Landscape and Urban Planning 181 (2019) 169–178

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Research Note

Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach

Tomás Rossetti[a,b,1], Hans Lobel[a,c], Víctor Rocco[a,b], Ricardo Hurtubia[a,b,d]

[a] Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Chile
[b] Centre for Sustainable Urban Development (CEDEUS), Chile
[c] Department of Computer Science, Pontificia Universidad Católica de Chile, Chile
[d] School of Architecture, Pontificia Universidad Católica de Chile, Chile

ARTICLE INFO

Keywords:
Perceptions
Discrete choice models
Machine learning
Public spaces

ABSTRACT

People's perceptions of the built environment influence the way they use and navigate it. Understanding these perceptions may be useful to inform the design, management and planning process of public spaces. Recently, several studies have used data collected at a massive scale and machine learning methods to quantify these perceptions, showing promising results in terms of predictive performance. Nevertheless, most of these models can be of little help in understanding users' perceptions due to the difficulty associated with identifying the importance of each attribute of landscapes. In this work, we propose a novel approach to quantify perceptions of landscapes through discrete choice models, using semantic segmentations of images of public spaces, generated through machine learning algorithms, as explanatory variables. The proposed models are estimated using the Place Pulse dataset, with over 1.2 million perceptual indicators, and are able to provide useful insights into how users perceive the built environment as a function of its features. The models obtained are used to infer perceptual variables in the city of Santiago, Chile, and show they have a significant correlation with socioeconomic indicators.

1. Introduction

Urban landscapes are experienced by their users in great part through visual perceptions. This has been explored in the literature, showing that perceptions can influence the intensity of use of public spaces (Khisty, 1994; Shriver, 1997) or encourage the use of certain transportation modes (Antonakos, 1995; Hunt & Abraham, 2007; Hyodo, Suzuki, & Takahashi, 2000; Jiang, Christopher Zegras, & Mehndiratta, 2012; Tilahun and Li, 2015; Zacharias, 2001). Given this, it is relevant to understand how public spaces are perceived, which allows to identify possible interventions that can nudge users' behavior towards more sustainable practices, such as preferring denser neighborhoods or active transportation.

The literature has reported multiple efforts to quantify these perceptions, dealing with costly data-collection processes. Nevertheless, a recent body of literature has made use of massive data-collection techniques to quantify these perceptions using machine learning models. In spite of the advantages these models present in terms of their data processing and predictive abilities, they do not provide information that directly explains the drivers behind respondents' decisions. Because they work as "black boxes," there is no straightforward way of verifying if certain attributes, such as the presence of vegetation or pedestrians, have positive or negative impacts over respondents' perceptions. Moreover, they cannot allow to infer elasticities between the presence of different features, which could help understand how people are willing to trade off between them.

In this work, we propose a novel methodology to quantify perceptions of images of public spaces using a massive dataset: Place Pulse (Salesses, Schechtner, & Hidalgo, 2013). This project takes advantage of the opportunity provided by platforms like Google Street View that freely make public a virtually global coverage of two-dimensional images of public spaces in several cities. While two-dimensional images are not capable of conveying the whole complexity of a public space, they provide a good proxy for this. With this dataset, we argue in favor of the methodological usefulness of discrete choice models whose primary inputs are semantic segmentations of images of public spaces. These not only are theoretically better suited for many datasets and efficient enough for massive amounts of data, but have interpretable

E-mail address: terosset@uc.cl (T. Rossetti).
[1] Address: Vicuña Mackenna 4860, Macul, Santiago, Chile.

https://doi.org/10.1016/j.landurbplan.2018.09.020
Received 21 March 2018; Received in revised form 21 September 2018; Accepted 24 September 2018
Available online 11 October 2018

parameters, which are not a straightforward output from machine learning models.

The proposed models predict respondents' perceptions in an adequate manner, while providing informative econometric parameters that can be used to better understand users' perceptual reactions to meaningful landscape features, as well as inform public policy. Moreover, perceptual maps obtained for the case of Santiago, Chile, show significant correlations between perceptions and socioeconomic variables, suggesting these qualitative attributes can effectively predict the characteristics of cities at a fine level.

The remainder of this article is organized as follows: Section 2 presents a brief review of previous efforts to quantify perceptions of public spaces. Then, Section 3 describes the source of the perceptual indicators used, the analysis of the images carried out, and the model estimation. Section 4 shows the results obtained and a brief analysis. Section 5 presents a validation of the model. Finally, Section 6 presents main conclusions.

2. Literature review

Since Kevin Lynch's (1960) seminal work "The image of the city," there has been an increasing interest among urban planners and designers to understand how people perceive their environment. Later studies have found significant links between perceptions and actual use of public spaces (Antonakos, 1995; Hyodo et al., 2000; Jiang et al., 2012; Khisty, 1994; Shriver, 1997; Zacharias, 2001). Latkin and Curry (2003) even established a relation between perceived neighborhood disorder and depression.

Having established a relation between perceptions and use of public spaces, the obvious question that arises is what elements affect these perceptions. There have been several studies that relate the presence of specific elements with perceptions, such as density (Bonaiuto, Fornara, & Bonnes, 2003; Cetintahra and Cubukcu, 2015; Van Dyck, Cardon, Deforche, & De Bourdeaudhuij, 2011), vegetation (Kuo et al., 1998; Sheets and Manzer, 1991), sidewalks (Foster, Hooper, Knuiman, Bull, & Giles-Corti, 2016; Gilderbloom, Riggs, & Meares, 2015), overhead cables and poles (Crystal and Brush, 1978), and litter (Mertens et al., 2016).

As Clifton, Ewing, Knaap, and Song (2008) state in their review, measuring perceptions of public spaces is particularly costly because the only way of collecting them is by asking people to assess a particular street or neighborhood. A group of recent studies have solved this problem by using online, crowdsourced surveys, which provide them with a massive number of perceptual indicators at an extremely low cost. These surveys usually present respondents with a pair of images and ask them to choose which one adjusts best to a qualitative attribute (i.e., "safe," "happy," "depressing"). Some examples are UrbanGems (http://urbangems.org/; Quercia et al., 2014) and Place Pulse (http://pulse.media.mit.edu/, Salesses et al., 2013). These studies group indicators (which in both cases mentioned are "votes" given by respondents) in different ways and use machine learning methods to quantify perceptions (Dubey, Naik, Parikh, Raskar, & Hidalgo, 2016; Krizhevsky, Sutskever, & Hinton, 2012; Liu, Silva, Wu, & Wang, 2017; Naik, Philipoom, Raskar, & Hidalgo, 2014; Ordonez & Berg, 2014; Porzi et al., 2015).

In spite of the advantages these models present in terms of their data processing and predictive abilities, they do not provide information that directly explains the drivers behind respondents' decisions. These "black box" models require ad-hoc analyses to indirectly infer if certain attributes, such as the presence of vegetation or pedestrians, have positive or negative impacts over respondents' perceptions. Moreover, they cannot allow to infer elasticities between the presence of different features, which could help understand how people are willing to trade off between them. The only exception found is the case of Porzi et al. (2015), that is able to identify through an ex post analysis which elements of an image have a more relevant effect over the perceptual score they constructed. Nevertheless, this process can hardly be systematized given the number of images necessary to estimate these models.

Even though some efforts have been made to open up these "black box" models (see, for example, Adler et al., 2016), we believe this weakness can be avoided in a more straightforward manner through the use of econometric models founded on behavioral theories, which have not been tested in this context. We propose to complement machine learning algorithms with non-linear econometric models, specifically discrete choice ones. These models find the correlations between individual responses to indicators with features of images through a non-linear regression. This allows not only to predict users' manifested perceptions, but also to inform which features of scenes have a significant effect over them.

