Aesthetic Capital: What Makes London Look Beautiful, Quiet, and Happy?

Daniele Quercia
Yahoo Labs, Barcelona, Spain
dquercia@acm.org

Neil O'Hare
Yahoo Labs, Barcelona, Spain
nohare@yahoo-inc.com

Henriette Cramer
Yahoo Labs, Sunnyvale, USA
henriette@yahoo-inc.com

ABSTRACT
In the 1960s, Lynch's "The Image of the City" explored what impression US city neighborhoods left on its inhabitants. The scale of urban perception studies until recently was considerably constrained by the limited number of study participants. We here present a crowdsourcing project that aims to investigate, at scale, which visual aspects of London neighborhoods make them appear beautiful, quiet, and/or happy. We collect votes from over 3.3K individuals and translate them into quantitative measures of urban perception. In so doing, we quantify each neighborhood's aesthetic capital. By then using state-of-the-art image processing techniques, we determine visual cues that may cause a street to be perceived as being beautiful, quiet, or happy. We identify effects of color, texture and visual words. For example, the amount of greenery is the most positively associated visual cue with each of three qualities; by contrast, broad streets, fortress-like buildings, and council houses tend to be associated with the opposite qualities (ugly, noisy, and unhappy).

Author Keywords
urban informatics; crowdsourcing; aesthetics; visual, urban aesthetics; cities; streets.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

General Terms
Human Factors; Design; Measurement.

INTRODUCTION
Appearances matter in people's perceptions and affective experience of their urban environments. Visual features can distinguish one locale from another, and help people perceive a city as unique and recognizable. Doersch *et al.* [4], for example, identify visual elements such as windows, balconies, and street signs that distinguish Paris or London from other cities (e.g., street signs in Paris, Victorian houses in London). The visual qualities of urban facades not only affect aesthetic responses and people's judgments of urban locale [34], but also have stark social and psychological effects on their inhabitants [14].

One famous theory of urban appearance and social effects was articulated by Kelling & Coles [12]. Their theory of 'broken windows' posits that cues of disorder in public are highly visible and constitute a salient marker of urban spaces. The idea behind the theory is that neighborhood *appearances* drive the reality of neighborhood *safety*: one broken window leads to another broken window and, in turn, to future crime. This theory was not confined to academic circles but made its way into public policy, starting from the city of New York. However, basing public policy on stark notions of aesthetics and cracking down on minor transgressions has notable social and cultural downsides [27]. Nevertheless, since the importance of visual features in perceptions of the built environment has been given a major role in public policy, understanding their effects is key.

To explore visual assessments at scale, we make a number of contributions. We have built a crowdsourcing website under `UrbanGems.org` to collect ratings about how beautiful, quiet, and happy London's streets are. It picks up two random locations from Google Street View and Geograph[1], and ask users which one is more beautiful, quiet, or happy: we choose those three qualities as proxies for urban aspects that have been consistently discussed in 1960s urban studies. We have collected subjective aesthetic judgments of different parts of the city from more than 3.3K users. The goal of this paper is to focus on a specific aspect of this crowdsourcing effort; that is, whether it is possible to *automatically* extract aesthetically-informative features of annotated city scenes. To this end, we apply three state-of-the-art computer vision techniques on each urban scene, and identified effects of color, texture and a number of 'visual words'. We find, for example, that nature-made elements (generally characterized by diagonal and smooth lines) tend to be associated with happy scenes. By contrast, man-made elements (generally characterized by horizontal and vertical lines) tend to be associated with ugly scenes. We also extract visual patches (called 'visual words') from our urban scenes and find that visual words in happy scenes are associated with Victorian houses, public gardens, red bricks, and residential trees, while those in unhappy scenes are associated with council housing and highway road signs. After this visual analysis, three architects are interviewed and asked to reflect on the findings

[1] `http://www.geograph.org.uk`: geographically representative photographs of the whole Great Britain and Ireland.

Preprint of a paper that will appear in *CSCW'14*

presented here to gain a perspective from practicing domain experts on the applicability and (dis)advantages of our approach.

Our analysis is an exploration of how various techniques could assist the design of urban environments by, for example, identifying features with positive or negative effects. It might also result in practical applications in the city context such as affective route recommendation [24, 23], and automatic aesthetic profiling of streets and neighborhoods - while also considering the limitations of such automatic approaches.

RELATED WORK

Beyond Lynch's [15] classic notion of legibility and the ease with which a city can be 'read' in terms of layout, aesthetics are of crucial importance in people's evaluation of their surroundings [19, 32, 34]. Various urban planning and urban design researchers have established criteria that affect people's preferences for certain environments, and their aesthetic responses [34, 18].

The perception of 'urban disorder' from trash, abandoned property, and decay, for example, has been taken as a signal of a breakdown of the local social order [30]. Visual qualities can also affect people's mental states in a physical environment. Lindal and Hartig [14] investigated the effects of architectural variation and physical building attributes (e.g., height) on judgments of the (imagined) psychologically restorative quality of a street scene. Even when considering the inherent social and cultural biases in what is considered to be aesthetically pleasing, researchers have found a persistent correlation of perceived disorder with individual-level affective outcomes, with psychological distress [25], and with perceived powerlessness [6].

Authors such as Weber [34] identified factors such as vegetation, uniformity in style, scale, and symmetry as primary factors in aesthetic judgments. Nasar [19] proposed aesthetic programming could help develop guidelines for visual qualities to gain specific factions from different populations in particular contexts. They find that, to optimize for 'pleasantness', moderate complexity, references to historical elements, removing artificial nuisances, and ordering elements would all be beneficial. To increase interestingness and excitement, more natural materials, higher complexity and atypicality, as well as ordering and familiar elements would help.

So far the most detailed studies of perceptions of urban environments and their visual appearance have, however, relied on personal interviews and observation of city streets: for example, some researchers relied on annotations of video recordings by experts [28], while others have used participant ratings of simulated (rather than existing) street scenes [14].

Using the web to survey a large number of individuals has been recently done. Place Pulse is a website that asks binary perception questions (e.g., "Which place looks safer?") across a large number of geo-tagged images [26]. The goal is to create quantitative measures of urban perception, and results for the city of Boston and New York have been made available[2]. The authors do not focus on visual features but rather on relating socio-demographic factors to user ratings. Quercia *et al.* proposed a web game, called `UrbanOpticon.org`, that puts the recognizability of London's streets to the test. It picks up random locations from Google Street View and tests users ability to judge the location (e.g., closest subway station) [22]. By analyzing the data from the site, the authors found that areas with low recognizability significantly suffer from social problems of housing deprivation, poor living conditions, and crime.

CROWDSOURCING URBAN AESTHETICS

The main idea behind `UrbanGems.org` is that players compare a series of urban scenes and vote on which one is more beautiful, quiet or happy. Users compare two side-by-side street views from various neighborhoods around London. They must decide which scene best represents one of three qualities: happy, beauty and quiet. They can also opt for "Can't Tell", if undecided on which picture to click on.

Choosing Rating Qualities

One might well wonder why we chose to study the three qualities of beauty, quiet, and happy. For this exploration, a choice was made to return to classic urban studies, as well as more popular discussions on 'city life'.

