Adaptive City Characteristics: How Location Familiarity Changes What Is Regionally Descriptive

Vikas Kumar
University of Minnesota
Minneapolis, Minnesota, USA
vikas@cs.umn.edu

Saeideh Bakhshi
Facebook Inc.
Menlo Park, California, USA
saeideh@gatech.edu

Lyndon Kennedy
Futurewei Technologies, Inc.
Santa Clara, California, USA
lyndonk@acm.org

David A. Shamma
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
aymans@acm.org

ABSTRACT

Proliferation of GPS-enabled mobile devices has brought a plurality of location-aware applications leveraging the location characteristics in the shared content, like photos and check-ins. While these applications provide contextual and relevant information, they also assume geo-tagged contents to be representative of the geo-bounded characteristics of location. In this paper, however, we show that the characteristics geo-tagged contents capture about a location can vary based on the familiarity of user (sharing the content) with the location. Using a large dataset of geo-tagged photos, we learn descriptive spatial photo characteristics and user temporal-location-familiarity to highlight unique characteristics photos capture of location, which vary significantly if taken by locals versus tourists. We then propose a ranking-approach to find most representative photos for a given city. A user-based evaluation shows photos are more diverse and characteristic of location compared to other popular baselines while being representative of how locals and tourists would describe the city.

CCS CONCEPTS

• **Information systems** → **Geographic information systems**; **Multimedia and multimodal retrieval**;

KEYWORDS

image content; retrieval; location-familiarity; tourists; locals

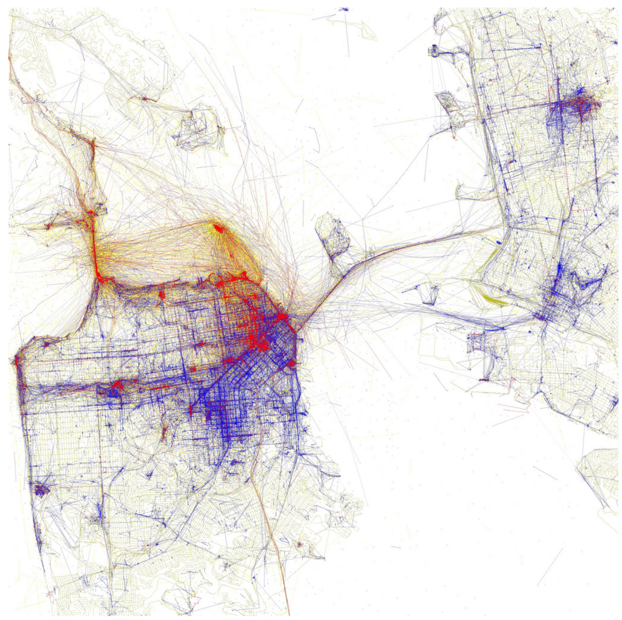


Figure 1: Distinct group of users in San Francisco. Spots of photos taken by locals are in blue and tourists spots are in red. Yellow trails might be either. Image ⓒⓘⓢ Eric Fischer on Flickr: https://flic.kr/p/87P5qP.

1 INTRODUCTION

Location plays a critical role in personalizing and identifying relevant content for users in online services. Mobile services such as Google Now, Yelp, and Foursquare have shown to elicit user experiences [24, 34] by leveraging the location. Other services, including online games [31] or shopping [18], have shown to provide better assistance while adapting to location. In the past decade, such location based services have grown further through geo-tagging and place-tagging. Users find it more easy now than ever to share contents tagged with location from their GPS-enabled mobile devices. This exponential growth of geo-tagged contents such as photos, check-ins etc provide valuable opportunity to study users' interaction patterns and their unique perceptions of surroundings.

Among many geo-tagged contents, online photo sharing popularized by systems like Instagram, Flickr, and Facebook, has made them an ubiquitous choice for users to learn and explore about a location. While browsing within limited screen-space on mobile

The work was conducted while the authors were at Yahoo Labs.

UMAP'17, July 9–12, 2017, Bratislava, Slovakia

978-1-4503-4635-1/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3079628.3079665

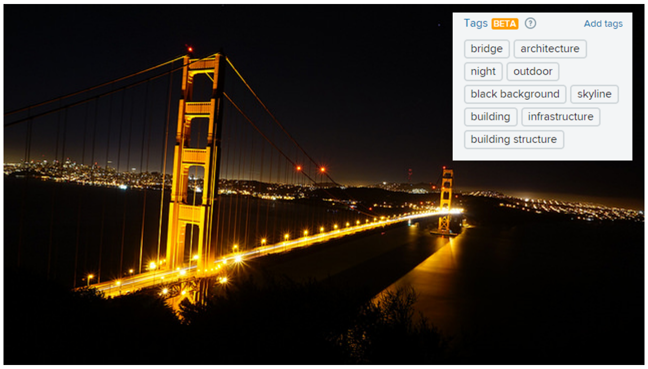


Figure 2: A sample photo of Golden Gate bridge in San Francisco and its *autotags* determined via computer vision.

devices, photos being more intuitive, illustrative, and far more succinct than words, allow users to peer into places that they would otherwise never see. They help choose travel destinations, influence decision making, planing, cognition as well as users' behavior at destinations [15]; as a result, it is critical for location-services to identify photos that capture diverse representation of location.

Geo-tagged photos have been studied extensively in the past decade [28]. They have been used in tag-aggregation research [3, 4, 19, 29], landmark detection [10, 20, 37], identifying representative icons or hero images for a city [5, 12]. In most of these applications, photos are judged by their social potential [1], content quality [5], or a combination of them [20] while assuming they capture *local* characteristics. However, in this paper, we question the very assumption of localness and provide a better understanding of how *well* photos capture the characteristics they represent of a location and if they vary based on who shares it. In other words, we analyze the effect location-familiarity of a user could have on photos they capture. For example, users who are less familiar with location, such as tourists, are likely to find popular landmarks or attractions to be more relevant to capture than locals. Similarly both locals and tourists are likely to be aware of popular landmarks in the city but some landmarks common among locals are likely to be unknown to most of the tourists. We show in this paper that geo-tagged contents result in different representations when users' familiarity with location is considered—challenging the assumption of *localness* in the geo-tagged content.

