Looking at Cities in Mexico with Crowds

Darshan Santani
Idiap Research Institute
EPFL, Lausanne, Switzerland
dsantani@idiap.ch

Salvador Ruiz-Correa
Center for Mobile Life
CNS-IPICYT, Mexico
src@cmls.pw

Daniel Gatica-Perez
Idiap Research Institute
EPFL, Lausanne, Switzerland
gatica@idiap.ch

ABSTRACT

Mobile and social technologies are providing new opportunities to document, characterize, and gather impressions of urban environments. In this paper, we present a study that examines urban perceptions of three cities in central Mexico (Guanajuato, Leon and Silao), which integrates a mobile crowdsourcing framework to collect geo-localized images of urban environments by a local youth community, and an online crowdsourcing platform (Amazon Mechanical Turk) to gather impressions of urban environments along twelve physical and psychological dimensions. Our study resulted in a collection of 7,000 geo-localized images containing outdoor scenes and views of each city's built environment, including touristic, historical, and residential neighbourhoods; and 156,000 individual judgments from MTurk. Statistical analyses show that outdoor environments can be reliably assessed with respect to most urban dimensions by the observers of crowdsourced images. Furthermore, a cross-city statistical analysis shows that outdoor urban places in Guanajuato (a touristic, cultural heritage site) are perceived as more quiet, picturesque and interesting compared to places in Leon and Silao, which are commercial and industrial hubs, respectively. In contrast Silao, is perceived to have lower accessibility than Leon. Finally, we investigate 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. Through the use of collective action, participatory sensing and mobile crowdsourcing, our study engages citizens to understand socio-urban problems in their communities.

Categories and Subject Descriptors: H.4.m [Information Systems Applications]: Miscellaneous

Keywords: Mobile Crowdsourcing; Urban Perception; Outdoor Places; Collective Action; ICTD; Mexico

1. INTRODUCTION

Community awareness and action on urban problems are longstanding practices in developing countries [13]. 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 [11].

In this context, mobile and social technologies are providing new opportunities to document, characterize, map, and ultimately address urban problems in developing cities. Mobile crowdsourcing 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 [28, 1, 2, 36]. Other recent approaches are studying cities in the developed world, and use online 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 [30]. 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 research examining urban impressions in developed cities using online crowdsourcing, judgments have been elicited using images obtained from Google Street View (GSV) [30, 25, 17]. 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. In a recent study, 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 [28]. Second, GSV images fail to capture the temporal aspects of a city: only static 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 [23, 20]. In contrast, mobile crowdsourcing represents an opportunistic, just-in-time way of documenting urban changes over time.

In this paper, we present a study on urban perception on three cities in central Mexico, which integrates: (1) mobile crowdsourcing involving a local youth community to collect first-person perspective images depicting urban issues that are defined by the community itself, and (2) online crowdsourcing using Mechanical Turk

ACM DEV 2015, December 01–02, 2015, London, United Kingdom .

ACM ACM 978-1-4503-3490-7/15/11 ... $15.00.
DOI: http://dx.doi.org/10.1145/2830629.2830638.

(MTurk), where US crowd-workers contribute their impressions on photos of the urban environment along twelve physical and psychological dimensions of a place. Our contributions are:

1. A mobile crowdsourcing 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 distinct geography, economic activity, and population (Section 3). The data set contains outdoor scenes and views of each city's built environment, including touristic, historical, and business sites, residential neighbourhoods, 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) (Section 4). 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 less accessibility than Leon. Overall, Guanajuato has the highest mean scores amongst all three cities on most positive labels and lowest scores on negative labels (Section 5.4). 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 crowdsourcing with respect to GSV-type surveying, we present a study about how the perceptions of urban environments vary across different times of the day using a small image sample and 12,000 individual judgments. The results show that places at evenings are perceived as less *happy*, *pleasant* and *preserved*, in comparison with mornings. We do not observe statistically significant differences in perception of *dangerous* between mornings and evenings for the studied places (Section 5.5).

Overall, our current work engages youth communities as actors of social change and contributes towards understanding socio-urban problems in cities through the use of collective action, participatory sensing, and crowdsourcing technologies.