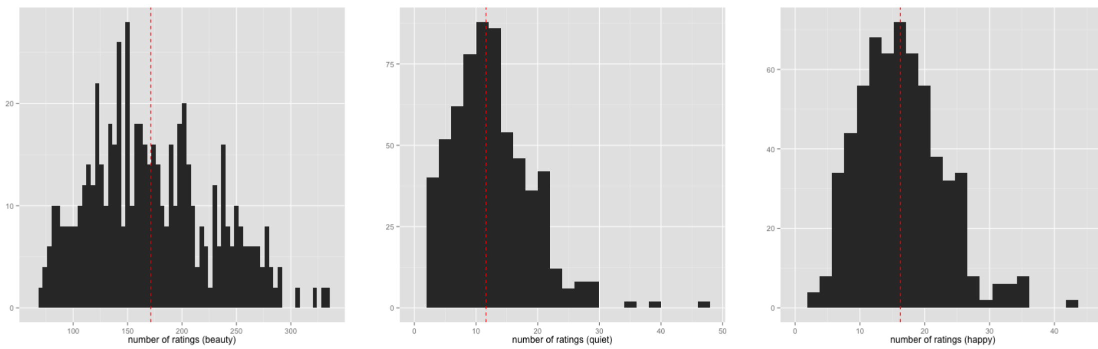
Quiet. In the noise of a city, it might be hard to find quiet places. That is why a variety of mobile applications has been built to discover such places in big cities. The Economist 'Thinking Spaces' application allows the global community of Economist readers to create, share and explore the spaces where they think and get new ideas[3]. This has resulted in interactive maps of major cities around the world features readers' beloved thinking spaces. More recently, sound artist Jason Sweeney proposed a platform where people crowdsource and geo-locate quiet spaces, share them with their social networks, and take audio and visual snapshots. It is called Stereopublic[4] and is "an attempt to both promote 'sonic health' in our cities and offer a public guide for those who crave a retreat from crowds" - both for those in need of quietness and for people with disabilities, like autism and schizophrenia.

Beauty. In *The Death and Life of Great American Cities*, Jane Jacobs [10] offered a critique of 1950s urban planning policy and of modernism. She argued against authoritarian top-down urban planning and separation of uses (i.e., residential, industrial, commercial), while instead emphasizing the importance of mixed functions and human activity. She dedicated an entire chapter on 'Visual Order', in which, amongst other topics, she discusses how streets provide the principal visual scenes in cities, and how 'visual interruptions' can used to create visual order without sacrificing the intensity and diversity that a city's functional order demand. Visual order and harmony are linked to 'beauty', which forms the basis of our second question as it is easy for people to grasp (as opposed to concepts such as 'visual order' or 'aesthetically pleasing') and is thus amenable to an open crowd-sourcing setting. Note

[2] http://pulse.media.mit.edu/results/

[3] `http://thinkingspace.economist.com/`

[4] `http://www.stereopublic.net/`



(a) Beauty (b) Quiet (c) Happy

Figure 1. Number of Answers for Each Urban Scene. The number of answers each scene has received are centered around means thanks to randomization: 171 answers per scene for beauty (default question), 12 for quiet, and 16 for happy.

that we are certainly not the first to measure perceptions of beauty. In 1967, Peterson proposed a quantitative analysis of public perceptions of neighborhood visual appearance [21]. He did so by choosing ten variables that reflected visual appearance (e.g., preference for the scene, greenery, open space, safety, beauty) and having 140 individuals rate 23 pictures of urban scenes in the Chicago metropolitan area for each of those ten variables. He found that beauty and safety are approximately collinear with preference for a scene, suggesting that “beauty of visual appearance is in fact synonymous with perception of visual pleasure and, hence, desirability of visual appearance” [21].

Happiness. Additional urban studies from the 1960s tried to systematically relate well-being in the urban environment (including happiness) to the desire for visual order, beauty, and aesthetics. In *The Image of the City* [15], Kevin Lynch illustrated what elements appeared to make cities more vivid and memorable to city dwellers. More recently, in *The Architecture of Happiness* [3], Alain de Botton dwells on how different architectural styles talk about cultural values. He analyzed how people’s needs and desires manifest their ideals of *beauty* and *happiness* in architecture, considering that there is an intimate relationship between our visual taste (e.g., what we consider quiet and beautiful) and our values (e.g., search for happiness).

While we most certainly do not claim to address the full gamut of observations and proposals of the classic authors above, we here aim to explore whether some prior observations related to visual characteristics of city streets can be reproduced using crowd ratings and visual analysis.

Selecting Scenes

To avoid sparsity problems (too few answers per picture), a random scene is selected within a 300-meter radius from a subway station and within the bounding boxes of census areas. This results in 258 Google Street views and 310 Geograph images whose ratings are not sparse (the number of images is different in the two sets as individual census areas contain more than one Geograph image). To see why we choose those two sources of pictures, consider that, in general, photos might not necessarily show what they are supposed to show (representative urban scenes in each neighborhood), and some pictures might be of better quality than others. For this reason, we did not opt for Flickr images, as people often upload pictures of extremely different levels of quality and character. We instead use two kinds of pictures for which quality is comparable: Google Street View pictures captured by camera-mounted cars, and Geograph pictures provided by volunteers with the goal of mapping the whole Great Britain and Ireland in a crowdsourcing fashion. We use multiple images of the two types at the *same* locations. We find that the visual features associated with our three qualities are almost the same for the two sets of pictures, and that ratings are not correlated with objective measures of image quality - there is no correlation between images’ ratings and their sharpness and contrast levels, both of which have been used as proxies for quality [36].

Setting up the site

To increase the likelihood that people will adopt the platform, we added some simple engagement strategies. Those strategies include giving points, creating a sense of freshness and of purpose. In our platform, with each selection, the user is asked to guess the percentage of other people who shared their view. The player then scores points for correct guesses. After being presented with 10 pairs of scenes, the player has completed one round and can share the resulting score on Facebook or Twitter with a single click. The purpose of the score is to facilitate the player’s assessment of their performance against previous game rounds or against other players [33]. After the first round, each player is also asked to complete a small questionnaire (e.g., age, gender, location). Participants engaging in multiple rounds are identified through browser cookies, which uniquely identify users[5]. Pictures are chosen randomly to create a sense of freshness and increase replay value. In addition, for experimental sake, randomization reduces biases and leads to reliable results, producing a distribution of answers for each picture that is

[5] Unless multiple individuals use the same computer with exactly the same username. This situation, however, represents an unlikely exception.

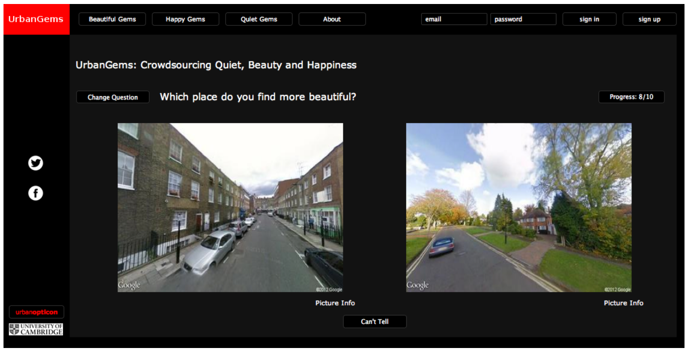


Figure 2. Screenshot of the Crowdsourcing Task. The question "Which place do you find more beautiful?" is displayed on top of the two urban scenes. By clicking on that question, the other two questions, about quietness and happiness, are made available.

