

The Young and the City: Crowdsourcing Urban Awareness in a Developing Country

Salvador Ruiz-Correa
Center for Mobile Life Studies
Centro Nacional de
Supercómputo-IPICYT-Idiap,
México
src@cmls.pw

Darshan Santani
Idiap Research Institute
École Polytechnique Fédérale
de Lausanne (EPFL),
Switzerland
dsantani@idiap.ch

Daniel Gatica-Perez
Idiap Research Institute
École Polytechnique Fédérale
de Lausanne (EPFL),
Switzerland
gatica@idiap.ch

ABSTRACT

We present a crowdsourcing study that investigates impressions of urban spaces by young inhabitants in a city in the developing world. Our goal is to obtain collective perceptions from the actual inhabitants of the city under study, and more specifically youth (16-18 year-olds) about issues like danger, accessibility, and dirtiness. We collect over 9000 judgments for 102 photos of outdoor urban spaces in a city in Central Mexico using standard scales in social sciences. We present reliability and response analyses and demonstrate how local youth can provide relevant urban insights in a crowdsourcing setting.

1. INTRODUCTION

The state of urban environments in developing countries is in part responsible for major challenges in attaining sustainable development and poverty reduction, as well as strong economic growth and social wellbeing. The use of methodologies that can lead to improved understanding of socio-urban problems and concerns in developing cities, neighborhoods and communities is consequently of great value. Here, we present a crowdsourcing study that investigates youth's perceptions of city places in a developing country (Guajuato, México). Categorization of urban spaces through psychological constructs (such as a dangerous, clean, conserved, nice) is the subject of our investigation, as opposed to the study of their intended functional purpose, as largely conducted in ubiquitous computing and computer vision research. How young people categorize their urban environment from a perceptual point of view is important, as this is closely related to the degree of their urban awareness. Urban awareness, which includes the ability to identify significant urban problems resulting in citizen concerns and action, is an essential ingredient for a community to develop holistically. All these issues are specially important in developing countries as they have booming youth populations, in many cases well acquainted with the use and impact of mobile and

social technologies.

The contributions of our work include: (1) a new image data set of urban scenes of a city in a developing country, which reflects the reality of cities who are insufficiently covered by state-of-art mapping efforts like Google Street View; (2) a crowdsourcing experiment, in which local youth recruited from a working-class high-school provided judgements of six urban dimensions (dangerous, dirty, nice, conserved, passable, interesting) based on the image data; (3) a total of 9027 collected judgements for each of the urban dimensions; and (4) detailed analyses with regard to descriptive statistics, reliability, and annotator judgments. We first found that the obtained annotations were reliable according to inter-class correlation (ICC) analysis. Furthermore, annotator ability and the difficulty to judge urban places were estimated through a latent variable model commonly used in psychometric studies. These analyses suggested that the participating young people are able to identify persistent urban problems affecting their city.

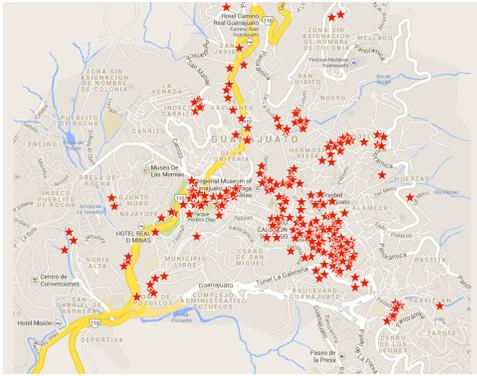
2. RELATED WORK

A few methods have been recently used to assess human perception of urban spaces. This includes works in psychology, ubiquitous computing, and social media. The emphasis of this emerging body of work is focused on psychological urban constructs [2] rather than on understanding the functional category of spaces – a thread of work that has been studied in ubiquitous computing or computer vision as place/scene categorization tasks, where spaces are first manually labeled (as home, work, office, etc.) and then recognized automatically using a variety of sensor data [1].