2. RELATED WORK

2.1 Systems for reporting urban issues

There are various existing systems that allow citizens to report urban issues. One in the developed world is FixMyStreet (FMS) [3], launched in the UK in 2007 [18], 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 [4], 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 concludes that only 11% of them contain images, but also that image uploading is a significant indicator of the actual response of the authorities to the reports and the commitment of reporters to keep contributing [34]. 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 [10], 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 [1], and various systems built around Ushahidi [5]. In Latin America, other examples include the work done to map informal settlements in Buenos Aires, Argentina [2], and in Rio de Janeiro, Brazil [12]. Another mapping effort is led in India by Humara Bachpan [36], an organization that conducts civic campaigns centred on "child clubs" to create maps of marginalized neighborhoods in India slums. 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.

Finally, 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 [22]. This is an attractive alternative, but it is limited to people who agree to join these services and accept corporate-driven terms of use.

2.2 Crowdsourced urban perception

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 [31, 23] and observation of the built environment using either actual or simulated images [19]. 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, in conjunction 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 [15, 32] and outdoor environments [30, 28, 25]. For outdoor environments, gathering perceptions typically involve the use of Google Street View (GSV) [30, 25, 9]; while GSV is widely available in the developed world, it is not so for the developing world [28]. In [30], 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) [25]. Our current study builds upon our previous work [28], where we carried out a crowdsourcing study to collect perceptions of six dimensions of urban awareness (dangerous, dirty, preserved, etc.) by local inhabitants of one Mexican city. Compared to [28], we collect and study data that is ten times larger and comes from three cities, define and study a larger number of urban constructs, conduct cross-city analyses, and analzye temporal effects as discussed next.

2.3 Temporal effects on urban perception

The work in the previous subsection uses 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 [23, 20, 16]. 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 [23]. 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 [20]. 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 [16]. We study not only the role of night-time or dark scenes on perceived safety, but also other psychological dimensions e.g., quietness, happy, pleasant, etc.

3. DATA COLLECTION FRAMEWORK

In this section we describe our data collection framework, including the criteria to select the studied cities in Mexico, the definition and selection of urban awareness dimensions, and the mobile crowdsourcing design to collect two image corpora.

3.1 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, which aims at addressing specific urban issues by young volunteers through the use of collective action, participatory sensing and mobile crowdsourcing. The project 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" [11].

Our research project 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 understand the urban environment of their city.

3.2 Selection of Cities

To collect images from 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 [26, 8].

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 neighbourhoods.

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 1 shows a sample of images from each city.

3.3 Definition and Selection of Dimensions

In order to select labels to characterize urban awareness for outdoor environments, we base our methodology on prior work [28, 30, 25]. 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 the paper, we will use the umbrella term "urban awareness" to refer to these labels. Images served as stimuli to rate perceptions for 12 urban awareness labels, along a seven-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (7), as typically done in psychology and urban planning research [15, 20].

We have 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 faces various problems including crime, prevalence of alcohol and drugs in some streets and alleys, 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 is essential to study and understand the role these perceptions play in these cities.

3.4 Mobile Crowdsourcing Design

The images used in this study were collected as part of an Urban Data Challenge (UDC). The UDC was co-designed with the aid 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 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 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.

Figure 1: Random selection of images from the *city-image* corpus. For privacy reasons, images showing faces have been pixelated.

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-referenced 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 paper.

3.5 Image Corpus

City Image Corpus: As stated before, as a result of the UDC, we collected an image corpus consisting of 7,000 geo-tagged images. For our current analysis, we focus 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 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 build environments. Figure 1 shows a sample of images from the 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, along the selected dimensions. 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,

(a) Morning (b) Evening

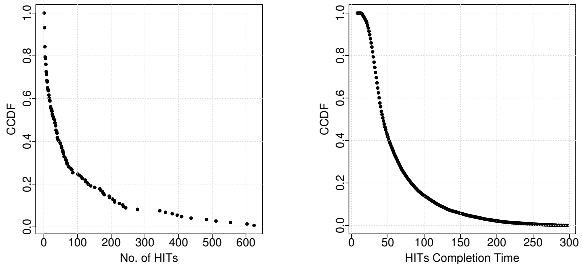
Figure 2: Random selection of images from the *evening-image* corpus, showing an image taken in a) Morning, and b) Evening. For privacy reasons, images showing faces have been pixelated.

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 2 shows two images of an urban place taken during morning and evening respectively.