Race (%)			
White 92.1	Asian 3.1	Black 0.6	Indian 0.6
Country (%)			
UK 65.1	France 4.5	USA 4.3	Netherlands 3.2
City (%)			
London 40.3	Cambridge 6.5	Manchester 1.6	Amsterdam 1.4
Profession (%)			
Student 15.5	IT Prof 12.3	Scientist 7.5	Architect 6.6

Table 1. Statistics of Participants about Race (census terminology used), Country of Origin, City of Origin, and Profession. This data is available for those 515 participants (at most) who have been willing to provide their personal information.

centered around a typical number of answers. In our analysis, the median is of 171 answers per scene for beauty, 12 for quiet, and 16 for happy (Figure 1). These mean values are different as the default question is that on beauty, and it thus preferentially attracts more votes, while the other two are accessible from a drop-down menu. Interestingly, despite the fact that the ordering has the happiness question last, that question has still attracted more votes than the question on quietness.

To increase the sense of belonging to a community, users could post their scores on Facebook and Twitter after each round. These posts are expected to not only foster a preliminary sense of community but also increase awareness among other social media users. Additionally, users' votes are used to rank the urban scenes, and top-ranked images are shown under three different pages on the site, each of which corresponds to beautiful, quiet, and happy scenes.

Statistics after Launch

After two beta tests, we have made the final version of the platform publicly available and issued a press release in September 2012. Shortly after that, the site was featured in major newspapers and news sites, including BBC News. After 4 months, we collected data from as many as 3,301 participants: 36% connecting from London (IP addresses), 35% from the rest of UK, and 29% outside UK. A fraction of those participants (515) answered a surveyed in which they specified their personal details. The percentage of male-female for those participants is 66%-34%. Their average age was 38 years old (range: 18 - 77 years old). Compared to the 2001 UK census, our sample was fairly representative (Table 1), in that, (as per census terminology[6]) White participants were slightly overrepresented by +6.4%, and participants of Asian descent, Black, Indian, Mixed and Irish were represented in a balanced way. The top country of origin was United Kingdom (65.1%) and the top city of origin was London (40.3%). Professions were quite diverse, the most common being Student, IT Professional, Academic/Scientist, and Architect/Urbanist.

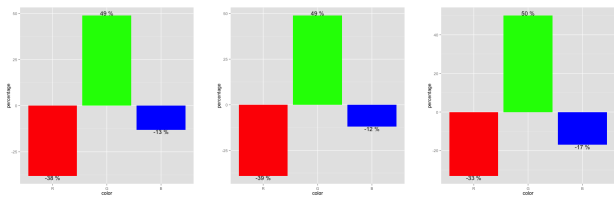
After processing 17,261 rounds of annotation (each of which annotates at most ten pairs of pictures), we rank pictures by their score for beauty, quiet, and happiness based on the fraction of votes they received. Based on those rankings, we are now able to quantify perceptions.

Quantifying urban perceptions has already been done for the concept of imaginability: in 1969, Milgram quantified recognizability (proxy for imaginability) for New York City by running a variety of small-scale experiments that resulted in collective mental maps of the city. More recently, web crowdsourcing has been used to take such an experimentation to a larger scale [22]. Since the data from this crowd-sourcing effort in London is publicly available, we are also able to correlate recognizability scores for 98 of our locations with their scores of beauty, quiet, and happiness. We find that recognizability is correlated with beauty ($r = 0.28$), quiet ($r = -0.35$), and happiness ($r = 0.14$), and those correlations are statistically significant with p-value < 10^{-3}. We are also able to compute the correlations between each pairwise combinations of the three qualities. All correlations are statistically significant (i.e., all p-values are 10^{-4}): happy-quiet has $r = 0.29$, quiet-beauty $r = 0.33$, and beauty-happy is $r = 0.64$. We find that the strongest affiliation happens to be between beauty and happiness. The intimate relationship between the two has been aptly crystallized by Stendhal in his book "On Love": "Beauty is the promise of happiness" [31].

VISUAL ANALYSIS OF SCENES

Stendhal also knew that there is not only one acceptable visual style: "There are as many styles of beauty as there are visions of happiness". In this vein, we should think about our urban environments as visual representation of our values (e.g., pursuit of happiness), that is, as the transubstantiation of our ideals into a material form. Next, we will explore this rendition of values for the city of London by studying what it is about certain neighborhoods that makes them appear to speak of beauty, quiet and, ultimately, happiness. To this end, we will start our analysis by focusing on colors and answering the following question: Which *colors* correlate the most/least with pictures of happy/quiet/beautiful scenes? After this, we analyze the texture of images to see which types of texture correlate with the same qualities. Finally, we do the same for visual words, which represent local, salient, patches in an image.

[6] http://en.wikipedia.org/wiki/Ethnic_groups_in_the_United_Kingdom



(a) Beauty (b) Quiet (c) Happy

Figure 3. The Relative (percentage) Importance of the Three Color Components. Red, green, and blue for scenes considered to be: a) beautiful; b) quiet; and c) happy. While the bars visually suggest the relative importance of the three color components, the actual values are reported in Section "Presence of Colors".

Presence of Colors

We use two methods to analyze the color distribution of images. The first method represents the color of each pixel as an RGB triplet (r, g, b), each component of which can vary from 0% (absent completely) to 100% (maximum value). If all the components are at 0% the result is black; if all are at 100%, the result is the brightest representable white. We then average each of these components separately, giving the average *redness*, *greenness* and *blueness* of the pixels in the image. Although other color spaces (e.g., HSV) are known to better approximate the human visual system, we chose the RGB here since it allows for an intuitive analysis of the amount of red, green and blue in the image.

To test each color component's contribution to the three qualities of beauty, quiet and happiness while controlling for interaction effects between colors, we build a simple linear model in which we represent a scene's rating for any given quality (e.g., beauty) as a linear combination of the i^{th} scene's RGB triplet (r_i, g_i, b_i). That is, the regression is of the form $beauty_i = \alpha + \beta_1 r_i + \beta_2 g_i + \beta_3 b_i + \epsilon_i$, where, say, β_1 reflects the importance of red, and the corresponding relative importance is β_1 normalized by the sum of the absolute values of the three beta coefficients.

We build three linear models, one for each of the three qualities, and we report the relative importance of the triplet (r, g, b) in Figure 3 (all the beta coefficients are statistically significant at least at level 0.001). The relative importance for the triplet (r, g, b) is (-38%, 49%, -13%) for beauty, (-39%, 49%, -12%) for quiet, and (-33%, 50%, -17%) for happiness. Green (second element) is the strongest positively associated color for all the three qualities, while red and, to a lesser extent, blue are slightly negatively associated (i.e., associated with the opposite qualities of ugly, noisy, and unhappy).

