Mining Frequent Trajectory Patterns and Regions-of-Interest from Flickr Photos

Guochen Cai
School of Business (IT)
James Cook University
Cairns QLD4870, Australia
Guochen.Cai@my.jcu.edu.au

Chihiro Hio
School of Business (IT)
James Cook University
Cairns, QLD4870, Australia
Chihiro.Hio@my.jcu.edu.au

Luke Bermingham
School of Business (IT)
James Cook University
Cairns, QLD4870, Australia
Lukeh-B@hotmail.com

Kyungmi Lee
School of Business (IT)
James Cook University
Cairns, QLD4870, Australia
Joanne.Lee@jcu.edu.au

Ickjai Lee
School of Business (IT)
James Cook University
Cairns, QLD4870, Australia
Ickjai.Lee@jcu.edu.au

Abstract

Flickr represents a massive opportunity to mine valuable human movement data from geo-tagged photos. However, existing Flickr trajectory data mining research has not considered mining frequent trajectory patterns whilst also considering the temporal domain. Therefore, a significant opportunity exists to demonstrate the application of a pattern mining algorithm to a large geo-tagged photo dataset. Thus, we present a novel application of the trajectory pattern mining algorithm to a 2012 Flickr dataset of Australia and encompassing state, Queensland. In our experiments we show that many interesting, previously unknown patterns discovered through our framework. Our framework is able to discover expected major landmarks such as cities and tourist attractions. In addition, we make the notable discover of what is theorized to be valuable tourist travel information about sequential movements between hot-spot attractions.

1. Introduction

A single photograph can capture a moment in time that describes where someone was and what he/she was doing. With this in mind, a series of photos, presented chronologically, can describe the approximate spatio-temporal movements of an individual. A collection of spatio-temporal entries connected to represent movement is known as a *trajectory*. Analysis of multiple photo trajectories can reveal valuable, previously unknown, information such as frequent travel patterns and Regions-of-Interest (RoI).

Only recently has the process of converting photo data into trajectories been a feasible activity. The reason for this change is attributed solely to geo-tagging. Geo-tagging is a technology that includes a geographic reference inside the meta-data of specific types of content: photos, videos, and SMS. The effect geo-tagging has had on the amount of valuable data available for extraction is profound. Flickr, one of the most popular photo sharing sites on the web, reported in 2011 that their photo upload count was “about 4.5 million daily”. Clearly, photo sharing and social media sites such as this present a potential ocean of valuable information that can be harnessed.

It is this assertion that forms the basis for this paper. Specifically, in this paper we first extract a mass of geo-tagged photographic information from the Flickr API. We extract two datasets containing different spatial regions: (1) the whole of Australia, (2) a finer grained snapshot of the tourist popular state of Queensland. We then use the data extracted from these regions to construct a series of spatio-temporal trajectories. Following this, we then present a novel application of the Trajectory Pattern Mining (TPM) algorithm to discover interesting frequent patterns and RoI contained in the data. Our results are promising and therefore validate the endeavor of mining frequent sequential patterns from Flickr datasets.

However, in contrast to our research the majority of existing Flickr trajectory mining research focus on clustering trajectories [15,16] or finding association rules [7]. However, to the best of our knowledge this is the first study to show the application of sequential

978-1-4799-2504-9/14 $31.00 © 2014 Crown Copyright
DOI 10.1109/HICSS.2014.188

IEEE computer society

trajectory pattern mining to Flickr datasets. Of course the mining of sequential trajectory patterns has been studied in several other domains in the recent decades [3,5,17]. For example both [3] and [17] mined people's travel sequences using GPS trajectory data. In addition, [5] conducted research very similar to ours that extracted sequential movement patterns from geo-tagged photos. However, the main difference between the previous studies and our research is the consideration of the temporal domain. In all the presented studies only the spatial attribute was considered when constructing frequent sequential patterns. Clearly the exclusion of the temporal dimension greatly limits the usefulness of the constructed patterns. Specifically, the temporal dimension is an important feature of trajectory data, which can provide rich and specialized detail about the underlying structural patterns in the data. Overall, it should now be evident that the application of TPM [2] to Flickr data is unique to our research.

Additionally, the application of TPM is not only unique but also quite significant. This is because the sequential trajectory patterns it produces highlight the most frequent sequence of locations visited by the photo-takers. By uncovering accurate information about photo-takers' movement behavior, various application areas can be extremely benefited. Specifically we propose the resultant patterns found by applying TPM to Flickr data would be of great interest to the industries of: tourism, retail, transport, and marketing.

The rest of paper is organized as follows. Section 2 briefly outlines preliminaries on Flickr data mining and TPM. Section 3 introduces our framework for mining frequent trajectory patterns. Section 4 reports the results found by applying TPM to the extracted Flickr data. Finally, Section 5 concludes some final remarks of the study and introduces the direction of possible future work.

2. Preliminaries

2.1. Flickr data mining

Massive amounts of photos are uploaded to Flickr each day. These photos often contain additional metadata that can be mined to form interesting patterns about photo-takers. Unsurprisingly, this has caused Flickr data mining to become a hot area of research. In general past Flickr data mining research can be categorized into the following four categories: clustering, Association Rules Mining (ARM), path recommendation and tag analysis.