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 (Human Intelligence Tasks). 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 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 to reduce potential bias and stereotyping associated to the city 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 judgements. 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,



(a) (b)

Figure 3: Worker Participation. Plot showing the complementary cumulative distribution function (CCDF) of a) Number of HITs per worker, b) HITs completion times.

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.

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 3a). We observe a long-tailed distribution in HIT completion times (mean: 59 secs, median: 43 secs, max: 297 secs), as shown in Figure 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.

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 [27]. 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 travelling 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 travelled to Mexico, which is not surprising given that the pool of crowd-workers is based in the US.

5. RESULTS AND DISCUSSION

5.1 Annotations Quality

We begin our analysis by assessing the reliability of annotations. We measure the inter-rater consensus by computing intraclass correlation (ICC) among ratings given by the worker pool [33]. Our annotation procedure requires every place to be judged by k annotators randomly selected from a larger population of K workers. $ICC(1,1)$ and $ICC(1,k)$ values, which respectively stand for single and average ICC measures, are computed for each label and city across all images.

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

Table 1 reports the $ICC(1,k)$ values for all cities for $k = 10$ (due to space constraints, we omit $ICC(1,1)$ values.) 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. This suggests that MTurk observers tend to agree on their perceptions of most dimensions. It is interesting to note that the combined label *quiet* 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.

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 [35]. 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 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 crowd-workers. 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 are study-

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 2: Means and standard deviations (in brackets) of annotation scores for each city and label.

ing and the intended goals of the crowdsourced collection, leaning towards documenting urban concerns.

From Table 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 Guanajuato and Leon and Silao all exceed 0.8, potentially suggesting differences in city perceptions. A systematic analysis to statistical testing of these differences are presented in Section 5.4.

5.3 Correlation and PCA Analysis

To understand basic statistical connections between urban awareness labels, we perform correlation analysis using the mean annotation scores for all labels. Figure 4 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 code 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 urban sites which are *quiet*. 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 can also observe significant negative correlations between dimensions in cluster one and cluster three.

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 [24]. In Figure 5, 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 [29].

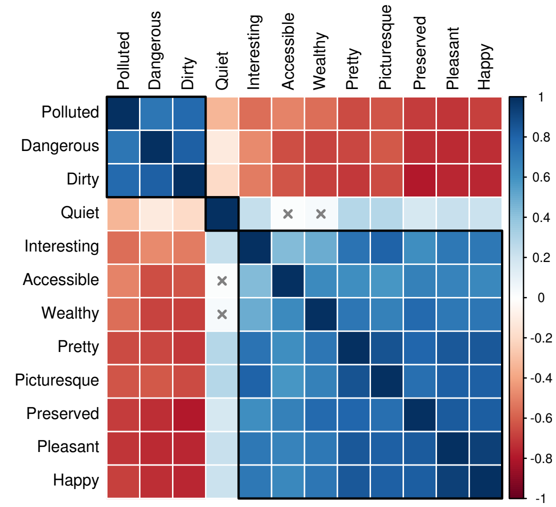


Figure 4: Plot showing the correlation matrix between dimensions. Matrix is color coded as per the palette shown in the right, with blue (resp. red) indicating positive (resp. negative) correlation coefficients. Black rectangular borders indicate the three distinct clusters found in the correlation matrix. Cells marked *X* are *not* statistically significant at $p < 0.01$.

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 [37]. We perform the HSD test to compute pairwise comparisons of mean values between cities for each label, which results in a total of 36 comparisons (3 city-wise pairs across 12 dimensions). Out of a total of 36 comparisons, we find 30 comparisons to be statistically significant at p-value < 0.01. In Table 3, we report the top-10 significant results of the Tukey's HSD test, where the differences in the observed means are 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 [14]. Additionally in Figure 6, we show the barplots comparing the mean annotation scores across all cities for two 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 are perceived as more *quiet*, *picturesque* and *interesting* compared to places in Leon and Silao (rows 1,2,3,4,5 in Table 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 except *dangerous*, as highlighted in Table 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 6a. 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*, but plots are omitted due to space constraints.
2. Silao is perceived to be more *polluted*, when compared to Guanajuato with differences in mean scores exceeding 0.6 (row 7 in Table 3). We believe these results are an effect of agricultural and industrial activity in Silao region, when compared to the historic and touristic nature of Guanajuato. When looking at the