3. Methodology and data collection

This section shows the methodology used to obtain the model results presented in Section 4. First, the source of the perceptual indicators is explained. Later, the image parametrization is detailed. Finally, the methodology behind model estimation is briefly described.

Readers should be aware that neither we nor the works cited in the previous section are directly quantifying perceptions. What we are doing is quantifying the relationship between observable perceptual indicators, that are a manifestation of cognitive processes, with observable indicators of landscapes. We are therefore quantifying landscape-perception relations and not directly quantifying perceptions. Nevertheless, and for the sake of simplicity, we will refer to this as quantifying perceptions in a general way.

3.1. Perceptual indicators

The Place Pulse 2.0 dataset (Dubey et al., 2016) contains the results of a large-scale effort to obtain perceptual indicators of urban and suburban public spaces. This platform seeks to survey people's perceptions around the world through crowdsourcing on the Place Pulse web site (http://pulse.media.mit.edu/). Every respondent here is presented with two images taken at random from within one of 56 predefined cities, and is asked to state which one adjusts better to a certain adjective. Note that responses are given with respect to a specific pair of images and a specific adjective. This dataset includes responses for the following qualitative attributes: "Beautiful," "Boring," "Depressing," "Lively," "Safe," and "Wealthy." All images are taken from Google Street View (GSV), which is why they generally show streets during the day that do not contain many pedestrians or cyclists. An example of a choice task is shown in Fig. 1.

The database, free for anyone to use, contained a total of 1,223,649 choices, or perceptual indicators, when downloaded for this study. For each choice, the geographical coordinates of each scene are informed, along with the choice that was made (left scene, right scene, or if both were deemed to be equal). We downloaded this dataset from the Place Pulse website. We also downloaded the GSV images considered using Google's API and following instructions provided by the creators of Place Pulse, which state that images should be downloaded using the coordinates provided and the default camera direction. Note that not all choices were considered in this study because some images were identified as outliers and eliminated (see Section 3.3).

Respondents were contacted through ads on social media, with a special interest in English-speaking countries. Over 65% of the responses came from the United States, India, the United Kingdom, Brazil, and Canada. On average, each respondent successfully provided 16.6 perceptual indicators (Dubey et al., 2016). Unfortunately, the dataset provided does not contain any information that may indicate respondents' characteristics.

Responses to these questions might have differed if these public spaces were presented in other formats, particularly if respondents were able to view them in person (Huang, 2009; Svobodova, Vojar, Sklenicka, & Filova, 2017). Nevertheless, and to the best of our knowledge, the literature has shown these biases exist in absolute terms and not in relative ones. This leads us to believe comparison tasks such as the ones presented in Place Pulse do not induce significant biases. Another limitation this dataset has is that only features from the specific two-dimensional view presented to respondents may be used for econometric analysis, while three-dimensional landscape characteristics are what is interesting for landscape and public space management. However, these two-dimensional characteristics have significant correlations with landscape characteristics (Dramstad, Tveit, Fjellstad, & Fry, 2006; de la Fuente de Val, Atauri, & de Lucio, 2006; Frank, Fürst, Koschke, Witt, & Makeschin, 2013; Schirpke, Tasser, & Tappeiner, 2013), and therefore are informative for environmental and urban policies.

3.2. Image parametrization

Interpretable and meaningful landscape features have to be extracted from two-dimensional representations of landscapes to obtain informative models. To do this, we decided to extract low and high-level features, which will be explained in more detail in the following sections.

3.2.1. Low-level features

Low-level features can be obtained at very low computational costs and give broad information on the image. We selected eight low-level features for their potential use. Each variable is defined as follows:

1. Edges: The percentage of pixels of each image that are determined to be an edge through the Canny Edge Algorithm (Canny, 1986). An example is shown in Fig. 2.
2. Blobs: The number of blobs (binary large objects) found in the image, defined as a group of connected pixels with homogeneous characteristics, such as color or texture. An example is shown in Fig. 3.
3. Averages for HSL channels: Average values for the hue (the characteristic of a color that makes it appear similar to red, yellow, green or blue), saturation (the attribute that makes a color appear more or less chromatic) and lightness (the attribute that makes a color appear to emit more or less light) channels of the HSL color space (Westland, 2016). This color space was chosen because it has a more intuitive interpretation than the RGB space, and because its use imposes a low computational cost (Lissner and Urban, 2012).
4. Standard deviations for the HSL channels: standard deviation for the hue, saturation and lightness channels of the HSL color space.

3.2.2. High-level features

One of this work's main objectives is to quantitatively identify which physical elements of public spaces affect qualitative perceptions. Because of this, a necessary step is to infer which physical elements are present in an image. We resort to machine learning-based visual recognition techniques, and particularly to semantic segmentation methods (Fulkerson et al., 2009), to do this.

The problem of semantic segmentation is defined as the assignation of each of the pixels in an image to one possible category, where categories are predefined and generally correspond to semantically meaningful visual elements, like cars, buildings, or pedestrians. In this work we use the SegNet method (Badrinarayanan et al., 2015) for semantic segmentation, which is based on a 27-layer Convolutional Neural Network architecture. The model's outputs are grayscale images of the same size as the originals, where each pixels' value indicates its corresponding category.

To generate the segmentations, we use the SegNet model trained on a subset of the CamVid dataset (Brostow, Fauqueur, & Cipolla, 2009), which consists of videos recorded near Cambridge, United Kingdom, using a car-mounted camera. The subset used to train the model contains 367 training images with a resolution of 360×480 pixels. The following visual categories or classes are considered in this dataset: "Building," "Car," "Cyclist," "Fence," "Pedestrian," "Pole," "Road," "Sidewalk," "Sky," "Traffic sign," and "Vegetation."

In order to capture the relative weights of the physical elements present in an image, we evaluated various specifications for the information contained in the semantic segmentation model's outputs. The one that produced best results in terms of fit to the data is presented in Eq. (1), where p_{ics} is the proportion of pixels of class c in section s of image i. We defined only two sections, lower and upper halves of the image, to retain interpretability. The logarithmic specification was used to account for a decreasing effect of the intensity of each element on perceptions.

$$HLF_{ics} = log(100p_{ics} + 1) \tag{1}$$

3.3. Methodology and model specification

This work will assume perceptions can be modeled through latent variables, following a long tradition in several fields such as engineering (Golob, 2003; Walker and Ben-Akiva 2002) and psychology (Bollen, 2002; Borsboom, Mellenbergh, & van Heerden, 2003). This framework, presented in Fig. 4, allows to determine which observable variables correlate with unobservable variables (in this case, perceptions) through the use of structural equations. The main intuition behind these models is that features of images affect their perceptual attributes, and that different features have different weights over these perceptions. What the estimation process does is find these weights in such a way that they mimic observed perceptual indicators as closely as possible. If a weight related to a specific feature is significantly different than zero, then we can conclude that this feature and the perceptual variable in question are correlated.

