countries, the number of tourists from the United States peaks in 2016 and 2017. Such an observation corroborates the idea that politics and tourism

in Cuba are correlated. In fact, the rapprochement between Cuba and the United States was established with Obama's visit to the island in 2017, but before that, there were several other policy changes: The United States dropped the ``State sponsor of terrorism'' designation for Cuba in 2015; regularly scheduled flights between the United States and Cuba restarted in 2016; and travel by sea restarted in 2016 as well.

In order to dig deeper into this matter, we took a closer look at the Regions of Interest, identified in the Section **Clustering** visited by tourists from the United States. Figure *5* shows the concentration of those tourists inside each ROI for each year, since 2014. The ROIs that received more attention were 114, 116 and 125, known as "Plaza San Francisco de Asís", "Plaza Vieja", and "Plaza de Armas" respectively, which are also depicted in Figure *5*. Interestingly, ROIs 114 and 116 are the regions with higher concentration for years 2017 and 2018 respectively. The increase of interest in those regions can be explained by the political changes with the United States previously described, such as the approval, in 2016, of the travel by sea from the United States, allowing cruise companies to make deals with the Cuban government. Such a change increased the activity in the port of Havana (in front of ROI 114) and nearby regions. This area is characterized by visits from tourists that arrive via the cruises, being the ROIs 116 and 125 indirectly affected by the activity in that region.

2.1.5 Discussions and Future Work

In this work we started looking at the tourism dynamics in Havana in the last few years, and we tried to correlate these dynamics with important events concerning the relationship between the Cuban government and the outside world, especially, the United States. Our results suggest that the changes in the United States' policies with regard to Cuba had a significant impact on the flow of US residents visiting Cuba. Moreover, it was possible to notice that the most affected area was nearby the port of Havana.

Given this tourism growth in Havana, a growth of income in the island has been happening as well, which is enabling the reconstruction of important parts of the city. In current/future work, we intend to analyze the tourism dynamics for all the ROIs in order to understand how these constructions and renovations affected the tourists' interest.

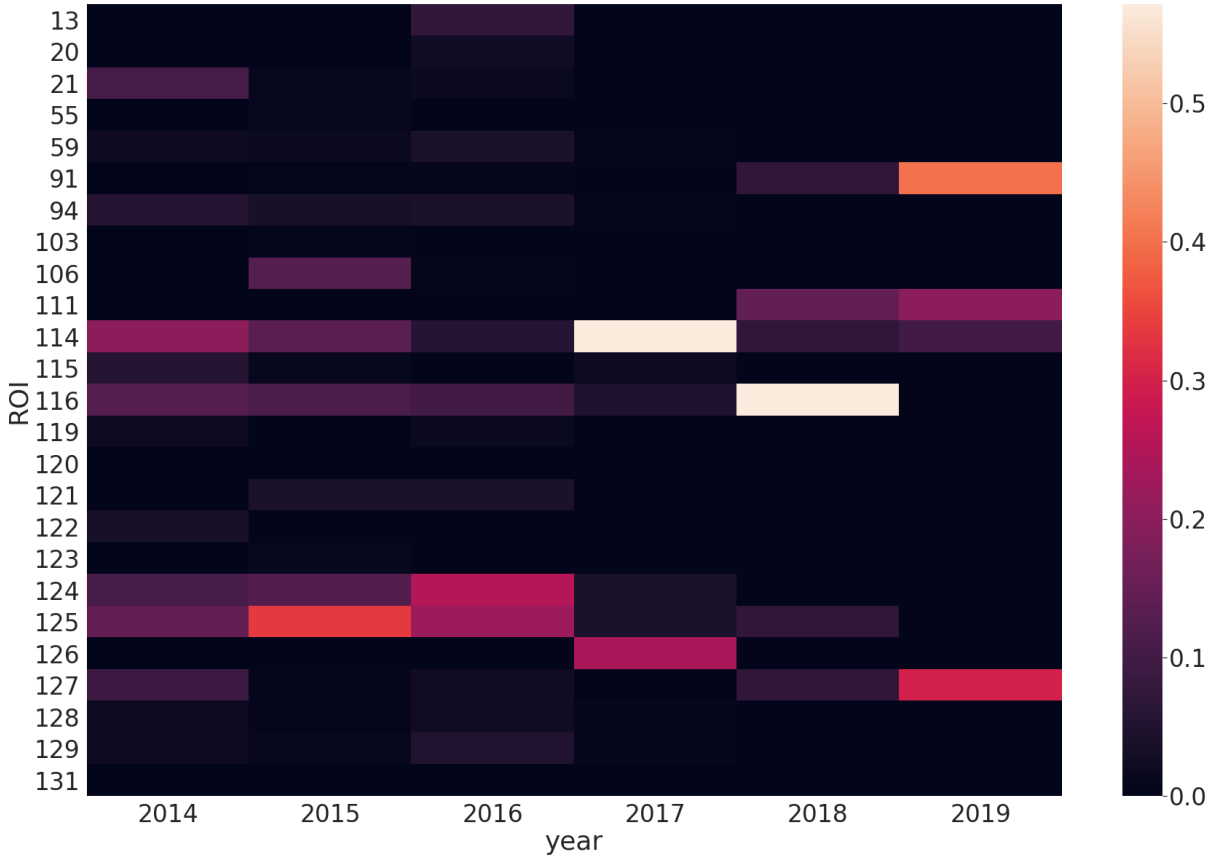

Figure 5. Temporal dynamics of visits to Regions of Interest in Havana by tourists from the United States. Values are normalized by the sum of columns

Due to the impact of the political event on the increase in the visits from Americans to the island, we can partially answer the question that gives title to this article, pointing that there is a connection between the political event and the tourism, and suggesting that there was a political tourism in that time lapse. Regarding policies to encourage tourism, this study did not observe strong evidence of impact on tourist visits to the city. As a future work we plan to analyze in a wider range the ROIs that were visited by the American tourists in a way to look for political interests.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and Fundação de Amparo à Pesquisa de Minas Gerais (Fapemig - Project APQ – 03763-12).

References

[1] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. 2013. Density-based clustering based on hierarchical density estimates. In Pacific-Asia conference on knowledge discovery and data mining. Springer, 160–172.

[2] Eduardo Salinas Chávez, Eros Salinas Chávez, and Lluís Mundet i Cerdan. 2019.El Turismo en Cuba: Desarrollo, Retos y Perspectivas/Tourism in Cuba: Develop-ment, Challenges, Perspectives.ROSA DOS VENTOS-Turismo e Hospitalidade11,1 (2019).