The strong association of green with the three qualities is no surprise. In our urban scenes, green is associated with trees, and, more generally with nature [13]. As one cited by Komar and Melamid puts it: "Green's my favorite color - and I'm wearing green. It reminds me of forests, trees, nature, nice things like that - everything opposite of New York City." More generally, the importance of green has been widely discussed in urban studies (see [1] as an example).

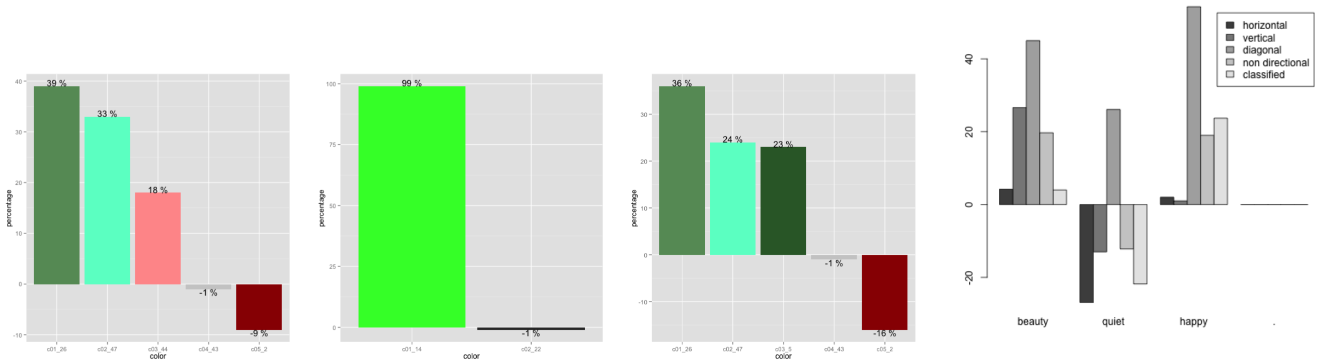
To then go beyond the three primary colors, we create a 64-bin color histogram in the RGB color space by quantizing each of the color components into 4 bins. The histograms count the number of pixels that belong in each of these 64 color bins; we can consider each bin as representing a unique color in a reduced palette. Given the ratings for all pictures for beauty, quiet, and happiness, we calculate the Pearson correlation coefficient between each of the 64 colors and these ratings. To remove spurious correlations, we retain the colors that have correlation coefficient $r > 0.15$, leaving 14 colors associated with beauty, 18 with quiet, and 12 with happiness. We then build a simple linear model in which a scene's rating for any given quality (e.g., beauty) is the linear combination of those remaining colors. Figure 4 reports the β coefficients (and corresponding colors) that are statistically significant in the linear regression. Different greens (reflecting nature) are positively associated with beauty, while gray (reflecting roads) and dark red are negatively associated with it (i.e., associated with the opposite quality - ugliness). For the quality of quiet, green becomes the only color that matters (gray's importance is as low as 1%). For happiness, green is again positively associated, and grey and dark red are negatively associated (are associated with unhappy scenes). These results match the previous ones, in that green and its shades are positively associated with our three qualities.

The main drawback of color histograms is that its representation is dependent on the color of the object being studied, ignoring shape and texture. Color histograms can potentially be identical for two images depicting different objects that happen to share similar color information. In the next section, we look at image texture and answer this question: Which types of *texture* (e.g., smooth vs. sharp lines) correlates the most/least with pictures of happy/quiet/beautiful scenes?

Texture

To represent the texture of images, we follow the approach of Park *et al.* [20] and extract a global edge histogram from the region-based MPEG-7 Edge Histogram descriptor [17]. The image is divided into a number of small blocks, and blocks corresponding to edges (i.e., local areas of high contrast) are detected. These edge blocks are then further classified into four types: vertical, horizontal, 45°and 135°diagonals, and non-directional. We also count the total number of blocks that are classified as edges as a percentage of all blocks. Images with a high number of these 'classified' blocks tend to contain fine-grained details and, as such, tend to be visually rich. We will see to which extent the level of detail in a scene contributes to its beauty, happiness and quietness.

As we have done for color in the previous section, we build a linear regression model in which the dependent variable is beauty (or quiet or happiness) and the predictors are the four types of edges (i.e., horizontal, vertical, diagonal, non-directional) plus the fraction of *classified* blocks as edges. The corresponding (normalized) beta coefficients, shown in Figure 4(d), will measure the relative contribution of a certain edge type to, say, beauty. Beautiful pictures tend to not contain horizontal edges but do contain vertical, diagonal and non-directional edges. Since images depicting buildings and



(a) Beauty (b) Quiet (c) Happy (d) Edge Types

Figure 4. The Relative (percentage) Importance of Colors. The urban scenes considered are: a) beautiful; b) quiet; and c) happy. Out of the 64 colors (our color space), we report only the colors that are correlated with the ratings for the three qualities. (d) Relative importance of different edge types.

man-made structures tend to contain many horizontal and vertical edges [16], it is quite likely that these horizontals corresponds to buildings. The fact that vertical edges contribute more than horizontal ones is somewhat puzzling: perhaps people are less likely to find vertical lines in buildings ugly than they are to find horizontal lines ugly. The positive contributions of diagonal and non-directional edges to beauty and happiness are likely explained by the presence of 'natural' scenes and shapes in those images [16]. Based on the number of classified blocks (as edges), we see that high level of detail does not contribute to beauty as much as it does to happiness. By contrast, for quiet scenes, the 'classified' bar is negative, suggesting that the more detailed a scene, the less quiet it is perceived to be. Again, we find that, as one expects, quiet scenes tend to also be "visually quiet" (i.e., they are smooth and flat).

Visual Words

The color and edge histograms discussed in the previous sections ignore spatial information in an image: they do not give us any hint as to the specific regions within images that correspond to beauty, happiness and quietness. To address this, we now set out to answer this last question: Which *visual elements* (e.g., building styles) correlate the most/least with pictures of happy/quiet/beautiful scenes? To answer this question, we use a popular approach for exploiting local information in images that creates *visual words* [11, 29]. These are *interest points* of an image and typically correspond to a change in intensity (i.e., an edge). These local interest points are represented by a feature vector, which describes "a small patch on the image (array of pixels) which can carry any kind of interesting information in any feature space (color changes, texture changes)" [35]. We use a codebook of visual words from these interest points by applying a clustering algorithm to the set of interest points. After that, an image can be considered as a bag of visual words, similar to bag of words in text retrieval, allowing text retrieval techniques to be applied to image collections.

In this work, we use Speeded Up Robust Features (SURF), a texture based descriptor based on Haar wavelets [2], to detect and describe interest points. We quantize the interest points into a set of 500 visual words using the K-Means algorithm. We represent each urban scene as a histogram of 500 visual words, and we calculate correlations between each visual word and the three qualities.