To learn the characteristics each photo captures about a location, we use computer-vision technique combined with information-theory method over a large-scale geo-tagged photos dataset publicly available from Flickr [32]. A set of descriptive spatial characteristics for each photos is determined, and the familiarity of the users sharing the photos are learned based on their temporal-spatial interactions. Two distinct categories of locals and tourists are identified, and results highlight how tourist with shorter engagement with location capture characteristics that differ significantly from those captured by locals. For example, in San Francisco, tourists tend to be interested in places of attraction such as monuments or bridges, whereas locals tend to find food or people more interesting to capture.

The characteristics learned from the model is then use to serve a common purpose similar to location-aware services, that is, retrieving representative photos of location. We devise a location-familiarity-based characteristic-score for each photo to re-rank photos of a city as seen by locals and tourists respectively. The re-ranked photos are then evaluated by human judges in an online survey where they assess the quality of photo sets compared to other popular baselines. The results from the survey (a) highlight the limitation of underlying assumptions of localness in traditional location-aware services, (b) emphasize the need to include users' familiarity with location, (c) show improved effectiveness from location-familiarity-based characteristic-score ranking for locals and tourists, and (d) demonstrate accuracy in identifying a diverse and accurate characteristics of locations while being representative of both locals and tourists respectively.

2 RELATED WORK

Online shared items with spatial footprints like geo-tagged photos, or check-ins, are shown to provide contextual meaningful content leading to more pleasant experiences in location-based services [24, 34, 35]. Such spatial limitations on items and users are shown to be be more efficient and accurate compared to traditional techniques [22, 23]; however, the location-based systems in their definition of *local-ness* of content overlook the affect of content owners' familiarity with location, that is, they consider the contents generated by locals and tourists to be equally representative of location.

While there exists several work in mining specific spatial characteristics from geo-tagged photos to learn about a location, such as, identification of landmarks [5], ranking characteristic photos for a landmark [20], characterizing preferences of international tourists [30], mining movement patterns [14], and identification of tourist hot spots [10]; there is rarely any focus on the distinction in these characteristics based on content owners' location-familiarity. It is important to note that if the purpose is to identify landmarks in a city then we already bias the analysis to places that are prevalent among tourists. In comparison, to identify distinct representations of a city to help users explore about unknown or novel destinations then systems are required to be adaptive to diverse crowds of the city i.e. locals and tourists. In a more recent work, Johnson et al. [17] in their analysis of localness of geo-tagged content suggest that only 70% of online shared contents represent what is local to the location and that this percentage decreases in more rural locations. In this work, we extend this result further by examining the content by learning descriptive characteristics to understand distinct representations users capture about the city based on their familiarity with the location.

The idea of distinct representations of a city is primarily motivated from existing literature in *geographical psychology*, also known as *psychogeography* [6]. It is a field that discusses the laws, methods, and inventive strategies for exploring an urban landscape. In its simplest form, it suggests that places in the same city can have different associations and effects on human emotions and behavior depending on their familiarity with location. For example, consumer psychology of tourism [21, 26] highlight differences in tourists' and locals' behavior depending on what a city has to offer. Tourists, in their short stay, are thus more likely to explore well known or famous attractions in the city while locals are likely to be interested in food, parks or other means of entertainments. For example, in San Francisco, tourists can be often seen at the ferry building, piers, or the Golden Gate bridge compared to locals who prefer to be at local restaurants, breweries, parks, stadiums etc. These differences are illustrated in maps of various cities in United States by Eric Fischer [8]. Figure 1 shows the areas of interest in San Francisco for locals and tourists by Fischer's metric; the map consists of points where a photo was taken with blue color representing a local user and red representing tourist.

Only a few recent studies measure the differences in users perception about a location [27, 36]. Using GPS traces they find specific routes of the city to be preferred by visitors, however, we expand the understanding of characteristics users capture in their photos with descriptive keyword for both locals and tourists. In this work, we thus aim to draw attention of existing location-based systems to reconsider the definition of localness of content while being representative to both locals and tourists respectively.

3 MODELING SPATIAL CHARACTERISTICS

Our first goal is to model descriptive spatial characteristics from geo-tagged photos. In this section, we start with an outline of the geo-tagged photos dataset available from Flickr. We then describe the computer vision technique along with information-theory metric we use to model the descriptive characteristics for each geo-tagged photo, to analyze the differences that users with different familiarity capture in the photos.

3.1 Dataset

We use the YFCC100M image dataset [32] consisting of 100 million publicly-available Creative Commons images from Flickr. The images have attributes such as the owner, acquisition timestamps, user-provided titles, descriptions, tags, and geo-tagging. For our analysis, we consider only the subset of geo-tagged images taken in United States. Further, we retrieved a multitude of social metrics from the Flickr API, such as the number of favorites, number of views and number of shares for each photo. For analysis, we only consider photos with at least 10 views and 10 favorites. We use these thresholds to consider photos shown to have potential for social engagement [1]. The resulting dataset consists of approximately 4.5 million images.

3.2 Descriptive Characteristics

To find the characteristics photos capture about a given location we leverage the visual content of each online shared geo-tagged photo using computer vision techniques. We learn meaningful descriptive characteristics from the photos in the form of keywords using a deep convolutional neural network that learns discriminative image representations, using large-scale collections of training examples pre-trained on the ImageNet dataset [7] provided by the Caffe framework [16]. The output of the last fully-connected layer (fc7) delivers a 4096-dimensional feature representation of each image. Using a linear support vector machine [33], the images are then classified along 1700 different ImageNet concepts. We refer to these automatically-detected keywords for the visual content as "characteristics" captured in the photo. A sample of the descriptive characteristics derived from a photo are shown in Figure 2.

