Insiders and Outsiders: Comparing Urban Impressions between Population Groups

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

Salvador Ruiz-Correa
IPICYT
Mexico
salvador.ruiz@ipicyt.edu.mx

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

ABSTRACT

There is a growing interest in social and urban computing to employ crowdsourcing as means to gather impressions of urban perception for indoor and outdoor environments. Previous studies have established that reliable estimates of urban perception can be obtained using online crowdsourcing systems, but implicitly assumed that the judgments provided by the crowd are not dependent on the background knowledge of the observer. In this paper, we investigate how the impressions of outdoor urban spaces judged by online crowd annotators, compare with the impressions elicited by the local inhabitants, along six physical and psychological labels. We focus our study in a developing city where understanding and characterization of these socio-urban perceptions is of societal importance. We found statistically significant differences between the two population groups. Locals perceived places to be more *dangerous* and *dirty*, when compared with online crowd workers; while online annotators judged places to be more *interesting* in comparison to locals. Our results highlight the importance of the degree of familiarity with urban spaces and background knowledge while rating urban perceptions, which is lacking in some of the existing work in urban computing.

CCS CONCEPTS

• **Information systems** → **Crowdsourcing**;

KEYWORDS

Urban Perception; Crowdsourcing; Mobile; Youth; Subjective Properties

ACM Reference format:
Darshan Santani, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2017. Insiders and Outsiders: Comparing Urban Impressions between Population Groups. In *Proceedings of ICMR '17, Bucharest, Romania, June 06–09, 2017 (ICMR'17),* 7 pages.
DOI: http://dx.doi.org/10.1145/3078971.3079022

ICMR'17, June 06–09, 2017, Bucharest, Romania

ACM ISBN 978-1-4503-4701-3/17/06. . . $15.00
DOI: http://dx.doi.org/10.1145/3078971.3079022

1 INTRODUCTION

Online crowdsourcing platforms are regularly being used to conduct behavioral studies in computer and social science. In the context of urban computing, recent research has established the feasibility of obtaining impressions of urban scenes for both indoor and outdoor environments using crowdsourcing [6, 7, 9, 11, 13, 16]. In [9], the authors conducted a study to measure the perception of outdoor urban scenes on safety, class and uniqueness, based on images in four cities – two each in the US and Austria. In a similar study on urban perception, judgments from over 3,000 individuals were collected to examine visual cues that could correlate outdoor places in London with three dimensions (beauty, quietness, and happyness) [6]. A recent work [16] explored the connections between visual attributes of the built environment and safety perception and found that familiarity with the environment is associated with the perception of safety.

In addition to using online crowdsourcing platforms, there are numerous studies in the domain of urban planning which have used qualitative methods including questionnaires and interviews to quantify urban perception. In this domain, researchers have examined the relationship between the visual characteristics of the built environment and perceptual attributes [3, 5, 10], including the effect of time on urban perception [2, 4]. In contrast with these works, our current work focuses on the observer attributes i.e., who is observing as opposed to what the observer sees. In this paper, we investigate how the impressions of outdoor urban spaces judged by online crowd annotators, compares with the impressions elicited by the local inhabitants, along six physical and psychological labels. In other words, we empirically test the observer differences from the point of view of urban perception of six subjective attributes of outdoor environment. Our current study is closely related and builds upon our earlier research, where we presented a mobile crowdsourcing methodology to collect perceptions of urban awareness labels (dangerous, dirty, preserved, etc.) by local inhabitants of Guanajuato city in Mexico [7].

Most of the existing studies on urban visual perception have focused on collecting impressions by "external" observers, however an issue which remains open in the literature is whether these perceptions match with the perceptions of locals who live there. By locals, we refer to citizens familiar with the outdoor places, streets and terrain under study; by context, we explicitly mean the knowledge people possess by being a "local". In this study, our objective is to compare the perceived impressions of outdoor urban scenes between

(a) **(b)** **(c)** **(d)**

Figure 1: Sample images from the manually collected image corpus.

two population groups: local inhabitants and non-local online crowd workers. We investigate whether the background knowledge of the observers has an impact on the generated impressions, i.e., between city inhabitants who have a context of the places and the characteristics of the built environment, and external observers who might not.

