Computational Analysis of Urban Places Using Mobile Crowdsensing

THÈSE N^{O} 7243 (2016)

PRÉSENTÉE LE 17 NOVEMBRE 2016

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE L'IDIAP
PROGRAMME DOCTORAL EN GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Darshan SANTANI

acceptée sur proposition du jury:

Prof. P. Frossard, président du jury
Prof. D. Gatica-Perez, directeur de thèse
Prof. K. Van Laerhoven, rapporteur
Dr A. Monroy-Hernandez, rapporteur
Prof. S. Susstrunk, rapporteuse

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

To Pitajee, Mata and Biwi

Acknowledgements

The long winding PhD journey began on a cold January day in 2012, when I first met Daniel Gatica-Perez for a chat at EPFL to explore the possibility of pursuing doctoral studies in his research group. Since then, I have always felt at home in Daniel's company. Throughout these years, Daniel has been supportive, available, patient, and honest as a mentor. I express my sincere gratitude to Daniel for sharing his research passion, giving me the space and freedom to express myself as a researcher, and guiding me through the intricacies of academic life. I thank Daniel for not just being an ideal research mentor but also personally as a friend.

But my research journey began long before I moved to EPFL. It began when C. Jason Woodard hired me as a research engineer at the School of Information Systems at SMU (Singapore) in 2006. Jason taught me to be independent, thorough, and persistent while walking me through the ropes of academic research. I am grateful to Jason for allowing me to dedicate part of my work time to pursue other research directions which eventually led to my doctoral research. Special thanks to Rajesh Krishna Balan at SMU for his guidance and advice, especially when I was applying for graduate school.

I would like to thank everyone in the social computing group (in no specific order) – Laurent, Minh-Tri, Joan, Oya, Dayra, Dinesh, Rui, Gulcan, Skanda, and Trung – for making the journey the one to remember. Special thanks to Minh-Tri, Joan, and Laurent for productive research discussions and brainstorming sessions. A shout-out to Laurent and Joan for being close friends, and a source of advice both at the professional and personal level.

Thanks to all the Idiapers, particularly Nikos, Thomas, Majid, Matthew, and David for good discussions over lunch and coffee breaks. I recognize and appreciate the (often latent) efforts of the entire systems crew at Idiap, in particular Frank Formaz and Louis-Marie Plumel, to cater to my last-minute requests. Thanks also to Nadine Rousseau and Sylvie Millius for making our lives easier administratively and painstakingly translating all the official documents to English. I apologize for not naming each Idiaper individually, but needless to say, I had a productive time at Idiap.

During my doctoral studies, I had the opportunity to work on various projects and collaborate with people from diverse backgrounds. But of many joint works, one of my memorable collaboration was with the SenseCityVity team in Mexico led by Salvador Ruiz Correa (Sal). I have always enjoyed Sal's company and admire his energetic and optimistic outlook towards life. Thank you Sal for making my Mexico trip a memorable one. I take this opportunity to

thank all the members of my doctoral thesis committee – Pascal Frossard, Andrés Monroy-Hernández, Kristof Van Laerhoven, Sabine Süsstrunk – for reviewing my thesis and providing insightful comments.

It would be suffice to say that the last four years have been eventful. But I would not have survived without the love, affection and sacrifice of all my family members – Pitajee, Mata, Neha, Kishi, and Pooja. My loving parents gave their unconditional support, encouragement and always stood besides my choices, both personal and professional ones. Kishi, you rock ! A special personal shout goes to Neha who was always there with me through the highs and lows of the PhD journey. I can not thank Neha enough for her understanding, selfless care, and a source of constant inspiration. Keep smiling Biwi ! I do not know whether I deserve the love showered by each one of them but I was definitely given more than I had asked for. *Tahe dil se shukriya* !

Martigny, 21 October 2016 Darshan Santani

Abstract

In cities, urban places provide a socio-cultural habitat for people to counterbalance the daily grind of urban life, an environment away from home and work. Places provide an environment for people to communicate, share perspectives, and in the process form new social connections. Due to the active role of places to the social fabric of city life, it is important to understand how people perceive and experience places. One fundamental construct that relates place and experience is *ambiance*, i.e., the impressions we ubiquitously form when we go out. Young people are key actors of urban life, specially at night, and as such play an equal role in co-creating and appropriating the urban space. Understanding how places and their youth inhabitants interact at night is a relevant urban issue.

Until recently, our ability to assess the visual and perceptual qualities of urban spaces and to study the dynamics surrounding youth experiences in those spaces have been limited partly due to the lack of quantitative data. However, the growth of computational methods and tools including sensor-rich mobile devices, social multimedia platforms, and crowdsourcing tools have opened ways to measure urban perception at scale, and to deepen our understanding of nightlife as experienced by young people.

In this thesis, as a first contribution, we present the design, implementation and computational analysis of four mobile crowdsensing studies involving youth populations from various countries to understand and infer phenomena related to urban places and people. We gathered a variety of explicit and implicit crowdsourced data including mobile sensor data and logs, survey responses, and multimedia content (images and videos) from hundreds of crowdworkers and thousands of users of mobile social networks. Second, we showed how crowdsensed images can be used for the computational characterization and analysis of urban perception in indoor and outdoor places. For both place types, urban perception impressions were elicited for several physical and psychological constructs using online crowdsourcing. Using low-level and deep learning features extracted from images, we automatically inferred crowdsourced judgments of indoor ambiance with a maximum R^2 of 0.53 and outdoor perception with a maximum R^2 of 0.49. Third, we demonstrated the feasibility to collect rich contextual data to study the physical mobility, activities, ambiance context, and social patterns of youth nightlife behavior. Fourth, using supervised machine learning techniques, we automatically classified drinking behavior of young people in an urban, real nightlife setting. Using features extracted from mobile sensor data and application logs, we obtained an overall accuracy of 76.7%.

While this thesis contributes towards understanding urban perception and youth nightlife patterns in specific contexts, our research also contributes towards the computational understanding of urban places at scale with high spatial and temporal resolution, using a combination of mobile crowdsensing, social media, machine learning, multimedia analysis, and online crowdsourcing.

Keywords : Mobile Crowdsensing, Ambiance, Urban Perception, Indoor Places, Outdoor Places, Visual Perception, Youth, Nightlife, Alcohol, Ubiquitous Computing, Urban Computing, Social Computing, Social Media, Foursquare

Résumé

En ville, les lieux publics constituent un habitat socio-culturel qui permet à tout un chacun de contrebalancer la routine quotidienne de la vie urbaine ; ces lieux constituent ainsi un environnement à la fois hors du travail et de la maison. Ces lieux sont une opportunité pour les gens de communiquer, partager des idées, ou encore former de nouveaux liens sociaux. Étant donné l'importance de ces lieux pour le tissu social d'une ville, la compréhension de la manière selon laquelle les gens les perçoivent et en font l'expérience constitue un problème qui mérite d'être abordé. Un construit social fondamental qui relie les lieux publics à l'expérience de ceux-ci est la notion d'*ambiance*, c'est-à-dire les impressions que nous nous faisons de manière omniprésente lorsque nous sortons. Les jeunes adultes comptent parmi les acteurs-clé de la vie urbaine, particulièrement de nuit ; dans ce sens, ils jouent un rôle égal dans la co-création et l'appropriation de l'espace urbain. Comprendre les interactions entre les lieux publics et le comportement des jeunes la nuit consttitue ainsi un problème urbain pertinent.

Jusqu'à récemment, notre capacité à évaluer les qualités visuelles et perceptuelles des espaces urbains et à étudier les dynamiques entourant le comportement des jeunes dans ces endroits a été partiellement limitée par le manque de données quantitatives. Cependant, les avancées dans le domaine des méthodes computationnelles combinée au développement d'outils tels que les appareils mobiles incluant des capteurs, les plateformes sociales de contenu multimedia, ou les systèmes de crowdsourcing (travail collaboratif de masse) ont permis de mesurer la perception urbaine à grande échelle, ainsi que d'approfondir notre compréhension de la vie nocturne expérimentée par les jeunes adultes.

Dans cette thèse, en tant que première contribution, nous avons conçu, mis en œuvre et analysé quatre études de crowdsensing mobile incluant des populations de jeunes adultes de pays divers dans le but de comprendre et inférer plusieurs phénomènes liés aux lieux publics urbains et aux personnes. Nous avons collecté diverses données crowdsourcées de manières implicite et explicite, comprenant des logs et des données de capteurs provenant de smartphones, des réponses à des questionnaires, ainsi que du contenu multimedia (images et vidéos) de centaines de volontaires et de milliers d'utilisateurs de réseaux sociaux. Deuxièmement, nous avons montré comment des images crowdsourcées peuvent être utilisées pour la caractérisation automatique et l'analyse de perception urbaine dans les lieux intérieurs et extérieurs. Pour chaque type d'endroits, les impressions de perception urbaine ont été collectées sur la base de plusieurs attributs physiques et psychologiques par le biais de crowdsourcing en ligne. En utilisant des features de bas niveau et des descripteurs de deep learning, nous avons inféré de manière automatique des jugements d'ambiance de lieux intérieurs avec un R^2 maximum de 0.53 et de 0.49 pour la perception de lieux extérieurs. Troisièmement, nous avons démontré la faisabilité de collecter des données contextuelles riches pour étudier la mobilité physique, les activités, les contextes d'ambiance, ainsi que le comportement social de jeunes adultes de nuit. Quatrièmement, en utilisant des techniques d'apprentissage statistique supervisé, nous avons classifié le comportement de consommation d'alcool de jeunes adultes dans un environnement nocturne et urbain réel. En utilisant des descripteurs dérivés de données de capteurs de smartphones ainsi que des logs d'utilisation d'applications, nous avons obtenu une précision générale de 76.7%.

Alors que cette thése contribue à une meilleure compréhension de la perception urbaine ainsi qu'au comportement des jeunes adultes de nuit dans des contextes bien définis, notre recherche contribue aussi à la compréhension computationnelle des lieux publics à grande échelle, avec une résolution spaciale et temporelle élevée, en utilisant une combinaison de crowdsensing mobile, de médias sociaux, d'apprentissage statistique, d'analyse de multimédia, et de crowdsourcing en ligne.

Mots-clé : crowdsensing mobile, ambiance, perception urbaine, lieux intérieurs, lieux extérieurs, jeunesse, vie nocturne, alcool, informatique omniprésente, informatique urbaine, informatique sociale, médias sociaux, Foursquare.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context and Motivation

CITIES are unique expressions of human activity and, at their core, are the intersection of physical spaces and the people who live in it. Cities are not only buildings and roads but also the people who use them and who, through the continuous exchange of ideas and intermingling, create new knowledge and innovate [77]. From a city perspective, public places have always played a central role in facilitating a socio-cultural habitat for people to counterbalance the daily grind of urban life, an environment away from home and work [149, 234, 40]. Places provide a social environment for individuals to communicate, share perspectives, and in the process form new friendships [68] or opportunities to find life partners [77]. With increasing urbanization, the role of public places in urban life has become more important than ever as they contribute towards the functioning and vitality of a city.

There is a growing body of research in urban studies and psychology to contextualize the understanding of places according to the perceptions of their inhabitants, and the socio-economic and psychological factors behind them [71]. In those domains, the literature has started investigating connections between psychological features of cities and key indicators like well-being and prosperity [167]. Given the inevitable growth of urban life worldwide, and the connections between places and well-being, an important goal is the development of scientific methodologies to provide “a better idea of how people perceive and experience places” [71].

One fundamental construct that relates place and experience is *ambiance*, defined as “the mood or feeling associated with a particular place” or “the character and atmosphere of a place” [137, 56]. As soon as we walk into a place or an unfamiliar neighborhood, we can tell if it is suitable for us. We ubiquitously judge restaurants, cafes, or bars according to their ambiance – whether a venue is energetic, bohemian, loud, or trendy. Similar perceptions are formed when visiting an unknown city – a lot can be told about a place from its appearance [98]. Overall, we form place impressions combining perceptual cues from the physical environment that involve most senses (vision and hearing, but also smell, taste, and touch) as well as prior knowledge of both the physical space and its inhabitants [81]. As urban dwellers, we rely on ambiance to make decisions that have long-term impact on how we interact with those places e.g., defining our favorite social hangouts and shaping our discoveries, including the kind of people we might end up meeting and interacting with.

Until recently, our ability to assess the visual and perceptual qualities of urban environments has been limited partly due to the lack of quantitative data. However, the growth of sensor-rich mobile devices, social media, and crowdsourcing platforms have opened ways to measure urban perception at scale. Social multimedia platforms (e.g., Flickr, Foursquare, Yelp, Instagram) that allow users to take and share photos using mobile devices have gained widespread adoption. These social media platforms represent a crowdsourced mechanism to document cities and places within them. Users are documenting not just the diversity and richness of places that city life offers, but also the character of the city including skyscrapers, monuments, touristic spots, public parks, etc. As a result, millions of images describing places and cities across the globe are available.

Complementary to this trend, due to an increased use of online crowdsourcing platforms to obtain judgments from diverse populations, scientists have started to use crowdsourcing as a medium to study urban perception for both indoor [83, 183] and outdoor environments [176, 171, 162], in a similar fashion as we form impressions of people [21]. In summary, mobile and social technologies are providing new opportunities to document, characterize, and gather impressions of urban environments.

In other disciplines it has been shown that cities have always been an attractive option for young people due to their better access to education, improved job opportunities, active nightlife, and better amenities [77]. However, urban public spaces are often seen as “adult space”, where young people and their practices are seen as dangerous, threatening the adult order in public spaces as well as the safety of others [210, 192, 211]. Youth are often stereotyped as “trouble-makers” creating public disturbance and disorder particularly during night time, resulting in increased surveillance and regulations [210, 211, 212]. Relatively little is known about the dynamics surrounding youth experiences in urban spaces particularly at night [212].

Mobile crowdsensing and social media provides the possibility to study questions related to youth populations and their environments that have been elusive in the past. Understanding nightlife, i.e., how cities and their youth inhabitants interact at night, is a relevant issue to multiple stakeholders including city officials, business associations, police departments, health and educational authorities. A vibrant nightlife scene can be simultaneously seen as an urban development strategy, an economic opportunity, a source of health and safety risks, and a way in which citizens co-create and appropriate the urban space [212]. Young people are key actors of nightlife, and as such become the focus of many of the above stakeholders, with respect to the design of strategies and policies that encourage the appropriation of the urban space while promoting healthy behaviors [100, 49]. We posit that the emergence of mobile and social technologies can contribute to the understanding of nightlife as experienced by young people, who use mobile and social technologies day and night.

In this dissertation, we take a multidisciplinary approach to the computational characterization of urban public places from the perspective of urban perception and study how young people can document and use these urban spaces, by integrating concepts from urban studies, human geography, social psychology, and computing, including mobile crowdsensing and online crowdsourcing, machine learning, and multimedia analysis.

1.2 Goals and Research Questions

In the thesis, we present the design, implementation and analysis of four mobile crowdsensing studies involving populations of young people contributing a variety of multifaceted crowdsourced data including mobile sensor data, mobile application logs, survey responses, and multimedia content (images and videos). We pursue the following objectives:

- **Objective 1**: To examine how crowdsensed images can be used for the analysis and automatic inference of urban perception in indoor and outdoor urban places. For both place types, urban perception are obtained for several physical and psychological constructs. Indoor places are judged along categories including *romantic*, *bohemian*, *formal* and *trendy*, among others. Outdoor perception are assessed along dimensions including *dangerous*, *dirty*, *happy*, *picturesque*, etc.
- **Objective 2**: To understand the going out behavior of young people during night time: the places where youth spend their weekend nights (physical mobility), the activities they perform (consumption of alcohol), the atmosphere of their hang-outs (ambiance context), and the people they hang out with (social context).
- **Objective 3**: To demonstrate the feasibility to automatically infer youth's nightlife activity of alcohol consumption based solely on mobile sensor data and application logs.

While this thesis contributes towards understanding urban perception and youth nightlife behavior in specific contexts, our research also contributes towards the computational understanding of urban places at scale with high spatial and temporal resolution using a combination of mobile crowdsensing, social media, machine learning, multimedia analysis, and online crowdsourcing. More specifically, we address the following research questions in this thesis:

RQ1: Data Collection via Mobile Crowdsensing: How can mobile crowdsensing be used to inform about the places, activity, and social context of youth nightlife patterns? How can a population of young people be engaged through mobile crowdsensing to collect images of urban environments? How can the correctness and validity of crowdsensed data, in particular images, be verified?

RQ2: Human Perception of Places: Can human observers reliably assess the perception of indoor and outdoor places using crowdsourced images? If so, what physical and psychological dimensions of perception can be consistently assessed? To what extent do crowdsourced annotations by external observers correspond to in-situ self-reports?

RQ3: Machine Perception of Places: Can crowdsourced judgments of place perception be automatically inferred using features extracted from images? To what extent do automatically extracted ambiance features represent the crowdsourced annotations by both in-situ observers and external online observers?

RQ4: Social Activity Recognition: Can alcohol consumption be automatically inferred from mobile sensor and log data in an uncontrolled, real-life nightlife setting? If so, what sensor features are more predictive of alcohol consumption?

1.3 Summary of Contributions

1.3.1 Mobile Crowdsensing

As a first contribution, we used diverse populations to collect rich contextual data using mobile crowdsensing to gain insights about people and urban places. Mobile Crowdsensing is a umbrella term used to describe sensing studies ranging from participatory mobile sensing to opportunistic mobile sensing for a variety of application scenarios including infrastructure sensing, personal analytics, and social computing [75, 123]. *Participatory sensing* involves the active participation of individuals to gather sensory data e.g., capturing images of dangerous looking streets, reporting civic issues (street lighting, illegal dumping of waste, potholes, etc.) [39]. *Opportunistic sensing* is an autonomous (or semi-autonomous) collection of data without the active involvement of individuals or explicit user interaction e.g., continuous sampling of mobile sensor data (such as GPS, Accelerometer, Bluetooth, etc.).

On one hand, opportunistic sensing provides scalability and diversity while reducing participants' response burden; on the other hand, participatory sensing provides valuable contextual information via experience sampling which otherwise is difficult to infer using raw sensor streams. Consequently, researchers have adopted a mix of participatory and opportunistic sensing approaches for studies which demand explicit user intervention to enrich contextual information, while leveraging the continuous collection of sensor data on participants' mobile devices [45, 216, 217].

Besides the use of smartphones for data collection, the ubiquity and popularity of mobile social multimedia platforms offers a promising avenue to obtain data about places and people. In all these platforms, users often take photos of many city areas and share them publicly resulting in millions of images across the globe. We refer to these platforms as a form of **implicit crowdsensing** where thousands of users are collecting and contributing data voluntarily without any external or definitive requirement. On the other hand, we refer to a mix of participatory and opportunistic sensing methodology as **explicit crowdsensing** as participants are explicitly instructed to intentionally gather certain types of data.

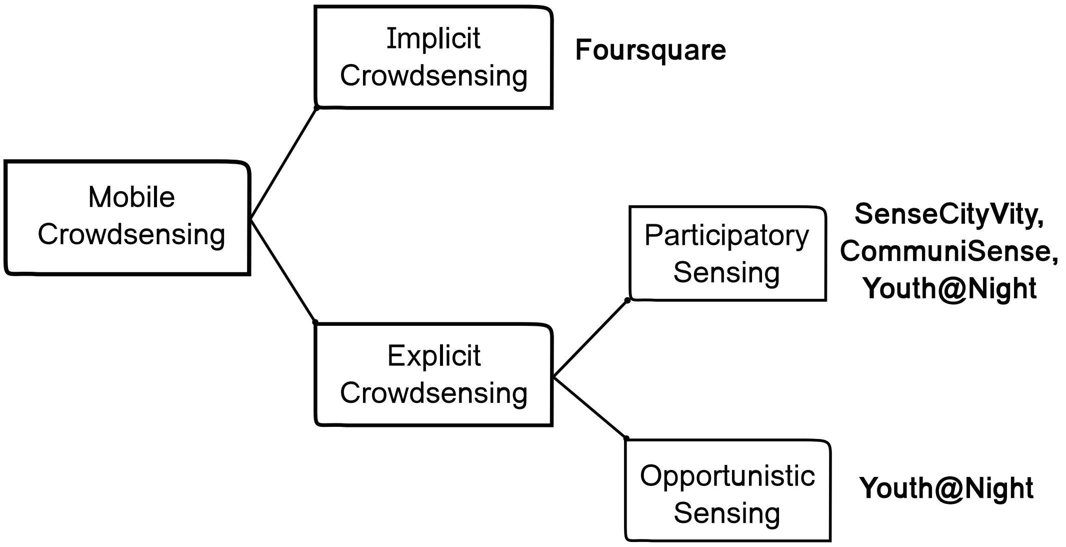


Figure 1.1 – Taxonomy of mobile crowdsensing used in the thesis.

In this thesis, we have adopted both implicit and explicit crowdsensing methodology to obtain data about people and places. We have successfully carried out one implicit crowdsensing using Foursquare data and three explicit mobile crowdsensing studies, ranging from a purely participatory approach (*SenseCityVity* and *CommuniSense*) to a combination of participatory and opportunistic sensing (*Youth@Night*). Figure 1.1 visually summarizes the different types of crowdsensing studies carried out in this thesis and Table 1.1 summarizes the data collected as part of these studies. Below we describe these studies and the collected data.

Foursquare Study

Foursquare is one of the world's most popular location-based social networks, with over 8 billion place visits (check-ins) at over 65 million places from over 50 million users worldwide [73]. Due to its large-scale place database and developer-friendly API, we collected over 50,000 images and 28,000 comments from 300 indoor places across six cities around the world (Table 1.1). The collected images were subsequently used to gather impressions of indoor places via online crowdsourcing as described in Section 1.3.2. The dataset and the analysis have been described in a paper published at the 23rd ACM International Conference on Multimedia (MM '15) [183].

SenseCityVity

SenseCityVity describes an approach aimed at documenting urban environments of cities in Mexico by youth through the use of mobile crowdsensing. The study was conducted in three cities, each one characterized by distinct geography, economic activity, and population. The project involved the development of a mobile crowdsensing platform and a deployment

Study	Feature	Location	Dataset
Foursquare	300 popular indoor places across six cities	Barcelona, Mexico City, New York City, Paris, Seattle, and Singapore	50,000 semantically-localized images; 28,000 comments
SenseCityVity	200 participants over 3 months duration	Guanajuato, Leon, and Sliao (Mexico)	7,000 geo-localized images; 380 videos; 13 video documentaries
CommuniSense	30 participants over 2 weeks	Nairobi (Kenya)	881 travel survey responses; 254 hazard reports; 101 geo-localized images
Youth@Night	241 participants over 3 months	Zurich and Lausanne (Switzerland)	8 million sensor points; 2,500 images; 894 video clips; 10,000 questionnaires; 40 interviews; 54,000 Foursquare check-ins

Table 1.1 – Summary of data collected during four mobile crowdsensing studies.

of an Urban Data Challenge (UDC) co-designed with young student volunteers to collect geo-localized images, audio, and video. More than 200 student volunteers participated in the UDC that resulted in a collection of over 7,000 geo-localized images, and 380 first-person perspective videos. The collected images contain outdoor scenes of each city's built environment, including touristic, historical, and business sites, residential neighborhoods, and areas with narrow streets and alleys. The image dataset was subsequently used to gather impressions of outdoor scenes as described in Section 1.3.2. Going beyond just a mapping exercise, the *SenseCityVity* project addresses the need for communities to become more aware of their urban environment and take collective action towards addressing some of the urban problems in their respective local communities. The data collected by and for the people provides an alternative and more comprehensive picture of the issues that matter to citizens. The *SenseCityVity* project and the collected data have been described in detail in a paper to appear at the IEEE Pervasive Computing 2016 [172].

CommuniSense

The use of mobile crowdsensing for infrastructure monitoring present exciting opportunities for developing cities, as they currently lack technologies to obtain reliable data on road infrastructure conditions. In this research thread, we built a mobile citizen-reporting application, *CommuniSense*, to support data collection and verification of road infrastructure conditions for the citizens of Nairobi. *CommuniSense* can be seen as a special case of the *SenseCityVity* project, where citizens turned the focus of their cameras downwards to capture images of road hazards, specifically potholes and speed-bumps. Using a two-week field study with 30 college students, we captured a total of 254 report submissions and 101 geo-localized images from different parts of Nairobi. The research conducted as part of the *CommuniSense* project has been published at the 17th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'15) [185].

Youth@Night

The Youth@Night project outlines a large-scale mobile crowdsensing study with more than 200 young participants over three months in two cities of Switzerland. The study was designed to capture the heterogeneity of youth behavior during night time and to examine the feasibility to automatically characterize alcohol consumption using mobile sensor data. For the study, we developed two smartphone applications that allowed participants to respond to various surveys using a participatory approach, while at the same time opportunistically collect sensor and log data. The study resulted in over 10,000 survey responses (in-situ and ex-situ), 8 million sensor data points, 2,500 images, and 894 videos.

During the study, participants recorded a total of 1,394 place visits (check-ins) at diverse places including personal homes, and non-commercial public spaces. Irrespective of place type, the study captured places along the full spectrum of social and ambiance variables, i.e., check-ins to public places alone, and house parties at private homes; or, public places which were reported to be quiet, and private places which were reported to be very loud. Using the video dataset, we conducted a computational analysis to measure the extent to which automatically extracted loudness and brightness of places represent the crowdsourced annotations by both in-situ participants and external online observers. To the best of our knowledge, this kind of analysis has not been reported earlier in the Ubicomp community.

Overall, we demonstrated a crowdsensing methodology to collect rich contextual data to improve our current understanding of youth nightlife practices in Switzerland. The proposed methodology and the analysis have been published in a paper to appear at the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16) [181].

1.3.2 Characterization and Inference of Urban Perception

One of the key challenges to understand how people perceive places is the difficulty in obtaining place impressions. Most of the related works in psychology are based on physically visiting a place and making observations about its atmosphere [83]. While this approach is ecologically valid, it is not scalable and does not capture the temporal features (e.g., a place that might be ideal for a business lunch, but that turns into a trendy loud bar in the night) and geographical aspects (e.g., places differ across cities). Our proposed approach is to gather place impressions based on images shared on social media, where observers rate ambiance after viewing images for that place in an online setting. This approach has the advantage of being scalable and can easily span national boundaries to help examining spatial and cross-cultural differences in ambiance perception. In addition, this approach also permits a better understanding of a place based on images taken during different times of the day.

To study ambiance of **indoor** places, we used the data collected from Foursquare (Table 1.1). To understand what type of images are most appropriate to describe ambiance and to assess how people perceive places, we designed online crowdsourcing experiments on MTurk using 50,000 images from 300 indoor places. Results showed that images with clear views of the environment were perceived by crowdworkers as being more informative of ambiance than other image categories. We demonstrated that reliable estimates of ambiance (as defined in the psychology literature) can be obtained using social media images for several of the investigated dimensions, suggesting the presence of visual cues to form place impressions. Furthermore, we found that most aggregate impressions of ambiance were similar across popular places in all six studied cities.

For **outdoor** environments, previous studies have examined urban perception using images from Google Street View (GSV) [176, 162, 23]. While GSV provides a scalable and automated way to obtain images, not every world region is equally represented in the platform [171]. Instead in this work, we used a subset of images collected during the *SenseCityVity* project. Using 144,000 individual judgments from MTurk, we gathered impressions of urban outdoor scenes along 12 dimensions. Statistical analysis showed that outdoor environments can be reliably assessed with respect to most urban dimensions in an online crowdsourcing setting. Furthermore, cross-city statistical analysis showed significant differences across cities. As a way of showing the additional value of mobile crowdsensing with respect to online street level imagery (which provides static and daytime views), we also investigated whether the perceptions of urban environments vary across different times of the day, and found that places in the evening are perceived as less happy, pleasant and preserved, when compared to the same place in the morning.

Our findings indicate the feasibility of using crowdsourced images to gather impressions of urban perception for both indoor and outdoor places. These findings further suggest the presence of visual cues used by raters to form place impressions. While there is a variety of potentially informative visual cues, the specific connections of visual features with urban perception still needs to be established. We hypothesized that some of the studied dimensions have the potential to be automatically recognized. To test this hypothesis, we conducted two studies to automatically infer place ambiance using visual cues from images. For inference, we build upon prior work in object classification and scene understanding using low-level image features including Color Histogram, GIST, HOG, LBP, SIFT and generic deep learning features (activation layer of convolutional neural network) [173, 165, 232]. Features extracted from deep learning with convolutional nets consistently outperformed individual and combinations of several low-level image features to infer the studied dimensions for both indoor and outdoor places. Our results further demonstrated the feasibility to automatically infer indoor ambiance with a maximum R^2 of 0.53 and outdoor perception with a maximum R^2 of 0.49.

To our knowledge, our work contributes one of the first results on how images collected from mobile crowdsensing can be used for characterization of impressions in indoor and outdoor places. From the perspective of multimedia computing, our work contributes towards developing automated ways to study social perception of urban places at scale. The applications derived from the above research are manifold. For indoor places, this could include hyper-local, ambiance-driven place search and discovery (e.g., a *trendy* place for a night-out or a *romantic* place for the wedding anniversary) to data-driven recommendations for place owners to improve the presentation of their venues (e.g., architecture design and style). For outdoor spaces, characterizing perception could potentially provide urban designers and city planners a data-driven and scalable approach to examine the physical appearance of cities and help design and evaluate urban policies informed by urban perception.

Much of our research on indoor places has been published at the 23rd and 24th ACM International Conference on Multimedia [183, 184]; while the work on outdoor perception has been published at the 6th ACM Symposium on Computing for Development (DEV'15) [186].

1.3.3 Characterizing Youth Drinking Behavior

Alcohol consumption is the number one risk factor for morbidity and mortality among young people in many countries in the developed world. Heavy drinking and related incidents in public on weekend nights are a concern for city authorities, and a nuisance for the general public [135, 221]. In late adolescence and early adulthood, excessive drinking and intoxication is more common than in any other life period, which carries a significant risk of adverse psychological, social, and physical health consequences [78]. To examine drinking habits, most previous studies in alcohol research have relied on self-reported assessments. This setting has three limitations. First, self-reported data on past alcohol consumption is prone to recall biases. Second, the environment in which drinking happens often differs from the one under which the self-assessment of drinking happens. Third, the collected data is limited by study design, focusing on single aspects of this complex phenomena. In order to increase research validity and advance understanding of contextual factors, smartphones offer a promising alternative to capture drinking-related phenomena.

Using the sensor data collected during the Youth@Night study (Table 1.1), we examined the feasibility to automatically classify drinking behavior of young adults in an urban, real nightlife setting. We found features extracted from accelerometer data to be the most informative among all features types with 75.8% accuracy, and that a combination of features results in an overall accuracy of 76.7%. Features extracted using location logs were the second best feature with an accuracy of 68.5%. We observed that Wifi and Bluetooth logs were also discriminant. To contextualize these findings, we explored the role of two potential confounding variables – going out (home *vs.* away) and gender – on the reported alcohol consumption. We found that these two variables alone cannot explain the more nuanced notions of drinking behavior observed in our data.

Our findings demonstrated the feasibility of classifying drinking behavior using smartphone data. From the mobile computing viewpoint, we believe it is a promising result.

1.4 Thesis Outline

The thesis is organized as follows. Chapter 2 outlines the Foursquare data collection, online crowdsourcing experiments to obtain human impressions, data analysis, and automatic inference of ambiance for indoor places. Chapter 3 follows a similar methodology to characterize and infer urban perception for outdoor places using the *SenseCityVity* dataset. Chapter 4 presents the design and implementation of the mobile crowdsensing study for road infrastructure monitoring in Nairobi, including the findings of a city-wide road quality survey. Chapter 5 outlines the design and implementation of the Youth@Night study including a detailed analysis of the questionnaire, survey, and video data, and proposes a machine learning pipeline to automatically classify alcohol consumption of young adults in Switzerland. Chapter 6 presents the conclusions of the thesis as well as possible directions for future work. Due to the multi-disciplinary nature of the thesis, the related work is discussed individually in each chapter.

1.5 Publications

This thesis outlines much of the research which have been previously published in the following peer-reviewed journals and conference papers. Additionally, the thesis reports research which is currently under review or unpublished work. The list of publications is provided in reverse chronological order below:

[P1] **Darshan Santani**, Rui Hu, Daniel Gatica-Perez. “InnerView: Learning Place Ambiance from Social Media Images”, in *Proceedings of the ACM International Conference on Multimedia (MM ’16)*, 2016

[P2] **Darshan Santani**, Joan-Isaac Biel, Florian Labhart, Jasmine Truong, Sara Landolt, Emmanuel Kuntsche, Daniel Gatica-Perez. “The Night is Young: Urban Crowdsourcing of Nightlife Patterns” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp ’16)*, 2016

[P3] **Darshan Santani**, Trinh-Minh-Tri Do, Florian Labhart, Sara Landolt, Emmanuel Kuntsche, Daniel Gatica-Perez. “DrinkSense: Characterizing Youth Drinking Behavior using Smartphones” *Currently Under Submission*, 2016

[P4] Salvador Ruiz-Correa, **Darshan Santani**, Beatriz Ramirez Salazar, Itzia Ruiz Correa, Fatima Alba Rendon-Huerta, Carlo Olmos Carrillo, Brisa Carmina Sandoval Mexicano, Angel Humberto Arcos Garcia, Rogelio Hasimoto Beltran and Daniel Gatica-Perez. “SenseCityVity: Mobile Sensing, Urban Awareness, and Collective Action in Mexico”, in *IEEE Pervasive Computing* (*Forthcoming*), 2016

[P5] **Darshan Santani**, Salvador Ruiz-Correa, Daniel Gatica-Perez “Looking at Cities in Mexico with Crowds ”, in *Proceedings of the ACM Symposium on Computing for Development (DEV ’15)*, 2015

[P6] **Darshan Santani**, Daniel Gatica-Perez. “Loud and Trendy: Crowdsourcing Impressions of Social Ambiance in Popular Indoor Urban Places” in *Proceedings of the ACM International Conference on Multimedia (MM’15)*, 2015

[P7] **Darshan Santani**, Jidraph Njuguna, Tierra Bills, Aisha W. Bryant, Reginald Bryant, Jonathan Ledgard, Daniel Gatica-Perez. “CommuniSense: Crowdsourcing Road Hazards in Nairobi”, in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI ’15)*, 2015

Funding Acknowledgments: This thesis was funded by the Swiss National Science Foundation (SNSF) through the *HAI* and *Youth@Night* projects. We also acknowledge the financial support of EPFL CODEV for the *SenseCityVity* project.

2 Characterizing Impressions of Ambiance in Indoor Places

2.1 Introduction

THERE is an increasing interest in social media and ubiquitous computing to characterize places beyond their function and towards psychological and emotional constructs. In this context, an area of active research is the development of “a better idea of how people perceive and experience places” [71]. As soon as we walk into a bar, cafe or a restaurant, we judge if the place is appropriate for us. In other words, we form place impressions combining perceptual cues that involve most senses (vision and hearing, but also smell, taste, and touch) as well as prior knowledge of both the physical space and its inhabitants [81]. As urban dwellers, we rely on ambiance to make decisions that have long-term impact, defining our favorite social hangouts and shaping our discoveries, including the kind of people we might end up meeting and interacting with.

One of the key challenges to understand how people perceive a place is the difficulty in obtaining place impressions. A standard approach involves physically visit a place and gather impressions making observations about its atmosphere [83]. Clearly, this approach is neither scalable nor captures contextual aspects of venues like time, by which a place might be ideal for a business lunch, but then turns into a trendy loud bar at night. In contrast, gathering place impressions based on images or videos shared on social media, where observers rate ambiance after viewing media items coming from a venue, has the advantages of being scalable, allowing the study of contextual factors like time, and spanning national boundaries to examine geographic and cultural differences in ambiance perception. In this work we conduct our ambiance analysis based on the social media approach.

Social media represents a crowdsourced mechanism to document urban places like restaurants, bars, clubs, and cafes. Due to the growth of sensor-rich mobile devices, online directory services like Yelp, Foursquare, and TripAdvisor are all popular today. These platforms provide users with functionality to search for places in a given region and to leave feedback in the form of reviews and comments highlighting their experiences [25]. These services co-exist with integrated services like Google Places, Facebook Places, and Instagram. In all of them, users often take photos at venues and share them publicly. As a result, millions of images documenting places across the globe are available. While many of the shared images are either personal or show food or drink related items, there are also images which provide views of the indoor environment and likely adequate to gauge the place atmosphere.

Relying on a combination of social media data collection, image crowdsourcing, data analysis, and automatic feature extraction we contribute research resources and findings towards the understanding of how social ambiance can be systematically studied. This chapter addresses the following research questions:

RQ2.1: What types of social media images best convey the ambiance of popular indoor places? (**RQ2.1** maps to thesis's **RQ2**)

RQ2.2: Can human observers reliably assess the perception of indoor places using crowdsourced images? If so, what physical and psychological dimensions of perception can be reliably assessed? (**RQ2.2** maps to thesis's **RQ2**)

RQ2.3: Can crowdsourced judgments of indoor ambiance be automatically inferred using low-level image and deep learning features extracted from images? (**RQ2.3** maps to thesis's **RQ3**)

In this chapter, we make the following five contributions:

1. Using Foursquare, we collected 50,000 images from 300 popular places from six cities worldwide in three world regions – North America (New York City, Seattle, and Mexico City), Europe (Barcelona and Paris) and Asia (Singapore) (Section 2.3). In addition to geographic and cultural diversity, these cities were chosen because of their active user population on Foursquare. The focus of our collection was on *popular* places in Foursquare rather than on arbitrary places.
2. We designed an image crowdsourcing experiment on Amazon's Mechanical Turk (MTurk) to assess the suitability of image categories that are most appropriate to describe the ambiance of a place (Section 2.4). Research in both psychology and computing has confirmed the feasibility of crowdsourcing as a way to conduct behavioral studies when appropriate incentives and mechanisms for quality control are established [113]. Using statistical tests, the results showed that images with clear views of the environment were perceived by crowdworkers as being more informative of ambiance than other image categories, like food and drinks, or groups of people. Based on the crowdsourcing results, we built a refined image corpus suitable for indoor ambiance characterization.
3. A priori, the social ambiance of places is not known to zero-acquaintance visitors or observers. We design a second crowdsourcing experiment to assess how people perceive places from the perspective of ambiance (Section 2.5). We asked crowdworkers to rate indoor ambiance along 13 different physical and psychological dimensions where images

served as stimuli to form place impressions. The ambiance categories include *romantic*, *bohemian*, *formal* and *trendy*, among others.

4. Based on the results obtained from the second experiment, we found that reliable estimates of ambiance can be obtained using social media images, suggesting the presence of strong visual cues to form place impressions. Furthermore, while we identify a few statistically significant differences across cities along four ambiance dimensions, most aggregate impressions of ambiance were similar across popular places in all cities, which open relevant questions about the roles that geography and background knowledge of observers might be playing.

5. We devised a methodology to automatically infer human impressions of place ambiance for the studied places, using a variety of low-level image features (including color histogram, texture, structural features, histogram of oriented gradients), and generic deep learning features. Our results indicated the feasibility to automatically infer indoor ambiance with a maximum R^2 of 0.53 using features extracted from a pre-trained convolutional neural network (CNN), precisely GoogLeNet model trained on *ImageNet* data. Furthermore, CNN features consistently outperformed individual and combinations of several low-level image descriptors for all the 13 studied dimensions.

The chapter is organized as follows. We begin with a review of related work (Section 2.2). Section 2.3 presents the methodology to select places and their associated images using Foursquare. Section 2.4 describes our first crowdsourcing experiment to identify the most adequate image categories for ambiance characterization. Next, Section 2.5 describes the design of the second crowdsourcing experiment to examine whether reliable estimates of indoor ambiance can be obtained using online crowdsourcing along 13 different physical and psychological dimensions. Using the aggregated annotations, Section 2.6 provides descriptive statistics and study differences of perceptual characteristics across cities for each ambiance label. Next, Section 2.7 proposes a machine learning methodology to automatically infer ambiance impressions using low-level image features and generic deep learning features. Section 2.7 was a joint work with Rui Hu of Idiap Research Institute. Finally, Section 2.8 concludes with a summary of our findings. This chapter describes much of the research that has been published here in these papers [183, 184].

2.2 Related Work

Given the multifaceted nature of our research questions, we review the related work along multiple research axes: ubiquitous and multimedia computing, social media, hospitality research, social psychology, and urban computing.

2.2.1 Place Characterization in Ubiquitous Computing

The existing work on place characterization in ubiquitous computing, computer vision, and audio processing has examined several aspects including physical properties of places like their geographic location [109]; place composition, including the scene layout and the objects present in the scene [161]; place function, i.e., home, work, or leisure places; and place occupancy and noise levels [214]. This research has used both automatic [130, 46] and semi-automatic approaches [230] and a variety of data sources often studied in isolation, including images, sensor data like GPS/Wifi, and RF data. Works like [130, 46] have used audio or audio-visual data to characterized places through phone apps. The studied place categories (personal places in [130], home, work in [46]) differ significantly from ours. A recent work [214] investigates the recognition of physical ambiance categories (occupancy, human chatter, noise and music levels) using standard audio features collected in-situ by users. In contrast, our work examines social images as source of data, impressions of people who are not physically at the places, and a much larger number of social ambiance categories.

2.2.2 Analysis of Social Media Images

The emergence of social multimedia platforms, which allow users to take and share photos using mobile devices, have gained wide spread adoption. In the social multimedia literature, the work in [231] studies the problem of recommending locations based on mobility traces extracted from GPS and social links, without using image information. In contrast, the work in [44] uses geo-localized images, travel blog text data, and manual user profiles to suggest trips. Other works involving social images include [61] and [48]. These cases are focused on outdoor places and do not address the atmosphere dimensions we studied in this chapter. Due to the availability of large amount of images on Flickr or Instagram, researchers have analyzed these platforms (recent examples include [93, 26]). In [26], using a corpus of one million Instagram images, the authors studied the relationship between photos containing a face and its social engagement factors and found that photos with faces are more likely to receive likes and comments. As it relates to our work, an interesting result is that only 20% of images were found to contain faces, which suggests that many other image categories (related to food, drinks, places, etc.) exist (for an example, refer to the small-scale coding study in [93]).

2.2.3 Study of Ambiance in Indoor Places

In hospitality and retail studies, there has been significant interest to examine the effect of physical ambiance cues or “atmospherics” such as color, lighting, or layout on customer perception and quality inferences [24, 208]. In a study conducted in a retail store, it was found that ambient (such as music, lighting, smell), design (such as color, ceilings, spatial layout) and social factors present in the store environment contribute towards higher merchandise and service quality [24]. In another study [90], the role of atmospherics across 10 full-service restaurants in Hong Kong was investigated. Using five dimensions of restaurant atmospherics (facility aesthetics, ambiance, spatial layout, employee factors, and view from the window) it was found that these dimensions have a significant influence on patrons' dining experience, and their willingness to pay more and recommend the restaurant to others. Similar results were obtained in another related work conducted across ethnic restaurants in the U.S. [127]. However, most of these studies are either done in controlled laboratory settings or based on questionnaires, which may have limitations with respect to ecological validity or recall biases.

Unlike the above research, we take a different direction examining the social perception of places (the ambiance impressions that people form about venues). Our proposed research is most closely related to work in social psychology [167, 81] which has investigated first impressions of places in connection to the personality of their inhabitants, mostly in controlled settings. A key first study [83] investigated the reliability (in terms of inter-rater agreement) of impressions of place ambiance and patron personality formed by (a) observers physically present at a number of indoor places, and (b) observers of Foursquare user profiles who visited those places (as opposed to views of places as we do). The results suggest that people do indeed form consistent impressions of ambiance and patrons traits. Extending this work using the same data, authors in [166] proposed to infer place ambiance using visual cues (including color and aesthetic features) extracted from the profile pictures of its patrons. Both these studies, however, only examined 49 places in one city (Austin, TX). In contrast, we study 300 places in six cities.

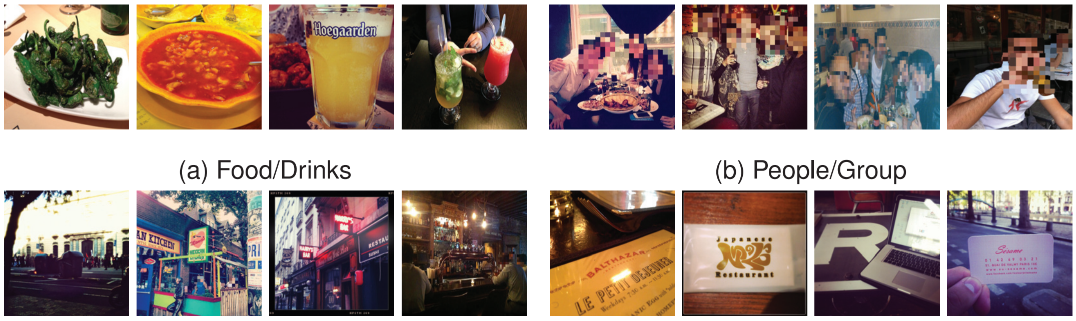
In the field of architecture and urban planning, there is a body of work to measure visual perceptions of outdoor built environments using qualitative research methods including interviews, visual preference surveys, and observations of the built environment using either actual or simulated images. More recently in urban computing, there is a recent push to measure urban perception of outdoor spaces using images from Google Street View [142, 151, 159] or Panoramio images to discover salient visual features in outdoor scenes in cities [61, 22, 233]. We review all the related work on outdoor environments in Section 3.2.3 of Chapter 3. In contrast with these works, we focus on indoor places and to the best of our knowledge, our work constitutes a first study to automatically infer ambiance impressions of indoor places from deep features learned from images shared on social media.

2.3 Dataset

In this section, we describe our methodology to select the list of popular places and their associated images across six cities in Foursquare.

2.3.1 Place Database

We ground our analysis on data collected from Foursquare. In Foursquare, users typically visit a place, announce their arrival (*check-in*) and share information about their visits to places.



(a) Food/Drinks

(b) People/Group

(c) Physical Environment

(d) None of these

Figure 2.1 – Sample images from *Random Image* corpus. Based on online annotations, a random set of four images which were annotated as (a) Food/Drinks, (b) People/Group, (c) Physical Environment, and (d) None of these. For privacy reasons, images showing faces have been pixelated. Best viewed in color.

City	Ratings	Photos	Visitors
Barcelona (BCN)	8.66 (0.67)	309.58 (383.53)	1,874.34 (2,371.43)
New York City (NYC)	9.31 (0.41)	463.62 (387.31)	8,272.16 (6,208.76)
Paris (PAR)	8.55 (0.63)	220.98 (254.16)	1,685.76 (1,433.14)
Seattle (SEA)	8.95 (0.38)	240.7 (147.94)	3,533.54 (1,815.34)
Singapore (SG)	8.29 (0.86)	304.88 (206.58)	3,457.64 (3,916.89)
Mexico City (MXC)	8.78 (0.49)	361.34 (374.85)	3,692.56 (3,578.84)

Table 2.1 – Summary statistics of Foursquare data. For each city, mean attribute scores of popular places are shown, along with their standard deviations (shown in brackets).

Each place on Foursquare maintains a profile page, which contains general information about the place (address, directions, phone number, etc.), in addition to place-specific data such as its popularity, total number of check-ins and past visitors, and a collection of images uploaded by users. As per Foursquare rules, a place or a venue is a geographical location with fixed spatial coordinates, i.e., latitude and longitude. Using the Foursquare public API, we collected all the relevant information for a selected place. Note that Foursquare has changed its mobile application and API significantly since our data collection. Throughout this chapter, we will use place and venue interchangeably in the context of Foursquare.

For our analysis, we studied popular indoor places on Foursquare for six cities around the world – Barcelona, Mexico City, New York City, Paris, Seattle, and Singapore. These cities were chosen for two reasons. First, they all are large cities in diverse world regions, and are known to have a vibrant urban life. Second, they all have an active user population on Foursquare. For each city, we chose 50 of the most popular places in each city which fall under the Foursquare-defined category of either being “Food” or “Nightlife Spots”, which means cafes, restaurants, or bars. Table 2.1 lists the mean values of Foursquare data for all 50 places in each city. For indoor ambiance work, we focused on studying popular indoor places as opposed to arbitrary places (i.e., indoor or outdoor, and that might or might not be represented on Foursquare).

Place selection was performed manually by me, taking into account place popularity, number of checkins, number of past visitors and number of available images. As image quality was an important criterion for selecting a place, we ignored all places which did not had any good-quality images such as dark images. In Table 2.1, using data obtained from Foursquare API, we noticed that the user-generated mean rating of places selected for the study is above 8.2 (on a scale from 1 to 10) for all cities, confirming the popularity of places. We also observe that these places are frequently visited by a large visitor population.

2.3.2 Image Corpora

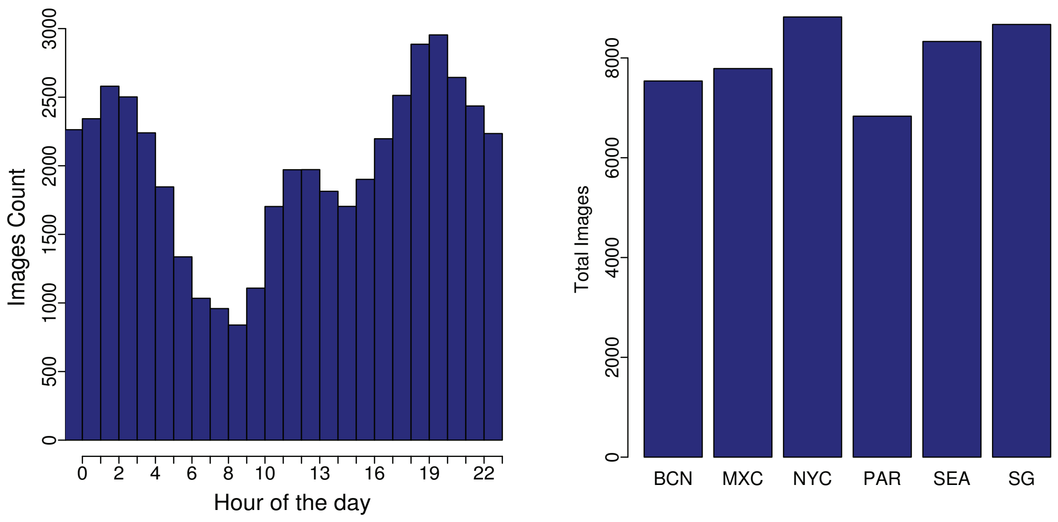
The second important consideration was the selection of images for each chosen place. Using the Foursquare public API, we collected all the publicly available images for each selected place to build the *50K image* corpus which is described below.