2.1.1. Flickr clustering. The basic concept of Flickr clustering is to partition geo-tagged photos into similar groups in order to maximize inter-group dissimilarity and minimizes intra-group dissimilarity. There have been several clustering methods proposed in the literature [1,5,8,16]. Geo-referenced photos indicate Points-of-Interest (PoI) where relatively a large amount of photos is taken. Zheng *et al*. [16] investigate regions of attractions that are similar to PoI, and use them for route analysis. Lee et al. [8] proposed one framework to discover geographical hierarchical PoIs and PoIs in several temporal categories. Moreover, based on the extracted PoI, [16] found the popular travel routes presented as a set of visited PoIs. *k*-means and DBSCAN variants are two popular approaches for PoI mining. Kisilevich *et al*. [5] used a DBSCAN variant in their study for POI identification whilst Crandall *et al*. [1] investigated a similar PoI problem with *k*-means variants.

2.1.2. Flickr ARM. Flickr ARM is to find positive associative patterns given user specified minimum support and confidence. There have been few Flickr ARM proposed in the literature [7]. Lee *et al*. [7] proposed one PoI ARM framework. They extracted popular PoI from geotagged photos first, and then applied ARM techniques to find PoI association rules. The main different between their research and this paper, is that this paper makes consideration of the temporal domain when mining sequential trajectory patterns.

2.1.3. Flickr path recommendation. In a somewhat different area, some recent studies have focused on route recommendation systems [6,9,10,14]. These approaches model geo-tagged images as a sequence of location points, and mine travel sequences. Shi *et al*. [14] combine user-landmark preference and category-based landmark similarity to provide personalized landmark recommendation. Similar work is seen in Lu *et al*. [9] where users are able to specify personal preference in the travel route planning. Kurashima *et al*. [6] integrate topic models into Markov models to provide travel route recommendations whilst Okuyama *et al*. [10] use trip models represented by the order sequences of tourist places to compute a travel plan. These approaches do not reveal sequential trajectory patterns, but mainly focus on travel route recommendations.

2.1.4. Flickr tag mining. Metadata tags associated with photos are a rich source of potential information. Kennedy *et al.* [4] use the concept of representative tags and tag-driven approach to extract place and event semantics. Rattenbury *et al.* [11,12] conduct similar research and investigate ways to extract place and event semantics from folksonomy. [13] uses language model with their own proposed tag-based smoothing and other techniques to predict the single most likely location of a photo to place image on a map

2.1.5. Discussion. Flickr data mining has been studied in a variety of previous works, all with their own specific goals of the type of information they wish to extract. However, there is no research on mining frequent sequential trajectory patterns from Flickr data. This is extremely significant because sequential trajectory patterns have the ability to demonstrate photo-takers movement behaviors. Understanding where people are moving when they take photos is highly beneficial to many industry and business entities. Specifically, we demonstrate that our framework produces results that are useful to: tourism related businesses, government and researchers.

2.2. Sequential TPM

The aim of sequential TPM is to identify frequent patterns of moving objects from identified RoI and temporal annotations; formally the produced result is called as a *trajectory pattern*. There are only few algorithms for spatiotemporal mining. TPM, which was proposed by Giannotti *et al.* [2], is the most widely adopted and this one is applied in this paper.

TPM first tessellates the study region into grid cells based on a user provided grid resolution. It computes grid densities from trajectories. Dense grid cells expand to neighboring grids forming a larger dense rectangle as long as the average density of the newly formed rectangle satisfies a user provided density threshold. Frequently visited newly formed rectangles become RoI. Each trajectory is now represented by a series of these newly formed RoI. Prefix pattern mining is finally applied to find interesting frequent trajectory patterns.

3. Framework of frequent trajectory pattern mining

Figure 1 displays the framework of frequent trajectory pattern mining from Flickr photos. Flickr API is used to collect geo-tagged photos in the framework consisting of three main components: pre-processing, TPM and visualization.

First, the geo-tagged photo data have noise and redundancy. These noisy data require pre-processing to clean and remove faulty data. Since time is an important factor of TPM, it is crucial to ensure that data have a correct time attached. Some data have an incorrect time attached like “2012-00-00”. Stripping those incorrect time annotated photos is used as a cleaning process in this paper. In addition to this, if a photo does not have an EXIF date, Flickr automatically sets a photo taken time to the time of upload, which causes time duplication. Another noise factor is from spatial outliers (extreme longitude and latitude). Trimming those outliers is essential. Moreover, Flickr API may lead to many duplicate data. Removing those superfluities based on photo-id and time is also required in order to remove unwanted deviation. A data format of preprocessed geo-tagged photos is converted into a specific format of TPM as described in Figure 2.

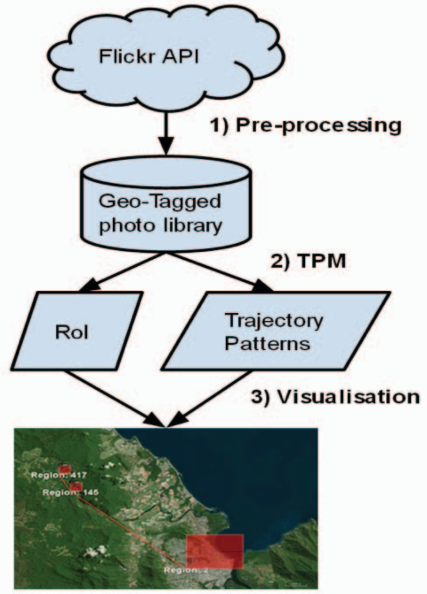

Figure 1: Framework of frequent trajectory pattern mining.

userid, # of photos, { [time, latitude, longitude] … }

Figure 2: Data format for TPM.

Second, after pre-processing TPM is in place to discover RoI and frequent trajectory patterns. TPM requires three user provided input parameters: *MinSup*, *CellSize, and Time Tolerance*. *MinSup* determines a threshold of support for RoI and trajectory patterns. The value of *MinSup* represents a ratio in total (e.g. 0.1 means 10% in total). *CellSize* is used to define a grid size for RoI. The value of *CellSize* represents a side of rectangle shape (e.g. 0.01 means 1k square area.). *Time Tolerance* determines an acceptable range of similarity of time annotation. (e.g. 1 day means the gap of the two time annotations must be the same or less than 1 day. Since 1 day is applied to both previous and next moving directions, the intersection of similarity range can be found with a 2 days gap.)