To capture the perceptions of urban spaces, we rely on a method where images are used to form impressions. We gather images which describe urban scenes and city's built environment in their natural setting, e.g., images showing different neighborhoods, alleys and streets, touristic and historical sites. In most of the recent studies, urban perceptions were elicited using images obtained from Google Street View (GSV) [6, 9]. Even though GSV provides a scalable and automated way to collect images, it suffers from two limitations. First, the image database of GSV is not exhaustive in spatial coverage, which is particularly evident in developing countries. In our earlier work, we found that 58% of our images were either unavailable or erroneous in the GSV database for Guanajuato city [7]. Second, due to the way Google collects the street views (via cameras mounted on top of a vehicle), GSV does not always contains images of narrow streets and winding alleys, which is often the case in this city. Consequently, we rely on a set of manually collected images in our study. We acknowledge that our current image collection methodology is not scalable, but we plan to investigate alternate means to obtain representative and diverse images in a scalable manner, as part of future work.

Our contributions are two-fold. First, we designed and conducted a crowdsourcing study to gather impressions of local residents and non-local online crowd-workers along six physical and psychological labels (*dangerous*, *dirty*, *interesting*, etc.), based on images of outdoor places (Section 3). Second, we performed statistical analysis to compare the perceptual impressions of online crowd workers and local residents (Section 4). We found statistically significant differences between the population groups. As key findings, locals perceived places to be more *dangerous* and *dirty*, when compared with online annotators; while online annotators judged places to be more *interesting* in comparison to locals. As a potential application, we could imagine embedding our findings into existing online photo-sharing platforms (such as Flickr or Instagram), to enhance the view of a place by integrating "insiders" knowledge for "outsiders", who otherwise might not have access to these kinds of resources e.g., an American tourist planning for a Mexican holiday.

In summary, our current work attempts to empirically test the observer differences from the point of view of urban perception of subjective attributes. Our work highlights the importance of context, background knowledge, and prior beliefs in urban perception studies.

2 IMAGE CORPUS

We ground our analysis on an image corpus collected as part of our previous study [7]. The image corpus was collected in Guanajuato, a mid-size Mexican city with a population of around 170,000 people. One of the paper's authors and about 10 volunteers, who were also residents of Guanajuato, visited different parts of the city and captured images of outdoor urban sites using their mobile phones. City areas covered included different neighborhoods, historical city alleys and streets, central plaza, and touristic and historical sites. Most of the images were collected during early mornings and weekends. Volunteers were asked to capture images of urban scenes in their natural setting and not necessarily capture beautified images or apply some form of digital filters, as usually the case with images found on social media platforms, like Flickr or Instagram. As a result of this data collection, we obtained a set of 99 images which we analyze in this study. Figure 1 shows a sample of images from the corpus. All the images were geo-tagged. A map showing the spatial coverage of the images is shown in our previous study [7]. We refer the readers to [8, 13] for a detailed description of the mobile crowdsourcing methodology deployed to gather images.

3 METHODS

3.1 Selection of Labels

In order to select labels to characterize urban awareness for outdoor environments, we base our methodology on our earlier work [7], where we proposed a rating instrument consisting of six labels to characterize urban awareness, which are: *accessible*, *dangerous*, *dirty*, *interesting*, *preserved*, and *pretty* (see Table 1). Throughout the paper, we will use the term "urban awareness" to refer to these labels. For both the population groups, images served as stimuli to rate perceptions for the selected six urban awareness labels, along a seven-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (7).

We have chosen this list of labels for two reasons. First, these labels encompass the physical and psychological constructs evoked while describing the characteristics of the studied build environment. Second, Guanajuato is a historical city and a UNESCO world heritage site, with a vibrant tourism industry but it also faces various socio-urban and civic problems including crime, street gangs, prevalence of alcoholism and drugs in streets and alleys, dirty streets with garbage and non-artistic graffiti and murals. In addition to affecting the prosperity and safety of citizens, these issues also hurt the city's image as a tourist destination. As a result, it is essential to study and understand the role these perceptions play in Guanajuato city.