In order to learn the spatial property of these characteristics and find the ones that *uniquely* identify with a location (a city), we model the keywords and locations into an information theory metric known as *conditional entropy*. The metric, defined as $H(X|Y)$, measures the certainty of variable X (bits of information) given the knowledge about variable Y. Smaller metric value implies higher certainty about variable X. We model set of characteristics C and the set of locations L as the random variables X and Y respectively; and formally define $h(c|l)$, see equation (1). This measures the certainty a characteristic (c) carries given the location (l).

$$h(c|l) = p(c,l) \times \log \frac{p(l)}{p(c,l)} : c \in C, l \in L. \tag{1}$$

In equation (1), $p(c)$ is the ratio of the number of photos with visual descriptor c to the total number of photos; and, $p(l)$ is the ratio of number of photos taken at the location l to the total number of photos. Finally, $p(c,l)$ is the joint probability of characteristic c and location l. A smaller entropy (higher certainty) implies a higher chance that the characteristic c is unique or highly certain given the location l. Likewise, higher entropy (less certainty) implies that the characteristic c is less likely or less certain to be representative of the location. For instance, the characteristic "outdoor", one of the most common visual descriptor in photos fails to uniquely identify with any given location having higher conditional entropy value. Whereas a tag like "latte" uniquely identifies with city of Seattle having a smaller conditional entropy value.

Note on Location: We evaluate the metrics with location referring to distinct cities in United States. The photo's location in form of latitude and longitude is converted to corresponding city using geocoding APIs[1]. Focusing on cities, instead of specific landmark or tourist spots allows us to understand the differences more accurately between the locals and tourists. It also allows to serve broadly for search retrieval purposes close to user queries exploring destinations that often starts at city level.

3.3 Location Familiarity: Locals Versus Tourists

To determine if users based on their familiarity capture unique photos of a location, we summarize photos into two distinct sets. The first set consists of all photos taken by locals (more familiar of the location) and the second set are photos taken by tourists (less familiar of the location).

The classification of users into locals or tourists is determined based on their temporal-spatial interaction patterns. The timestamps of photos taken in succession by user was leveraged to identify their temporal associations with each location. For a given location, we visualize the distribution of differences in timestamp of first and last photos taken by the user and segregate users into two sets with activity periods either to be (1) under 30 days, or (2) more than 30 days. Users with shorter activity at a location (city in our case) suggesting temporary presence are classified as *tourists*, and the latter group of users, suggesting longer presence, as *locals*. The choice of 30 days is also found to be consistent with the prior definitions used to identify locals [8, 11].

An individual is likely to be recognized as local in more than one location in the above classification. These locations can be the places where a user may have spent time during her childhood, or

[1] https://developer.yahoo.com/maps/rest/V1/geocode.html

Table 1: Top 5 tags unique for Seattle and San Francisco sorted by increased value of conditional entropy.

San Francisco		Seattle	
Locals	*Tourists*	*Locals*	*Tourists*
texture	urban	ocean	latte
graffiti	architecture	sunset	urban
people	skyline	sidewalk	architecture
monochrome	ferrybuilding	urban decay	skyline
portrait	bridge	biking	outdoor

college, or currently as a resident. Likewise, the same individual is likely to be a tourist in more than one location, where she may have shorter periods of interactions. A user do not identify as both a local and a tourist for the same location. Furthermore, we find percentage of locals and tourists are different for different cities. For instance, in San Francisco, the percentage of local users is 68%, while tourists constitute the remaining 32%. And, the percentage of local users decreases from larger urban cities to more rural cities with moderate population. In smaller cities, most users are identified as tourists due to lack of active periods of photography [17].

In this paper, we focus only on urban cities where we find fair representation of both locals and tourists from the dataset. For each location we segregate the photos taken by users into local and tourist sets. Using the conditional entropy metric, unique characteristics for each location is determined for both locals' and tourists' photos respectively. In Table 1, the top 5 descriptors for Seattle and San Francisco are shown with increasing metric value of conditional entropy (or decreasing certainty). The top descriptors in Seattle for locals are "ocean", "sunset" and "sidewalk" compared to those of tourists' "latte", "urban" and "architecture". A visual inspection of the photos from Seattle taken by locals are shown to include sunset and sunrise photos. We believe that this is due to time of a day when residents are involved in casual walk or jog along the lakeshores and parks. Whereas tourists' photos include shots of Pike Place Market, Starbuck's first cafe, the urban-architecture of Space Needle and Seattle skyline (popular tourist destination). Likewise, for San Francisco, photos taken by locals are found to consist of pictures of local pride parades, or the graffiti in Clarion Alley (a popular area known for street art), while photos taken by tourists, similar to those in Seattle, include urban settings and architecture influenced by the skyline, famous Golden Gate Bridge, and Bay Bridge—implying a clear evidence of varying characteristics that users capture in their photos based on how they associate with location.

3.4 A Comparison

There exist many approaches to learn unique spatial characteristics from geo-tagged photos. Kennedy et al. [20] used clustering to identify landmarks within a city, then TF-IDF to determine representative photos for the given landmark. Similarly, Chen et al. [5] in their work use both TF-IDF as well as conditional entropy in part to identify iconic places and a representative image icon from geo-tagged photos within the city. Other approaches include characterizing preferences for tourists only [9, 30, 37], or discovering landmarks [13, 14, 25].

However, we emphasize and invite future work to reconsider the assumption of localness of the geo-tagged content. A proposition that a popular landmark in a city could indeed be well-known to both locals and tourists but there exist certain landmarks and characteristics that are likely only known to the true locals of the city. Our approach accounting for the location-familiarity of user in geo-tagged content show that inclusion of users' temporal patterns lead to different conclusions for the characteristics content represent about the location.