[3] Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z Sui. 2011. Exploringmillions of footprints in location sharing services. ICWSM2011 (2011), 81–88.

[4] Herbert Edelsbrunner, David Kirkpatrick, and Raimund Seidel. 1983. On the shape of a set of points in the planeside Transactions on information theory29,4 (1983), 551–559.

[5] Birgit Elias. 2003. Extracting landmarks with data mining methods. International Conference on Spatial Information Theory. Springer, 375–389.

[6] Sheryl Marie Elliott and Lisa Delpy Neirotti. 2008. Challenges of tourism in adynamic island destination: The case of Cuba.Tourism Geographies10, 3 (2008),375–402.

[7] Yingjie Hu, Song Gao, Krzysztof Janowicz, Bailang Yu, Wenwen Li, and Sathya Prasad. 2015. Extracting and understanding urban areas of interest using geo-tagged photos. Computers, Environment and Urban Systems54 (2015), 240–254.

[8] Chiao-Ling Kuo, Ta-Chien Chan, I Fan, Alexander Zipf, et al.2018. Efficient Method for POI/ROI Discovery Using Flickr Geotagged Photos. ISPRS International Journal of Geo-Information7, 3 (2018), 121

[9] Linna Li and Michael F Goodchild. 2012. Constructing places from spatial foot-prints. In Proceedings of the 1st ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information. ACM, 15–21

[10] Linna Li, Michael F. Goodchild, and Bo Xu. 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and

Geographic Information Science (2013). https://doi.org/10.1080/15230406.2013.777139

[11] Felicity Picken. 2017. The SAGE International Encyclopedia of Travel and Tourism.

[12] Eros Salinas, Lluis Mundet, and Eduardo Salinas. 2018. Historical evolution and spatial development of tourism in Cuba, 1919–2017: What is next? Tourism Planning & Development15, 3 (2018), 216–238.

[13] Andrew Stadlen. 2018. Why we're changing Flickr free accounts. https://blog.flickr.net/en/2018/11/01/changing-flickr-free-accounts-1000-photos/

[14] Hee Jeong Yun, Dong Jin Kang, and Myong Jae (MJ) Lee. 2018. Spatio-temporal distribution of urban walking tourists by season using GPS-based smartphone application. Asia Pacific Journal of Tourism Research(2018). https://doi.org/10.1080/10941665.2018.1513949

.

2.2 ARTICLE II: Analysis of the international tourism in Havana through their ROIs dynamics.

Laura Pérez Vera. Giovanni Comarela. Jugurta Lisboa-Filho

ABSTRACT

The metadata of the spatial footprints obtained from Flickr of the geographical region of Havana is used to find the Regions of Interest within the city. We conduct a spatio-temporal analysis of the changes in the Regions of Interest visited by the tourists in order to understand the impact of political relations in the tourism. This analysis is guided by the following research question: Is there a connection between the dynamics on the visits to the country and the political events that happened on the country? Hierarchical Agglomerative clustering, DBSCAN, and alpha-shape algorithms are used to obtain the Regions of Interest. Our results show that there exists a connection between the political events and changes in the tourism dynamics.

2.2.1 Introduction

The development of Web 2.0 has allowed the web community to interact with each other, providing information to central sites and with this the user becomes the source of information either voluntarily or involuntarily. Among the various types of information contributed and shared by users, the ones that have geographic properties are called Volunteered Geographic Information (VGI) [5].

In recent years, VGI has led to an explosive growth in the availability of geographic information generated by the user. The most common providers of VGI are Flickr, OpenStreetMap, Twitter, Facebook, YouTube, Wikimapia, Foursquare, among others [12].

Spatial footprints are a special type of VGI on social media platforms. By posting stories or sharing information on social media, people can georeference their posts by attaching their locations. Compared to traditional VGI platforms (e.g.: OpenStreetMap) in which people directly contribute geographic information, the spatial footprints can be considered indirectly as VGI, that is, the main objective of the user is to share

information instead of contributing data [1]. This is a kind of spatial information that provides a rapid and cost-efficient alternative to the traditional surveys [6]

With this type of data, different temporal analyzes can be carried out, where different economic sectors can be highly benefited. Tourism as one of the most economically important industries in the world [13]; knowing the state of the touristic areas is a major priority for countries whose economy depends almost entirely on the tourism industry. This industry can be affected by several events such as natural disasters, change in the political relations and the pass of the time in the tourism infrastructure.

In order to better understand the changes in the tourism industry of a geographic area and how these changes are related to the different kinds of events in time, it is necessary to make a spatio-temporal analysis of the data. There are traditional surveys that are carried out by the tourism agencies in order to obtain that kind of information from the travelers, but these methods have proven to be time consuming and do not reflect reality because most of the time failed to obtain all the information needed.

Methods have been developed to obtain the places where people focus their attention using geotagged photos, one of the most recent paperwork to develop an efficient method for POI/ROI discovery from Flickr [8]. Most of these articles use known techniques of data mining clustering (such as DBSCAN or ST-DBSCAN methods) for obtaining the POI/ROI in the different geographic areas [7].

Cuba as one of the countries with the tourism industry being one of the most important economic polo [4], and being subjected to some important political changes in terms of economic relations serves this work as a provider for the VGI data needed. Based on different sources of VGI this study identifies the different touristic ROIs in a geographic area and establishes how the dynamics between the ROIs are related to different events, to make a better contribution of information in the tourism scope of each ROI.

The objective of this work is to analyze the spatial distribution and temporal dynamics of the ROIs obtained from geotagged photos. The research questions that guide this paper are: Is there a connection between the dynamics of the visits to the country and the political events that happened in the country? If so, how this connection affects the interest on the main ROIs on a temporal basis.

To that end, first, we collect the metadata from the geo-tagged photos inside the geographic area of Havana using Flickr's public API (Application Programming Interface). Second, the data is filtered, eliminating those photos with voting values lower than the 0.85 percentile of the concentration of the data for each year. Thirdly an agglomerative clustering followed by the application of DBSCAN method in each cluster to find the spatial separation inside the cluster, this is applied on the data for each year, and we apply the **alpha-shape** method on the clustered data to obtain the ROI for each cluster in each year. With the ROIs of each year obtained, a similarity step is implemented to find and eliminate the ROIs that represent the same ROI in different years.

Then, we make a temporal analysis based on the origin of the photo's owner that visited the city through the years. Finally, we analyze the dynamic of the ROIs.

As a result, this work shows that the politics events in Cuba affects the dynamic in the tourism area, concluding that there is a connection between the political events and the tourism flow.