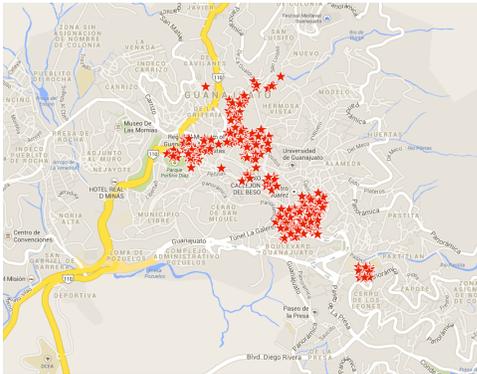
The perception of *indoor* spaces from images has been studied in social psychology [3]. More recently, the work in [4] defined a number of indoor ambiance categories and study the level of agreement reached by visitors to the places themselves. The study was done on a US city. This line of work was recently extended in [10], using images shared in Foursquare where users captured snapshots of the venues. The study was conducted on places from six large cities, but only one belonged to a developing country. More closely related to our work is that by [9] and [7], which selected images of *outdoor* urban scenes using Google StreetView and later crowdsourced the collection of urban impressions using the images as stimuli. The work by [9] used images from four developed cities (in the US and Austria) and collected labels for three urban dimensions, namely uniqueness, safety, and social class. The labels were produced via crowdsourcing

Urb-IoT'14, October 27-28 2014, Rome Italy.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission from Permissions@acm.org.
Copyright 2014 ACM 978-1-4503-2966-08/14/10 ...\$15.00
<http://dx.doi.org/10.1145/2666681.2666695>.



(a)



(b)

Figure 1: Locations where geo-tagged images were collected. Our data is composed of two sets of three hundred images each. (a) First set: outdoor images captured at touristic spots, key historical sites, traditional neighborhoods, main squares, plazas, and thoroughfares. (b) Second set: images of historical city alleys.

by attributing relative judgments to pairs of images, i.e., a pairwise ranking approach. The work by [7] presented a similar approach, using also Google StreetView images and annotations produced by pair-wise image ranking, but diverging with respect to the dimensions of interest (beauty, quietness, happiness) and the studied city.

Our work differs significantly in four key aspects. First and foremost, our interest is in studying cities in the developing world. Due to this, the constructs of highest relevance might differ from those in developed cities. For example, danger might not be the first idea that comes to mind when one thinks of Zurich, in the same way quietness might not be the first thought about Mexico City. Second, images that accurately capture representative areas of developing cities are often not available on Google StreetView for many areas, despite the wide penetration of the company’s approach to mapping. This is due both to physical conditions of certain neighborhoods in developing cities as well as socio-economic ones: in certain rough neighborhoods, Google data capture vehicles might get in but not get out. Third, our goal is in obtaining collective perceptions from the actual inhabitants of the cities under study, and more specifically youth, rather than from external observers, as done in both [9] and [7]. Finally, we collect judgments from outdoor urban spaces using

Table 1: Image data set used for our study (manually collected) compared to content available on Google Street View.

Google Street View Status	Number of images (%)
Available	42.1
Unavailable	46.1
Erroneous	11.8

standard scales in social psychology for individual images as opposed to collecting relative rankings between image pairs. Our approach is therefore amenable for standard reliability and response analyses widely used in the social sciences.

3. DATA AND METHODS

We begin this section by describing our methods for selecting urban sites and images. We then explain our approach to crowdsourcing urban perception. Finally, we present our approach to model annotators ability in rating urban spaces.