4 EVALUATION

Photos capture diverse representations of what cities are popularly known for and play an illustrative and intuitive role to allow other users to peer into different places within the city. They help users plan and choose their destinations[2] [15]. With such a critical role of photos in representation of city, we evaluate if the descriptive characteristics learned from photos are effective in identifying diverse representative photos of location. In this section, we describe a location-aware characteristics score to rank photos and explain evaluation setup including the assessment of quality of the photo sets compared to other popular baselines by human judges. We also evaluate how well these photos capture locals' and tourists' views of the city.

4.1 Location-Aware Characteristics Score

To identify and rank relevant photos, it is important to optimize for both content and perceived relevance of items [2]. Using the descriptive characteristics (c) derived from image content, and social engagement potential ($\#favorites$) as proxy for their relevance, we devise a characteristic-score charScore$_g$, for each geo-tagged photo (g) at location (l) as:

$$\text{charScore}_g = \frac{\log(\#\text{faves}_g)}{\sum\limits_{c \in \text{chars}(g)} \frac{\text{h}(c|l)}{\text{size}(\text{chars}(g))}} \tag{2}$$

The metric ranks photos with higher social engagement potential and unique characteristics of the location to the top. Higher social engagement, (more number of favorites[3]) implies higher relevance; and smaller the sum of conditional entropy implies higher the chances that candidate photo captures the representative characteristics of location. The score is calculated for each geo-tagged photo in the locals' and tourists' set respectively. Ranking photos in decreasing order of the score provide us the relevant, representative photos for the location, while unique to locals and tourists respectively.

[2]We find this to be evident in YFCC100M dataset too; The distribution of timestamps of user likes/favorites for photos at a location, and the distribution of timestamps of their uploads from the same location are compared. We find the first distribution to have statistically smaller ($p < 0.001$) mean than the later.

[3]We also investigated other social metric signals to rank photos such as number of views, number of shares etc. in characteristic score. However, in preliminary analysis by a group of 8–10 users, photos are found to be very similar to recommendations based on number of favorites. We choose to keep the social metric that is more available and intuitive across other domains.

Table 2: Age distribution of Survey Participants. Majority of our participants are in the age group of 30–40

Age Group	Total	Male	Own DSLR
0–20	8	62.5%	62.5%
20–30	97	70.0%	37.1%
30–40	106	67.0%	51.9%
40–50	30	60.0%	46.7%
50–60	20	50.0%	45.0%
Above 60	10	60.0%	40.0%

Table 3: Most survey participants are male with a higher percentage of participants sharing photos on Facebook. Flickr users tend to be more likely to own professional level cameras, like DSLRs.

Participants	Percent	Female	Own DSLR
Male	64.2%	—	51.0%
Female	35.8%	—	56.5%
Shares on Facebook	68.6%	38.7%	42.4%
Shares on Flickr	34.3%	29.1%	63.4%
Shares on Google Photos	17.7%	25.0%	41.7%

Table 4: A break down of the 271 survey participants.

City	Total	Local/Resident	Own DSLR
Boston	32	62.5%	34.3%
Los Angeles	39	51.2%	38.4%
Seattle	27	48.1%	25.9%
San Francisco	75	48.0%	61.3%
New York	47	42.6%	46.8%
Chicago	51	38.4%	41.1%

4.2 Survey Setup

To gather insights on quality of scoring the candidate photos and understand how well photos capture the diverse representation of a city, we design an online survey for a large-scale human-based evaluation. The participants are asked to choose a city from set of choices and assess four different unique set of photos for each city. The four sets are described below:

localR: Candidate set based only on photos taken by users identified as locals and sorted by the characteristic score.

touristR: Candidate set based only on photos taken by users identified as tourists and sorted by characteristic score.

geo-popularR: A baseline set that contains the most socially relevant photos of location, sorted by number of favorites.

popularR: A naive baseline set that contains photos irrespective of location, sorted by number of favorites.

We use top 20 photos in each set for evaluation. In our preliminary analysis, photos at top of the lists are found to be taken by expert users often with specific content like wedding, landscape, or portraits. The fan-following of these experts further resulted in unusual higher social engagement bias for these photos bringing in lack of *diversity* in the sets [38]. Although the quality and aesthetics of these photos are un-questionable, the high correlation in their main theme made photos being more repetitive in candidate set. To address this issue, we, (a) filter out photos with similar characteristics in the set (determined using similarity between their characteristics-based feature vector), and (b) limit the number of photos to 2 per photographer in a given set. This helps achieve a diverse candidate photos set (see Figure 3(a) for an example set).