50K Image Corpus

Images for a place listed on Foursquare can be obtained via the API, but due to rate limits imposed by the API, we have access to at most 200 publicly visible images per place. We gathered a total of 50,023 images for all 300 places. This gives an average of 166 images per place, which is lower than the average estimated from the profile metadata ($\geq$ 300), yet it remains a large number. In addition to gathering the images, the API also provides information on the image source (i.e., the application used to generate the photo), creation time, user metadata, and other attributes such as image height and width. However, due to API restrictions we had access to metadata information for only 47,980 (96%) images, which were fairly distributed across all six studied cities (Figure 2.2b)

Using the metadata information, we found that the median height and width of an image in our collected corpus is 720 pixels, with 55% of images taken via iPhone, 19% via an Android device, 2% by Blackberry devices, and 22% uploaded via Instagram. 42% of images were uploaded by females, 54% of images by male users on Foursquare, while the rest 4% of users chose not to disclose their gender information on Foursquare. We also plot the distribution of image creation times in Figure 2.2a. We identify three distinctive peaks – the first one occurs during the lunch hour (11am–1pm), the second peak around dinner time (6–8pm), and the last one occurs after midnight and early hours of the morning (nightlife). This result confirms our intuition that social media images provide a well-suited medium to capture places during different times of the day.

Given that each place in our database has on average more than 300 user-contributed images (Table 2.1), we decided to select a small number of images per place to illustrate the



(a)

(b)

Figure 2.2 – Plots showing the image metadata for the *50K Image Corpus* a) Histogram showing the frequency of images taken during different times of the day. b) Barplot showing the total number of images per city.

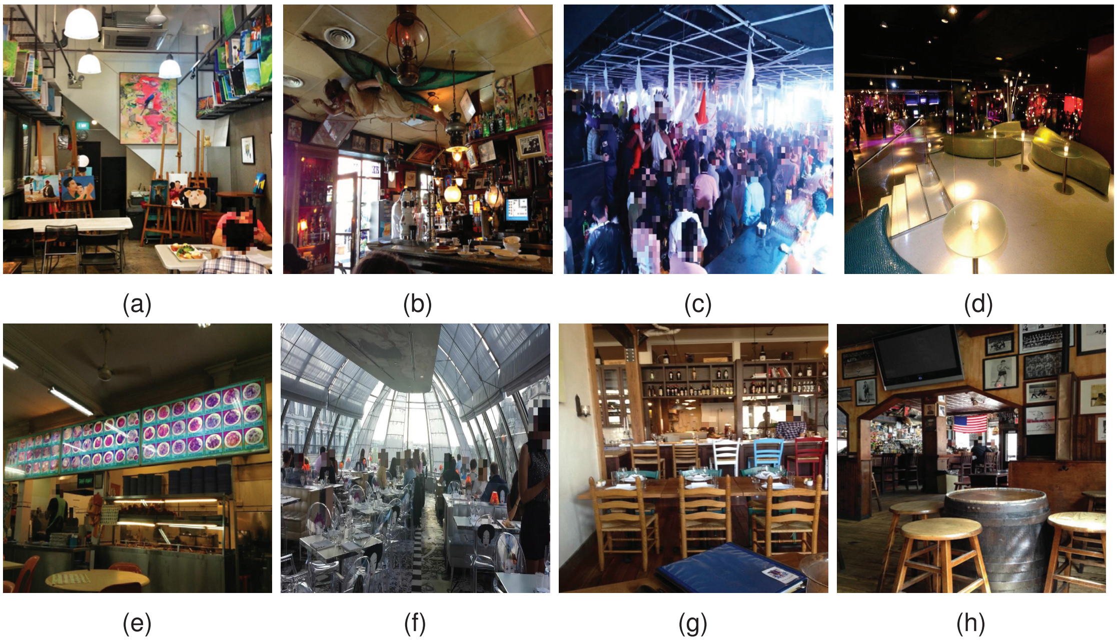
place's atmosphere. This decision was motivated by the need to account for the variability in image quality, while at the same time providing general views of a place, without complicating the annotation process. Moreover, having more than one image for a place gives us the flexibility to show the place at different times of the day. Our hypothesis is that images of the physical environment will often be more representative of the place ambiance, so to test our hypothesis we built two image corpora, which are described next.

Random Image Corpus

It is challenging to select a small number of representative images that can accurately describe the ambiance of a place. One approach is to randomly select them given the collection of all images available for a place. We follow this approach to build a *random image* corpus. For the study described in this section, since we are interested in only a few images to represent each place, we randomly sampled three images per place from the *50K image* corpus, to build a *random image* corpus of 900 images for all 300 places. Figure 2.1 highlights a sample of selected images from this corpus.

Physical Environment Image Corpus

Our second approach is to build an image corpus with clear views of the physical environment. We manually select a small number of images per place that satisfy this condition. Although this task can potentially be automated (Section 2.7), we have chosen to manually control the



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 2.3 – Sample images from *Physical Environment Image* corpus. Based on MTurk annotations, images which scored the **highest** on (a) Artsy, (b) Creepy, (c) Loud, and (d) Trendy; and **lowest** on (e) Artsy, (f) Creepy, (g) Loud, and (h) Trendy. For privacy reasons, images showing faces have been pixelated. Best viewed in color.

quality of data for the crowdsourcing experiment. The selection was performed by me after browsing through all the user-contributed images. During the process, we opted for images with a view clearly showing the space from different angles (with or without the presence of visitors). To the best of our ability, we avoided images where one can potentially identify faces, to protect the privacy of individuals. Moreover, images that showed the venue name or any other information that explicitly revealed the identity of the place were discarded e.g., an image showing Starbucks or Hard Rock Cafe logos, to reduce any bias while characterizing the place ambiance. Figure 2.3 presents a sample of selected images from this corpus. Note that all these attributes cannot be controlled for while choosing images randomly.

The *physical environment image* corpus also contains 900 images for all 300 places (three images per place). The manual selection was constrained by the quality of Foursquare images. At times, we encountered images which were not optimally bright or clear. However, this setting is realistic due to the absence of any beautified or vendor-provided images, which can potentially add biases to the impressions.

2.4 Experiment 1: Suitability of Images to Convey Ambiance

In this section we address RQ2.1, i.e., we use both image corpora to judge which approach results in better "ambiance quality", that is, images which are more adequate to convey

Method	**Physical Environment**			
	Top 2		Bottom 3	
	Ambiance	Phy. Env.	Ambiance	Phy. Env.
Majority Vote	91.7%	95.8%	8.3%	4.2%
Median	89.7%	94.7%	10.3%	5.3%

(a) Physical Environment Image Corpora

Method	**Random**			
	Top 2		Bottom 3	
	Ambiance	Phy. Env.	Ambiance	Phy. Env.
Majority Vote	5.8%	2.7%	94.2%	97.3%
Median	4.1%	2.0%	95.9%	98.0%

(b) Random Image Corpora

Table 2.2 – Table showing the summary statistics for each aggregation method for a) Physical Environment, b) Random image corpora. For each method, we show the percentage of images from both image corpora which are either in Top 2 ranks (rank 1 or 2), or ranked in Bottom 3 (ranks 3,4,5).

ambiance according to crowd judgments. On one hand, random selection of images is a realistic “in the wild” setting that provides an automated way to collect images. However, it will represent a valid approach only if it results in a collection of images which provide sufficient visual cues to characterize place ambiance. On the other hand, the manual selection of physical environment images is a controlled setting that satisfies the criteria described in the previous section.

2.4.1 Crowdsourcing Image Impressions

Our hypothesis is that many images in the *random image* corpus might not be perceived by crowdworkers as ideal to characterize a place’s ambiance, as they do not contain visual cues to gauge a place’s physical environment. In our exploratory inspection, most of these random images contain food items or show groups of people. To answer RQ1, we conducted a crowdsourcing study to gather the perceived ability of both image corpora – *random* and *physical environment*, to describe a place’s ambiance and physical environment. For crowdsourcing, we used MTurk and chose US-based workers with at least 95% approval rate for historical HITs (Human Intelligence Tasks). In addition, to increase the potential reliability of MTurk annotations, we only chose “Master” annotators, which typically involves a worker pool with an excellent track record of completing tasks with precision.

For each HIT annotation task, we picked a set of five images per place, consisting of two from the *physical environment image* corpus, and three from the *random image* corpus. We ensured that images from the two sets do not overlap. In each HIT, workers were asked to view these five images and then answer three questions. In the first question, the workers were asked to rank the images based on how informative they were of the ambiance of the place. In the second one, workers were asked to rank the same set of images based on their degree of information about the physical environment of the place. The third question asked workers to categorize the images in four classes: a) Food/Drinks, b) People/Group, c) Physical Environment, and d) None of these. For these questions, no explicit definitions of ambiance, physical environment, food/drinks or people were provided, as we wanted the workers to rely on their internal representation of these concepts.

In the two ranking questions, images cannot be given the same rank, each image needed to have a different rank. For the image categorization task, the workers were asked to choose exactly one category for each image. If the images had the same rank or fell into one or more categories, we asked the annotators to pick the rank or category that for them was the best choice. We collected 10 annotations for each HIT across all 300 places, for a total of 3,000 responses for every question. Every worker was reimbursed 0.15 USD per HIT.

We also gathered crowdworkers' demographics via an email-based survey. We asked workers about their age group, gender, education level, current place of residence (categorized as either rural, suburbs, small-sized town, mid-sized town or city), and any experience of living in a big city. We also inquired them about their typical frequency to go out for food or drinks (almost every day, 2-3 times per week, once a week, 1-2 times a month, or less than once a month). These questions were designed to understand the ability of workers to rate images for ambiance and physical environment based on previous experiences.

2.4.2 Results

In this subsection, we present the results of our first crowdsourcing experiment.

Worker Participation and Demographics

For a total number of 3,000 HITs available for this experiment, we observe that a typical worker completed an average of 39 HITs. While 50% of the workers submitted less than 9 HITs, the worker with the highest number of HITs completed 295 assignments. We observe a long-tailed distribution in HIT completion times (mean: 114 secs, median: 88 secs, max: 593 secs). It is worth noting that we allocated a maximum of 10 minutes per HIT.

We had a pool of 101 workers who responded to our HITs. Of all HIT respondents, 32% replied to our demographics survey. We notice a balanced gender ratio (50% of workers being female), which corroborates earlier findings in the literature [169]. While only 34% of our worker pool currently lives in a big city, 75% of them had already experienced city living in the past. Furthermore, 56% of them go out for food or drinks at least once a week. These findings provide evidence that the majority of respondents are likely capable to assess ambiance in urban environments. We also notice that the worker population is relatively not so young with the most popular category (53%) being the age group of 35-50 years old. Note that the worker demographics reported here encompasses the worker population in both online crowdsourcing experiments of this chapter.

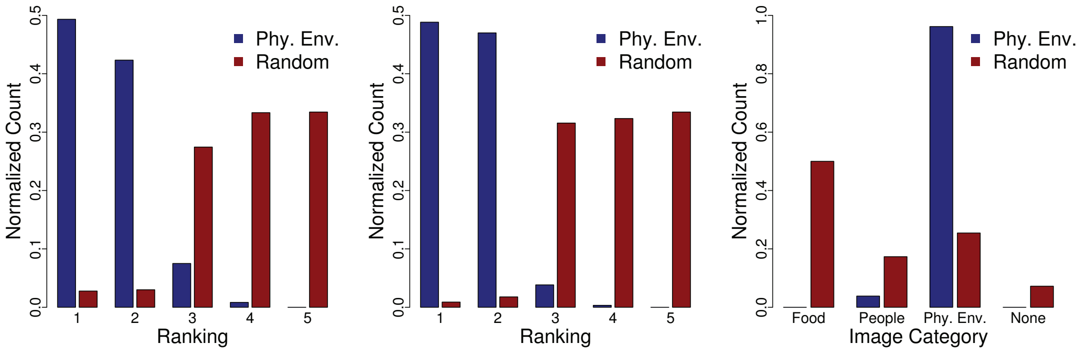
Analysis of Annotations

Now we turn our focus towards assessing the suitability of each image corpus to convey ambiance. As mentioned earlier, for each HIT we collected 10 impressions per place. Thus, it becomes essential to consider the role of different aggregation methods in analyzing the results. Aggregation is used to create a composite score per place given the 10 responses for each question. In other words, for every question, aggregation is performed at the place-level. We use two different aggregation techniques. The first one is the *majority vote*, where we compute the majority score given the 10 impressions for each place. We then summarize the results based on 300 majority impressions. For the *median* method, we compute the median as the composite score across 10 impressions for each place.

Table 2.2 lists the summary statistics for the two aggregation techniques. For each aggregation technique and each corpus, we report the percentage of images which are in Top 2 ranks (ranks 1,2) and Bottom 3 ranks (ranks 3,4,5). We list these statistics for both the ambiance and physical environment questions. For the *majority vote* technique, manually selected images (*physical environment image* corpus) are in Top 2 ranks 91.7% for ambiance and 95.8% for physical environment, whereas random images are in Bottom 3 ranks for 94.2% and 97.3%, respectively. Note that a random ranking method would assign the manually selected image in the Top 2 rank with a probability of $1/10$ $(1/\binom{5}{2})$. We also plot the histogram of rankings for image sets from both corpora in Figures 2.4a and 2.4b. These results show that manually selected images are associated with higher ranks, while the random set of images are associated with lower ranks for both ambiance and physical environment, irrespective of the aggregation technique.

In addition to asking annotators to rank images, we also asked them to classify images into one of the four categories (food/drinks, people/group, physical environment, and none of these), as described in Section 2.4.1. In Figure 2.4c, we plot the assigned category for the *majority vote* technique. We observe that images from the *physical environment image* corpus are labeled as describing the physical environment in 96.2% of the cases. In contrast, images from the *random image* corpus are representative of either food/drinks, or people, or other in 74.6% of cases combined, and showing food items or people in 67% of cases.

Statistical Comparison: We perform a statistical comparison of rankings between both image corpora for ambiance and physical environment dimensions. We performed the Wilcoxon signed-rank statistical test, which is a non-parametric test to compare the mean ranks of two populations [220]. At the 99% confidence level, we obtained a p-value $<$ 2.2×10^{-16} for both dimensions, validating the hypothesis that manually selected images are perceived by crowdworkers as better describing ambiance and physical environment.



(a) Ambiance (b) Physical Environment (c) Image Category

Figure 2.4 – Results for *Majority Vote* aggregation technique. Plot showing the histograms for a) Ambiance, b) Physical Environment, and c) Image Category, for both *Physical Environment Image* and *Random Image* corpus.

In summary, these results provide an answer to RQ2.1, validating that images with clear views of the environment from Foursquare places are perceived as more suitable to characterize indoor ambiance than other image categories, as they contain visual cues to gauge a place's ambiance and physical environment. Note that in this section we report the summary statistics across all cities combined. Individual trends for each city are similar to the overall trends. Section 2.7.3 investigates the use of automatic techniques to explore visual categories representative of the studied places in connection to indoor ambiance.

2.5 Experiment 2: Crowdsourcing Place Ambiance

A priori, the ambiance of places is not known to zero-acquaintance observers. In this section, we address the second question (RQ2.2) i.e., whether reliable estimates of ambiance can be obtained using Foursquare images. Based on the *physical environment image* corpus of 900 images across 300 places, we design a second online crowdsourcing experiment and asked crowdworkers to rate indoor ambiance along 13 different physical and psychological dimensions where images served as stimuli to form place impressions.

2.5.1 Selection of Ambiance Categories

In order to select dimensions to characterize place ambiance, we base our methodology on prior work [83] in which the authors proposed a rating instrument consisting of 41 dimensions for ambiance characterization. In our work, we chose 13 dimensions for which the corresponding intraclass correlations were amongst the highest as reported in [83]. Note that many dimensions in [83] did not reach sufficient inter-annotator agreement. We used a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5) to judge the ambiance labels, while [83] used a 3-point categorical scale (yes, maybe, no). In the rest of

Label	BCN	NYC	PAR	SEA	SG	MXC	Combined	Graham [83]
Artsy	0.81	0.66	0.69	0.72	0.80	0.76	0.76	0.63
Bohemian	0.63	0.58	0.54	0.54	0.66	0.66	0.62	0.67
Conservative	0.67	0.77	0.78	0.67	0.70	0.85	0.76	0.77
Creepy	0.54	0.62	0.60	0.57	*0.32*	0.62	0.59	0.81
Dingy	0.74	0.81	0.59	0.69	0.67	0.81	0.74	0.74
Formal	0.76	0.93	0.93	0.89	0.89	0.90	0.91	0.82
Sophisticated	0.68	0.91	0.90	0.85	0.80	0.87	0.86	0.70
Loud	0.80	0.81	0.76	0.74	0.82	0.82	0.80	0.74
Old-fashioned	0.82	0.46	0.78	0.45	0.72	0.67	0.72	0.67
Off the beaten path	0.58	0.62	0.39	0.54	0.61	0.59	0.58	0.73
Romantic	0.38	0.84	0.86	0.80	0.83	0.86	0.82	0.63
Trendy	0.69	0.71	0.50	0.43	0.68	0.85	0.69	0.58
Up-scale	0.69	0.91	0.90	0.85	0.81	0.85	0.86	0.76

Table 2.3 – $ICC(1,k)$ scores of 13 ambiance dimensions for each city. $ICC(1,k)$ scores obtained in [83] are also shown in the last column for comparison. Cells marked in *italic* are not statistically significant at $p < 0.01$.

the chapter, we will use the terms dimensions and labels interchangeably in the context of ambiance categories. The list of selected labels is shown in alphabetical order in Table 2.3.

2.5.2 Crowdsourcing Ambiance Impressions

To answer RQ2.2, crowdsourcing was employed to gather ambiance impressions. We used MTurk with the same worker qualification requirements as the first experiment (Section 2.4.1). In each HIT, the workers were asked to view three images corresponding to a place, and then rate their personal impressions of the place ambiance based on what they saw. As part of the annotation interface, we ensured that workers viewed images in high resolution (and not just the image thumbnails). People were given a previous definition of each ambiance category. Moreover, workers were not informed about the city under study to reduce potential bias and stereotyping associated to the city identity. We collected 10 annotations for each dimension across all 300 places, for a total of 3,000 responses.

2.5.3 Results

For the 3,000 available HITs in this experiment, a typical worker completed an average of 56 HITs, with one worker completing 270 HIT assignments. When compared to the first experiment, similar results were obtained for HIT completion times (mean: 97 secs, median: 68 secs, max: 596 secs).

We turn our focus towards assessing the reliability of the annotations. We measure the inter-annotator consensus by computing intraclass correlation (ICC) among ratings given by the worker pool. Our annotation procedure requires every place to be judged by k annotators

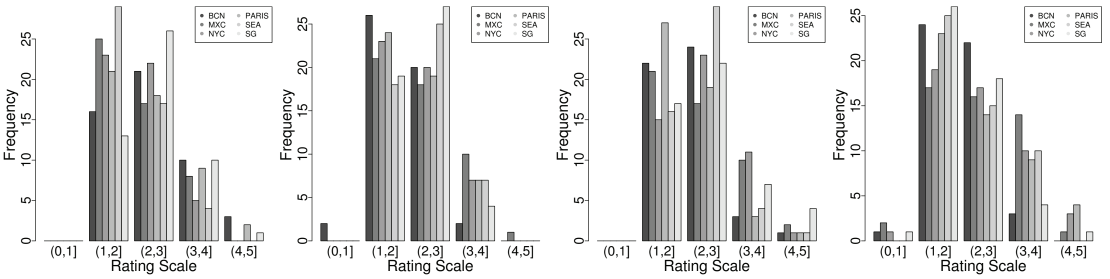
Label	BCN	MXC	NYC	PAR	SEA	SG	Combined
Artsy	2.54	2.20	2.14	2.36	2.05	2.46	2.29
Bohemian	2.34	1.94	2.07	2.09	1.99	2.04	2.08
Conservative	2.04	2.36	2.33	2.17	2.37	2.28	2.26
Creepy	1.33	1.37	1.21	1.20	1.21	1.18	1.25
Dingy	1.68	1.61	1.60	1.49	1.57	1.49	1.57
Formal	1.60	2.13	2.14	2.01	1.95	1.62	1.91
Sophisticated	2.09	2.41	2.42	2.37	2.20	2.15	2.27
Loud	2.30	2.45	2.51	2.09	2.33	2.49	2.36
Old-fashioned	2.20	2.30	2.33	1.90	2.44	2.16	2.22
Off the beaten path	2.27	1.88	2.06	1.89	1.99	1.96	2.01
Romantic	1.77	2.09	1.95	1.92	1.86	1.80	1.90
Trendy	2.34	2.55	2.55	2.49	2.45	2.54	2.49
Up-scale	1.93	2.36	2.39	2.36	2.13	2.01	2.20

Table 2.4 – Means of annotation scores for each city and label.

randomly selected from a larger population of K workers ($k = 10$, while K is unknown as we have no means to estimate the MTurk "Masters" worker population). Consequently, $ICC(1,1)$ and $ICC(1,k)$ values, which respectively stand for single and average ICC measures [189], are computed for each ambiance dimension across all cities.

Table 2.3 reports the $ICC(1,k)$ values for all cities. In addition to listing the individual scores for each city and label, we also report the combined $ICC(1,k)$ scores for each label for the whole dataset, where we have combined all places across cities. We observe acceptable inter-rater reliability for many labels, with all the scores being statistically significant (p-value $<$ 0.01), with the exception of *creepy* label in Singapore. Furthermore, we notice that the inter-rater reliability for labels *formal*, *sophisticated*, *romantic*, and *up-scale* is typically high (above 0.8) for most of the cities. Using correlation analysis between labels (which is presented in Section 2.6.2), we find that these four labels are collinear, with pairwise correlations exceeding 0.8. It is interesting to note that label *loud* achieved high agreement from images not showing any sound (0.8 combined score). On the other hand, labels *creepy* and *off the beaten path* are the labels with the lowest ICC (below 0.6 for the combined score).

Importantly, these reliability scores are comparable to the ones obtained by Graham et al. [83] (last column of Table 2.3), who conducted a similar study, but where the raters physically visited every venue; while in our case online images act as a stimuli. To summarize, these results provide an answer to RQ2.2 as they suggest that consistent impressions of place ambiance can be formed based upon images contributed in social media, which further suggests that there might be strong visual cues present for annotators to form accurate place impressions. The investigation of automatically inferring ambiance impressions using visual cues from images is presented in Section 2.7.



(a) Artsy (b) Conservative (c) Loud (d) Sophisticated

Figure 2.5 – Barplots comparing the mean annotation scores across all cities for a) Artsy, b) Conservative, c) Loud, and d) Sophisticated.

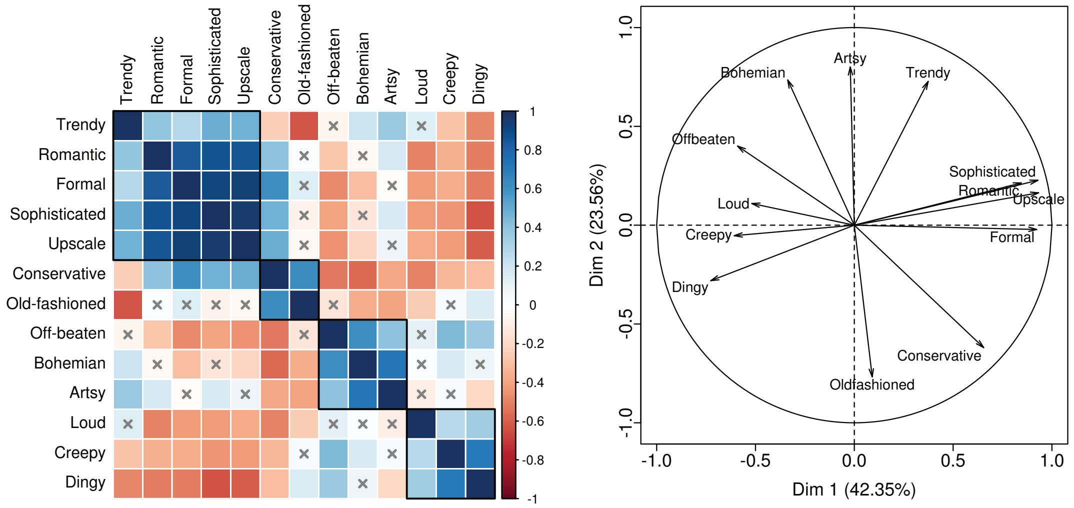
2.6 Characterizing Ambiance Impressions Across Cities

In this section, we present descriptive statistics and study differences of ambiance impressions across six studied cities for each ambiance label.

2.6.1 Descriptive Statistics

Table 2.4 lists the descriptive statistics (mean annotation score) for each city and label. The mean scores are derived as follows: first, for every place we compute the mean score for each ambiance label, using the 10 annotations per label for each place; we then compute the mean scores and standard deviations for each city and label using the 50 places for each city. At the level of individual annotations, minimum and maximum values are 1 and 5 respectively for all each city and label, showing that the full scale is used by the crowdworkers. The mean value obtained for all labels and all cities is below 3, which indicates a trend towards disagreement with the corresponding label. On the other hand, each city has venues that score high for each dimension.

In all cities, except Barcelona, the mean score for *trendy* is the highest amongst all labels; Barcelona places score the maximum on being *artsy*. *Creepy* scores the lowest (along with the lowest variance) for all cities, which is not surprising given that all places are popular places in their respective cities. From Table 2.4, we do not observe much variation in the mean values across cities, but a few differences stand out. For instance, the mean differences of the *formal* attribute between NYC and Barcelona, and the *old fashioned* attribute between Paris and Seattle exceed 0.5, potentially suggesting differences in place perceptions. We explore this further in Section 2.6.3. To visually aid the comparison between cities, we show the barplots of the binned mean annotation scores for four the ambiance labels in Figure 2.5, where finer differences can be observed across cities and relative ratings.



(a) Correlation Matrix

(b) PCA

Figure 2.6 – a) Plot showing the correlation matrix between ambiance dimensions. Black rectangular borders indicate the four distinct clusters found in the correlation matrix. Cells marked **X** are **not** statistically significant at $p < 0.01$. b) Plot showing the first two principal components on aggregated place annotation scores across all cities.

2.6.2 Correlation and PCA Analysis

To look for linear relationship between ambiance labels, we perform correlation analysis using the mean annotation scores for all ambiance labels. Figure 2.6a visualizes the correlation matrix across all ambiance dimensions. We have used hierarchical clustering to re-order the correlation matrix in order to reveal its underlying structure. We color code the matrix instead of providing numerical scores to facilitate the discussion. We observe four distinct clusters. Starting from top left in the first cluster, labels *formal*, *sophisticated*, *romantic* and *up-scale* are highly collinear with pairwise correlations exceeding 0.8. The second cluster consists of places which are either *conservative* or *old-fashioned*, and the third cluster consists of *off-beaten*, *bohemian* or *artsy* places. The fourth cluster (bottom-right) lies on the opposite spectrum with respect to cluster one, and consists of *loud*, *dingy* and *creepy* places. Each of these four clusters clearly correspond to different ambiances. Furthermore, we can also observe significant negative correlations between dimensions in cluster one and cluster four and between clusters two and three.

To further explore the relationship between labels, we perform principal component analysis (PCA) on the aggregated annotation scores for all 300 places. PCA is a statistical method to linearly transform high dimensional data to a set of lower orthogonal dimensions that best explains the variance in the data [157]. Figure 2.6b shows the first two principal components which in total explain 66% of the variance in the annotation scores. The first component, which accounts for 42% of variance, contains dimensions that resemble either the positive (cluster 1 in Figure 2.6a) or negative attributes (cluster 4 in Figure 2.6a) respectively, on the right and left side of X-axis. The second principal component explain 24% of variance and contain labels associated with *trendiness* or *artsy* nature of places on the positive Y-axis, whereas the negative Y-axis contains labels on the opposite spectrum (*conservative*, *old fashioned*), as shown in Figure 2.6b. Overall, these findings corroborate the correlation analysis, potentially suggest the presence of halo effect in indoor ambiance impressions [144], and have support from research in environmental psychology [174].

2.6.3 Statistical Comparison

To better understand whether mean differences across cities for some of these ambiance labels are statistically significant, we perform the Tukey's honest significant difference (HSD) test. Tukey's HSD test is a statistical procedure for groups which compares all possible pairs of mean values for each group, the null hypothesis being that the mean values being compared are drawn from the same population [207]. We perform the HSD test to compute pairwise comparisons of mean values between cities for each ambiance label, which result in a total of 195 comparisons (15 city-wise pairs across 13 dimensions). Table 2.5 lists only the significant results of the Tukey's HSD test, where the differences in the observed means are statistically significant at p-value $<$ 0.01. Based on these statistics and commenting only on results where differences is larger than 0.4, we observe that:

1. Popular places in Seattle are perceived as less *artsy* compared to places in Barcelona (Figure 2.5);
2. Popular places in Paris are perceived as less *old fashioned* compared to New York City and Seattle.

To validate the statistical significance of the Tukey's HSD test, we perform a series of pairwise Kolmogorov-Smirnov test (KS test) across all cities and labels. The KS test is a non-parametric test to compare the empirical distributions of two samples, with the null hypothesis being that the two samples are drawn from the same distribution [133]. We perform the KS test to compare the cumulative distribution functions of each city-pair across each dimension (195 comparisons). We report the $p-$values for the KS test in Table 2.5 for a statistical level $\alpha = 0.01$. Results from the KS test confirms most of the results from the Tukey's HSD test. It is worth noting that if we increase the significance level (α) to 0.05, we observe differences across other city-pairs to be statistically significant as well.

To conclude this subsection, our study shows that most of the differences across cities for each of the ambiance dimensions are not statistically significant. This result is interesting in itself as it might suggest that popular places in social media in cosmopolitan cities have many points in common. To our knowledge, this is a result that has not been reported before in social media research, but that could have some support from literature that discusses the "uniformization of taste" in globalized cities [50] and social media content. This said, any possible interpretation would have to be further validated with more data and a combination of further data analysis and ethnography.

Label	**City Pair**	**Mean Difference**	$p-$**value** $\times 10^{-3}$
Old Fashioned	SEA–PAR	$+0.544$	0.099 (0.007)
Artsy	SEA–BCN	-0.492	4.26 (6.18)
Old Fashioned	PAR–NYC	-0.434	4.09 (0.67)
Bohemian	MEX–BCN	-0.398	3.70 (*39.68*)
Off the beaten path	MEX–BCN	-0.386	1.43 (0.051)
Off the beaten path	PAR–BCN	-0.376	2.11 (0.67)

Table 2.5 – Tukey's HSD and KS test statistics. $p-$values obtained from KS test are shown in brackets in the last column. Values marked in *italic* are not statistically significant at $p < 0.01$.

2.7 Automatic Inference of Indoor Ambiance Perception

In Section 2.5, reliable estimates of ambiance were obtained for several of the dimensions, suggesting the presence of visual cues that allow to create such impressions. In this section, we address RQ2.3 to examine the feasibility of automatically inferring human ambiance impressions using visual cues from images. To automatically infer human perception of outdoor spaces, recent works have used a variety of low-level image features including Color Histogram, GIST, HOG, LBP, SIFT and generic deep convolutional activation features [142, 151, 159]. For indoor places, we build upon these works and prior work in object classification and scene understanding using deep learning techniques [173, 165, 232].

For automatic inference, we used the *50K image* and *physical environment image* corpora which were described in detail in Section 2.3.2. Due to changes in the Foursquare API [2], we had API access to only 280 places (as opposed to 300 places examined so far) resulting in a total of 45,848 (out of 50,023) images for the *50K image* corpus. On the other hand, *physical environment image* contains 900 manually selected images for all 300 places. All the subsequent analysis is based on these image corpora.

2.7.1 Visual Feature Extraction and Aggregation

Building upon recent work in the literature, we extracted the following set of low-level and deep visual features as described below:

1. **Color Histogram**: We computed global color histogram in RGB space. Each channel was quantized into 8-bins, resulting in 8^3 possible color combinations and a 512-dimensional feature vector for color descriptor.
2. **GIST**: This descriptor captures the dominant spatial structure of a scene from a set of perceptual dimensions (e.g., naturalness, openness, roughness, expansion, ruggedness) [150]. We use the standard setting of this descriptor, resulting in a 512-dimensional vector.
3. **Gradient (HOG)**: Histogram of oriented gradients (HOG) computes occurrences of gradient orientations in localized region of an image [104]. We apply the pyramid HOG

implementation, where images are first represented in pyramid hierarchies, then the HOG descriptor is computed on each level, and finally the final descriptor is the concatenation of vectors across all levels [33]. We compute the pyramid HOG descriptor for levels $l = 0$ to $l = 3$. Images are divided into 2^{2*l} regions, and a 8-bin histogram is computed within each region, which results in a 680-dimensional feature vector.

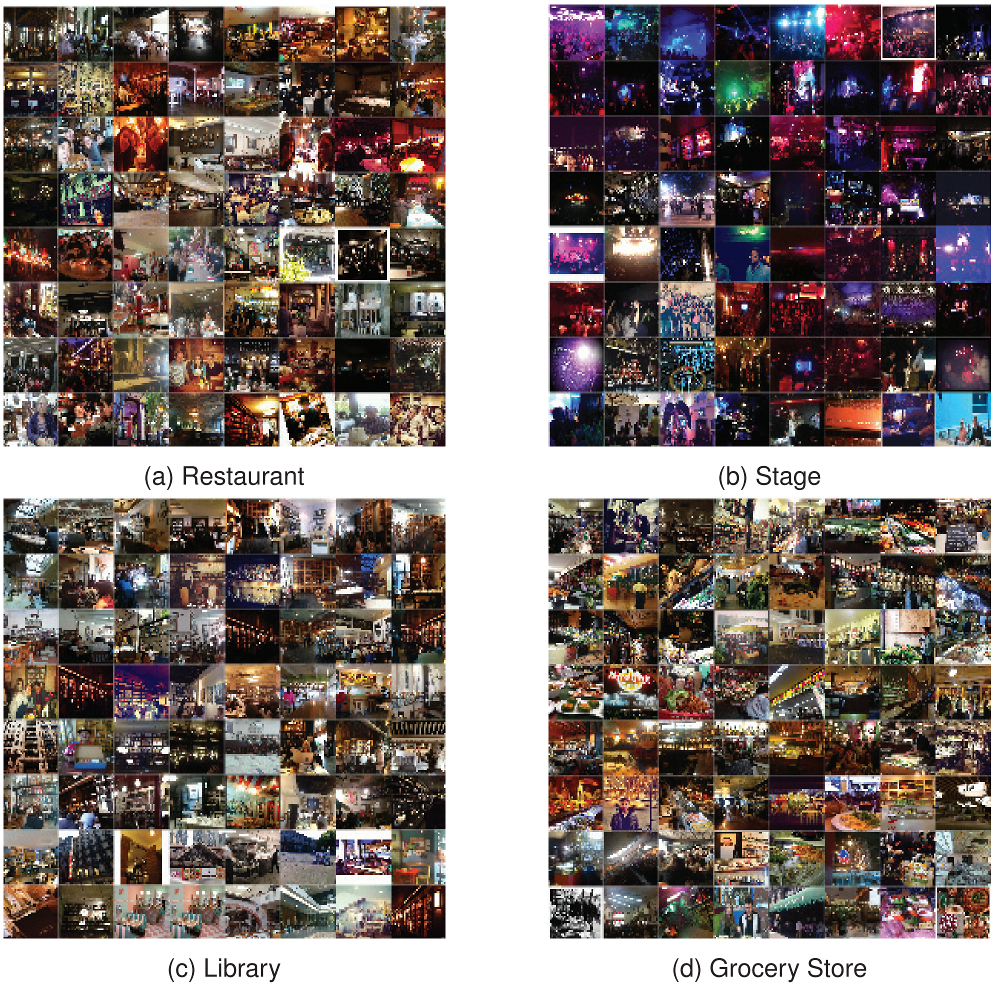
4. **Texture (LBP)**: Texture captures the spatial arrangement of color and intensities in an image. We apply the local binary pattern (LBP) descriptor [146], which encodes local texture information (such as spots, edges, and corners) by comparing each pixel with its neighborhood pixels, resulting in a 256-dimensional vector.
5. **CNN**: The availability of large-scale image datasets [173] and the performance of deep neural networks for object classification and scene understanding [165, 232], have opened opportunities to explore these features for our problem. We have used the features extracted using a pre-trained convolutional neural networks model (CNNs) using the Caffe framework [101]. Specifically, we used the GoogLeNet CNN [194] trained on *ImageNet* data. *ImageNet* data contains over 14 million images across 1,000 categories. To extract the CNN descriptors, for each image, we obtained the final layer class probabilities across all 1,000 *ImageNet* classes, resulting in a 1000-dimensional feature vector.

   We chose to use a pre-trained model trained on a large and diverse data (*ImageNet* in our case), as opposed to training a CNN model on our data from scratch for two reasons. First, it has been shown that features extracted using pre-trained CNN models can potentially provide discriminative features for multiple visual recognition tasks [62, 165]. *ImageNet* categories are well suited for our problem as they are descriptive of visual cues typically present in restaurants, bars, etc. (refer to Figure 2.3, Figure 2.7, and Figure 2.8). Second, using a pre-trained model avoids the need to train, adapt or fine-tune a CNN on our dataset, which can be computationally expensive and resource intensive.

Feature Aggregation: As previously stated in Section 2.5.2, ambiance ratings were given for each place and each place had an average of over 160 images. We extracted all the above described visual features for each image. Then, in order to obtain a representative feature vector for each place, we apply an early feature fusion approach by computing the mean feature vectors across all images describing the same place, for each feature set.

2.7.2 Inference Method and Evaluation

We formulate the inference of indoor ambiance perception as a regression problem where our objective is to predict aggregated human impressions using visual cues extracted from images. For regression, we used Random Forest which is a tree-based supervised learning method that guards against overfitting to the training data [36]. For model validation, we performed m repetitions of a k-fold stratified cross-validation approach. For all experiments, we set $m = 10$, and $k = 10$. After the model run and validation, we computed the mean of the evaluation metric across mk runs as the model output. To evaluate the performance of different feature sets, we used two standard measures: the root-mean-square error ($RMSE$) and coefficient of



(a) Restaurant

(b) Stage

(c) Library

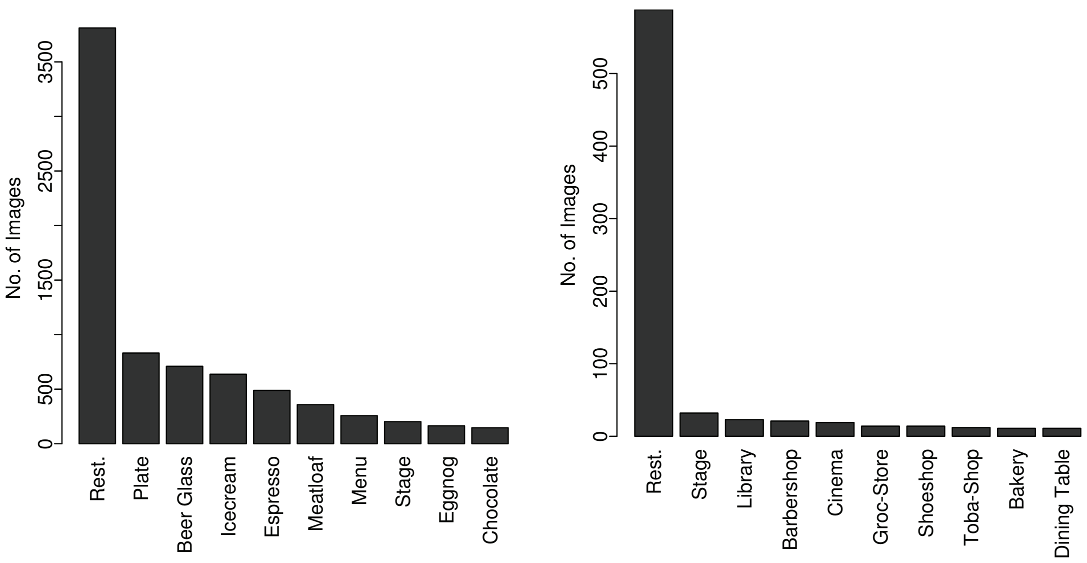
(d) Grocery Store

Figure 2.7 – Sample of random images from the *50K image* corpus which were automatically classified as (a) Restaurants/Eating Place, (b) Stage, (c) Library, and (d) Grocery Store. For each class, from top to bottom, images are sorted in decreasing order of *ImageNet* dominant class probability. Best viewed in color.

determination (R^2) between the perceived ground-truth (i.e., aggregated MTurkers judgments) and predicted ambiance scores for each label and feature set. Furthermore, we also examined the variable importance measures from random forests to understand the relative importance of specific visual cues for each dimension (see Section 2.7.4). To understand and compare the predictive performance of each feature set, we choose the baseline model to be the mean annotated score as the predicted value for each label.

2.7.3 Visual Categories

We begin our analysis by examining the distribution of the most likely *ImageNet* class assigned to each image. As stated before, the last layer of the GoogLeNet CNN model



(a) 50K Corpus

(b) Phy. Env. Image Corpus

Figure 2.8 – Histogram of Top 10 recognized *ImageNet* classes for a) *50K Image* Corpus, and b) *Physical Environment Image* Corpus.

outputs the probability distribution of the image across all 1,000 *ImageNet* classes. Given this probability distribution, we chose the *ImageNet* class with the highest probability as the dominant class for each image. In Figure 2.8, we show the distribution of the top-10 dominant classes for both image corpora. For the *50K image* corpus, most of the top ten dominant categories are associated with either food (e.g., plate, meatloaf, ice-cream, chocolate) or drinks (e.g., beer glass, espresso, eggnog), as shown in Figure 2.8a. These results corroborate previous findings reported in Section 2.4.2, where we found that randomly sampled images from Foursquare showed food-related items in close to 50% of cases (Figure 2.4c). Furthermore, these findings are expected given that all selected places in the study are either restaurants, bars, cafes, or nightclubs (Section 2.3.1). To elucidate the recognized classes visually, Figure 2.7 shows a sample of images selected randomly across four of the top-10 *ImageNet* categories from the *50K image* corpus.

While analyzing the top-10 class distribution for the *physical environment image* corpus, we observe that most of the dominant classes relate to the physical attributes of the indoor environment (e.g., stage, library, dining table, bakery, grocery store) and do not contain food or drinks categories, which is in contrast with the class distribution for the *50K image* corpus (Figure 2.8b). These findings are not surprising given that all images in this corpus were manually selected to show clear views of the indoor scene (Section 2.3.2). Some of the automatically recognized categories in *physical environment image* corpus may seem intriguing at first glance (e.g., library, grocery store, or barbershop), but after manually browsing the images belonging to these categories, we found that these classes describe various attributes of the indoor environment and are misclassified yet makes sense visually.

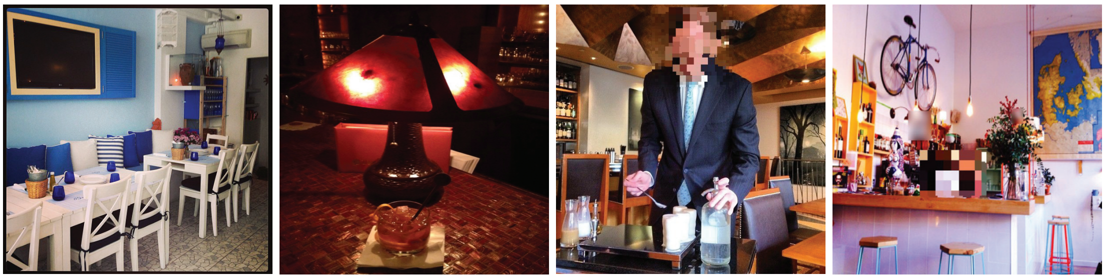
	Baseline 50K	**Color 50K**	**GIST 50K**	**HOG 50K**	**LBP 50K**	**LLC 50K**	**CNN-Phy-Env**		**CNN-50K**	
	$RMSE$	R^2	R^2	R^2	R^2	R^2	R^2	$RMSE$	R^2	$RMSE$
Artsy	0.69	0.01	0.02	0.04	0.05	0.05	0.12	0.66	**0.22**	0.63
Bohemian	0.55	0.08	0.05	0.11	0.09	0.14	0.08	0.54	**0.24**	0.50
Conservative	0.67	0.21	0.20	0.19	0.11	0.25	0.24	0.60	**0.30**	0.57
Creepy	0.29	0.05	0.04	0.01	0.00	0.04	0.06	0.29	**0.14**	0.28
Dingy	0.50	0.04	0.02	0.01	0.05	0.05	0.05	0.50	**0.17**	0.47
Formal	0.82	0.10	0.07	0.03	0.10	0.14	0.28	0.72	**0.37**	0.70
Loud	0.73	0.33	0.29	0.26	0.31	0.42	**0.53**	0.51	0.52	0.51
Off-beaten	0.61	0.05	0.01	0.01	0.00	0.01	0.15	0.47	**0.17**	0.47
Old-fashioned	0.50	0.16	0.11	0.10	0.08	0.16	**0.24**	0.54	0.22	0.55
Romantic	0.67	0.10	0.15	0.03	0.08	0.17	0.36	0.57	**0.39**	0.56
Sophisticated	0.79	0.11	0.10	0.04	0.10	0.17	0.26	0.72	**0.38**	0.67
Trendy	0.64	0.19	0.1	0.12	0.15	0.21	0.17	0.61	**0.32**	0.54
Up-scale	0.78	0.14	0.11	0.03	0.13	0.19	0.29	0.69	**0.40**	0.65

Table 2.6 – Inference results for 13 ambiance dimensions for all feature sets, using R^2 and $RMSE$ as evaluation measures. Cells marked in **bold** correspond to the best R^2 result obtained for each dimension across all feature sets. LLC-50K refers to the model with all low-level image features combined.

For instance, most of the images belonging to “library” class contain images showing wall shelves typically found in cafes and bars (Figure 2.7c); while some of the “grocery store” images contain transparent window shelves displaying food or drink items (Figure 2.7d).

Further, we observe that the top class in both image corpora is “restaurant/eating place”. For the *50K* and *physical environment* image corpora, this class respectively contains 8% and 65% of total images. At first, it looks like a generic class but after manually browsing these images (Figure 2.7a), we found that most of the images represent the physical environment well, providing clear and unoccluded views of the indoor space. These findings point towards the feasibility to automate the selection of images for indoor ambiance characterization, and avoiding the need to manually select images, as was carried out for the *physical environment image* corpus (Section 2.4). We previously reported that human observers preferred views of the physical environment of places to make impressions of ambiance (Figure 2.4 described in Section 2.4.2). However, as we show in the following subsection, other image categories are also exploited by the learning algorithm.

While the plots shown in Figure 2.8 are focused on illustrating the dominant visual class, many other classes are also informative of ambiance for a given image (e.g., a photo of someone eating an ice cream in a restaurant or drinking an espresso in a coffee shop are not likely to be generated in a club). This is the reason why we use the 1000-dimensional probability vector to infer the ambiance dimensions. Overall, these findings point towards the nature of images shared on Foursquare, which to our knowledge has not been analyzed before at such a scale from the perspective of ambiance.



(a) Dining Table (b) Table Lamp (c) Suits (d) Altar

Figure 2.9 – Sample of images from the *50K image* corpus which are recognized to belong to the visual category of a) Dining Table, b) Table Lamp, c) Suits of clothes, and d) Altar. Best viewed in color.

2.7.4 Predicting Ambiance Impressions

In this section, we evaluate the performance of both low-level image and deep CNN features to automatically infer human impressions of place ambiance. Table 2.6 reports the R^2 values between perceived ground-truth and inferred ambiance scores for each label and feature set. Note that Table 2.6 reports the results across all cities combined. As previously stated in Section 2.7.2, we assigned the mean annotated scores as the predicted value to be the baseline model for each label. For the baseline model, we only report the $RMSE$ values as it has a R^2 of 0. For the model learned using CNN features, we report both the R^2 and $RMSE$ values. All the low-level visual features are computed on the *50K image* corpus, while the CNN features are computed for both image corpora. A model having a better feature representation to estimate ambiance impressions would result in higher R^2 and lower $RMSE$ values, when compared to other models.

Overall, the results indicate that a maximum R^2 of 0.53 can be obtained by the CNN-based regressor (for the *loud* dimension), while low R^2 values are obtained for other dimensions (*creepy*, *dingy*, *off the beaten path*). For seven out of 13 dimensions, the obtained R^2 exceeds 0.30. On the *50K image* corpus, CNN features consistently outperform individual and combinations of several low-level image features (including Color Histogram, GIST, HOG, and LBP) for all labels, which is consistent with the results reported in the vision and multimedia literature [165]. The results highlight the suitability of pre-trained CNN models to infer perceptual judgments in indoor places; our finding complements the works in recent literature adopting deep learning to infer high-level perception of outdoor scenes [151, 159].

The performance of low-level features indicates that all feature sets perform the best for the *loud* feature (Table 2.6). Amongst the low-level features, color features achieve the highest R^2 for most of the ambiance dimensions, including *trendy* and *up-scale* places, which corroborates results reported in [166], that attempted to infer ambiance for a 50-place dataset from profile pictures of Foursquare users rather than using photos of the venues. HOG and GIST also perform moderately for the *conservative* label.