3.1 Selection of urban sites and images

Guanajuato City (pop. 170,000) is the capital city of Guanajuato state in Central Mexico. The city is known for its art scene and tourism industry, yet it also has a variety of socio-urban problems. The city is located in a little valley with narrow and winding streets. Most are pedestrian alleys, some being long sets of stairs on the mountain sides leading to houses that do not have car access. Many main streets crossing the city are partially underground. The historic, colonial center has numerous small plazas and mansions, temples, theatres, and government constructions built using traditional pink or green sandstone. Pedestrian alleys are one of the key urban features of the city, which over the years, determined a unique lifestyle. One could argue that the city’s quality of life and tourism heavily depend on the preservation of the social fabric and traditional urban spaces that evolved around city alleys. Overall, we believe this city reflects a common situation in cities in Latin America, where a combination of historical downtown areas, urban sprawls, increasing populations, and large socio-economic disparities all co-exist.

The first author and a group of volunteers collected two groups of three hundred geo-tagged images each, using a variety of locations in Guanajuato City (Figure 1). All images were taken between 9 am and 2 pm, when most people are at work or school and the traffic load is relatively low. Different criteria were used to select locations for each group. The first consists of outdoor images captured at touristic spots, key historical sites, traditional neighborhoods, main squares, and thoroughfares. The second group consists of pedestrian images of downtown alleys that not only have historical value, but are actively used by people in social and religious activities such as traditional celebrations and religious pilgrimages. In order to accommodate the time constraints of volunteer annotators (high school students) participating in our crowdsourcing experiment, a smaller set of a set of fifty-one images (*Image set 1*) was obtained by random sampling the first group of three hundred images described previously. A second set of fifty-one images (*Image set 2*) was similarly obtained from the second group (city alleys).

We remark that images for some of the sites used in our study can be obtained through Google Street View (as done in [9],[7].) However, we choose to manually collect the im-



El lugar que se muestra en la imagen es (o está):

CONSERVADO :	<input type="radio"/> totalmente en desacuerdo	<input type="radio"/> en desacuerdo	<input type="radio"/> un poco en desacuerdo	<input type="radio"/> no estoy seguro	<input type="radio"/> un poco de acuerdo	<input type="radio"/> de acuerdo	<input type="radio"/> totalmente de acuerdo
SUCIO :	<input type="radio"/> totalmente en desacuerdo	<input type="radio"/> en desacuerdo	<input type="radio"/> un poco en desacuerdo	<input type="radio"/> no estoy seguro	<input type="radio"/> un poco de acuerdo	<input type="radio"/> de acuerdo	<input type="radio"/> totalmente de acuerdo
INTERESANTE :	<input type="radio"/> totalmente en desacuerdo	<input type="radio"/> en desacuerdo	<input type="radio"/> un poco en desacuerdo	<input type="radio"/> no estoy seguro	<input type="radio"/> un poco de acuerdo	<input type="radio"/> de acuerdo	<input type="radio"/> totalmente de acuerdo
TRANSITABLE :	<input type="radio"/> totalmente en desacuerdo	<input type="radio"/> en desacuerdo	<input type="radio"/> un poco en desacuerdo	<input type="radio"/> no estoy seguro	<input type="radio"/> un poco de acuerdo	<input type="radio"/> de acuerdo	<input type="radio"/> totalmente de acuerdo
PELIGROSO :	<input type="radio"/> totalmente en desacuerdo	<input type="radio"/> en desacuerdo	<input type="radio"/> un poco en desacuerdo	<input type="radio"/> no estoy seguro	<input type="radio"/> un poco de acuerdo	<input type="radio"/> de acuerdo	<input type="radio"/> totalmente de acuerdo
BONITO :	<input type="radio"/> totalmente en desacuerdo	<input type="radio"/> en desacuerdo	<input type="radio"/> un poco en desacuerdo	<input type="radio"/> no estoy seguro	<input type="radio"/> un poco de acuerdo	<input type="radio"/> de acuerdo	<input type="radio"/> totalmente de acuerdo

Para [Continuar](#) completa toda la forma por favor.