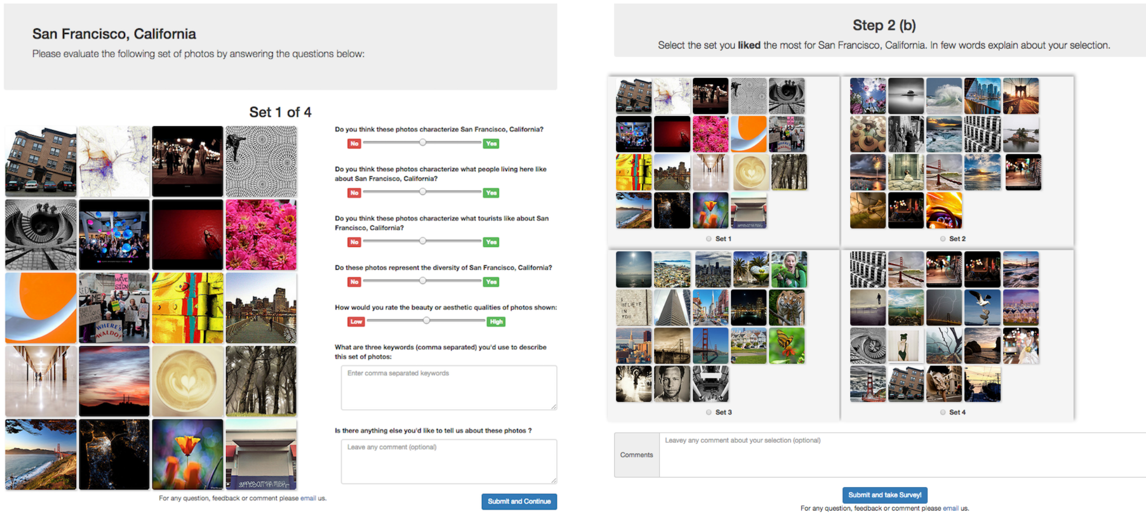
In the survey, participants answer a set of specific questions on a slider scale for each of four sets to assess the quality of retrieved photos. These questions primarily evaluate two qualities: (1) their diverse and characteristic representation of the city, and (2) whether they characterize what locals or tourists would think of the city. Correspondingly, as shown in Figure 3(b), each participant chooses a location (all within United States) and record her feedback for (a) "How characteristic are these photos of the location?", (b) "Rate the beauty or aesthetic quality of photos", and (c) "Do these photos represent the diversity of location?". Finally, whether photos are reflective of locals or tourists they answer: (d) "Do you think these photos characterize what locals think about the location?", and (e) "Do you think these photos characteristic what tourists think about the location?". As a final step, participants are asked to pick a set they think *best* represents the location. We keep the order of candidate photo sets randomized for every user, and a response is considered complete only when assessment for all four sets are recorded. In addition to quality assessment, we ask participants to provide their familiarity with location as a local or as a tourist; and, keywords that they think characterize the city in their own view.

5 RESULTS AND DISCUSSION

To participate in the survey, participants were invited via Amazon Mechanical Turks (limited to United States) and social media platforms such as Twitter, Facebook, and Google+. A total number of 271 participants completed the survey. Among those, 154 participants self-identify themselves as local to the cities they evaluate while 157 self-identify as tourists. The selection of cities were limited to six of the major locations in United States with balanced representation of both locals and tourists. Table 2 shows the age distribution of participants; and Table 3 shows gender, and platform these participants often use for photo sharing. Table 4 shows percentage of participants who self-identify as locals for each of six cities. There are no significant differences in our findings based on gender or age of participants.

5.1 User Keywords Versus Descriptive Characteristics

We evaluate the keywords users provided in survey describing the city in their own view. Locals are found to be more diverse in their description of cities compared to tourists. The 154 locals provided 217 unique keywords compared to only 159 keywords from the 157 tourists; that is 58% statistically more keywords per user. We also determine the most frequent occurring keywords (pre-processed with removal of stopwords and Porter stemming) for each city. The 20 most frequent keywords used by locals and tourists for San Francisco are illustrated in Figure 4. The high frequency of "food", "ocean" and "culture" among locals' descriptions compared to tourists' "architecture", "bridge" and "street" for San Francisco



(a) Participant records their responses assessing each recommendation set on a slider scale.

(b) In final step of the survey, participants are asked to choose the set of photos that they prefer the most for given location.

Figure 3: Evaluation of Recommendation Sets by Participants

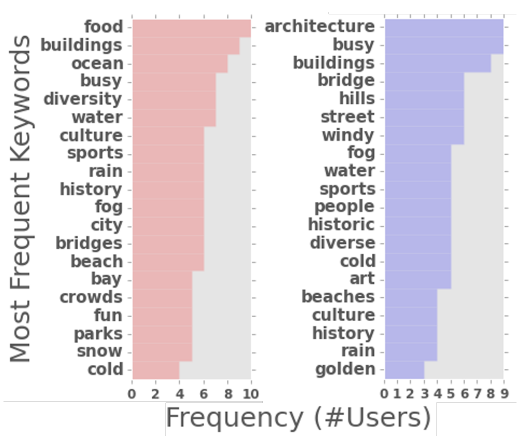


Figure 4: Most frequent keywords to describe a city by residents (left) and visitors (right).

emphasize the different ways users characterize the same city based on their familiarity—a result coherent to the descriptive spatial characteristics we learn about city from the geo-tagged photos. Similar distinctions are observed among other cities evaluated in the survey[4].

5.2 Photos Assessment

In this section, we evaluate the effectiveness of the characteristic score in ranking the diverse representative photos of a location. From the survey responses, participants are found to overall *prefer the candidate set based on tourists' photos*, with 52% selecting *touristR*, 23% *geo-popularR*, followed by 20% and 5% for *localR* and *popularR* respectively. The difference in percentages are found to be statistically significant from each other ($p < 0.01$) using a Chi-square proportion test. Nevertheless, we observe these percentages to be slightly different for each city. However, the order is same. The inter-rater agreement statistic or Cronbach's alpha is also found to be significant with value of 0.84.

Participants' response for each individual set is further examined to better understand their assessment for quality, diversity and how well they think photos capture the relevant meaningful characteristics of location. In Figure 5, the cumulative percentage of participants are shown for different scales of agreements (the continuous values of slider scale are converted to respective five segments similar to Likert scales). The percentages on the right reflect the percentage of participants who "Strongly Agree" or "Agree"; in the middle are the percentage of participants with *neutral* response; and in left are percentage of users who "Strongly Disagree" or "Disagree". The color shades represent each of the five Likert scales.

Result shows 86% of **participants strongly agree *touristR* being more characteristic of location** compared to only 70% agreement for *geo-popularR* and 65% for *localR*. The scores on characteristics, aesthetics, as well as diversity of photos from *touristR* are found to be higher and statistically significant to other sets using Kruskal-Wallis non-parametric test ($p < 0.001$). Furthermore, participants find *localR* photos to be more characteristic of what locals think about location (4th plot from top in Figure 5: 75% vs others) and *touristR* photos to be more characteristic of what tourists think about location (bottom of Figure 5: 80% versus others).

5.2.1 Discussion. The overall higher preference for tourists set highlights the importance tourists play for these locations. In their short stay, these users are likely to capture destinations that are well-known and often better representative of what is popular within the city — an implicit design implication for location-based services. Services could become more accurate and efficient if only tourists

[4]We do not discuss each city due to page limitations.