While examining individual labels, the *loud* label achieves the highest performance with $R^2 \geq 0.26$ for all feature sets, suggesting visual patterns such as texture cues associated with the perception of loudness e.g., presence of crowd, elevate stage, etc. (Figure 2.7b). On the other hand, *creepy* achieves the lowest predictive performance ($R^2 \leq 0.14$). The low predictive power for *creepy* can be attributed to a combination of lower inter-rater agreement and lowest mean scores out of all labels (Table 2.3 and Table 2.4). Overall, positively phrased (*romantic*, *up-scale*, *sophisticated*) and negatively phrased labels (*creepy*, *dingy*), which likely correspond to different ambiances, achieve in each case similar R^2 values.

Moreover, the *romantic* label achieves promising prediction performance ($R^2 = 0.39$) with CNN features extracted using the *50K image* corpus, and relatively poor performance with low-level image features. We further analyze the relative importance of visual categories for different labels using the variable importance measures taken from random forests [36]. Using these measures, we find that "dining table" and "table lamp" are the top two discriminative visual categories for *romantic* places (see Figure 2.9a and 2.9b); "suit of clothes" and "red wine" for places perceived as *upscale* and *formal* (see Figure 2.9c); and "altar" for *artsy* places (see Figure 2.9d).

Comparing CNN-Phy-Env and CNN-50K: While evaluating the performance of deep features between the *50K* and *physical environment* image corpora, we observe that the CNN-50K model outperformed the CNN-Phy-Env model in terms of higher R^2 and lower $RMSE$ values for most of the dimensions except the *loud* and *old-fashioned* dimensions. We found that for some labels (e.g., *loud*, *romantic*, *old-fashioned*, *off the beaten path*), R^2 values are comparable, while for some dimensions (e.g., *artsy*, *bohemian*, *trendy*), the difference between R^2 values across image corpora is relatively high. Note that the ambiance ratings were obtained using just three manually selected images (Section 2.5.2). These findings suggest that it is difficult to automatically represent some dimensions with just three images (compared to the human ability to infer and generalize from very few examples), and that the availability of more images provides additional informative visual cues to characterize some of these labels at high level.

To further investigate the effect of the amount and type of images on the predictive performance, we conducted a series of experiments. How can we train machines to identify image types which can describe the atmosphere of indoor places? From the *50k image* corpus, we selected all images belonging to the top k dominant classes of the *physical environment image* corpus. For the experimental setup, we varied the parameter $k \in \{1, 10, 20, 50\}$ e.g., for $k = 10$, we first computed the top 10 dominant visual classes of *physical environment image* corpus and then selected all images from the *50K images* corpus which were classified as belonging to either of these top 10 categories. To detect the dominant visual class for each image, we followed the methodology outlined in Section 2.7.3. After image selection, we ran the regression experiments for each k using the CNN-based features. Table 2.7 reports the R^2 values and the number of images (N) selected for each k. Note that due to the long-tail characteristics of the dominant class distribution (Figure 2.8), the number of images (N) do

	Top-1 N=3,811	**Top-10** N=7,605	**Top-20** N=9,029	**Top-50** N=11,306	**All** N=45,848
Artsy	0.11 (0.09)	0.18 (0.11)	0.17 (0.14)	0.20 (0.12)	0.22
Bohemian	0.12 (0.08)	0.18 (0.08)	0.21 (0.16)	0.23 (0.11)	0.24
Conservative	0.24 (0.20)	0.31 (0.23)	0.29 (0.21)	0.30 (0.26)	0.30
Creepy	0.07 (0.06)	0.06 (0.08)	0.11 (0.13)	0.10 (0.08)	0.14
Dingy	0.09 (0.06)	0.11 (0.10)	0.11 (0.09)	0.15 (0.10)	0.17
Formal	0.25 (0.16)	0.30 (0.21)	0.32 (0.29)	0.27 (0.30)	0.37
Loud	0.38 (0.32)	0.54 (0.37)	0.56 (0.43)	0.56 (0.37)	0.52
Off-beaten	0.10 (0.05)	0.12 (0.07)	0.16 (0.12)	0.18 (0.08)	0.17
Old-fashioned	0.17 (0.16)	0.24 (0.15)	0.26 (0.18)	0.23 (0.18)	0.22
Romantic	0.27 (0.16)	0.37 (0.21)	0.38 (0.33)	0.37 (0.27)	0.39
Sophisticated	0.25 (0.12)	0.32 (0.20)	0.33 (0.32)	0.32 (0.30)	0.38
Trendy	0.13 (0.16)	0.27 (0.22)	0.29 (0.26)	0.29 (0.23)	0.32
Upscale	0.27 (0.15)	0.32 (0.22)	0.33 (0.32)	0.31 (0.29)	0.40

Table 2.7 – Inference results obtained when varying the amount and type of images of the *50k image* corpus for all studied dimensions.

not increase linearly with increasing k. For comparison, the last column of Table 2.7 lists the results for the entire *50K image* corpus, which is identical to the one reported in Table 2.6. Using these statistics, we can make the following observations:

- Overall, as k increases, R^2 also increases for most of the labels, which is expected due to the availability of more images (N). We report the results for four different values of k but similar trends were observed for other values.
- When comparing the performance of different k, comparable results were obtained for $k = 50$ and $k = 20$, and competitive results for $k = 10$ relative to the model trained on the entire corpus. With $k = 1$, moderate results are obtained for *conservative*, *loud*, and *old-fashioned*. Note that $k = 1$ corresponds to “restaurant/eating place” class, which was the dominant class for both image corpora (see Figure 2.8 and Figure 2.7a). Most of the images in this category represent the physical environment well, which might likely suggest the competitive performance for some of the ambiance labels.
- To diagnose the role of image types, Table 2.7 also shows the R^2 values (in brackets) when the regression experiment was run with a random selection of images not belonging to the the top-k classes for each k. In other words, N was held constant but random image selection was performed. For all k, we consistently observe higher R^2 using images from top k classes relative to random image selection.

In summary, we have presented results on the effect of the amount of data and image types used for learning ambiance. These findings potentially point towards the feasibility to automate the selection of images for indoor ambiance characterization, and avoiding the need to manually select images, as was carried out for the *physical environment image* corpus (Section 2.4).

Overall, the regression results suggest the feasibility to automatically infer place ambiance with promising prediction performance for some of the dimensions. To the best of our knowledge, our work reports the first results on how images collected from popular places on Foursquare can be used to automatically infer indoor ambiance impressions from deep features. Most of the research on prediction of perceptual dimensions have focused on outdoor places, so at present we can not compare our results with prior work in the literature.

2.8 Conclusion

In this chapter, we presented a study on perception and inference of psychological dimensions of indoor places from social media content under the construct of ambiance. Using about 50,000 Foursquare images collected from 300 popular indoor places across six cities, we first assessed the suitability of social media images as data source to convey place ambiance, and found that images with clear views of the environment were perceived as being more informative of ambiance than other image categories. Second, we demonstrated that reliable estimates of ambiance can be obtained for several dimensions using Foursquare images, suggesting the presence of visual cues to form place impressions. We further found that most aggregated impressions of places were similar across cities. Third, we demonstrated the feasibility to automatically infer ambiance impressions using visual cues extracted from images. Using pre-trained convolutional neural network models, we extracted visual features and obtained a maximum R^2 of 0.53 and lower root-mean-squared errors for all dimensions relative to the baseline method.

The applications derived from this research are manifold. On the user side, this could include hyper-local, ambiance-driven place search and discovery in online platforms, where users could search for places by its ambiance e.g., a *formal* place for a family dinner, a *romantic* place for a date, or a *trendy* place for a night out. This could complement existing sources of information like place reviews and user comments. As a second application domain, on the side of venue owners, this could include data-driven tools to deepen the understanding of the impressions that their venues evoke in potential and real customers, and recommendations about improving the appearance and architecture style of their venues.

3 Characterizing Urban Perception in Outdoor Places

3.1 Introduction

COMMUNITY awareness and action on urban problems are long-standing practices in developing countries [67]. The ability to reflect and act upon concerns defined by a community's interests and values for its own benefit takes on special relevance in Latin America due to the local governments' inability to realize the full potential of both human and economic resources and a historically slow (when not absent) response by the authorities. Civic engagement and action with the local environment have educational, social, and economic aspects [41].

In this context, mobile and social technologies are providing new opportunities to document, characterize, map, and ultimately address urban problems in developing cities. Mobile crowdsensing efforts for urban mapping and surveying conducted by citizens equipped with mobile phones (who generate reports, take pictures, or create maps) are emerging, often concentrated in informal settlements and other problematic regions [171, 7, 14, 178]. Other recent approaches are studying cities in the developed world, and use custom-built crowdsourcing platforms to establish the feasibility of obtaining reliable estimates of urban impressions of physical and psychological constructs like safety, beauty, and quietness, elicited by images of the city taken at street level [176]. The possibility of obtaining crowdsourced perceptions of the socio-urban image of a city by non-residents is valuable for developing cities, specially when the cities have large flows of visitors (tourists, students or business people), as it could help to understand the choices that non-locals make regarding the use of the public space, or to identify misperceptions due to the lack of local context. In this chapter, we present a study on computationally characterizing urban perception of outdoor environments of three cities in central Mexico.

In research examining urban impressions in developed cities using online crowdsourcing, judgments have been elicited using images obtained from Google Street View (GSV) [176, 162, 87]. Even though GSV provides a scalable and automated way to collect images, it suffers from two limitations. First, the GSV image database is not exhaustive in spatial coverage in developing countries due to accessibility and safety issues. For instance, due to the way Google collects street views (via cameras mounted on top of a vehicle), GSV does not always contains images of narrow streets and winding alleys. For instance, we found in a small sample of images taken in this type of area that more than half of GSV images were either unavailable or erroneous within a mid-size touristic city in Mexico [171]. Second, GSV images fail to capture the temporal aspects of a city: only static and daytime views are available, and it can take years before images are updated. This does not facilitate studying the effect of time of the day in the perception of the urban environment, which is a key aspect as discussed in urban studies literature [155, 128]. In contrast, mobile crowdsensing represents an opportunistic, just-in-time way of documenting urban changes over time.

Outdoor spaces contain architectural elements that are visually and perceptually different relative to indoor places. Consequently, the study of outdoor places involves the collection of visual data which can describe these elements in the built environment. In this chapter, we present the *SenseCityVity* project which integrated: (1) mobile crowdsensing involving a local youth community to collect first-person perspective images depicting urban issues that were defined by the community itself, and (2) online crowdsourcing using MTurk, where US crowd-workers contributed their impressions on photos of the urban environment along twelve physical and psychological dimensions. The *SenseCityVity* project was a joint collaboration with Prof. Salvador Ruiz-Correa (IPICYT, Mexico).

We address the following research questions in this chapter:

RQ3.1: How can a population of young people be engaged through mobile crowdsensing to collect images of urban environments? (**RQ3.1** maps to thesis's **RQ1**)

RQ3.2: Can human observers reliably assess the urban perception of outdoor environments using images? If so, what physical and psychological dimensions of urban perception can be reliably assessed? (**RQ3.2** maps to thesis's **RQ2**)

RQ3.3: Can crowdsourced judgments of urban perception be automatically inferred using low-level image and deep learning features extracted from images? (**RQ3.3** maps to thesis's **RQ3**)

In this chapter, we make the following contributions:

1. A participatory mobile crowdsensing framework involving over 70 local students that resulted in a data set of 7,000 geo-localized images collected in three cities in Guanajuato state (Guanajuato, Leon, and Silao), each one characterized by its distinct geography, economic activity, and population. Guanajuato is a touristic city, Leon is a business and industrial hub, and Silao is an agricultural town. The images contain outdoor scenes and views of each city's built environment, including touristic, historical, and business sites, residential neighborhoods, and areas with narrow streets and alleys.

2. An online crowdsourcing study on MTurk to gather impressions of crowd-workers along 12 physical and psychological labels including (*dangerous*, *dirty*, *interesting*, *happy*, *polluted*, etc.), based on 1,200 images (400 images per city). The studied dimensions include and extend those studied in recent literature. Statistical analyses on 144,000 individual judgments show that the outdoor scenes in the investigated cities can be reliably assessed by observers of crowdsourced images with respect to most urban dimensions.
3. A cross-city statistical analysis shows that outdoor urban places in Guanajuato city are perceived as more *quiet*, *picturesque* and *interesting* compared to places in Leon and Silao. In contrast, Silao is perceived to have higher accessibility than Guanajuato, but lower accessibility than Leon. Overall, Guanajuato has the highest mean scores amongst all three cities on most positive labels and lowest scores on negative labels. This finding is relevant as it could inform both citizens and authorities about significantly different perceptions that could lead to action.
4. As a way of showing the additional value of mobile crowdsensing with respect to GSV-type surveying, we present a study to compare the perceptions of urban environments across different times of the day using a small sample of 100 images and 12,000 individual MTurk judgments. The results show that places at evenings were perceived as less *happy*, *pleasant* and *preserved*, in comparison with mornings. We did not observe statistically significant differences in perception of *dangerous* between mornings and evenings for the studied places.
5. Using a similar methodology that was adopted for indoor places, we automatically inferred human perceptions of outdoor places using a variety of low-level image, and generic deep learning features. CNN features consistently outperformed all the individual low-level image features for the 12 studied dimensions. We obtained a maximum R^2 of 0.49 using CNN features; for 9 out of 12 labels, the obtained R^2 values exceeded 0.44.

In this chapter, we have focused on the empirical analysis and automatic characterization of urban perception for outdoor places. Going beyond just the collection and analysis of images, the *SenseCityVity* project addresses the need for local communities to become more aware of the realities of their urban environment and take collective action towards addressing some of the urban civic issues in their respective local communities. In addition to geo-localized images, the project also resulted in a collection of videos of urban scenes and video-recorded interviews of locals which was subsequently used for community reflection and artistic creation as described here [172].

The chapter is organized as follows. We begin with a review of related work in Section 3.2. Section 3.3 describes the data collection framework, including the criteria to select the studied cities in Mexico, the definition and selection of urban awareness dimensions, and the mobile crowdsensing design to collect two image corpora. The development of the mobile phone application, recruitment of participants and collection of raw image data was carried out by a team led by Salvador Ruiz-Correa. Section 3.4 outlines the design of the online crowdsourcing experiment on MTurk to gather impressions of crowdworkers along 12 labels. Using the

MTurk annotations, Section 3.5 presents the descriptive statistics, provides statistical results on cross-city comparison, and study differences in urban perceptions across different times of the day. Section 3.6 proposes a machine learning methodology to automatically infer urban perception of outdoor environments using low-level image and generic deep learning features. Finally, Section 3.7 concludes with a summary of our findings, and the potential impact of our work. This chapter some of the research that has been published here [186, 172].

3.2 Related Work

In this section, we begin by reviewing the related work on existing systems for reporting civic issues. Next, we discuss studies describing the crowdsourced and automatic perception of outdoor places in the field of architecture, urban computing and computer vision. Finally, we review research that relates to the temporal effects on urban perception.

3.2.1 Systems for Reporting Civic Issues

There are various existing systems that allow citizens to report urban issues. One in the developed world is FixMyStreet (FMS) [16], launched in the UK in 2007 [112], and later implemented in other countries (mainly in Europe) with varying degrees of success. FMS allows people to share and map text reports about problems; the system allows uploading images as an optional feature. SeeClickFix [17], was launched in the US in 2008. These systems have not generally been adopted in Latin America, among other reasons, because they require an authority committed to take ownership for the system and respond timely to the reports. A recent analysis of six years of FMS reports concluded that only 11% of them contain images, but also that image uploading was a significant indicator of the actual response of the authorities to the reports and the commitment of reporters to keep contributing [190]. These findings support our choice of mediating participation via geo-localized photo taking. On the other hand, in contrast to these systems, which by design promote individual participation [29], our work is community-oriented and puts community interests and action at the center.

In this sense, our work is closer to a number of open mapping initiatives in developing regions. Notable examples include the Kibera slum in Kenya [7], and various systems built around Ushahidi [18]. In Latin America, other examples include the work done to map informal settlements in Buenos Aires, Argentina [14], and in Rio de Janeiro, Brazil [222]. Another mapping effort is led in India by Humara Bachpan [178], an organization that conducts civic campaigns centered on “child clubs” to create maps of marginalized neighborhoods in India. Two differences between these initiatives and our work are (1) the engagement of communities of youth in both data collection and data appropriation exercises; and (2) the development of a methodology to produce crowdsourced assessments of the conditions of photographed urban places.

Recently, social media channels are being used to generate reports of urban-related concerns, sometimes containing photos. In Mexico, Twitter has been notably used for real-time, eyewitness reports of insecurity and drug-related crimes in towns and cities [139]. This is an attractive alternative, but it is limited in terms of representativeness and spatio-temporal coverage [206].

3.2.2 Crowdsourced Perception of Outdoor Places

In the field of architecture and urban planning, many of the studies about visual perceptions of built environments have used qualitative research methods including interviews, visual preference surveys [179, 155] and observation of the built environment using either actual or simulated images [126]. However, most of these studies are either conducted in controlled laboratory settings or are based on questionnaires, which may have limitations with respect to scalability, ecological validity, or recall biases.

With the popularity of social media and mobile phones, together with an increased use of online crowdsourcing platforms to obtain judgments from diverse populations, scholars have started to explore crowdsourcing as a medium to obtain estimates of urban perception for both indoor [83] (including our work on indoor places in Chapter 2) and outdoor environments [176, 171, 162]. For outdoor environments, gathering perceptions typically involve the use of Google Street View (GSV) [176, 162, 23]; while GSV is widely available in the developed world, it is not so for the developing world [171]. In [176], the authors conducted a study to measure the perception of outdoor urban scenes on *safety*, *class* and *uniqueness*, based on geo-tagged images obtained via GSV in four developed cities (in the US and Austria). In a similar study on urban perception, judgments were collected to examine visual cues that could correlate outdoor places in London with three dimensions (*beauty*, *quietness*, and *happiness*) [162].

3.2.3 Automatic Inference of Outdoor Perception

To automatically infer human perception of places, most of the recent work have focused on outdoor places. Recent works have used a variety of low-level image features including Color, GIST, HOG, LBP, SIFT and more recently, generic deep convolutional activation features. Using these features on geo-tagged images from GSV, high-level attributes for outdoor scenes are inferred in two US cities [142, 151]. Using the same dataset, in [159], a CNN architecture is proposed to predict and discover mid-level visual patterns which correlate with the perceived safety of an outdoor scene. Using images from Google Street View, in [61], authors proposed a discriminative clustering methodology to identify visual elements (e.g. windows, balconies, and street signs) unique to the cities of Paris and London (HOG and Color features). Building upon this work, authors in [22], proposed a scalable visual processing framework to identify the relationships between the visual appearance of a city and its non-visual attributes (e.g. crime statistics, housing prices, etc.). In another study to identify attributes specific to a city [233], authors conducted an analysis of 2 million geo-tagged Panoramio images from 21 cities to discover salient visual features of outdoor scenes.

3.2.4 Temporal Effects on Urban Perception

The works described in the previous subsections use temporally fixed image stimuli to obtain perceptions. However, cities are dynamic and the time of day might play an important role with respect to urban perception. In the field of urban planning, there has been significant interest to quantify perceived safety and fear of crime during night-time [155, 128, 86]. Using on-street pedestrian surveys, the authors found that 90% of the respondents felt that the improvements in street lighting lead to a reduction in the perceived fear of crime, an increased sense of personal safety, and increased pedestrian use after dark [155]. No image stimuli were used in this study. Another study compared 16 images of outdoor scenes taken during day-time and night-time to test the effect of visibility on making a place feel safe [128]. Respondents were undergraduate psychology students who were asked to rate these scenes. The authors found statistically significant differences on ratings on perceived safety during different times of the day. A similar study was conducted using 20 photographs of night-time locations from a university campus [86]. In our work, we study not only the role of night-time or dark scenes on perceived safety, but also other psychological dimensions e.g., *quietness*, *happy*, etc.

In contrast with the previous works, we followed a mobile crowdsensing approach to collect images of outdoor scenes and document them statically and dynamically over time. In this chapter, we define and study a large number of urban constructs, and examine how perceptions of urban environments vary across different times of the day. In this chapter, we have focused on the related work on outdoor environments. To review the literature on the visual perception of indoor places, please refer to Section 2.2.3 of Chapter 2.

3.3 Data Collection Framework

In this section, we describe the *SenseCityVity* data collection framework, including the criteria to select the studied cities in Mexico, the definition and selection of perceptual dimensions, and the mobile crowdsensing design to collect two image corpora.

3.3.1 SenseCityVity Project Design

The image corpus used in our study was collected with an approach aimed at exploring urban environments of cities in Latin America, with an initial emphasis in Mexico. The approach was developed in the context of a larger research initiative, *SenseCityVity*, which aims at addressing specific urban issues by young volunteers through the use of collective action, and mobile crowdsensing. *SenseCityVity* emphasizes that empowering citizens through technological means that increase awareness and deepen the understanding of socio-urban concerns is of crucial importance. This is so because the state and evolution of their cities strongly depend on the capacity of their populations and the existence of institutional policies to create the structural conditions for sustainable development. By development, we understand the “process by which people individually and collectively enhance their capacities to improve their lives according to their values and interests” [41].

SenseCityVity followed a transdisciplinary approach to explore the urban environment involving computer scientists and other experts on one side, and social actors on the other. Our team included specialists in computer science, social media, psychology, and visual arts, who conducted the technical and social design, as well as the development and execution of experiments on mobile sensing and crowdsourcing jointly with student volunteers, who were recruited from a local technical high school. Participating students were altruistically motivated and eager to contribute their knowledge and experiences, and to co-design all project activities to better understand the urban environment of their city.

3.3.2 Selection of Cities

To collect images of outdoor urban spaces, we grounded our work in three small to mid-size cities in central Mexico: Guanajuato (pop. 170,000), Leon (pop. 1.5 million) and Silao (pop. 147,000). Guanajuato is a touristic city in central Mexico, and the capital of a state of the same name. Guanajuato occupies a valley, forming a complex network of narrow roads, pedestrian alleys, and stairways running uphill. Most pedestrian alleys have no car access, and other major roads run underground. Guanajuato is a historical city and a UNESCO world heritage site, with a vibrant tourism industry that is centred around the city’s historical downtown area (dating from the Spanish colonial times) and several large art festivals. Guanajuato city often appears as one of top destination to visit in Mexico [168, 47].

The city of Leon is a business and industrial hub in the state of Guanajuato that drives a large part of the economical activity of the state. Leon has a strong leather industry, offering products both to the national and international markets. Leon also receives a large number of tourists. In contrast, Silao is a local hub of agricultural and industrial activity in the region, with a wide variety of farm crops, dairy packaging plants and a major car assembly plant. Due to its relatively larger size, some areas in Leon are quite inaccessible either due to safety concerns or because of the presence of large walls which typically surround up-scale neighborhoods.

The three cities reflect a common situation in Latin American urbanization, which produces complex environments with historical sites, suburban sprawl, affluent neighborhoods, and informal settlements. For the three cities, images were captured from areas that included different neighborhoods reflecting the characteristics of each city, as well as touristic and historical sites. Figure 3.1 shows a sample of images from each city.

3.3.3 Definition and Selection of Dimensions

To select labels to characterize urban awareness for outdoor environments, we base our methodology on prior work [176, 171, 162]. The list of selected labels in our study encompasses the labels studied in the literature, in addition to new ones. We have chosen the following 12 dimensions in alphabetical order: *accessible*, *dangerous*, *dirty*, *happy*, *interesting*, *pleasant*, *picturesque*, *polluted*, *preserved*, *pretty*, *quiet* and *wealthy*. Three labels (*dangerous*, *dirty* and *polluted*) have a negative connotation, while the rest have a neutral or a positive connotation. Throughout this chapter, we use the umbrella term "urban awareness" to refer to these labels. As was the case with indoor places, images served as stimuli to rate perceptions for 12 urban awareness labels. Impressions were elicited along a seven-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (7), as typically done in psychology and urban planning research [83, 128].

We had chosen this list of labels for several reasons. First, these labels encompass physical and psychological constructs evoked while describing socio-urban characteristics of the built environment. Second, all the three cities face various problems including crime, prevalence of alcohol and drugs, and streets with garbage and non-artistic graffiti, to mention a few. These issues not only affect the well-being and safety of its citizens, but also hurt the image of a city e.g., as a tourist destination. Thus, it was essential to study and understand the role these perceptions play in these cities.

Note that the list of dimensions studied for outdoor spaces were different from the ones chosen for indoor places. From the perspective of urban design, studying elicited impressions for outdoor spaces should involve examination of different variables when compared to indoor places. For instance, "*picturesque*" is a meaningful perception construct for an outdoor place, but not necessarily so for an indoor place.

3.3.4 Urban Data Challenge

The images used in this study were collected as part of an Urban Data Challenge (UDC). The UDC was co-designed with the help of student volunteers from a technical school in Guanajuato city. The technical school was founded to provide high-quality education on science, technology and humanities to working class youth (about 600 students in the age group of 16–18 years old) who live in Guanajuato and surrounding suburbs. The UDC was carried out by the team in Mexico led by Salvador Ruiz-Correa.

The data challenge was carried out during a 12-week period and consisted of weekend camps, workshops, data collection, and creative use of collected data with follow-up activities. Student volunteers were organized into ten teams of ten members, each consisting of seven students, two parents and a teacher. Teams sought support from their classmates to achieve their goals, with the objective of involving a larger community in the data collection. Workshop activities included discussions about ethics, privacy, urbanism, basic techniques of

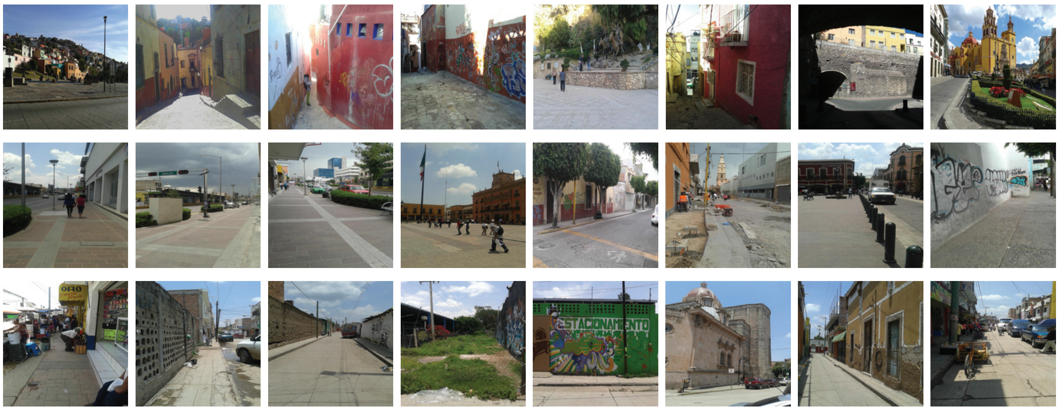


Figure 3.1 – Random selection of images from the *city image* corpus. Top row contains all images from Guanajuato city, middle row describes all images from Leon city, and bottom row shows images from Silao city. For privacy reasons, images showing faces have been pixelated. Best viewed in color.

photography, and the use of mobile devices for participatory sensing in urban environments. To cover Leon and Silao, student teams visited these cities in person, which are 56km and 25km respectively from Guanajuato.

Teams explored each city to document their urban concerns by photographing urban places via mobile phones. Each team was given an Android-based smartphone. However, students also used their own mobile devices for data collection. We developed a mobile application that enabled students to take pictures and upload them to our image server. The collected images covered not only urban concerns, but also captured the ebb and flow of the city highlighting different facets of urban life. Mechanisms to incentivize participation included creation of study circles to raise awareness about the importance of understanding urban phenomena through the use of mobile technology, and the role of citizens in proposing creative, community-based solutions to prevalent urban problems. The UDC produced over 7,000 geo-tagged images. All the images were taken from a first-person perspective, corresponding to the natural situation in which a person navigates and perceives the urban environment. Twelve hundred images were then selected for the crowdsourcing experiments reported in this chapter.

3.3.5 Image Corpora

City Image Corpus: UDC resulted in an image corpus consisting of 7,000 geo-tagged images. For the analysis reported in this chapter, we focused on a random selection of 1,200 images with 400 images per city, which we call the *city image* corpus. All images were taken between 9 AM and 5 PM during workdays. The collected image set consists of outdoor images captured at touristic hotspots, key historical sites, traditional neighborhoods, main squares, thoroughfares, main/commercial streets and downtown areas. Volunteers were asked to capture images of urban scenes in their natural setting and to avoid beautified

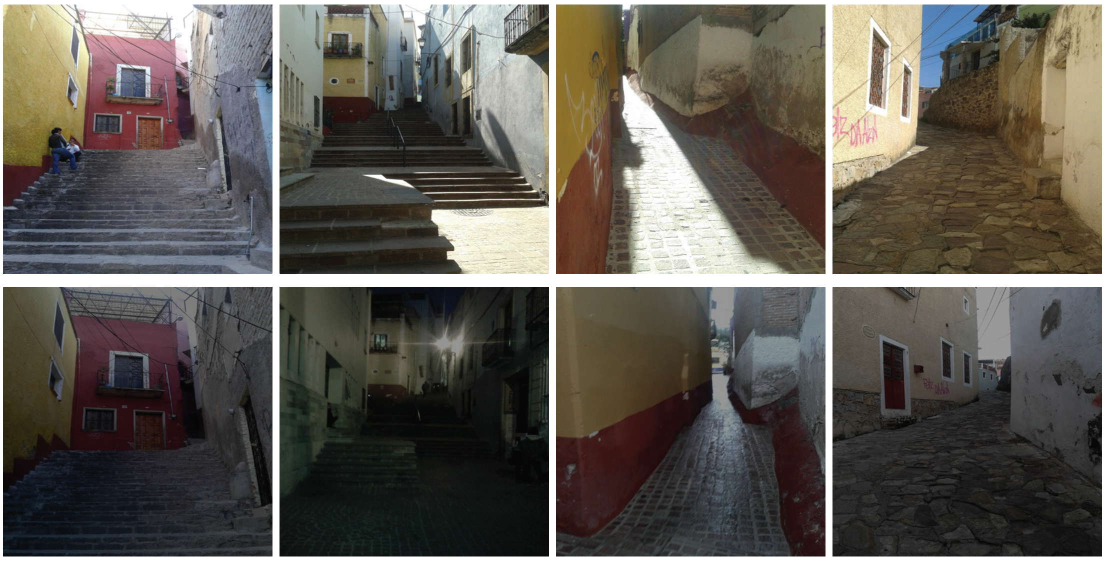


Figure 3.2 – Random selection of images from the *evening image* corpus. Top row indicates images of outdoor scenes taken during the morning, and bottom row indicates the same places photographed during the evening. For privacy reasons, images showing faces have been pixelated. Best viewed in color.

images or applying digital filters, as is usually the case with images found in social media platforms, like Flickr or Instagram. It is important to note that the *city image* corpus contains not only those images that document an urban concern, but also images which capture the ebb and flow of the city while depicting different aspects of urban life and built environments. Figure 3.1 shows a sample of images from this corpus for each city.

Evening Image Corpus: In addition, people participating in the UDC also collected images of urban areas during different times of the day in order to test if perceptions of the urban environments vary across different times of the day. This illustrates the benefits of just-in-time crowdsourced photo taking compared to static approaches, like Google Street View. We focused our analysis on 50 urban sites in Guanajuato city, where volunteers captured images of the same place during two different times of the day: first during the morning (between 10-11 AM), and then in the evening (between 6-7:30 PM). As a result, for the *evening image* corpus, we obtained 50 images per time slot, resulting in a total of 100 images. Figure 3.2 shows two images of an urban place taken during morning and evening respectively.

3.4 Crowdsourcing Impressions

To gather impressions of online annotators, we designed a crowdsourcing study on Mechanical Turk (MTurk). We chose US-based workers with at least 95% approval rate for historical HITs. To increase the reliability of annotations, we only chose "Master" annotators, a worker pool with an excellent track record of completing tasks with precision. In each HIT, the workers

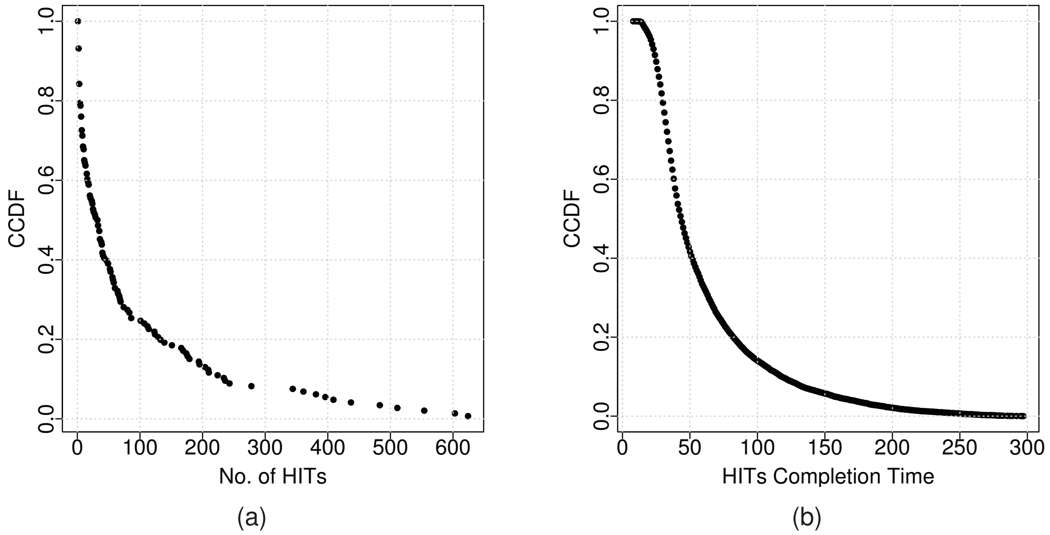


Figure 3.3 – Worker Participation. Plot showing the complementary cumulative distribution function (CCDF) of a) Number of HITs per worker, b) HITs completion times.

were asked to view an image of an urban space, and then rate their personal impressions based on what they saw along 12 labels. Additionally, workers were required to view images in high-resolution (and not just the image thumbnails). Workers were not given any information of the studied city or the country to reduce potential bias and stereotyping associated to the place identity. We collected 10 annotations for each image and label, resulting in a total of 13,000 responses (12,000 for the *city image* corpus and 1,000 for the *evening image* corpus) and a total of 156,000 individual judgments. Every worker was reimbursed 0.10 USD per HIT.

We also gathered crowdworkers' demographics via an email-based survey. We asked workers about their age group, gender, level of education, current place of residence (categorized as rural, suburbs, small town, mid-size town, or city), and any experience of visiting any developing countries, in any region including Latin America, Asia and Africa.

3.4.1 Worker Participation

For a total number of 12,000 HITs available for *city image* corpus, we observe that workers completed an average of 82 HITs, while they could potentially undertake 1,200 HITs (400 HITs per city). We had a pool of 146 workers who responded to our tasks. While 50% of the workers submitted less than 30 HITs, the worker with the highest number of HITs completed 624 assignments (Figure 3.3a). We observe a long-tailed distribution in HIT completion times (mean: 59 secs, median: 43 secs, max: 297 secs), as shown in Figure 3.3b. It is worth noting that we allocated a maximum of 5 minutes per HIT. Similar statistics were obtained for the 1,000 HITs available for *evening image* corpus.

3.4.2 Worker Demographics

Of all 146 HIT respondents, 53% replied to our demographics survey. We notice a slightly skewed gender ratio (58% of workers being female), which corroborates earlier findings in the online crowdsourcing literature [169]. 80% of respondents reported their ethnicity as White/-Caucasian, with 12% participants being Asian, and 3% each belonging to Hispanic/Latino and Black/African American. 45% of respondents are college graduates. Furthermore, we also notice that the worker population is relatively middle age with the most popular category (43%) being the age group of 35-50 years old (18–24: 3%, 25–34: 32%, 50+: 22%)

While only 18% of our worker pool reported to live in a big city, the majority of them (45%) are sub-urban (for the remaining categories: rural: 18%, small-sized town: 9%, mid-sized town: 9%). Only a minority (23%) of the survey respondents reported having experience visiting any country in the developing world. For those with traveling experience in developing countries, holidays and tourism were the main purposes of the visit (55%). Amongst the visited countries, 44% of these subset of respondents have traveled to Mexico, which is not surprising given that the pool of crowdworkers is based in the US.

3.5 Crowdsourcing Results

In this section, we report the results on crowdworkers' annotation quality, present descriptive statistics of crowdsourced annotations, provides statistical results on cross-city comparison, and study differences in urban perceptions across different times of the day.

3.5.1 Annotations Quality

We begin our analysis by assessing the reliability of annotations. We measured the inter-rater consensus by computing intraclass correlation (ICC) among ratings given by the worker pool [189]. Our annotation procedure required every place to be judged by k annotators randomly selected from a larger population of K workers. The average ICC measures, $ICC(1,k)$, were computed for each label and city across all images. Table 3.1 reports the $ICC(1,k)$ values for all cities for $k = 10$. In addition to listing the individual scores for each city and label, we also report the combined $ICC(1,k)$ scores for each label and the whole dataset, where we have combined all places across cities. We observe acceptable inter-rater consensus for most labels, with all values being statistically significant (p-value < 0.01).

We notice that the inter-rater reliability for labels is above 0.7 for most of the labels and cities, suggesting that MTurk observers tend to agree on their perceptions of most dimensions. The *quiet* label achieves high agreement from images not showing any sound (0.73 combined score). On the other hand, the label *polluted* is the one with lowest combined ICC (0.64). We also observe that *accessibility* has low ICC for two of the three cities. Silao has overall received the lowest ICC scores compared to the other two cities.

Label	Guanajuato	Leon	Silao	Combined
Accessible	0.86	0.55	0.36	0.72
Dangerous	0.83	0.65	0.73	0.76
Dirty	0.85	0.72	0.70	0.78
Happy	0.82	0.76	0.61	0.78
Interesting	0.61	0.70	0.60	0.70
Pleasant	0.83	0.77	0.66	0.79
Picturesque	0.77	0.69	0.64	0.76
Polluted	0.68	0.56	0.57	0.64
Preserved	0.82	0.75	0.63	0.77
Pretty	0.80	0.69	0.66	0.76
Wealthy	0.84	0.73	0.57	0.76
Quiet	0.71	0.65	0.53	0.73

Table 3.1 – $ICC(1,k)$ scores of 12 dimensions for each city. All values are statistically significant at $p < 0.01$.

3.5.2 Descriptive Statistics

Given the multi-annotator impressions, it is necessary to create a composite score for each image, given a label. To gather the individual ratings, we used an ordinal scale, which implicitly describes a ranking. It is known that the central tendency of an ordinal variable is better expressed by the median [193]. Thus, we compute the median score for each label given the 10 responses per image. Given the median scores, we then compute the mean scores and standard deviations for each label using all 400 images for each city.

Table 3.2 lists the descriptive statistics for each city and label, in addition to showing the aggregated scores for each label across all cities. At the level of individual annotations, the minimum and maximum values are 1 and 7 respectively for each label and city, indicating that the full scale was used by the crowdworkers. The mean scores for the majority of labels is below 4 for each city, which indicates a trend towards disagreement with the corresponding label. On the other hand, each city has urban sites that score high and low for each dimension.

In all cities, the mean scores for *accessible* are the highest amongst all labels. On all labels phrased positively (except *accessible* and *wealthy*), Guanajuato scores the highest amongst all cities, which is not surprising given that Guanajuato is a UNESCO world heritage site with a vibrant tourism industry. In contrast, *wealthy* has the lowest mean score for all cities (combined mean 2.64), which is not surprising either given that the type of cities we studied and the intended goals of the crowdsourced collection, were leaned towards documenting urban concerns.

From Table 3.2, we observe variation in the mean values across cities for some of the labels, but a few differences stand out. For instance, the mean differences of the *picturesque* and *interesting* attributes between Guanajuato and Silao, and the *quiet* attribute between Guana-

Label	Guanajuato	Leon	Silao	Combined
Accessible	4.62 (1.1)	5.02 (0.8)	4.41 (0.7)	4.69 (0.9)
Dangerous	2.92 (1.2)	2.86 (0.8)	3.17 (0.9)	2.98 (1.0)
Dirty	3.00 (1.2)	3.05 (0.9)	3.44 (1.0)	3.16 (1.1)
Happy	3.97 (1.1)	3.69 (0.8)	3.36 (0.8)	3.67 (1.0)
Interesting	4.38 (1.0)	3.63 (0.8)	3.50 (0.8)	3.84 (0.9)
Pleasant	4.13 (1.1)	3.83 (0.8)	3.48 (0.8)	3.82 (1.0)
Picturesque	3.55 (1.2)	3.00 (0.9)	2.73 (0.8)	3.09 (1.1)
Polluted	2.55 (0.9)	2.93 (0.8)	3.19 (0.9)	2.89 (0.9)
Preserved	4.04 (1.2)	4.00 (0.9)	3.48 (0.9)	3.84 (1.0)
Pretty	3.41 (1.2)	3.10 (0.9)	2.80 (0.9)	3.11 (1.0)
Quiet	4.08 (0.9)	3.24 (0.8)	3.10 (1.0)	3.47 (1.0)
Wealthy	2.58 (1.0)	2.90 (0.8)	2.43 (0.7)	2.64 (0.9)

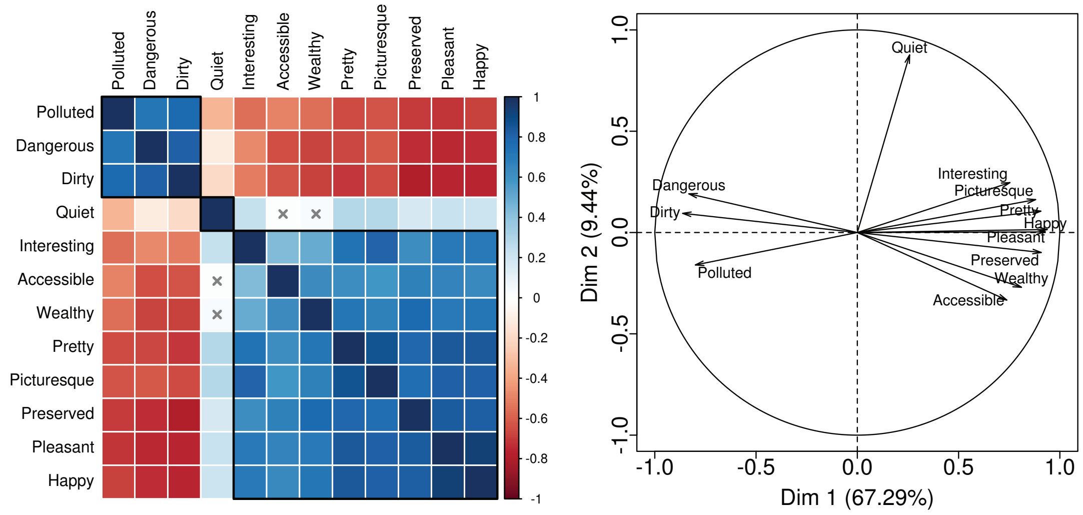
Table 3.2 – Means and standard deviations (in brackets) of annotation scores for each city and label.

juato and Leon and Silao all exceed 0.8, potentially suggesting differences in city perceptions. A systematic statistical analysis of these differences are presented in Section 3.5.4.

3.5.3 Correlation and PCA Analysis

To understand basic statistical connections between urban awareness labels, we performed correlation analysis using the mean annotation scores for all labels. Figure 3.4a visualizes the correlation matrix across all dimensions using the aggregated data for all cities. We have used hierarchical clustering to re-order the correlation matrix in order to reveal its underlying structure. We color coded the matrix instead of providing numerical scores to facilitate the discussion. We observe three distinct clusters. Starting from the bottom right in the first cluster, all the positive labels *happy*, *preserved*, *pretty*, *picturesque*, *pleasant*, *interesting* and *wealthy* are highly collinear with pairwise correlations exceeding 0.7. The second cluster consists of *quiet* dimension. The third cluster (top-left) lies on the opposite spectrum with respect to cluster one, and consists of *dangerous*, *dirty* and *polluted*. Each of these clusters correspond to different aggregate impression, the first and third somewhat resemble “sentiment” i.e., positive/negative. As such, we also observe significant negative correlations between dimensions in cluster one and cluster three.

Furthermore, the strong correlation between *dangerous* and *dirty* points towards the validation of the well-known Broken Window Theory in urban sociology [224], which postulates that the presence of physical disorder (garbage, grafitti, etc.) in urban environments leads to social disorder (crime, fear). In other words, in our study, we found that images which were perceived to be *dirty* (indicating some form of physical disorder) were also perceived as *dangerous* (indicating some form of social disorder), with high correlation (Figure 3.4a).



(a)

(b)

Figure 3.4 – a) Plot showing the correlation matrix between dimensions. Matrix is color coded as per the palette shown in the right, with blue (resp. red) indicating positive (resp. negative) correlation coefficients. Black rectangular borders indicate the three distinct clusters found in the correlation matrix. Cells marked *X* are *not* statistically significant at $p < 0.01$. b) Plot showing the first two principal components on aggregated annotation scores on 1,200 images across all cities.

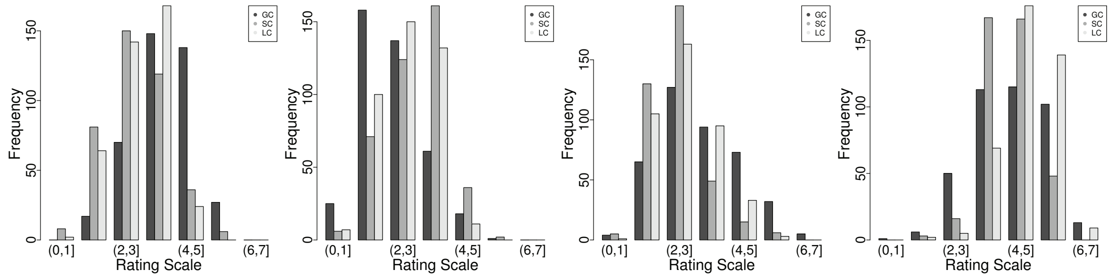
PCA Analysis: To further explore the relationships between labels, we perform principal component analysis (PCA) on the aggregated annotation scores for all 1,200 images. PCA is a statistical method to linearly transform high dimensional data to a set of lower orthogonal dimensions that best explains the variance in the data [157]. In Figure 3.4b, we show the first two principal components which explain 77% of the variance in the annotation scores along the 12 dimensions. Note that before applying PCA, the labels were scaled to unit variance. We observe that the first component, which accounts for 67% of the variance, contains labels that resemble either the positive or negative "sentiment", respectively, on the right and left side of X-axis. Furthermore, component two primarily contains label *quiet*. These results corroborate the findings from correlation analysis and have clear support from early work in environmental psychology [174]. Note that similar findings were observed for ambiance impressions of indoor places in Section 2.6.2 (Figure 2.6).

3.5.4 Statistical Comparison across Cities

To better understand whether mean differences across cities for some of these urban awareness labels are statistically significant, we perform the Tukey's honest significant difference (HSD) test. Tukey's HSD test is a statistical procedure for groups which compares all possible pairs of mean values for each group, with the null hypothesis stating that the mean values being compared are drawn from the same population [207]. We performed the

Label	City Pair	Mean Difference ± CI
Quiet	SC-GC	-0.98 ± 0.19
Interesting	SC-GC	-0.88 ± 0.18
Quiet	LC-GC	-0.84 ± 0.19
Picturesque	SC-GC	-0.82 ± 0.21
Interesting	LC-GC	-0.75 ± 0.18
Pleasant	SC-GC	-0.65 ± 0.19
Polluted	SC-GC	$+0.63 \pm 0.18$
Happy	SC-GC	-0.62 ± 0.19
Accessible	SC-LC	-0.61 ± 0.19
Pretty	SC-GC	-0.61 ± 0.20

Table 3.3 – Tukey's HSD test statistics showing the top–10 significant results, where the differences in the observed means across cities for labels exceed 0.6. GC, LC, SC respectively refers to Guanajuato City, Leon City and Silao City. All mean differences are statistically significant at $p < 0.01$.



(a) Quiet (b) Polluted (c) Picturesque (d) Accessible

Figure 3.5 – Barplots comparing the mean annotation scores across all cities for a) Quiet, b) Polluted, c) Picturesque, and d) Accessible.