Figure 2: Website used to collect urban impressions. Annotators were asked to view an image and then select a score in a 7-point Likert scale that reflects their personal perception based on what they saw, ranging from strongly disagree (‘totalmente en desacuerdo’) to strongly agree (‘totalmente en acuerdo’). Six descriptors were used to qualify urban spaces: dangerous (‘peligroso’), dirty (‘sucio’), nice (‘bonito’), conserved (‘conserved’), passable (‘transitable’) and interesting (‘interesante’).

ages for two reasons that we believe are common to developing cities. First, 46% of the collected places in our study are not available on Google. This probably due to street inaccessibility or security issues in some areas of Guanajuato (Table 1). Furthermore, 11.8% of the images of sites found in Google Maps do not correspond to those of the actual sites. Namely, we observed that a site found in the Google map that is not associated with a set of actual images, results in a street view that either shows black images or images from other (sometimes distant) places. The second reason is also key: our photos emphasize aspects of the urban environment that are not clearly shown in images captured for Google Maps due the sinuous nature and roughness of some streets in Guanajuato. In addition, the device used to collect Google Street View images is mounted on a car at a height that provides a large field of view. One disadvantage of this setup is that the resulting views cannot capture details of the urban environment that can be taken with a cellphone camera located at pedestrian level. If the street is narrow and sinuous, as is the case in many Guanajuato streets, some of the images taken with by Google look distorted, and some of them have artifacts produced by sun glare, or occlusions due to tree branches, loading trucks or buses.

3.2 Crowdsourcing data

Perception data was collected using a website designed for the study. We conducted two experiments. In the first one, each image from Image Set 1 (IS1) was shown to each annotator and asked to provide a 7-point Likert rating ranging from strongly disagree (1) to strongly agree (7) for six descriptors of the shown place (dirty, dangerous, nice, conserved, passable and interesting). The second experiment was similar but considered images from Image set 2 (IS2) (Figure 2). Experiment 1 was conducted with a population of 101 annotators. Experiment 2 was conducted with a population of 76 annotators, some of which (about 2%) also participated in Experiment 1. In each annotation task, students were asked to view an image and then rate their per-

sonal impressions based on what they saw. We ensured that students viewed high-resolution images to form their impressions. Students were not given any information about the urban place being displayed to reduce potential stereotyping associated with the different areas of the city. Although we expected that as local inhabitants the annotators might have been to some of the places before, only about 2 % of them were able to identify one or two city sites from the images used in the experiments. Each student worked independently with a high resolution laptop computer during the execution of the task, which lasted 1.2 hours on average. Once the data collection was completed, students participated in a workshop about urban awareness in developing countries.

Participating annotators were youngsters of age 16 - 18 years old chosen at random from a population of 600 hundred students from a public high-school in Guanajuato, the Colegio de Estudios Científicos y Tecnológicos del Estado de Guanajuato (CECYTEG), campus Guanajuato. This technical college was created to provide high-quality education on science, technology, and humanities to low-income youth who live in Guanajuato City and surrounding suburbs. Students engaged in the crowdsourcing experience were altruistically motivated and eager to contribute with their knowledge and experiences to better understand the urban environment of their city. Support for the experiment was achieved through a multi-institutional partnership that included school authorities, teachers, and parents.

3.3 Modeling annotator impressions

Consider a crowdsourcing study in which annotator s expresses a judgment about an image showing an urban space by answering a question q regarding a description of the image. For example, the image may show a pedestrian alley and the annotator is asked to tell if the alley looks dangerous or not. If the alley is considered dangerous (on the grounds of crime statistics for instance), the annotator’s answer may coincide $x_{qs} = 1$ or not $x_{qs} = 0$ with the fact that the place is dangerous. For a set of S annotators and Q questions