Third, in order to clarify and analyze mining results, a geo-overlaid visualizer over the World Wind Java SDK[1] is developed to present detected patterns. In this paper, a software package written in JAVA programming language was developed for the second step TPM algorithm and the third step geo-overlaid visualizer.

4. Results and discussion

Tourism is one of the biggest industries in Australia, and Queensland (the second largest state) is a popular tourist destination hosting the Great Barrier Reef and tropical rain forest. It is located in tropics and famous for various water activities and tropical adventurous activities. Understanding the photo-takers' behaviour in particular spatial-temporal movement patterns, not only can provide people who plan to visit Australia or Queensland, recommendations about RoI and frequent trajectories the previous travelers taken, but also help government and business make strategies to serve tourists well.

4.1. Parameter selection

The TPM algorithm that is applied in this paper is highly sensitive to parameter selection as with many other data mining algorithms. It relies heavily on the minimum support (*MinSup*) value for cell to become a region of interest and also on the size of cell (*CellSize*) that is used to partition the study region. It is a non-trivial problem to choose best values of parameters which will produce meaningful and informative patterns. Thus, the approach adopted during experimentation was a systemic trial and error approach, in which the *MinSup*, *CellSize* and resultant number of RoI were recorded. After thorough experimentation of diverse values of parameters it was concluded that the following parameters produced the most interesting and comparable results. Only experimental results of these values are presented in this paper.

Table 1: Values of parameters used in this study.

	Australia	Queensland
# of Entries	383,335	62,290
# of Trajectories	7319	2445
MinSup	0.025	0.025, 0.001
CellSize	0.08	0.02, 0.001
Time Tolerance	1, 10, 60 days	1 day

The number and the size of RoI mined are affected by the values of parameters *MinSup* and *CellSize*. To estimate best values for these two parameters, several experiments have been performed which suggest *MinSup* = 0.0025 and *CellSize* = 0.08 for the RoI mining in the entire Australia. At the value 0.08 which means 8km for the parameter *CellSize*, clear separations can be identified between major land marks. For the parameter *MinSup*, 0.0025 which equals to 0.25% is the appropriate value which makes the size of attraction locations not too small.

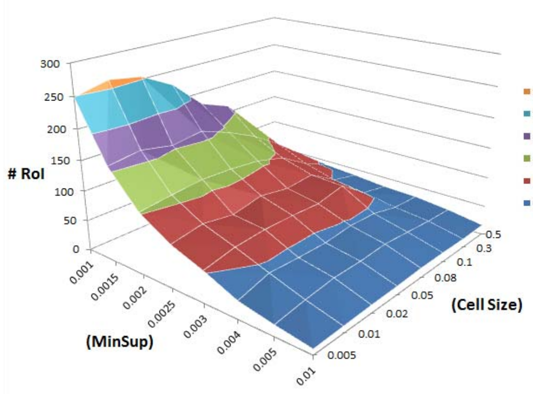

Figure 3: Number of RoIs in different values of *MinSup* and *CellSize* (Australia data).

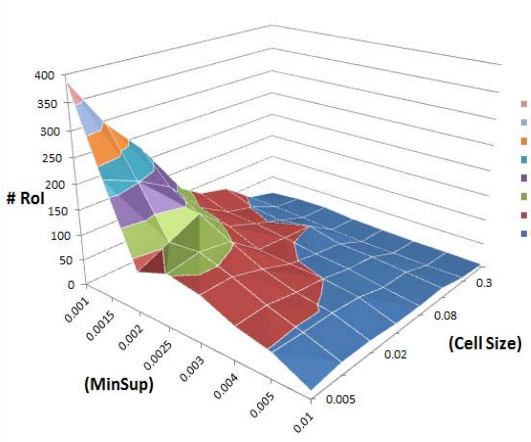

Figure 4: Number of RoIs in different values of *MinSup* and *CellSize* (Queensland data).

Figure 3 demonstrates the trend of changes of number of RoI with different values of *MinSup* and *CellSize* for Australia dataset and Figure 4 shows the trend for Queensland dataset. As a result of *MinSup* and *CellSize* affecting the number of RoI, the same trend can be noted that decreasing the values of *MinSup* and *CellSize* can generate more RoI while the

[1] http://worldwind.arc.nasa.gov/java/

change is sharp in Queensland dataset. This is particularly evident when the parameters are set to be small. The effect is that the majority of data becomes a RoI. Using 0.0025 as the value of parameter *MinSup* and 0.08 as the value of parameter *CellSize*, we test several values of parameter *Time Tolerance* to find the interesting patterns and the changes of number and length patterns.

Like the trend of number of RoI, the number of patterns is affected by the values of parameter *MinSup*, *CellSize* and *Time Tolerance*. The first two parameters can determine the size and number of RoI, whilst the first and third parameters can determine the number and length of patterns. Specifically, decreasing the values of *MinSup* and *CellSize*, small size and more RoI can be found, and with this value of *MinSup*, increasing the value of *Time Tolerance* can generate more patterns and longer length patterns can be found.

Figure 5 shows that the increase of value of tolerance makes the running time grow dramatically. Using long tolerance, more trajectories will be calculated as the frequent patterns.