This approach's first assumption is that qualitative attributes, such as perceptions, can be explained by observable variables through a stochastic structural relation such as the one shown in Eq. (2). This means that features of scenes, represented by vector X_i, and characteristics of observers, represented by vector X_n, are related to a (latent) perceptual variable L^*_{in} through equation L and parameters λ. An error term, ε_{in}, is included to account for measurement errors and variables not included in L_{in}. Note that relation L is analogous to indexes found in the literature (see, for example, Frank et al., 2010), which construct a score as a function of the weighted sum of different observable variables. In those cases, X_i and X_n represent the input variables and λ represent the weights used.

$$L^*_{in} = L(X_i, X_n; \lambda) = L_{in} + \varepsilon_{in} \tag{2}$$

The framework's second assumption is that perceptions manifest themselves through observable variables. This establishes a measurement relation between L^*_{in} and an indicator observed with respect to scene i and observer n, I_{in}, as is shown in Eq. (3). This measurement relation states that I_{in} can be described as a function of L^*_{in} through I and parameters α. Note that, in this specific case, indicators are "choices" given in the Place Pulse survey. Indicators may take other forms in different data collection settings.

$$I_{in} = I(L^*_{in}; \alpha) \tag{3}$$

The Place Pulse dataset presents two scenes and ask users to state which one adjusts better to a certain qualitative attribute. Because we are modeling these attributes stochastically, it is impossible to deterministically state which one will be chosen. Rather, our models will only be able to estimate with what probability one image will be selected. More formally, this probability is shown in Eq. (4), which in turn

Fig. 1. Example of choice experiment presented in Place Pulse.

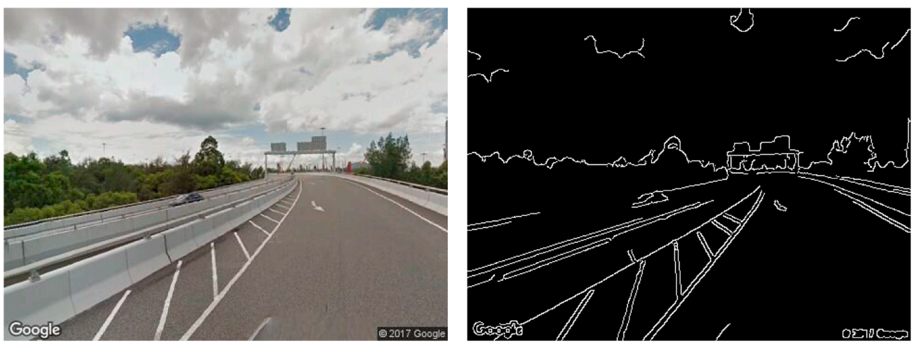


Fig. 2. Example of images presented in the Place Pulse dataset (left) and edges detected in it (right).

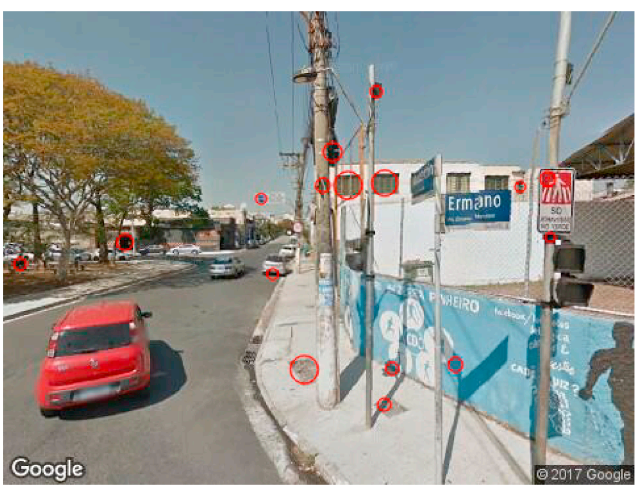


Fig. 3. Example of 16 blobs, shown inside red circles, found in an image presented in Place Pulse.

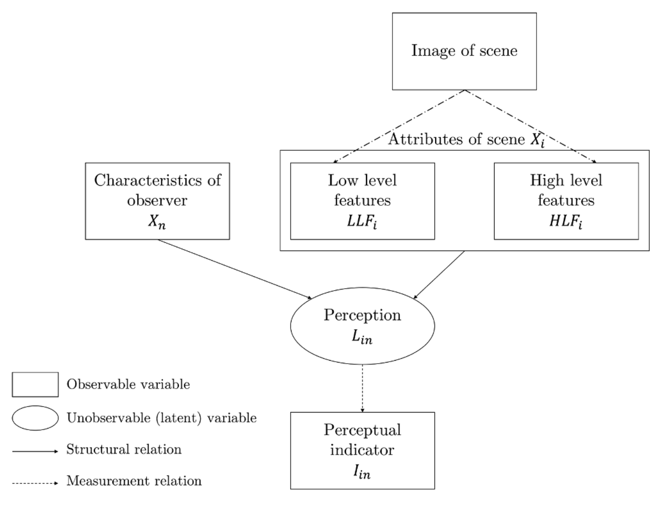


Fig. 4. Framework.

can be rewritten as (5). Note that in this case we are assuming scene i will be selected over scene j if their latent variables differ by an amount greater than δ, which is a parameter to be estimated.

$$P_i = \Pr(L_{in}^* > L_{jn}^* + \delta) \tag{4}$$

$$= \Pr(L_{in} + \varepsilon_{in} > L_{jn} + \varepsilon_{jn} + \delta)$$

$$= \Pr(\varepsilon_{in} - \varepsilon_{jn} > L_{jn} - L_{in} + \delta) \tag{5}$$

Different assumptions on the distribution of the error terms ε yield different specifications for P_i. In this work we will assume they have a Type I Extreme Value distribution, which yields a logit model (McFadden, 1974). This kind of model is widely used because of its lower estimation costs and negligible differences in parameter estimates when compared to the probit model that arises from the assumption that ε have a Normal distribution (Lee, 1982; Ruud, 1983). These probabilities are shown in Eqs. (6) and (7), where i and j are the two scenes shown, e is the indifference alternative (selected when both scenes are deemed to have a similar amount of the qualitative attribute in question), and δ is the indifference threshold. μ is a scale parameter related to the error terms.

$$P_i = \Pr(L_{in}^* > L_{jn}^* + \delta) \tag{6}$$

$$= \frac{1}{1 + \exp(\mu(L_{jn} - L_{in} + \delta))}$$

$$P_e = \Pr(|L_{in}^* - L_{jn}^*| < \delta) \tag{7}$$

$$= 1 - P_i - P_j = (\exp(2\mu\delta) - 1)P_{in}P_{jn}$$

Table 1
Parameters obtained for models that quantify qualitative attributes.