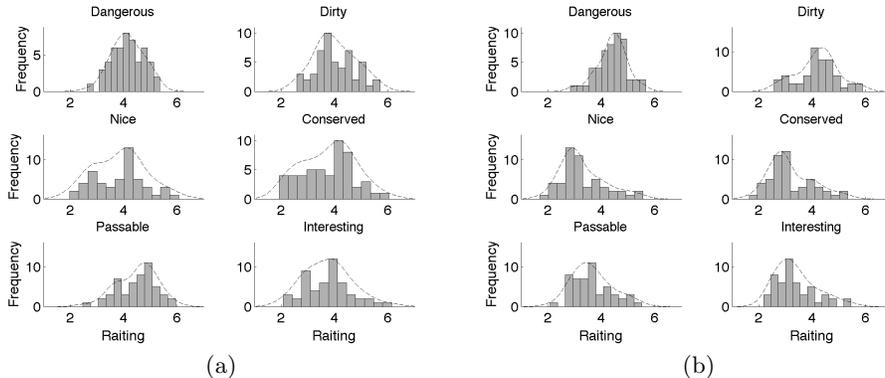


Figure 3: Annotator’s mean ratings histograms for urban descriptors associated with (a) IS1 and (b) IS2.

(related to a set of corresponding images), the performance of all annotators in subjectively judging the actual qualities of an urban space (v. gr. dangerous, dirty, accessible, etc.) is given in the $\mathcal{Q} \times \mathcal{S}$ binary matrix \mathbf{X} . In an urban context this ability could be related to the annotator sensibility or awareness of his urban environment. Here, we assume that the actual qualities of a place have been previously measured by quantitative means and have not been disclosed to the annotators. Based on \mathbf{X} alone we wish to evaluate the ability of each annotator in judging the qualities of a place from an image.

One approach is to define the ability as the fraction of the questions for which the annotator’s answers reflect the actual site qualities. A more subtle analysis is to accept that some qualities are more difficult to judge than others so that an annotator who answered difficult questions should be awarded more highly than an annotator who answered the same number of easy questions. A priori, however, we do not know which are difficult questions (i.e. which judgments are more difficult to make) and this needs to be estimated based on \mathbf{X} . To account for inherent differences in question difficulty, we model the probability that an annotator’s answer to question q reflects the actual site qualities based on the annotator’s *latent ability* α_s and the *latent difficulty* of the question δ_q . A generative model of the response is

$$p(x_{qs} = 1 | \alpha_s, \delta_q) = \sigma(\alpha_s - \delta_q), \quad (1)$$

for $s = 1, \dots, |\mathcal{S}|$ and $q = 1, \dots, |\mathcal{Q}|$, where $\sigma(x) = 1/(1 + e^{-x})$. Under this model, the higher the ability is above the latent difficulty of the question, the more likely is that the annotator’s judgment will correspond to the actual qualities of the urban site. Model parameters can be found through maximum likelihood for observed data \mathbf{X} or by extending the model through a Bayesian technique by setting prior models on the α_s and δ_q ’s. Here we use off-the-shelf tools developed in R language (Test Analysis Modules or TAM package) to analyze our perception data set.

The model in (1), which we borrowed from Rasch [8], has been extensively used in psychometric studies. It was developed to determine the probability of a specified response (e.g. correct/incorrect answer) as a function of person and item parameters. In our study, however, answers to questions cannot be considered “correct” or “incorrect” as an annotator response depends upon an internal representation of

Table 2: Descriptive statistics for annotation ratings: *mode*, *mean*, and *standard deviation* computed over a population of k annotators.

Description	IS1 ($k = 101$)	IS2 ($k = 76$)
Dangerous	4, 4.15 ± 0.59	4, 4.41 ± 0.56
Dirty	5, 4.02 ± 0.77	5, 4.20 ± 0.78
Nice	5, 3.84 ± 0.96	2, 3.29 ± 0.84
Conserved	5, 3.77 ± 0.98	2, 3.20 ± 0.82
Passable	6, 4.49 ± 0.73	4, 3.65 ± 0.74
Interesting	5, 3.74 ± 0.86	2, 3.38 ± 0.76

a quality that is reflected by a purely subjective judgment.