The remainder of this paper is structured as follows. Section 2.2.2 reviews related works on analyzing geotagged photos, their source and the methods used for obtaining the POI and ROI. Section 2.2.3 presents the workflow and methods proposed by this work. Section 2.2.4 presents the case study and the research outcomes. Lastly, Section 2.2.6 concludes this article.

2.2.2 Related Works

There are similar works that obtained the ROIs on a geographic area using geotagged photos as data source.

A three-layer framework for extract the AOI from geotagged photos and understanding their spatio-temporal dynamics is presented by Hu et al. [7], retrieved data is used from Flickr through their public API (Application Programming Interface), the data pre-processing is made on the data layer, the spatio-temporal layer, focuses on extracting information about the spatial extents as well as the temporal growth of the ROI, finally in this study the semantic layer serves the purpose of discovering knowledge from the extracted ROI, DBSCAN clustering method is used in order to identifying point clusters with 200m for Eps and 2\% for MinPts as parameters.

Kuo et al [8] proposed a framework to obtain the ROIs, using as data source the geotagged photos from Flickr. The attractive footprints discovery is conducted to eliminate noises and select valid footprints with a local maximum, using a voting value equation. In order to obtain the attractive footprints a threshold is set. The clustering step is performed by processing those spatio-temporal properties and attributes of those attractive footprints.

The study of Li et al. [9] conducted an exploratory analysis to identify tourist attractions as hot spots in Flickr usage patterns, using partial least squares regression.

This study illustrates the relationships between the tweet and photo densities and the characteristics of people in different counties of California, finding that after removing probable tourist raw data, well-educated people in management, business, science and the arts are more likely to be involved in the generation of georeferenced tweets and photos.

Yun et al. [14] finds the distribution of urban walking tourists by season, using as a data source a combined GPS-based smartphone application with an additional questionnaire to track urban walking tourists` spatio-temporal movement in the Bukchon Hanok Village and collect their demographic information. Kernel density is used to identify the density features in the neighborhood, with a radius of 10 meters based on an ArcGIS guideline, and a 1-meter cell size as the spatial unit for Kernel density analysis.

Dimobe et al. [2] conducts a study to find the dynamics to uncover ecological effects of recent land management on savanna habitats including tourism. They collected data from satellite images and used different landscape metrics to find the dynamics.

This study differs from the previous ones mentioned in that the construction of ROIs is carried out in a spatio-temporal manner, separated by each year. The clustering method used is the Agglomerative Hierarchical clustering and a study of the dynamics between the ROIs is carried out, taking into account the origin of the users and the main political events.

2.2.3 Materials and Methods

This section is divided as follows: Section 2.2.3.1 describes the method used by this paper for obtaining the metadata of the photos from Flickr and how the data is cleaned in order to eliminate the errors that commonly came with this type of data. Section 2.2.3.2 specifies the clustering process followed by this work to obtain the spatio-temporal ROIs. Section 2.2.3.3 explains the process that was used by this paper in order to study and obtain the dynamic between the ROI.

2.2.3.1 Data collection and Pre-processing

This study analyzes the spatio-temporal dynamics of the POI and ROI obtained from geotagged photos. Millions of photos are uploaded to the site on a daily basis from all around the world and of every subject. For millions of budding photographers, it's a place to freely share their photos with friends, family and categorized communities of other camera wielders around the world.

Flickr is a great digital tool that has been responsible for the curing of images from the ultra-mundane to the graphic masterpiece. There have been different articles that have used Flickr as a data source, providing an information background on terms of data quality and spatial distribution of the data, because of that Flickr is used as a data source for this paper. For the data extraction process the Flickr's public API[3] was used providing a wide range of information of each photo that has been uploaded.

The **flickr.photos.search** method has a limitation of return at most the first 4,000 results for any given search query according to the specifications of the method itself. In order to obtain more than that amount a method to the process of data collection was developed.

The method developed use as entry the bounding box of the area (in our case was the bounding box of Havana) and divide the bounding box for the x and y axis alternatively in a recursive way until the area divided had less than the minimum amount required. Photo metadata contains information about photo ID, photo title,

3 www.flickr.com/services/api/

description, tags, upload time, time when a photo was taken, location (in the form of latitude and longitude), and owner ID.

2.2.3.2 *Clustering*

The first steps of the data clustering process in this study are mainly based on the first stages of the Spatial Overlap method (SO) described by Kuo et al. [8]. The process of obtaining those spatial footprints which represent attractive footprints, in order to eliminate some noise and to identify those spatial footprints that are going to form each ROI is described in Section **Error! Reference source not found.**. The p attern discovery step described in Section 0 explains the process of identifying those spatial footprints that are going to belong to the same ROI. The Section 2.2.3.2.3 specifies the hierarchical density-based spatial clustering method that is used to obtain the clusters.

2.2.3.2.1 Attractive Footprints Discovery

The discovery of attractive footprints is conducted under the assumption that people will gather in areas that they are interested in, for each spatial footprint within each subset is calculated the voting value using the equation shown in Equation *1*, representing the summation of the weight function between that spatial footprint and those that are from different users and in a distance less than a radius r.

$$v_{pi=\sum_{i=1}^{j} \quad w_{ij},w \quad ij=e^{\frac{-\vee i-j\vee}{2\sigma^2}}}$$

Equation 1. Voting Value Equation

The spatial footprints with voting values higher than a threshold are going to be considered as attractive footprints. The value of this threshold is set according to the density of the voting values on each subset. This value is correlated with the amount of people that had visited the city in that year.

In the study of Kuo et al. [8] the threshold is set to 100, because if at least 100 people are around that area is going to be considered as an attractive area, according to the study.

In order to set the first threshold for obtaining the attractive footprints, a threshold needs to be set for finding those spatial footprints that received special attention on each subset.

For that manner the Equation *2* establishes the threshold for discovering those attractive footprints within each subset, being ***n*** the number of spatial footprints of each subset, ***p*** the quantile (value between 0 and 1) to be calculated, and **Q** represents the index of the voting value for the data. According to the results it is decided to use 0.85 as **p** value.

$$Q = \big(p * (n+1)\big)^{th} Term$$

Equation 2. Quantile equation

2.2.3.2.2 Pattern Discovery

In order to identify those points that belong to the same cluster, it is necessary to discover the pattern for each attractive footprint that is going to be clustered. The pattern discovery process is based on the one proposed by Kuo et al. [8].