HSD test to compute pairwise comparisons of mean values between cities for each label, which resulted in a total of 36 comparisons (3 city-wise pairs across 12 dimensions). Out of a total of 36 comparisons, we found 30 comparisons to be statistically significant at p-value $<$ 0.01. In Table 3.3, we report the top-10 significant results of the Tukey's HSD test, where the differences in the observed means were 0.6 or higher i.e., greater than half a point on the rating scale. We refrain from making claims for all smaller effect cases, following recent discussions in the statistical literature [76]. Additionally in Figure 3.5, we show the barplots comparing the mean annotation scores across all cities for four labels to elucidate some of the significant results from Tukey's HSD statistics. Based on these statistics we observe that:

1. Outdoor urban places in Guanajuato were perceived as more *quiet*, *picturesque* and *interesting* compared to places in Leon and Silao (rows 1,2,3,4,5 in Table 3.3). Overall, Guanajuato has the highest mean scores amongst all three cities on most positive labels, except *accessible* and *wealthy*, and the lowest mean scores on most negative labels

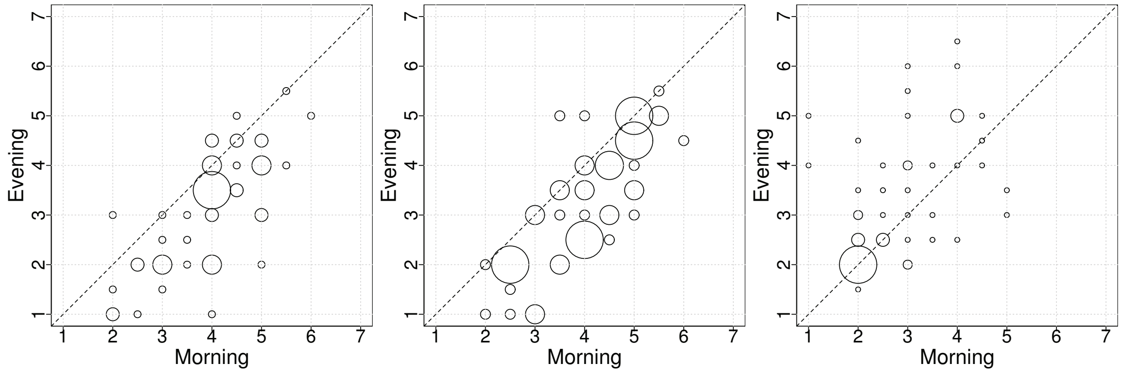
Label	Mean Scores		Mean Diff. ± CI	$p-$value
	Morning	Evening		
Accessible	4.40 (0.93)	4.38 (1.26)	-0.02 ± 0.58	0.928
Dangerous	2.86 (0.99)	3.38 (1.29)	$+0.52 \pm 0.60$	0.026
Dirty	2.89 (1.09)	3.33 (1.29)	$+0.44 \pm 0.63$	0.069
Happy	3.92 (0.98)	3.11 (1.18)	-0.81 ± 0.57	**0.0003**
Interesting	4.35 (0.74)	3.78 (1.18)	-0.57 ± 0.52	**0.005**
Picturesque	3.63 (0.97)	3.00 (1.21)	-0.63 ± 0.58	**0.005**
Pleasant	4.01 (1.03)	3.24 (1.20)	-0.77 ± 0.59	**0.0008**
Polluted	2.48 (0.86)	2.79 (1.12)	$+0.31 \pm 0.52$	0.123
Preserved	4.03 (1.02)	3.32 (1.26)	-0.71 ± 0.60	**0.003**
Pretty	3.58 (1.05)	3.03 (1.23)	-0.55 ± 0.60	0.018
Quiet	4.12 (1.01)	3.79 (1.16)	-0.33 ± 0.57	0.132
Wealthy	2.71 (1.03)	2.15 (0.92)	-0.56 ± 0.51	**0.005**

Table 3.4 – Descriptive statistics and Tukey's HSD test statistics showing the results for morning and evening times of the day. Value marked in **bold** are statistically significant at $p < 0.01$.

except *dangerous*, as highlighted in Table 3.2. To highlight the differences in perception between Guanjuato and the other two cities, we present the barplot comparing the mean annotation scores of *quiet* dimension across all cities in Figure 3.5a. We observe that the relative percentage of places in Guanajuato which are rated higher than 4 is significantly larger than those corresponding to other cities. Similar patterns are observed for *interesting* and *picturesque* (Figure 3.5c).

2. Silao was perceived to be more *polluted*, when compared to Guanajuato with differences in mean scores exceeding 0.6 (row 7 in Table 3.3). We believe these results are an effect of agricultural and industrial activity in Silao, when compared to the historic and touristic nature of Guanajuato. When looking at the barplot in Figure 3.5b, we observe a similar pattern as highlighted above, where the relative proportion of places which were rated on a higher scale for being polluted in Silao are significantly larger than those in Guanajuato.
3. Silao was perceived to have lower *accessibility* than Leon (row 9 in Table 3.3 and Figure 3.5d). We believe this result is due to the fact that Leon, as a larger and more modern city, has many broad avenues and multi-lane streets. In contrast, Silao has small town streets. See examples in Figure 3.1.

To further validate the statistical significance of the Tukey's HSD test, we performed a series of pairwise Kolmogorov-Smirnov test (KS test) across all cities and labels. We performed the KS test to compare the cumulative distribution functions of each city-pair across each label (36 comparisons) and found 32 comparisons to be statistically significant for a statistical level $\alpha = 0.01$. Results from the KS test corroborates the findings from the Tukey's HSD test.



(a) Happy (b) Preserved (c) Dangerous

Figure 3.6 – Scatter plots showing the pair-wise annotator scores across two different times of the day for a) Happy, b) Preserved, and c) Dangerous. Each dot corresponds to an image, with the size of the dots proportional to the number of observations. The 45° line is also shown in all the plots.

3.5.5 Analysis of Evening Image Corpus

In this subsection, we analyze the *evening image* corpus to statistically test if the perceptions of the urban environments in one of the studied cities vary across different times of the day, along the selected dimensions. As described in Section 3.3.5, we collected 50 images of places during the morning (between 10-11 AM), and the evening (between 6-7:30 PM) in Guanajuato. In Table 3.4, we list the mean scores for each label across times of the day. To aggregate the impressions, we followed the same procedure as explained in Section 3.5.2 for each image and time slot. Using Table 3.4, we notice that the mean scores for the majority of labels is below 4 for each time slot, which indicates a trend towards disagreement with the corresponding label, analogous to the results obtained with the *city image* corpus (Section 3.5.2). Furthermore, we observe that the mean values of the perceptual ratings in mornings are similar to the ones observed for Guanajuato in Table 3.2. This is consistent with the fact that all the images from the *city image* corpus were taken between between 9 AM and 5 PM (Section 3.3.5).

For the *evening image* corpus, the mean annotation scores for all positive (resp. negative) labels are higher (resp. lower) for images taken in the morning, compared to the ones taken during the evening. In Table 3.4, we also report the results of of the Tukey's HSD statistics to test whether the mean scores differ across different times of the day for all labels. Based on these statistics we observe that:

1. Evenings were perceived as less *happy*, *pleasant* and *preserved*, in comparison with morning time (differences: -0.81, -0.77, -0.71 respectively). See Figure 3.6a and 3.6b.
2. We do not observe statistically significant differences in the perception of *dangerous* between mornings and evenings (p-value $=$ 0.026). This result is in contrast with findings from the literature [128, 86]. To put our finding in perspective, it is important to note that

places were overall not perceived as *dangerous* in the *evening image* corpus (combined mean: 3.12). See also Figure 3.6c.

Pair-wise Analysis: After observing that some of the perceptual ratings differ across times of the day, and in order to understand the variability of ratings for individual places, we examined the pair-wise ratings of each image between morning and evening time. We focus our pair-wise analysis on two statistically significant labels (*happy* and *preserved*), in addition to examining the non-significant *dangerous* label. Figure 3.6 shows the respective plots. If the perception ratings were similar between morning and evening, most of the points would have fallen on the $45°$ line. On the contrary, we observe that a significant majority of points lie below the line for *happy* and *preserved* (Figure 3.6a and 3.6b), indicating that places in the morning were perceived on a higher scale compared to evenings. Furthermore, it is interesting to observe a more mixed trend for the *dangerous* label (Figure 3.6c). These plots further validate our findings on the *evening image* corpus.

3.6 Automatic Inference of Outdoor Perception

In Section 3.5, reliable estimates of perception was obtained for several of the urban awareness dimensions, suggesting the presence of visual cues that allow to create such impressions. In this section, we address RQ3.3 to examine the feasibility to automatically infer these perceptual impressions using visual cues from images of outdoor places. For automatic inference we used the *city image* corpus consisting of 1200 images across three cities (Section 3.3.5). Due to the relatively small size of the evening image corpus, it was not used in the subsequent analysis.

3.6.1 Visual Feature Extraction

For automatic outdoor perception, we extracted a similar set of low-level and deep learning visual features as was done earlier for the inference of indoor ambiance perception. Table 3.5 summarizes the list of visual features extracted from images. A detailed explanation of these features has been described in Section 2.7.1.

To extract the low-level features on the *city image* corpus, we used the same configuration as described earlier. With regard to deep learning features, there is one key difference – the choice of the image database on which the deep CNN architecture was trained. For outdoor places, we chose a pre-trained CNN model which uses the GoogLeNet architecture trained on *Places205* database [232], in contrast to *ImageNet* that was used for indoor places. *Places205* database is a large-scale scene-centric database of 2.5 million images of indoor and outdoor scenes across 205 categories, while *ImageNet* is an object-centric database. It has been shown that *Places205* database contains more images per scene category compared to *ImageNet* [232]. Recently, authors in [159] reported that features extracted

Visual Feature	Description	Dimensionality
Color	Color histogram in RGB space	512
GIST	Dominant spatial structure of a scene	512
HOG	Histogram of oriented gradients	680
LBP	Spatial arrangement of color and intensities	256
CNN-CP	Final layer class probabilities using a GoogLeNet CNN pre-trained on *Places205*	205
CNN-FC	Fully connected layer of a GoogLeNet CNN pre-trained on *Places205*	1024

Table 3.5 – Summary of the visual features extracted from images.

from a pre-trained Places-CNN achieve higher performance accuracy relative to a similar CNN architecture trained on *ImageNet* for outdoor scenes. Given the kind of images that was collected during the *SenseCityVity* project (e.g., outdoor scenes of city's built environment, residential neighborhoods, streets and alleys, etc.), we believe *Places205* database is more suited for outdoor places compared to *ImageNet*.

With this configuration, besides extracting the final layer class probabilities (CNN-CP model), we also extracted the output of the fully-connected layer (FC) of the GoogLeNet architecture as additional feature representation (Table 3.5). To extract CNN features, all images were resized to 256 $\times$ 256 pixels and subjected to mean image subtraction. To conclude, for both CNN models we used the same CNN architecture (GoogLeNet) trained on *Places205* database, these models differ only in terms of derived activation features.

3.6.2 Inference Method and Evaluation

As done in Chapter 2, we formulate the inference of urban outdoor perception as a regression problem where our objective is to predict aggregated human impressions using visual cues extracted from images. We used Random Forest as the regression technique. For model validation, we performed m repetitions of a k-fold stratified cross-validation approach. For all the experiments, we set $m = 10$, and $k = 10$. After the model run and validation, we computed the mean of the evaluation metric across mk runs as the model output. To evaluate the predictive performance between feature sets, we used root-mean-square error ($RMSE$) and coefficient of determination (R^2) as evaluation metrics. These evaluation metrics were computed between the perceived ground-truth (i.e., human judgments) and predicted perceptual scores obtained for each label and feature set. For the baseline model, we chose the mean annotated score as the predicted value for each label to better understand and compare the predictive performance of each feature set individually.

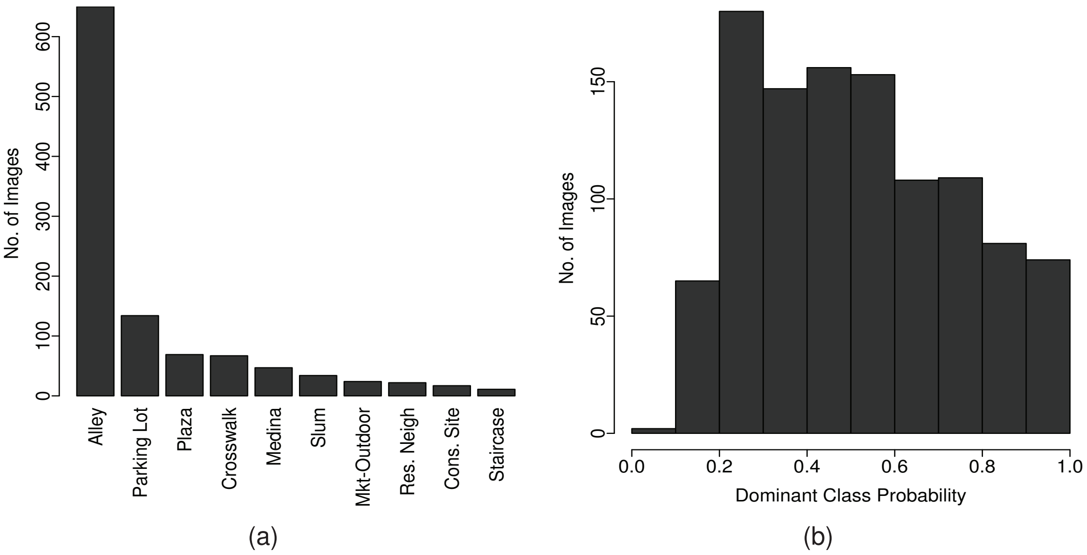


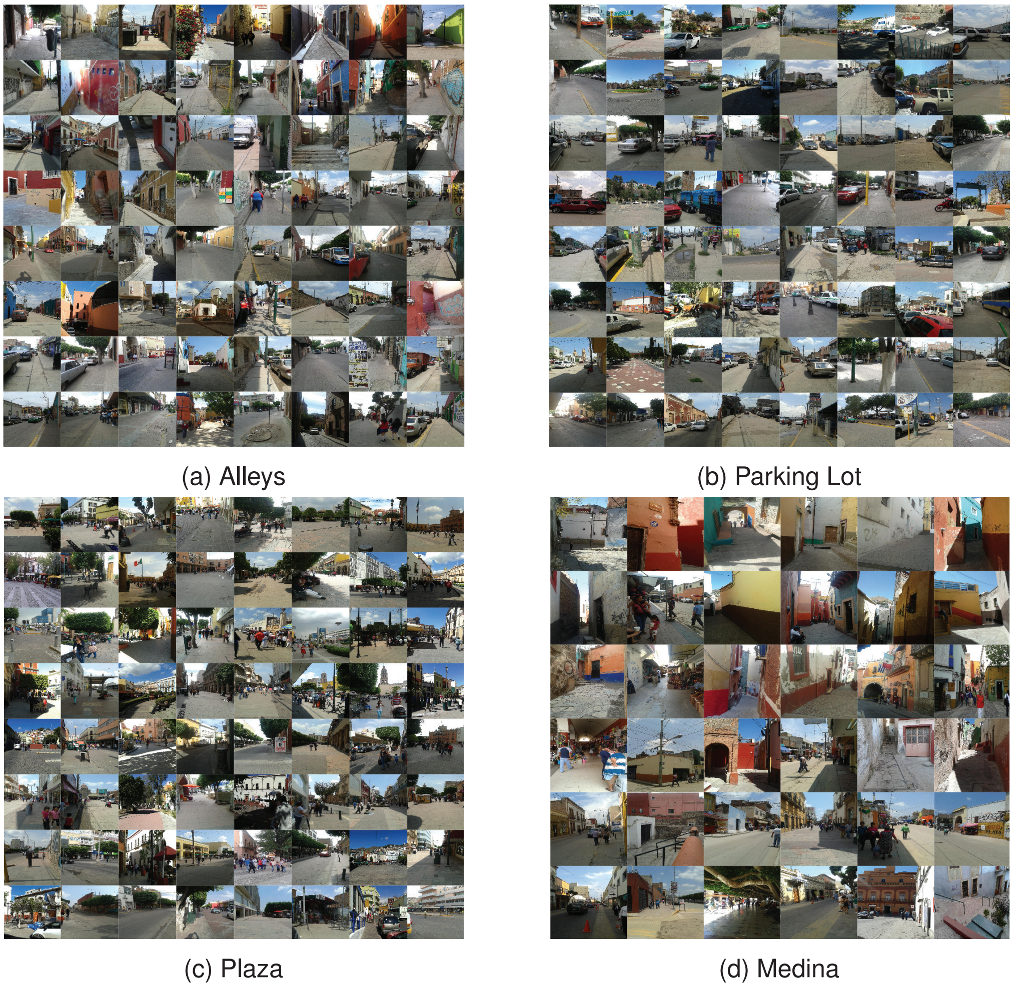
Figure 3.7 – a) Histogram of top-10 recognized classes for the *city image* corpus, b) Histogram of the class probabilities for the top-10 recognized categories for the *city image* corpus.

3.6.3 Visual Categories

We begin our analysis by examining the distribution of the most likely *Places205* category assigned to each image. Using the CNN model, for each image in the *city image* corpus, we obtained the vector containing the class probabilities across 205 scene categories. Given this vector, we chose the scene category with the highest probability as the dominant class for each image. Figure 3.7a shows the histogram of the top-10 recognized *Places205* categories. To visually illustrate the recognized classes, Figure 3.8 shows a mosaic of images across four of the top-10 *Places205* categories from the *city image* corpus.

Overall, the dominant class distribution exhibits long-tail characteristics, with a total of 52 unique scene categories. 90% of the images were assigned these top-10 categories (out of 205 possible categories). More than 54% of the images were classified as an “alley” as their dominant class (see Figure 3.8a for a visual illustration). Most of the top ten dominant categories are associated with the physical attributes of the outdoor environment (e.g., alley, parking lot, plaza, crosswalk, residential neighborhood, construction site, etc.) which potentially suggest an automated way to validate the images collected as part of the *SenseCityVity* study (Figure 3.8). These findings further point towards the differences between images describing indoor (*50K image* corpus) and outdoor environments (*city image* corpus). Refer to Figure 2.8 and Figure 3.8 for a visual comparison.

As a second observation, some of the recognized categories in Figure 3.7a may seem surprising at first glance e.g., “medina”, but after manually browsing the images belonging to these categories (Figure 3.8d), we found that these classes describe various attributes of the outdoor environment specific to the studied cities and are likely misclassified yet makes sense



(a) Alleys

(b) Parking Lot

(c) Plaza

(d) Medina

Figure 3.8 – Sample of random images from the *city image* corpus which were classified as (a) Alley, (b) Parking Lot, (c) Plaza, and (d) Medina. For each class, from top to bottom, images are sorted in decreasing order of dominant class probability. Best viewed in color.

visually. For instance, most of the images belonging to "medina" category contain images showing narrow and windings alleys and streets, which is typical to the city of Guanajuato. Of all images classified as "medina", 66% of images were from Guanajuato.

The distribution of dominant class captured similarities and differences in urban characteristics across the three studied cities. For all the three cities, a similar proportion of images were labeled as *alleys*. However, more *crosswalks* and *plazas* were identified in Leon city; more *medinas* and *staircase* were recognized in the city of Guanajuato; while in Silao city, more *outdoor markets* were classified relative to other two cities. These findings corroborate the functional and urban characteristics of each city (Section 3.3.2).

Furthermore, we noticed that the probabilities associated with the dominant classes were not similar across images. In Figure 3.7b, we plot the histogram of the dominant class

	Baseline	**Color**	**GIST**	**HOG**	**LBP**	**CNN-FC**		**CNN-CP**	
	$RMSE$	R^2	R^2	R^2	R^2	R^2	$RMSE$	R^2	$RMSE$
Accessible	0.94	0.21	0.23	0.16	0.25	**0.48**	0.69	0.46	0.70
Dangerous	0.99	0.16	0.16	0.09	0.15	**0.38**	0.79	0.35	0.80
Dirty	1.06	0.14	0.15	0.10	0.14	**0.37**	0.85	0.35	0.87
Happy	0.95	0.26	0.21	0.13	0.19	**0.46**	0.71	0.43	0.73
Interesting	0.94	0.25	0.30	0.18	0.25	**0.45**	0.70	0.41	0.73
Pleasant	0.97	0.24	0.21	0.15	0.20	**0.45**	0.73	0.42	0.75
Picturesque	1.06	0.23	0.25	0.15	0.22	**0.49**	0.76	0.44	0.79
Polluted	0.92	0.13	0.13	0.09	0.09	**0.30**	0.77	0.28	0.78
Preserved	1.05	0.20	0.20	0.15	0.18	**0.45**	0.78	0.40	0.81
Pretty	1.02	0.23	0.21	0.15	0.19	**0.46**	0.75	0.43	0.77
Wealthy	0.88	0.22	0.20	0.13	0.19	**0.46**	0.65	0.44	0.66
Quiet	1.00	0.32	0.30	0.18	0.18	**0.48**	0.73	0.45	0.75

Table 3.6 – Inference results for 12 ambiance dimensions for all feature sets, using R^2 and $RMSE$ as evaluation measures. Cells marked in **bold** correspond to the best R^2 result obtained for each dimension.

probabilities for the top-10 recognized categories. Some images were classified with higher probabilities; while others were fairly difficult to classify which might be due to scarcity of *city image* corpus kind of images in *Places205* database [202]. Overall, manually browsing these images highlight the various aspects of urban life and built environment of the studied cities – narrow and winding alleys, wall graffiti, clutter of open wires hanging low in the streets, unmarked pavements, etc. Note that the UDC participants were instructed to capture images of urban scenes in their natural setting and focused on the general environment (rather than only on detected problems).

3.6.4 Predicting Urban Perception

In this subsection, we evaluate the predictive performance of both low-level image descriptors and deep learning features to infer the urban perception of outdoor scenes. Table 3.6 reports the R^2 values between the human impressions (i.e., ground-truth) and predicted perceptual scores for each feature set over all 12 dimensions. For the baseline model, we only report the $RMSE$ values as it has a R^2 of 0.

For automatic inference, we built two CNN models which differ in the way visual features were extracted for each image using the GoogLeNet CNN. For the CNN-FC model, image features correspond to the output of the fully-connected layer of the GoogLeNet CNN, while CNN-CP model contains the final layer class probabilities across all 205 *Places205* categories as the feature vector (Table 3.5). For both the CNN models, we also report the $RMSE$ values. A model having a better feature representation to estimate urban perception would result in higher R^2 and lower $RMSE$ values, when compared with other models. Using the results reported in Table 3.6, we can make the following observations:

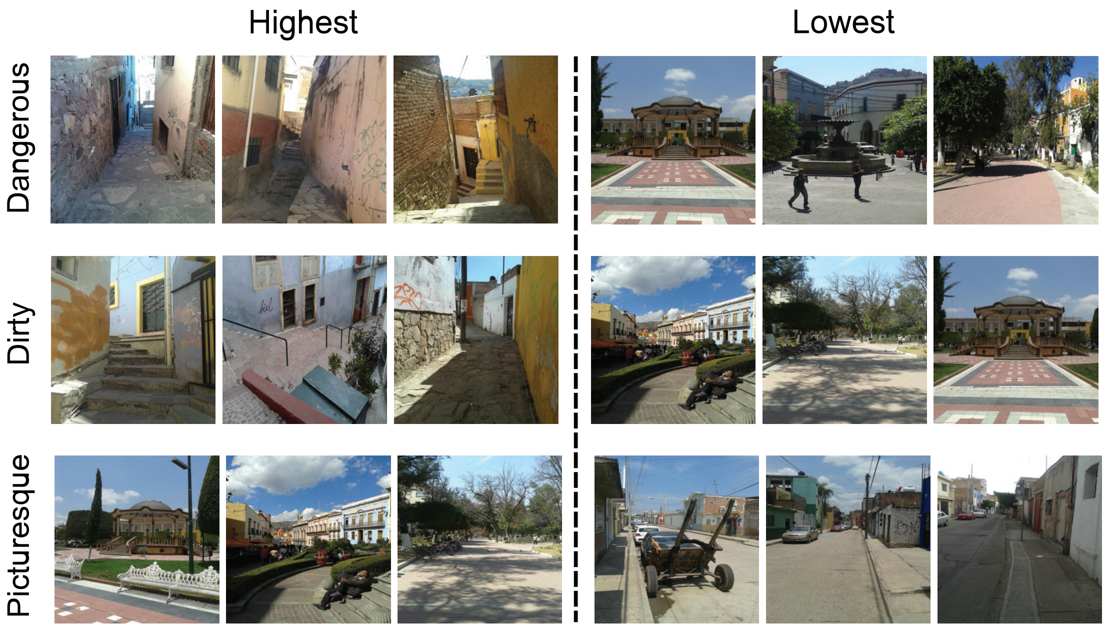
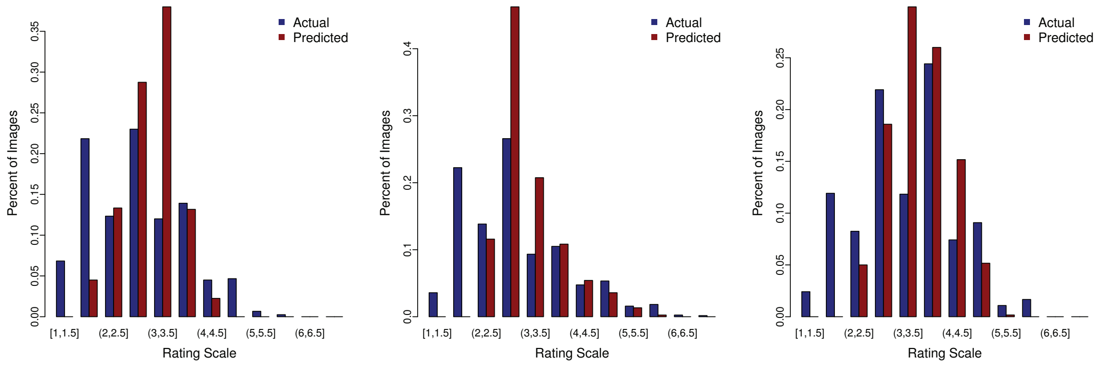


Figure 3.9 – Sample of images with the highest and lowest predicted scores for *dangerous*, *dirty*, and *picturesque* dimensions using CNN-FC model. Best viewed in color.

- CNN based features consistently outperformed low-level image descriptors (including Color Histogram, GIST, HOG, and LBP) for all dimensions. These findings are consistent with the results reported for indoor place ambiance (Table 2.6) and computer vision literature [165]. For outdoor places, we did not conduct the experiments combining low-level image features as we believe that the overall trends would not change significantly, as observed for indoor place ambiance (Table 2.6).
- Overall, the results indicate that a maximum R^2 of 0.49 for *picturesque* dimension can be obtained using the CNN-FC regressor. For 9 out of 12 variables, the obtained R^2 values exceeded 0.44, with the lowest one obtained for *polluted* label.
- Among the low-level features, Color Histogram, GIST, and LBP have comparable predictive performance, while HOG performs relatively poorly for most of the dimensions. Color and GIST features achieved the best performance for the *quiet* label ($R^2 \geq 0.30$). The performance of HOG features, which have performed well for scene recognition [226], potentially suggest that the shape context within an image holds low predictive power to characterize the perceptual attributes of the outdoor environments in our dataset.
- While examining individual labels, *polluted* achieved the lowest predictive performance across all feature sets ($R^2 \leq 0.30$), which might be associated with the labeling noise during the annotation process i.e., low inter-rater agreement (Table 3.1). Positively phrased labels (*happy*, *pretty*, *picturesque*, etc. – cluster 1 in Figure 3.4a) have higher R^2 values than negatively phrased labels (*dangerous*, *dirty* and *polluted* – cluster 3 in Figure 3.4a), though all the dimensions in both clusters achieve similar R^2 values.
- Similar to the *loudness* trait for indoor places ($R^2 \geq 0.26$), the *quietness* of outdoor places achieved $R^2 \geq 0.18$ for all feature descriptors.



(a) Dangerous (b) Picturesque (c) Quiet

Figure 3.10 – Plots comparing the histograms of both actual and predicted values for a) Dangerous, b) Picturesque, c) Quiet.

- While comparing the performance of CNN-FC and CNN-CP model, we notice that CNN-FC model outperformed CNN-CP model, though the R^2 and $RMSE$ values are relatively comparable for all dimensions. To visually assess the performance of the CNN-FC model, Figure 3.9 shows example images with the highest and lowest predicted scores for three studied dimensions.

Overall, our findings suggest the suitability of using pre-trained CNN models to infer high-level human perceptual judgments for outdoor scenes. Our findings are comparable to what have been reported in the literature. On the Place Pulse dataset, Naik et al. built a predictor combining low-level features from geometric texton histograms, GIST, and geometric color histograms to achieve an R^2 of 0.54 for the *safety* dimension [142]. Using the same dataset, the authors in [151] reported correlation coefficients (r) ranging from 0.4 to 0.7 to predict *safety*, *uniqueness*, and *wealth* using generic deep convolutional activation features. On the Place Pulse data, Porzi et al. using a different CNN architecture, reported an accuracy of 70% when predicting users' votes on image pairs for the *safety* label [159].

To further diagnose the performance of our top performing feature representation, CNN-FC model, we plot the histogram of the ground-truth and predicted values for three labels in Figure 3.10. Using these plots, for each dimension, we observe that the model is over-estimating lower values and under-estimating higher values i.e., even though we obtained promising R^2 values, the model is biased towards the mean value for each dimension. In other words, for images which were perceived by humans as high on *dangerous* or *picturesque*, was predicted by the model as less *dangerous* or *picturesque* and vice-versa. Even though we obtained promising R^2 values the model seems to be sensitive or biased towards outliers. Recent work has proposed a CNN model that weights extreme values during training as means to counter the inherent data bias due to unbalanced distribution [102]. It seems like a promising approach to increase the model performance for future work.

	Guanajuato			Leon			Silao		
	E_{BL}	E_{RF}	R^2	E_{BL}	E_{RF}	R^2	E_{BL}	E_{RF}	R^2
Accessible	1.13	0.7	0.62	0.8	0.64	0.38	0.75	0.67	0.22
Dangerous	1.17	0.85	0.49	0.84	0.69	0.34	0.92	0.75	0.36
Dirty	1.22	0.91	0.46	0.93	0.79	0.29	0.96	0.82	0.31
Happy	1.11	0.77	0.54	0.81	0.62	0.41	0.8	0.67	0.33
Interesting	0.95	0.77	0.36	0.82	0.62	0.44	0.80	0.64	0.37
Picturesque	1.23	0.87	0.51	0.88	0.66	0.45	0.83	0.68	0.34
Pleasant	1.12	0.78	0.54	0.84	0.65	0.41	0.81	0.68	0.32
Polluted	0.91	0.77	0.30	0.82	0.71	0.25	0.92	0.81	0.25
Preserved	1.19	0.82	0.54	0.93	0.70	0.45	0.90	0.76	0.31
Pretty	1.18	0.84	0.51	0.87	0.67	0.43	0.88	0.73	0.33
Quiet	0.94	0.77	0.35	0.82	0.68	0.33	0.95	0.74	0.42
Wealthy	1.01	0.69	0.55	0.83	0.65	0.41	0.70	0.58	0.31

Table 3.7 – Inference results for each city across 12 ambiance dimensions using CNN-FC model. For each city, E_{BL} and E_{RF} refer to the $RMSE$ for baseline and CNN-FC model respectively.

Urban Perception across Cities: Until now, we have discussed the inference results using the combined data for all cities. Now, we examine the predictive performance of CNN-FC model on each city. Table 3.7 reports the inference results for each city across all 12 ambiance dimensions. For each city, we report the baseline $RMSE$ (E_{BL}) and CNN-FC model $RMSE$ (E_{RF}), in addition to reporting their respective R^2 values. Using the results listed in Table 3.7, we observe similar trends to the findings reported for inter-rater reliability scores for each city (Table 3.1). Amongst all the three cities, Guanajuato achieves the best performance for all dimensions (except *interesting* and *quiet*), while Silao has received the lowest R^2 scores for most dimensions. As a second observation, the obtained R^2 values for Guanajuato are higher when compared to the values for all cities combined for most of the dimensions (Table 3.6).

3.7 Conclusion

In this chapter, we presented the design and implementation of the *SenseCityVity* project co-designed with over 200 young volunteers contributing over 7,000 geo-localized images, and 380 first-person perspective videos documenting three cities in Mexico. The collected images describe outdoor scenes and views of each city's built environment. Using a subset of 1,200 images, we presented a computational analysis and automatic characterization of urban perception of outdoor places for three cities in central Mexico. Using the aggregated annotations, we concluded that outdoor environments can be reliably assessed with respect to most urban dimensions. We further demonstrated the feasibility to automatically infer human perceptions of outdoor places using a variety of low-level image and deep learning features with promising accuracy. To demonstrate the additional value of collecting images via crowdsensing relative to street-level imagery, we investigated the perception of urban environments during different times of the day and found that places in the evening were perceived as less *happy*, *pleasant* and *preserved* relative to the same place in the morning. Overall, our findings are promising that could potentially provide urban designers and city planners a data-driven and scalable approach to examine the physical appearance of cities and help design urban policies informed by urban perception.

We conclude the chapter by discussing the potential impact of our work. Today, the purpose of improving the urban life in developing cities depends on the collective action of citizens, communities, and governments. It is through synergistic interactions between government and self-organized citizens that complex urban issues can be tackled in the midst of chaotic urban growth and prevailing conditions of economic inequality. In this regard, educating citizens to develop a more discerning understanding of urban problems is crucial. Our work addresses this matter (beyond scientific inquiry) by contributing tools that communities could use to generate benefits for themselves. The mobile crowdsensing approach used to collect the data enabled participating volunteers to become more aware of their urban environment. The data collected by and for the people provides an alternative and more comprehensive picture of the issues that matter to citizens, beyond a mapping exercise conducted by professional surveyors, which is often expensive and less detailed. As a final point, our proposed methodology was recently adopted by other researchers in other cities in Mexico, including Leon, Merida, and Queretaro.

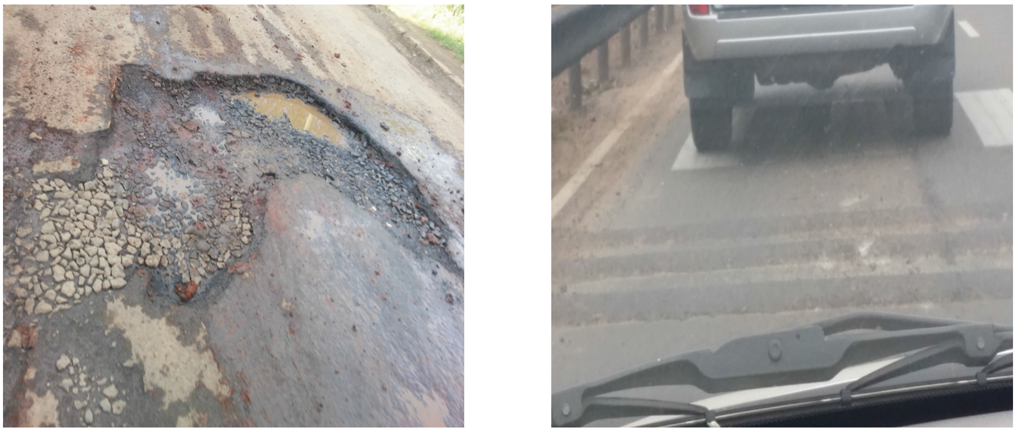
4 Crowdsensing Road Infrastructure Conditions

4.1 Introduction

URBANIZATION is increasing at a faster pace than ever. For the first time in human civilization, a majority of world's population lives in urban areas [209]. Rural and transnational populations are constantly migrating towards cities in search for better opportunities and quality of life. Rapid urbanization has put a considerable amount of strain on urban infrastructure including public transportation and road surface conditions. This is a common scenario in most of the emerging economies including cities in Africa, India, China, and Latin America.

In this regard, Nairobi is no different. Nairobi is one of the fastest growing metropolitan cities in Africa and a major business and technology powerhouse in East Africa. In the last decade, Nairobi (population of 3.1 million in 2009) has experienced rapid urbanization, which has led to a rise in traffic congestion and long commute times resulting in growing frustration amongst commuters [97]. While there has been significant growth in car ownership and informal bus transit (known as *matatus*), the transportation infrastructure has not kept pace with this growth. It is estimated that traffic congestion in Nairobi costs the economy an estimated 37 billion Kenyan Shillings annually (equivalent to 413 million USD) [13]. This trend is unsustainable and detrimental to Kenya's 2030 development plans [203].

In addition to this growth in travel demand, Nairobi has not received adequate attention with regard to long term transportation planning [79]. Nairobi roads are known for their hazardous conditions, which include gaping potholes, unregulated speed-bumps and abrupt road surface changes. Figure 4.1 shows some of the road hazards typically seen on the streets of Nairobi. In the rest of the chapter, by "road hazards", we specifically refer to potholes and speed-bumps. Although speed-bumps are traditionally used for traffic calming and speed mitigation, in Nairobi they are frequently unlabeled, poorly (and often irrationally) placed, and are not accompanied with proper signage (Figure 4.1b). For example, in one of our field tests on a 2.4km stretch of road where the speed limit is 60km/h, under free flow conditions we encountered 13 speed-bumps resulting in an average travel speed of 20km/h.



(a) Potholes

(b) Speed Bumps

Figure 4.1 – Road surface conditions in Nairobi (a) Potholes (b) Speed Bumps.

As is the case with most developing cities, Nairobi lacks monitoring technologies to obtain reliable data on road infrastructure conditions in urban areas. The costs associated with deploying sensing infrastructure to monitor road quality in developing urban areas are often prohibitive, and impractical to implement; therefore it becomes imperative to leverage locally available resources (i.e., people) to collect this information. We posit that mobile crowdsensing offers a promising avenue to obtain reliable data on urban infrastructure conditions.

In this chapter, we present the design and implementation of a mobile prototype system, called *CommuniSense*, to address the problem of documenting Nairobi's road infrastructure conditions. *CommuniSense* allowed citizens to locate, describe, and take pictures of road hazards using smartphones. A crowdsensing approach, using smartphones, is promising due to the widespread penetration of mobile devices (78.2% mobile penetration in 2013 [203]) and the increasing popularity of smartphones in Kenya. Smartphone penetration has been fueled by the introduction of low-cost Android phones and the trend is expected to continue as different vendors including Google, Huawei, and LG plan to roll out more low-cost smartphone devices in Kenya [204, 205]. This trend is similarly observed in other developing countries [160].

Prior research on route recommendation has typically focused on computing the shortest route with minimal commute times. Existing mapping services only take into account traffic congestion. However, they fail to account for poor road surface conditions, which are prevalent in many developing cities including Nairobi. Our current work builds on recent work [111] that introduces a mobile framework to collect GPS traces from probe vehicles, to examine speed implications caused by poor road quality in Nairobi. Our work extends this work by providing a tool to collect labeled road hazard data. Together, these two strategies (collection of mobile sensor data, and labeled road hazard data) can be used to predict the locations of various road hazards. Our broader objective is to build a travel model to estimate travel speeds, fuel consumption, and vehicle emissions, as a function of road surface conditions.

CommuniSense advocates a model of citizen participation that facilitates the push towards a public-private partnership for small-scale community infrastructure maintenance activities. In this regard, *CommuniSense* can be seen as a special case of the *SenseCityVity* project (Chapter 3). In *SenseCityVity*, citizens collected images of city's built environment with a focus on urban issues, whereas in *CommuniSense* citizens turned the focus of their cameras downwards to capture images of road hazards. This chapter addresses the following research questions:

RQ4.1: What are the perceptions of citizens in large developing cities towards the state of existing road infrastructure conditions? (**RQ4.1** maps to thesis's **RQ2**)

RQ4.2: How can mobile crowdsensing be leveraged to support citizen-based data collection of road infrastructure conditions in developing cities, like Nairobi? (**RQ4.2** maps to thesis's **RQ1**)

RQ4.3: How can the correctness and validity of crowdsensed data, in particular images, be verified? (**RQ4.3** maps to thesis's **RQ1**)

This research was partly done during my research internship at IBM Research Africa in Nairobi, together with Jidraph Njuguna, Tierra Bills, Aisha W. Bryant, and Reginald Bryant. The chapter makes the following four research contributions:

1. Our first contribution is a travel survey to understand the existing state of road quality conditions in Nairobi. The survey questionnaire collected demographics, weekly travel practices, perception of current road quality conditions and impact on their travel experience. To account for socio-economic bias, we conducted the survey via two different channels: web-based and SMS-based. To the best of our knowledge, our survey is the first survey with publicly available results to assess the opinions of people on road surface conditions and their impact in Nairobi.
2. Based on the survey's findings, we then developed a mobile crowdsensing application, called *CommuniSense* to collect data on road surface conditions. The application allows users to submit road hazard reports where they locate, describe, and take pictures of road hazards. We test our application through a two-week field study amongst 30 participants, who submitted a total of 254 reports characterizing various forms of road hazards from different areas in Nairobi.
3. Our third contribution is the verification of the authenticity of user-contributed reports from our field study. For this, we designed and conducted an online crowdsourcing study using MTurk, to verify whether submitted reports indeed depicts road hazards, as recorded by smartphone users. We found 92% of user-submitted reports to match the MTurkers judgments.
4. Our fourth contribution is the design and development of a layered and an interactive web-based visualization framework to visualize the location of road hazards and other geo-localized information in Nairobi.

While the mobile prototype was designed and tested on a specific city, our methodology is equally applicable to other developing cities. Our overall contribution lies in designing, implementing and understanding the adoption of a mobile crowdsensing platform in a world region which is often under-represented and under-studied in mobile computing research.

The chapter is structured as follows. Section 4.2 presents a review of the related work. Section 4.3 describes the design and implementation of the road quality survey and demonstrates its key findings. Using the survey's findings, Section 4.4 outlines the design and the backend architecture of the mobile application and reports the results of the field study. The mobile application was developed by Jidraph Njuguna. Section 4.6 proposes the image verification methodology for user-contributed submissions using online crowdsourcing. Section 4.7 presents the visualization framework. Section 4.8 discusses the technical challenges faced during the design and implementation of *CommuniSense*. Finally, Section 4.9 concludes with a summary of the findings. This chapter outlines research that has been published here [185].

4.2 Related Work

Given the multifaceted nature of our work, we review the related work that speaks to three research domains, as described below.

4.2.1 Data Collection in Developing World

In the developing world, one of the most common ways to collect data is via text messages or SMS. The low cost of feature phones and wide availability of SMS service has enabled various SMS-based data collection systems including RapidSMS [10], FrontlineSMS [27], and Ushahidi [147]. FrontlineSMS has been designed to gather unstructured data, while RapidSMS has been designed primarily for structured data. The Ushahidi platform extended FrontlineSMS and was deployed first in Kenya during the 2007 post-election violence. The platform allowed Kenyans to submit violence related reports using SMS (and email).

Despite the popularity of SMS, SMS-based tools are often unreliable and expensive. The costs associated with sending 1Mb data over SMS is over 3600 times more expensive than GPRS (General Packet Radio Service) [12] (as cited in [88]). Furthermore, SMS-based tools cannot provide fine-grain location and high-quality image data. Although these platforms have been successfully deployed in the past, they provide a bare minimum support for user interactivity and are designed to be deployed in environments experiencing financial, social, political, or natural disaster hardships.

In the recent past, the increasing popularity of smartphones and increasing investment in cellular infrastructure has generated excitement for smartphone-based crowdsourcing solutions in developing regions. This growth has provided major opportunities to collect data in a cost effective manner. Tools like OpenDataKit (ODK) [88] and Nokia Data Gathering [8], have been designed primarily for the developing world. ODK is a smartphone-based platform designed to build data collection solutions for organizations with limited financial and technical resources (e.g., NGOs). In a more recent work, the team behind ODK, redesigned its architecture to simplify the process of creating and managing data collection pipelines for individuals with limited technical experience [37].

Technically, *CommuniSense* is similar to ODK 2.0. We designed and build our system from scratch to integrate incentive (financial/social) [84, 229], gamification, crowdsourced verification and social media modules for future needs. To the best of our understanding, incorporating these modules in ODK would require systemic changes to its core architecture.

4.2.2 Road Surface Monitoring

In field of mobile sensing, there has been research interest to automatically detect potholes and monitor road surface conditions using accelerometer sensor data [66, 170, 136]. In *Pothole Patrol*, the authors presented a machine learning based approach to detect potholes, using accelerometer and GPS data [66]. Mednis et al. described a pothole detection algorithm using accelerometer data obtained using Android based smartphones [136]. In contrast with the mobile sensing domain, the transportation research community has proposed road surface monitoring using camera-based systems [107]. Most of the systems and algorithms described above rely on manually collected ground truth data which serve as training data. Collecting training data this way requires careful planning and experimentation, which typically involves repeatedly driving a set of road segments and manually labeling the location of potholes and other forms of road anomalies. Relying on hand-labeled datasets severely limits the scale, spatial coverage and amount of data available to train classifiers. Therefore, we believe that a crowdsourced platform like *CommuniSense* provides an alternate scalable solution to collect labeled data for developing future automated platforms.

4.2.3 Crowdsourcing in Intelligent Transportation Systems

In intelligent transportation systems (ITS) and traffic research domain, the use of crowdsourcing methods have began to receive attention [20, 103, 154]. In *CrowdITS*, the authors proposed a hybrid system to integrate crowd-based reporting with GPS-based navigation system, to suggest congestion-free routes [20]. In [103], the author advocated the use of mobile social media and collaborative applications to increase social interactions on the road. Crowdsourcing in ITS present exciting opportunities for developing cities, as they lack monitoring technologies to obtain reliable data on traffic and road infrastructure conditions, as argued in Section 4.1. In this regard, our proposed system, *CommuniSense*, advances the research in this direction by integrating citizen-based reporting with online crowd-based verification for quality control.

4.3 Road Quality Survey

Most surveys in Kenya have focused on either traffic congestion [203], travel choice behavior [177] or mobile penetration and usage [225]. To the best of our knowledge, no digitized survey has been conducted to understand the opinions of people on the state of road quality in the context of Nairobi or Kenya in general. We conducted a travel survey in Nairobi with two goals. First, we wanted to understand what Nairobi travelers think of the existing state of road quality in their city. Second, we wanted to gauge their willingness to engage and participate in reporting information on road hazards to support government in urban road maintenance. The survey questionnaire had a series of questions asking respondents about themselves, how they travel on a weekly basis, and how they rate the current road quality conditions based on their daily travel experience. Specifically, our survey had four themes:

- **Mode of transport**: In this section, we asked respondents about their frequency of usage of different transportation modes on a weekly basis. We focused on four transport mode choices: personal vehicle, matatus or bus, taxis, and walking. Matatus are privately-owned informal minibuses that form the backbone of transportation network in Nairobi.
- **Status quo on road quality**: In this section, we explored the current state of road quality in respondents' residential neighborhood, workplace neighborhood, and Nairobi at large, on a five-point Likert scale ranging from *very poor* (1) to *very good* (5). We also asked participants to rate potholes and speed bumps as major road nuisances on a five-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5).
- **Overall impact of road hazards**: In this section, we asked users about the impact of road hazards on their travel discomfort and their personal vehicle's wear and tear (if they owned a personal vehicle). In the survey, we used "road hazards" as an umbrella term to refer to potholes, speed bumps, cracks on the road surface, abrupt pavement changes, or uneven road surface conditions; we made this definition explicit to the respondents.
- **Reporting road hazards**: In this section we quizzed users on their knowledge about how to report a road hazard to the city council, and if they had reported any in the past. We also asked them about their preferred choices and motivations for reporting road hazards.
- **Demographics**: We asked participants about their demographic characteristics (age and gender), living status, and whether they own an Android-based smartphone.

The majority of the survey questions were multiple choice where respondents chose from a list of options. Additionally, we had two open-ended questions where respondents were asked to name their residential and workplace neighborhood (as free-form text). All the multiple-choice questions were mandatory, whereas the open-ended questions were optional. In total, the survey consisted of 18 mandatory and 2 optional questions. Responses were anonymous.

For conducting the survey, we used two different channels: web-based (online) and SMS-based (offline). We used two different channels to account for any demographics bias. On one hand, we believe that an online survey would typically target upper class, upper middle class, and expatriate communities; while on the other hand, a SMS-based survey would cater more to working class and non-smartphone users who typically do not have easy access to the internet. We acknowledge that our surveys are not representative of population of Nairobi as no stratification technique or demographic sampling was applied while selecting users.

4.3.1 Online Survey

In this channel, we used an online platform (Google Forms) to conduct the survey. The survey was distributed via email to mostly university students, and internally within my host organization in Nairobi (IBM Research Africa). In addition to asking the respondents to complete the survey, we also asked them to share the survey on various social media channels (including Twitter and Facebook) to reach a larger audience. We also posted the survey on IBM Africa's Twitter and Facebook official pages. No monetary incentives were provided for answering the survey.

4.3.2 SMS-based Survey

Our second distribution channel was a SMS-based mobile survey platform using mSurvey [9]. mSurvey is a Nairobi-based company which provides a mobile platform to conduct surveys and market research in Kenya. In order to have a wider reach and specifically address a population that does not have easy access to the internet, we used mSurvey's platform, where respondents receive each question per SMS on their mobile devices.

The mSurvey platform used a sample of 500 users randomly selected from their existing worker population in Nairobi. No stratification techniques and demographics filters were applied while selecting the user sample in our survey. Respondents received 40 Kenyan Shillings (equivalent of 0.5 USD) as financial reward to complete the full survey.

4.3.3 Results

For the online survey (gSurvey), we received a total of 442 responses, while for the SMS-based survey (mSurvey), we received a total of 439 completed responses. In total we have a pool of 881 respondents to our survey. In this section we describe the results of both surveys. We focus on five survey themes, which are relevant to the scope of our work.