In order to use the model in (1) to analyze our data recorded using 7-point Likert scales, annotator responses were recoded using a binary scale. We use a simple example to illustrate the recoding process. Consider an image of a street alley that was qualified as dangerous by an annotator with a score above 4. If the alley is dangerous in actuality, the annotator score is set to $x_{qs} = 1$, otherwise it is set to $x_{qs} = 0$. Similarly if an image of a street alley was qualified as dangerous with a score above 4, and the alley is not actually dangerous, the annotator score is set to $x_{qs} = 0$, otherwise it is set to $x_{qs} = 1$.

4. RESULTS AND DISCUSSION

We start our analysis by examining the descriptive statistics of crowdsourced annotations for each of the urban descriptors and image sets. We then turn our focus on assessing annotation quality by means of intraclass correlation (ICC) techniques, and analyze annotators ability and questions difficulty through the model described in Section 3.3. Finally we present qualitative observations from some of the participating students.

4.1 Descriptive statistics

Histograms of mean annotator ratings show a noticeable degree of skewness for IS2 (Figure 3). Annotators tend to agree that the urban places depicted in pedestrian alleys are dangerous and dirty. Similarly, they have a tendency to disagree that the places are nice, conserved, passable and interesting. Therefore, our youngster population perceive a degree of urban disarray around these historical places (Table 2). The story is different for IS1. Histograms of mean

Table 3: Inter-Rater Reliability using two-way random effects, consistency, average-measures $ICC(C, k)$, agreement, average-measures $ICC(A, k)$ and corresponding confidence Intervals with $p < 0.001$.

Description	IS1	IS2
	ICC(C,101)	ICC(C,76)
Dangerous	0.94 (0.91, 0.96)	0.93 (0.89, 0.95)
Dirty	0.96 (0.94, 0.97)	0.96 (0.94, 0.95)
Nice	0.98 (0.97, 0.98)	0.97 (0.96, 0.98)
Conserved	0.98 (0.97,0.99)	0.97(0.96, 0.98)
Passable	0.96 (0.94, 0.97)	0.96 (0.94, 0.97)
Interesting	0.97 (0.96, 0.98)	0.96 (0.95, 0.98)
	ICC(A,101)	ICC(A,76)
Dangerous	0.92 (0.89,0.95)	0.91 (0.86,0.94)
Dirty	0.95 (0.94,0.97)	0.95 (0.93,0.97)
Nice	0.97 (0.96,0.98)	0.96 (0.94,0.97)
Conserved	0.97 (0.96,0.98)	0.69 (0.94,0.97)
Passable	0.95 (0.92,0.96)	0.94 (0.92,0.96)
Interesting	0.95 (0.94,0.97)	0.95 (0.93,0.97)

ratings suggest that annotators have a hard time determining if the city places are dangerous (Figure 3). This result is surprising and requires further investigation, which could be guided by the findings presented in [6]. Annotators are inclined to agree that IS2 images show places that are passable, but are not that dirty. Sites do not look nor nice nor conserved for some annotators of this data set but some others are not sure.

4.2 Annotations quality

Inter rater reliability (IRR) was assessed using a two-way mixed, consistency, average-measures $ICC(C, k)$ [5] to assess the degree that coders provided consistency in their descriptor ratings across subjects. The resulting $ICC(C, k)$ was in the excellent range shown in Table 3, indicating that annotators had a high degree of agreement and that urban descriptors tested are rated similarly across coders. Computation of agreement, average-measures, $ICC(A, k)$ [5] lead to similar results (Table 3). The high ICC values suggests that the independent annotators introduced a minimal amount of measurement error, and therefore statistical power for subsequent analyses is not substantially reduced. Image ratings were therefore deemed to be suitable for use in the hypothesis tests of the study.