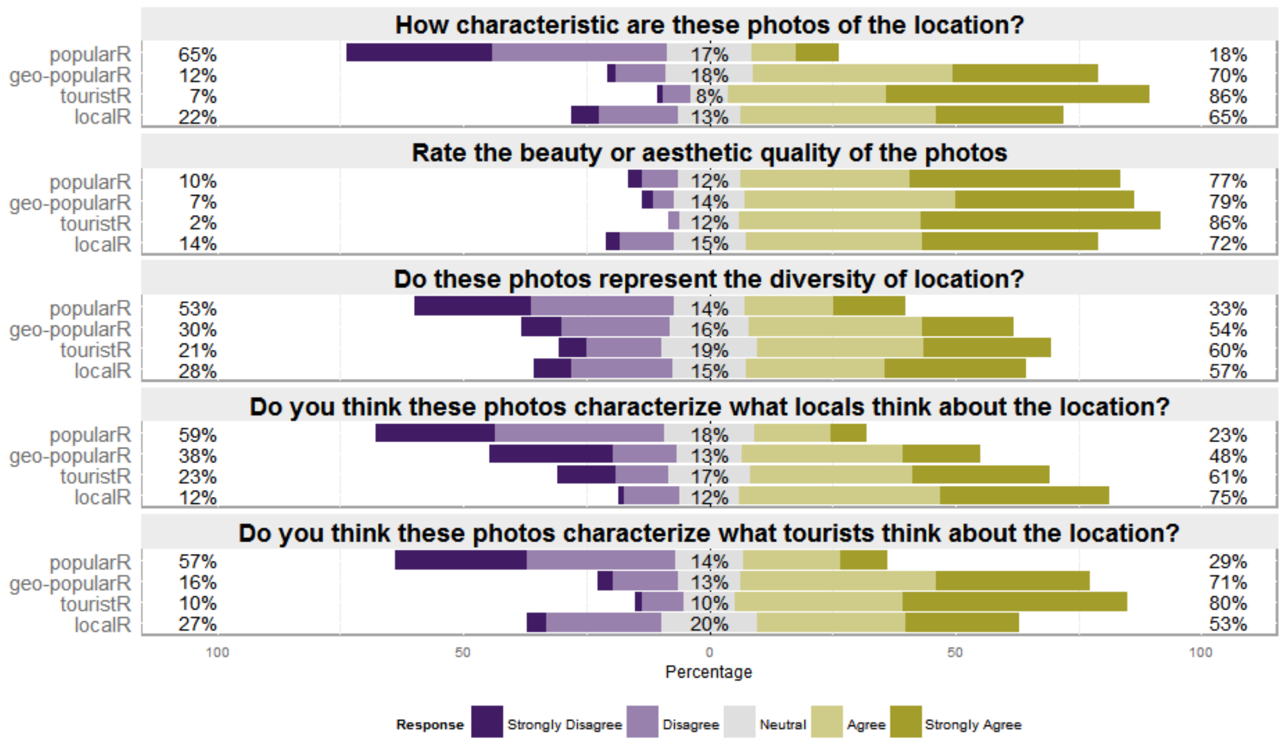


Figure 5: Cumulative Survey responses from recommendation assessment. Users unanimously agreed that *touristR* photos capture the best characteristics of location. However, users in the survey are shown to recognize the sets that characterize what locals and tourists think about location.

generated content is analyzed for content retrieval or ranking purposes instead of mining the whole dataset. One such scenario is seen in online systems when no pre-existing information available about the requesting user for any personalization, a phenomena also known as coldstart. In such scenario, we hypothesize that ranking items based on other similar users who in past were tourists at the location could be more effective. To understand this implication better, we investigate in next section if participants who identify as tourists have any significant difference in their opinion to the photo sets compare to those who identify as locals.

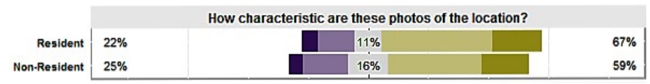
We draw two more conclusions from the results above: (1) the photos selected based on their visual content are able to capture representations of city that is more recognizable to users than the popular baselines; and, (2) their ability to distinguish the representations of city that matches with locals' and tourists' views respectively. These findings highlight the efficacy we achieve by adapting to the difference in preferences that exist between locals and tourists for a location.

5.3 Location Familiarity

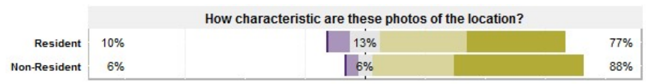
We now analyze if the familiarity of participants with location affect their assessment of the photos. Among 154 participants, who self-identify as locals to the location, there is still larger percentage of users (44.8%) who prefer the *touristR*. Only 31.42% of these users prefer the local set i.e. *localR*. However, we notice a significant difference in percentage of participants, who self-identify as tourists, in their preference for *localR* and *touristR*. There are only 9.7% of these participants who prefer the local set (*localR*) compared to 31.42% of the participants identified as locals. A major percentage of tourist participants prefer the *touristR* i.e. 60.3%. This difference in agreement between local and tourist participants underlines the **difference in perception among these users**. The awareness/familiarity of the participants who are locals are likely to recognize the characteristics in photos only known to locals than to the participants who are tourists. For example, trails, distinct coffee shops, unique restaurants, parks etc. We find the evidence of these differences in the feedback locals shared about the *localR* set:

> *So far, this set seems most "boston". Shows boston scenes (or scenes that could be from boston) around the year and in a variety of light and weather* — Resident, Boston.
>
> *These feel more authentic and have a better NYC feel. Too many coffee cups, but we do love our coffee in NY. Definitely feels like a local took these* — Resident, New York.
>
> *At least a few in this set that are clearly Los Angeles (e.g. union station) and blondes (hey, its LA), but still can't figure out how they relate at all to each other* — Resident, Los Angeles.