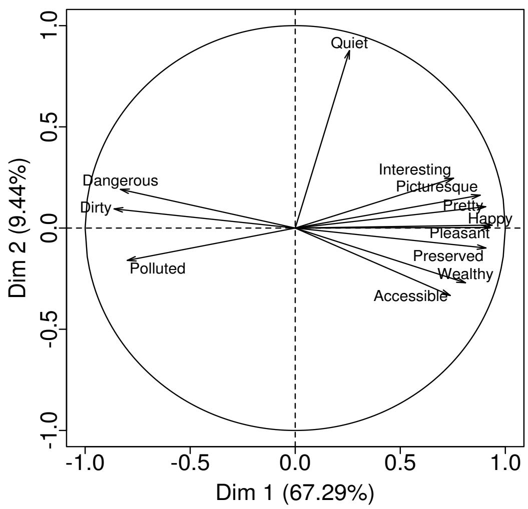


Figure 5: Plot showing the first two principal components on aggregated annotation scores on 1,200 images across all cities.

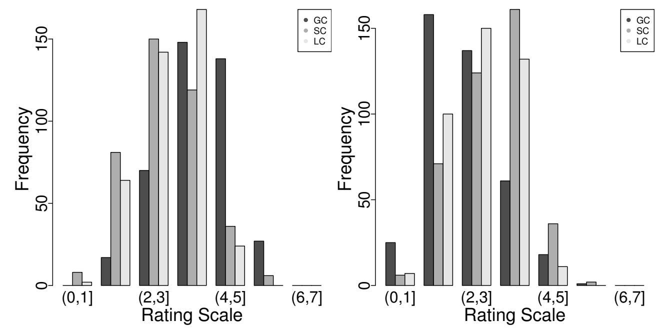
Label	City Pair	Mean Difference ± CI
Quiet	SC-GC	-0.98 ± 0.19
Interesting	SC-GC	-0.88 ± 0.18
Quiet	LC-GC	-0.84 ± 0.19
Picturesque	SC-GC	-0.82 ± 0.21
Interesting	LC-GC	-0.75 ± 0.18
Pleasant	SC-GC	-0.65 ± 0.19
Polluted	SC-GC	$+0.63 \pm 0.18$
Happy	SC-GC	-0.62 ± 0.19
Accessible	SC-LC	-0.61 ± 0.19
Pretty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

Table 3: 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 $p < 0.01$.

barplot in Figure 6b, we observe a similar pattern as highlighted above, where the relative proportion of places which are rated on a higher scale for being polluted in Silao are significantly larger than those in Guanajuato.

3. Silao is perceived to have lower accessibility than Leon (row 9 in Table 3). 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 1.

To further 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 [21]. We perform 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) (b)

Figure 6: Barplots comparing the mean annotation scores across all cities for a) Quiet, and b) Polluted.

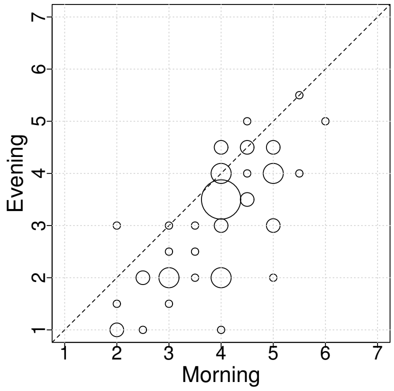
Label	Mean Scores		Mean Diff. ± CI	p-value
	Morning	Evening		
Accessible	4.40 (0.93)	4.38 (1.26)	-0.02 ± 0.58	0.928
Dangerous	2.86 (0.99)	3.38 (1.29)	$+0.52\pm0.60$	0.026
Dirty	2.89 (1.09)	3.33 (1.29)	$+0.44\pm0.63$	0.069
Happy	3.92 (0.98)	3.11 (1.18)	-0.81 ± 0.57	**0.0003**
Interesting	4.35 (0.74)	3.78 (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)	$+0.31\pm0.52$	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 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 $p < 0.01$.

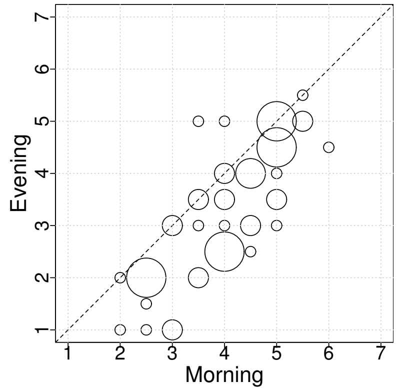
5.5 Evening Corpus Analysis

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 3.5, 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 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 5.2 for each image and time-slot. Using Table 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 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 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 3.5).