Since the individual visual words do not have an intuitive interpretation, we highlight them (filtering away those with correlations $r < 0.1$) on example images that rank highly for beauty, happiness and quietness (we do the converse for bottom-ranked images). One should consider that visual words are based not on colors but on texture. Hence, if associated with trees, visual words reflect the trees' texture and do not reflect the presence of green. Despite this, we still find that greenery is positively associated with our three qualities in the analysis of visual words.

For beauty, Figure 5 shows the top-ranked scenes (top row) and bottom-ranked scenes (bottom row). Each scene contains red dots next to the visual words that positively (top row) correlate with beauty or negatively (bottom row) correlate with beauty (i.e., they correlated with the opposite quality - ugly). Visual words affiliated with beauty include those around: Figure 5*(a)* Victorian houses; *(b)* public gardens; *(c)* red bricks; *(d)* residential trees. Those affiliated with low value of beauty include those present on *(e)* council housing; *(f)* bridges; *(g)*, again, council housing; and *(h)* highway road signs and guardrails. The red dots on top-ranked pictures and those on bottom-ranked ones mean two different things - the former reflect positive (e.g., happy) visual words, while the latter reflect negative (e.g., ugly) ones. This is best exemplified by the bottom-ranked picture in Figure 5(e), in which no red dot is present on trees, as opposed to what happens in top-ranked pictures, where trees and, more generally, greenery, receive many red dots.

For quiet, Figure 6 shows the top-ranked scenes (top row) and bottom-ranked scenes (bottom row). Visual words affiliated with quiet include those on *(a,c,d)* trees and hedgerows; and *(b)* residential windows. Those affiliated with low values of quiet include those on *(e)* construction sites and public buses;

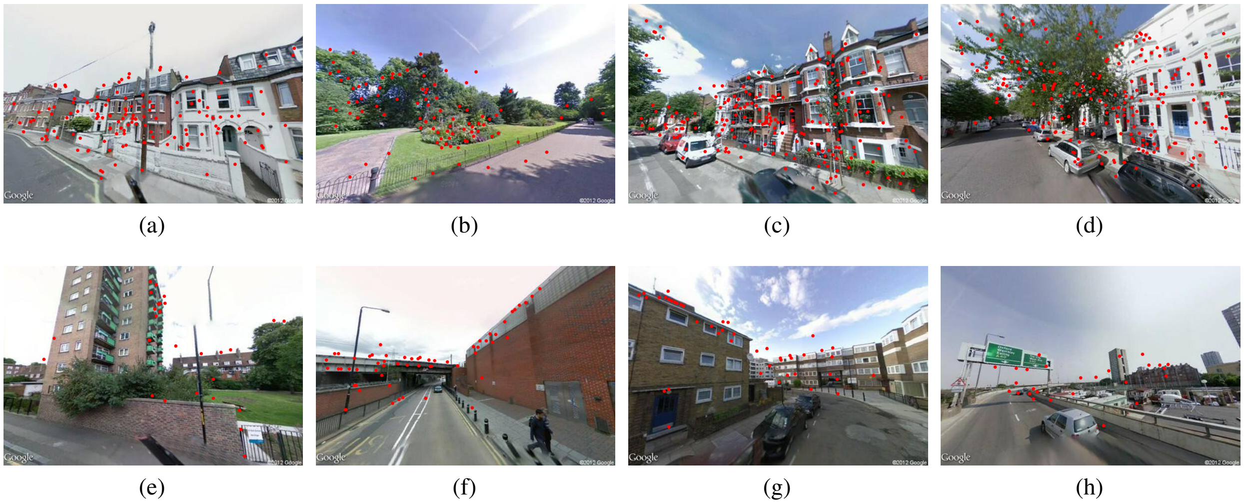


Figure 5. Visual Words for Beauty (top row) and Ugly (bottom row).

(f) council housing; *(g)* typical architecture of central London buildings; and *(f)* guardrails. Scenes considered to be quiet are characterized not only by fewer salient edges (as we have seen in the previous section, e.g., in Figure 4(d)) but also by fewer salient visual words. We might say that quiet scenes are also quiet (smooth) in the visual sense. Beauty and quiet do not go always hand in hand, however. Buildings with texture similar to that of the building in Figure 6(g) are generally associated with noisy scenes (because they are in central London) yet are also considered beautiful.

Finally, Figure 7 shows the top-ranked scenes (top row) and bottom-ranked scenes (bottom row) for happiness. Visual words affiliated with happiness include those on *(a)* trees and buses; *(b)* trees; *(c)* trees and London letterbox; and *(d)* people. Less happy visual words include those on *(e)* street billboard; *(f)* construction site; *(g)* chain link fence; and *(h)* bridge.

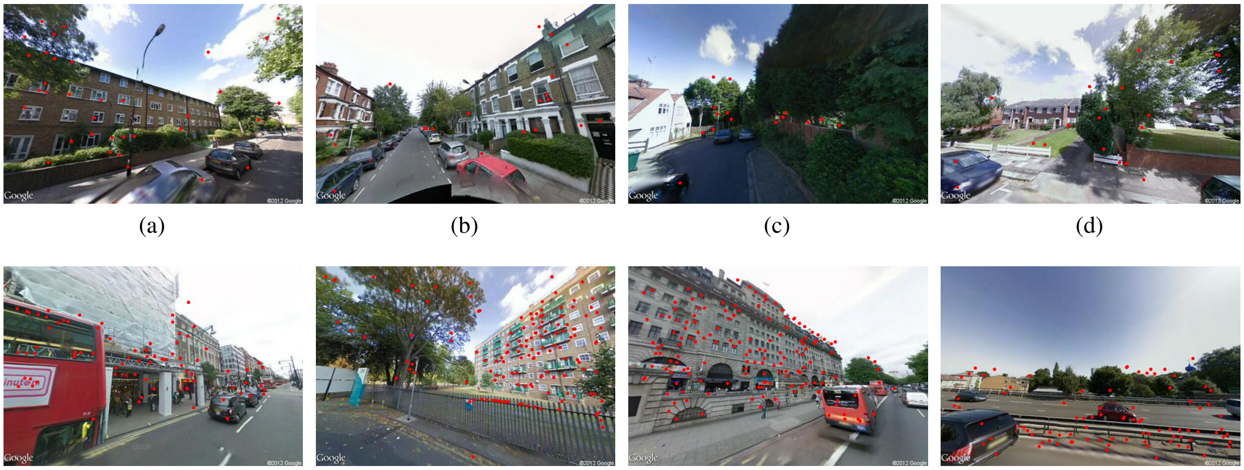
PRACTITIONER REACTIONS

To gain a perspective from practitioners in the realm of the built environment as a starting point for discussion, we asked three architects (2 males, 1 female, NY and Barcelona-based) to reflect on the work presented here, and the role of 'visual words'. They were interviewed in person or over Skype. First, they were given a basic overview of the project and were asked to, according to their own preference, either draw or describe a 'happy scene in London'. Then, they were primed with 10 red-dotted street views indicating visual words (such as those in Figure 5) that were highly ranked on happiness. They then were asked to either redraw their scene or explain how they would change it. They were asked to which extent relationships between visual features and people's reactions currently play a role in their work, and how visual words such as those identified here could help, or perhaps hinder.