Demographics

In the online survey, 62% of respondents were male, and 37% of them were female, while the remaining participants chose not to share their gender identity. On the other hand, amongst the mSurvey participants, 58% of respondents were male, and 42% of them were female. For the age distribution, we observe that amongst the gSurvey (resp. mSurvey) population,

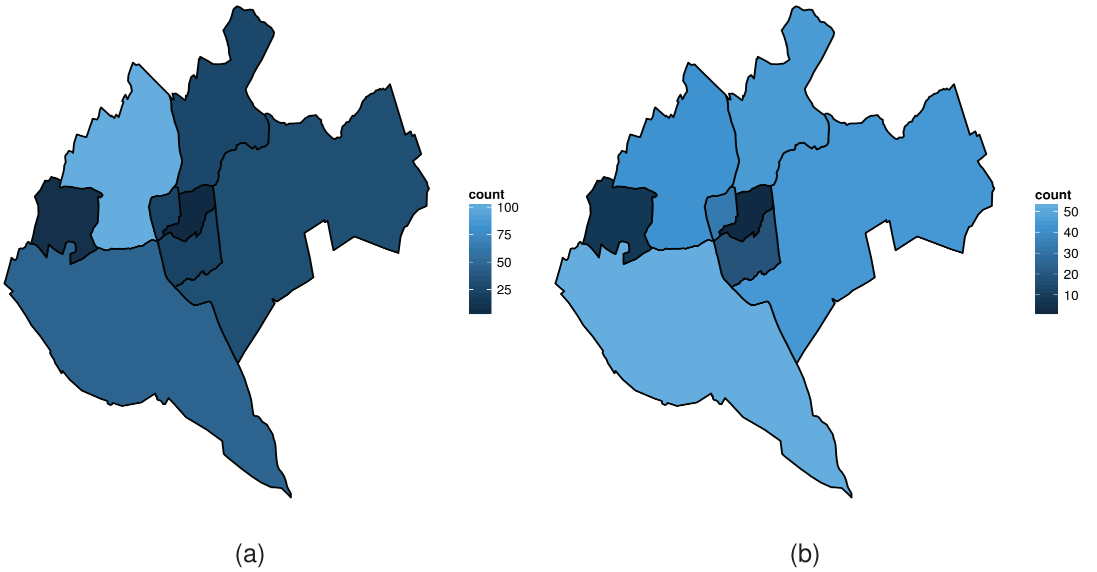


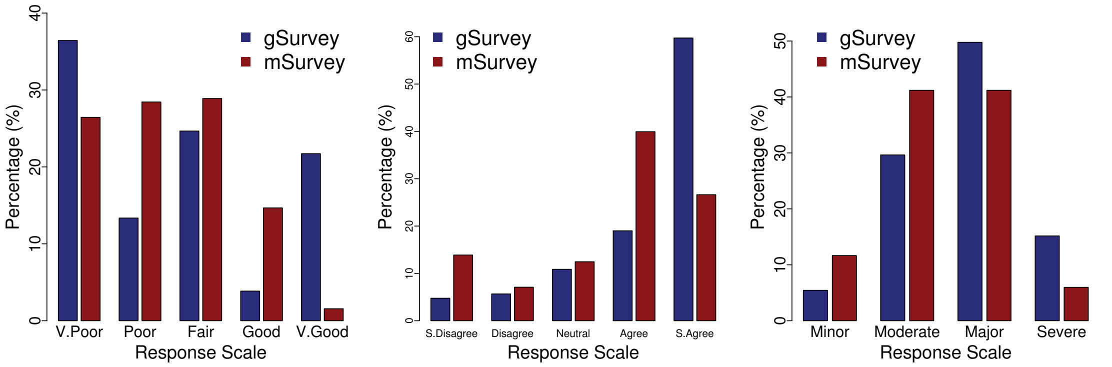
Figure 4.2 – Nairobi residential neighborhoods of respondents for a) web-based survey, and b) SMS-based survey. Spatial information of different administrative areas and neighborhoods of Nairobi are obtained via [4]. Best viewed in color.

23% (resp. 66%) belong to 18–24 age group, 47% (resp. 26%) belonged to the age category of 25–34 years, and 25% (resp. 7%) belong to 35–50 age segment. For both surveys, we did not had a single participant below 18 years old. From these results, it is clear that both surveys cater to different population demographics. In terms of smartphone ownership, 76% of gSurvey respondents owned an Android-based smartphone, while for the mSurvey population 50% of them owned one.

As mentioned in the previous paragraph, participants were asked to list the name of their residential and workplace neighborhood. Of all the users who provided a response, we geocoded their neighborhood addresses to geographic coordinates (latitude and longitude pairs). Figure 4.2 shows the spatial coverage of participants' residential neighborhoods in Nairobi, for both surveys. Based on the local knowledge of the city, we observe that high-income neighborhoods (like Karen, Kilimani and Kileleshwa) are represented more in the web-based survey, when compared to SMS-based survey.

Status quo on road quality

In the online survey, 47% (resp. 30%) of respondents rated road quality as either *poor* or *very poor* in their residential (resp. workplace) neighborhoods (see Figure 4.3a). Consistent with the online survey, the majority of mSurvey respondents 55% (resp. 38%) rated the quality of roads in their residential (resp. workplace) neighborhoods, as either *poor* or *very poor*. Figure 4.3a shows the distribution of road quality in residential neighborhoods across the entire response scale, which clearly highlights that both survey populations find the state of



(a) (b) (c)

Figure 4.3 – Plots showing the histograms for a) Road quality in residential neighborhood, b) Potholes as a major road nuisance, and c) Impact of road hazards on personal travel discomfort, for both online and SMS-based surveys.

road quality at places where they live to be dismal, with a more negative perception for the mSurvey participants.

When survey takers were asked to rate the road quality in Nairobi at large (i.e., not only for home and work neighborhoods), 45% of online respondents found the existing road surface conditions to be bad (*poor* or *very poor*). Surprisingly, only 20% of SMS-based respondents considered the overall Nairobi roads to be in bad condition, with an overwhelming 42% found the roads in good shape (*good* or *very good*). This is in contrast to their perception of their personal neighborhoods discussed in the previous paragraph. There might be some aspirational factors at play here; this would have to be investigated in future work.

Of all the online respondents, 79% *agreed* or *strongly agreed* that potholes are a major road nuisance, while 67% of the mSurvey population acknowledged this fact. Figure 4.3b compares this trend across both populations and the entire scale. 42% of web-based and 29% SMS-based respondents *agreed* or *strongly agreed* that speed-bumps are a major road inconvenience. These findings substantiate our intuition that potholes and speed-bumps are indeed perceived as road hazards, with the SMS population being less sensitive to this issue.

Impact of road hazards

While considering the impact of road hazards, 65% of gSurvey and 46% of mSurvey respondents considered road hazards to cause either *major* or *severe* impact on their personal travel comfort, as shown in Figure 4.3c. Of all the online survey takers who own a personal vehicle, 77% of people considered road hazards to have a *major* or *severe* impact on their vehicle's wear and tear. Note that while asking survey questions in this category, we explicitly defined "road hazards" as potholes, speed bumps, road surface cracks, abrupt pavement changes, or uneven road surface conditions.

Reporting road hazards

Amongst the gSurvey population, 96% of respondents did not know the process of reporting the road hazard to Nairobi's city council. For the 4% who were aware of the process, 55% have ever reported information on a road hazard to the local administration. In contrast, 23% of SMS-based population were aware of the hazard reporting process, and 59% of them have reported one or more of these road hazard complaints.

Furthermore, in this category we also asked respondents about their preferred choice to report road hazards in Nairobi, even if they had never done it before. While asking this question in the online survey, users were given the freedom to choose more than one option as their response. We found that 70% of respondents chose mobile application as their preferred choice, while the second option was to report hazards via social media channels such as Twitter and Facebook (56%).

In contrast, for the SMS-based survey, we formulated this question as a ranking question, where users were asked to rank their top-3 preferred choices. Due to the lack of ranking feature in Google Forms, we formulated this question in the online survey as a multiple choice question with multiple responses. For their top preferred choice, only 4% and 26% of respondents chose mobile application and social media respectively, while 43% of respondents choosing a personal visit to the city council as their top choice to report hazards in Nairobi.

These results point towards a clear difference in the populations; the SMS respondents are probably making their choices based on their most common interaction practice with government (face-to-face) in combination with lower degrees of mobile internet connectivity.

Free-form user comments and feedback

While conducting the online survey, respondents were given the option to voice their opinion and leave comments towards the end of the survey. It was not possible to provide that option to the mSurvey respondents due to the SMS inherent nature on how the survey was conducted. Out of 442 gSurvey participants, 101 left feedback encompassing different topics. We used topic modeling to perform basic content analysis and discover underlying topics from user comments. We used Latent Dirichlet Allocation (LDA) model [31], where each user comment was treated as a single document. After experimenting with different model parameters, the resultant topics did not provide any clear insights due to data scarcity.

As an alternative, we manually coded each comment in order to reveal common concepts and themes. We found that comments varied from general praise for the survey; personal commentary on the current state of traffic situation in Nairobi; negligence and the lack of any hope for a visible feedback from the local city council; and general advice on how to improve the existing situation. In the words of few users, here are some comments we find insightful:

> "*Speed bumps are fine as long as they are marked so you don't just "discover" them with your head on the ceiling and stuff flying in the car. Look at the roads around the hospitals. I am sure a patient is half killed before they even get to the hospital for treatment ...*"

> "*The state of our roads is dismal at best. Networks that were designed for a 90's population are being used, unchanged in the second decade of the 21st century*"

> "*Road hazards is a major cause of road accidents in Kenya that should be addressed.*"

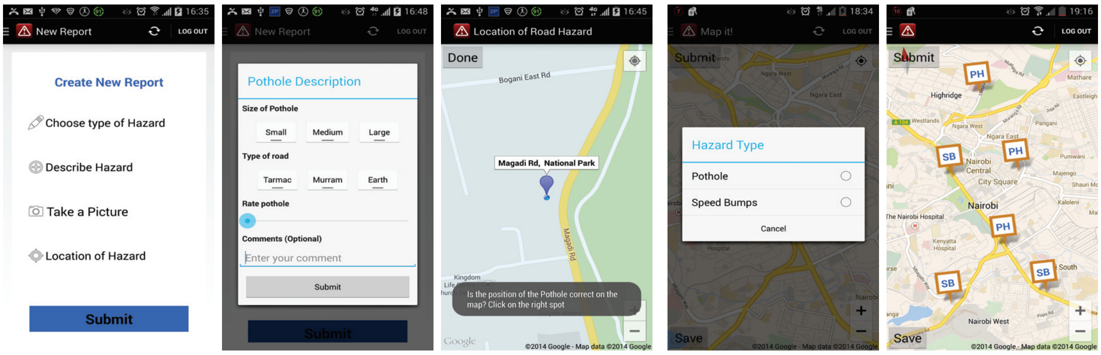
> "*Good job, keep it up. I look forward to seeing a site where I can report hazards and visualize whether the report has been received or not by the relevant authorities, and track of whether reported hazards are being fixed or not.*"

4.4 Mobile Application

In developing cities, the lack of reliable infrastructure, limited connectivity, and inadequate resources make data collection difficult. Paper-based systems are a perennial favorite for city and government administrations. In Nairobi, the city council relies on these systems to handle road quality related complaints as well. However, the reasons that make paper popular are also its liabilities. These paper-based reporting systems lack transparency, accountability, and the speed at which reports are handled is very slow, leaving residents frustrated.

As part of our research, we visited the Nairobi city council engineering offices to learn more about the current reporting system. We found out that residents can use three options to report road hazards: phone calls, postal letters, or walk-in reports. Once the report is submitted, they are sent out to the engineering department for assessment. The engineer goes to the field to assess the hazard reported, takes pictures, documents exact location and severity, and determines how to best fix the hazard. When the engineer gets back to the office, they file a request for supplies which takes time to be fulfilled. This process usually lasts on average between three and six months and in some cases longer before appropriate actions are taken. We designed our crowdsensing solution to improve this current process.

CommuniSense is a mobile crowdsensing application, build on the Android platform, that is designed to collect data on road surface conditions in Nairobi (Figure 4.4). It is a relatively low-cost solution that leverages mobile technology. We chose to build *CommuniSense* on Android since it is cost effective, provides rich programmable interface, offers in-built graphics support and is supported across multiple devices. Using the Android platform, we can collect rich data including multimedia (images), location (GPS) and a myriad of other sensor data (accelerometer, Bluetooth, WiFi, etc.) which is not possible via a typical feature phone. Furthermore, the Android platform automatically optimizes the user interface (UI) experience on each device, while allowing as much control of the UI on different mobile device types.



Is the position of the Pothole correct on the map? Click on the right spot

(a) Hazard Report Submission

(b) Mapping Hazards

Figure 4.4 – Screenshots of the mobile app showing the sequence of stages for (a) Hazard report submission, and (b) Mapping hazards (in the map interface, PH stands for Potholes and SB stands for speed-bumps).

Crowdsourcing the execution of microtasks to a diverse group of people offers unique advantages when combined with a highly motivated pool of workers. From the previously discussed survey, we found that the citizens wants to be engaged and are willing to participate in data collection. As per a travel survey conducted in [79], the mode of daily commutes in Nairobi is 47% by walking, 29% by matatu or mini-bus, 15% by private auto, 7% by other buses or shuttles, and 1% by other modes of transportation. This distribution avails a large segment of commuters that can provide manual reporting via *CommuniSense*. We are well aware that our smartphone-based approach presents constraints to our collection methodology in terms of reaching a much wider audience. However, recent trends in smartphone penetration and subsidized cost of smartphones in Kenya, demonstrate that they can be used to achieve sufficient data diversity [148, 204, 205].

This data collection platform provides us with hand-labeled hazard locations for two purposes. First, the geo-referenced images are valuable to document road hazards when displayed on a map. Second, we plan to use the geo-located data as training data for future work to detect and locate hazards using other phone sensor data, as done in other recent work [66, 138, 30]. The mobile application provides users with two reporting options which are described below.

4.4.1 Hazard Report Submission

In this option, users submitted a completely documented report which included the type of road hazard, its description (including hazard's severity and road type), a picture showing the hazard, and its corresponding location (Figure 4.4a). To capture the location of the hazard, GPS sensor was triggered as soon as the user started the application, so that when the user submitted the report, we automatically captured hazard's location.

While GPS provides accurate location estimates in the order of few meters, it suffers from few limitations, including urban canyon errors due to bad radio reception in areas surrounded by tall buildings (applicable in Nairobi downtown) and cold-start problems which result in inaccurate location estimates when a device is initially switched on. During some initial field tests, we found that the GPS coordinates in urban areas were in most cases off by 100–250m. As a result, the mobile application gave users the functionality to update the location of the hazard (relative to its GPS-inferred location) by clicking and dragging the marker on the map, as shown in Figure 4.4a.

In Nairobi, mobile data is relatively expensive, and in remote areas the signal strength is weak to sustain a reliable data connection. Consequently, when the user was done completing the report they were given two upload choices. Users could immediately submit the reports or save them locally on the device and upload them later, when reliable mobile data connection or access to Wi-Fi was available. Note that if the user decided to submit from the field and the data failed to get to the servers, it was automatically saved locally.

The reason to provide the offline reporting functionality was motivated by an initial discussion with a small set of commuters who raised concerns over the cost associated with uploading an image from the field. As a result, we performed image compression as a second mean to reduce the costs associated with transferring the hazard report. We used Android's in-built `base64` encoding for image compression. When a user took an image, we compressed the image locally on the device and then sent the compressed image to the backend server for storage. While compressing images, we took into account the orientation of the mobile device (portrait or landscape mode) and its native resolution.

4.4.2 Mapping Hazards

The second option (*MapIt* in short) provided users with a quick way to report the location of a road hazard. Users were shown a map interface centered and zoomed to their current location (inferred from GPS). A click on the map prompted a dialog with a list of different hazards. Once the user selected a hazard name, a marker was placed on the map. As with the complete reporting, the user could adjust the location by clicking and dragging the marker. Figure 4.4b shows the data capturing process.

The idea behind this option was to give users the flexibility to report road hazards who were unable to fully document them during their commute. Drivers and commuters typically have intimate knowledge of the routes they frequently take and therefore can offer insights on the road surface conditions at a later time. This is a way to utilize the local knowledge of people to document road quality conditions in areas they are most familiar with, without requiring them to go through the process of submitting a complete report.

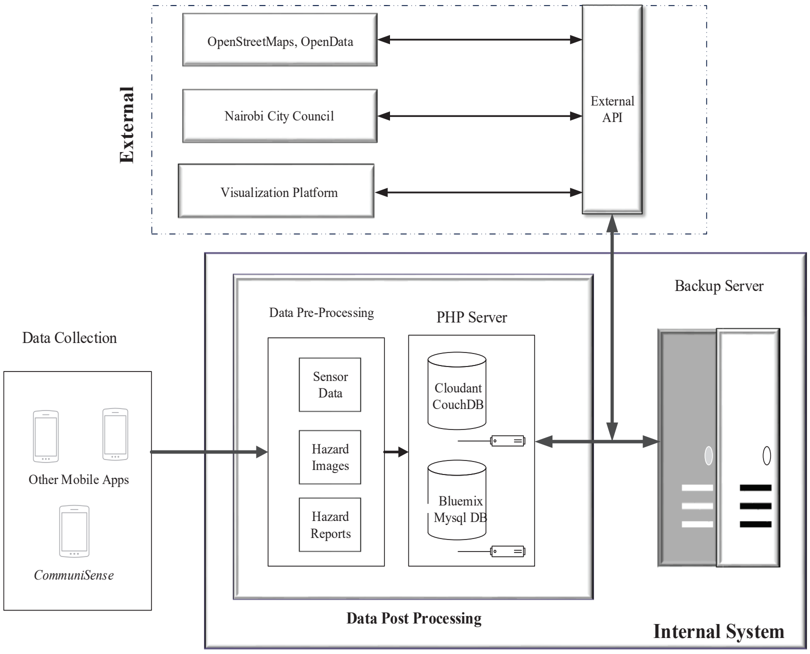


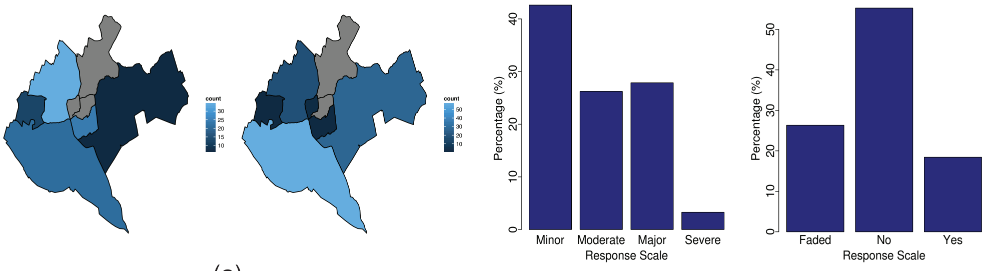
Figure 4.5 – CommuniSense Backend Architecture

4.4.3 Backend Architecture

The application was linked to a cloud PHP server that handled user authentication, received reports, and handled all device to server communications (Figure 4.5). Users were required to create an account using their phone number. We anonymized the phone number and other identifying information to maintain user privacy. A unique `userID` was generated and associated with any activity between a user and the server. The hazard metadata was saved in a MySQL database and images were saved as binary large objects.

4.5 Field Study

To test the functionality of *CommuniSense*, we conducted a two-week pilot user study. We performed a limited release of our mobile application to a selected number of participants. The application was published on Google Play Store but it was not open to everyone, as we wanted to test the functionality of the app with a limited set of users, before making it open for everyone. We restricted the access to our app via a private Google Plus community. Only users who were part of the community had access to the *CommuniSense* application. To enroll participants in our study, we emailed 150 users, mostly college students from local universities and visiting students. Once a participant showed their willingness to be part of the study, we invited the participant to join our Google Plus group. After joining the group, participants had access to download and install the application on their devices. To motivate the users, we promised to award 500 Ksh (equivalent of 5.5 USD) to the top five contributors



(a)

(b)

Figure 4.6 – Results from the field study (a) Spatial coverage of user-contributed submissions for full reports (left), and mapping hazards submissions (right). Regions colored gray do not have any user contributed submissions. (b) Histogram showing hazard attributes for severity of potholes (left), and labeling of speed-bumps (right), as reported by field users.

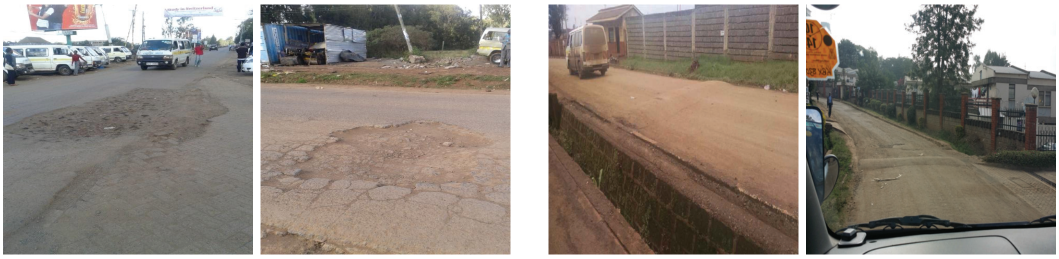
towards the end of the study. The top contributors were chosen based on the maximum number of legitimate and unique reports covering different neighborhoods in Nairobi.

During our field experiment, of the 150 email invites sent, we had a total of 41 users who accepted our invitation to join the Google Plus community. Out of those 41 users, 30 installed the application (20% response rate). During the two weeks of the trial, we had a total of 101 full report submissions and 153 *MapIt* submissions. Of all the full reports, 62% submission were of potholes, and the remaining 38% were of speed-bumps. Of all the *MapIt* submissions, 42% submission were of potholes, and the remaining 58.17% were of speed-bumps.

Out of 101 full reports submitted, 99 of them came from Nairobi county (61 potholes and 38 speed-bumps), while for the *MapIt* submissions, 109 came from Nairobi county. Figure 4.6a shows the spatial coverage of the reports from within Nairobi city limits. One can observe that there are at least a few reports from most neighborhoods. although some of the regions are missing from our field experiment. Of the 61 full reports of potholes submitted, we observe that 43% of potholes were rated *minor*, and 31% were rated either *major* or *severe* on the severity scale, as shown in Figure 4.6b. Note that field users rated the severity of potholes on a 4-point scale ranging from *minor* (1) to *severe* (4). When observing the speed-bumps, we note that 55% of field users encountered an unlabeled speed-bump (Figure 4.6b).

4.6 Image Verification Experiment

Crowdsourcing offers opportunities for people to supplement their income in developing countries. However, the openness of access to crowdsourcing platforms often leads to malicious and spam behavior, and sometimes sabotage. As an example, for the well-known DARPA network challenge, the winning entry received 80% of malicious submissions [195]. In another example Ushahidi, the crowdsourcing platform for social activism and crisis mapping, shut down their operations during the 2011 Arab spring due to growing concern



(a) Potholes

(b) Speed Bumps

Figure 4.7 – Sample images collected from the field study. Two random images from our dataset reported as (a) Potholes, and (b) Speed Bumps. Images showing faces and license plate numbers have been blurred or masked. Best viewed in color.

that governments official might use the platform to track the activities of people [143]. For most of the tasks available on crowdsourcing services like Mechanical Turk, Crowdflower, or MobileWorks, financial incentives need to be in place to motivate the worker population to participate. However, when the crowdsourcing task involves monetary incentives, users might put in only minimum effort to secure the financial reward. As a result, quality control in crowdsourced data cannot be neglected. This presents several research challenges.

In our case, we had to verify that a) a contributed submission indeed depicted the road hazard as reported by the user, and b) it was located at the claimed location. The second verification task is conceptually feasible as the location information was automatically captured via the GPS location sensor most times. However, the authenticity of the reported road hazard and its details is more subjective to evaluate. We present a crowdsourced approach for this in the next subsection.

4.6.1 Crowdsourcing Image Verification

We designed and conducted a crowdsourcing study to assess whether the images obtained via the mobile application can display road hazard properties. As done in previous chapters (Chapter 2 and Chapter 3), we used MTurk for online crowdsourcing. We chose US-based "Masters" workers with at least 95% approval rate for historical HITs. For each HIT annotation task, the annotators were shown one image and asked to classify the image as either pothole or speed-bump. Given the annotators' choice, they were further asked to describe the chosen hazard. If the user categorized an image as a pothole, users were asked to further describe it in terms of size and severity. For size, users were given the option to choose from *small*, *medium* and *large*; for severity, users were given a four-point scale ranging from *minor* (1), *moderate* (2), *major* (3) and *severe* (4). If the user chose speed-bump as the option we asked them to describe its size, number of bumps, and whether the speed-bump was labeled or painted. If the user was unable to classify an image as either containing a pothole or a speed-bump, then an option was given to mark whether the image contained both a pothole and a speed-bump; or showed uneven or cracked road surface; or the image showed a smooth road surface. For the MTurk experiment, we randomly chose 50 images from the set collected in the field experiment. We collected 10 different annotations for each image. Consequently, we collected a total of 500 responses for every question. Every worker was reimbursed 0.15 USD per HIT (i.e., per image)

The questions asked to describe the hazard were identical to the questions shown to users reporting from the wild i.e., using our mobile application. For these questions, no explicit definitions of a pothole or speed-bump were provided, so online annotators needed to rely on their internal representation. All the images shown to the users were anonymized. To the best of our ability, we avoided images where one could potentially identify faces or skin color, to protect the privacy of individuals and reduce any potential bias while characterizing the road quality. Moreover, we ensured that images that showed the license plate numbers or any other information that could explicitly reveal the identity of the city under study were masked e.g., an image showing street banners with the word Nairobi in it. Image examples are shown in Figure 4.7.

4.6.2 Results

In this section we present the results of our image verification experiment.

Completion Rate

For the MTurk experiment, we had a pool of 39 workers who responded to our HITs. For a total number of 500 HIT assignments available in this experiment, we observe that a typical worker completed an average of 13 HITs, while they could potentially undertake 50 HITs. One worker completed the highest number of 41 HIT assignments. We observe a typical heavy tail-like distribution in HIT completion times (mean: 37.8 secs, median: 21 secs, max: 290 secs). Each HIT was allocated a maximum of 5 minutes for completion.

Image Label Quality

Aggregation was used to create a composite score per image given the 10 different responses for each question. We explore two different aggregation techniques. The first one is the *majority vote* where we computed the majority score given the 10 annotations for each image. The second one is the *median* method, where we computed the median across the 10 annotations for each image. Table 4.1 lists the summary statistics for both aggregation methods. For each aggregation technique, we computed the total number of correctly classified images for both road hazards where the consensus between the MTurk population and the field user matched. Out of 50 images which were verified via MTurk, 34 were verified as potholes, and 12 as speed-bumps, where the MTurk population and the field user labeled the image in the

Method	Potholes	Speed Bumps
Majority Voting	34 (100%)	12 (75%)
Median	34 (100%)	12 (75%)

Table 4.1 – Table showing summary statistics for aggregation methods. For each method, we show the total number and percentage (shown in brackets) of correctly classified images for both road hazards i.e., where the ratings between the MTurk population and the field experiment were identical.

	Min (Max)	Mean (Median)	$ICC(1, k)$
Size of potholes	1.0 (3.0)	2.16 (2.0)	0.90
Severity of potholes	1.0 (4.0)	2.26 (2.0)	0.91
Size of speed-bumps	1.0 (3.0)	1.92 (2.0)	0.73

Table 4.2 – $ICC(1, k)$ scores of hazard attributes (All values were statistically significant at $p < 0.01$.) Mean and median (in brackets) values of each hazard attribute is also shown.

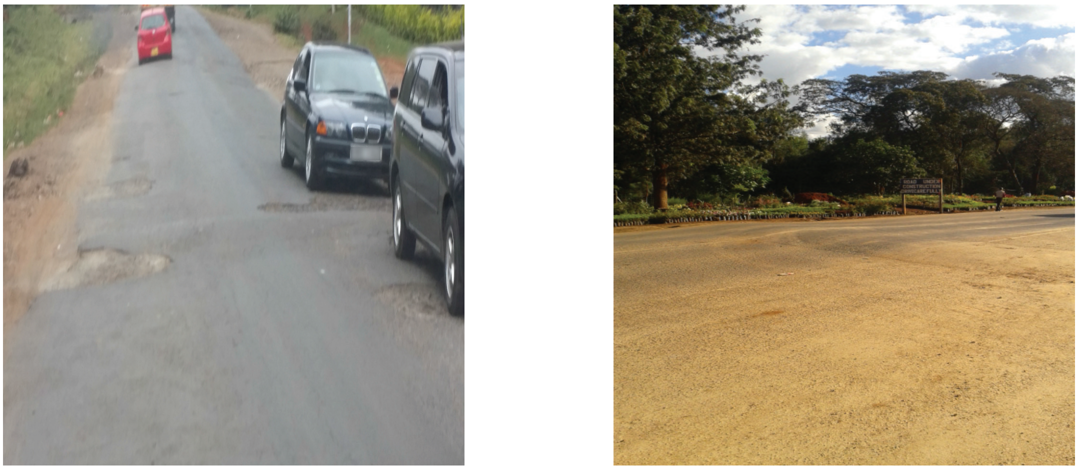
same category (see Table 4.1 and image examples in Figure 4.7.)

In terms of agreement with the mobile app user, 92% of images were verified with the same label as reported by the user, i.e., 46 images out of 50. Four images were labeled as ambiguous. Based on manual inspection, we found that two out of those four images contained both a pothole and a speed-bump (Figure 4.8a shows an example); while the remaining two images contained an unlabeled speed-bump which was not clearly visible, and hence was classified as ambiguous (Figure 4.8b demonstrates an example of this type).

Now we turn our focus towards assessing the reliability of annotations for hazard attributes (e.g., severity of potholes, size of speed-bumps, etc.). Note that in addition to asking users about image category, we also asked them to describe the attributes of the chosen hazard. As done previously, to measure the inter-annotator consensus for hazard attributes, we computed the intraclass correlation ($ICC(1, k)$), given the 10 annotations per image. Table 4.2 reports the $ICC(1, k)$ values for all correctly verified images (i.e., 46 out of 50 images). Table 4.2 lists the ICC values for three key hazard attributes: size and severity of potholes, and size of speed-bumps. We observe high inter-rater reliability for all hazard attributes, with all the scores being statistically significant (p-value $<$ 0.01). Similar results were obtained for other hazard attributes. These results highlight the potential of using a crowdsourcing approach as means to verify the authenticity of the reported road hazard and its attributes.

4.7 Visualization Framework

Our visualization framework is a web-based application which provides a layered and an interactive (zooming and map navigation) interface, where geo-localized information from



(a) (b)

Figure 4.8 – Misclassified images where the consensus differed between the MTurk population and the field user. Images showing faces and license plate numbers have been blurred or masked. Best viewed in color.

varied data sources is overlaid on top of the base map layer in an interactive fashion. It was developed using existing open-source web technologies built on top of OpenstreetMap (OSM) data. The framework follows a layered architecture where the underlying base layer (or map layer) consists of map data from OSM, while additional layers are overlaid on top of the base layer. Using this framework, we visualize the location of road hazards in Nairobi in Figure 4.9. The location of road hazards (potholes and speed-bumps) was contributed by the field users.

The visualization platform is data agnostic and any spatial information can be rendered as an additional layer. Layers can be rendered in their raw form (latitude/longitude pairs) or visualized in processed form (e.g., heatmaps), as in Figure 4.9. Moreover, the platform has been designed to handle large-scale datasets. The framework has been presented using Nairobi as the use case but it can be easily extended to any other city with minimal changes.

Besides the purpose of the visualization interface to provide a platform for local Nairobians to browse through the crowdsourced data, we believe that it can serve as a platform to engage citizens, increase awareness and initiate a public dialogue on the state of road quality in Nairobi. The visualization platform was designed to give the citizen-contributed data back to the community which helped create the data at the first place. In the process, the platform could facilitate a reliable, independent source of information about potholes and speed-bumps that can be used to alert municipal officials and allows citizens to monitor progress in resolving these hazards.

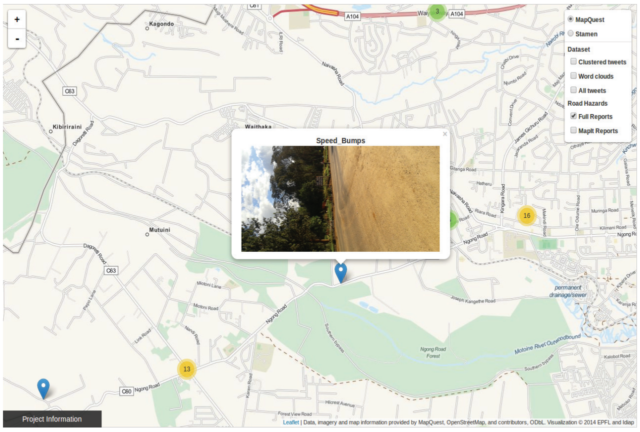


Figure 4.9 – Visualization Framework

4.8 Discussion

In this section, we first describe the technical challenges and lessons learned while deploying *CommuniSense* in Nairobi. Then, we assess the potential of social media as an alternative medium to obtain road hazard datasets. We conclude this section by discussing *CommuniSense*'s possible role in promoting citizen engagement in Nairobi.

4.8.1 Technical Challenges

During the field study, we faced three major technical issues. First, due to the myriad of affordable grey market devices, we found that certain devices did not handle the mapping and location functionality well. As a result, users found it difficult to interact with the location marker on the map (Figure 4.4a). Second, we observed that a significant number of smartphones were still using older versions of Android (2.2 and 2.3). These versions required a different UI design, when compared to the Android version (3.0 and above) on which *CommuniSense* was developed. Although, Android provides backward compatibility, certain devices were not able to render the UI properly, causing inconvenience to users when interacting with the app. Third, for some user-submitted reports, the `base64` compression caused loss of image quality, dependent on the way the device was oriented (landscape vs. portrait mode).

We used Google Play store to deploy *CommuniSense*. The processes associated with performing a limited release of the app using Google Play store, proved to be daunting for non-technical users. The process required participants to be added to a private Google Plus (G+) community. Access to G+ requires users to have a GMail account and activate alerts from their G+ profiles. Note that only those users who were part of our private G+ community, were given access to the app. Many users complained of not receiving the G+ invitation, only to discover they had not activated alerts on their G+ profiles. Even when the user successfully became part of the G+ community, they cannot search and install the app via the play store. The only way to install the app is to click *Install* on the web-interface, which then prompted the app to be installed on the device automatically (only when the device was connected to the internet). This is not the typical way users install mobile apps and so this process created confusion among early users. We believe that more work needs to be done to simplify, and streamline the process of conducting limited app release distribution via Google Play and other app distribution channels.

4.8.2 Comparison with Social Media

As discussed in Section 4.2.1, there exist systems which allow citizens to report civic issues (e.g., SeeClickFix [17], FixMyStreet [16], Citizens Connect [1], etc.), but none of these systems exist for Kenya. Due to the lack of any real-time traffic monitoring and broadcast systems, one of the systems which has gained popularity in Kenya and Nairobi in particular, is *ma3route* [6]. ma3route is a mobile and web platform that allows citizens to report and share information on existing traffic conditions in their city. *ma3route* publishes all user submissions on their Twitter channel [6]. As of writing (August 2016), *ma3route* has more than 495K followers and has posted 506K tweets that contain in excess of 83K images and videos.

To examine the potential of social media as an alternative medium to obtain road hazard datasets, we manually coded the most recent 300 tweets from *ma3route*'s Twitter feed (most recent date: February 2, 2014). We found that 45% of tweets contained information on traffic conditions and jams, 7% described road accidents, 8% of tweets reported street protests and how they were impeding the traffic flow, 2% of tweets reported road hazards, and the rest 38% discussed other topics (e.g., corruption, high fuel prices, suggestions to improve infrastructure, etc.) Out of 300 tweets, 81 (27%) of them contained an image. Only seven tweets in our sample contained information on road hazards, and out of those seven tweets, only three of them (1%) posted road hazard information with an image. Based on these findings, even though Twitter as a data collection medium looks promising, but it currently lacks the spatial coverage and topical focus offered by specialized mobile crowdsensing. We plan to investigate the role of Twitter to collect road hazards data, as part of future work.

4.8.3 Citizen Engagement

Nairobi residents have been frustrated and lost faith in the city council's ability to improve road conditions. The words of few users highlight this sentiment (obtained via online survey as described in Section 4.3):

> "*Anything to do with city council would require a major overhaul of the personnel. Otherwise this would not be possible.*"

> "*And I do not trust that the city council would take our complaints seriously. They first need to fix the roads properly instead of patching them up year after year!*"

> "*Actual or visible feedbacks would motivate me to report even paying some costs.*"

> "*I would only report road issues if I thought something would be done about it. I am not sure that's currently the case.*"

These sentiments are shared among residents in many developing cities. We believe that the city council could benefit by leveraging the data collected by *CommuniSense*. The design of this application provides a channel to gather direct input from citizens on the condition of urban infrastructure. This would save time and money involved in manually documenting road hazards, as currently done by government engineers. The platform would also offer a mechanism to engage users into reporting hazards as well as providing accountability structures to show residents that their tax money is being used effectively.

4.9 Conclusion

In this chapter, we examined the use of mobile crowdsensing to document Nairobi's road quality information. First, we presented the key findings of a road quality survey in Nairobi. The survey examined key local issues including weekly travel practices, perception of current road quality conditions and their impact on daily travel experience. Second, we developed a mobile crowdsensing application to locate, describe, and photograph road hazards. We tested the application through a two-week field study amongst 30 participants who documented a total of 254 road hazards from different areas in Nairobi. Third, we demonstrated the use of online crowdsourcing to verify the authenticity of user-contributed reports i.e., to verify whether contributed images from the field indeed depicted road hazards. Overall, *CommuniSense* advances the research in the domain of citizen-based reporting, by integrating it with online crowd-based verification for quality control.

To conclude, we believe that the effectiveness of existing governance systems can be substantially enhanced by applying mobile crowdsensing solutions, which facilitate real-time data collection, categorization, verification, and dissemination. As developing countries start looking forward towards improving social welfare and quality of life, it is important to funnel meaningful feedback from community stakeholders on local needs [34].

5 Crowdsensing of Urban Nightlife Patterns

5.1 Introduction

UNDERSTANDING nightlife, i.e., how cities and their inhabitants interact at night, is a relevant issue to multiple stakeholders including city officials, business associations, police departments, health and educational authorities, and NGOs. A vibrant nightlife scene can be simultaneously seen as an urban development strategy, an economic opportunity, a source of health and safety risks, and a way in which citizens co-create and appropriate the urban space. [212]. Young people, including older teenagers and young adults, are key actors of nightlife, and as such become the focus of many of the above stakeholders, with respect to the design of strategies and policies that encourage the appropriation of the urban space while promoting healthy behaviors.

We posit that Ubicomp research can contribute to the understanding of nightlife as experienced by young people, who happen to use mobile and social technologies day and night. This can provide social and health scientists with ecologically valid, in-situ contextual data that can provide information about venues, mobility, activities, and social patterns, with high spatial and temporal resolution, and potentially at scale. Moreover, ubiquitous computing adds the possibility of collecting sensor data and media in addition to more traditional mobile survey based data (e.g. via experience sampling done in a number of disciplines including social psychology, epidemiology, etc).

In cities, mobile crowdsensing provides the possibility to study questions related to populations and their environments that have been elusive in the past. The engagement of mobile crowds to collect everyday life data using smartphones has followed two main directions in the literature. One approach relies on crowdsensing, i.e., the use of sensors in mobile devices to collect data (either at pre-defined regimes or opportunistically) without requiring human intervention [115, 45]. While many sensor data types (location, accelerometer, WiFi) can be reasonably processed using data coming from smartphones in pockets or bags, other sensors like camera and microphones suffer from it. The second trend in the literature involves requesting explicit actions from crowdworkers, including photo-taking and audio recording, where the sensors are unoccluded for data collection [112, 227, 186].

In this chapter, we present the development and implementation of a mobile crowdsensing study, called *Youth@Night*, to capture and examine the nightlife patterns of youth populations (aged between 16–25 years) in Switzerland. The study was funded by the Swiss National Science Foundation and consisted of a multidisciplinary team from Addiction Switzerland, Research Institute (Emmanuel Kuntsche and Florian Labhart), Department of Geography at the University of Zurich (Sara Landolt and Jasmine Truong), and the Social Computing Group and the development team at Idiap Research Institute (Flavio Tarsetti, Olivier Bornet, and Hugues Salamin).

For the study, we developed two smartphone applications that captures data on where young people hang out, their social context, and their activities during night time. The field study was conducted in two Swiss cities (Zurich and Lausanne), which are recognized as the two national nightlife hubs among young people in the German and French-speaking regions of the country, according to city officials [132, 145]. The study served two key objectives. First, to capture the heterogeneity and complexity of going out behavior during night time: the places where youth spend their weekend nights (physical mobility), the activities they perform (consumption of alcohol), and the people they hang out with (social context). Second, to examine the feasibility of using mobile sensor data to automatically characterize alcohol consumption behavior of young adults in an urban, in-the-wild nightlife setting. Our research questions are the following:

RQ5.1: How can mobile crowdsensing be used to study nightlife patterns of urban youth populations? How can the collection of mobile videos be integrated as part of this crowdsensing task? (**RQ5.1** maps to thesis's **RQ1**)

RQ5.2: What are the places and social contexts in which young people hang out? How is the use of private and public places distributed among youth? (**RQ5.2** maps to thesis's **RQ1**)

RQ5.3: Can alcohol consumption be automatically inferred from mobile sensor and log data in an uncontrolled, real-life setting of youth's nightlife activity? If so, what sensor features are more predictive of alcohol consumption? (**RQ5.3** maps to thesis's **RQ4**)

RQ5.4: To what extent do automatically extracted ambiance features represent the crowd-sourced annotations by both in-situ observers and external online observers? Do crowdsourced annotations by external observers correspond to in-situ self-reports? (**RQ5.4** maps to thesis's **RQ2** and **RQ3**)

The chapter makes the following contributions:

1. We developed two custom Android-based smartphone applications: a survey logger and a sensor logger application. These applications respectively allowed participants to respond

to various surveys using a participatory approach, while at the same time collect sensor and log data in a non-intrusive and privacy-preserving manner. For each place visit (check-in), the application also requests users to self-report their alcohol consumption (including type, size and alcohol volume) and capture a video providing a panoramic view of the checked in place.

2. We conducted an “in-the-wild” study with more than 200 participants aged 16–25 years old in two Swiss cities (Zurich and Lausanne), which are recognized as the two national nightlife hubs among young people in the German and French-speaking regions of Switzerland. Each participant contributed data through the use of the mobile applications over multiple weekends between 8PM and 4AM on Friday and Saturday nights for three months (Section 5.3). As part of the design, we used a data-driven approach to define recruitment zones based on their nightlife activity using Foursquare data, to increase the likelihood of finding potential study participants.
3. The crowdsensing study resulted in a collection of close to 1,400 place visits accumulating over 8 million sensor data points over 2,934 user-nights, and 894 videos that spread across different place categories, and diverse social and ambiance settings. Private places were self-reported to be more bright, and less crowded compared to public places. Bars and clubs were reported to be relatively more crowded, louder, and darker compared to the rest of public place categories.
4. We build a machine learning methodology to infer whether a participant consumed alcohol in a given night, based on sensor and application logs, and determine what are the most informative phone-derived cues. We found that accelerometer data is the most informative single cue, and that a combination of features results in an overall accuracy of 76.7%.
5. Using the video data, we examined the reliability of automatic feature extraction relative to different crowd-workers (in-situ and ex-situ). Automatic feature extraction was performed to infer the ambiance (i.e., loudness and brightness) of documented places at scale. We found that these features were reliable with respect to the video content, but that the videos did not always reflect the place ambiance reported in-situ. We found that automatically extracted ambiance features described more accurately the perception of ex-situ annotators than that of in-situ participants.

We believe that the developed methodology provides an attractive opportunity to improve our current understanding of patterns of physical mobility, activities, and social context of youth population, as they experience nightlife.

The chapter is organized as follows: Section 5.2 presents a review of related work along multiple research axes. Section 5.3 outlines the design of the two mobile applications, the recruitment of participants, and the protocol as part of the crowdsensing study design. The mobile application was developed by Flavio Tarsetti and Olivier Bornet, while the recruitment was carried out by a team of research assistants, led by the partners in Addition Switzerland and University of Zurich. Section 5.4 introduces the data collection framework to gather different data types including questionnaires, in-situ surveys, sensor and media (video) data. Section 5.5 presents a descriptive analysis of the questionnaire and survey data to understand the demographics of the participants, the places they visit and the contextual factors surrounding their place visits. Section 5.6 proposes a machine learning pipeline to automatically infer a binary state of alcohol consumption for single nights, describing the various steps including the pre-processing of sensor data, feature extraction and regression analysis. Section 5.6 was a joint work with Trinh Minh Tri Do of Idiap Research Institute. Section 5.7 uses the video dataset to investigate the potential of automatically extracted features to estimate the loudness and brightness of places at scale. Section 5.7 was a joint work with Joan-Isaac Biel of Idiap Research Institute. Section 5.8 describes the overall experience of participants during the study including compliance with the video recording task, and their experiences as shared on the qualitative interviews. Qualitative interviews were conducted by Jasmine Troung and Sara Landolt. Finally, we discuss and conclude with a summary of findings in Section 5.9. This chapter outlines some of the research that has been published here [181].

5.2 Related Work

Given the multifaceted nature of our work, we review the related work along four domains: mobile sensing for data collection and its applications to pervasive healthcare, mobile crowdsourcing, computational modeling of places, the intersection of urban nightlife, youth studies and alcohol epidemiology.

5.2.1 Mobile Sensing for Data Collection

In mobile sensing, few groups worldwide have collected mobile sensor data that is at the same time rich, longitudinal, and that covers a large population. One of the earliest work was done as part of the MIT's Reality Mining initiative [65]. Another was the Nokia Mobile Data Challenge in Switzerland, which showed the feasibility of collecting continuous smartphone data from 200 users over one year [115, 125]. Most mobile sensing studies have focused on gathering sensor data including accelerometer, GPS, WiFi, Bluetooth. Other studies have also collected perceptual data including audio and still images for place characterization [46, 214, 130], life-logging [91], visual perception [186], etc. In contrast, fewer crowdsensing studies have collected visual data in the form of mobile videos. A recent study has proposed a crowdsourcing framework to acquire and transmit mobile videos under resource constraints for disaster response scenarios [200].

5.2.2 Mobile Sensing for Ubiquitous Healthcare

Research has also started to examine the role of smartphones and wearable devices to monitor health-related variables. In these domains, the recognition of stress levels has been an area of active research. In *StressSense*, the authors identified stress rate from human voice recorded using smartphones with an accuracy of 76% and 81% resp. in outdoor and indoor environments [129]. Another stress related study was conducted using five days of data from 18 participants in [180]. Using features derived from a variety of in-built smartphone sensors including accelerometer, mobile phone usage (call, SMS, location and screen events), their method achieved an accuracy of 75% in classifying the stress levels of participants (under stress or not). Recently, the authors have developed a stress model (called cStress), to establish a gold standard for continuous stress assessment in the mobile environment [92].

Another application which has received attention in ubiquitous healthcare research is the detection of smoking gestures [196, 19, 156]. In [196], Tang et al. conducted a study with six participants, all wearing activity monitor devices in their wrists data, to automatically detect puffing and smoking behavior. The authors developed a detection model incorporating temporal and high-level smoking topographic features. In a similar work, Parate et al. proposed to capture changes in arm orientations to detect smoking gestures in real-time [156].

In the field of alcohol consumption detection, there have been relatively few works. In [105], the authors proposed a phone-based system to detect the gait anomalies of walking under the influence of alcohol. Based on accelerometer data from three participants, the authors extracted gait features to differentiate intoxicated walking patterns from regular patterns. In the domain of drunk driving detection, a study on early detection of drunk driving using a smartphone was described in [51]. Using accelerometer data, the authors developed a mobile application to detect whether the driving patterns match with the cues for drunk driving gathered from real driving tests (such as lateral acceleration, lane positioning, speed control, etc.). To support people with alcohol addiction, Wang et al. proposed SoberDiary, a phone-based support system to help users monitor and and manage their alcoholic intake and remain sober in their daily lives [215]. The study was conducted on 11 clinical patients over 4 weeks, using a bluetooth-enabled breathalyzer that was paired with smartphones. The authors found SoberDiary system helped patients reduce their alcohol consumption.

5.2.3 Mobile Crowdsourcing Platforms

The wide spread adoption of mobile devices has lead to the emergence of mobile marketplaces where mobile users are paid to perform tasks in the physical world. In these marketplaces, mobile users are asked to perform tasks which are characterized by users' physical mobility (location-based) or their real-time nature (e.g., surveys, performing household chores, etc). Mobile marketplaces are different from online crowdsourcing platforms like Mechanical Turk or CrowdFlower, which do not impose these physical constraints. Notable mobile marketplace platforms include GigWalk [5], TaskRabbit [11], and FieldAgent [15]. Following the rise of these platforms, recent research has examined the practices and dynamics of mobile marketplaces [141, 197, 198]. On one hand, these platforms provide relatively easy access to on-demand workforce (both online and mobile), however it is difficult to recruit crowdworkers using these platforms for a sustained period of time. Consequently, in some of the large-scale mobile sensing campaigns including in our current work, researchers themselves recruit participants to gain a more fine-grained control on demographics, location, study duration, etc. [46, 115, 65].

5.2.4 Automatic Place Characterization

Previous work has modeled places automatically using data obtained from mobile sensors [46, 214]. In [214], the authors used automatically extracted features from audio signals to infer the level of occupancy, human chatter, music, and noise of places. In [46], the authors addressed the task of place categorization based on the automatic processing of opportunistically captured audio signals and still images. In this regard, our work is closely related to work by Chon et al. [45], who carried out a two month deployment of a crowdsensing platform to collect 48,000 place visits from 85 participants in Seoul, to examine the coverage and scalability of place-focused crowdsensing.

Our research differs from prior work in three aspects. First, in addition to capturing similar mobile sensor modalities, we collected videos of places combined with location and time based surveys to gain additional context. As we show, videos provide a highly detailed window into the physical and social experience of the participants, closer to “being there” than what other sensors can provide. Second, the place data was collected using a mix of opportunistic and participatory sensing, as opposed to a purely opportunistic approach [46, 214]. Our methodology facilitated the continuous collection of sensor data in addition to in-situ survey and media data to enrich the contextual information about places. As we explain later, the intentionality of our crowdsourcing task also allowed to study issues related to the perception of social acceptability of mobile video recording in everyday life. Finally, our study covers a much larger geographic area than [45, 217, 214], including two cities with linguistic and cultural differences, but also many areas around each city.

5.2.5 Urban Nightlife, Youth and Drinking

In urban studies and human geography, researchers acknowledge that little attention has been given to understanding the dynamics surrounding youth experiences and urban nightlife [212]; this is certainly so in Ubicomp research [201, 89, 35]. Activities during daytime have often been the primary source of investigation to understand topics ranging from human mobility and experiences, to how spaces are used and regulated in urban areas [134, 43, 192, 94, 191]. In these domains, urban public spaces often are seen as ‘adult space’, where young people and their practices are seen as dangerous, threatening the ‘adult’ order in public spaces as well as the safety of others [210, 192, 211]. Youth are often stereotyped as “trouble-makers” creating public disturbance and disorder particularly during night time, resulting in increased surveillance and regulations [210, 211, 212]. Less attention has been given to understanding the dynamics surrounding youth experiences in urban spaces at night [212].