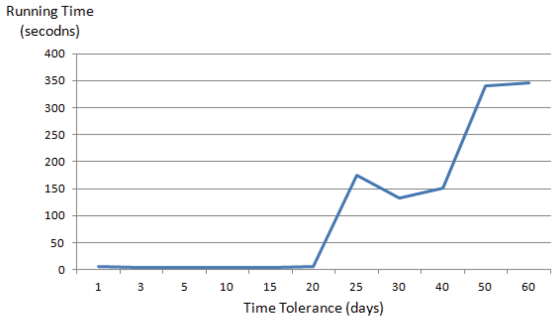

Figure 5: Running time for different values of *Time Tolerance* (*MinSup = 0.0025, CellSize = 0.08*).

4.2. Australia

In Section 4.1, it is shown that the Australian data has 383,335 records, of which there are 7,319 trajectories in total. Figure 6 illustrates the distribution of the data throughout Australia. Unsurprisingly, the majority of dense data correlates to the highly populated coastal towns.

4.2.1 RoI. Using parameters defined in Section 4.1, the RoI found are shown in Figure 7. Specifically, Figure 7 displays 74 RoI in the Australian study region. The red rectangles on the map show the RoI and their size. We can clearly see that there are 8 big size RoI (Perth, Adelaide, Melbourne, Sydney, Gold Coast, Brisbane, Sunshine Coast and Cairns). As mentioned, the majority of major cities in Australia are along the coast. Thus, it validates the effectiveness of our approach to see that the majority of the resultant RoI that are discovered are distributed along the coast.

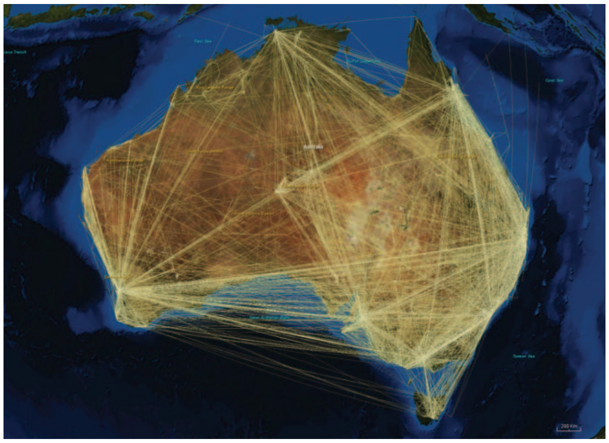

Figure 6: Distribution of the entire trajectories (|*n*| = 7,319).

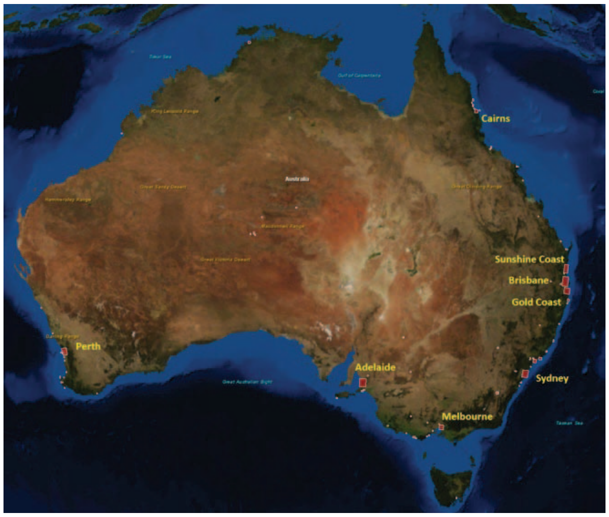

Figure 7: RoI in entire Australia (*MinSup* = 0.0025, *CellSize* = 0.08, number of RoI = 74).

Table 2 illustrates the top 10 RoI from the entire Australian dataset. The first five places are the capital cities of New South Wales state, Queensland state, Western Australia state, Victoria state and South Australia state respectively. Gold Coast is for Surfers Paradise and many fun parks, Sunshine Coast is a center for tourism which contains Australia Zoo and several annual sporting events, Cairns is a home for the Great Barrier Reef, and Byron Bay is a beachside town where there are several beaches which are popular for surfing. It is interesting to find that the majority of last five RoI are all in Queensland, most likely because of its abundance of tourist destinations.

Table 2: Top ten RoI.

	RoI	Support
Rank1	Sydney	1,237
Rank2	Brisbane	1,105
Rank3	Perth	905
Rank4	Melbourne	820
Rank5	Adelaide	658
Rank6	Gold Coast	585
Rank7	Sunshine Coast	325
Rank8	Cairns	231
Rank9	Newcastle	224
Rank10	Byron Bay	155

4.2.2 Frequent patterns. Using parameters defined in Section 4.1, we perform experiments to find the frequent patterns in the Australia dataset. In this experiment, it has been found that when the *Time Tolerance* is less than 10 days most of the frequent patterns found are about the movement between nearby RoI. Moreover, at these values, all the patterns found are 2-length which means all the patterns found are the movement between two RoI.