3.2 Crowdsourcing Impressions

MTurkers: To gather impressions of online annotators, we designed a crowdsourcing study online on Amazon's Mechanical Turk (AMT). We chose US-based "Master" annotators with at least 95% approval rate for historical HITs (Human Intelligence Tasks). In each HIT, workers were asked to view an image of an urban space, and then rate their perceptual impression based on what they saw, for six labels. Additionally, annotators were given the option to describe how the urban space made them feel (as free-form text). 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, resulting in a total of 990 responses. Every worker was reimbursed 0.10 USD per impression.

Locals: In addition to obtaining the image corpus, we also obtained the associated "local" annotations of the same corpus from our previous work [7]. Images were annotated by a separate group of volunteers, which was different from the one which collected the data. To collect the impressions, a website was designed to allow volunteers to submit their annotations. Unlike the AMT study, each image was annotated more than 10 times, but to enable a fair comparison across the population groups, we randomly sampled 10 impressions per image. As a result, we collected a total of 990 responses for 99 images. No financial incentives were provided to the volunteers for their participation. Volunteers were intrinsically motivated to contribute towards the study. Note that local annotators knew that the images being annotated, were taken in Guanajuato.

All the local annotators are high-school students aged 16–18 years old. 95% of them are born in Guanajuato city; while those who were not born in Guanajuato, have lived in Guanajuato for at least 8 years. 85% of the annotators currently live either in Guanajuato city or suburban areas close to the city (within aprrox. 6KM).

Labels	Locals		MTurkers	
	Mean±SD	ICC	Mean±SD	ICC
Accessible	4.16±1.16	0.69	4.13±1.29	0.81
Dangerous	4.43±0.91	0.63	3.19±1.20	0.83
Dirty	4.33±1.24	0.68	3.25±1.26	0.85
Interesting	3.55±1.23	0.70	4.14±1.10	0.63
Preserved	3.54±1.28	0.76	3.63±1.30	0.78
Pretty	3.47±1.38	0.80	3.25±1.36	0.83

Table 1: Means, standard deviations and $ICC(1,k)$ of annotation scores for each label and group.

4 RESULTS

4.1 Annotations Quality

We begin our analysis by assessing the reliability of annotations given by locals and MTurkers. We measure the inter-rater 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 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 [14], are computed for each label across all images.

Table 1 reports the $ICC(1,k)$ values for both groups (due to space constraints, we omit $ICC(1,1)$ values.) We observe acceptable inter-rater consensus for most labels, with all values being statistically significant at p-value < 0.01. We notice that the inter-rater reliability for all labels is reasonably high (above 0.6) for both the groups. From Table 1, we find that label *pretty* (resp. *dirty*) achieved high agreement for locals (resp. MTurkers). Further, we find that for all labels (except *interesting*), the consensus between MTurkers is higher relative to locals. In other words, locals tend to disagree more for most of the urban awareness labels, when compared to online annotators.

4.2 Descriptive Statistics

As noted in Section 3.2, for each label and image, we collected 10 impressions. So, it becomes relevant 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 as median [15]. Thus, we computed the median score for each label given the 10 responses per image. Given the median scores, we then computed the mean scores and standard deviations for each label using all 99 images for both groups.

Table 1 lists the descriptive statistics for each population group and label. At the level of individual annotations, minimum and maximum values are 1 and 7 respectively for each label and population group, indicating that the full scale was used by locals and MTurkers alike. The mean scores for majority of labels is below 4, which indicates a trend towards disagreement with the corresponding label. Recall that annotators were asked to label images along a seven-point Likert

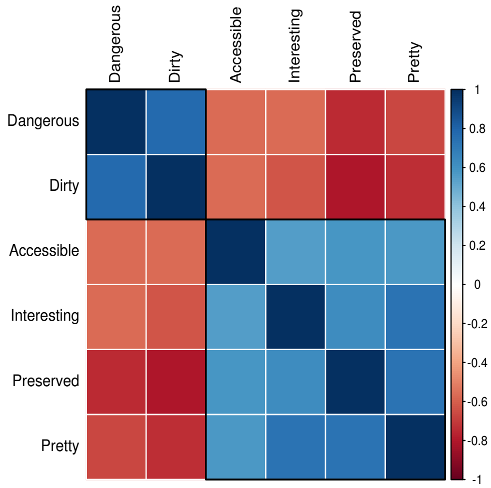


Figure 2: Plot showing the correlation matrix between all labels for MTurkers. Black rectangular borders indicate the two distinct clusters found in the correlation matrix. All cells are statistically significant at $p < 0.05$.

scale ranging from *strongly disagree* (1) to *strongly agree* (7). Based on the overall mean scores, locals annotators perceived images to be more *dangerous* and *dirty*, while MTurkers perceived them to be more *interesting*. A statistical analysis of these observed differences are presented later in Section 4.3.