The **distance_radius_subset** function on Figure *6* shows how the distance is calculated between all attractive footprints within a radius of 50m for each subset analyzed. Notice that the latitude and longitude of the attractive footprints is converted to radians in order to apply the haversine distance metric that requires the data in the form of [latitude, longitude] and both inputs and outputs are in units of radians.

The function **pattern_yearly** calculates the pattern of each attractive footprint that is being analyzed for each yearly subset, returning a vector of twelve positions with the number of spatial footprints that had being taken in that month from different owners inside a radius r, **dict_owner** is the dictionary that contains the spatial footprint by owner that is closer inside a radius r from the photo is being analyzed.

```
def distance_radius_subset(subset):
        RADIANT_TO_KM_CONSTANT = 6367
        radius_km = 50/1e3                  # 50 meters of radius
        radius_radian = radius_km / RADIANT_TO_KM_CONSTANT
        lat_long=[np.radians(list(zip(x.long, x.lat))) for x in subset]
        A = radius_neighbors_graph(lat_long, radius_radian,mode='distance',
            metric='haversine')
        return A
def pattern_yearly(sp):
    X = [0]*12
    for n in sp.dict_owner.values():
        m = month of n
        X[m-1] += 1
    return X
```

Figure 6. Pattern calculation for the attractive footprints

2.2.3.2.3 Regions of Interest

With the clustering process finished, the process of generating the ROI for each subset is the next task, for generating the polygons objects from the points inside each cluster the **alpha-shape** algorithm is used. **alpha-shape** algorithm is a generalization of the convex hull for a finite set of points introduced by Edelsbrunner et al. [3].

An issue that is found on the result from the clustered data is that some points from different clusters were spatially close, which causes that the polygons generated for some datasets were intercepted. In order to cleaning some of these polygons that spatially was giving no information, the function **roi** was implemented, the clusters are ordered by number of points in decreasing order and only added to the final list if they do not intercept in a percent major than 30 of the polygons already on it. The Figure *7* shows the method used to build the ROIs.

```
def roi(subset):
        for cluster in subset.clusters:
            points=[x.geom for x in cluster.sp]
            alpha_shape = alphashape(points)
            if alpha_shape.geom_type == 'Polygon':
                patches.append(alpha_shape)
        if len(patches)>=1:
            polygons=[]
            polygons.append(patches[0])
            for pol1 in range(1,len(patches)):
                flag=False
                for pol2 in range(0,len(polygons)):
                    if patches[pol1].intersects(polygons[pol2]):
                        por=(patches[pol1].intersection(polygons[pol2])


                                                .area/polygons[pol2].area)*100
                        if por>30:
                            flag=True
                            break
                if not flag:
                    polygons.append(patches[pol1])

          return polygons
```

Figure 7. ROI calculation

2.2.3.3 Dynamics of the ROI

In order to understand the dynamics of each ROI obtained as the result of the clustering method described on the Section 2.2.3.2, this part of the work is divided on two main steps: first analyze the changes of the density inside of each ROI through time, and validate those changes with the historic events happened on the geographic area. The density of each ROI is related to the number of spatial footprints inside each of the ROI obtained from the previous step. A matrix is constructed to analyze the number of points inside each polygon through the years.

2.2.3.3.1 Identification of each ROI

With this approach of analysis, the first problem is how to find the uniqueness of each polygon in order to be able to identify them inside each subset. The problem to identify a polygon through time is reduced to the problem of how to identify if two polygons are equal or if they are not equal in what degree they are similar, in other words a spatial distance metric needs to be used.

Hausdorff distance is defined as the maximum distance of a set to the nearest point in the other set by Shonkwiler [10], as a metric it finds the similarity between two geometries. For this paper two polygons are going to be considered as equals if the Hausdorff distance is less than 0.003.

2.2.3.3.2 Density matrix

The density of each ROI per year is related to how many spatial footprints were taken inside that polygon in a given year. The matrix to be analyzed is formed by the ROI obtained from the Section a and the number of photographs inside each polygon through the 15 years to be analyzed.

2.2.4 Results

The data was collected from the geographic area of Havana using the coordinates belonging to the bounding box: [-82.4551, 23.0499, -82.3025, 23.1470]. In total 136,970 records within the bounding box of Havana were retrieved, with a total of 7323 users, using the **flickr.photos.search** method.

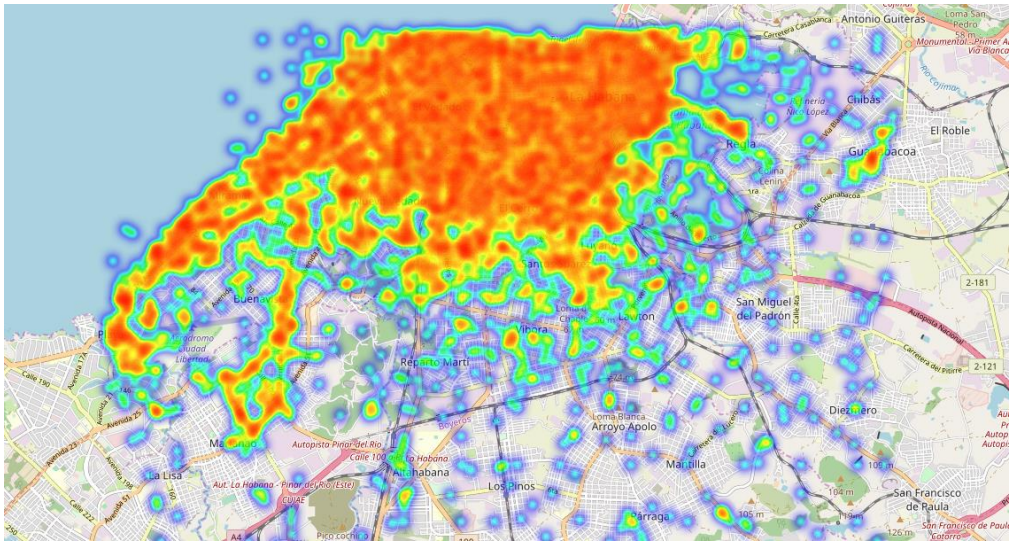

Figure 8. Heatmap of the e photos collected from Flickr

The Figure *8* shows the spatial distribution of the photos obtained from Flickr, which is consistent with the geographical location of Havana and shows that the higher concentration of points is in the Central Havana area, corresponding to the main touristic area in the city.

The date field (time when a photo was taken) from the data collected contained errors with years previous to 1900 and after the actual date, after eliminating those errors, the concentrations of the data is between the years of 2008 and 2015.