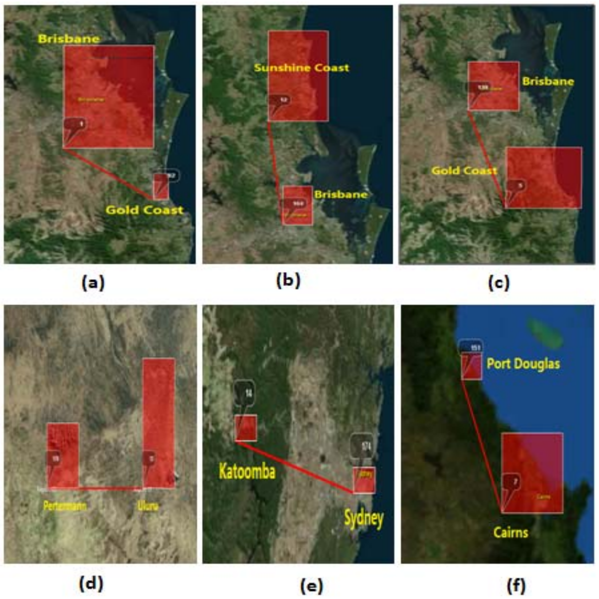

Figure 8: Frequent patterns with 1 day tolerance: (a) from Brisbane to Gold Coast; (b) from Sunshine Coast to Brisbane; (c) from Gold Coast to Brisbane; (d) from Uluru to Petermann; (e) from Katoomba to Sydney; (f) from Cairns to Port Douglas (*MinSup* = 0.0025, *CellSize* = 0.08, *Time Tolerance* = 1 day).

Figure 8 depicts 6 frequent trajectories patterns with 1 day as the *Time Tolerance*. 3 out of these 6 patterns are:

- Uluru [0 day, 1 day]→ Petermann;
- Katoomba [1 day, 1 day] → Sydney;
- Cairns [0 day, 1 day] → Port Douglas.

Photo-takers move to Petermann from Uluru in the same day and also to Sydney from Katoomba. People visit Cairns and then go to Port Douglas in the same day or the second day. Each of these three patterns has only one time annotation, but for the other three patterns, each has more than 1 time annotation that is illustrated in Table 3. Brisbane, Gold Coast and Sunshine Coast are three nearby popular destinations for tourists that lead to many people' visit and many random movements generated by these people. These photo-takers may spend some different days in one place and go to another place.

Table 3: Frequent patterns with more than 1 time annotation (1 day tolerance).

Frequent Sequence	Time Annotation
• Brisbane → Gold Coast	[8 days, 8 days] [5 days, 6 days] [0 days, 3 days]
• Sunshine Coast → Brisbane	[4 days, 4 days] [5 days, 6 days] [2 days, 2 days] [6 days, 6 days] [0 days, 2 days] [2 days, 2 days]
• Gold Coast → Brisbane	[12 days, 12 days] [57 days, 57 days] [15 days, 17 days] [23 days, 24 days] [24 days, 24 days] [24 days, 25 days] [13 days, 15 days] [12 days, 12 days]

However, when the *Time Tolerance* is 10 days, sizable patterns between far from RoI come out. Actually, most of the patterns are about the movement between the top ten RoI. During these frequent patterns, 7 patterns start from Brisbane. Figure 9 shows the patterns starting from Brisbane. Besides the patterns from Brisbane to nearby popular RoI which are Sunshine Coast, Gold Coast, Byron Bay, there are also from Brisbane to other three far away cities which are Cairns, Sydney and Melbourne.

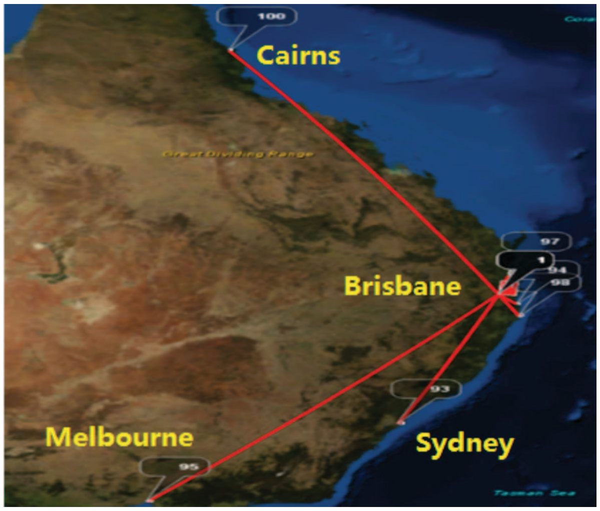

Figure 9: Patterns starting from Brisbane.

As highlighted in Section 4.1, experiments with parameters, specifically using 60 days as *Time Tolerance* did produce some interesting 3-length patterns. By decreasing the values of these three parameters, many patterns between small attraction points can be found. Interestingly, when using 0.001 as value of *MinSup*, 0.01 as the value of *CellSize* and 1 day as the value of *Time Tolerance*, two 4-length patterns are found. These two patterns are both about the sequences between Kuranda and Cairns. The two 4-length patterns are:

- Cairn city [0 day, 1 day] → Barron Falls [0 day, 1 day] → Kuranda [0 day, 1 day] → Barron Falls;
- Skyrail Tjapukai [0 day, 1 day] → Barron Falls [0 day, 1 day] → Kuranda [0 day, 1 day] → Barron Falls.

Kuranda is a picturesque mountain retreat 25 km northwest of Cairns hosting rainforest skyrail, scenic railway, local shops and markets. The other terminal of skyrail is in Tjapukai which is an Australian aboriginal cultural destination. People take skyrail and railway which traffic between Kuranda and Cairns both via Barron Falls, a spectacular falls near Kuranda.

Figure 10 demonstrates the number of patterns with different values of *MinSup* and *Time Tolerance*. More patterns can be mined using smaller *MinSup* that means photo-takers move randomly and not so many typical trajectories attract most photo-takers' movements.

4.2.3 Conclusion. The results found from the entire Australia data reveal interesting patterns. To begin with, the results show that most of the identified frequent patterns are nearby movement, i.e. Brisbane to Gold Coast. A logical theory for this kind of trend is that in general more photo-takers are driving to the next major landmark and taking photos, rather than covering large distances by air travel. In addition, the context of visited locations provides some insight into this behavior as well. Specifically, many of the nearby frequent patterns are trips between major cities and local tourist attractions. Another interesting observation was that when a 1 day temporal window was used there was nearly no 3-length patterns. Due to the nature of TPM long distance patterns are difficult to uncover.