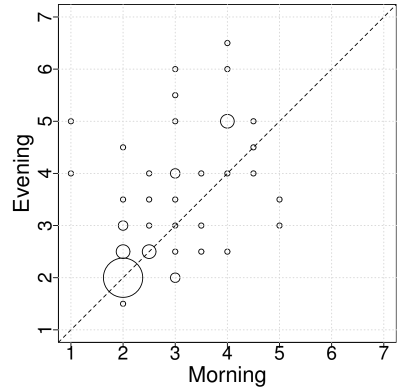
For the *evening-image* corpus, the mean annotation scores for all positive (resp. negative) labels are higher (resp. lower) for images taken in the morning, compared to the ones taken during the evening. In Table 4, we also report the results of of the Tukey's HSD statistics to test whether the mean scores differ across different times of the day for all labels. Based on these statistics we observe that:



(a) Happy



(b) Preserved



(c) Dangerous

Figure 7: 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.

1. Evenings are perceived as less *happy*, *pleasant* and *preserved*, in comparison with morning time (differences: -0.81, -0.77, -0.71 respectively). See also Figure 7a and 7b. In addition, in Figure 8, we show a pair of images of a place where the rating between morning and evening for *Happy* differ by three ordinal scales (rated 5 and 2 respectively).
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 [20, 16]. To put our finding in perspective, it is important to note that places are overall not perceived as *dangerous* in the *evening-image* corpus (combined mean: 3.12). See also Figure 7c.

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 7 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 7a and 7b), indicating that places in the morning are perceived on a higher scale compared to evenings. Furthermore, it is interesting to observe a more-mixed trend for the *dangerous* label (Figure 7c). These plots further validate our findings reported in the previous section.

(a) Morning

(b) Evening

Figure 8: Images where the perceptual ratings differ across times of the day for *Happy*. For privacy reasons, images showing faces have been pixelated.

6. LIMITATIONS AND FUTURE WORK

In this paper, we have presented a study to examine urban perceptions in three cities in central Mexico. Our study involves data collected by locals via mobile crowdsourcing, while assessments on collected data are undertaken by an external non-local population via online crowdsourcing. In other words, we have examined the perception of places as seen by "others" rather than "locals". In our study, external observers are all US-based crowd-workers.

To elicit impressions of urban perceptions, the observer population plays an important role, whether the population is external (as is the case in our study) or local (local community who is familiar with the environment). It can be argued that the collected assessments by external observers induce bias in the ratings and thus limit the generalizability of our findings. In our survey, most of the external observers reported not to have travelled to any developing country in the past (77% of our worker pool as reported in Section 4.2). We acknowledge this is one of the limitations of our work. However, we believe that the possibility of obtaining external perceptions of a city is valuable in and of iteself to quantitatively characterize the urban landscape, specially when the cities have large influx of visitors (i.e. travellers, tourists, students, business people, etc.) As part of future work, we plan to engage local communities to gather responses from them and compare their impressions with the ones obtained via online crowd annotators.

As a second limitation, we believe that the lack of ground-truth on perceptual ratings makes it difficult to further contextualize some of our findings. For most of the psychological dimensions (e.g., *happy*, *pleasant*, etc.) there exist no unique ground-truth, while for the physical dimensions (e.g. *dangerous*, *polluted*, etc.), there might be proxy measures. Previous studies have examined the relationship between the perceptions of *safety* and *class* and homicides rate in New York City [30]. However, due to the lack of publicly available data in the studied cities, such analysis was not feasible. Future work could include partnerships with the city or police to investigate whether this information could be available for research.

Due to the nature of our data collection, the spatial coverage of our approach can be seen as a potential third 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 our 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. Our data collection methodology was limited to areas that could be reached by our local community. However, in the context of development, it is more relevant for people to explore the urban area where they live and work in order to achieve 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 cities.

We conclude 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 crowdsourcing 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. With the objectives of informing urban planners and designing interventions in the local communities, we are teaming up with several NGOs and the local government to address some of the highlighted issues. Future development of both the academic results and community effort of the SenseCityVity project can be found here [6, 7].