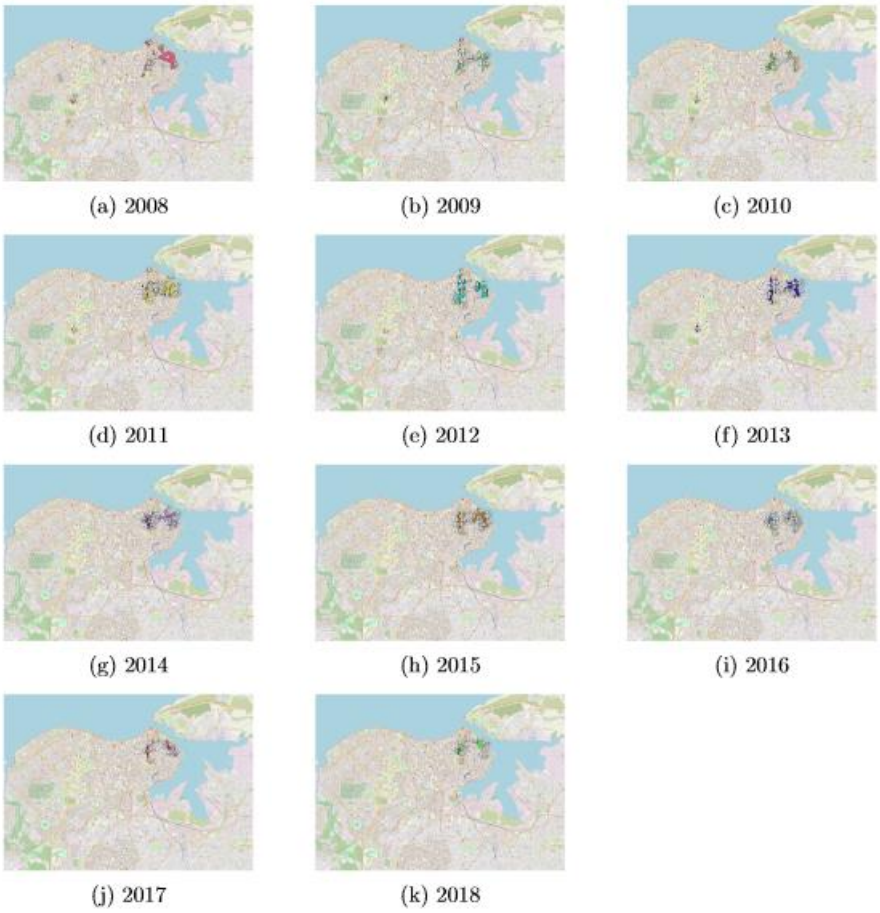

Figure 9. Result of the spatio-temporal method per year

The Figure *9* shows the results of the clustering method in the data for each year. From this figure it can be seen that the main activity and concentration of the ROIs are in the Old and Central Havana areas, which is persistent with the main touristic activity in the city. The Figure *9* shows that in years 2008 to 2013, polygons were obtained in the area of the *"Plaza de la Revolución"*, however these polygons disappear in the years from 2014 to 2018, being the time when the main changes in economic policies to Cuba occur at the attention and activities are on the polygons nearby the port area.

In the years 2016 to 2017 among the most interesting ROIs spatial transformations is the increase in size of the ROI 248 in 2016 to the ROI 252 in 2017, in 2017 Obama as the president of United States visits the island, as a result of the

rapprochement between the two countries, in 2016 travel by sea is allowed, allowing cruise companies to enter the country.

After applying the Hausdorff distance we obtain the final ROIs shown in Figure *10*.

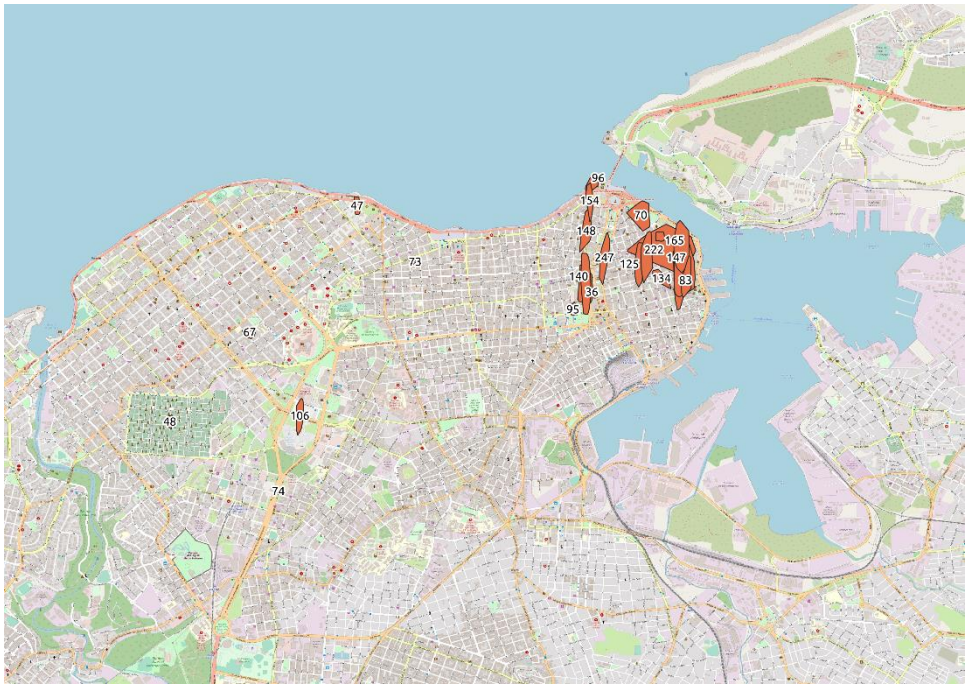

Figure 10. ROIs obtained after apply similarity section

2.2.5 Temporal Analysis

In this section, an analysis is made upon the results obtained of the previous sections, in order to understand the relation between the main political events in the country and in this way answer the research question of this work.

Figure *11* shows the temporal distribution of tourists visiting Havana, from the year 2004 to 2019, for the 10 countries that most originated visitors to the city during such a period. This image shows the dynamics in the visits of the Unites States visitors that reaches its highest point in the year of 2017, being consistent with the approach between this country and the Cuba government, which was consolidated with Obama's visit in 2016, but implied the application of different economic and political decrees that were the elimination of "State sponsor of terrorism" in 2015, the regularly scheduled

flights between the U.S. and Cuba in 2016 and the and the beginning of travel by sea in 2016.