It is natural to examine the relationship between the perceptual ratings and official government statistics, as in previous studies [9]. But, as is the case with most developing cities, due to a lack of publicly available data, a statistical analysis between these two sources, e.g. dangerous label ratings and crime rates of Guanajuato neighborhoods, was not feasible.

Correlation Analysis: To understand the statistical relationship between labels, we perform correlation analysis between all labels, for both the population groups. For both groups, we find that *accessible*, *interesting*, *preserved* and *pretty* labels are positively correlated, with pairwise correlations exceeding 0.7 (bottom-right rectangular box in Figure 2); furthermore *dangerous* and *dirty* labels are also positively correlated, with pairwise correlations exceeding 0.65 (top-left rectangular box in Figure 2). Figure 2 shows the correlation matrix between urban awareness labels for MTurkers. Similar matrix is obtained for locals, but we are not showing it due to space constraints. These findings corroborate earlier findings reported in the literature [13].

It is interesting to observe the correlation between the *dirty* and *dangerous* labels. Clearly, no causal relation can be inferred from here. Historically, some literature in urban sociology (now under much criticism) has postulated that the presence of physical disorder (garbage, grafitti, etc.) in urban environments might lead to social disorder (crime, fear) [18].

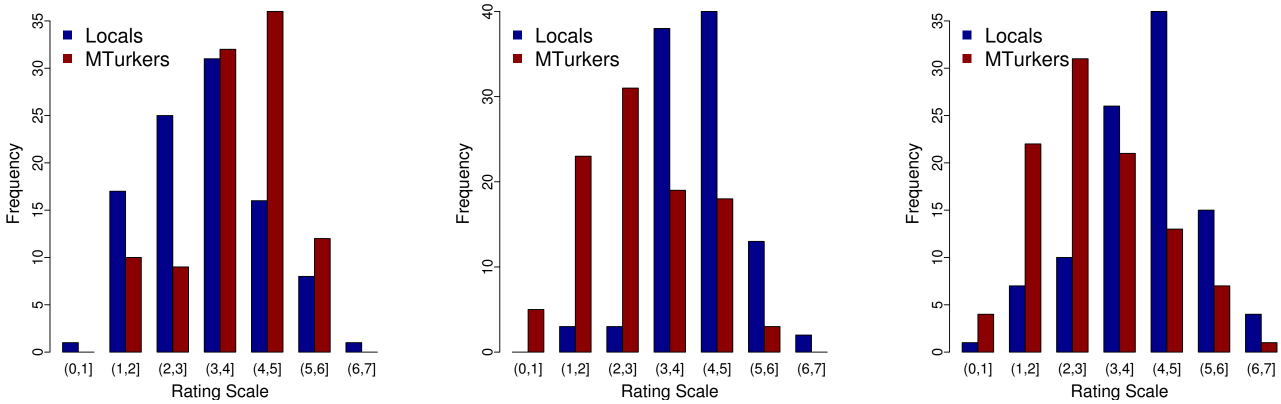
4.3 Comparing Impressions

Now we turn our attention towards comparing impressions between locals and MTurkers. From Table 1, we observe that mean values of perceptual ratings for all labels differ across both groups. To understand whether mean differences between groups for some of these labels are statistically significant, we perform the Tukey's honest significant difference (HSD) test. Tukey's HSD test is a statistical procedure to examine whether mean values are significantly different across groups [17]. We perform the Tukey HSD test to compute the pairwise comparisons of mean values between population groups for each label. Table 2 present the results. To complement Tukey's HSD statistics, we also show the plot comparing the distributions of perception ratings across both the population groups in Figure 3. Based on these statistics, we observe that:

(1) Images were perceived to be significantly more *dangerous* and *dirty* by locals when compared to MTurkers. When looking at the individual ratings, we found that 87% (resp. 81%) of images were rated by locals as more *dangerous* (resp. *dirty*) in comparison to MTurkers (Fig 3b and 3c).
(2) Images were perceived to be significantly more *interesting* by MTurkers than locals, a contrasting trend relative to the previous observation. 57% of images were rated on a higher *interesting* scale by MTurkers than locals, when looking at the individual ratings (Figure 3a).
(3) For the rest of the labels, the range of perceptions elicited are not statistically different between locals and MTurkers. Both groups found images to be equally *accessible*, *preserved* and *pretty*.