There is a significant body of work at the intersection of youth, drinking, and urban spaces, where researchers have examined drinking places including pubs and bars [63], house parties [99] and public spaces [53]. Alcohol use is commonly considered a social activity, especially among adolescents and young adults, where peers are seen as the most consistent and strongest factor in the initiation and maintenance of alcohol use in adolescents and young adults [158]. From the perspective of young people, research has also studied various aspects of alcohol consumption, ranging from "pre-loading" (a phenomenon where youth consume alcohol before going out for the night), to health risks associated with excessive drinking [212]. Recent in situ studies suggest that young adults tend to drink more alcohol when in company than alone [116], that the number of persons present tended to be higher in heavier drinking nights [122] and that the higher the number of drinks consumed at a given time during the course of the evening [199]. Other factors such as gender, composition of the group, activities, or drinking norms are also known to influence alcohol use among groups of peers [223, 153]. To our knowledge, no previous study has investigated the link between engaging in alcohol use and an automatic measure of social context.

Only a few studies have examined the link between alcohol use and the number of drinking locations visited. For example, in a study of persons arrested for drunk driving, Wieczorek et al. found that multi-location drinkers had higher blood alcohol levels at arrest than those who drank at a single location [219]. More recently, Dietze and colleagues reported that about two third of young adults visited at least 2 different locations, while 39% drank in only one location on their last big night out [57]. However, since these studies endorsed a public health perspective, they focused on the potential harms related to the quantity of alcohol consumed through self-reported data. To the best of our knowledge, ours is a first study which examines the link between automatically measured mobility and drinking behavior.

Self-reported data is potentially prone to recall biases. In the alcohol research literature, it has been shown that people forget to report about half of their actual consumption. In Switzerland, each person drinks about 3.4 liters of pure alcohol annually as per the survey data, while the alcohol sales data indicate an average of 6.8 liters of alcohol consumption per person annually [118]. In [118], the authors reported statistically significant differences between self-reported retrospective alcohol consumption and mobile phone based self-reports among young people. As a result, the feasibility to automatically characterize drinking behavior using smartphone data presents unique opportunities to study the urban nightlife of young adults.

In contrast with these works, our work captures the heterogeneity and complexity of the going out behavior of youth during night time using mobile crowdsensing, which to the best of our knowledge has not been studied in urban studies and human geography.

5.3 Mobile Crowdsensing Study Design

In this section, we describe the design of our crowdsensing study, including the development of two mobile applications, the specific urban context and protocol of the field study, and the recruitment of participants.

5.3.1 Mobile Application

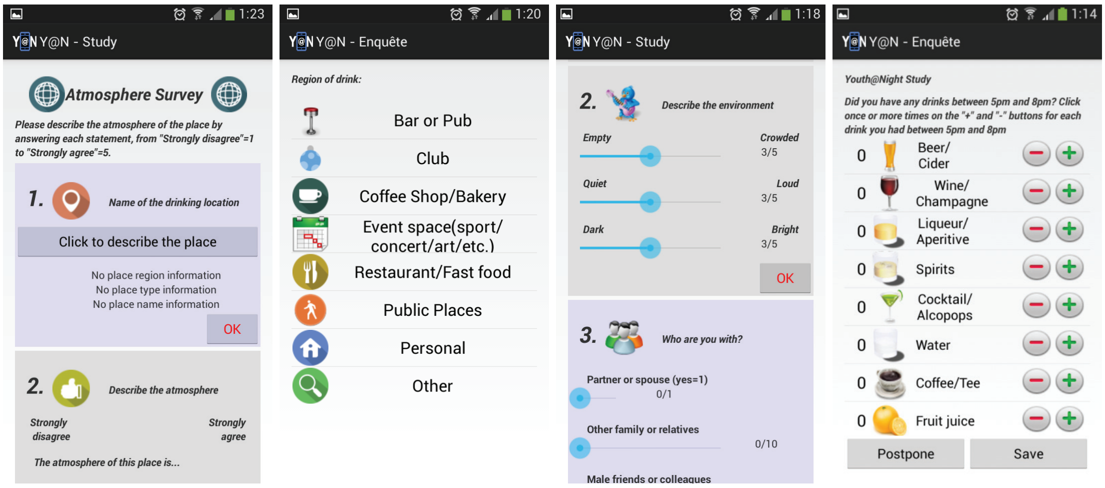
We developed two custom Android-based smartphone applications: a survey logger and a sensor logger application. These applications allowed participants to respond to various surveys, while at the same collect sensor and log data in a non-intrusive and privacy-preserving manner. We used Android 4.0.3 version as it represents compatibility with over 95% of the total Android-based phones at the time of design [55]. During our recruitment campaign, we did not encounter any user having an Android phone below this version. Throughout the chapter, “weekend nights” refer to Friday and Saturday nights between 8PM until 4AM the following day. Below, we describe these two applications in detail.

Survey Logger: The survey logger application let participants respond to various surveys in real-time on weekend nights. Surveys include place survey, video survey, and drink survey. We describe the questions asked in these three surveys in the next section. Figure 5.1 shows the screenshots of the application. Due to the multilingual Swiss population, the mobile interface was designed in three languages (English, German, and French), where users could choose their language of choice.

Sensor Logger: We developed a second application to collect different types of sensor and log data. It was designed to run as a background process without any user interaction. As with the survey logger application, all data was recorded only during the weekend nights in a non-intrusive manner. By design, it did not appear in the list of running applications on the user’s device. So, if users wanted to close the service, they had to manually terminate the application. We describe the type of data collected using the sensor logger in Section 5.4.

5.3.2 Study Context

Location: The study was conducted in two cities in Switzerland, namely Zurich (German-speaking city with population of 400,000) and Lausanne (French-speaking city with population of 140,000). These cities were chosen to capture regional diversity and because both cities are the two major hubs for nightlife activities in Switzerland [132, 145]. Switzerland has a national population of 8 million people, i.e., about the same as New York City. Zurich and Lausanne are the first and fourth largest Swiss cities. The main nightlife areas in both cities are walkable, and many of them are in close physical proximity. Both cities provide excellent public transportation including during night time [187]. Due to these factors, both cities receive an influx of youth from neighboring towns and even other cantons on weekend nights.



Please describe the atmosphere of the place by answering each statement, from "Strongly disagree"=1 to "Strongly agree"=5.

1. Name of the drinking location

No place region information
No place type information
No place name information

2. Describe the atmosphere

The atmosphere of this place is...

2. Describe the environment

Empty	Crowded	3/5
Quiet	Loud	3/5
Dark	Bright	3/5

3. Who are you with?

Partner or spouse (yes=1) 0/1
Other family or relatives 0/10
Male friends or colleagues

Youth@Night Study

Did you have any drinks between 5pm and 8pm? Click once or more times on the "+" and "-" buttons for each drink you had between 5pm and 8pm

0	Beer/ Cider
0	Wine/ Champagne
0	Liqueur/ Aperitive
0	Spirits
0	Cocktail/ Alcopops
0	Water
0	Coffee/Tee
0	Fruit juice

Figure 5.1 – Screenshots of the survey logger mobile application.

Nightlife and Youth in Switzerland: Nightlife for youth in Switzerland reflects patterns that are common in other western European countries. First, young people (especially those living on their own) often spend part or all of the night at home. This is partly related to the high cost of going out: the typical cost of a night out can easily reach 50 Swiss Francs (CHF), equivalent to 52 USD. Based on a pre-study survey conducted with 367 potential study participants (16–25 yo), we found that on average, young people spend 45 CHF per night when going out, with 17% of respondents reported to be spending more than 75 CHF per night. Note that unlike the US, the legal alcohol purchase age in Switzerland is 16 years for beer and wine and 18 years for spirits. Second, youth in Switzerland also spend part of the night in public spaces other than bars, restaurants, or clubs. This includes public squares, parks, train stations, but also streets and areas outside nightclubs [53]. It is important to note that drinking in public spaces (an activity associated with nightlife) is not criminalized in Switzerland. Third, young people (below 25 years old) are allowed subsidized travel from 7PM on all forms of public transportation, which encourages the use of public transportation and reduces the risks associated to driving cars.

In summary, this setting provides opportunities to examine the physical mobility, activities, and social patterns of young Europeans as they experience various aspects of nightlife.

5.3.3 Recruitment of Participants

For our field study, we recruited participants in Zurich and Lausanne, all between 16 and 25 years old. The field study was planned to run from September to December 2014, and participants were recruited in the month of September 2014. This period was chosen as it represents the time of the year after the summer holidays and before the Christmas break. For recruitment, young people on the streets were approached by our recruitment team on weekend nights between 8PM and midnight. Recruitment was done by a team of research assistants in groups of two to four people. Before the first recruitment session, the assistants got familiar with the smartphone application, practiced their introduction speech, and were reminded of the recruitment process. An authorization from the city authorities was required to conduct recruitment on the streets. In order to identify the recruitment zones to help the assistants, we used a place dataset extracted from Foursquare [182]. Using the Foursquare data, recruitment areas were identified based on their nightlife activities (e.g., bars, clubs, public parks, streets, etc.) which were further discussed and validated with local experts (social workers and police.)

During the recruitment process, teams of research assistants carried a field diary to document their experiences. We manually coded these notes and found some common themes. First, our choice to wear a lime-green t-shirt having the study logo, worn by all recruiters, intrigued passers-by and assisted the team to engage in conversation. Furthermore, it legitimized the recruiters approaching people on the streets. Second, recruiters observed that the study payment of 100 CHF was a key factor for participants aged between 16 and 18 years old. Third, the recruitment team noted that young people were typically going out earlier than their older counterparts. Fourth, our team noted that many iPhone users appeared frustrated and even mentioned being “discriminated” as they appeared quite interested to participate in the study. Fifth, the weather played a role during the second week of recruitment: few people were going out as it was raining during that week, which reduced the pool of potential participants. We believe that these findings could be useful to Ubicomp researchers recruiting participants for similar crowdsensing campaigns.

5.3.4 Study Protocol

We ran our field field study from September to December 2014. Before the study began, participants downloaded the two mobile applications (survey and sensor logger) and installed them on their own phones. During a weekend night, whenever participants were in a new location, they were asked to describe the place, its environment, and record a short 10-second video to capture their current environment using the survey logger application. While participants were at a given place, they were also asked to document their drinks (alcoholic and non-alcoholic), and describe the people they were with. In parallel, sensor data was continuously collected in the background using the sensor logger app. Note that we did not perform any real-time place detection on users’ mobile devices, so we had no means to know if the participant had indeed moved to a new location during the night. It was left to the participants to self-report if they had moved to a new venue.

Participants were sent hourly prompts to remind them to report new locations. Participants could stop these prompts for a given evening (e.g., when going to bed) or snooze them (e.g., if they were in a theater) at any time. At the end of the study, participants were given a monetary incentive of 100 CHF, if they completed at least 10 weekend nights of participation. Participants who volunteered for less than 10 evenings were compensated on a pro-rated basis, with a minimum of three nights. All participants were informed of the type of data collected as part of the study, as well as on all other aspects of the data collection. Our study was approved by the ethical review board of Vaud and Zurich cantons, respectively, for the cities of Lausanne and Zurich in Switzerland.

5.4 Mobile Data Collection Framework

In this section, we describe the mobile data collection framework together with the different types of data collected, data transmission, and privacy.

5.4.1 Data Types

Pre-Study and Exit Questionnaires

After the recruitment phase and before the study, participants were asked to complete a questionnaire about their demographics, weekend nightlife habits, smartphone usage, and social media usage. After the field study concluded, participants were also asked to complete an exit questionnaire about their experiences during the study.

Survey Data

Place Survey: This survey was designed to document the functional attributes of the place and its in-situ atmosphere. Participants could answer the survey only when they reported to be in a new location during a night. If participants decided not to answer the survey when getting to a new place, they could have answered it any time while they were in that place. The place survey had two goals. First, to capture the place attributes including its city and place category (e.g., bar, restaurant, nightclub, public spaces, homes, etc.). We chose nine high-level categories adapted from Foursquare's place category hierarchy [3], as previously done in the literature [45]. Second, users were also asked to document the environment along three dimensions: occupancy, loudness, and brightness. These ratings were given on a five-point Likert scale, ranging from *very low* (1) to *very high* (5). For the place survey, we received a total of 1,394 responses from 206 participants (Table 5.1). In our data, public places refer to all places that are not private, including bars, restaurants, cafes, clubs and other outdoor public spaces (e.g., parks, plazas, lakeside, etc.). To avoid confusion, we refer to this later category as *PBS* (PuBlic Space) in the rest of the chapter.

Video Survey: In this survey, participants were asked to record a short 10-second video capturing the environment of their current place. Participants were instructed to capture a panorama by slowly recording a video turning from left to right with the phone in the landscape (horizontal) mode. Participants could take the video survey only after completing the place survey. If participants were unable or reluctant to take a video, we asked them

Dataset	# Records	# Users
Place Survey	1,394	206
Video Survey	1,323	204
Drink Survey	2,532	218
Combined Survey	1,323	204
Sensor Data	8 million	241
Interviews	40	40

Table 5.1 – Summary of the collected data during the *Youth@Night* study.

to specify the reasons in form of a multiple-choice questionnaire. We gave participants the following five reasons to choose from: 1) Ethical ("It is not appropriate to record a video now"); 2) Legal ("Recording a video is not allowed in this place"), 3) Safety ("I don't feel safe recording a video now"), 4) Social ("I was asked by someone not to record a video"), and 5) Other. Participants were allowed to choose multiple reasons (and at least one) if they decided not to record a video. We obtained a total of 1,323 responses to the video survey from 204 participants (Table 5.1).

Drink Survey: This survey was designed to log participants' nightlife activities and social context of their place visits. Users were asked to describe their activities in form of drink consumption, both alcoholic and non-alcoholic beverages. Drink attributes included information on their current drink, including the type of drink (beer/cider, wine/champagne, whisky, tea/coffee, fruit juice, water, etc.), the size of the drink (small, medium and big), and the alcohol quantity (light, medium, strong, and alcohol-free). All these attributes were selected based on existing literature on alcohol consumption [152]. To inform their social context, users reported the people they were with, including friends, colleagues, or family members. Participants were asked to answer this survey any time during a weekend night, with an automatic reminder every hour prompting users if they wanted to report a new drink. For this survey, we received a total of 2,532 responses from 218 participants (Table 5.1).

It is common for users to forget reporting every consumed drink. As a result, we designed an additional survey for users to report any forgotten drinks, which otherwise should have been reported using the drink survey. As with the drink survey, this survey could be answered any time during the weekend night, with an automatic reminder every hour prompting users to add their unreported drinks. While answering this survey, users were asked to indicate only the number of forgotten drinks for each drink type, so we have no information of the drink attributes (such as size of the drink, alcohol quantity, etc.), as was the case with the drink survey. For this survey, we received a total of 942 responses from 163 participants. Note that the responses to the *forgotten drink survey* was used only to complement the information of alcohol consumption for automatic recognition (Section 5.6), and not considered for the descriptive analysis (Section 5.5).

Combined Survey Dataset: During the study, we received a variable number of responses for each of the survey as reported in Table 5.1. For some place visits, participants responded to the place survey, but not to the video survey, while for other visits participants only responded to the drink survey, but did not provide answers to the place or video survey. This is inevitable given the “in-the-wild” nature of our study; similar trends have been reported in previous mobile data campaigns [115]. Due to these missing records, we combined the survey responses to include only those check-ins for which all the three surveys have been answered, resulting in a total of 1,323 check-ins from 204 participants. Consequently, for each check-in in the combined dataset, we have the complete information about the place functional attributes, activities (social and alcohol consumption), and the responses to the video survey. In the rest of the chapter, a “check-in” refers to the act of recording place information (via the place survey) and responding to the video and drink survey.

Sensor Data

In addition to collecting participants’ self-reported survey data, we also collected data from various mobile sensors and recorded application usage using the sensor logger. The data types included accelerometer, application, battery, bluetooth, location, screen and wifi logs. Table 5.2 lists all data types, along with their respective sampling rate. Sampling rate for various sensors was optimized using rigorous experimentation to optimize battery life, following prior work [115]. During the field study, we gathered over 8 million sensor data records from 241 participants (Table 5.1). Note that due to the automated nature of sensor data collection in the background, we obtained sensor data from more participants relative to those who responded to different surveys i.e., not every participant who contributed sensor data equally responded to different surveys (see last column of Table 5.1).

Interviews

After the data collection study was concluded, we conducted semi-structured interviews with 40 participants. While conducting these interviews, participants’ recorded videos were used as stimuli to talk about their going out practices during weekends. Participants were also asked about their video-taking experience. The purpose of these interviews was to gain personal insights about the way young people engage in urban nightlife as a complex way of enjoying themselves while negotiating the dangers of the city at night [95].

5.4.2 Data Transmission

To preserve users’ cellular data and optimize battery life, data transmission from phones to the backend server was performed when the device was connected to a WiFi access point. Automatic data upload was scheduled for every Monday at a random time between midnight and 6AM. When the data was successfully transmitted it was deleted from the

Sensor Data Group	Description	Sampling Rate
Accelerometer	Set of acceleration values along three axes	10 seconds continuously every minute at 50Hz
Application	List of all background applications	Once per minute
Battery	Battery charging status	Whenever a change in battery status is detected
Bluetooth	List of bluetooth devices in range	Once every 5 minutes
Location	Set of location estimates using GSM or GPS	1 minute continuously every 2 minutes
Screen	Status of the phone screen (on/off)	Whenever the screen state is changed
WiFi	List of all visible WiFi hotspots	Once every 5 minutes

Table 5.2 – List of all types of sensor and log data collected during the study.

device. Participants were provided an option to force a manual upload, if automatic uploads did not succeed.

5.4.3 Data Privacy

As stated before, the study was approved by the ethical review board of Vaud and Zurich cantons in Switzerland. Given the potentially sensitive nature of the collected dataset, we requested the consent of participants to share their data only within the research team. This restriction has implications on who can view and analyze the data, in addition to how the data can be manually coded.

5.5 Descriptive Analysis of Survey Data

In this section, we first introduce the descriptive statistics of the study questionnaires, then we list the findings and insights from the survey data.

5.5.1 Analysis of Pre-Study Questionnaires

Participants Demographics: Of the total study participants, 201 of them responded to the study long questionnaire. Using the demographic information, we observe a fairly balanced gender ratio (52% male, 48% female). We found that the majority of participants (62%) are below the age of 20, as shown in Figure 5.2a. From an occupational point of view, 62%

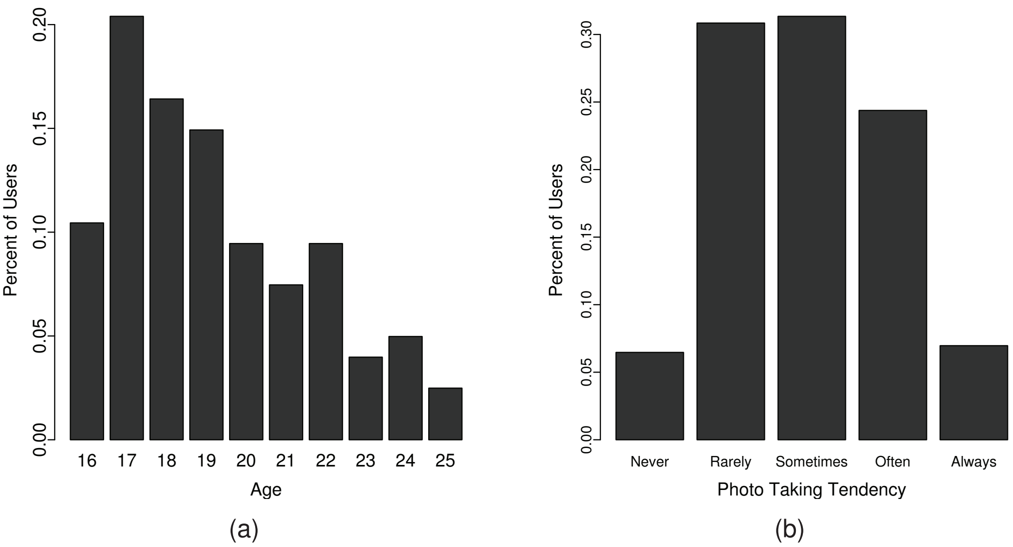


Figure 5.2 – Plots showing the barplots for a) Age, and b) Photo taking tendency when going out at night for study participants.

of participants reported being students, 24% reported apprenticeship as their occupation, while 4% declared to be working full-time. Over 83% of participants reported to be living with their parents, while only 10% reported to be living in either shared housing or a student residence. From these findings, it is clear that the demographics of participants are inclined towards teenage students and young adults living with their parents. During recruitment, our aim was to have a nightlife population as representative as possible. This population is significantly different than those reported in previous Ubicomp research, e.g., undergraduate and graduate students in Korea [45] or the US [217].

Smartphone Devices and Usage: Now we examine the diversity of mobile devices used by participants. The device information indicates eight different mobile manufactures and 51 model versions, with Samsung being the dominant phone manufacturer (63%), followed by HTC (18%) and Sony (9%). Samsung Galaxy S4 and Galaxy S3 (released in 2014 and 2013 respectively) are the two most popular model versions.

In terms of usage, more than 92% of participants reported to be using a smartphone for at least two years. 90% of them rated themselves as either a heavy or medium smartphone user. Over 81% of respondents send text messages more than half of the time when gone out during the night. When asked about the frequency of photo-taking when going out at night, 31% of respondents reported to take photos often or always (Figure 5.2b). It is clear that the demographics of the participants align with their aggregated smartphone use.

Home Location and Going Out Behavior: 63% respondents reported going out at least once per weekend while spending an average of 42 CHF (equivalent to 43 USD) a night. Participants were asked to list the postal code of their residence as well. After geo-coding

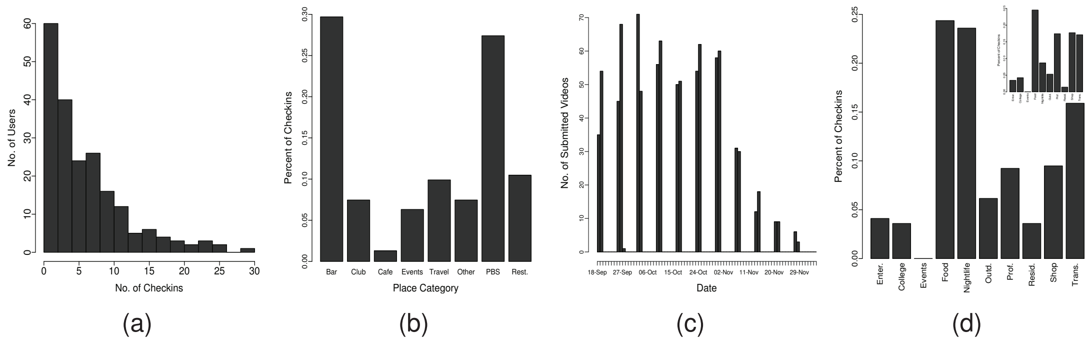


Figure 5.3 – Plots showing the histograms for a) Number of contributed check-ins per user, b) Distribution of categories for public places, c) Number of submitted videos every weekend night, and d) Distribution of categories for Foursquare check-ins for night time (In the inset, the overall distribution is shown).

a total of 128 unique postal codes, we found that participants live in 11 different cantons (Switzerland has 26 cantons), with 52% of participants reported living in the canton of Vaud, 37% in the canton of Zurich, while the rest (11%) resides in the neighboring cantons. Interestingly, only 40% reported living within the city limits of either Lausanne or Zurich, while the rest of participants commute from neighboring towns for nightlife. As argued before, these findings confirm that both cities receive an influx of youth from neighboring cities and cantons on weekend nights. The spatial coverage of our data spans the main east-west corridor of Switzerland (distance between Lausanne and Zurich is 226 km), which differs from previous works (typically limited to one city.)

5.5.2 Analysis of Survey Data

In this subsection, we analyze the combined survey data consisting of 1,323 check-ins from 204 participants to better understand the different places participants visited and the contextual factors surrounding their place visits.

User Contributions

We received a total of 1,323 check-ins from 204 participants. On average, a participant contributed 6.5 place check-ins, with one participant submitting a maximum of 29 check-ins. Users' place check-ins follows a typical long tail distribution (Figure 5.3a). Similar heavy tail characteristics have been previously reported in the literature [45]. From 1,323 check-ins, participants submitted a total of 894 videos, while for the rest 429 check-ins participants did not to take a video. Figure 5.3c shows the distribution of submitted videos for the study duration. Note that the video dataset, discussed in Section 5.7, is different than [200] in that it is intentional, focused on nightlife patterns, and spans a variety of places.

Place Analysis

Now, we examine the diversity and coverage of places visits: (a) How well are different place types represented in our study? (b) How well the distribution of check-ins compare with the check-ins distribution from social media data? In this subsection, we report the results using the semantic location data gathered via the survey logger, and not the physical location inferred via GPS (collected via the sensor logger). To examine the spatial coverage of mobility patterns, as inferred via GPS traces, refer to Figure 5.6.

As reported earlier, participants recorded place information including the city of the check-in and place category using the place survey. We use the place category information to infer whether participants were at a public or a private place. When participants checked into a private place, we further asked them to specify whether the check-in was at their own homes, a friend's home, or other private venues.

Of all the 1,323 check-ins, 626 (47.3%) were at private places, and the rest (52.7%) at public places (Table 5.3). Of the 626 check-ins at private venues, 62% were reported from their own home, 30% from their friends' home, while the rest (8%) occurred at either their workplace or other private venues (e.g., student hostel, someone else's home while baby-sitting, etc.). A large number of check-ins at private venues might be due to two factors: a) the majority of participants (83%) reported living with their parents (Section 5.5.1), and b) spending a night outside is relatively costly given the demographics and income earning status of participants. After manually browsing through some of the videos taken in private places, we found that some videos indeed show young people in large family homes (with a large living room and kitchen), but also studios and small apartments that correspond to living alone and shared accommodation. Videos recorded at homes clearly have an intimate, unfiltered flavor. One can observe personal items, bedrooms, presence of friends, without the beautification often found in social media content. The video dataset offers a novel way to conceptualize such places. The variety of private places captured on video is novel in and of itself, as previous studies have focused on college students (who in the US typically do not live with parents.)

For check-ins at public places, 30% were at bars, followed by 27% at PBS as shown in Figure 5.3b. Restaurants and travel each contributed around 10% of all check-ins. We observe that a significant portion of check-ins happened at PBS, which suggests that youth spend a considerable amount of time hanging out in these spaces away from mainstream nightlife areas. Recall that PBS refers to outdoor public spaces e.g., parks, plazas, lakeside, etc. (Section 5.4.1). This provides support for the qualitative work done with youth in the US context [35]. The PBS videos are specially interesting as they are unfiltered. Videos in dark parks or squares where people hang out, and video taken on streets outside commercial venues, are commonly found in our data. These places also provide support for qualitative work on the practices of Swiss youth in these venues [53]. Note that both Lausanne and Zurich have a scenic lakeside used often for recreational activities. Overall, these findings reflect that the study indeed captured different patterns of participants' nightlife behavior.

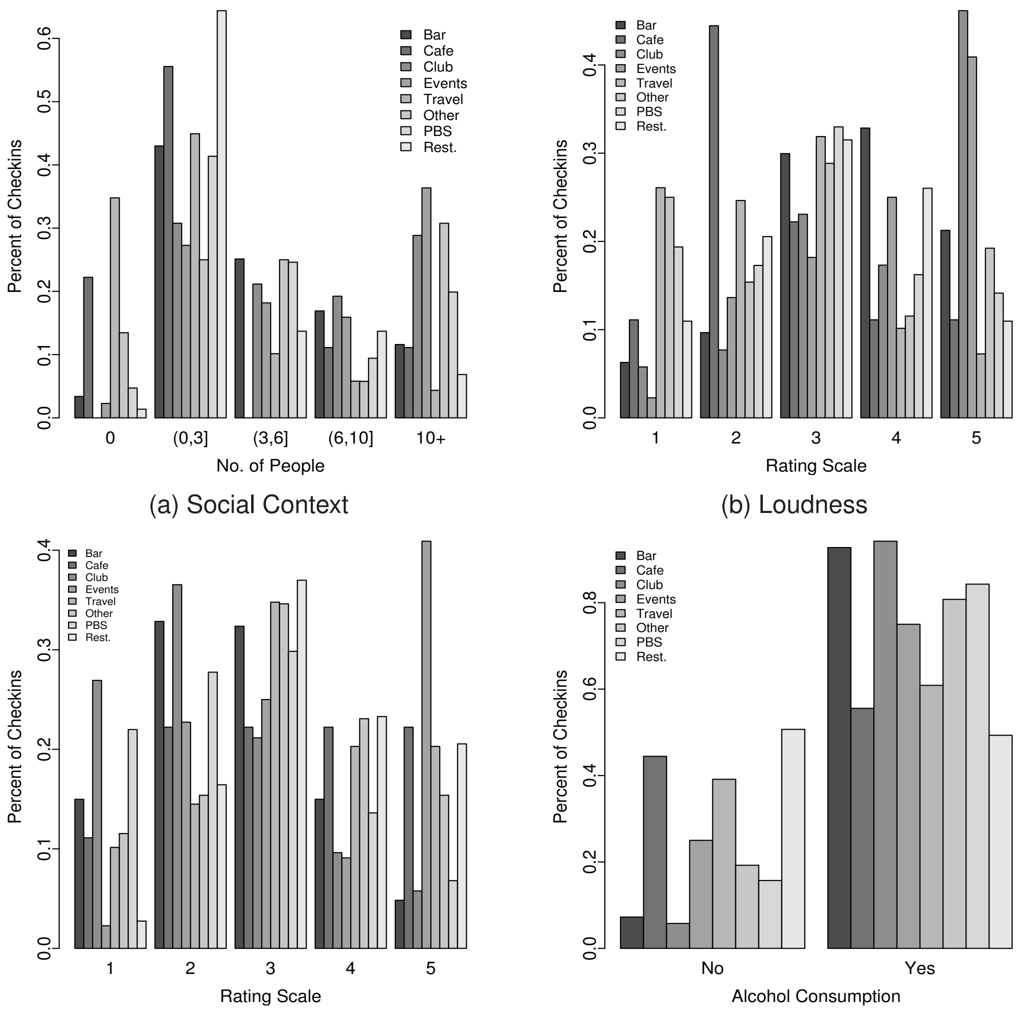
Context	Attributes (Features)	Private N=626	Public N=697
Social	Alone	149	51
	At most 5 people	325	403
	More than 5 people	152	243
Ambiance	Crowdedness	1.88	3.0
	Loudness	2.14	3.19
	Brightness	3.14	2.84
Activity	Consumed Alcohol	299	560
	Not Consumed Alcohol	327	137
Videos	# Videos Taken	416	478
	# Videos Not Taken	210	219

Table 5.3 – Summary statistics for public and private places.

Comparison with Foursquare: Independent of the *Youth@Night* study, we collected geo-localized Foursquare check-ins from Switzerland. This dataset was collected in addition to the data for 300 places across six cities (Chapter 2). For the two studied cities (Zurich and Lausanne), we obtained a total of 54,184 publicly available check-ins between December 2011 and February 2014. For all Foursquare check-ins for which place type information was available, we plot the distribution across 10 Foursquare categories, temporally filtered between 8PM to 4AM in Figure 5.3d. In the inset of the same figure, we plot the overall Foursquare check-in distribution (i.e., without any temporal filtering). For both the temporally filtered and overall distribution, food places received the most number of check-ins, which is in contrast with the findings from *Youth@Night* survey data. Similar to the previous work [45], we observe that places visited during night were more represented in our crowdsensing study compared to Foursquare e.g., events category did not contain a single check-in for temporally filtered Foursquare data. For some of the categories, the check-in distribution of our study (Figure 5.3b) is similar to the temporally filtered Foursquare check-ins; however it is significantly different with respect to the overall Foursquare check-in distribution. These findings point towards limitations of social media in terms of representativeness and temporal resolution at least in the context of Switzerland [206].

Activities and Social Context

We now examine the activity and social context of participants at night, i.e., with whom and how many people participants were at the time of reporting their check-ins. For 24% (resp. 7%) of check-ins to private (resp. public) places, participants reported to be spending the night alone (see Table 5.3). For 24% of private place check-ins, participants reported to be with more than five people. Digging further into these cases, we found that for 72% of these check-ins, alcohol was reported to be consumed, potentially suggesting social occasions or house parties [99, 72]. Manually browsing through videos confirmed this point.



(a) Social Context

(b) Loudness

(c) Brightness

(d) Alcohol Consumption

Figure 5.4 – Plots showing the histograms for a) Social Context, b) Loudness, c) Brightness, d) Alcohol Consumption across all public place category types.

For check-ins to public places, we found that for most of the categories (except events and other), the majority of check-ins were reported to be with fewer than three people (Figure 5.4a). By investigating further, we found that for bars and clubs participants reported to be with more than five people for 34% and 52% of check-ins respectively. Not surprisingly for event spaces, participants reported to be with more than 10 people for 36% of event check-ins; while for the travel category, 35% of check-ins were reported to be alone. The travel category corresponds to situations when people were either walking, traveling in public transportation or using their personal vehicles.

Overall for 65% of place check-ins, participants reported consuming alcohol. We found that for check-ins at private venues, users reported drinking alcohol for 48% of cases (Table 5.3), suggesting the trend of home drinking [72]. For check-ins at public venues, we found that users reported drinking alcohol for 80% of cases. Digging further into public place categories, we found that 84% of check-ins were reported with alcohol consumption in the PBS category (see Figure 5.4d). This finding provides support for qualitative work on the prevalence of "street square" drinking amongst Swiss youth [53]. Note that consuming alcohol in public places including public transportation is not criminalized in Switzerland.

Ambiance Context

In the place survey, participants were also asked to judge the environment in-situ along three dimensions: place occupancy, loudness, and brightness. All ratings were given on a five-point Likert scale (1-*very low*, 5-*very high*). Using this data, we observe that public places are in general more crowded, louder, and darker relative to private places (see Table 5.3). These findings are not surprising. Figure 5.4b and 5.4c plots the distribution of loudness and brightness across all public place categories. On average, bars and clubs were reported to be relatively more crowded, louder and darker, compared to other categories.

Overall, the survey data captured diverse places during night including private homes, public squares and public transportation, which potentially suggests that youth spend a considerable amount of time hanging out away from mainstream nightlife areas. Furthermore, irrespective of place type, the study captured places along the full spectrum of social and ambiance variables, i.e., check-ins to public places alone, and house parties at private homes; or, consuming alcohol alone, and staying sober in the company of more than 10 people; or public places which were reported to be quiet, and private places which were reported to be very loud. This poses interesting challenges towards automatic analysis of youth's social activity (Section 5.6) and videos (Section 5.7), as the places covered are diverse in terms of overall social and physical ambiance dimensions.

5.6 Automatic Recognition of Social Activity

In the previous section, we demonstrated the heterogeneity of youth going out behavior using self-reported survey data. We also we provided a detailed account of participants' alcohol consumption habits, and the specific places where they drink, using the survey data. In this section, we focus our attention towards automatically classifying their alcohol consumption using mobile sensor data.

While alcohol consumption is one of the key social activities of youth going out behavior, it is the number one risk factor for morbidity and mortality among young people in many countries. Heavy drinking and related incidents in public on weekend nights are a concern for city councils, policy makers, and a nuisance for the general public [135, 221]. There is a recent push to study alcohol consumption patterns using mobile phones [119] to increase research validity and reliability and to widen the understanding of contextual factors. In this section, we propose a machine learning pipeline to automatically infer a binary state

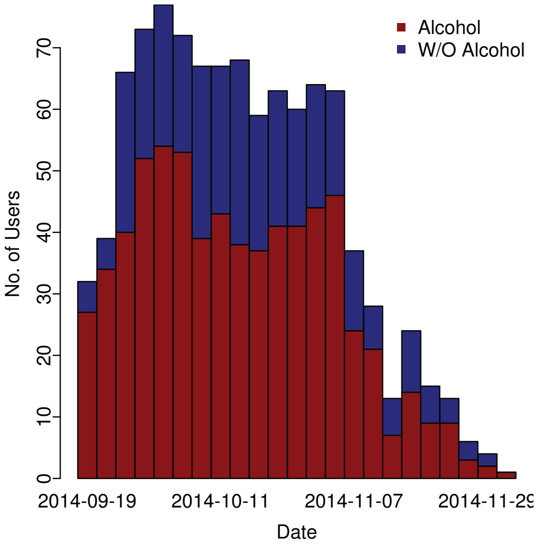


Figure 5.5 – Temporal distribution of participants contributing data to the field study. Numbers in red indicate users who reported consuming alcohol on a specific night, while in blue we show the participants who had reported no consumption of alcohol throughout the night. Best viewed in color.

of alcohol consumption for single nights, describing the various steps including the pre-processing of sensor data, feature extraction and regression analysis. Our objective is to study the discriminative power of smartphone-derived cues related to physical motion, location, connectivity, and application usage for alcohol consumption in an “in-the-wild” nightlife setting.

5.6.1 Data Filtering

For automatic classification, we conducted our analysis on a *user-night*. A user night indicates a night-out per user. If the field study run for m nights, where for each night we have n participating users contributing sensor data, the total number of user-nights will be mn. In our study, we obtained data for 2,934 user-nights from 241 participants (Table 5.1). On average, a participant contributed sensor data for 12 nights, with more than half contributing 14 nights or more. The maximum number of nights contributed by a single user is 30.

However, we observe that many user-nights missed one or more sensor data type. This is inevitable given the real-time and in-situ nature of our study; similar trends have been reported in previous large-scale mobile data campaigns [124]. Table 5.4 lists the number of user-nights with missing records for all sensor data types. Due to inherent sparsity, we filtered the data to select only those user-nights for which sufficient sensor data is available. Consequently, for our analysis and inference framework, we selected only those user-nights which satisfied the following three criteria: a) the participant had responded to either the *drink* or the *forgotten drink* survey at least once during a given night (Section 5.4.1); b) the

Sensor Data Group	% User Nights with Available Records in Raw Data	% User Nights with Available Records in Filtered Data
Accelerometer	76.18%	99.51%
Application	38.24%	99.31%
Battery	76.24%	100%
Bluetooth	75.66%	98.52%
Location	55.69%	100%
Screen	69.33%	92.78%
WiFi	65.13%	93.67%

Table 5.4 – Percentage of user-nights with available data in the raw and filtered datasets.

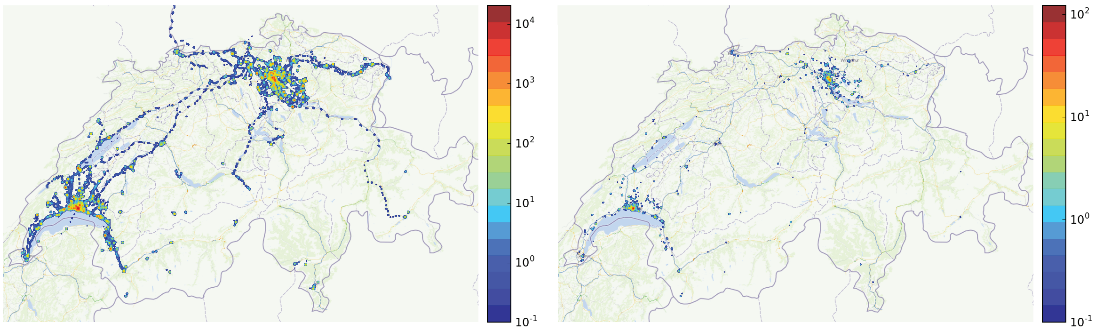
participant had at least one data sample for any sensor data type during the night; and c) the participant had at least one stay-point based on location sensor logs. (A stay-point is a stable location defined later in Section 5.6.2.)

As a result of filtering, we were left with a total of 1,011 user-nights from 160 users consisting of 12 Fridays and 11 Saturdays nights between September and December 2014. In Figure 5.5, we show the temporal distribution of participants for each night, in addition to showing the overall participation rate. We observed 12 weekend nights where we had more than 50 participants contributing data with drinking events to our study. Figure 5.6 shows the spatial movement of the participants based on their accumulated GPS traces and stay-points, plotted at the country level. The study participants restricted their movements mostly to the two studied cities – Lausanne and Zurich; but sometimes they also spent their nights in neighboring cities as a reflection of real-life mobility. These mobility plots also corroborate the self-reported data on participants' home location as reported in Section 5.5.1.

Even after filtering, the data contains user-nights with missing data for some of the sensor data types. In Table 5.4, we show the percentage of user-nights with available data in the raw and filtered data respectively. For instance, 64 user-nights have missing WiFi sensor logs in the filtered dataset. It points towards the challenges of in-situ studies. For the subsequent analysis in this section (pre-processing, feature extraction, regression analysis and classification tasks), we used the filtered 1,011 user-nights dataset, unless otherwise stated.

5.6.2 Feature Extraction and Analysis

In this subsection, we describe the pre-processing of data and the procedure for extraction of features. As mentioned in the previous subsection, our unit of analysis is a *user-night*. Consequently, we aggregated all the data informing different aspects of phone usage per user on a nightly level. Table 5.6 lists all the different features extracted. Features were selected based on prior work [121, 163, 228, 140, 180, 60]. In addition, we generated features that



(a) Raw GPS Traces

(b) Stay Points

Figure 5.6 – Heatmap showing spatial distribution of raw GPS traces (left) and stay-points (right) of participants. Zurich (north-east of the country) and Lausanne (south-west) are clearly represented in the map. Many other places in the country were also visited. Red color indicates high density of stay-points, while blue represents low density. Best viewed online in color and high-resolution.

could potentially characterize levels of drinking activity. Below, we describe each data type and list all the extracted features.

Accelerometer Logs: We obtained accelerometer data every minute for 10 seconds, with a sampling rate of 50Hz resulting in a total of 500 readings per minute. Each individual reading contained the raw acceleration values along the three axes (x_i, y_i, z_i). Given the raw readings, we computed basic features including signal magnitude area (SMA_i), raw acceleration (a_i), resultant acceleration (a_i^r), and relative resultant acceleration with respect to the previous reading (a_i^{rr}), as shown below (feature names are taken from previous literature):

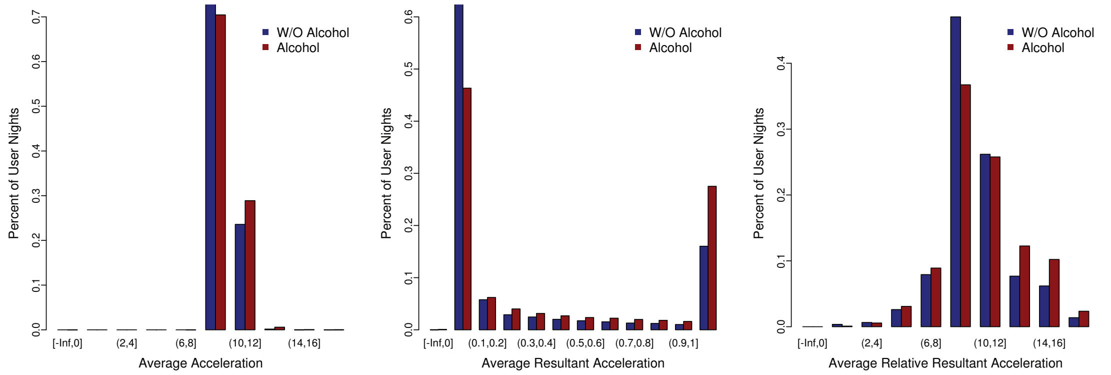
$$\bar{x} = \sum_{i=1}^{500}(x_i), \quad \bar{y} = \sum_{i=1}^{500}(y_i), \quad \bar{z} = \sum_{i=1}^{500}(z_i),$$

$$SMA_i = \sum_{i=1}^{500}|x(i)| + |y(i)| + |z(i)|, \quad a_i = \sqrt{x_i^2 + y_i^2 + z_i^2},$$

$$a_i^r = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2},$$

$$a_i^{rr} = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2}.$$

We defined each scan per minute as a *segment*. Our data contains a total of 391,717 segments. To transform the raw readings per segment, we further computed basic summary statistics (mean, standard deviation, min., max.) over each segment, in addition to calculating the angle between the g vector and the vector containing the average acceleration value for each axis ($[\bar{x}\ \ \bar{y}\ \ \bar{z}]$). To obtain the aggregated features per user-night, we generated



(a) a_i (b) a_i^r (c) a_i^{rr}

Figure 5.7 – Plots showing the histograms for a) Acceleration, b) Resultant Acceleration, and c) Relative Resultant Acceleration for both night types.

histograms of segment features, as pointed in Table 5.6. In total, we generated a total of 79 acceleration features (all variants of five basic features.) Most of the features extracted from accelerometer logs are derived and adapted from prior work [121, 163, 28]. For instance, the SMA feature is designed to distinguish between static and dynamic activities, while the angle provides cues to inform the tilt and orientation of the mobile device. In Figure 5.7, we the distribution of raw acceleration, resultant acceleration and relative resultant acceleration for both alcoholic and non-alcoholic user-nights.

Accelerometer data has widely been used to recognize physical activities of users ranging from standing, sitting, walking, jogging, and climbing [28, 121, 163]. We are not inferring categories of physical activity, as we did not collect activity labels as part of the field study.

Application Logs: There have been numerous studies which have examined mobile applications logs to understand user context ranging from location, temporal and social context [59, 58, 32]. In this work, we extracted features to explore the role of mobile application usage (in terms of spatial and social context) on alcohol consumption. Application usage might be informative of drinking-related activity as it has been shown that differentiated usage of apps occur in real life depending on where and with whom the phone users are.

Through the sensor logger, we scanned the list of all running applications every minute (Table 5.2). For each scan, we obtained a list of up to 50 applications. As opposed to dedicated apps designed to monitor application usage [32], the sensor logger was designed to gather only the list of currently running apps. As a result, our analysis of application logs is limited to examine fine-grained application usage behavior (e.g., time spent on Facebook, number of messages and pictures shared using WhatsApp, etc.). In our dataset, we observed a total of 1,299 unique applications, with WhatsApp being the most frequently used application. Running applications can either be native (i.e., pre-installed by the device manufacturers), default Android apps or user-installed apps. As noted earlier, due to the multi-region nature of our study, the mobile interface was designed in three different languages. Consequently,

Id	Category	Total Apps	Top-3 Apps
C1	Systems/Native	39	Contacts, Accueil TouchWiz, Parametres
C2	Communication	12	Messages, InCallUI, S Voice
C3	Entertainment	8	Youtube, Musique, Shazam
C4	Social	5	WhatsApp, Facebook, Snapchat
C6	Travel	5	Maps, Mobile CFF, SBB Mobile
C5	Other	28	Chrome, Internet, Recherche Google
C7	Below100	1,200	Alben, Applications, Actives
C8	Y@N	3	Y@N-Enquete, Y@N-Studie, Y@N-Study

Table 5.5 – Categorization of mobile applications. Note that apps in multiple languages appear due to the multi-region nature of the study.

apps having the same functionality appear in different names due to the translated interface e.g., Telephone and Telefon both refer to the Android application handling the call activity in English and French interface respectively.

Application Categorization: To deal with the diversity of applications, we manually coded the top 100 applications (based on their usage frequency) into eight categories (Table 5.5) which are: System/Native, Communication, Entertainment, Social, Travel, Other, Below100, and Youth@Night (in short Y@N, our app for this study). Similar app categorization systems has been used to group mobile apps in the literature [58, 32]. In Table 5.5, we show the total number of applications classified in a given category, besides showing the Top-3 apps for that category. The Below100 category contains applications that were not frequently used by the participants in the study, while Y@N denotes applications designed for our study. Note that the Y@N category contains only the survey logger application; the sensor logger app, being a daemon service, did not appear in the list of running applications. Applications falling under Y@N category also belongs to the Top 100 applications. Note that each application was assigned to the closest possible category, though an application in principle could belong to multiple categories.

For each user-night, in addition to the basic features (e.g., number of records or scans, count of unique applications, duration), we generated a set of three additional count features using the categorized applications. First, for all the applications obtained per scan, we incremented their respective category count (*Count 1*). Second, we examined the top-3 applications (based on their recency order) in a scan and incremented their respective category count (*Count 2*). It is important to note that while returning the list of running apps, the Android API returns the list in order of recency, i.e., the most recently app used is returned first, and so on. Using the Android API, we have no means to examine for how long an app has been used, thus we used the recency order as a proxy to understand temporal app usage. For the last and third count feature (*Count 3*), we count the number of times an app category had improved its rank (based on the recency order), with respect to the previous scan. For *Count 3*, we looked at only the top-3 applications and their respective categories. We incremented the count only if there was an improvement in the rank with respect to the previous rank. For each count feature, we also added their normalized count version as features. Count normalization was performed based on the number of scans in a user-night. In total, we generated a set of 52 features using application logs (Table 5.6).

For all eight categories, we do not observe any significant differences for all count features between alcohol and non-alcohol user-nights. To illustrate it, in Figure 5.8b, we show the distribution of normalized count (*Count 1*) for “Social” applications (C4 category) across several bins for both types of user-nights.

Battery Logs: The battery sensor returns a set of values whenever a change in the battery status is detected. From this sensor, we obtained features including battery status, level, temperature, and whether the phone was plugged to a power source. The battery can be in one of five states (charging, discharging, full, not charging, and unknown), depending on its power status [54]. A total of 10 features were generated using the battery logs.