All three indicated that such visual words could be useful in providing insight in features affecting people's reactions. Asked which features stood out, they specifically noted the markings related to the importance of green spaces (well-known in their field); the positive effect of pedestrians and cyclists (one architect remarked that the presence of people humanized London and reminded him of Jane Jacobs' work, noting her emphasis on the presence and activity of people, without referencing a specific scholarly work); suggestions of the variety of color, type of materials and features (e.g., traditional middle-class housing material) associated with happiness & beauty; the appearance of mostly private buildings, buildings leaving more space for walkway, and buildings whose facades included windows and ornaments; and features that influence familiarity and comfort (such as common 'artisanal' lamp posts) or create the 'London identity' with typical red letterboxes, red buses, and red telephone cabins.

They also remarked on potential business or investment opportunities. Knowing what makes people - customers - happy, and then designing for such 'happiness' could increase profits. It could guide investment in urban landscaping by business and property owners. Ratings would indicate appeal to a greater population, and could steer the adding of value to space in a certain way, potentially having consequences for real estate development.

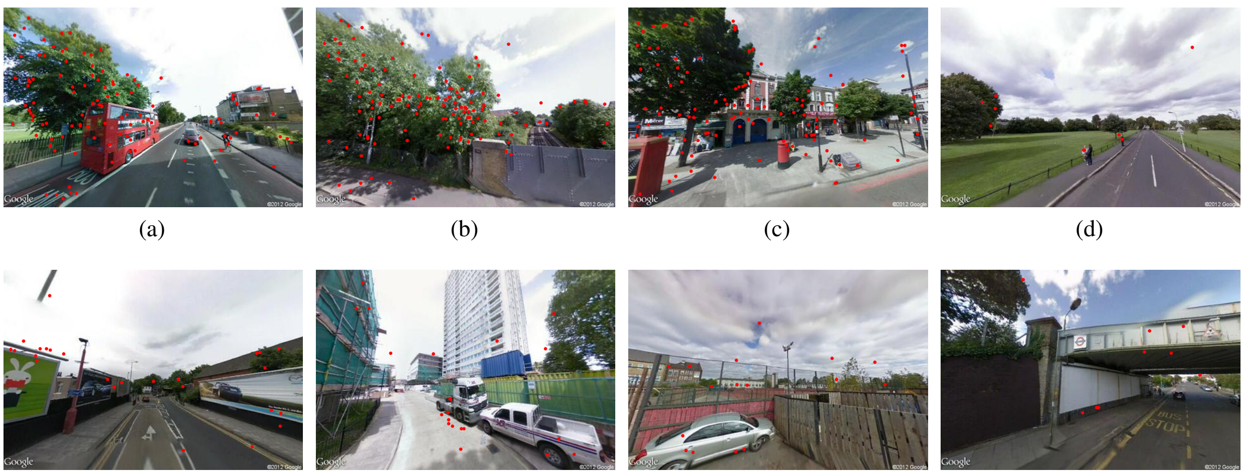
Limitations were also noted. The composition of features, for example, plays a role in how urban facades are rated. One noted that visual features would be useful to know, but that she would not rely on them as they might not necessarily reflect the experience of 'being there'. Additional drawbacks of visual words from crowdsourced ratings include that they tend to reflect traditional styles of architecture and, as such, reflect a 'democratic' view of English beauty, which may not necessarily match with modernist or more forward-looking architecture. Similarly to critique on 'democratization of art' [13], it was noted that democratizing architecture may have considerable drawbacks. When asked for a reaction, the third practitioner noted that algorithms might start to pigeonhole certain types of buildings, yet algorithms might also be an interesting reference, opening the floor for discourse.



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

(e) (f) (g) (h)

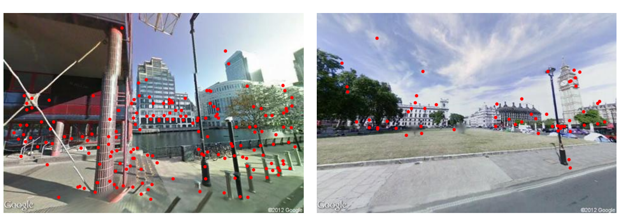
Figure 6. Visual Words for Quiet (top row) and Noisy (bottom row).



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

(e) (f) (g) (h)

Figure 7. Visual Words for Happy (top row) and Sad (bottom row).



(a) (b)

Figure 8. Visual Words for Tall Buildings in Happy Scenes.

PUTTING RESULTS INTO CONTEXT

To rephrase our findings in the broader context of the literature, we consider three simple elements that have been found to be universally present in most urban environments: tall buildings, cars, and trees.

Tall Buildings

One architectural element that has been widely discussed in the literature is tall buildings:

Four-story Limit (`Pattern 21` in A Pattern Language [1]) "There is abundant evidence to show that high buildings make people crazy". Therefore, the authors of A Pattern Language suggested that: "In any urban area, no matter how dense, keep the majority of buildings four stories high or less. It is possible that certain buildings should exceed this limit, but they should never be buildings for human habitation."

Connected Buildings (`Pattern 108`) "Isolated buildings are symptoms of a disconnected sick society. "

In Chapter *Salvaging Projects* [10], Jane Jacobs wrote: "All housing projects with tall buildings are especially handicapped in supervision of children . . . The corridors of the usual high-rise, low-income housing building are like corridors in a bad dream: creepy lit, narrow, smelly, blind."

We have indeed found that tall *residential* buildings are negatively associated with beauty and happiness. For example, the visual words in Figure 5(e) and Figure 5(h) are associated with the opposite qualities, that is, they are associated with ugly and unhappy scenes. Two specific types of tall buildings are exceptions to this general rule: tall *office* buildings (e.g., Figure 8(a)) and landmarks (e.g., Figure 8(b)). The former type is generally represented by glass buildings of the " international skyscraper style": in London, it is common to see ancient bricks contrasting thrillingly with soaring verticals of glass and steel. These buildings are considered to be beautiful by our participants, and the corresponding visual words tend to occur on windows and, more generally, on glass materials. The latter type of tall buildings (i.e., landmarks) and their importance have been discussed at length in the literature. A case in point is the paragraph in Chapter *Visual Order* [10] in which Jacobs noted: "Landmarks, as their name says, are prime orientation clues. But good landmarks in cities also perform two other services in clarifying the order of cities. First, they emphasize (and also dignify) the diversity of cities; they do this by calling attention to the fact that they are different from their neighbors, and important because they are different. This explicit statement about themselves carries an implicit statement about the composition and order of cities. Second, in certain instances landmarks can make functional fact but need to have that fact visually acknowledged and dignified. "

Cars

The second element frequently discussed in the literature is that of cars; more generally, that of transportation systems.

Local Transport Areas (`Pattern 11`) "Cars give people wonderful freedom and increase their opportunities. But they also destroy the environment, to an extent so drastic that they kill all social life."