Canada has been one of the countries that has always maintained relations with Cuba and Canadian tourism has always been one of the constants on the island. Figure *11* shows also that Canada is the second country of origin of visitors to the city, this figure shows that the dynamics of this country starts to increase after the year 2007, being 2007 the same year that Canadian companies began to promote the transfer of patients from Canada to Cuba to receive medical attention through so-called health tourism.

Spain has had relations with Cuba on a regular basis, maintaining the investment leadership in Cuba since 1988. Figure *11* shows an intense dynamic for Spain since the year 2006, reaching the highest point in the year 2007, which is consistent with the main political and economic events between Spain and Cuba. In 2007, the official cooperation ties were reestablished with the visit to the island of the Minister of Foreign Affairs as part of the new policy of Spanish President Rodriguez Zapatero. As part of this new relationship "Sol Melia", which is a Spanish company that has several hotels around Cuba, opens its 25th hotel in Havana. In this sense, collaboration and development aid actions are agreed between both countries, which will be carried out during the period 2007-2010.

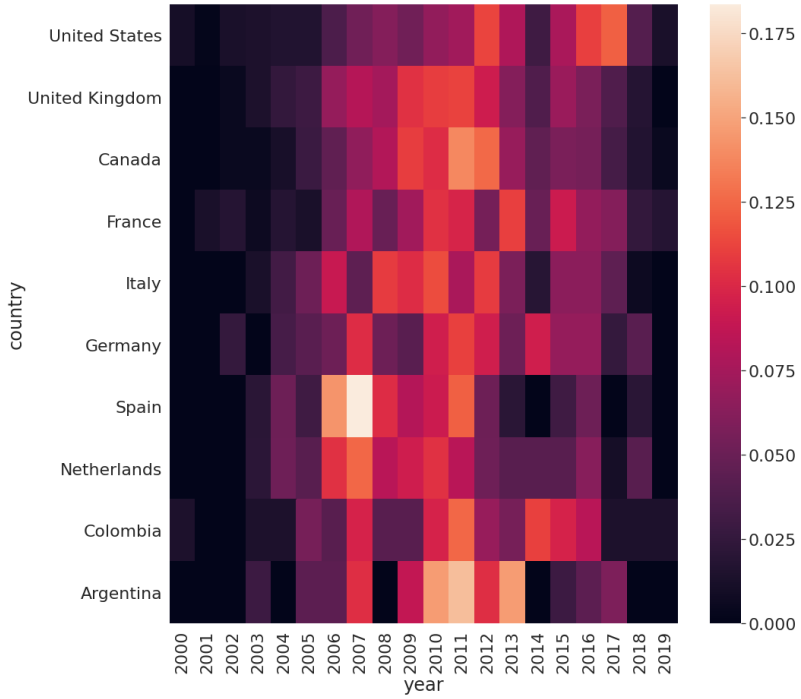

Figure 11. Temporal distribution of visits to Havana per country of origin.

The dynamics in Figure *11* for all the countries start to increase after the years 2006 and 2007, according to the Cuba government itself they started to implement a series of measures to make Cuban tourism more competitive than the Caribbean. As part of the measures was working on new investments and repairing hotels of historic interest in the cities, add new services and make our offer more competitive in general.

Figure *12* shows the dynamic of the ROIs through the years, resulting from Section 2.2.3.3. In this figure ROI 61 is the polygon that most shows activity through the years, the area represented by polygon 61 is the area of Old Havana, being the main area of tourism in Havana. A dynamic shown in this figure is between the ROI 140 and ROI 147, in the year 2017 and 2018, the ROI 147 is the ROI representing the port area, which increases in 2017, which is consistent with the activity in the port area due to income of cruises as a result of the relations between Cuba and the United States. The ROI 67 is located in an area of Vedado in Havana, this area is well known for house rental activity for foreigners, the interesting in this dynamic is that until the 2010 this activity was illegal in the country, but it is in 2010 when this ROI loses interest, this is mainly an effect that happened in Cuba when the activity becomes legal many of the rental houses does not meet the requirements and they had to close.

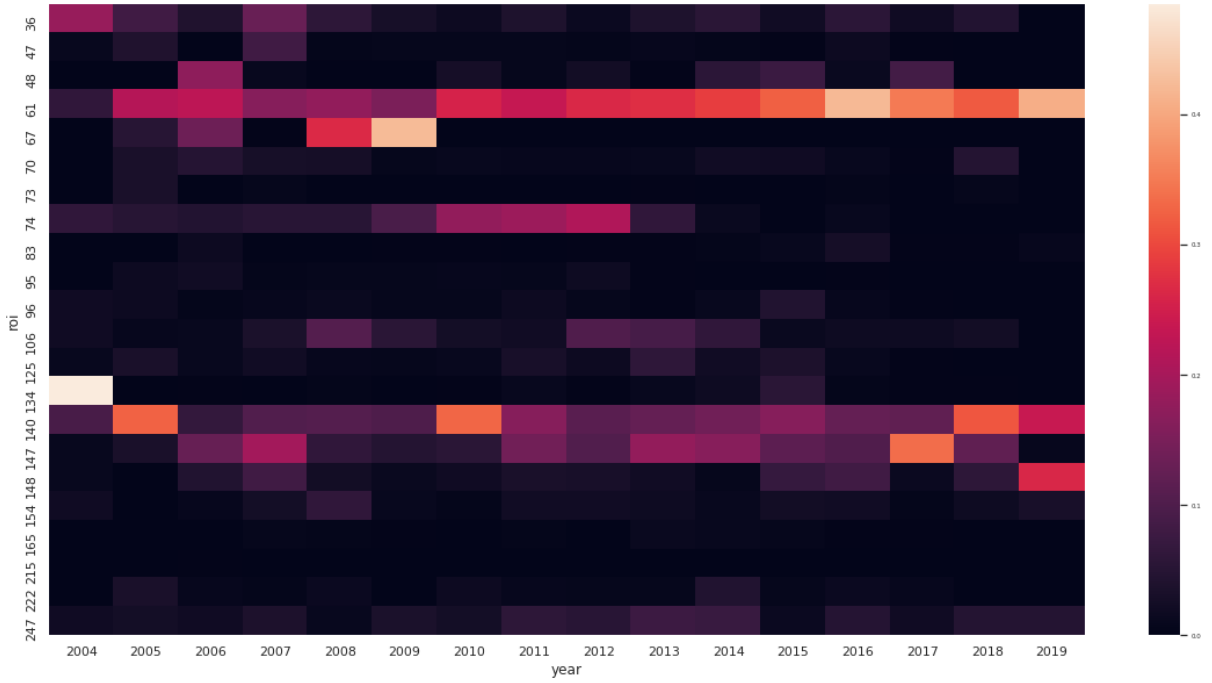

Figure 12. Temporal distribution of visits to Havana per ROIs. Values are normalized according to the sum of each row.