	Beautiful	Boring	Depressing	Lively	Safe	Wealthy
Constant (left)	−0.0374** (−7.63)	−0.00708 (−1.27)	−0.0168** (−3.04)	−0.0378** (−9.65)	−0.0272** (−8.21)	−0.0279** (−5.38)
Building (top)	−0.0982** (−25.03)	−0.0525** (−10.64)	0.0399** (8.86)	0.0571** (16.43)	−0.0210** (−6.97)	−0.0162** (−3.63)
Building (bottom)	−0.0698** (−11.72)	0.0303** (4.74)	0.0544** (8.13)	−0.0391** (−8.22)	−0.0309** (−7.69)	−0.0477** (−7.72)
Car (top)	−0.0738** (−10.58)	0.0367** (4.59)	0.0675** (8.58)	−0.0517** (−9.15)	−0.0480** (−10.03)	−0.0935** (−12.41)
Car (bottom)		−0.0450** (−9.11)	−0.0508** (−10.15)	0.105** (28.42)	0.0637** (20.56)	0.0675** (14.06)
Cyclist (top)	0.256** (4.58)				0.133** (3.54)	0.200** (3.34)
Cyclist (bottom)	−0.0966** (−5.56)	−0.0494* (−2.42)	0.0381 (1.91)	0.0794** (5.57)	−0.0272* (−2.25)	−0.0405* (−2.14)
Fence (top)	0.0996** (17.00)	−0.0492** (−7.37)	−0.0754** (−11.46)	0.0688** (14.59)	0.0749** (18.60)	0.0919** (14.96)
Fence (bottom)	0.0485** (8.44)	−0.0543** (−8.28)	−0.0644** (−9.91)	0.0290** (6.19)	0.0147** (3.74)	
Pedestrian (top)		−0.0682* (−2.80)	−0.0687** (−2.88)	0.0595** (3.50)	0.0879** (5.98)	0.0906** (3.95)
Pedestrian (bottom)		−0.175** (−14.19)	−0.0414** (−3.30)	0.139** (15.62)	0.0200* (2.66)	−0.0321* (−2.71)
Pole (top)					−0.0184* (−2.45)	0.0330** (2.87)
Pole (bottom)	−0.0900** (−8.42)		0.0516** (4.19)	−0.0840** (−9.55)	−0.152** (−21.24)	−0.179** (−16.08)
Road (top)				−0.0185* (−2.01)		−0.0298* (−2.44)
Road (bottom)	0.113** (12.18)	−0.0241* (−2.36)	−0.118** (−11.53)	0.121** (16.06)	0.188** (29.82)	0.181** (18.44)
Sidewalk (top)	−0.104** (−10.58)	0.0589** (5.24)	0.111** (10.00)	−0.105** (−11.82)	−0.151** (−22.26)	−0.138** (−11.75)
Sidewalk (bottom)	−0.0347** (−4.52)			0.0491** (7.60)	0.0585** (10.82)	0.0461** (5.52)
Sky (top)	−0.0470** (−10.75)	0.0311** (6.09)	0.0398** (8.10)	−0.0140** (−3.94)	−0.0147** (−4.74)	−0.0313** (−6.49)
Sky (bottom)						
Traffic sign (top)	0.0894** (11.58)	−0.0453** (−5.26)	−0.0630** (−7.22)	0.0526** (8.44)	0.0480** (8.66)	0.0765** (8.82)
Traffic sign (bottom)	−0.0299** (−3.34)		0.0408** (4.07)	0.0279** (3.87)		
Vegetation (top)		−0.0183* (−2.12)		0.0301** (4.92)	0.0211** (3.98)	
Vegetation (bottom)	0.126** (24.97)	0.0474** (8.06)	−0.0856** (−14.52)	−0.0426** (−9.89)	0.0755** (20.92)	0.0554** (9.85)
Edges	6.87** (28.30)	−3.51** (−12.36)	−6.48** (−22.65)	4.76** (23.57)	4.64** (26.79)	6.07** (22.38)
Blobs		−0.00747** (−8.42)	−0.00191* (−2.17)	0.00785** (12.57)	0.00465** (8.86)	0.00616** (7.43)
Hue average	0.00350** (7.99)	−0.00149** (−3.13)	−0.00294** (−6.26)	0.000302 (0.84)	0.000674* (2.22)	0.00356** (7.68)
Saturation average	0.00252** (3.88)	0.00495** (6.64)	−0.00136 (−1.87)	−0.00253** (−4.82)	0.00197** (4.38)	0.000510 (0.73)
Lightness average	0.00108** (3.54)	0.00212** (5.60)	−0.00161** (−4.30)	0.0000444 (0.17)	0.00250** (10.98)	0.00130** (3.69)
Hue std. deviation	0.0142** (17.58)	−0.000936 (−1.03)	−0.00764** (−8.43)	0.00517** (7.95)	0.0110** (20.05)	0.00797** (9.30)
Saturation std. deviation	0.000963 (1.33)	−0.00633** (−7.64)	−0.00228* (−2.80)	0.00589** (10.10)	0.00119* (2.42)	0.00437** (5.65)
Lightness std. deviation	0.00624** (11.23)	−0.00418** (−6.57)	−0.00555** (−8.81)	0.00327** (7.28)	0.00336** (8.86)	0.00516** (8.80)
Threshold (δ)	0.284** (152.22)	0.343** (148.38)	0.348** (150.73)	0.248** (179.00)	0.259** (216.41)	0.286** (145.25)
Number of observations	161,472	117,453	122,048	246,329	339,885	140,330
Indifference choices	12.8%	16.3%	16.2%	11.7%	12.3%	13.4%
Number of parameters	24	25	26	29	29	28
Log-likelihood	−150,814.91	−117,580.73	−120,600.49	−232,350.21	−325,593.85	−134,565.59
First preference recovery [a]	64.5%	59.8%	62.0%	60.7%	59.5%	61.6%
LLF likelihood ratio test [b]	6,370.27**	1,388.89**	2,327.71**	4,278.15**	6,772.63**	3,086.46**

"(top)" and "(bottom)" variables correspond to the upper and lower halves of the images respectively.
[a] Proportion of observed choices, excluding indifference choices, that have the greatest estimated utility.
[b] Log-likelihood ratio test statistic when comparing full model with model without high-level features. **p < 0,01; *p < 0,05.

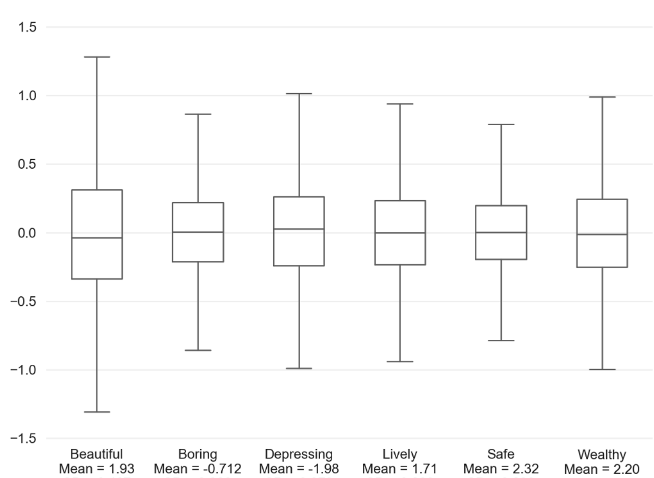


Fig. 5. Box plot of the qualitative attributes estimated, along with their means and standard deviations (SD). They were zero-mean normalized because they have incommensurate scales, which makes comparison between different perceptual variables impossible.