Ring Roads (`Pattern 17`) "It is not possible to avoid the need for high speed roads in modern society; but it is essential to place them and build them in such a way they they do not destroy communities or countryside"

In Chapter *Lost areas* [15], Kevin Lynch added: "Many [Los Angeles] subjects had difficulty in making a mental connection between the fast highway and the remainder of the city structure . . . A high-speed artery may not necessarily be the best way of visually delimiting a central district"

The presence of moving cars and buses is indeed *negatively* associated not only with quiet scenes (as one would expect and, for example, sees in the pictures at the bottom of Figure 6) but also with beauty (e.g., Figure 5(h)). By contrast, despite the presence of cars, some scenes are still considered to be quiet (pictures at the top of Figure 6).

Trees

Among all colors, green enjoyed the strongest associations with all the three qualities. Green emerged not only in the color histogram analysis but also in the visual word analysis, which has nothing to do with colors but is based on image texture. Visual words in happy scenes tend to occur on trees (e.g., pictures at the top of Figure 7), and that, again, relates to what the literature would suggest:

Tree Places (`Pattern 171`) " . . . shape the nearby buildings in response to trees, so that trees themselves, and the trees and buildings together, form places which people can use."

Again, in the chapter *Visual Order* [10], Jacobs observed that: "There are some city streets which need unifying devices, to suggest that the street, with all its diversity, is also an entity. . . . One of the simplest such devices is trees along the stretch to be unified . . . "

The presence of those three elements in our findings speaks to their external validity and potential generalizability.

DISCUSSION

Beyond the visual and 'pleasant'. Cities are not just collections of buildings and views. Recreating a visual 'identical copy' of a city would not 'feel' like the original. Even with all the visually pleasant elements in place, it it not the same. Aesthetics extend beyond the visual; experiencing a city is not about seeing singular viewpoints, or looking at buildings, it is about moving through the city, experiencing it with all of the senses. This entails moving through a city [8] and developing judgments over time through familiarity and social changes [19]. The potential of cities lies in the 'life between buildings' [5] and their role as meeting places of people [7]. Perceptions about whether this potential is being met are not just about in-the moment sights, sounds and smells; memories, culture, history and all human activity could not be captured by simply analyzing visual properties. Platforms such as ours should be enriched to investigate additional factors at play such as stylistic references, personality, affective state, cultural background and to ultimately further refine existing models [19] of aesthetic responses to the built environment. Our work is simply a first step, and analyzes one element: the effects of visual characteristics on people's perceptions of the streets they encounter. We have here started with three specific qualities, but our approach could be used to assess other aspects and perceptions of urban surroundings as well; for example safety, excitement, expectations of the availability of certain services and businesses, expectations of demographics, or prices. An intended contribution of this work is the investigation of the merits - and limitations - of the presented crowdsourcing and visual analysis approach, and the comparison with other work addressing perceptions of the urban environment. Input from architecture, urban design, and social geography is crucial in going further than just these first steps. As pointed out by Hillier, "buildings are not just objects, but transformations of space through object" [9], and they need to be considered as systems of spatial relations and not as just physical objects. Jacobs [10] points out the essential interplay of the 'bits and pieces' of a city, warns against bland consensus, and points out that unifying design elements should not be so ubiquitous that they be rendered ineffective.

Variety matters, and opposing qualities such as diverse excitement and calm order are both necessary to support pleasant experiences. In addition, contemporary urban designers such as Jan Gehl [5] offer a wealth of criteria for designing 'good public spaces', and these are not just about visual qualities alone. While visual aesthetics are a definite influential feature, the visual words in this work still have to be translated for use by non-computational-oriented audiences. We need to further develop the ways to get at these essential characteristics, to determine how their interplay affect perceptions, and to combine larger-scale quantitative and in-depth qualitative projects that reflect diversity.

Tool, not a directive. There are a number of potential applications of crowdsourced scores such as those generated by this platform. From the perspective of mobile applications, these scores can be used as input for generating local recommendations that go beyond recommending local venues and that consider the aesthetic value of the trajectory. We are currently working on techniques that automatically generate routes that are not only short but also emotionally pleasant [24, 23]. To quantify the extent to which urban locations are pleasant, we combine the data generated by our crowdsourcing project with Flickr metadata. This can lead to practical applications in the city context such as affective route recommendation, aesthetic profiling of streets and neighborhoods, and affective-aesthetic extensions of local 'walk-scores' (see, e.g., `walkscore.com`). On the scale of the built environment itself, perception scores could be mapped and used to identify areas that are most in need of improvement. With an increased understanding of social perception of urban environments, it may be possible for targeted changes. This could benefit local authorities, city planners, real estate agents, architects and homeowners, and members of urban participatory platforms deciding on which interventions to adopt that would make a place appealing. However, this raises the question of whether striving for the visually pleasant 'mean' is truly the way to go. The identification of elements correlated to perceptions of beauty may appear to pose a risk of striving for 'cultural continuity and reduplication', rather than 'architectural design' as defined by Hillier [9]. Revolutionary designs, as well as preservation of urban history, may well be considered 'ugly' to the algorithm. Just blindly applying a computational judgment would not suffice, and more work is thus needed to understand when using these algorithms is appropriate and when it is not. At the same time, these types of platforms allow for much wider audiences to run, and contribute to, studies of perceptions of the urban environment. They can help identify features that affect people's perceptions and affective effects, and add to existing small-scale studies using qualitative ratings of desirable features for specific contexts. Human interpretation is still necessary to apply the outcomes of the process presented here. In this regard, it is interesting to note that the architects' interpretation of what was marked as important, for example, may not necessarily match the exact feature identified by the algorithm. All this goes to suggest that the work presented here should not be taken as a directive for urban design. Rather, it should be taken as a contribution to the discussion on computational approaches in profiling streets, neighborhoods and cities; and as an addition to the toolkit available to those disciplines engaged with people's perceptions of outdoor settings.

TECHNICAL LIMITATIONS

Limitations of Visual Analysis. A limitation of this study is that 'what is perceived' is not necessarily 'what is there', nor 'what would be perceived in-place and in-the-moment'. Users' votes might be influenced by picture quality, cultural and socio-economic factors, and shared priors (e.g., reputation of a neighborhood built over the years). Urban experiences are highly socially constructed, and do not depend on visual characteristics alone. Different kinds of people may engage with cities in different ways and (visual) characteristics may convey different meanings to various demographic groups. Visual characteristics do not necessarily reflect human activity and history, and social or political meanings that require in-depth local knowledge and experience to be understood. Future studies should extend the analysis presented here and further tease out such differences and the effects of a multitude of factors.

Limitations of Image Processing Algorithms. Although our results show that it is possible to use automatic image analysis to gain real insights into why people find certain urban scenes to be happy, quiet or beautiful, there are limitations to this analysis. While the color analysis show interesting associations between certain colors and beauty, quietness and happiness, average color and histograms give a very coarse representation of the content of an image, and the color distribution is very dependent on factors like the time of day. The edge histogram results suggest that people prefer scenes containing more 'natural' shapes. This feature can be affected by the distortion in the Google Street View images caused by the fish-eye lens, however, and correcting the images for this distortion would improve the reliability of this feature. Finally, while the visual words help us to localize the most beautiful, happy and quiet areas in a scene, the result of this are perhaps *too* local, in that, they represent points in the scene, whereas a single object can be made up of the interplay of many such points. An intermediate representation of areas, such as that used by Doersch *et al.* [4], might offer further insights into the types of objects within a scene corresponding to each of the features. Also, given that these local correspondences were calculated based on a relatively small dataset, based on annotations of entire scenes, a more reliable identification of the happy, quiet or beautiful areas within a scene could be achieved by soliciting sub-scene annotations from users, and from obtaining a larger dataset.