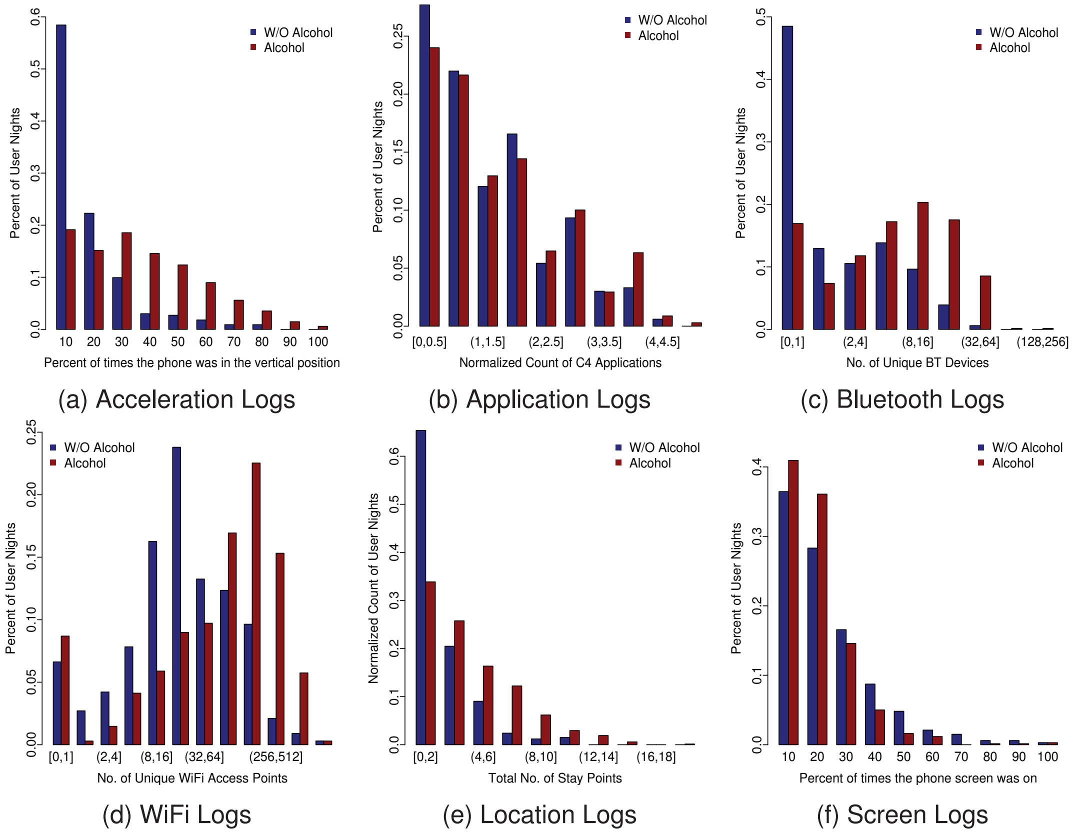
Bluetooth and Wifi Logs: From the Bluetooth and WiFi sensors, we obtained the list of the nearby Bluetooth (BT) devices and WiFi access points (APs) respectively. For these sensors, the scan was performed every 5 minutes. For every scan, we computed the list of features as described in Table 5.6. All the generated features are self-explanatory and were computed based on prior work [228], and correspond to a total of 6 features.

Bluetooth device density has often been used as a proxy to the social context of the user’s environment e.g., being alone or in the vicinity of a nearby group of people [64, 58]. When a large number of BT devices are detected, it likely signals the presence of people in the surrounding. Analysing BT logs becomes relevant to our work as we are interested in understanding the social context of participants and how it might potentially impact their drinking behavior. In our study, participants encountered an average of 9.6 unique BT devices and 128 WiFi APs in a typical user-night. For alcohol user-nights, 12.3 BT devices and 162 WiFi APs were encountered, while for non-alcohol user-nights only 3.95 devices and 59.3 WiFi APs were detected on average. Figure 5.8c and Figure 5.8d shows the distribution of BT devices and WiFi APs, respectively for both alcoholic and non-alcoholic user-nights.

Location Logs: The location sensor returns the set of location estimates using either GPS or GSM based on Android API calls. Location data was collected continuously for one minute,

Feature Group	Features (Dimensionality)	β
Accelerometer	Number of records (1)	$+2.43 \times 10^{-4}$
	Histogram of angle between g vector and X, Y, and Z-axis (6x3)	-14.17×10^{-2}
	Histogram of std. dev. of acceleration along X, Y and Z-axis (6x3)	$+69.75 \times 10^{-2}$
	Histogram of mean, median and max resultant acceleration (6x3)	-17.98×10^{-2}
	Histogram of mean, median and std. of relative resultant acceleration (6x3)	-16.08×10^{-2}
	Histogram of signal magnitude area (SMA) (6)	$\ldots$
Application	Number of records (1)	-2.39×10^{-6}
	Count of apps in different categories (8x2)	$+1.24 \times 10^{-2}$
	Count of Top-3 apps per category (8x2)	$+1.45 \times 10^{-1}$
	Relative ranking of Top-3 apps per category (8x2)	$+1.70 \times 10^{-1}$
	Duration between the first and last scan (1)	-6.3×10^{-6}
	Count of unique and total apps (2)	$\ldots$
Battery	Number of records (1)	-2.74×10^{-5}
	Count of various battery status (5)	$+3.57 \times 10^{-3}$
	Min. and Max. level of battery (2)	-7.75×10^{-6}
	Diff. between min. and max. battery levels (1)	-5.04×10^{-4}
	Plug time (1)	-3.8×10^{-3}
Bluetooth (BT)	Number of records (1)	-5.2×10^{-4}
	Number of BT IDs in range (1)	$\ldots$
	Number of unique BT IDs (1)	$+9.55 \times 10^{-3}$
	Percent of empty BT scans (1)	$\ldots$
Location	Number of stay-points (1)	$+2.92 \times 10^{-2}$
	Sum of duration at stay-points (1)	$\ldots$
	Sum of travel distance between consecutive stay-points (1)	$+4.5 \times 10^{-8}$
	Sum of travel time between consecutive stay-points (1)	$\ldots$
	Hist. of computed speed (8)	$+1.21$
Screen	Number of records (1)	$+4.3 \times 10^{-4}$
	Percent of time screen was *on* (1)	-0.41
	Screen count after midnight (1)	$\ldots$
	Duration of screen events (*on/off*) (2)	$+0.65$
WiFi	Number of records (1)	-6.73×10^{-5}
	Number of unique visible WiFi hotspots (1)	$+8.9 \times 10^{-4}$

Table 5.6 – List of all features extracted (organized by feature group) for each user-night. For each feature, β refers to the coefficient estimate of the penalized logistic regression (PLR) model. All β values are statistically significant at $p-$value < 0.001.



(a) Acceleration Logs (b) Application Logs (c) Bluetooth Logs

(d) WiFi Logs (e) Location Logs (f) Screen Logs

Figure 5.8 – Plots showing the histograms for a) Percent of times the phone was in the vertical position inferred using accelerometer logs, b) Normalized *Count 1* of “Social” applications (C4 category), c) Number of unique Bluetooth devices in range, d) Number of unique WiFi access points, e) Total number of stay-points, and f) Percent of times the phone screen was on inferred using screen logs for both alcoholic and non-alcoholic user-nights.

every two minutes. For each data point, we collected the location coordinates (longitude and latitude) and provider information (GSM-based or GPS-based). We also computed user speed using the location data.

In addition, for each user-night, we extracted the sequence of stay-points using location pairs (longitude and latitude). A stay-point is a region (radius of d meters) where one stays for a given duration of time (t minutes). It is a standard unit of analysis for place extraction [140]. In our analysis, we have used $d = 200$ meters and $t = 5$ minutes, which is similar to the ones reported in prior work [60]. Stay-points are extracted independently for each user. Using stay-points, we extracted a series of features such as the duration of stay at stay-points, travel distance, and time between consecutive stay-points, as reported in Table 5.2. The total number of extracted location-related features is 12.

Mobility using Location Logs: In a user-night, on average a participant had stayed on 3.8 stay-points, with the mean stay duration of 2.1 hours in a given stay-point. For alcohol user-

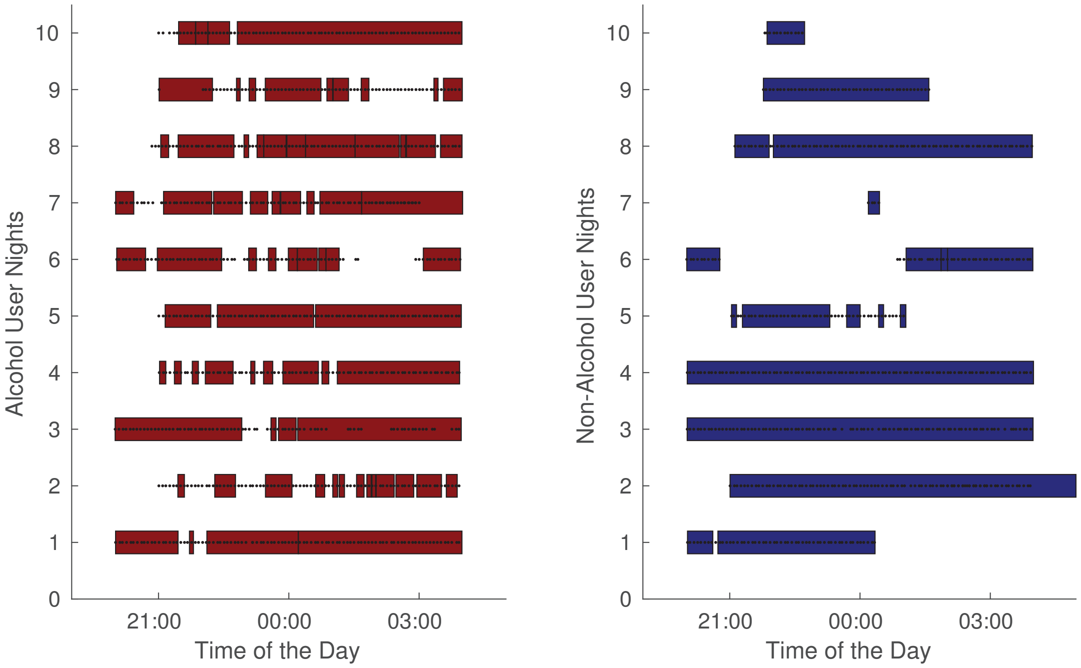


Figure 5.9 – Stay-points for a random selection of 10 alcoholic (left) and non-alcoholic (right) user-nights. A rectangle bar indicates a stay-point, and black dots indicate the raw location traces.

nights, a participant spent an average of 1.6 hours per stay-point, while for nights where alcohol was not consumed, a user spent an average of 3 hours per stay-point. We also observe significant differences with respect to the number of stay-points between nights spent consuming alcohol (4.5) and otherwise (2.5). Figure 5.8e shows the distribution of total number of stay-points across several bins for both alcoholic and non-alcoholic user-nights, which highlights the role of mobility on alcohol consumption. To contextualize these results and further understand the role of mobility, we examined the distance traveled between consecutive stay-points. We found that on average, a participant traveled a distance of 3.7 kms (resp. 1.4 kms) in an alcoholic user-night (resp. non-alcoholic nights). These findings suggest potential links between mobility and drinking behavior. Besides, these results complement our understanding of the physical mobility of participants with respect to what was observed using self-reported check-in data (Section 5.5)

Figure 5.9 shows the stay-points for a random selection of 10 alcoholic and non-alcoholic user-nights. In the figure, a rectangle bar indicates a stay-point, and black dots indicate the raw location traces. It was not always the case that we continuously obtained location traces for the entire user-night, as evident in Figure 5.9. For most user-nights, we observe missing location data for parts of the night. For some user-nights, the data was received only for few minutes (see non-alcoholic user-nights 7 and 10). These findings point towards the inherent missing data in our collected dataset as provided by the Android API and the subsequent challenges involved in using this data for an inference task.

Screen Logs: Using the screen logs, we measured any change in the state of the screen i.e., whether the screen was *on* or *off*. Using these logs, we computed a set of five features following previous work [180], as highlighted in Table 5.6. The Android API only provides the information when the state of the mobile screen changes from *on* to *off* and vice-versa. Using this information, we computed the percent of time the screen was *on* (or *off*) by normalizing it based on the time difference between the first and last screen events in a given user-night.

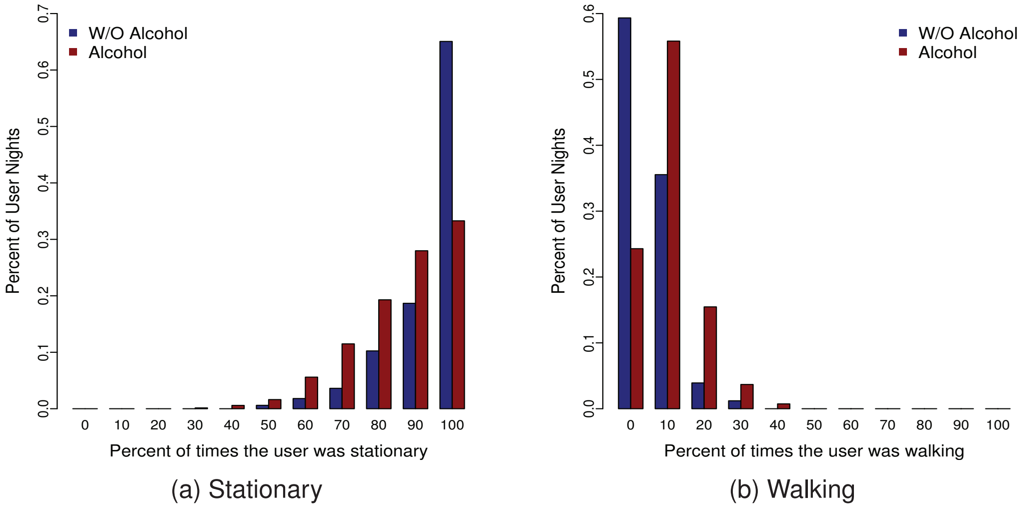
We found that on average, participants had their mobile screen *on* for 15.6% of the time, indicating some form of interaction with their mobile devices. When examining differences between user-nights, we found that for alcohol user-nights (resp. non-alcoholic nights), a participant kept their screen on for an average of 14.4% (resp. 18.3%) of times (see Figure 5.8f for a visual comparison).

Ground-truth Alcohol Consumption: In order to obtain the ground-truth data on alcohol consumption i.e., whether a participant had drunk alcohol during a user-night, we used the responses on the drink and forgotten drink survey (Section 5.4.1). Participants reported their drinks or forgotten drinks using the drink logger application. Participants reported both alcoholic and non-alcoholic beverages. If a participant reported consuming both types of drinks, we considered the night as alcoholic user-night. If a participant had not reported a drink using the drink survey i.e., while consuming it in-situ, but instead reported it the next day using the forgotten drink survey, we considered those user-nights in our analysis as well.

In total, 67% of user-nights had reported alcohol consumption, while 33% of user-nights had no alcohol consumption. This imbalance signals potential biases of the participants towards reporting mostly alcohol-related activities. After the in-situ data collection was over, we conducted a series of qualitative interviews with 40 participants to gain insights of their nightlife experience (Section 5.4.1). During the interviews, some participants mentioned perceiving the study as an “alcohol study” where they did not consider reporting consumption of non-alcoholic beverages (e.g., water, juice, etc.). This is interesting given the fact, that the instructions given to participants clearly stated them to report both alcoholic and non-alcoholic drinks (i.e., without any bias). Few other interviewees concluded that they were only allowed to report if they had bought the drinks themselves. Such narrations provide insights into the ways a “drink” was contextualized by participants during the data collection. Moreover, they offer a starting point to reflect on the observed reporting bias. Qualitative analysis of the interviews is presented in detail in Section 5.8.

5.6.3 Regression Analysis

In this section, we perform regression analysis using different feature groups. To examine the relationship between alcohol consumption and feature groups, we used penalized logistic regression (PLR) [74]. PLR measures the relationship of a binary response variable (consuming alcohol or not) as a function of the explanatory predictor variables (feature sets derived from sensors and log data). In contrast with traditional logistic regression, PLR guards against



(a) Stationary (b) Walking

Figure 5.10 – Plots showing the histograms for a) Stationary, and b) Walking inferred using location logs for both night types.

collinearity amongst features via regularization. Regularization is implemented using the lasso technique, which performs feature selection by forcing some of the coefficient estimates to zero, thereby increasing the model interpretability. When fitting the model, we scaled the parameters and performed a 10-fold cross-validation to find the best regularization parameter (λ), that minimizes the error function.

Results

For each feature group, we fit the penalized logistic model involving all features of that group against the binary alcohol consumption. Table 5.6 shows the results of the PLR model, which report the coefficient estimates (β) for each feature group. All β values are statistically significant at $p-$value < 0.001. For feature sets involving multiple dimensions (e.g., histogram of mean relative acceleration), we report the maximum β value for that feature set. As a result of regularization, features with zero coefficients are reported blank in the Table 5.6. We found that, while most features have either a positive or negative relationship with alcohol consumption, some features have zero β values (e.g., number of BT devices in range, sum of duration at stay-points, etc.). Features with zero β values suggest that either these features do not possess any predictive relationship with alcohol consumption or these features are correlated with other significant predictive feature(s) within the same feature group.

Most of the regression results in Table 5.6 corroborates the exploratory analysis reported in Section 5.6.2. For acceleration logs, the standard deviation of raw acceleration values (a_i) along the Y-axis has a positive relationship, while the maximum of the resultant acceleration values (a_i^r) has a negative association with alcohol consumption. When examining the features extracted using location logs, we found that the number of stay-points (Figure 5.8e) and computed user speed are positively associated with alcohol consumption. The percent of time the mobile screen was *on* is negatively associated with alcohol consumption with β-value of -0.41 (Figure 5.8f).

Amongst all features within all feature groups, user speed computed using location logs has the highest positive β value (1.21). To contextualize this finding, we examined the distribution of user speeds for both alcoholic and non-alcoholic nights. A participant was considered stationary if the computed speed was below 1 km/h, and a participant was assumed to be walking if the speed was in the range of 2–5 km/h [69]. In Figure 5.10, we show the percentage of time (as a fraction of a user-night duration) when the user was stationary or walking. For more than 65% (resp. 33%) of non-alcoholic (resp. alcoholic) user-nights, a participant was stationary throughout the night, shown in Figure 5.10a.

5.6.4 Inference of Drink Status

Classification Method: We formulated the inference of drink status as a binary classification task using Random Forests [36]. In our experimental setup, we set the number of trees to $N = 500$. We experimented with different values of N, but the classification accuracy remained fairly stable.

Performance Evaluation: Each user-night was represented by a feature vector from one or multiple feature groups. Since the user-nights were imbalanced (67% of user-nights reported alcohol consumption), we randomly sub-sampled the majority class (i.e., alcohol consumption user-nights) to build 10 different balanced datasets. By training and evaluating the classifiers on balanced datasets, we built a classifier to predict binary alcohol consumption. For each balanced set, we used the 10-fold cross-validation approach. To guard against the case where the data of one user was distributed between both the training and the test sets, the 10-folds were created based on disjoint users (as opposed to the user-night). The final classification accuracy was computed using the cross-validation accuracy averaged over 10 balanced sets. Using this experimental setting, the baseline accuracy for the classification task is 50% (i.e., random guess).

Results

To highlight the contribution of each feature group, we first report the classification accuracy using each feature group separately in Table 5.7. As can be seen, accelerometer data is the most informative feature among all features types, with 75.8% accuracy. Features extracted using location logs are the second best feature with an accuracy of 68.5%. Next, we see that Wifi and Bluetooth logs are also discriminant with the number of unique wifi access points and nearby bluetooth devices observed during a user-night selected as the most important features (see Figure 5.8d and Figure 5.8c). Furthermore, features extracted using screen, application, and battery logs are relatively less good features.

Feature Group	Accuracy	Feature Combination	Accuracy
Accel.	75.8%		
Location	68.5%	A+L	76.3%
Bluetooth	64.2%	A+L+B	76.4%
WiFi	65.2%	A+L+B+W	76.7%
Apps	61.1%	A+L+B+W+Ap	76.2%
Battery	58.8%	A+L+B+W+Ap+Ba	76.4%
Screen	61.1%	A+L+B+W+Ap+Ba+S	76.2%

Table 5.7 – Classification accuracy of alcohol consumption using each and different combination of sensor data types.

Table 5.7 also reports the classification accuracies of alcohol consumption using a combination of different feature groups. Using all the feature groups, we obtained an overall accuracy of 76.2%, which is a modest increase over the accuracy achieved using just accelerometer features (75.8%). Accelerometer features, when combined with location, bluetooth and wifi features achieve the highest overall accuracy of 76.7%. In summary, these results demonstrate the feasibility to automatically classify drinking behavior with promising accuracy using smartphone data.

5.6.5 Role of Going Out and Gender on Drinking Events

To contextualize and reflect on the regression and inference results, in this section we discuss the role of two potential confounding variables – going out and gender – on the reported alcohol consumption.

Role of Going Out on Drinking Behavior

For many people, consuming alcohol is inherently a social activity [116, 117], which typically happens outside the home environment in pubs, bars, restaurants, nightclubs, etc. In the previous sections (Sections 5.6.3 and 5.6.4), it has been shown that accelerometer, location, bluetooth, and wifi are the top predictive groups for alcohol consumption. Accelerometer features are associated with users' physical activities (e.g., standing, sitting, etc.); location features inform the physical mobility of users; bluetooth and wifi features can be related to the social context of users' environment (e.g., alone or in a group). Given the discriminative nature of these feature groups to characterize alcohol consumption, it can be argued that the list of features extracted for the regression and inference model are in reality capturing participants' social and physical activity, and only indirectly informing drinking behavior. As a result, we examine the role of going out behavior (i.e., going away from home) together with alcohol consumption.

	Alcohol Consumption	
	Negative	Positive
Not Gone Out	104	84
Gone Out	76	365

(a) Going Out Behavior

	Alcohol Consumption	
	Negative	Positive
Female	190	317
Male	141	361

(b) Gender

Table 5.8 – Role of a) Going out behavior, b) Gender on alcohol consumption.

As part of the study protocol when the participants documented their drinks, we also asked them to report the place (e.g., home, bar, restaurant, etc.) where they were having their drink (Section 5.4). As a result, using this self-reported information, we knew the type of place where the participants were at the time of reporting drinks. A participant could be at multiple venues during a night. To handle this case, if participants had reported at least one place outside their home during a night, they were considered as gone out for the night; and if participants had reported staying at home through out the night, they were considered as staying at home. Of all 1,011 user-nights, participants self-reported place information for 629 (62.2%) user-nights. Note that using the place survey, we received a total of 1,394 check-ins (Table 5.1), but due to sensor data filtering (Section 5.6.1), we were left with only 629 check-ins for which self-reported place information was available.

Table 5.8a shows the total number of user-nights of drinking contrasted with going out behavior. Out of 629 user-nights, participants had gone out for 441 (70%) user-nights, while they had stayed at home for 30% of nights. If a participant had gone out, it was highly likely that they had consumed alcohol (82.8% of user-nights). On the other hand, for 44.7% of user-nights, participants stayed at home and consumed alcohol there, supporting the increasing trend of home drinking [72]. In [72], using focus groups spanning different age segments, the authors reported that cost, convenience, safety, and social occasions are the key reasons of drinking at home. In the same study young people aged 16–25 years old (the age group of our study participants) reported social occasions to be the principal reason for home drinking. Given that the likelihood of consuming alcohol at home (44.7%) and not consuming alcohol away from home (17.2%) is not negligible, it is fair to say that the extracted features might be capturing more nuanced notions of drinking behavior, and not just informing the social and physical context of users.

Role of Gender on Drinking Behavior

As stated before, it is known that gender plays a determinant role in alcohol use. In [223], using 16 population surveys across 10 countries, the authors found that males drank more frequently and consumed more alcohol relative to females. Similar findings were reported in another large-scale study conducted amongst an adolescent population in Europe [120]. To inspect the gender differences on alcohol consumption, in this section we present results of reported alcohol consumption differentiated by gender as observed in our study.

From a total of 160 participants who contributed sensor data for 1,011 user-nights (Section 5.6.1), gender information was available for 159 participants. We observed a fairly balanced gender ratio with 85 males (53.45%) and 74 females (46.55%). Males contributed a total of 502 user-nights, and females provided slightly higher number of 507 user-nights. Table 5.8b reports the total number of user-nights by each gender and their reported alcohol (and non-alcohol) consumption. Of all the user-nights reported by the female cohort, 62.5% (resp. 37.5%) of them were reported with alcohol (resp. non-alcohol) consumption. On the other hand, males reported 72% of nights with alcohol consumption, which is higher relative to the female population. These results show that gender alone cannot explain the patterns observed in Sections 5.6.3 and 5.6.4.

In summary, Table 5.6 and 5.7 provide evidence of the feasibility to characterize and classify drinking behavior using smartphone data. These results highlight the discriminative power of physical activity (accelerometer), physical mobility (location) and social context (wifi and bluetooth) features as partly predictive of alcohol consumption. From an alcohol epidemiology point of view, we believe it is a promising result. Smartphones, which can collect a wide range of sensor data (activity, location, proximity), event logs (including all phone applications), and media (photos, audio, video), offer a promising alternative to improve the way in which drinking-related phenomena can be studied. Furthermore, the recurrent use of retrospective surveys, which induces recall biases, can be potentially reduced using smartphones.

With the increasing availability and acceptability of wrist-worn wearable devices (like Jawbone, Fitbit, Apple Watch), which typically embed accelerometers inside them, our findings point to interesting ways to track drinking behavior and promote self-reflection based on behavioral data. As potential applications, near real-time information on consumption could potentially influence alcohol intake and could assist them in their decision-making to drink more or not. It has been shown that self-monitor and real-time feedback can help reduce alcohol consumption and heavy drinking [215].

5.7 Video Content Analysis

In the previous sections, we have presented a detailed statistical and inference analysis using survey and sensor data. In this section, we turn our attention to examine the video dataset to answer RQ5.4. Concretely, we explored the use of signal processing to measure two dimensions of place ambiance – loudness and brightness – automatically from collected videos. The automatic processing and characterization of place ambiance is a scalable alternative to manual data labeling. In contrast with prior work on computational modeling of places [46, 214], we study to what extent automatically extracted features represent the in-situ levels of loudness and brightness of nightlife places, as perceived by both study participants and external observers of videos. Furthermore, we also compared in-situ self-reports with the crowdsourced annotations by external online observers.

To answer these questions, we formulated the problem in terms of Brunswik's lens model [38], a model often used in human perception research [82]. The model requires computing *cue utilization*, *cue validity* and *observer accuracy* of different features. In Brunswik's terms and as described here [82], cue validity (r_v) is the correlation between automatically extracted ambiance and perceived in-situ ambiance (via self-reports); cue utilization (r_v) refers to the correlation between automatically extracted ambiance and manually coded perceived ambiance by external observers after watching the videos; and, observer accuracy (r_{acc}) is the correlation between perceived in-situ and manually coded ambiance. Higher r_v and r_u is obtained if what was perceived by the machine was equally perceived by both in-situ and external observers respectively; while, higher r_{acc} indicates that the visual cues were both utilized and valid resulting in similar perceptions by both in-situ and ex-situ observers.

On one hand, in-situ self-reports may be closer to the ground-truth, but they also include individual biases from participants, in part due to the explicit context of nightlife and potential alcohol use. On the other hand, manual coding may resemble automatically processed video content. Manual coding is also affected by individual biases, but these biases are smoothed by aggregating annotations by multiple coders. Our hypothesis was that cue utilization would be higher than cue validity in our data.

We believe it is important to examine the reliability of different crowd-workers (in-situ and ex-situ) in comparison with automatic feature extraction to better understand their strengths and biases to inform various aspects of nightlife. Are in-situ self-reports reliable when self-reporting itself might be affected by the situation being studied (e.g., nightlife and potential alcohol use)? In these situations, what can be considered the "ground-truth"? To the best of our knowledge, this kind of analysis has not been reported earlier in the Ubicomp community.

5.7.1 Video Dataset

In our study, participants recorded a total of 894 videos, 51 of which were either corrupted or had null size. The remaining 843 (94%) videos had a mean duration of 9.4 seconds. 73% of the videos lasted exactly 10 seconds (the default setting while capturing videos via the app). In what concerns this analysis, the duration of videos limits the amount of information captured in videos, and thus, how accurate they may represent nightlife places and their ambiance. For the rest of the analysis, we use the collection of 843 videos.

5.7.2 Manual Coding and Agreement

To annotate the video corpus, we asked two research assistants to rate the ambiance dimensions (i.e., occupancy, loudness, and brightness) after watching the videos. Annotators were also asked assign categories to places documented in the videos, as was done for the place survey. We measured the inter-annotator agreement of external observers and their agreement with in-situ self-reports. For ordinal ratings, we used intraclass correlation coefficients

$(ICC(1,k))$ [189], and for categorical responses, we used Fleiss' Kappa coefficients (κ) [70].

We observed fairly high agreement between external observers for all ambiance and place attributes including occupancy ($ICC = 0.90$), loudness ($ICC = 0.86$), brightness ($ICC = 0.75$), and place category ($\kappa = 0.75$). The aggregated annotation scores for external observers also achieved high agreement with self-reports for place category ($\kappa = 0.72$); moderate agreement for occupancy ($ICC = 0.50$) and loudness ($ICC = 0.55$); and, no agreement for brightness ($ICC = -0.07$). For brightness (resp. loudness), 50% (resp. 55%) of the videos had an ordinal scale difference of 1 or 2 points with respect to self-reports and 17% (resp. 8%) of them had a difference larger than 2 points, on a five-point scale. These findings suggest that external observers annotated some ambiance attributes differently from in-situ self-reports. We investigate these differences further in Section 5.7.4.

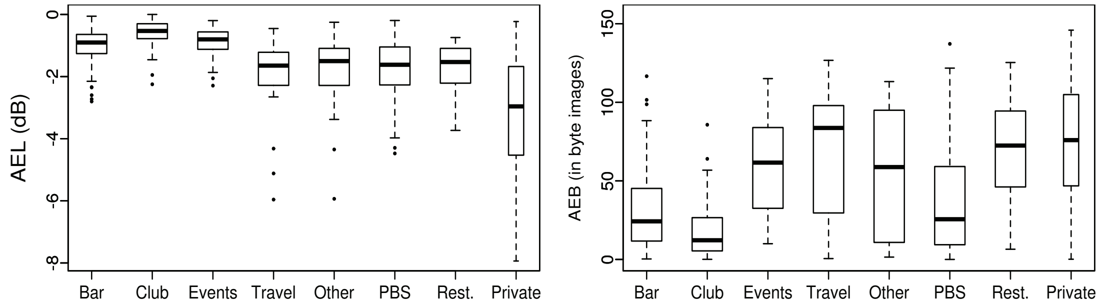
5.7.3 Feature Extraction and Analysis

We computed the loudness and brightness of videos using standard features from audio and image processing, as described below:

Automatically Extracted Loudness (AEL): We extracted the loudness of places as the audio power (AP) using the audio channel of videos [110]. The AP coefficients are computed as the average square of digital audio signal $s(n)$ within successive non-overlapping frames. For each lth frame, AP is computed as: $AP(l) = \frac{1}{N_{hop}} \sum_{n=0}^{N_{hop}-1} |s(n + lN_{hop})|^2$, where L is the total number of time frames (each of time duration L_w) and N_{hop} is the number of time samples corresponding to the time interval between consecutive frames. For each video, AEL was the mean $AP(l)$ values across all frames, with L_w set to 128ms. The higher the value of AEL, the louder the video.

Automatically Extracted Brightness (AEB): We computed the average brightness of a video using a typical measure in image processing [80]. The brightness of a frame B is determined as the average intensity of the luminance channel $Y(x,y)$ computed across all N image pixels in the YUV color space, i.e. $B = \frac{1}{N} \sum_{(x,y)} Y(x,y)$. For each video, we computed B for every frame, and then take the mean to obtain AEB. The higher the value of AEB, the brighter the video.

Figure 5.11 shows the boxplots of AEL and AEB for each place type. Overall, we observe that private places are quieter compared to public places, which is consistent with the findings reported earlier using self-reported data (Table 5.3). When comparing the median values of AEL across public place categories, clubs, events, and bars were found to be the loudest places with low statistical dispersion (Figure 5.11a). In contrast, for the videos taken in private places, the distribution of AEL shows a wide spread with varying loudness profiles. The diversity in loudness of home environments may reflect different social settings e.g., private parties, family dinners, or being alone at home (see Table 5.3).



(a) Loudness

(b) Brightness

Figure 5.11 – Boxplots of AEL and AEB by place category. AEL values are normalized with respect to the loudest place and log-scaled. AEB is measured in byte images, a 8-bit integer ranging from 0 (dark) to 255 (bright). *PBS* refers to public spaces.

When comparing the median values of automatically extracted brightness, we found clubs and bars to be the darkest places together with PBS across all place types (Figure 5.11b). These findings are not surprising as clubs and bars at night are naturally expected to be darker than other indoor places such as restaurants (see Figure 5.4c). Among public places, restaurants have comparable distributional spread as private places. Finally, videos taken in the travel category have the highest median brightness, which mostly reflects on public transportation vehicles or stations that tend to be well-lit during night.

To validate these findings, we conducted a series of pairwise Kolmogorov-Smirnov test [133] across all place categories for AEL and AEB. For loudness, all the tests were significant at $p < 0.01$, except pair-wise tests between restaurant, travel and PBS. Similar results were obtained for AEB – all the tests were statistically significant at $p < 0.01$ except tests between bar and PBS, and between restaurant, private and travel categories.

5.7.4 Feature Reliability

To measure the reliability of different features, we first measured the cue validity (r_v) and cue utilization (r_u) of AEL and AEB by computing their pair-wise Pearson's correlations with self-reports and manual coding of loudness and brightness, respectively. Table 5.9 shows these measures for each place type. Using these statistics, we observe the following trends. First, both AEL and AEB show significant cue utilization with moderate to high effect sizes ($0.48 \leq r_u \leq 0.83$) for all places types. Second, we obtain significant cue validity with moderate effect sizes ($0.25 \leq r_v \leq 0.48$) for both AEL and AEB for some of the categories. Third, cue utilization effect sizes are overall higher than for cue validity for both AEL and AEB, i.e., automatic ambiance features describe more accurately the perception of ex-situ annotators than that of participants in-situ. Fourth, the effect sizes of public and private places are comparable. Finally, AEB values are generally higher than AEL for cue utilization.

Place Type	r_v	$p-$**value**	r_u	$p-$**value**	r_{acc}	$p-$**value**
Loudness (AEL)						
Bar	0.34	2.26e-04	0.54	3.34e-08	**0.24**	1.88e-02
Club	**0.31**	4.65e-02	0.62	1.89e-04	**0.24**	2.01e-01
Restaurant	0.44	6.27e-03	0.53	8.05e-04	0.56	2.23e-04
PBS	0.44	1.99e-07	0.48	1.68e-07	0.46	2.53e-07
Events	**0.06**	7.33e-01	0.61	2.62e-04	**0.19**	3.07e-01
Travel	**0.29**	4.24e-02	0.62	8.31e-06	**0.32**	3.46e-02
Public	0.48	8.03e-26	0.67	7.83e-50	0.50	8.19e-26
Private	0.45	5.77e-19	0.82	9.66e-82	0.47	1.21e-21
Brightness (AEB)						
Bar	0.25	5.56e-03	0.71	2.03e-15	**0.19**	6.77e-02
Club	**0.01**	5.35e-01	0.60	4.09e-04	**0.02**	9.16e-01
Restaurant	**0.34**	3.7e-02	0.83	3.08e-10	0.42	7.59e-03
PBS	0.38	8.08e-06	0.78	5.72e-23	0.38	3.25e-05
Events	**0.25**	1.63e-01	0.83	8.38e-09	**0.33**	6.76e-02
Travel	**0.31**	3.2e-02	0.72	3.94e-08	0.45	1.99e-03
Public	0.43	1.00e-20	0.80	5.63e-82	0.42	4.62e-18
Private	0.35	7.85e-12	0.73	2.78e-56	0.31	1.16e-09

Table 5.9 – Cue validity (r_v), cue utilisation (r_u), and observer accuracy (r_{acc}) for AEL and AEB (N=843). Values marked in **bold** are **not** statistically significant at $p < 0.01$.

Table 5.9 also reports the observer accuracy (r_{acc}) which was computed by correlating perceived in-situ and manually coded ambiance for each place type. Overall, public places obtained higher r_{acc} relative to private places for both loudness and brightness. Of all place types, bars, clubs, and events obtained the lowest r_{acc} which were not statistically significant at $p < 0.01$. These results indicate that the experience of physically present in those places was not perceived in the same manner after watching the video of those places.

Apart from these findings, we observe that for place types with no significant cue validity, both AEL and AEB show statistically significant cue utilization (e.g., events, clubs). For these places, we watched a sample of videos where manually coded ratings were different from self-reports. While watching these videos, we did not find any obvious explanation for the difference in ratings, yet we found manual coding to be more reliable (i.e., correcting for biases and mistakes while self-reporting). For some other videos, we found that the presence of flashy lighting (common in clubs and events) may alter the perception of brightness between what is experienced in-situ and what is perceived manually. In general, we noted that videos tend to capture more diversity of content compared to audio content, which may explain higher cue utilization of AEB than AEL.

Overall, the analysis confirms the diversity of our video dataset, and the value of understanding data quality using a combination of automatic and manual methods. In the field of Ubicomp research and mobile crowdsensing studies, it is important to examine the reliability of different crowd-workers (in-situ and ex-situ) in comparison with automatic feature extraction to better understand their strengths and biases to inform various aspects of places and people. Our findings advances the research in this direction. We believe that the video dataset will open the door to additional research questions.

5.8 Participants Experiences of Mobile Crowdsensing Study

In this subsection, we first report the overall experience of participants, then we list the results on participants' compliance with the video recording task, and finally highlight a few observations from the qualitative interviews.

5.8.1 Overall Experience

After the field study concluded, participants were asked to complete an exit questionnaire about their experiences during the study (Section 5.4). For the exit questionnaire we received responses from a total of 201 users. From an overall usability point of view, more than 80% of participants *agreed* or *strongly agreed* that the mobile application (i.e., survey logger) was intuitive and easy to use. For 49% of participants, the use of the application became a routine after a while that participants did without having to think about it too much. From the perspective of answering in-situ questionnaires, 34% of participants found them hard as the app interfered with their night outs. Using the application, 38% of participants reflected on the place ambiance for the first time. When asked whether the mobile application affected their alcohol consumption, 79.6% of participants stated that the app did not had any effect on how much they drank compared to their usual consumption. Overall, these findings suggest that most of the participants had an acceptable experience with the study.

5.8.2 Compliance of Video Crowdsourcing

Now, we report the results on participants' compliance with the video recording task. We refer to compliance as the extent to which participants carried out the assigned task (video taking). We measured compliance on two aspects: 1) participants recording the video after check-in to a place, and 2) participants following the instructions given for video recording.

Recorded Videos: As stated before, participants recorded a total of 894 videos. Of all check-ins to public (resp. private) places, 68% (resp. 66%) resulted in a video (Table 5.3), suggesting that the video taking was not significantly different based on participants' location. When participants did not take a video, they had to specify one or more of the five predefined reasons. Of all the 429 check-ins with no video, safety, ethical, and social were the top three cited reasons (around 30% each). Only for 5% of cases, legal reasons were indicated for not recording a video, which in itself points towards ways in which participants perceived and conceptualized video taking in public and private places. When comparing the differences across place types, we found that social reason was specified more frequently in private places than in public ($\chi^2 = 16.56; p - value = 4.7e - 05$); while safety reason was specified more in public places than in private ($\chi^2 = 5.69; p - value = 0.017$).

Video Attributes: Participants were instructed to take a 10-second video capturing a panorama with the phone in landscape (horizontal) mode. We used manual coding to annotate: 1) video orientation, and 2) if the video was captured with the camera panning. Based on the aggregated ratings, we found that 72% of the videos were recorded vertically, and 76% of videos captured a panoramic scene. Overall, we observe low to moderate compliance for these video attributes.

5.8.3 Qualitative Experience of Participants

To conclude this section, we report some key observations of participants' video-taking experience as shared on the qualitative interviews (Section 5.4). One participant felt awkward taking videos in indoor and dark environments, but felt comfortable in outdoor places: "*It was somewhere indoor, and it was quite dark and it was pretty calm ... I can't just pull out my phone and film with a strong flash. I think people would have wondered what I was doing. But when I was simply outdoor, I didn't care*". Interestingly, some other participants did not differentiate between places to record, which could be interpreted as their indifference towards place type for video capturing. When asked about how recording a video made them feel unsafe, one participant noted: "*Because people don't like it. It could have led to conflicts. For example, people I don't know, they would have the feeling I film them, then they go nuts*" Apart from the stated reasons, few participants reported being too drunk, forgetful, or low battery for not taking a video.

5.9 Conclusion

In this chapter, we presented a mobile crowdsensing study to capture and examine the nightlife patterns of young people in Switzerland. The study resulted in a collection of 1,394 place visits accumulating over 8 million sensor data points and 894 videos over three months. Using place visits, we gained insights into the heterogeneity of youth hangouts during night, which were diverse from the perspective of social and physical ambiance dimensions. Using features extracted from sensor data, we demonstrated the feasibility to automatically classify drinking behavior with an overall accuracy of 76.7%. Finally, using the video corpus, we concluded that automatically extracted loudness and brightness features describe more accurately the perception of ex-situ annotators than that of in-situ participants. To conclude, we believe that the developed methodology demonstrated the promise of collecting rich data to improve our current understanding of patterns of physical mobility, activities, and social context of youth populations, as they experience nightlife.

6 Conclusion and Future Work

Understanding the psychological dimensions of urban spaces – the feeling associated with a particular place – under the construct of visual perception is an emerging subject in social and multimedia computing. The growth of mobile and social technologies are providing new opportunities to document, characterize, and gather impressions of urban environments at scale with high spatial and temporal resolution. In this thesis as a first contribution, we took a multidisciplinary approach to quantify the perception of indoor and outdoor environments, which until recently had largely remained unexplored. Urban impressions were elicited using images as visual stimuli along several psychological and physical dimensions that included and extended those studied in recent literature. Using recent advancements in deep learning, we proposed a methodology to automatically infer urban perception using state-of-the-art visual cues extracted from images. Our results advance the research of visual perception of places and raise interesting possibilities to conduct global studies of urban perception.

As a second contribution, the thesis proposed a mobile crowdsensing methodology to collect rich contextual data to deepen our current understanding of youth nightlife practices in Switzerland. Adopting a mix of participatory and opportunistic sensing, we examined the physical mobility, activities, and social context of urban nightlife behavior of over 200 young people for three months in two Swiss cities. Using automatically derived cues from mobile sensor and logs data, we introduced an inference framework to classify the binary state of alcohol consumption for single nights in an urban "in-the-wild" nightlife setting. Overall, the *Youth@Night* project filled the research gaps towards capturing and examining the the heterogeneity and complexity of the youth going out behavior during night time.

The chapter is structured as follows. Section 6.1 summarizes the main conclusions of each chapter of the thesis. Section 6.2 discusses the limitations of our work and suggests potential future research directions.

6.1 Conclusion

In Chapter 2, we presented a study on perception and inference of psychological dimensions of indoor places from social media content under the construct of ambiance. Using about 50,000 Foursquare images collected from 300 popular indoor places across six cities, we first assessed the suitability of social media images as data source to convey place ambiance, and found that images with clear views of the environment were perceived as being more informative of ambiance than other image categories. Second, we demonstrated that reliable estimates of ambiance can be obtained for several dimensions using Foursquare images, suggesting the presence of visual cues to form place impressions. We further found that most aggregated impressions of places were similar across cities. Third, we demonstrated the feasibility to automatically infer ambiance impressions using visual cues extracted from images. Using pre-trained convolutional neural network models, we extracted visual features and obtained a maximum R^2 of 0.53 and lower root-mean-squared errors for all dimensions relative to the baseline method.

In Chapter 3, we presented the design and implementation of the *SenseCityVity* project co-designed with over 200 young volunteers contributing over 7,000 geo-localized images, and 380 first-person perspective videos documenting three cities in Mexico. The collected images describe outdoor scenes and views of each city's built environment. Using a subset of 1,200 images, we presented a computational analysis and automatic characterization of urban perception of outdoor places for three cities in central Mexico. Using the aggregated annotations, we concluded that outdoor environments can be reliably assessed with respect to most urban dimensions. We further demonstrated the feasibility to automatically infer human perceptions of outdoor places using a variety of low-level image and deep learning features with promising accuracy. To demonstrate the additional value of collecting images via crowdsensing relative to street-level imagery, we investigated the perception of urban environments during different times of the day and found that places in the evening were perceived as less *happy*, *pleasant* and *preserved* relative to the same place in the morning. Overall, our findings are promising that could potentially provide urban designers and city planners a data-driven and scalable approach to examine the physical appearance of cities and help design urban policies informed by urban perception. As a final note, *SenseCityVity* project was carried out in a developing country, but we believe some of the findings are equally transferable to the developed cities.

In Chapter 4, we examined the use of mobile crowdsensing to document Nairobi's road quality information. First, we presented the key findings of a road quality survey in Nairobi. The survey examined key local issues including weekly travel practices, perception of current road quality conditions and their impact on daily travel experience. Second, we developed a mobile crowdsensing application to locate, describe, and photograph road hazards. We tested the application through a two-week field study amongst 30 participants who documented a total of 254 road hazards from different parts of Nairobi. Third, we demonstrated the use of online crowdsourcing to verify the authenticity of user-contributed reports i.e., to verify whether contributed images from the field indeed depicted road hazards. Overall, *CommuniSense* advances the research in the domain of citizen-based reporting, by integrating it with online crowd-based verification for quality control.

In Chapter 5, we presented a mobile crowdsensing study to capture and examine the nightlife patterns of young people in Switzerland. The study resulted in a collection of 1,394 place visits accumulating over 8 million sensor data points and 894 videos over three months. Using place visits, we gained insights into the heterogeneity of youth hangouts during night, which were diverse from the perspective of social and physical ambiance dimensions. Using features extracted from sensor data, we demonstrated the feasibility to automatically classify drinking behavior with an overall accuracy of 76.7%. Finally, using the video corpus, we concluded that automatically extracted loudness and brightness features describe more accurately the perception of ex-situ annotators than that of in-situ participants. To conclude, we believe that the developed methodology demonstrated the promise of collecting rich data to improve our current understanding of patterns of physical mobility, activities, and social context of youth populations, as they experience nightlife.

6.2 Limitations and Future Work

In this section, we discuss the potential limitations of our work and possible future research directions along three main research themes presented in this thesis: mobile crowdsensing, urban perception of places, and youth nightlife patterns.

6.2.1 Mobile Crowdsensing

In this subsection, we highlight potential research directions to address two fundamental challenges in mobile crowdsensing studies.

Incentives Mechanisms: Except the *SenseCityVity* project, in both the *Youth@Night* and *CommuniSense* studies participants were given monetary incentives for data collection. In *SenseCityVity*, participating volunteers were altruistically motivated to contribute their time and resources for the study. To get longitudinal data over months and possibly years, active volunteer participation is a must for the crowdsensing study to succeed. Future work could investigate the design of incentive mechanisms – social, financial, reputation or a combination – to maximize sustained user participation. There is a growing body of work which have looked into incentive schemes under different settings [188, 114, 131]. One could incorporate the findings from these works towards the design of future crowdsensing studies. In mobile crowdsensing research and to the best of our knowledge, incentive mechanisms have been less studied for participant recruitment. A significant amount of effort, time and resources were spent during the recruitment process in all our crowdsensing studies. Future work could also design effective incentive strategies (e.g., hierarchical referral scheme [42, 213]) to enroll a large and diverse population.

Crowdsourced Data Verification: In Chapter 4, we demonstrated the potential of using a crowdsourcing approach to verify the authenticity of the crowdsourced data. We used an online crowdsourcing platform for verification, but future work could integrate these quality control techniques in citizen-based mobile reporting applications, such as *SenseCityVity* or *Youth@Night* mobile apps. This would also involve designing appropriate incentives, as proposed above, to maximize user participation.

6.2.2 Urban Perception of Places

In this subsection, we highlight the potential limitations and future directions for the work on urban perception of indoor and outdoor places, as described in Chapter 2 and Chapter 3.

Lack of Ground Truth: As a first limitation, we believe that the lack of ground-truth or "gold standard" on perceptual ratings makes it difficult to contextualize some of the findings reported in this thesis. Perceptual ratings are inherently subjective in nature, and there are no objective measures to quantify and evaluate them. This can be regarded as one of the main limitations of our work.

For most of the psychological dimensions (e.g., *romantic*, *happy*, *pleasant*, etc.) studied in this thesis, there exist no unique ground-truth, while for some of the physical dimensions (e.g. *dangerous*, *polluted*, *loud*, etc.), there might be proxy measures to quantify them. Previous studies in outdoor places have examined the relationship between the perceptions of *safety* and *class* and homicides rates in New York City [176]. However, due to the lack of publicly available data in the studied cities in Mexico, such analysis was not feasible on the *SenseCityVity* images. Future work could include partnerships with the city or police to investigate whether this information could be available for research purposes.

Furthermore, to contextualize some of the findings or evaluate the applications of the perception work, an interesting analysis would be to gather impressions by domain "experts" (e.g., designers, architects, city planners), who are responsible for designing these urban spaces. This would facilitate creation of a "gold standard" for visual perception research in urban places, in addition to comparing and quantifying the predictive validity of some of the proposed techniques presented in this thesis.

Examination of "Other" Places: For indoor places, our analysis showed that popular places in the studied cities evoked similar perceptions for most of the ambiance dimensions. This result is interesting in itself as it might suggest that popular places in social media in cosmopolitan cities have many points in common, potentially suggesting a "uniformization of taste" in globalized cities [50]. However, any possible interpretation would have to be further validated with more data and qualitative studies. These results also highlight the need to study other venues, including the long tail of not so popular places and venues not represented on Foursquare because of well-known socio-economic biases in social media [206, 175]. The intrinsic biases of social media result in few places being well represented (as the ones studied in this thesis), but the majority of urban indoor venues that are poorly represented (if at all) could be also studied under the ambiance lens. Could significant differences in impressions between popular and non-popular venues be quantified? Could one devise transfer learning approaches that could be used to learn from popular places and adapted for non-popular ones? These are just a couple of research questions that could be investigated as part of future work. The ability to automate the image selection process (as demonstrated in Section 2.7.4) provides potential avenues to pursue these research directions by increasing the number and diversity of places.