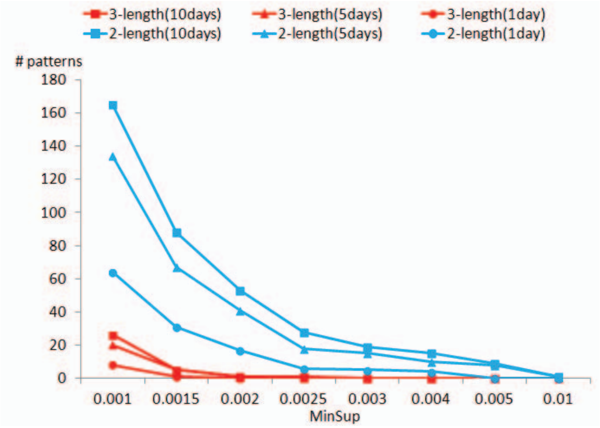

Figure 10: Number of patterns changing along with different values of *MinSup* & *Time Tolerance* (*CellSize* = 0.08).

4.3. Queensland

Dataset of Queensland is extracted from the entire Australia dataset in 2012. The dataset consists of 2,445 trajectories in total, made up of 63,290 of photo-taker locations.

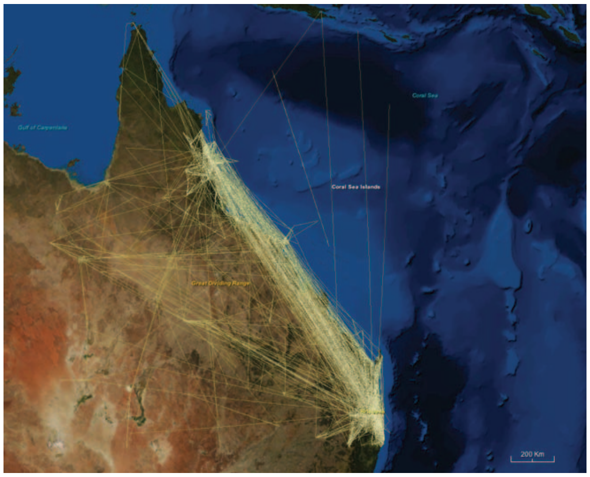

Figure 11: Distribution of the entire trajectories in Queensland in 2012 (|*n*| = 2,445).

Figure 11 displays the distribution of trajectories in Queensland in 2012. It is observed that the coastline is high density area, especially Brisbane and Cairns areas, notably the inland areas are far less dense.

4.3.1 RoI. Same as the entire Australia case, several experiments have been performed which suggest *MinSup* = 0.0025 and *CellSize* = 0.02 for the RoI mining in the Queensland. A distribution of RoI with these parameters clearly distinguishes the central coastal cities and their frequently visited photo-taker locations, i.e. tourist hotspots.

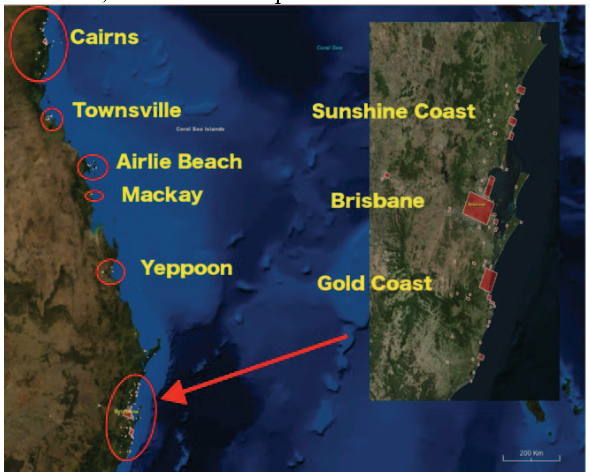

Figure 12: Distribution of RoI in Queensland.

Figure 12 displays 103 RoI in the Queensland region. These RoI are distributed along the major coastal cities of Queensland. To be specific the RoI located are: Brisbane, Gold Coast, Sunshine Coast, Cairns, Townsville, Airlie Beach, Mackay, and Yeppoon. Analysis of these RoI reveals the frequent locations of photo-takers in Queensland. To present an overview of this finding the top five RoI for the Queensland data are highlighted in Table 4.

Table 4: Top five RoI in Queensland.

	RoI	Support
Rank1	Brisbane	986
Rank2	Gold Coast	384
Rank3	Cairns	156
Rank4	Byron Bay	135
Rank5	Redcliffe	128

4.3.2 Frequent patterns. Based on the parameters selected in Section 4.1 TPM was able to generate five 3-length patterns and sixty 2-length patterns. The results show that a common trend in the 3-length patterns is *returning trips*. Returning trips are classified as sequences where a photo-taker has a starting place, goes somewhere, and then finally returns to the starting point. In Figure 13(a) and (b) both illustrate this case. This returning trip is interesting because these sequences are highly supported and occur within one day. Interestingly, Kuranda and Fitzory Island are two very popular tourist destinations. The former hosts many nearby tourist attractions such as SkyRail, Kuranda Scenic Rail, local market, Barron falls and zoos whilst the latter hosts Great Barrier Reef where tourists enjoy snorkeling and swimming with tropical fishes. Thus, it can be concluded that Cairns to Kuranda/Fitzroy Island must both be common day-trips in the region.