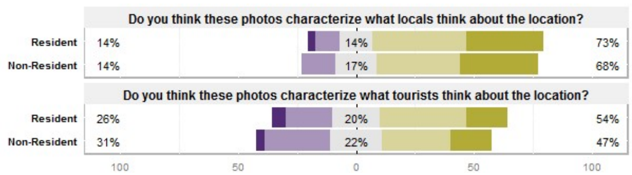
Similar differences are observed in individual assessment of each candidate photo sets. As shown in Figure 6(a) and 6(b), higher percentage of participants who are locals (67%) find *localR* to be



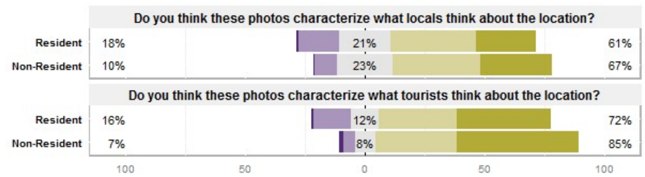
(a) *localR*



(b) *touristR*



(c) *localR*



(d) *touristR*

Figure 6: Difference in opinion of Locals (marked as Resident) Vs Tourists (marked as Non-Resident): (a) Locals find *localR* more characteristic/representative of location ($p < 0.05$), (b) Tourists find *touristR* more characteristic ($p < 0.01$), (c) Both locals and tourists find *localR* to be more representative of locals, and (d) Both locals and tourists find *touristR* to be more representative of tourists.

more characteristic of location than participants who are tourists (59%); while, higher percentage of tourists (88%) find *touristR* to be more characteristic of location than participants who are locals (77%)[5]. Moreover, the local participants find the local set (*localR*) to represent more of "what locals think about the location" than the tourist set (*touristR*)—73% vs 61%, while no difference observed in tourist participants response to the same question. However, both local and tourist participants agreed upon *touristR* to be highly characteristic of "what tourists would think about the location".

5.3.1 Discussion. In summary, participants reflect difference in their opinion for candidate set based on how familiar they are with location. The participants who are more familiar with location identify photos in local set more easily and find it delightful compared to tourists. On the other hand, participants who self-identify as tourists find tourist set highly representative of location compared to other baselines. This result highlight the accuracy of our metric in retrieving photos that uniquely representative of location while being adaptive to views of locals and tourists respectively. Result further underlines the importance of being adaptive to location-familiarity with geo-tagging as well as provide an alternative for problems such as cold-start in online location-based systems.

6 CONCLUSION AND FUTURE WORK

While location-based services in mobile social web provide multitudes of possibility to produce and consume information they also create an unique opportunity to better understand user interactions and their perceptions of surroundings. To our knowledge, this is the first work to not only emphasize that perceptions differ for individuals at a location but also be able to describe these perceptions with descriptive spatial characteristics. We demonstrate that geo-tagged content can vary based on how group of users perceive given their familiarity with location. We then devise a location-familiarity-aware characteristic-score that improves the effectiveness in retrieval of representative photos of location. Assessed by human judges, the photos are shown to be diverse, representative, and adaptive to characteristics familiar to locals and tourists respectively; an important result challenging the underlying assumption of *localness* in content for location-based services.

[5]The differences found to be significant to each other using the non-parametric Wilcoxon rank sum test ($p < 0.05$)

Nevertheless, there are some limitations to our approach that we would like to address in our future work. First, even though our approach is able to distinguish characteristics for large number of cities in the dataset, we only be able to discuss six cities for our evaluation. This is due to known limitations of time and effort involved in online human-based evaluations (a better approach would be to conduct an A/B test within large scale systems). The other limitation is that we did not consider context of users in our implementations such as time, year, or even season. Since a user could visit a city for different reasons and at different times of the year we believe contextual information could play more important role. For example, a tourist traveling for leisure compared to traveling for business may seek to explore different destinations in the city. Similarly, a tourist behavior could vary based on season of the year like snowy winters versus hot summers. We believe that adapting to such contexts could further help recognize and understand contextual characteristics captured in the content.

7 ACKNOWLEDGMENT

We thank Yahoo Flickr team for their support and feedback during this work, and our reviewers for their valuable feedback and suggestions on multiple aspects of this paper.

REFERENCES

[1] Saeideh Bakhshi, David A Shamma, Lyndon Kennedy, and Eric Gilbert. 2015. Why We Filter Our Photos and How It Impacts Engagement. In *Ninth International AAAI Conference on Web and Social Media.*
[2] Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12, 4 (2002), 331–370.
[3] Iván Cantador, Ioannis Konstas, and Joemon M Jose. 2011. Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web* 9, 1 (2011), 1–15.
[4] Iván Cantador, Martin Szomszor, Harith Alani, Miriam Fernández, and Pablo Castells. 2008. Enriching ontological user profiles with tagging history for multi-domain recommendations. (2008).
[5] Wei-Chao Chen, Agathe Battestini, Natasha Gelfand, and Vidya Setlur. 2009. Visual Summaries of Popular Landmarks from Community Photo Collections. In *Proceedings of the 17th ACM International Conference on Multimedia (MM*

'09). ACM, New York, NY, USA, 789–792. DOI: http://dx.doi.org/10.1145/1631272.1631415

[6] Guy Debord. 1955. Introduction to a critique of urban geography. *Critical Geographies A Collection of Readings* (1955).

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.

[8] Eric Fischer. 2014. Locals and tourists. *A+ U-ARCHITECTURE AND URBANISM* 530 (2014), 30–33.