Discussion: The results presented above point towards a clear difference between groups across three dimensions. We believe that locals are using familiarity with the place, background information and prior beliefs, while judging places on being *dangerous* and *dirty*. Recall that locals knew that images were taken in Guanajuato city (Section 3.2). On the other hand, external observers viz. MTurkers form their impressions based on visual cues present in an image. For online annotations, we have used US-based MTurkers, so it is probable that they might not have seen these kinds of urban scenes or terrain in the past. This behavior is particularly evident in the way MTurkers judged places to be on a higher *interesting* scale compared to locals.

To further examine this behavior, we analyzed some of the comments made by MTurkers. Remarks such as "I want to wander through the street" and "This space gives me a sense of wonder, I feel as if i need to explore whats around the corner" were made on images which were rated significantly high on being *dangerous* by locals, but low by MTurkers. Additionally, comments like "This space makes me feel warm because of the brick and the ambiance" and "A picturesque backdrop to step into ancient history!" were given on images which were rated significantly high on being *interesting* by MTurkers, and vice-versa for locals. These comments further highlight the crucial role of background knowledge and beliefs in forming urban perceptions.



(a) Interesting **(b) Dangerous** **(c) Dirty**

Figure 3: Plots comparing the distributions of perception ratings for a) Interesting, b) Dangerous, and c) Dirty, for both locals and MTurkers.

Label	Group Pair	Mean Difference	$p-$value
Accessible	LO–MT	+0.03	0.89
Dangerous	LO–MT	+1.24	**0.00**
Dirty	LO–MT	+1.08	**0.00**
Interesting	LO–MT	−0.59	$\mathbf{5 \times 10^{-3}}$
Preserved	LO–MT	−0.09	0.62
Pretty	LO–MT	+0.22	0.24

Table 2: Tukey's HSD statistics. LO and MT respectively stands for locals and MTurkers. Values in bold are statistically significant at $p < 0.01$

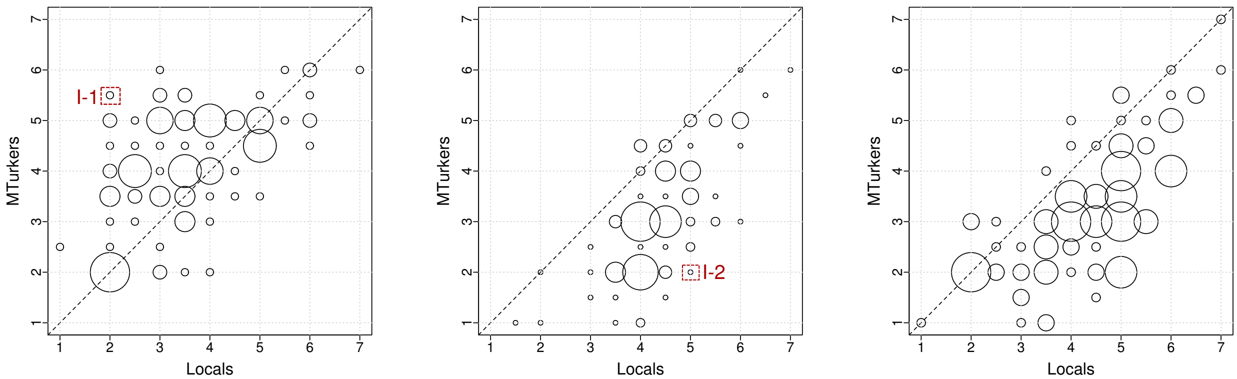
4.4 Pair-wise Analysis

In the previous section, we have established that perceptual ratings differ between groups. Now in order to understand the variability of ratings for individual images, we examined the pair-wise ratings of each image between groups. We focus our pair-wise analysis on the statistically significant labels. Figure 4 shows the respective plots. If the perception ratings were similar between groups, 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 *dangerous* and *dirty* (Figure 4b and 4c), indicating that locals perceive these labels on a higher scale compared to MTurkers. Reverse trend is observed for *interesting* label (Figure 4a). These plots further validate our findings reported in Section 4.3.