7. ACKNOWLEDGEMENTS

This work has been supported by the EPFL-CODEV SenseCityVity and SNSF Youth@Night projects. We thank all the members of the SenseCityVity collective. We also thank all the student volunteers for their participation and the school authorities of CECYTE for the support to our project.

8. REFERENCES

[1] `http://mapkibera.org`.

[2] `https://www.caminosdelavilla.org/`.

[3] `https://www.fixmystreet.com`.

[4] `http://seeclickfix.com`.

[5] `http://www.ushahidi.com`.

[6] `www.idiap.ch/project/sensecityvity/`.

[7] `www.facebook.com/SenseCityVity-1420771584831936`.

[8] Top 5 cities in mexico: Readers' choice awards 2013. `http://cntrvlr.com/1f77Jbg`.

[9] M. D. Bader et al. Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions. *Health & place*, 2015.

[10] B. Baykurt. Redefining citizenship and civic engagement: political values embodied in fixmystreet. com. *Selected Papers of Internet Research*, 2012.

[11] M. Castells et al. *Reconceptualizing Development in the Global Information Age*. Oxford University Press, 2014.

[12] W. Connors. Google, microsoft expose brazil's favelas. `http://on.wsj.com/1V3X2qI`.

[13] G. Esteva et al. *Grassroots post-modernism: Remaking the soil of cultures*. Palgrave Macmillan, 1998.

[14] A. Gelman et al. The Statistical Crisis in Science. *American Scientist*, 2014.

[15] L. Graham et al. Can the ambiance of a place be determined by the user profiles of the people who visit it. In *Proc. ICWSM*, 2011.

[16] K. Hanyu. Visual properties and affective appraisals in residential areas after dark. *Journal of Environmental Psychology*, 1997.

[17] K. Hara et al. Combining crowdsourcing and google street view to identify street-level accessibility problems. In *Proc. CHI*, 2013.

[18] S. F. King et al. Fix my street or else: using the internet to voice local public service concerns. In *Proc. of the 1st international conference on theory and practice of electronic governance*, 2007.

[19] P. J. Lindal et al. Architectural variation, building height, and the restorative quality of urban residential streetscapes. *Journal of Environmental Psychology*, 2013.

[20] L. J. Loewen et al. Perceived safety from crime in the urban environment. *Journal of environmental psychology*, 1993.

[21] F. J. Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 1951.

[22] A. Monroy-Hernández et al. The new war correspondents: The rise of civic media curation in urban warfare. In *Proc. CSCW*, 2013.

[23] K. Painter. The influence of street lighting improvements on crime, fear and pedestrian street use, after dark. *Landscape and urban planning*, 1996.

[24] 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*, 1901.

[25] D. Quercia et al. Aesthetic capital: what makes london look beautiful, quiet, and happy? In *Proc. CSCW*, 2014.

[26] R. Reid. Top 8 places to (safely) visit in mexico now. `http://bit.ly/1hIjkce`.

[27] J. Ross et al. Who are the crowdworkers?: Shifting demographics in mechanical turk. In *CHI*, 2010.

[28] S. Ruiz-Correa et al. The young and the city: Crowdsourcing urban awareness in a developing country. In *Proc. of URB-IOT*, 2014.

[29] J. A. Russell et al. A description of the affective quality attributed to environments. *Journal of personality and social psychology*, 1980.

[30] P. Salesses et al. The collaborative image of the city: Mapping the inequality of urban perception. *PLoS ONE*, 2013.

[31] R. J. Sampson et al. Seeing disorder: Neighborhood stigma and the social construction of "broken windows". *Social psychology quarterly*, 2004.

[32] D. Santani et al. Loud and Trendy: Crowdsourcing Impressions of Social Ambiance in Popular Indoor Urban Places. In *Proc. ACM Multimedia*, 2015.

[33] P. E. Shrout et al. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 1979.

[34] F. M. Sjoberg et al. The effect of government responsiveness on future political participation. *Available at SSRN 2570898*, 2015.

[35] S. S. Stevens. On the theory of scales of measurement. *Science*, 1946.

[36] S. Sturgis. Kids in india are sparking urban planning changes by mapping slums. `http://bit.ly/1LtS9S4`.

[37] J. W. Tukey. The philosophy of multiple comparisons. *Statistical science*, 1991.