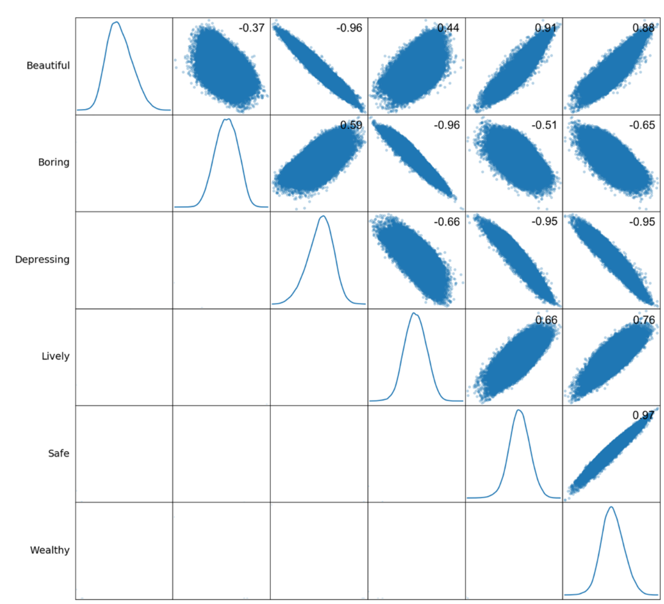
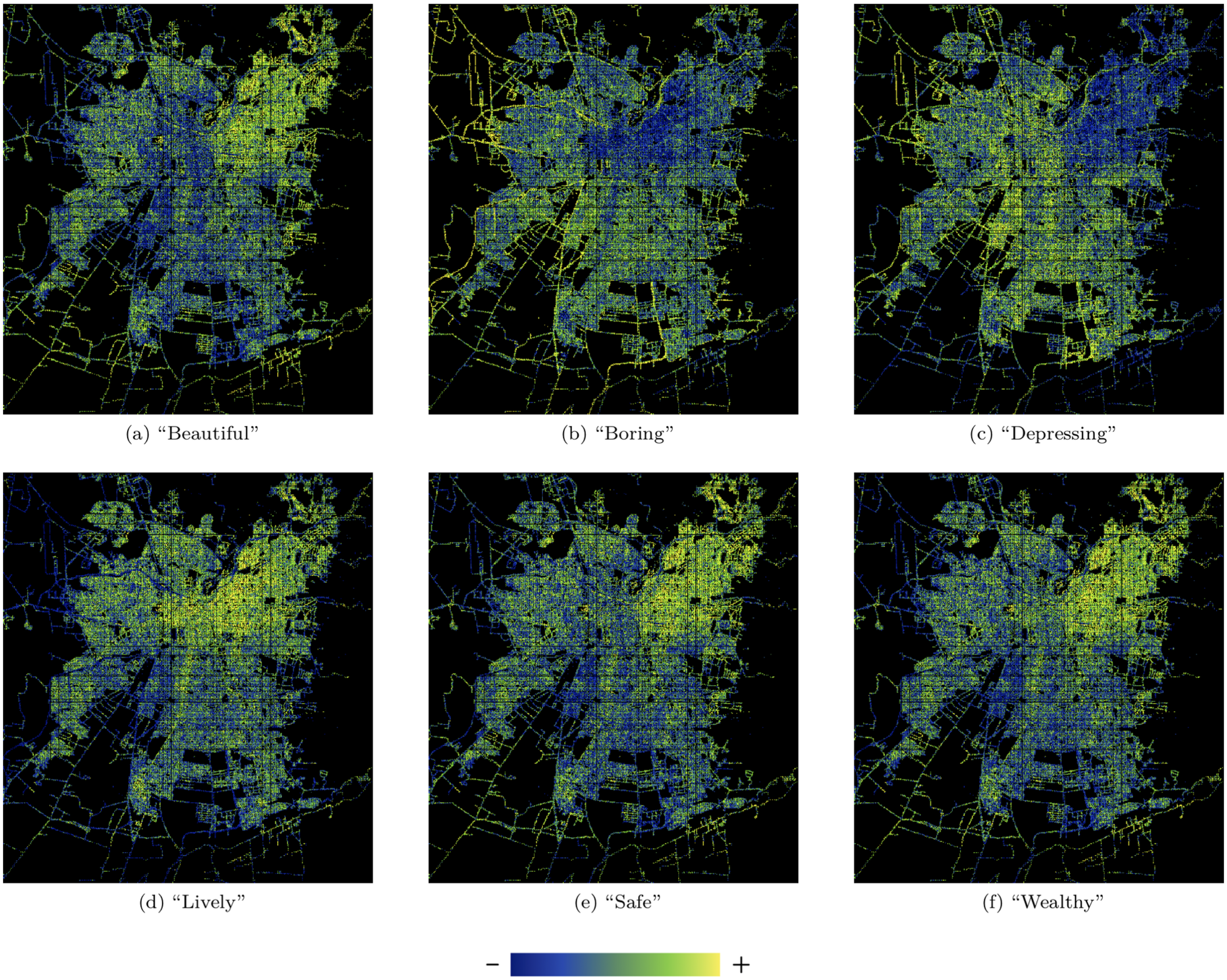


Fig. 6. Scatterplot matrix for the estimated qualitative attributes, with correlation coefficients shown in upper-right corners and density plots shown in the diagonal.



(a) “Beautiful” (b) “Boring” (c) “Depressing”

(d) “Lively” (e) “Safe” (f) “Wealthy”

Fig. 7. Perceptual maps for Santiago, Chile.

Note that the resulting model is analogous to a logit model derived from the Random Utility theory (Domencich and McFadden, 1975; McFadden, 1974). This specific case, with indifference alternatives, is less common in the discrete choice literature, but has been studied earlier in the work of Krishnan (1977) for binary choices and later extended by Lioukas (1984) for multinomial cases. The inclusion of this element allows to consider indifference observations, which represent between 11% and 16% of observed choices in this dataset. This, in turn, induces gains in efficiency (i.e., more precise parameters).

We decided to model the deterministic part of the latent variable (L_i) linearly, as shown in Eq. (8). Here, LLF_{ik} is the k-th low-level feature of scene i, and HLF_{ics} describes the presence of objects of class c in section s of the image, as mentioned in the previous section.

$$L(X_i;\lambda) = \lambda_0 l_i + \sum_k \lambda_k^{LLF} LLF_{ik} + \sum_c \sum_s \lambda_c^{HLF} HLF_{ics} \tag{8}$$

The latent variable does not contain subscript n because respondents’ characteristics were not available. Nevertheless, if this information were available it could be easily accommodated in this framework. While this data could be important to address perceptual heterogeneity, there is evidence that suggests perceptions tend to be universal and do not vary substantially across different groups (Cañas, Ayuga, & Ayuga, 2009; Kearney et al., 2008 Ode, Fry, Tveit, Messager, & Miller, 2009), even in the case of the Place Pulse 1.0 dataset (Salesses et al., 2013).

Parameters λ and δ were estimated through likelihood maximization using Biogeme (Bierlaire, 2003). Because these functions are linear, μ cannot be estimated independently, which is why it was set to be equal to one. The dummy variable l_i was set equal to one if scene i is presented on the left and zero otherwise, and was included to account for possible systematic biases towards scenes presented on the left or right-hand sides. Multicollinearity issues among the models’ input variables (features of two-dimensional images) were discarded after checking the database’s eigenvalues and vectors. Images detected as outliers were omitted from the estimation. Outliers were defined as being at a distance greater than three standard deviations from the average for any low-level feature, or at a distance greater than six standard deviations for more than one high-level feature. A total of 3,729 images corresponded to these cases, which represent 3.37% of the total amount of available images, and were mostly distorted, dark, or indoors scenes, or images that have important errors in the segmentation output.

4. Results

Table 1 shows the result for the discrete choice models calibrated for the six qualitative attributes considered in Place Pulse. They show a significant improvement with respect to models containing only low-level features, and most parameters are highly significant ($p < 0.01$).