Qualitative Analysis: In Figure 4, we observe that some of the images differ in ratings significantly. For some of the images, the perceptions between groups differ by more than three ordinal scales. Figure 5 shows two such images. One of the authors of our study is a local resident of Guanajuato, and we asked him to provide possible explanations to interpret some of these disparities.

Figure 5a (marked as I-1 in Figure 4a) depicts an image which was rated 2 (resp. 5.5) by locals (resp. MTurkers) on the *interesting* scale. The image shows a view taken from south to north. If one sees in the opposite direction, the street connects to the busiest street in Guanajuato, which is possibly the reason why local people have judged it as not very interesting; while, MTurkers found the view quite interesting as it conveys a picturesque town or as one of the MTurkers remarked: "This space makes me feel serene." The second image (Figure 5b) shows a street that is very close to a dangerous part of the city. It is not surprising that locals have rated it as 5 on the *dangerous* scale, while MTurkers only 2, as one would by just looking at the image. To the naked eye, the image does not look unsafe at all, which was corroborated by one of the MTurkers: "This area looks moderately well-off, though by no means wealthy; I feel that this area is relatively safe."

Further, it is interesting to observe the disparity in ratings for the *dirty* label. As shown in Figure 4c, for some of the images the difference in ratings between locals and MTurkers is high (by two or more ordinal scales). As with other labels, we want to examine the possible causes for the difference in ratings. We manually browsed few images, which are rated low by MTurkers and high by locals, and found those images to be relatively *clean* to our naked eyes. We show one such image in Figure 1c, which was rated 5 (resp. 2) by locals (resp. MTurkers) on the *dirty* scale. For this particular image, one of the MTurkers commented: "This space makes me feel relaxed, like I'm taking a stroll through a quaint European city on my way to brunch or lunch.", which corroborates with the ratings given by MTurkers. A plausible hypothesis to support this finding might be the expectations of locals (i.e., high-school students in our study) to be sensitive about the *dirty* issue. Being a youth resident, one has high expectations for his neighbourhood or the city to be clean and thus would rate the images of his city *dirty* on a higher scale, with the expectation that the urban spaces should be more cleaner than their status quo. On the contrary, one might also argue that since locals are used to seeing the place as is (i.e. being dirty or clean), there is no obvious reason to explain why they would rate images as more *dirtier* than external observers. We plan to investigate this issue in more depth, as part of future work.



(a) Interesting (b) Dangerous (c) Dirty

Figure 4: Scatter plots showing the pair-wise annotator ratings by Locals and MTurkers for a) Interesting, b) Dangerous, and c) Dirty. Each dot corresponds to an image, with the size of the dots proportional to the number of images at a given rating pair. 45° line is also shown in all the plots. Two dots highlighted in the plots as I-1 and I-2 are enlarged in Figure 5a and 5b respectively.

5 LIMITATIONS AND FUTURE WORK

In this section, we address the three limitations of our work and outline potential future directions. As a first limitation, while comparing the observer differences, we did not control for observers' demographics attributes e.g., gender, age, ethnicity, education status, or familiarity with living or experiencing urban environments in developing countries. We will examine the effect of demographics differences on the perceptual ratings as part of future work. Moreover, we also plan to conduct an additional crowdsourcing study to explore the differences between youth populations of Mexico and other developed countries e.g., Switzerland. Second, in this paper we have examined the observer differences using 99 images. We plan to extend our analysis to include more images and annotations as well as extending the study to multiple cities. Third, due to the subjective nature of the labels, it is difficult to contextualize some of the findings reported in the paper. For some of the studied labels (e.g., *interesting*, *preserved*, *pretty*), there exist no ground truth or "gold standard". In order to contextualize some of the findings or evaluate the applications of the perception work, an interesting future analysis will 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 of urban places.