CONCLUSION

Our aim has been to identify the visual cues that are generally associated with concepts that are difficult to define, such as beauty, happiness, and quietness. The difficult task of deciding what makes a building beautiful, or what is sought after in a quiet location, is outsourced to the users of our site using comparisons of pictures. Using three existing image processing techniques that extracted colors, edges, and visual words, we were able to find interesting visual associations with the three qualities of beauty, quietness, and happiness.

More generally, we have offered a tool to better understand people's visual perceptions of the urban environment, starting with visual ratings of urban sceneries. Rather than aiming for a 'consensus' in what 'is averagely pleasing', we here aimed to explore the usage of crowdsourcing tools and visual analytics in identifying visual urban features that people appear to react to. Those initial steps call for new research and a critical debate across a variety of disciplines, including urban informatics, planning and architecture. Ultimately, the vision behind this project is that, with a comprehensive list of aesthetic features at hand, we would be more likely to systematically understand people's reactions to their surroundings, without striving for a lovable but bland mean.

Acknowledgments. We thank Adam Barwell for his role in building the site; Jon Crowcroft of Cambridge and the whole NetOS for continuous support; Fred Lewsey for contributing to the launch of the site; the anonymous reviewers for shepherding the submission; and Michael Batty and Licia Capra of UCL for their feedback.

REFERENCES

1. Alexander, C., Ishikawa, S., and Silverstein, M. *A Pattern Language: Towns, Buildings, Constructions*. Oxford University Press, 1977.
2. Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. Speeded-up robust features (surf). *Computer Vision and Image Understanding 110*, 3 (June 2008), 346–359.
3. De Botton, A. *The Architecture of Happiness*. Knopf Doubleday, 2008.
4. Doersch, C., Singh, S., Gupta, A., Sivic, J., and Efros, A. A. What makes Paris look like Paris? *ACM Transaction Graphics* (2012).
5. Gehl, J. *Cities for People*. Island Press, 2010.
6. Geis, K., and Ross, C. A new look at urban alienation: The effect of neighborhood disorder on perceived powerlessness. *Social Psychology Quarterly* (1998).
7. Glaeser, E. *Triumph of the City*. Penguin Press, 2011.
8. Heft, H., and , J. L. Evaluating environmental scenes using dynamic versus static displays. *Environment and Behavior 32*, 3 (2000), 301–322.
9. Hillier, B. *Space is the machine, e-version*. Cambridge University Press, 2006.
10. Jacobs, J. *The Death and Life of Great American Cities*. Random House, 1961.
11. Jurie, F., and Triggs, B. Creating Efficient Codebooks for Visual Recognition. In *Proceedings of the* 10^{th} *IEEE International Conference on Computer Vision (ICCV)*, IEEE Computer Society (2005).
12. Kelling, G., and Coles, C. *Fixing Broken Windows: Restoring Order And Reducing Crime In Our Communities*. Free Press, 1998.
13. Komar, V., Wypijewski, J., and Melamid, A. *Painting by Numbers: Komar and Melamid's Scientific Guide to Art*. University of California Press, 1997.
14. Lindal, P. J., and Hartig, T. Architectural variation, building height, and the restorative quality of urban residential streetscapes. *Journal of Environmental Psychology* (2012).
15. Lynch, K. *The Image of the City*. Urban Studies. MIT Press, 1960.
16. Malobabic, J., le Borgne, H., Murphy, N., and O'Connor., N. Detecting the presence of large buildings in natural images. In *CBMI 2005 - 4th International Workshop on Content-Based Multimedia Indexing* (2005).
17. Manjunath, B. S., Ohm, J. R., Vasudevan, V. V., and Yamada, A. Color and Texture Tescriptors. *IEEE Transactions on Circuits and Systems for Video Technology 11*, 6 (2001), 703–715.
18. Nasar, J. L. Perception, cognition, and evaluation of urban places. *Human Behavior and Environment* (1989), 31–56.
19. Nasar, J. L. Urban design aesthetics: The evaluative qualities of building exteriors. *Environment and Behavior 26*, 3 (1994), 377.
20. Park, D. K., Jeon, Y. S., and Won, C. S. Efficient use of local edge histogram descriptor. In *Proceedings of the ACM Workshops on Multimedia* (2000).
21. Peterson, G. L. A Model of Preference: Quantitative Analysis of the Perception of the Visual Appearance of Residential Neighborhoods. *Journal of Regional Science 7*, 1 (1967), 19–31.
22. Quercia, D., Pesce, J. P., Almeida, V., and Crowcroft, J. Psychological Maps 2.0: A web gamification enterprise starting in London. In *Proceedings of ACM International Conference on World Wide Web (WWW)* (2013).
23. Quercia, D., Schifanella, R., and Aiello, L. M. Shortest Urban Paths or Shortcuts to Happiness? *IEEE Computer Society Special Technical Community on Social Networking E-Letter* (2013).
24. Quercia, D., Schifanella, R., and Aiello, L. M. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. *To Appear* (2014).
25. Robert J. Sampson and Jeffrey D. Morenoff and Thomas Gannon-Rowley. Assessing 'Neighborhood Effects': Social Processes and New Directions in Research. *Annual Review of Sociology* (2002).
26. Salesses, P., Schechtner, K., and Hidalgo, C. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS ONE* (July 2013).
27. Sampson, R. J. When Things Aren't What They Seem: Context and Cognition in Appearance-Based Regulation. *Harvard Law Review Forum 125* (May 2012).
28. Sampson, R. J., and Raudenbush, S. W. Seeing Disorder: Neighborhood Stigma and the Social Construction of Broken Windows. *Social Psychology Quarterly 67*, 4 (2004).
29. Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A., and Freeman, W. T. Discovering objects and their location in images. In *In IEEE International Conference on Computer Vision* (2005).
30. Skogan, W. *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*. University of California Press, 1990.
31. Stendhal. *De L'Amour (On Love)*. Penguin Classics, 1822.
32. Taylor, N. Legibility and aesthetics in urban design. *Journal of Urban Design 14*, 2 (2009), 189–202.
33. Von Ahn, L., and Dabbish, L. Designing games with a purpose. *Communications of ACM* (2008).
34. Weber, R., Schnier, J., and Jacobsen, T. Aesthetics of streetscapes: Influence of fundamental properties on aesthetic judgments of urban space 1, 2. *Perceptual and motor skills 106*, 1 (2008), 128–146.
35. Yang, J., Jiang, Y.-G., Hauptmann, A. G., and Ngo, C.-W. Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the* 9^{th} *ACM SIGMM International Workshop on Multimedia Information Retrieval* (2007).
36. Yeh, C.-H., Ho, Y.-C., Barsky, B. A., and Ouhyoung, M. Personalized photograph ranking and selection system. In *Proceedings of the International Conference on Multimedia (MM)* (2010).