[9] Juan Carlos García-Palomares, Javier Gutiérrez, and Carmen Mínguez. 2015. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and {GIS}. *Applied Geography* 63 (2015), 408 – 417. DOI: http://dx.doi.org/10.1016/j.apgeog.2015.08.002

[10] Juan Carlos GarcÃŋa-Palomares, Javier GutiÃĺrrez, and Carmen MÃŋnguez. 2015. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and {GIS}. *Applied Geography* 63 (2015), 408 – 417. DOI: http://dx.doi.org/10.1016/j.apgeog.2015.08.002

[11] Fabien Girardin, Francesco Calabrese, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. 2008. Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE* 7, 4 (2008), 36–43.

[12] Fabien Girardin, Andrea Vaccari, Re Gerber, and Assaf Biderman. 2009. Quantifying urban attractiveness from the distribution and density of digital footprints. *Journal of Spatial Data Infrastructure Research* (2009).

[13] Yingjie Hu, Song Gao, Krzysztof Janowicz, Bailang Yu, Wenwen Li, and Sathya Prasad. 2015. Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems* 54 (2015), 240 – 254. DOI: http://dx.doi.org/10.1016/j.compenvurbsys.2015.09.001

[14] Piotr Jankowski, Natalia Andrienko, Gennady Andrienko, and Slava Kisilevich. 2010. Discovering Landmark Preferences and Movement Patterns from Photo Postings. *Transactions in GIS* 14, 6 (2010), 833–852. DOI: http://dx.doi.org/10.1111/j.1467-9671.2010.01235.x

[15] Olivia H Jenkins. 1999. Understanding and measuring tourist destination images. *The International Journal of Tourism Research* 1, 1 (1999), 1.

[16] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093* (2014).

[17] Isaac L Johnson, Subhasree Sengupta, Johannes Schöning, and Brent Hecht. 2016. The Geography and Importance of Localness in Geotagged Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 515–526.

[18] Gerrit Kahl, Lübomira Spassova, Johannes Schöning, Sven Gehring, and Antonio Krüger. 2011. IRL SmartCart-a user-adaptive context-aware interface for shopping assistance. In *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 359–362.

[19] Marius Kaminskas, Francesco Ricci, and Markus Schedl. 2013. Location-aware music recommendation using auto-tagging and hybrid matching. In *Proceedings of the 7th ACM conference on Recommender systems*. ACM, 17–24.

[20] Lyndon S. Kennedy and Mor Naaman. 2008. Generating Diverse and Representative Image Search Results for Landmarks. In *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*. ACM, New York, NY, USA, 297–306. DOI: http://dx.doi.org/10.1145/1367497.1367539

[21] Metin Kozak, Enrique Bigné, Ana González, and Luisa Andreu. 2004. Cross-cultural behaviour research in tourism: a case study on destination image. *Consumer psychology of tourism, hospitality and leisure* 3 (2004), 303–317.

[22] Vikas Kumar, Daniel Jarratt, Rahul Anand, Joseph A. Konstan, and Brent Hecht. 2015. "Where Far Can Be Close": Finding Distant Neighbors In Recommender Systems. In *Proceedings of LocalRec Workshop in ACM conference on Recommender Systems*. 13–20. http://ceur-ws.org/Vol-1405/paper-03.pdf

[23] Justin J Levandoski, Mohamed Sarwat, Ahmed Eldawy, and Mohamed F Mokbel. 2012. Lars: A location-aware recommender system. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. IEEE, 450–461.

[24] Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. 2013. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1043–1051.

[25] Silvia Paldino, DÃąniel Kondor, Iva Bojic, Stanislav Sobolevsky, Marta C. GonzÃąlez, and Carlo Ratti. 2016. Uncovering Urban Temporal Patterns from Geo-Tagged Photography. *PLOS ONE* 11, 12 (12 2016), 1–14. DOI: http://dx.doi.org/10.1371/journal.pone.0165753

[26] RR Perdue, HJP Immermans, and M Uysal. 2004. *Consumer psychology of tourism, hospitality and leisure*. Vol. 3. CABI.

[27] Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. 2014. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 116–125.

[28] Gillian Rose. 2008. Using photographs as illustrations in human geography. *Journal of Geography in Higher Education* 32, 1 (2008), 151–160.

[29] Pavel Serdyukov, Vanessa Murdock, and Roelof Van Zwol. 2009. Placing flickr photos on a map. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 484–491.

[30] Shiliang Su, Chen Wan, Yixuan Hu, and Zhongliang Cai. 2016. Characterizing geographical preferences of international tourists and the local influential factors in China using geo-tagged photos on social media. *Applied Geography* 73 (2016), 26 – 37. DOI: http://dx.doi.org/10.1016/j.apgeog.2016.06.001

[31] Ning Tan, Gaëtan Pruvost, Matthieu Courgeon, Céline Clavel, Yacine Bellik, and Jean-Claude Martin. 2011. A location-aware virtual character in a smart room: effects on performance, presence and adaptivity. In *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 399–402.

[32] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. YFCC100M: The New Data in Multimedia Research. *Commun. ACM* 59, 2 (Jan. 2016), 64–73. DOI: http://dx.doi.org/10.1145/2812802

[33] Vladimir Vapnik. 1998. *Statistical Learning Theory*. Wiley, New York.

[34] Mao Ye, Peifeng Yin, and Wang-Chien Lee. 2010. Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 458–461.

[35] Hongzhi Yin, Yizhou Sun, Bin Cui, Zhiting Hu, and Ling Chen. 2013. Lcars: a location-content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 221–229.

[36] Vincent W Zheng, Yu Zheng, Xing Xie, and Qiang Yang. 2010. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th international conference on World wide web*. ACM, 1029–1038.

[37] Yan-Tao Zheng, Zheng-Jun Zha, and Tat-Seng Chua. 2012. Mining Travel Patterns from Geotagged Photos. *ACM Trans. Intell. Syst. Technol.* 3, 3, Article 56 (May 2012), 18 pages. DOI: http://dx.doi.org/10.1145/2168752.2168770

[38] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving Recommendation Lists Through Topic Diversification. In *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*. ACM, New York, NY, USA, 22–32. DOI: http://dx.doi.org/10.1145/1060745.1060754