Note that the ICC values described above were obtained using a very large number of annotators per data sample (101 and 76 for IS1 and IS2, respectively.) We recomputed IRR measures using a random sample of $k = 10$ raters in order to contrast results obtained in related crowdsourcing studies, where obtaining ratings from a significant number of annotators is impractical. We found that although ICC values are lower, the resulting consistency and agreement are acceptable. For IS1, ICCs are between 0.22 and 0.8 for the six dimensions of interest. Direct comparison with the works in [7] and [9] is not possible as those papers did not report these reliability measures due to the way impressions were collected (pairwise ranks, not absolute scores.) In contrast, these values are comparable to other works [4] [10], which have reported ICCs for social perception constructs with similar numbers of observers ($k = 10$) for indoor places.

4.3 Hypothesis testing: perception of danger

As pointed out in Section 4.1, annotators had difficulty

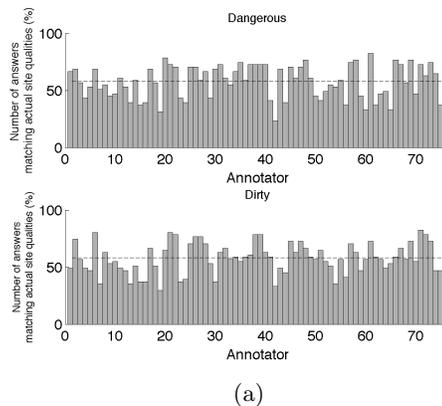


Figure 4: Number of annotator answers matching actual quality site (%) for (a) dangerous and (b) dirty descriptors of urban places depicted in IS2. Dashed lines represent mean values corresponding to the bar plots in (a) $\mu = 59.72$ and (b) $\mu = 58.06$. Means are not statistically different ($p = 0.95$).

rating dangerous sites shown in IS1 images. We computed the mean ratings from images depicting touristic places (T) separately from those not representing such kind of places (NT). We tested the hypothesis that the mean of the distributions of T vs. NT were statistically different ($p < 0.001$). A t-test showed that non-touristic places have a significant larger score of perception ($\mu_{NT} = 4.49$) of danger compared to touristic places ($\mu_T = 3.69$).

4.4 Assessing annotator ability and image difficulty

We proceed to examine the ability vs. difficulty results by focusing on dangerous and dirty descriptor ratings for IS2, for which ground truth assessments we obtained from city’s crime statistics for the second semester of 2013 (which were formally requested to the City’s Municipality by the first author) and careful in-situ inspection. This analysis was not conducted for the remaining descriptors (passable, nice, conserved an interesting), as ground truth data is not available yet. The number of annotator answers matching the actual quality site for each of the descriptors are shown in Figure 4. Dotted lines in the figure depict mean values corresponding to the bar plots. These values are not statistically different suggesting that both categories are equally difficult to rate on the basis of percent matching answers. On the other hand the results obtained with the latent variables model of Section 3.3 are summarized in Figure 5. We note that the origin (i.e., zero) of the ability scale is set as the mean ability. Histograms of ability and difficulty distributions for both, dangerous and dirty descriptors are shown in the bottom row of Figure 5. The standard deviation and (2.5 25 50 97.5) percentiles of annotator’s ability related to the dangerous descriptor are 0.67 and (-1.16 , -0.57, 0.04, 0.59, 0.98), respectively. The standard deviation and same percentiles of annotator’s ability related to the dirty descriptor are 0.65 and (-1.13, -0.43, -0.07, 0.49 , 1.16), respectively. These results suggest that the annotators population is slightly less able to quality dangerous places than dirty ones. Difficulty histograms for qualifying images in terms of dangerous and

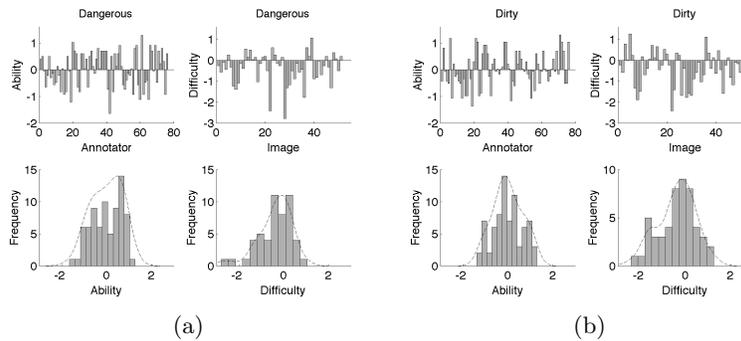


Figure 5: Ability and difficulty results obtained with the latent variable model.

dirty descriptors are shown in the same image. The corresponding means and standard deviations are -0.38 ± 0.75 and -0.40 ± 0.82 , respectively. Mean values are not statistically different.