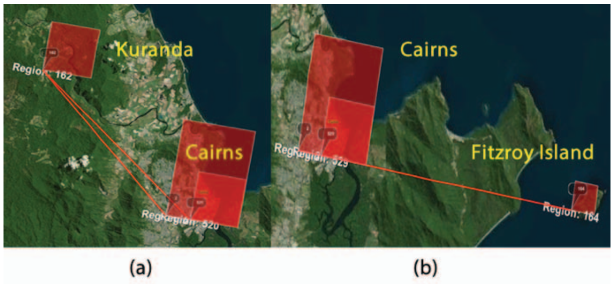

Figure 13: Two returning patterns.

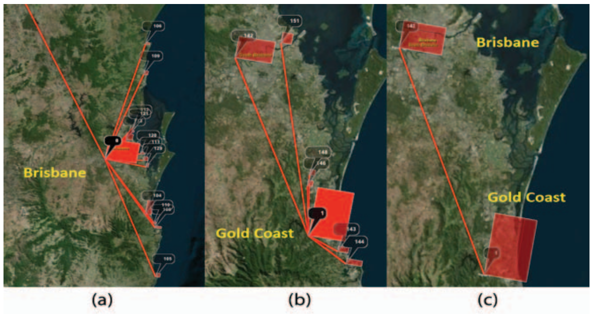

Figure 14: Notable 2-length patterns.

Figure 14 displays nearby movements are dominant over distance movements in 2-length patterns. Figure 14(a) has only one distance movements from Brisbane to Cairns. Most of movements in Figure 14(a) and (b) are within 2 hours driving range.

Gold Coast to Brisbane shown in Figure 14(c) has the highest support (132) and vice versa Brisbane to Gold Coast has a support of 125. The varieties of time annotations are observed from 0 day to 203 days duration time with a few days tolerance (e.g. [200 days, 203 days]). It implies that there are many types of people's movements. We can expect that local people use this path for work. Another possible explanation is local people go to attractions in Gold Coast and similarly Gold Coast residents go to Brisbane to shop in the city. Or when tourists go to either to this region of Queensland they will be inclined to visit both cities.

Although we set parameters to large values, the longest pattern in the experiment was found in Cairns area with parameters: *MinSup* 0.001, *CellSize* 0.001 and *Time Tolerance* 1 day.

.

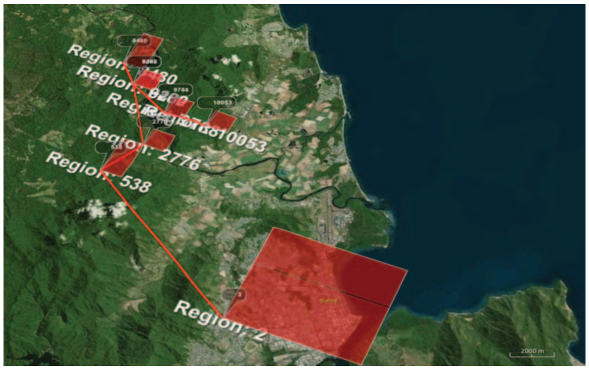

Figure 15: A 8-length pattern in Cairns area.

The support of this pattern is 3 out of 2,445 trajectories and the length of the pattern is 8. All the time annotations between regions are [0 day, 1 day]. To interpret it based on local background knowledge, it is a popular day trip tour in Cairns. Especially movements [2], [538], [2776], [6482], and [8480] are similar to the path of Kuranda Scenic Railway, and movements [8480], [9260], [9788], and [10053] are similar to the path of SkyRail. Therefore, fixed time transport service attributes for photos being taken at the same approximate times.

4.3.3 Conclusion. Due to the smaller study region, TPM was applied with granular parameters. Subsequently, many more 3-length patterns were uncovered. One interesting trend, which can be identified from these 3-length patterns, is that the majority of them are "day trips", returning to the starting point within 1 or 2 days. Another notable pattern that has emerged is that city centers are often used as a hub for photo-takers, who will then disperse amongst the suburbs of the city during the day. A further, interesting observation is the result of the overall most frequent 2-length pattern in the study region. Specifically, the bi-directional pattern from Brisbane to Gold Coast (and vice versa) constitutes the most common sequence of photo-taker movement. This pattern is highly realistic as both are major nearby cities, which have many attractions for tourists (fun parks and beaches), and are popular weekend destinations for locals living in either city.

4.4. Comparison

Many interesting comparisons can be drawn from the results of the two datasets. To briefly summarize the sequential patterns found from each area the analogy of "micro" and "macro" can be adopted. Namely, the Australian data highlights many spatially distant sequences, whereas the Queensland data discovers a variety smaller same city sequences. This result is not unexpected, as a direct consequence of the Queensland data being spatially and temporally more granular parameters applied than the entire Australian data. Specifically, as a result of the parameter difference between the two studied regions five 3-length patterns were found in Queensland, yet only one was found in Australia. Ideally, a greater number of longer length patterns could potentially be identified in Australia with different parameters. However, due to the density of the underlying data either: more intelligent pre-processing would have to occur, or the TPM algorithm needs to be modified to overcome to be time efficient to find long length patterns. Finally, an important similarity between the results of the two study regions is that both experiments were able to identify major cities, tourist locations, and expected tourist routes without any external user specification. Clearly this highlights the validity, robustness and applicability of our approach.

5. Final remarks

In this paper we have shown the effectiveness and applicability of our approach. Notably, interesting sequential patterns between expected major cities and hotspots were identified in both experiments conducted. In addition, majorly visited areas were identified without external user input. This supports the validity and applicability of our approach and it potentially demonstrates its usefulness in a wide variety of application areas. A notable finding of this research is the unexpected patterns discovered about human behavior, which can dramatically impact the effectiveness of the results. For example, it was necessary to perform pre-processing for data because of the photo-takers' tendency frequently taking many photos in the same area. A final remark about the usefulness of our approach is evident from the results that were uncovered demonstrate a previously unknown order in which photo-takers travel between regions. Namely, photo-takers visiting tourist locations may not necessarily follow the shortest path, but rather seem to have some other, previously unknown, preference in the order they visit attractions.