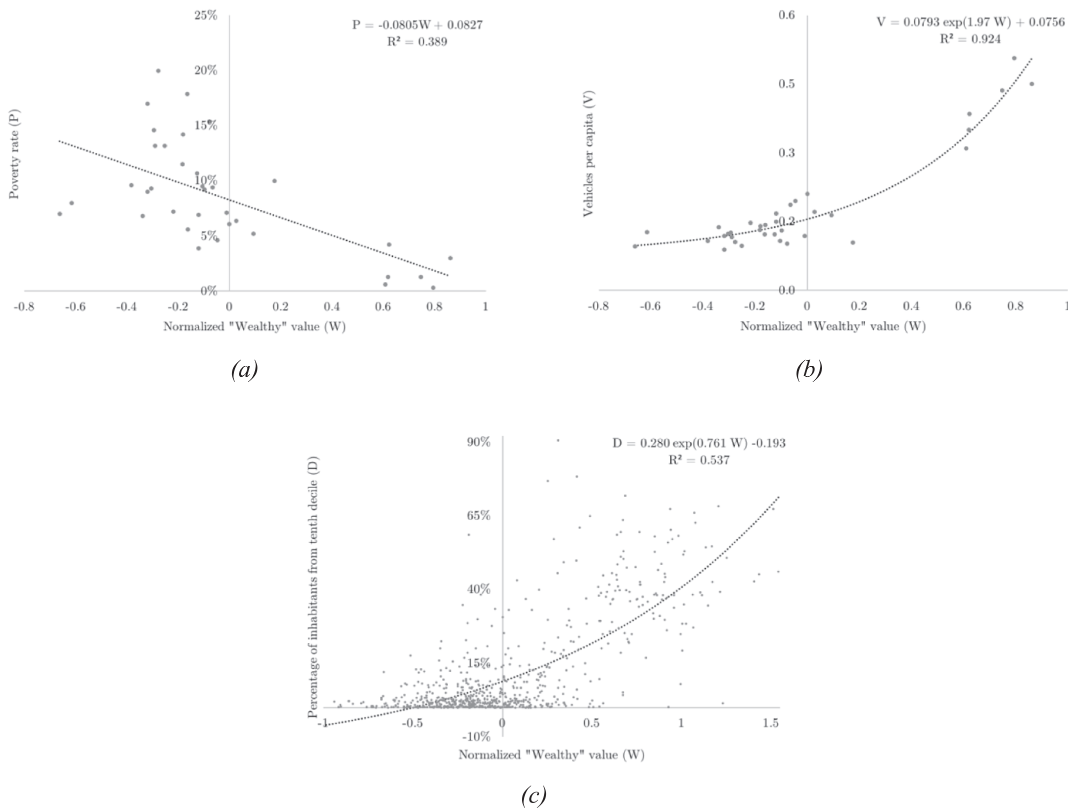


Fig. 8. Relation between predicted perception of wealth and socioeconomic indicators. (a) Poverty rate at the commune level. Data taken from the Chilean government's income survey (Ministerio de Desarrollo Social, 2014). (b) Vehicle ownership at the commune level. Data taken from the Chilean government's income survey (Ministerio de Desarrollo Social, 2014). (c) Percentage of high-income population at the traffic analysis zone level. Data produced by Niehaus et al. (2016) using synthetic populations, based on Santiago's mobility survey (SECTRA, 2015).

Goodness-of-fit measures are relatively lower than the ones obtained by Dubey et al. (2016) using the same dataset, indicating there is relevant information that was not included in the models' specifications. First preference recovery measures (the percentage of tasks where the observed choice corresponds to the one with greatest estimated probability) measures for these models vary between 60% and 65%.

Some interesting takeaways may be obtained from these results:

- Images with buildings were perceived as livelier but less beautiful and safe, and more boring and depressing.
- The presence of cars and pedestrians improved the qualitative attributes of images.
- Images with presence of bicycles and motorcycles, on average, were less boring, less depressing and livelier. Nevertheless, they were also perceived as less safe, less wealthy and less beautiful.
- Fences had an overall positive impact on qualitative attributes of images, which may be related to order and transparency (fences, unlike walls, allow to see through them).
- Images with more sidewalks were deemed to be safer, livelier and wealthier, but less beautiful on average.
- Higher proportions of "Road" pixels are associated to better qualitative attributes.
- Traffic signs had an overall positive effect over perceptions.
- Greenery in the lower halves of images (usually grass or bushes) made images feel more boring and less lively on average, but at the same time more beautiful, safer, wealthier and less depressing. Vegetation in the upper halves of images (usually trees) made places feel safer, livelier and less boring on average.
- Clearer views of the sky had a negative impact on perceptions.
- Edges and blobs had a positive impact on perceptions. Lighter images were considered to be safer, and images with a higher contrast were deemed to be better in terms of the surveyed qualitative attributes.

Fig. 5 shows a box plot of the qualitative attributes measured in the Place Pulse images. Perceived beauty has a higher variance than the rest of the variables, showing there are more diverse settings in terms of aesthetic beauty in the dataset used to estimate the models. On the other hand, perceived safety has the least variance.

Fig. 6 shows that some qualitative attributes, as quantified in this work, are highly correlated between them. This probably happens because these six concepts are not orthogonal to each other. "Lively" and "Boring," for example, can be considered antonyms, which is why it is expected that the quantification of these adjectives have a strong negative correlation.

5. Validation

To check the validity of the perceptions estimated with these models, the images with the greatest and least quantified perceptions were retrieved, shown in Fig. 9. All images were selected systematically: Perceptual variables were calculated with the maximum likelihood estimators shown in Table 1, and then were selected by choosing the six top and six bottom ones that were at a distance of less than 1.5 standard deviations from the mean.

These images allow to interpret results in a different way. For example, more "Beautiful" images have more vegetation and a higher contrast, while less "Beautiful" images are grayer. More "Boring,"