With respect to outdoor places, due to the nature of our data collection, the spatial coverage of our approach can be seen as a potential limitation. Spatial coverage includes two aspects. The first one is the spatial sampling of regions to select urban areas. We did not perform any uniform sampling to select places for the *SenseCityVity* study, which we plan to do as future work for comparison purposes. The second aspect is the spatial scalability, which involves reaching to diverse geographical regions. Even though *SenseCityVity* images were visually diverse, our data collection methodology was limited to areas that could be reached by the participant community. Having said that, in the context of development, it is more relevant for people to explore the urban area where they live and work in order to realize solutions to the problems they face on a daily basis. We plan to engage other communities in the future to collect diverse datasets in other developing cities.

Modeling Crowdworkers: While aggregating annotations from multiple workers, we did not take into account either the (hidden) quality of workers (i.e., different annotators have different work qualities), or the inherent difficulty of certain images for annotation. Both of which might affect the consensus score [218, 164]. To guard against noisy and (possibly) adversarial annotations, a promising future direction could involve application of probabilistic methods to learn these latent worker and task attributes and weigh their relative importance during aggregation [218, 164, 106]. It would also be worth investigating the effect of worker quality and task difficulty on the automatic inference of urban perception.

Besides these latent attributes, the demographics (age, gender, socio-economic status) including the “locality” of the worker population also plays an important role. By locality, we mean if the observer population is external (as is the case in our study) or locals who are familiar with the indoor/outdoor environments. It can be argued that the collected assessments by external observers induce bias in the ratings and thus limit the generalizability of our findings. We acknowledge this is one of the limitations of our work. Future work will investigate the impressions elicited by various demographics segments.

Machine Perception of Places: In this thesis, we established that it is feasible to automatically infer human perceptions of places using visual cues. However, the understanding of what specific visual cues raters used to judge a place and form ambiance impressions is an open issue. For indoor places, these cues could be subtle including color, lighting scheme, spatial layout, flooring and carpeting, wall decorations or patrons [24, 208, 90]). For outdoor places environmental cues could include architectural and landscape elements such as wall color, greenery, trees and hedgerows, corporate buildings, residential houses, street signs, graffitis, to mention a few [61, 22, 162]. Future work could investigate these specific connections.

Furthermore, a holistic study of place perception would involve the analysis and interpretation of multiple modalities including images, videos, sound, and text. Each of these modalities involve real challenges. In this thesis, we have used visual cues from images. However, there is a variety of potentially other informative cues e.g., visual cues from videos, acoustic cues that could be collected in-situ [214], text from user-generated comments, and metadata. The specific connections of these features with ambiance still need to be established. With respect to cues extracted from images, future work could also investigate fusion of other visual features such as presence of faces, demographics of patrons, image quality, or computational aesthetics features to deepen our understanding and improve the predictive performance [85, 166, 108].

For extracting deep learning features, we have used a pre-trained CNN model. Future work could fine-tune an existing pre-trained CNN or train a CNN model from scratch to see if improvements can be made relative to the reported results [159, 52]. From a machine learning perspective, we presented early results on the effect of the amount of data and image types used on learning ambiance (Section 2.7.4), but future work could inspect their roles in more depth. Future work could also examine the role of the inherent data bias caused due to an unbalanced distribution of annotation scores [102].

6.2.3 Urban Crowdsourcing of Nightlife Patterns

In this subsection, we highlight two promising research directions for the *Youth@Night* study.

Contextualizing Alcohol Consumption: The data collected as part of the *Youth@Night* study provides numerous opportunities to explore different angles of drinking behavior. In this thesis, we studied one aspect of drinking behavior. Alcohol use is context dependent, so from a scientific point of view, understanding the temporal aspects of consumption and how it affects intake is an interesting area of future exploration. One could also explore how the social context and mobility cues affect drinking behavior e.g., does increased mobility or large social groups increase the likelihood of alcohol consumption? The *Youth@Night* data would allow to empirically test these hypotheses, some of which have been analyzed in the literature using questionnaires [116, 122].

Besides mobility and social context, does the socially perceived ambiance of a place affect alcohol consumption? Do people drink more in dark and dingy places? To address these questions, the automatic identification of place ambiance as being ‘dark’, ‘dingy’, ‘trendy’, or ‘formal’ from smartphone sensor data and images (as demonstrated in this thesis) could provide an entirely new type of insights for alcohol research [96].

Computational Characterization of Places and Other Nightlife Activities: In Chapter 5, we found that 47% of check-ins and video recordings occurred in private venues including personal homes. The rich data generated in a private setting (survey questionnaires, videos, and mobile sensor data) provides a unique opportunity to study private places. Using a combination of mobile sensor and visual data, future work will propose a computational methodology to automatically discriminate private *vs.* public places at night. Moreover, we could infer certain activities of nightlife, other than alcohol consumption, like dancing in a club.

Bibliography

[1] Citizens Connect: Making Boston Beautiful. http://bit.ly/1a4asN3. Retrieved August 15, 2016.

[2] Foursquare API August 2014 Update. https://developer.foursquare.com/docs/2014update. Retrieved July 24, 2016.

[3] Foursquare Category Hierarchy. https://developer.foursquare.com/categorytree. Retrieved July 15, 2016.

[4] GADM Database of Global Administrative Areas. http://www.gadm.org/. Retrieved August 15, 2016.

[5] Gigwalk. http://www.gigwalk.com/. Retrieved July 15, 2016.

[6] ma3route. https://twitter.com/ma3route. Retrieved August 15, 2016.

[7] Map Kibera. http://mapkibera.org. Retrieved July 24, 2016.

[8] Microsoft Data Gathering. https://www.microsoftdatagathering.net/. Retrieved August 15, 2016.

[9] mSurvey. https://www.msurvey.co. Retrieved August 15, 2016.

[10] RapidSMS. https://www.rapidsms.org/. Retrieved August 15, 2016.

[11] TaskRabbit. http://www.taskrabbit.com/. Retrieved July 15, 2016.

[12] Vodacom Tanzania. http://vodacom.co.tz. Retrieved August 15, 2016.

[13] Nairobi City traffic jam costs Kenya 37 billions shillings annually. http://bit.ly/1ssjate, 2014. Retrieved August 15, 2016.

[14] Caminos de la Villa. https://www.caminosdelavilla.org/, 2016. Retrieved July 24, 2016.

[15] Field Agent Mobile Market Research & Audits. https://www.fieldagent.net/, 2016. Retrieved July 15, 2016.

[16] FixMyStreet: Report, View, and Discuss Local Street-related Problems. https://www.fixmystreet.com, 2016. Retrieved July 24, 2016.

[17] SeeClickFix. http://seeclickfix.com, 2016. Retrieved July 24, 2016.

[18] Ushahidi. http://www.ushahidi.com, 2016. Retrieved July 24, 2016.

[19] A. A. Ali, S. M. Hossain, K. Hovsepian, M. M. Rahman, K. Plarre, and S. Kumar. mPuff: Automated Detection of Cigarette Smoking Puffs from Respiration Measurements. In *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*, pages 269–280. ACM, 2012.

[20] K. Ali, D. Al-Yaseen, A. Ejaz, T. Javed, and H. S. Hassanein. CrowdITS: Crowdsourcing in Intelligent Transportation Systems. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 3307–3311. IEEE, 2012.

[21] N. Ambady, M. Hallahan, and R. Rosenthal. On Judging and Being Judged Accurately in Zero-acquaintance Situations. *Journal of Personality and Social Psychology*, 69(3):518, 1995.

[22] S. M. Arietta, A. A. Efros, R. Ramamoorthi, and M. Agrawala. City Forensics: Using Visual Elements to Predict Non-Visual City Attributes. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2624–2633, 2014.

[23] M. D. Bader et al., S. J. Mooney, Y. J. Lee, D. Sheehan, K. M. Neckerman, A. G. Rundle, and J. O. Teitler. Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions. *Health & place*, 31:163–172, 2015.

[24] J. Baker, D. Grewal, and A. Parasuraman. The Influence of Store Environment on Quality Inferences and Store Image. *Journal of the Academy of Marketing Science*, 22(4):328–339, 1994.

[25] S. Bakhshi, P. Kanuparthy, and E. Gilbert. Demographics, Weather and Online Reviews: A Study of Restaurant Recommendations. In *Proceedings of the 23rd international conference on World wide web*, WWW ’14, pages 443–454. ACM, 2014.

[26] S. Bakhshi, D. A. Shamma, and E. Gilbert. Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, pages 965–974. ACM, 2014.

[27] K. Banks and E. Hersman. FrontlineSMS and Ushahidi – A Demo. In *Information and Communication Technologies and Development (ICTD)*, pages 484–484. IEEE, 2009.

[28] L. Bao and S. S. Intille. Activity Recognition from User-annotated Acceleration Data. In *Pervasive computing*, pages 1–17. Springer, 2004.

[29] B. Baykurt. Redefining Citizenship and Civic Engagement: Political Values Embodied in FixMyStreet.com. *Selected Papers of Internet Research*, 1, 2012.

[30] T. Bills, R. Bryant, and A. W. Bryant. Towards a Frugal Framework for Monitoring Road Quality. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 3022–3027. IEEE, 2014.

[31] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[32] M. Bohmer, B. Hecht, J. Schoning, A. Kruger, and G. Bauer. Falling Asleep with Angry Birds, Facebook and Kindle: A Large Scale Study on Mobile Application Usage. In *Proceedings of the 13th international conference on Human computer interaction with mobile devices and services*, pages 47–56. ACM, 2011.

[33] A. Bosch, A. Zisserman, and X. Munoz. Representing Shape with a Spatial Pyramid Kernel. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, CIVR '07, pages 401–408, New York, NY, USA, 2007. ACM.

[34] M. Bott and G. Young. The Role of Crowdsourcing for Better Governance in International Development. *Praxis: The Fletcher Journal of Human Security*, 2:47–70, 2012.

[35] D. Boyd. *It's Complicated: The Social Lives of Networked Teens*. Yale University Press, 2014.

[36] L. Breiman. Random Forests. *Machine learning*, 45(1):5–32, 2001.

[37] W. Brunette, M. Sundt, N. Dell, R. Chaudhri, N. Breit, and G. Borriello. Open Data Kit 2.0: Expanding and Refining Information Services for Developing Regions. In *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, page 10. ACM, 2013.

[38] E. Brunswik. *Perception and the Representative Design of Psychological Experiments*. Univ of California Press, 1956.

[39] J. BURKE. Participatory Sensing. In *Proc. Workshop on World Sensor Web at SenSys (WSW'06), Oct.*, 2006.

[40] S. Carr. *Public Space*. Cambridge Press, 1992.

[41] M. Castells and P. Himanen. *Reconceptualizing Development in the Global Information Age*. Oxford University Press, USA, 2014.

[42] M. Cebrian, L. Coviello, A. Vattani, and P. Voulgaris. Finding Red Balloons with Split Contracts: Robustness to Individuals' Selfishness. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 775–788. ACM, 2012.

[43] P. Chatterton and R. Hollands. Theorising Urban Playscapes: Producing, Regulating and Consuming Youthful Nightlife City Spaces. *Urban studies*, 39(1):95–116, 2002.

[44] Y. Chen, A. Cheng, and W. Hsu. Travel Recommendation by Mining People Attributes and Travel Group Types From Community Contributed Photos. *IEEE Transactions on Multimedia*, 2013.

[45] Y. Chon, N. D. Lane, Y. Kim, F. Zhao, and H. Cha. Understanding the Coverage and Scalability of Place-centric Crowdsensing. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 3–12. ACM, 2013.

[46] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao. Automatically Characterizing Places with Opportunistic Crowdsensing Using Smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 481–490. ACM, 2012.

[47] CNTraveler. Top 5 Cities in Mexico: Readers' Choice Awards 2013. http://cntrvlr.com/1f77Jbg, 2013. Retrieved July 24, 2016.

[48] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *Proceedings of the 18th international conference on World wide web*, pages 761–770. ACM, 2009.

[49] A. Crawford. Criminalizing Sociability through Anti-social Behaviour Legislation: Dispersal Powers, Young People and the Police. *Youth Justice*, 9(1):5–26, 2009.

[50] H. Cuadra Montiel. *Globalization Approaches to Diversity*. InTech, 2012.

[51] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan. Mobile Phone Based Drunk Driving Detection. In *4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pages 1–8. IEEE, 2010.

[52] M. De Nadai, R. Vieriu, G. Zen, S. Dragicevic, N. Naik, M. Caraviello, C. Hidalgo, N. Sebe, and B. Lepri. Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life. In *Proceedings of the 24th Annual ACM Conference on Multimedia Conference (Forthcoming)*, MM '16. ACM, ACM, 2016.

[53] J. Demant and S. Landolt. Youth Drinking in Public Places: The Production of Drinking Spaces in and Outside Nightlife Areas. *Urban Studies*, 2013.

[54] A. Developers. Android BatteryManager. http://developer.android.com/reference/android/os/BatteryManager.html. Retrieved July 15, 2016.

[55] A. Developers. An overview of device characteristics and platform versions that are active in the android ecosystem. https://developer.android.com/about/dashboards/index.html#Platform, 2016. Retrieved July 15, 2016.

[56] O. Dictionary. Definition of Ambiance in English. http://bit.ly/1dAy80r. Retrieved July 24, 2016.

[57] P. M. Dietze, M. Livingston, S. Callinan, and R. Room. The big night out: What happens on the most recent heavy drinking occasion among young Victorian risky drinkers? *Drug and alcohol review*, 33(4):346–353, 2014.

[58] T. M. T. Do, J. Blom, and D. Gatica-Perez. Smartphone Usage in the Wild: A Large-scale Analysis of Applications and Context. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 353–360. ACM, 2011.

[59] T.-M.-T. Do and D. Gatica-Perez. By Their Apps You Shall Understand Them: Mining Large-scale Patterns of Mobile Phone Usage. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, page 27. ACM, 2010.

[60] T. M. T. Do and D. Gatica-Perez. The Places of Our Lives: Visiting Patterns and Automatic Labeling from Longitudinal Smartphone Data. *IEEE Transactions on Mobile Computing*, 13(3):638–648, March 2014.

[61] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. A. Efros. What makes Paris look like Paris? *ACM Trans. Graph.*, 31(4):101, 2012.

[62] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In *ICML*, pages 647–655, 2014.

[63] C. Duff. Accounting for Context: Exploring the Role of Objects and Spaces in the Consumption of Alcohol and Other Drugs. *Social & Cultural Geography*, 13(2):145–159, 2012.

[64] N. Eagle and A. Pentland. Social Serendipity: Mobilizing Social Software. *IEEE Pervasive Computing*, 4(2):28–34, 2005.

[65] N. Eagle, A. S. Pentland, and D. Lazer. Inferring Friendship Network Structure by Using Mobile Phone Data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.

[66] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan. The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring. In *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, pages 29–39. ACM, 2008.

[67] G. Esteva et al. *Grassroots Post-modernism: Remaking the Soil of Cultures*. Palgrave Macmillan, 1998.

[68] S. L. Feld. The Focused Organization of Social Ties. *American Journal of Sociology*, 86(5):1015–1035, 1981.

[69] K. Fitzpatrick, M. Brewer, and S. Turner. Another Look At Pedestrian Walking Speed. *Transportation Research Record: Journal of the Transportation Research Board*, (1982):21–29, 2006.

[70] J. L. Fleiss. Measuring Nominal Scale Agreement Among Many Raters. *Psychological bulletin*, 76(5):378, 1971.

[71] R. Florida, P. Rentfrow, K. Sheldon, T. Kashdan, and M. Steger. Place and Well-Being. *Designing Positive Psychology: Taking Stock and Moving Forward*, pages 385–395, 2011.

[72] J. Foster, D. Read, S. Karunanithi, and V. Woodward. Why do people drink at home? *Journal of Public Health*, 32(4):512–518, 2010.

[73] Foursquare. About Us Page. https://foursquare.com/about, 2016. [Online; accessed updated: 15-Aug-2016].

[74] J. Friedman, T. Hastie, and R. Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1):1, 2010.

[75] R. K. Ganti, F. Ye, and H. Lei. Mobile Crowdsensing: Current State and Future Challenges. *IEEE Communications Magazine*, 49(11):32–39, November 2011.

[76] A. Gelman and E. Loken. The Statistical Crisis in Science. *American Scientist*, 102(6):460, 2014.

[77] E. Glaeser. *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier and happier*. Macmillan, 2011.

[78] G. Gmel, J. Rehm, and E. Kuntsche. Binge Drinking in Europe: Definitions, Epidemiology, and Consequences. *Sucht: Zeitschrift fur Wissenschaft und Praxis*, 2003.

[79] E. J. Gonzales, C. Chavis, Y. Li, and C. F. Daganzo. Multimodal Transport Modeling for Nairobi, Kenya: Insights and Recommendations With an Evidence-based Model. *UC Berkeley Center for Future Urban Transport: A Volvo Center of Excellence*, 2009.

[80] R. C. Gonzalez and R. E. Woods. Digital Image Processing, 2002.

[81] S. D. Gosling, S. Gaddis, and S. Vazire. First Impressions Based on the Environments We Create and Inhabit. *First impressions*, pages 334–356, 2008.

[82] S. D. Gosling, S. J. Ko, T. Mannarelli, and M. E. Morris. A Room With A Cue: Personality Judgments Based on Offices and Bedrooms. *Journal of Personality and Social Psychology*, 82(3):379, 2002.

[83] L. Graham and S. Gosling. Can the Ambiance of a Place be Determined by the User Profiles of the People Who Visit It? In *Proceedings of AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2011.

[84] A. Gupta, W. Thies, E. Cutrell, and R. Balakrishnan. mClerk: Enabling Mobile Crowdsourcing in Developing Regions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1843–1852. ACM, 2012.

[85] M. Gygli, H. Grabner, H. Riemenschneider, F. Nater, and L. Van Gool. The Interestingness of Images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1633–1640, 2013.

[86] K. Hanyu. Visual Properties and Appraisals in Residential Areas After Dark. *Journal of Environmental Psychology*, 17(4):301–315, 1997.

[87] K. Hara, V. Le, and J. Froehlich. Combining Crowdsourcing and Google Street View to Identify Street-level Accessibility Problems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 631–640. ACM, 2013.

[88] C. Hartung, A. Lerer, Y. Anokwa, C. Tseng, W. Brunette, and G. Borriello. Open Data Kit: Tools to Build Information Services for Developing Regions. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, page 18. ACM, 2010.

[89] T. Hatuka and E. Toch. Being Visible in Public Space: The Normalisation of Asymmetrical Visibility. *Urban Studies*, 2016.

[90] V. Heung and T. Gu. Influence of Restaurant Atmospherics on Patron Satisfaction and Behavioral Intentions. *International Journal of Hospitality Management*, 31(4):1167–1177, 2012.

[91] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. SenseCam: A Retrospective Memory Aid. In *International Conference on Ubiquitous Computing*, pages 177–193. Springer, 2006.

[92] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment.

In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 493–504. ACM, 2015.

[93] Y. Hu, L. Manikonda, and S. Kambhampati. What We Instagram: A First Analysis of Instagram Photo Content and User Types. In *Proceedings of AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2014.

[94] P. Hubbard. Regulating the Social Impacts of Studentification: a Loughborough Case Study. *Environment and Planning A*, 40(2):323–341, 2008.

[95] P. Hubbard, J. Davidson, L. Bondi, and M. Smith. The Geographies of 'Going Out': Emotion and Embodiment in the Evening Economy. *Emotional Geographies*, pages 117–134, 2005.

[96] K. Hughes, Z. Quigg, L. Eckley, M. Bellis, L. Jones, A. Calafat, M. Kosir, and N. Van Hasselt. Environmental Factors in Drinking Venues and Alcohol-related Harm: The Evidence Base for European Intervention. *Addiction*, 106(s1):37–46, 2011.

[97] IBM. A Vision of a Smarter City. How Nairobi can lead the way into a prosperous and sustainable future. http://bit.ly/1pg5hAN, 2012. Retrieved August 15, 2016.

[98] A. B. Jacobs. Looking at Cities. *Places*, 1(4), 1984.

[99] M. Jarvinen and J. Ostergaard. Governing Adolescent Drinking. *Youth & society*, 2008.

[100] M. Jayne, G. Valentine, and S. L. Holloway. *Alcohol, Drinking, Drunkeness: (Dis)Orderly Spaces*. Ashgate Publishing, Ltd., 2011.

[101] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional Architecture for Fast Feature Embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.

[102] B. Jin, M. V. Ortiz-Segovia, and S. Susstrunk. Image Aesthetic Predictors based on Weighted CNNs. In *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016.

[103] O. Juhlin. Social Media On the Road. *Kista, Sweden: Springer*, 2010.

[104] F. Jurie and B. Triggs. Creating Efficient Codebooks for Visual Recognition. In *Proceedings of the Tenth IEEE International Conference on Computer Vision*, ICCV '05, pages 604–610, Washington, DC, USA, 2005. IEEE Computer Society.

[105] H.-L. C. Kao, B.-J. Ho, A. C. Lin, and H.-H. Chu. Phone-based Gait Analysis to Detect Alcohol Usage. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 661–662. ACM, 2012.

[106] Y. E. Kara, G. Genc, O. Aran, and L. Akarun. Modeling Annotator Behaviors for Crowd Labeling. *Neurocomputing*, 160:141–156, 2015.

[107] J. Karuppuswamy, V. Selvaraj, M. M. Ganesh, and E. L. Hall. Detection and Avoidance of Simulated Potholes in Autonomous Vehicle Navigation in an Unstructured Environment. In *Intelligent Systems and Smart Manufacturing*, pages 70–80. International Society for Optics and Photonics, 2000.

[108] A. Khosla, A. S. Raju, A. Torralba, and A. Oliva. Understanding and Predicting Image Memorability at a Large Scale. In *International Conference on Computer Vision (ICCV)*, pages 2390–2398, 2015.

[109] D. H. Kim, J. Hightower, R. Govindan, and D. Estrin. Discovering Semantically Meaningful Places from Pervasive RF-beacons. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, UbiComp ’09, pages 21–30. ACM, 2009.

[110] H.-G. Kim, N. Moreau, and T. Sikora. *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. John Wiley & Sons, 2006.

[111] A. Kinai, R. E. Bryant, A. Walcott-Bryant, E. Mibuari, K. Weldemariam, and O. Stewart. Twende-twende: A Mobile Application for Traffic Congestion Awareness and Routing. In *Proceedings of the 1st International Conference on Mobile Software Engineering and Systems*, MOBILESoft 2014, pages 93–98. ACM, 2014.

[112] S. F. King and P. Brown. Fix My Street or Else: Using the Internet to Voice Local Public Service Concerns. In *Proceedings of the 1st International Conference on Theory and Practice of Electronic Governance*, pages 72–80. ACM, 2007.

[113] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–456. ACM, 2008.

[114] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton. The Future of Crowd Work. In *Proceedings of the ACM 2013 Conference on Computer Supported Cooperative Work*, pages 1301–1318. ACM, 2013.

[115] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila. Towards Rich Mobile Phone Datasets: Lausanne Data Collection Campaign. *Proc. ICPS, Berlin*, 2010.

[116] H. Kuendig and E. Kuntsche. Solitary versus Social Drinking: An Experimental Study on Effects of Social Exposures on In Situ Alcohol Consumption. *Alcoholism: Clinical and Experimental Research*, 36(4):732–738, 2012.

[117] E. Kuntsche, R. Knibbe, G. Gmel, and R. Engels. Why do Young People Drink? A Review of Drinking Motives. *Clinical Psychology Review*, 25(7):841–861, 2005.

[118] E. Kuntsche and F. Labhart. Investigating the Drinking Patterns of Young People Over the Course of the Evening At Weekends. *Drug and Alcohol Dependence*, 124(3):319–324, 2012.

[119] E. Kuntsche and F. Labhart. ICAT: Development of an Internet-Based Data Collection Method for Ecological Momentary Assessment Using Personal Cell Phones. *European Journal of Psychological Assessment*, 29(2):140–148, 2013.

[120] E. Kuntsche, M. Wicki, B. Windlin, C. Roberts, S. N. Gabhainn, W. van der Sluijs, K. Aasvee, M. G. de Matos, Z. Dankulincová, A. Hublet, et al. Drinking motives mediate cultural differences but not gender differences in adolescent alcohol use. *Journal of Adolescent Health*, 56(3):323–329, 2015.

[121] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity Recognition Using Cell Phone Accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[122] F. Labhart, S. Wells, K. Graham, and E. Kuntsche. Do individual and situational factors explain the link between predrinking and heavier alcohol consumption? An event-level study of types of beverage consumed and social context. *Alcohol and Alcoholism*, 49(3):327–335, 2014.

[123] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. A Survey of Mobile Phone Sensing. *IEEE Communications Magazine*, 48(9):140–150, Sept 2010.

[124] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T. M. T. Do, O. Dousse, J. Eberle, and M. Miettinen. From big smartphone data to worldwide research: the mobile data challenge. *Pervasive and Mobile Computing*, 9(6):752–771, 2013.

[125] J. K. Laurila, D. Gatica-Perez, I. Aad, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, M. Miettinen, et al. The Mobile Data Challenge: Big Data for Mobile Computing Research. In *Pervasive Computing*, 2012.

[126] P. J. Lindal and T. Hartig. Architectural Variation, Building Height, and the Restorative Quality of Urban Residential Streetscapes. *Journal of Environmental Psychology*, 33:26–36, 2013.

[127] Y. Liu and S. S. Jang. The Effects of Dining Atmospherics: An Extended Mehrabian–Russell Model. *International Journal of Hospitality Management*, 28(4):494–503, 2009.

[128] L. J. Loewen, G. D. Steel, and P. Suedfeld. Perceived Safety from Crime in the Urban Environment. *Journal of environmental psychology*, 13(4):323–331, 1993.

[129] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting Stress In Unconstrained Acoustic Environments Using Smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 351–360. ACM, 2012.

[130] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell. SoundSense: Scalable Sound Sensing for People-centric Applications on Mobile Phones. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, pages 165–178. ACM, 2009.

[131] A. Mao, E. Kamar, Y. Chen, E. Horvitz, M. E. Schwamb, C. J. Lintott, and A. M. Smith. Volunteering Versus Work for Pay: Incentives and Tradeoffs in Crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.

[132] D. Marquard. Zurich, the Party Town. http://bit.ly/28M8nbk, 2014. Retrieved July 15, 2016.

[133] F. J. Massey Jr. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[134] H. Matthews, M. Limb, and B. Percy-Smith. Changing Worlds: the Microgeographies of Young Teenagers. *Tijdschrift voor economische en sociale geografie*, 89(2):193–202, 1998.

[135] F. Measham and K. Brain. Binge Drinking, British Alcohol Policy and the New Culture of Intoxication. *Crime, media, culture*, 1(3):262–283, 2005.

[136] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo. Real Time Pothole Detection using Android Smartphones with Accelerometers. In *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, pages 1–6. IEEE, 2011.

[137] Merriam-Webster. Definition of Ambience in English. http://bit.ly/1Ivgs23. Retrieved July 24, 2016.

[138] P. Mohan, V. N. Padmanabhan, and R. Ramjee. Nericell: Rich Monitoring of Road and Traffic Conditions Using Mobile Smartphones. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, SenSys '08, pages 323–336. ACM, 2008.

[139] A. Monroy-Hernandez, E. Kiciman, M. De Choudhury, S. Counts, et al. The New War Correspondents: The Rise of Civic Media Curation in Urban Warfare. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1443–1452. ACM, 2013.

[140] R. Montoliu, J. Blom, and D. Gatica-Perez. Discovering Places of Interest In Everyday Life From Smartphone Data. *Multimedia tools and applications*, 62(1):179–207, 2013.

[141] M. Musthag and D. Ganesan. Labor Dynamics in a Mobile Micro-Task Market. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 641–650. ACM, 2013.

[142] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo. Streetscore – Predicting the Perceived Safety of One Million Streetscapes. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, CVPRW '14, pages 793–799, Washington, DC, USA, 2014. IEEE Computer Society.

[143] V. Naroditskiy, N. R. Jennings, P. Van Hentenryck, and M. Cebrian. Crowdsourcing Contest Dilemma. *Journal of The Royal Society Interface*, 11(99), 2014.

[144] R. E. Nisbett and T. D. Wilson. The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of personality and social psychology*, 35(4):250, 1977.

[145] C. of Lausanne. Drinking in Public Space in City of Lausanne: A Challenge for Police. http://bit.ly/28M8X8X, 2010. Retrieved July 15, 2016.

[146] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002.

[147] O. Okolloh. Ushahidi, or 'Testimony': Web 2.0 Tools for Crowdsourcing Crisis Information. *Participatory Learning and Action*, 59(1):65–70, 2009.

[148] Okuttan Mark. Safaricom spends Sh400m to boost data revenue. http://bit.ly/1ug5Qxi, 2014. Retrieved August 15, 2016.

[149] R. Oldenburg. *The great good place: Café, coffee shops, community centers, beauty parlors, general stores, bars, hangouts, and how they get you through the day*. Paragon House Publishers, 1989.

[150] A. Oliva and A. Torralba. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[151] V. Ordonez and T. L. Berg. Learning High-Level Judgments of Urban Perception. In *European Conference on Computer Vision*, pages 494–510. Springer, 2014.

[152] W. H. Organization et al. *Global Status Report on Alcohol and Health 2014*. World Health Organization, 2014.

[153] J. Ostergaard. Mind the Gender Gap! When Boys And Girls Get Drunk at a Party. *Nordic Studies on Alcohol and Drugs*, 24(2):127, 2007.

[154] M. Ostergren and O. Juhlin. Road Talk: a Roadside Location-Dependent Audio Message System for Car Drivers. *J. Mobile Multimedia*, 1(1):47–61, 2005.

[155] K. Painter. The Influence of Street Lighting Improvements on Crime, Fear and Pedestrian Street Use, After Dark. *Landscape and urban planning*, 35(2):193–201, 1996.

[156] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis. RisQ: Recognizing Smoking Gestures With Inertial Sensors On a Wristband. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, pages 149–161. ACM, 2014.

[157] K. Pearson. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[158] J. Petraitis, B. R. Flay, and T. Q. Miller. Reviewing Theories of Adolescent Substance Use: Organizing Pieces in the Puzzle. *Psychological bulletin*, 117(1):67, 1995.

[159] L. Porzi, S. Rota Bulo, B. Lepri, and E. Ricci. Predicting and Understanding Urban Perception with Convolutional Neural Networks. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, pages 139–148. ACM, 2015.

[160] J. Poushter. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. http://pewrsr.ch/1RX3Iqq, 2016. Retrieved August 15, 2016.

[161] A. Quattoni and A. Torralba. Recognizing Indoor Scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 413–420. IEEE, 2009.

[162] D. Quercia, N. K. O'Hare, and H. Cramer. Aesthetic Capital: What makes London Look Beautiful, Quiet, and Happy? In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 2014.

[163] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity Recognition from Accelerometer Data. In *AAAI*, volume 5, pages 1541–1546, 2005.

[164] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy. Learning from Crowds. *Journal of Machine Learning Research*, 11(Apr):1297–1322, 2010.

[165] A. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN Features off-the-shelf: an Astounding Baseline for Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813, 2014.

[166] M. Redi, D. Quercia, L. Graham, and S. Gosling. Like Partying? Your Face Says It All. Predicting the Ambiance of Places with Profile Pictures. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

[167] P. J. Rentfrow. The Open City. *Handbook of Creative Cities*, 2011.

[168] Robert Reid. Top 8 places to (safely) visit in Mexico now. http://bit.ly/1hIjkce, 2011.

[169] J. Ross, L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson. Who Are the Crowdworkers?: Shifting Demographics in Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI EA '10, pages 2863–2872. ACM, 2010.

[170] V. Rouillard. Remote Monitoring of Vehicle Shock and Vibrations. *Packaging Technology and Science*, 15(2):83–92, 2002.

[171] S. Ruiz-Correa, D. Santani, and D. Gatica-Perez. The Young and the City: Crowdsourcing Urban Awareness in a Developing Country. In *Proceedings of the First International Conference on IoT in Urban Space*, URB-IOT '14, pages 74–79, 2014.

[172] S. Ruiz-Correa, D. Santani, B. Ramirez Salazar, I. Ruiz Correa, F. Alba Rendon-Huerta, C. Olmos Carrillo, B. Carmina Sandoval Mexicano, A. Humberto Arcos Garcia, R. Hasimoto Beltran, and D. Gatica-Perez. SenseCityVity: Mobile Sensing, Urban Awareness, and Collective Action in Mexico. *IEEE Pervasive Computing (Forthcoming)*, 2016.

[173] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[174] J. A. Russell and G. Pratt. A Description of the Affective Quality Attributed to Environments. *Journal of Personality and Social Psychology*, 38(2):311, 1980.

[175] D. Ruths and J. Pfeffer. Social Media for Large Studies of Behavior. *Science*, 346(6213):1063–1064, 2014.

[176] P. Salesses, K. Schechtner, and C. A. Hidalgo. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS ONE*, 8(7), 07 2013.

[177] D. Salon and S. Gulyani. Mobility, Poverty, and Gender: Travel 'Choices' of Slum Residents in Nairobi, Kenya. *Transport Reviews*, 30(5):641–657, 2010.

[178] Sam Sturgis. Kids in India Are Sparking Urban Planning Changes by Mapping Slums. http://bit.ly/1LtS9S4, 2015. Retrieved July 24, 2016.

[179] R. J. Sampson et al. Seeing Disorder: Neighborhood Stigma and the Social Construction of “Broken Windows”. *Social psychology quarterly*, 67(4):319–342, 2004.

[180] A. Sano and R. W. Picard. Stress Recognition Using Wearable Sensors And Mobile Phones. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 671–676. IEEE, 2013.

[181] D. Santani, J.-I. Biel, F. Labhart, J. Truong, S. Landolt, E. Kuntsche, and D. Gatica-Perez. The Night is Young: Urban Crowdsourcing of Nightlife Patterns. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Forthcoming)*, UbiComp ’16. ACM, 2016.

[182] D. Santani and D. Gatica-Perez. Speaking Swiss: Languages and Venues in Foursquare. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM ’13, pages 501–504. ACM, 2013.

[183] D. Santani and D. Gatica-Perez. Loud and Trendy: Crowdsourcing Impressions of Social Ambiance in Popular Indoor Urban Places. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, MM ’15, pages 211–220. ACM, ACM, 2015.

[184] D. Santani, R. Hu, and D. Gatica-Perez. InnerView: Learning Place Ambiance from Social Media Images. In *Proceedings of the 24th Annual ACM Conference on Multimedia Conference (Forthcoming)*, MM ’16. ACM, ACM, 2016.

[185] D. Santani, J. Njuguna, T. Bills, A. W. Bryant, R. Bryant, J. Ledgard, and D. Gatica-Perez. CommuniSense: Crowdsourcing Road Hazards in Nairobi. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI ’15, pages 445–456. ACM, 2015.

[186] D. Santani, S. Ruiz-Correa, and D. Gatica-Perez. Looking at Cities in Mexico with Crowds. In *Proceedings of the 2015 Annual Symposium on Computing for Development*, pages 127–135. ACM, 2015.

[187] SBB/CFF/FFS. Travel home safely with the nighttime network. http://bit.ly/1UtdH8O. Retrieved July 15, 2016.

[188] A. D. Shaw, J. J. Horton, and D. L. Chen. Designing Incentives for Inexpert Human Raters. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, pages 275–284. ACM, 2011.

[189] P. E. Shrout and J. L. Fleiss. Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological bulletin*, 86(2):420, 1979.

[190] F. M. Sjoberg et al. The Effect of Government Responsiveness on Future Political Participation. *Available at SSRN 2570898*, 2015.

[191] T. Skelton. Young People’s Urban Im/Mobilities: Relationality and Identity Formation. *Urban Studies*, 50(3):467–483, 2013.

[192] T. Skelton and G. Valentine. *Cool Places: Geographies of Youth Cultures*. Routledge, 2005.

[193] S. S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, 1946.

[194] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going Deeper with Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[195] J. C. Tang, M. Cebrian, N. A. Giacobe, H.-W. Kim, T. Kim, and D. B. Wickert. Reflecting on the DARPA red balloon challenge. *CACM*, 54(4):78–85, 2011.

[196] Q. Tang, D. J. Vidrine, E. Crowder, and S. S. Intille. Automated Detection of Puffing and Smoking With Wrist Accelerometers. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 80–87, 2014.

[197] R. Teodoro, P. Ozturk, M. Naaman, W. Mason, and J. Lindqvist. The Motivations and Experiences of the On-Demand Mobile Workforce. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 236–247. ACM, 2014.

[198] J. Thebault-Spieker, L. G. Terveen, and B. Hecht. Avoiding the South Side and the Suburbs: The Geography of Mobile Crowdsourcing Markets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 265–275. ACM, 2015.

[199] J. Thrul and E. Kuntsche. The Impact of Friends on Young Adults' Drinking Over the Course of the Evening – An Event-level Analysis. *Addiction*, 110(4):619–626, 2015.

[200] H. To, S. H. Kim, and C. Shahabi. Effectively Crowdsourcing the Acquisition and Analysis of Visual Data for Disaster Response. In *IEEE International Conference on Big Data*, pages 697–706. IEEE, 2015.

[201] E. Toch and I. Levi. Locality and Privacy in People-nearby Applications. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 539–548. ACM, 2013.

[202] A. Torralba and A. A. Efros. Unbiased Look at Dataset Bias. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1521–1528. IEEE, 2011.

[203] Transport and Urban Decongestion Committee. Interim Report. http://bit.ly/1vQfRnk, June 2014. Retrieved August 15, 2016.

[204] Trefis Team. Google's Android One Platform About More Than Just Phones. http://onforb.es/1yo1s1X, 2014. Retrieved August 15, 2016.

[205] S. Tshabalala. Africa's smartphone market is on the rise as affordable handsets spur growth. http://bit.ly/1O3B026, 2015. Retrieved August 15, 2016.

[206] Z. Tufekci. Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. 2014.

[207] J. W. Tukey. The Philosophy of Multiple Comparisons. *Statistical Science*, pages 100–116, 1991.

[208] L. W. Turley and R. E. Milliman. Atmospheric Effects on Shopping Behavior: A Review of the Experimental Evidence. *Journal of Business Research*, 49(2):193–211, 2000.

[209] United Nations. 2014 Revision of World Urbanization Prospects. http://bit.ly/2bcPLDc, 2015. Retrieved August 15, 2016.

[210] G. Valentine. Angels and Devils: Moral Landscapes of Childhood. *Environment and Planning D: Society and space*, 14(5):581–599, 1996.

[211] G. Valentine. Public Space and the Culture of Childhood, 2006.

[212] I. van Liempt, I. van Aalst, and T. Schwanen. Introduction: Geographies of the Urban Night. *Urban Studies*, 52(3):407–421, 2015.

[213] A. Vashistha, E. Cutrell, and W. Thies. Increasing the Reach of Snowball Sampling: The Impact of Fixed Versus Lottery Incentives. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work &#38; Social Computing*, CSCW '15, pages 1359–1363. ACM, 2015.

[214] H. Wang, D. Lymberopoulos, and J. Liu. Local Business Ambience Characterization Through Mobile Audio Sensing. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 293–304. ACM, 2014.

[215] K.-C. Wang, M.-C. Huang, Y.-H. Hsieh, S.-Y. Lau, C.-H. Yen, H.-L. C. Kao, C.-W. You, H.-H. Chu, and Y.-C. Chen. SoberDiary: A Phone-based Support System for Assisting Recovery from Alcohol Dependence. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 311–314. ACM, 2014.

[216] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell. StudentLife: Assessing Mental Health, Academic Performance And Behavioral Trends of College Students Using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 3–14. ACM, 2014.

[217] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell. SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 295–306. ACM, 2015.

[218] J. Whitehill, T.-f. Wu, J. Bergsma, J. R. Movellan, and P. L. Ruvolo. Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise. In *Advances in neural information processing systems*, pages 2035–2043, 2009.

[219] W. F. Wieczorek, B. A. Miller, and T. H. Nochajski. Multiple and Single Location Drinking among DWI Offenders Referred for Alcoholism Evaluation. *The American Journal of Drug and Alcohol Abuse*, 18(1):103–116, 1992.

[220] F. Wilcoxon. Individual Comparisons by Ranking Methods. *Biometrics*, 1(6):80–83, 1945.

[221] C. Wilkinson and M. Livingston. Distances To On-and Off-premise Alcohol Outlets and Experiences of Alcohol-related Amenity Problems. *Drug and Alcohol Review*, 31(4):394–401, 2012.

[222] Will Connors. Google, Microsoft Expose Brazil's Favelas. http://on.wsj.com/1V3X2qI, 2014. Retrieved July 24, 2016.

[223] R. W. Wilsnack, N. D. Vogeltanz, S. C. Wilsnack, and T. R. Harris. Gender Differences in Alcohol Consumption and Adverse Drinking Consequences: Cross-Cultural Patterns. *Addiction*, 95(2):251–265, 2000.

[224] J. Q. Wilson and G. L. Kelling. Broken Windows. *Critical Issues in Policing: Contemporary Readings*, pages 395–407, 1982.

[225] World Bank. Mobile Usage at the Base of the Pyramid in Kenya. http://bit.ly/1vQfRnk, December 2012. Retrieved August 15, 2016.

[226] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, pages 3485–3492. IEEE, 2010.

[227] T. Yan, M. Marzilli, R. Holmes, D. Ganesan, and M. Corner. mCrowd: A Platform for Mobile Crowdsourcing. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 347–348. ACM, 2009.

[228] Z. Yan, J. Yang, and E. M. Tapia. Smartphone Bluetooth Based Social Sensing. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, pages 95–98. ACM, 2013.

[229] D. Yang, G. Xue, X. Fang, and J. Tang. Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing. In *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, Mobicom '12, pages 173–184. ACM, 2012.

[230] M. Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz. On the Semantic Annotation of Places in Location-Based Social Networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 520–528. ACM, 2011.

[231] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma. Recommending Friends and Locations Based on Individual Location History. *ACM Transactions on the Web (TWEB)*, 5(1):5, 2011.

[232] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning Deep Features for Scene Recognition Using Places Database. In *Advances in Neural Information Processing Systems*, pages 487–495, 2014.

[233] B. Zhou, L. Liu, A. Oliva, and A. Torralba. Recognizing City Identity via Attribute Analysis of Geo-tagged Images. In *European Conference on Computer Vision (ECCV)*, pages 519–534. Springer, 2014.

[234] S. Zukin. *The cultures of cities*. Blackwell Oxford, 1995.

Darshan Santani

Idiap Research Institute, Switzerland
E-mail: dsantani@idiap.ch
Homepage: `http://www.idiap.ch/~dsantani/`

Education

PhD in Electrical Engineering July'12 - Present
EPFL, Lausanne, Switzerland
Thesis Advisor: Daniel Gatica-Perez

MSc in Management, Technology, & Economics March 2012
ETH, Zurich, Switzerland
Thesis: "Urban Information System for the Real-Time City"

Bachelor of Engineering (B.E.) in Computer Science & Engg. July 2006
Manipal Institute of Technology, India
Thesis: "Internationalization and Architecture Enhancement on Printing Mechanisms in Firefox Web browser"

Research Experience

Social Computing Group, Idiap Research Institute, EPFL July'12 - Present
Research Assistant
Supervisors: Daniel Gatica-Perez, Idiap Research Institute, Switzerland

IBM Research Africa, Nairobi June'14 - Sept'14
Research Intern, "CommuniSense: Crowdsourcing Road Hazards in Nairobi"
Supervisors: Reginald Bryant, Osamuyi Stewart, IBM Research

Wearable Computing Lab, ETH, Zurich June'11 - Sept'11
Research Intern, "Design and Development of CoenoSense Backend Platform"
Supervisors: Martin Wirz, Prof. Gerhard Troster, ETH Zurich

SENSEable City Lab, MIT, Boston Oct'10 - April'11
Research Student, "LIVE Singapore!"
Project Page: `http://senseable.mit.edu/livesingapore/`
Supervisors: Carlo Ratti, Krisitian Kloeckl, MIT

Microsoft Research Asia, Beijing June'10 - Oct'10
Research Intern, "Towards Urban Computing with GPS-Equipped Taxis"
Supervisor: Yu Zheng, Lead Researcher, MSRA

School of Information Systems, SMU, Singapore Sep'07 - Aug'09
Research Engineer, "Spatio-temporal efficiency of a Taxi Dispatch System"
Supervisors: Rajesh K. Balan & C. Jason Woodard, SMU

School of Information Systems, SMU, Singapore Feb'07 - Aug'09
Research Engineer, "Kuala: Computational Modeling Tools"
Supervisor: C. Jason Woodard, SMU

Industrial Experience

RedHat Inc. Feb'06-Aug'06
Software Engineer, Engineering Services and Operations Group, India
Project Title: "Internationalization and Architecture Enhancement on Printing Mechanisms in Firefox Web browser"

Publications

- Darshan Santani, Rui Hu, Daniel Gatica-Perez. "InnerView: Learning Place Ambiance from Social Media Images", in *Proceedings of the ACM International Conference on Multimedia (MM'16)*, 2016
- Darshan Santani, Joan-Issac Biel, Florian Labhart, Jasmine Truong, Sara Landolt, Emmanuel Kuntsche, Daniel Gatica-Perez "The Night is Young: Urban Crowdsourcing of Nightlife Patterns" in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*, 2016
- Salvador Ruiz-Correa, Darshan Santani, Beatriz Ramirez Salazar, Itzia Ruiz Correa, Fatima Alba Rendon-Huerta, Carlo Olmos Carrillo, Brisa Carmina Sandoval Mexicano, Angel Humberto Arcos Garcia, Rogelio Hasimoto Beltran and Daniel Gatica-Perez "SenseCityVity: Mobile Sensing, Urban Awareness, and Collective Action in Mexico" in *IEEE Pervasive Computing* (*Forthcoming*), 2016
- Darshan Santani, Salvador Ruiz-Correa, Daniel Gatica-Perez "Looking at Cities in Mexico with Crowds " in *Proceedings of the ACM Symposium on Computing for Development (DEV'15)*, 2015
- Darshan Santani, Daniel Gatica-Perez "Loud and Trendy: Crowdsourcing Impressions of Social Ambiance in Popular Indoor Urban Places" in *Proceedings of the ACM International Conference on Multimedia (MM'15)*, 2015
- Darshan Santani, Jidraph Njuguna, Tierra Bills, Aisha W. Bryant, Reginald Bryant, Jonathan Ledgard, Daniel Gatica-Perez "CommuniSense: Crowdsourcing Road Hazards in Nairobi" in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)*, 2015
- Salvador Ruiz-Correa, Darshan Santani, Daniel Gatica-Perez "The Young and the City: Crowdsourcing Urban Awareness in a Developing Country" in *Proceedings of the Urb-IoT*, 2014
- Darshan Santani, Daniel Gatica-Perez "Speaking Swiss: Languages and Venues in Foursquare" in *Proceedings of the ACM Multimedia (MM'13)*, 2013
- Darshan Santani, Daniel Gatica-Perez "Revisiting the Generality of the Rank-based Human Mobility Model" in *Adjunct Proceedings of Ubicomp'13*, 2013
- Stephan Karpischek, Darshan Santani, Florian Michahelles, "Usage Analysis of a Mobile Bargain Finder Application" in *Proceedings of the 13th International Conference on Electronic Commerce and Web Technologies (EC-WEB)*, 2012
- Siddharth Sanan, Darshan Santani, K. Madhva Krishna, Henry Hexmoor, "Extension of Reeds & Shepp Paths to a Robot with Front and Rear Wheel Steer" in *Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA)*, 2006

Posters and Demos

- Jidraph Njuguna, Darshan Santani, Tierra Bills, Aisha W Bryant, Reginald Bryant, "Citizen Engagement and Awareness of the Road Surface Conditions in Nairobi, Kenya" in *Proceedings of the Fifth ACM Symposium on Computing for Development (DEV'14)*, 2014
- Darshan Santani, Daniel Gatica-Perez "#-grams: Twitter Pulse from Hashtag Co-Occurrence Networks" in *ACM Web Science Data Visualization Challenge* (WebSci'14), 2014 (**Best Student Award**)
- Darshan Santani, Rajesh Krishna Balan, C Jason Woodard, "Understanding and Improving a GPS-based Taxi System" at the *Sixth International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2008

Technical Reports

- Joan-Issac Biel, Darshan Santani, Trung Phan Thanh, Daniel Gatica-Perez "Road Tweets: Analyzing Social Media Citizen Transit Reports in Nairobi", 2016
- Yin Zhu, Yu Zheng, Liuhang Zhang, Darshan Santani, Xing Xie, Qiang Yang "Inferring Taxi Status Using GPS Trajectories", 2011
- Darshan Santani, Yu Zheng, Chih-Chieh Hung, Wen-Chih Peng, Xing Xie "Towards Urban Computing with GPS-Equipped Taxis", 2010
- Darshan Santani, Rajesh Krishna Balan, C Jason Woodard "Spatio-temporal Efficiency in a Taxi Dispatch System", 2007

Patent

Darshan Santani, Dr.B.H.S. Thimmappa, MIT Manipal "An Improved Process for the Recovery of Metallic Silver from Photographic Wastes", Indian Patent (532/CHE/2004A)

Teaching

Teaching Assistant, Computational Social Media (PhD Course) Spring 2015, 2016
Department of Electrical Engineering (EDEE), EPFL, Switzerland

Scholarships and Awards

- ACM Multimedia (ACM MM) Travel Grant, 2015
- "Best Student Award" in ACM Web Science Data Visualization Challenge, 2014
- "SUN Scholarship", CEU Summer University, 2008
- Red Hat "Lord of the Codes" Scholarship, awarded by RedHat in collaboration with IIT-Bombay, 2005
- "All India Scholarship" EEI(Delhi), 2000

Technical Skills

- Programming Languages: R, Java, Python, Processing
- Operating Platforms: Linux/Unix, Windows