Figure *13* shows the number of photos taken by users of origin of the countries that most visit the island in the ROIs that were shown in Figure *10*. This figure shows that the ROI 61, in the countries of United Kingdom, France, Germany and Australia increase the activity in 2016 and then decreases from 2017, however in the case of the United States the activity for this polygon increase in the year of 2017, this event means that in 2017 the bigger activity in the port area and Old Havana was the activity by American tourists.

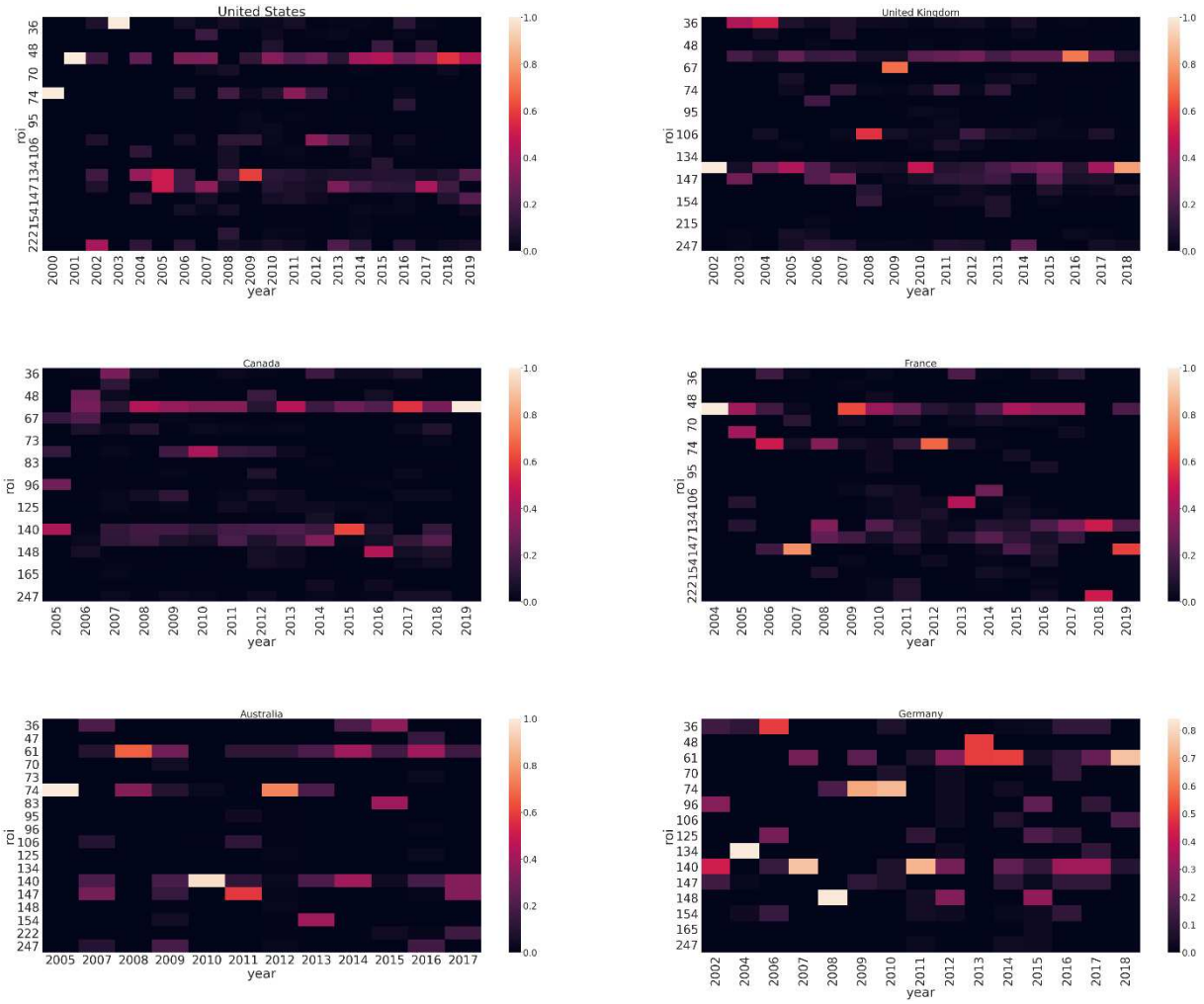

Figure 13. Temporal distribution of visits to Havana per country of origin and per ROIs. Values are normalized according to the sum of each row.

2.2.6 Conclusions

This work starts by implementing a clustering method in order to find the spatio-temporal polygons of the region of Havana, then a temporal analysis is made to try to find a relation between the main political and economic events and tourism activity.

The result of this work suggests that the changes in the United States policies regarding Cuba had a significant impact on the flow of the American tourists visiting Havana. In the same way Canada's policies towards the island, in response to the series of measures that were applied on the island to increase the tourism, had an effect in the flow of Canadians visiting the city concluding not only that these policies had an effect in the Canadian tourism but that the measures applied by the government of Cuba were effective regarding to increase the flow of Canadians visitors. Also, the

visit of the president of France in 2015 to Cuba had an impact and increased the visits in 2016 where agreements were signed between the two countries.

Due to the result of the analysis, this work concludes that there is a relation between the political events of the main foreign countries that visited the island and the tourism in Cuba. This work also suggests that the internal policies of the country of the approval of the activity of rental houses in Cuba, had an effect on existing businesses of that type in the city of Havana.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and Fundação de Amparo à Pesquisa de Minas Gerais (Fapemig - Project APQ – 03763-12).

References

[1] Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Z. (2011). Exploring millions of footprints inlocation sharing services.ICWSM, 2011:81–88.

[2] Dimobe, K., Goetze, D., Ou´edraogo, A., Forkuor, G., Wala, K., Porembski, S., and Thiom-biano, A. (2017). Spatio-temporal dynamics in land use and habitat fragmentation within aprotected area dedicated to tourism in a sudanian savanna of west africa.Journal of LandscapeEcology, 10:75 – 95.

[3] Edelsbrunner, H., Kirkpatrick, D., and Seidel, R. (1983). On the shape of a set of points inthe plane.IEEE Transactions on information theory, 29(4):551–559.