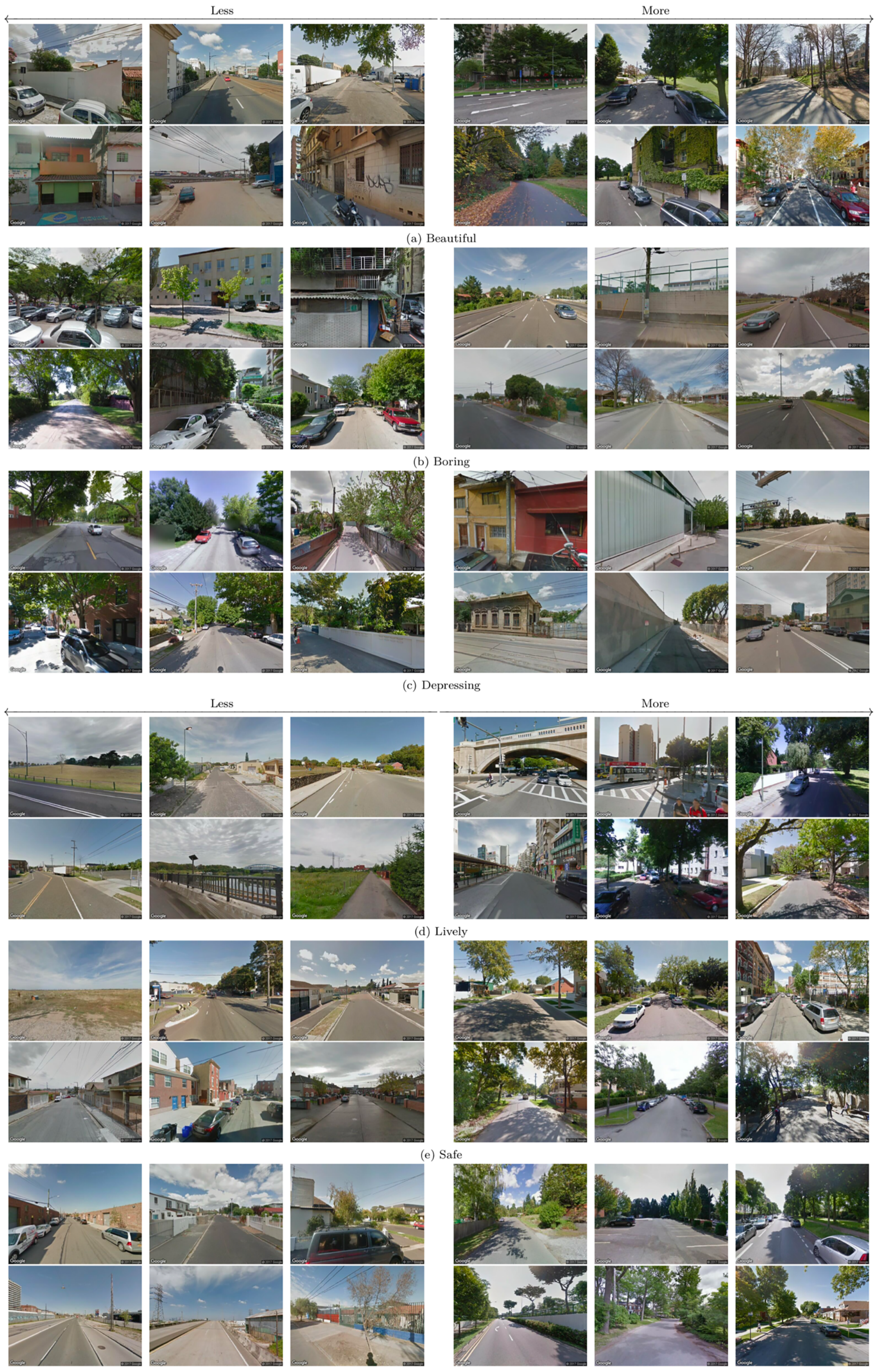


Fig. 9. Images with higher and lower utilities associated with each qualitative attribute surveyed.

“Lively” and “Safe” images are darker and have less cars and vegetation. Less “Depressing” images have more human activity and bluer skies and less buildings, meaning they are generally more open. More “Wealthy” images have less presence of skies and more vegetation, especially lawns and bushes.

To further check if these estimated perceptions correspond to real perceptions or actual data, perceptual maps were created for the city of Santiago, Chile. To do this, 126,286 images of Santiago, covering its whole surface, were downloaded from Google Street View and parametrized. Then, the quantified perceptions were calculated for each image using the parameters shown in Table 1, and then plotted geographically and color-coded. Results are shown in Fig. 7.

Santiago is a very segregated city (Tapia, 2011; Tiznado-Aitken, Muñoz, & Hurtubia, 2018). While most high-income households live in its northeastern area, middle and low-income households cover the rest of the city. This can be clearly seen in the perceptual maps, where higher-income municipalities have a clearly different coloring than the rest. The city’s central business district is stretched out between its historic city center and some areas of its northeastern area. Other attractors of activities are two main boulevards that go from the city center towards the southern areas (Vicuña Mackenna and Gran Avenida). These areas can also be noted in the “Lively” map. The city has some urban highways that are clearly marked in the “Boring” map.

To verify if there is any correspondence between these estimated perceptions and the actual characteristics of the city, perceived wealth was compared to socioeconomic indicators. This perceptual variable was selected because of its direct relation with socioeconomic characteristics of the city. Fig. 8 shows there are significant relations between these indicators and perceived wealth.

All of the above suggests that the perceptual models presented in this work are able to quantify real landscape-perception relations with some degree of accuracy, and that some of these have significant relations with actual characteristics of urban areas. Further work is needed to robustly validate these variables.

6. Conclusions

This work proposes a novel methodology to quantify qualitative attributes of public spaces using images at a massive scale. This is done complementing the advantages of discrete choice models with a semantic segmentation model, based on machine learning techniques. The resulting models show greater explanatory capabilities when compared to models solely based on machine learning techniques found in the literature, with little sacrifice in predictive power. Furthermore, results are generally consistent with previous findings in the literature regarding the effect of landscapes’ elements on individuals’ perceptions (see Section 2).

The results obtained were used to map modeled perceptions in Santiago, Chile, obtaining qualitative attributes of public spaces that are highly correlated with measurable socioeconomic spatial variables. Although this correlation is tested for only one attribute, and further validation is required for the others, this result suggests that the proposed method can capture essential characteristics of public spaces. This, in turn, can be used to inform public policy and urban planning. Moreover, these modeled qualitative attributes could be used as explanatory variables in models dealing with other aspects of urban phenomena, such as location choice and route choice, intensity of use of public spaces, and walkability, among others. The results obtained could also be used to inform some urban policies, such as increasing urban trees and sidewalks. Nevertheless, we believe this work’s most useful aspect to public policy is its methodological proposal. It allows local governments to map perceptions, identify areas that need improvements, and even identify features that may be lacking in these areas with a low cost and freely available method.

We believe these models show great potential and open an interesting research path that can help researchers and decision-makers better understand the way people perceive and process the environment that surrounds them.

Acknowledgements

This research was partially funded by the Complex Engineering Systems Institute, ISCI (ICM: P-05-004-F, CONICYT:FBO16), the Centre for Sustainable Urban Development, CEDEUS (CONICYT/FONDAP 15110020), and the BRT + Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF).

References

Adler, P., Falk, C., Friedler, S. A., Rybeck, G., Scheidegger, C., Smith, B., & Venkatasubramanian, S. (2016). Auditing black-box models for indirect influence. *2016 IEEE 16th international conference on data mining (ICDM)* (pp. 1–10). IEEE.

Antonakos, C. L. (1995). Environmental and travel preferences of cyclists. *Transprtation Research Part A: Policy and Practice, 29*(1), 85.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). SegNet: a deep convolutional encoder-decoder architecture for image segmentation. pp. 1–14.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. *Swiss transport research conference* (pp. 1–27). Switzerland: Ascona.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605–634.

Bonaiuto, M., Fornara, F., & Bonnes, M. (2003). Indexes of perceived residential environment quality and neighbourhood attachment in urban environments: A confirmation study on the city of Rome. *Landscape and Urban Planning, 65*(1–2), 41–52.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219.

Brostow, G. J., Fauqueur, J., & Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters, 30*(2), 88–97.

Cañas, I., Ayuga, E., & Ayuga, F. (2009). A contribution to the assessment of scenic quality of landscapes based on preferences expressed by the public. *Land Use Policy, 26*(4), 1173–1181.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-, 8*(6), 679–698.

Cetintahra, G. E., & Cubukcu, E. (2015). The influence of environmental aesthetics on economic value of housing: An empirical research on virtual environments. *Journal of Housing and the Built Environment, 30*(2), 331–340.

Clifton, K., Ewing, R., Knaap, G., & Song, Y. (2008). Quantitative analysis of urban form: A multidisciplinary review. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability, 1*(1), 17–45.

Crystal, J. H., & Brush, R. O. (1978). Measuring scenic quality at the urban fringe. *Landscape Research, 3*(3), 9–14.

de la Fuente de Val, G., Atauri, J. A., & de Lucio, J. V. (2006). Relationship between landscape visual attributes and spatial pattern indices: A test study in Mediterranean-climate landscapes. *Landscape and Urban Planning, 77*(4), 393–407.

Domencich, T., & McFadden, D. (1975). *Urban travel demand: A behavioral analysis.* Amsterdam: North-Holland Publishing Company.

Dramstad, W. E., Tveit, M. S., Fjellstad, W. J., & Fry, G. L. (2006). Relationships between visual landscape preferences and map-based indicators of landscape structure. *Landscape and Urban Planning, 78*(4), 465–474.

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9905*, 196–212.