4.5 Qualitative observations

Some of the students participating in the crowdsourcing experiment were asked their general impressions after they completed the annotation task. Many students agreed that rating images of urban sites not only was fun, but also gave them a wider perspective about the need for caring more deeply for the city where they live. One student told us that Guanajuato city has many faces, some of which she does not like: “I have learn to see Guanajuato from different side. In the past I ofte saw Guanajuato as a touristic city that you visit to make pictures. However, I see now that Guanajuato is not the way people paint it. It has a side that is not as beautiful. Guanajuato also has problems and forgotten places; problems can be solved only if people get involved.”

Some students became aware to urban problems that affect their daily life. One student said: “I have learn to identify certain aspects of the city that I did not noticed in the past. Some city alleys are quite dirty, and people do not pay attention to this. Perhaps one gets used to this kind of problem after some time. I think this happened to me. Every day I walk to school, I have to deal with a road that is full of trash, because people are lazy and do not care.”

5. CONCLUSIONS

In this paper, we studied youth impression of urban places in a developing country on the grounds of perceptual categorization assessments, which are closely related to the urban awareness of the population under study. We conducted a crowdsourcing study to gather over 9000 impressions of urban sites involving over 150 high-school students, using standard scales in social psychology for 102 images, which allowed us to use standard inter-rater reliability and response analyses. We found that the crowdsourced annotations have good levels of reliability for all dimensions, which show the feasibility of collecting this type of perceptions from local populations. We successfully fitted a latent variable model that allowed us to characterize annotator’s ability to make judgment and the difficulty in rating images depicting urban sites.

Acknowledgments. This work was funded by EPFL Cooperation and Development Center (CODEV) *Seed Money Program*

through the SenseCityVity project. We also acknowledge the enthusiastic participation of CECYTEG’s students, teachers and school authorities in SenseCityVity’s Urban Data Challenge.

6. REFERENCES

- [1] T. Do and D. Gatica-Perez. The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Trans. on Mobile Computing, Vol. 13, No. 3, Mar. 2014*.
- [2] M. C. Florida, R. and P. J. Rentfrow. The happiness of cities. *Regional Studies*, 47:613–627, 2013.
- [3] K. S. J. Gosling, D and T. Mannarelli. Room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82(3), 2002.
- [4] L. T. Graham and S. D. Gosling. Can the ambiance of a place be determined by the user profiles of the people who visit it? In *Proc. ICWSM*, 2011.
- [5] K. O. McGraw and S. P. Wong. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(30), 1996.
- [6] J. R. R. Naik N, Philipoom and C. Hidalgo. Streetscore: Predicting the perceived safety of one million streetscapes. *CVPR Workshop on Web-scale Vision and Social Media.*, 2014.
- [7] D. Quercia, N. K. O’Hare, and H. Cramer. Aesthetic capital: What makes london look beautiful, quiet, and happy? In *Proc. CSCW*, pages 945–955, 2014.
- [8] G. Rasch. *Probabilistic models for some intelligence and attainment tests*. The University of Chicago Press, 1980.
- [9] P. Salesses, K. Schechtner, and C. A. Hidalgo. The collaborative image of the city: Mapping the inequality of urban perception. *PLoS ONE*, 8(7):e68400, 07 2013.
- [10] D. Santani and D. Gatica-Perez. Loud and trendy: Crowdsourcing impressions of social ambiance in popular urban places. *arXiv*, (1403.1168), 2014.