In addition to addressing these limitations, an interesting future direction will be to perform visual content analysis by training a machine learning classifier to infer perceptual scores of different observer populations using automatically extracted visual cues, following recent work [1, 12].

(a) I-1 (b) I-2

Figure 5: Images where the perceptual ratings differ significantly between local and MTurk population.

6 CONCLUSION

In this paper, we investigated how the impressions of outdoor urban spaces judged by online annotators, compares with the impressions elicited by local residents, along six psychological labels. We focus our study in a developing city where understanding and characterization of these socio-urban perceptions is of societal importance. We found statistically significant differences between the two population groups. Locals perceived places to be more *dangerous* and *dirty*, when compared with non-local online annotators; while online annotators judged places to be more *interesting* in comparison to locals. Our findings can be potentially integrated with existing online photo-sharing platforms (Flickr, Instagram, etc.), to enhance the view of a place by using "insiders" knowledge for "outsiders". While our study was conducted using data collected from one city, it clearly shows that there is a need to better characterize the observers of crowdsourced urban perception studies as biases due to background knowledge and context are significant.

7 ACKNOWLEDGEMENTS

This work has been supported by the SNSF Youth@Night and EPFL-CODEV SenseCityVity 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 their support.

REFERENCES

[1] Abhimanyu Dubey, Nikhil Naik, Devi Parikh, Ramesh Raskar, and Cesar A Hidalgo. 2016. Deep Learning the City: Quantifying Urban Perception at a Global Scale. In *European Conference on Computer Vision (ECCV)*. Springer.

[2] Kazunori Hanyu. 1997. Visual Properties and Affective Appraisals in Residential Areas After Dark. *Journal of Environmental Psychology* (1997).

[3] Pall J. Lindal et al. 2013. Architectural Variation, Building Height, and the Restorative Quality of Urban Residential Streetscapes. *Journal of Environmental Psychology* (2013).

[4] Laura J. Loewen et al. 1993. Perceived Safety from Crime in the Urban Environment. *Journal of Environmental Psychology* (1993).

[5] Kate Painter. 1996. The Influence of Street Lighting Improvements on Crime, Fear and Pedestrian Street Use, After Dark. *Landscape and Urban Planning* (1996).

[6] Daniele Quercia, Neil Keith O'Hare, and Henriette Cramer. 2014. Aesthetic Capital: What Makes London Look Beautiful, Quiet, and Happy? *CSCW* (2014).

[7] Salvador Ruiz-Correa, Darshan Santani, and Daniel Gatica-Perez. 2014. 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)*.

[8] S Ruiz-Correa et al. 2017. SenseCityVity: Mobile Crowdsourcing, Urban Awareness, and Collective Action in Mexico. *IEEE Pervasive Computing* (2017).

[9] Philip Salesses, Katja Schechtner, and Cesar A Hidalgo. 2013. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS ONE* (2013).

[10] Robert J Sampson and Stephen W. Raudenbush. 2004. Seeing Disorder: Neighborhood Atigma and the Social Construction of "Broken Windows". *Social Psychology Quarterly* (2004).

[11] Darshan Santani and Daniel Gatica-Perez. 2015. 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)*. ACM.

[12] Darshan Santani, Rui Hu, and Daniel Gatica-Perez. 2016. InnerView: Learning Place Ambiance from Social Media Images. In *Proceedings of the 2016 ACM on Multimedia Conference (MM '16)*. ACM.

[13] Darshan Santani, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2015. Looking at Cities in Mexico with Crowds. In *Proceedings of the 2015 Annual Symposium on Computing for Development*. ACM.

[14] Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin* (1979).

[15] S. S. Stevens. 1946. On the Theory of Scales of Measurement. *Science* (1946).

[16] Martin Traunmueller, Paul Marshall, and Licia Capra. 2016. ... when you're a Stranger: Evaluating Safety Perceptions of (un) familiar Urban Places. In *Proceedings of the Second International Conference on IoT in Urban Space*. ACM.

[17] John W Tukey. 1991. The Philosophy of Multiple Comparisons. *Statist. Sci.* (1991).

[18] James Q Wilson et al. 1982. Broken Windows. *Atlantic Monthly* (1982).