[4] FERRER, L. I. (2018). Mintur: 2019 ser´a un a˜no de retos para el turismo cubano

.[5] Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal,69(4):211–221

[6] Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., and Toivonen,T. (2017). User-generated geographic information for visitor monitoring in a

national park:A comparison of social media data and visitor survey.ISPRS International Journal of Geo-Information, 6(3):85.

[7] Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., and Prasad, S. (2015). Extracting andunderstanding urban areas of interest using geotagged photos.Computers, Environment andUrban Systems, 54:240–254.

[8] Kuo, C.-L., Chan, T.-C., Fan, I., Zipf, A., et al. (2018). Efficient method for poi/roi discoveryusing flickr geotagged photos.ISPRS International Journal of Geo-Information, 7(3):121.

[9] Li, L., Goodchild, M. F., and Xu, B. (2013). Spatial, temporal, and socioeconomic patternsin the use of Twitter and Flickr.Cartography and Geographic Information Science.

[10] Shonkwiler, R. (1991). Computing the hausdorff set distance in linear time for any lp pointdistance.Information Processing Letters, 38(4):201–207.

[11] Spiegel, J., Gonzalez, M., Cabrera, G., Catasus, S., Vidal, C., and Yassi, A. (2008). Promot-ing health in response to global tourism expansion in cuba.Health Promotion International,23(1):60–69.

[12] Sun, Y., Fan, H., Bakillah, M., and Zipf, A. (2015). Road-based travel recommendationusing geo-tagged images.Computers, Environment and Urban Systems, 53:110–122.

[13] Yan, Y., Eckle, M., Kuo, C.-L., Herfort, B., Fan, H., and Zipf, A. (2017). Monitoring and as-sessing post-disaster tourism recovery using geotagged social media data.ISPRS InternationalJournal of Geo-Information, 6(5):144.

[14] Yun, H. J., Kang, D. J., and Lee, M. J. M. (2018). Spatiotemporal distribution of urbanwalking tourists by season using GPS-based smartphone application.Asia Pacific Journal ofTourism Research.

3 CONCLUSIONS

This work addressed an important issue for countries where tourism is the first sector of the economy. It started looking at the tourism dynamics in Havana in the last few years, and we tried to correlate these dynamics with important events concerning the relationship between the Cuba government and the main countries that visited the island.

Spatial footprints are the type of data used for this work, obtained from Flickr as a type of VGI data. The results from the first article suggest that the changes in the United States' policies with regard to Cuba had a significant impact on the flow of US residents visiting Cuba. Due to this impact on the increase in the visits from Americans to the island, it was concluded that there is a connection between the political event and the tourism, and suggesting that there was a political tourism in that time lapse. However, regarding the policies applied by the Cuban government to encourage tourism, the first article concludes that there was no evidence of impact on tourist visits to the city.

Regarding the clustering method used in the second article this work concludes that the spatio-temporal polygons of the region of Havana obtained suggest that the series of measures that were applied on the island to increase the tourism, had an effect in the Canadian tourism to Havana, also the rapprochement between the Cuban government and Canadian had an effect in the tourism in the city.

This work concludes that there is a relation between the political events of the main foreign countries that visited the island and the tourism in Cuba, suggesting also that the internal policies of the country of the approval of the activity of rental houses in Cuba, had an effect on existing businesses of that type in the city of Havana.

As future work it is suggested to be analyzed in greater depth the ROIs that were visited by the American tourists in a way to look for political interests. Also, these ROIs can be analyzed in order to understand how the new constructions and renovations affected the tourists' interest.

4 REFERENCES

AGGARWAL, C. C. **Data mining:** the textbook. [S.l.]: Springer, 2015.

CHENG, Z; CAVARLEE, J; LEE, K. SUI, D. Z. Exploring millions of footprints in location sharing services. **ICWSM**, v. 5, n. 1, p. 81–88, 2011.

DE TRÉ, G.; VAN BRITSOM, D.; MATTHÉ, T.; BRONSELAER, A. Automated cleansing of POI databases. In: _____ **Quality Issues in the Management of Web Information**. [S.l.]: Springer, 2013. p. 55–91

ELIAS, B. Extracting landmarks with data mining methods. In: INTERNATIONAL CONFERENCE ON SPATIAL INFORMATION THEORY,2003, Ittingen. **Proceedings…** Kartause Ittingen, Switzerland: Springer Berlin Heidelberg.2003. p. 375–389

FERRER, L. I. MINTUR: 2019 será un año de retos para el turismo cubano. **Cubadebate contra el terrorismo mediatico**, Havana, 27 December 2018. Disponivel em: <http://www.cubadebate.cu/noticias/2018/12/27/mintur-2019-sera-un-ano-de-retos-para-el-turismo-cubano/>. Acesso em: 15 jan. 2020.

GOODCHILD, M. F. Citizens as sensors: the world of volunteered geography. **GeoJournal**, v. 69, n. 4, p. 211–221, 2007.

HEIKINHEIMO, V.; MININ, E. D.; TENKANEN, H.; HAUSMANN, A.; ERKKONEN, J.;TOIVONEN, T. User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. **ISPRS International Journal of Geo-Information**, v. 6, n. 3, p. 85, 2017.

HU, Y.; GAO, S.; JANOWICZ, K.;YU, B.;LI, W.; PRASAD, S. Extracting and understanding urban areas of interest using geotagged photos. **Computers, Environment and Urban Systems**, v. 54, p. 240–254, 2015.

KUO, C. L., CHAN, T. C., FAN, I. C., ZIPF, A. Efficient Method for POI/ROI Discovery Using Flickr Geotagged Photos. **ISPRS International Journal of Geo-Information**, v. 7, n. 3, p. 121, 2018.

MIN, D. E. N. G.; ZHILIN, L.; XIAOYONG, C. Extended Hausdorff distance for spatial objects in GIS. **International Journal of Geographical Information Science**, v. 21, n. 4, p. 459–475, 2007.

RAO, K. V.; GOVARDHAN, A.; RAO, K. V. C. Spatiotemporal data mining: Issues, tasks and applications. **International Journal of Computer Science and Engineering Survey**, v. 3, n. 1, p. 39, 2012.

SUN, Y.; FAN, H; HELBICH, M.; ZIPF, A.. Road-based travel recommendation using geo-tagged images. **Computers, Environment and Urban Systems**, v. 53, p. 110–122, 2015.

YAN, Y.; ECKLE, M.; KUO, C. L; HERFORT, B.; FAN, H.; ZIPF, A. Monitoring and assessing post-disaster tourism recovery using geotagged social media data. **ISPRS International Journal of Geo-Information**, v. 6, n. 5, p. 144, 2017.