Foster, S., Hooper, P., Knuiman, M., Bull, F., & Giles-Corti, B. (2016). Are liveable neighbourhoods safer neighbourhoods? Testing the rhetoric on new urbanism and safety from crime in Perth, Western Australia. *Social Science and Medicine, 164*, 150–157.

Frank, L. D., Sallis, J. F., Saelens, B. E., Leary, L., Cain, K., Conway, T. L., & Hess, P. M. (2010). The development of a walkability index: Application to the Neighborhood Quality of Life Study. *British Journal of Sports Medicine, 44*(13), 924–933.

Frank, S., Fürst, C., Koschke, L., Witt, A., & Makeschin, F. (2013). Assessment of landscape aesthetics - Validation of a landscape metrics-based assessment by visual estimation of the scenic beauty. *Ecological Indicators, 32*, 222–231.

Fulkerson, B., Vedaldi, A., & Soatto, S. (2009). Class segmentation and object localization with super pixel neighborhoods. In 2009 IEEE 12th International Conference on Computer Vision, pp. 670–677.

Gilderbloom, J. I., Riggs, W. W., & Meares, W. L. (2015). Does walkability matter? An examination of walkability’s impact on housing values, foreclosures and crime. *Cities, 42*(PA), 13–24.

Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological, 37*(1), 1–25.

Huang, S.-C. L. (2009). The validity of visual surrogates for representing waterscapes. *Landscape Research, 34*(3), 323–335.

Hunt, J. D., & Abraham, J. E. (2007). Influences on bicycle use. *Transportation, 34*(4), 453–470.

Hyodo, T., Suzuki, N., & Takahashi, K. (2000). Modeling of bicycle route and destination choice behavior for bicycle road network plan. *Transportation Research Record: Journal of the Transportation Research Board, 1705*(1), 70–76.

Jiang, Y., Christopher Zegras, P., & Mehndiratta, S. (2012). Walk the line: Station context, corridor type and bus rapid transit walk access in Jinan, China. *Journal of Transport Geography, 20*(1), 1–14.

Kearney, A. R., Bradley, G. A., Petrich, C. H., Kaplan, R., Kaplan, S., & Simpson-Colebank, D. (2008). Public perception as support for scenic quality regulation in a nationally treasured landscape. *Landscape and Urban Planning, 87*(2), 117–128.

Khisty, C. (1994). Evaluation of pedestrian facilities: Beyond the level-of-service concept. *Transportation Research Record, 1438*, 45–50.

Krishnan, K. S. (1977). Incorporating thresholds of indifference in probabilistic choice models. *Management Science, 23*(11), 1224–1233.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems,* 1–9.

Kuo, F. E., Bacaicoa, M., & Sullivan, W. C. (1998). Transforming inner city landscapes: trees, sense of safety, and preference.

Latkin, C. A., & Curry, A. D. (2003). Stressful neighborhoods and depression: A prospective study of the impact of neighborhood disorder. *Journal of Health and Social Behavior, 44*(1), 34–44.

Lee, L. F. (1982). Specification error in multinomial logit models. Analysis of the omitted variable bias. *Journal of Econometrics, 20*(2), 197–209.

Lioukas, S. K. (1984). Thresholds and transitivity in stochastic consumer choice: A multinomial logit analysis. *Source Management Science, 30*(1), 110–122.

Lissner, I., & Urban, P. (2012). Toward a unified color space for perception-based image processing. *IEEE Transactions on Image Processing, 21*(3), 1153–1168.

Liu, L., Silva, E. A., Wu, C., & Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems, 65*, 113–125.

Lynch, K. (1960). *The image of the city.* MIT Press.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in econometrics* (pp. 105–142). Academic Press.

Mertens, L., Van Dyck, D., Ghekiere, A., De Bourdeaudhuij, I., Deforche, B., Weghe, N. V. D., & Van Cauwenberg, J. (2016). Which environmental factors most strongly influence a street's appeal for bicycle transport among adults? A conjoint study using manipulated photographs. *International Journal of Health Geographics, 15*(1), 31.

Ministerio de Desarrollo Social (2014). Resultados Encuesta Casen 2013.

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore – predicting the perceived safety of one million streetscapes. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops,* 793–799.

Niehaus, M., Galilea, P., & Hurtubia, R. (2016). Accessibility and equity: An approach for wider transport project assessment in Chile. *Research in Transportation Economics, 59*, 412–422.

Ode, Å., Fry, G., Tveit, M. S., Messager, P., & Miller, D. (2009). Indicators of perceived naturalness as drivers of landscape preference. *Journal of Environmental Management, 90*(1), 375–383.

Ordonez, V., & Berg, T. L. (2014). Learning high-level judgments of urban perception. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8694 LNCS*(PART 6), 494–510.

Porzi, L., Buló, S. R., Lepri, B., & Ricci, E. (2015). Predicting and Understanding Urban Perception with Convolutional Neural Networks. In Proceedings of the ACM International Conference on Multimedia (ACM-MM), pp. 139–148.

Quercia, D., O'Hare, N., & Cramer, H. (2014). Aesthetic Capital: What Makes London Look Beautiful, Quiet, and Happy? In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, pp. 945–955.

Ruud, P. A. (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice Models. *Econometrica, 51*(1), 225–228.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: mapping the inequality of urban perception. *PLoS ONE, 8*(7).

Schirpke, U., Tasser, E., & Tappeiner, U. (2013). Predicting scenic beauty of mountain regions. *Landscape and Urban Planning, 111*(1), 1–12.

SECTRA (2015). Encuesta origen destino de viajes 2012. Technical report, Ministerio de Transportes y Telecomunicaciones, Santiago, Chile.

Sheets, V. L., & Manzer, C. D. (1991). Affect, cognition and urban vegetation: Some effects of adding trees along city streets. *Environment and Behavior, 23*(3), 285–304.

Shriver, K. (1997). Influence of environmental design on pedestrian travel behavior in four Austin neighborhoods. *Transportation Research Record: Journal of the Transportation Research Board, 1578*(1), 64–75.

Svobodova, K., Vojar, J., Sklenicka, P., & Filova, L. (2017). Presentation matters: Causes of differences in preferences for agricultural landscapes displayed via photographs and videos. *Space and Culture,* 1–15.

Tapia, R. (2011). Vivienda social en Santiago de Chile. Análisis de su comportamiento locacional, período 1980–2002. *INVI, 26*(73), 105–131.

Tilahun, N., & Li, M. (2015). Walking access to transit stations: Evaluating barriers using stated preference. *Transportation Research Record, 2534*, 16–23.

Tiznado-Aitken, I., Muñoz, J. C., & Hurtubia, R. (2018). The role of accessibility to public transport and quality of walking environment on urban equity: The case of santiago de Chile. *Transportation Research Record: Journal of the Transportation Research, Board* page 036119811878203.

Van Dyck, D., Cardon, G., Deforche, B., & De Bourdeaudhuij, I. (2011). Do adults like living in high-walkable neighborhoods? Associations of walkability parameters with neighborhood satisfaction and possible mediators. *Health and Place, 17*(4), 971–977.

Walker, J., & Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences, 43*(3), 303–343.

Westland, S. (2016). Color communication. In J. Chen, W. Cranton, & M. Fihn (Eds.). *Handbook of visual display technology* (pp. 153–160). Sheffield, United Kingdom: Springer International Publishing 2nd ed.

Zacharias, J. (2001). Pedestrian behavior and perception in urban walking environments. *Journal of Planning Literature, 16*(3), 3–18.