5.1. Future work

In our research we have shown that our framework is highly effective and applicable to large photo-taker datasets. This encourages us to continue research in the area of Flickr data mining. Specifically we identify the following interesting future works and improvements for our framework:

- *Global Flickr Mining:* It is theorized that improving our framework with suitable noise reduction procedures and filtering tactics would allow for the processing of global Flickr data. We expect this kind of Flickr mining could potentially produce informative patterns about international flight routes and global travellers;
- *Photo-taker classification*: To produce more specific results it is proposed that classification can be applied to the data first. In such a way the resultant dataset can be wholly of one class, i.e. all tourists. This kind of data filtering would vastly improve the quality of results by reducing noise;
- *Temporal Filtering:* Reducing the studied data to only a specific time period will allow for more information rich results to be extracted. We propose temporal filters such as seasonal changes, morning-evening, and weekends;
- *Tag Filtering:* In order to aid classification and reduce noise data the tags from Flickr photos should be used to extract only relevant entries. For instance, only photos tagged with “holiday” or “vacation” are acceptable;
- *TPM improvement:* Current TPM only extracts rectangular shapes of RoI, which are unrealistic in real world where arbitrary shapes of RoI are present. This could be applied to wider application areas such as animal trajectory analysis and smart city planning where RoI are not well defined.

6. References

[1] D.J. Crandall, L. Backstrom, D.P. Huttenlocher, and J.M. Kleinberg, “Mapping the World’s Photos”, *In Proc. of Int. Conf. on WWW*, 2009, pp. 761-770.

[2] F. Giannotti, M. Nanni, D. Pedreschi, and F. Pinelli, “Trajectory Pattern Mining”, In *Proc. of the Conf. on Knowledge Discovery and Data Mining*, ACM, New York, NY, 2007, pp. 330 - 339.

[3] J. Kang, and S-H. Yong, “Mining Spatio-Temporal Patterns in Trajectory Data”, *Journal of Information Processing Systems*, 2010, Vol. 6, No. 4.

[4] L. Kennedy, M. Naaman, S.Ahern, R. Nair, and T. Rattenbury, “How Flickr Helps Us Make Sense of the World: Context and Content in Community-Contributed Media Collections”, In *Proc. of the Conf. on Multimedia*, ACM, New York, NY, 2007, pp. 631-640.

[5] S. Kisilevich, D. Kein, and L. Rokach, “ A Novel Approach to Mining Travel Sequences Using Collections of Geotagged Photos”, In *Proceeding of the 13th AGILE Int. Conf. on Geographic Information Science*, 2010, pp. 163 - 182.

[6] T. Kurashima, T. Iwata, G. Irie, and K. Fujimura “Travel Route Recommendation Using Geotags in Photos in Photo Sharing Sites”, In *Proc. of the Int. Conf. on Information and Knowledge Management*, 2010, pp. 579-588.

[7] I. Lee, G. Cai, and K. Lee, “Mining Points-of-Interest Association Rules from Geo-tagged Photos”, In *Proc. of the 46th HICSS*, 2013, pp. 1580 - 1588.

[8] I. Lee, G. Cai, and K. Lee, “Points-of-Interest Mining from People’s Photo-Taking Behavior”, In *Proc. of the 46th HICSS*, 2013, pp. 3129 - 3136.

[9] X. Lu, C. Wang, J.-M. Yang, Y. Pang, and L. Zhang “Photo2Trip: Generating Travel Routes from Geo-tagged Photos for Trip Planning”, In *Proc. of the Int. Conf. on Multimedia*, 2010, pp. 143-152.

[10] K. Okuyama, and K. Yanai “A Travel Planning System Based on Travel Trajectories Extracted from a Large Number of Geotagged Photos on the Web”, In *Proc. of the Pacific-Rim Conf. on Multimedia*, Sydney, Australia, 2011, pp. 531-540.

[11] T. Rattenbury, N. Good, and M. Naaman, “Towards Automatic Extraction of Event and Place Semantics from Flickr Tags”, In *SIGIR*, 2007, pp. 103-110.

[12] T. Rattenbury, N. Good, and M. Naaman, “Towards Extracting Flickr Tag Semantics”, In *Proc. of the 16th Int. Conf. on WWW*, 2007, pp. 1287-1288.

[13] P.Serdyukov, V. Murdock, and R. van Zwol, “Placing Flickr Photos on a Map”, In *SIGIR*, 2009, pp. 484 - 491.

[14] Y. Shi, P. Serdyukov, A. Hanjalic, and M. Larson, “Personalized Landmark Recommendation Based on Geotags from Photo Sharing Sites”, In *AAAI Conf. on Weblogs and Social Media*, 2011, pp. 622-625.

[15] Z. Yin, L. Cao, J. Han, J. Luo, and T. Huang, “Diversified Trajectory Pattern Ranking in Geo-Tagged Social Media”, In *Proc. of the SIAM Conf. on Data Mining*, pages 334 - 342, 2011.

[16] Y-T. Zheng, Z-J. Zha, and T-S. Chua, “Mining Travel Patterns from Geotagged Photos”, *ACM Transactions on Intelligent Systems and Technology*, 2012, Vol. 3 (3): 56.

[17] Y. Zheng, L. Zhang, X. Xie, and W-Y. Ma, “Mining Interesting Locations and Travel Sequences from GPS Trajectories”, *WWW*, 2009, pp. 791 - 800.