

Like Partying? Your Face Says It All. Predicting the Ambiance of Places with Profile Pictures

Miriam Redi
Yahoo Labs
redi@yahoo-inc.com

Daniele Quercia
University of Cambridge
dquercia@acm.org

Lindsay T. Graham
University of Texas, Austin
lindsaytgraham@gmail.com

Samuel D. Gosling
University of Texas, Austin
samg@mail.utexas.edu

Abstract

To choose restaurants and coffee shops, people are increasingly relying on social-networking sites. In a popular site such as Foursquare or Yelp, a place comes with descriptions and reviews, and with profile pictures of people who frequent them. Descriptions and reviews have been widely explored in the research area of data mining. By contrast, profile pictures have received little attention. Previous work showed that people are able to partly guess a place's ambiance, clientele, and activities not only by observing the place itself but also by observing the profile pictures of its visitors. Here we further that work by determining which visual cues people may have relied upon to make their guesses; showing that a state-of-the-art algorithm could make predictions more accurately than humans at times; and demonstrating that the visual cues people relied upon partly differ from those of the algorithm.

1 Introduction

The Internet is going local. Location-based sites like Foursquare are becoming local search engines, in that, they recommend places based on where users (and their friends) have been in the past. State-of-the-art data mining tools produce those recommendations by automatically analyzing ratings and reviews (Vasconcelos et al. 2014).

As we shall see in Section 2, those tools are well-established and make numbers (ratings) and pieces of text (reviews) relatively easy to mine. By contrast, mining pictures has been proven somewhat harder. Most of the computer vision research has been active in making algorithms more accurate. One of its subareas is called computational aesthetics and, interestingly, is concerned with proposing new ways of automatically extracting visual features that are good proxies for abstract concepts such as beauty and creativity (Redi et al. 2014). It comes as no surprise that, being only at their early stages, computation aesthetics algorithms have not been widely used on social-networking sites.

Here we set out to study whether social-networking sites such as Foursquare might benefit from analyzing pictures with computation aesthetics techniques. To determine 'for what' pictures might be useful, consider the work done by (Graham and Gosling 2011). The two researchers showed that people are able to guess place ambiance (e.g., whether

a restaurant is romantic, whether a coffee shop is friendly) by looking at the profile pictures of visitors. They did so by comparing two types of scores: ambiance ratings given by survey respondents who looked only at profile pictures of visitors; and ambiance ratings given by study participants who actually visited the places. That work showed that people are partly able to determine the ambiance of a place only by looking at the profile pictures of its visitors. It did not show, however, which visual cues the respondents may have relied upon to make their guesses.

Our goal is to determine whether state-of-the-art vision techniques could automatically infer place ambiance. In so doing, we make six main contributions by:

- Analyzing the sets of ambiance ratings collected by (Graham and Gosling 2011) (Section 3). We find that, by considering the pairwise correlations between the 72 ambiance dimensions, one can group those dimensions into 18 fairly orthogonal ones.
- Implementing a variety of state-of-the-art computer vision tools (Section 4). These tools extract the visual cues that the literature of computational aesthetics has found to correlate most strongly with subjective qualities of pictures (e.g., beauty, creativity, and emotions).
- Determining which facial cues people appear¹ to be using to make their guesses about place ambiance (Section 5). To this end, we carry out a correlation analysis between the presence of our visual features and ambiance ratings. We find that colors play an important role (e.g., posh and pretentious places are associated with the presence of pink, strange and creative places with that of yellow), and that computational aesthetics features such as uniqueness effectively capture the associations people make with creative places.
- Showing that our algorithms make accurate predictions (Section 6). We find that they show a precision error at most of 0.1 on a scale [0,1].
- Determining which visual cues our algorithms extract from profile pictures to make their predictions about place ambiance (Section 7). We find that they tend to mostly rely on image quality, face position, and face pose.

¹We say 'appear to be using' simply because we measure correlations and, as such, we do not know what people actually use.

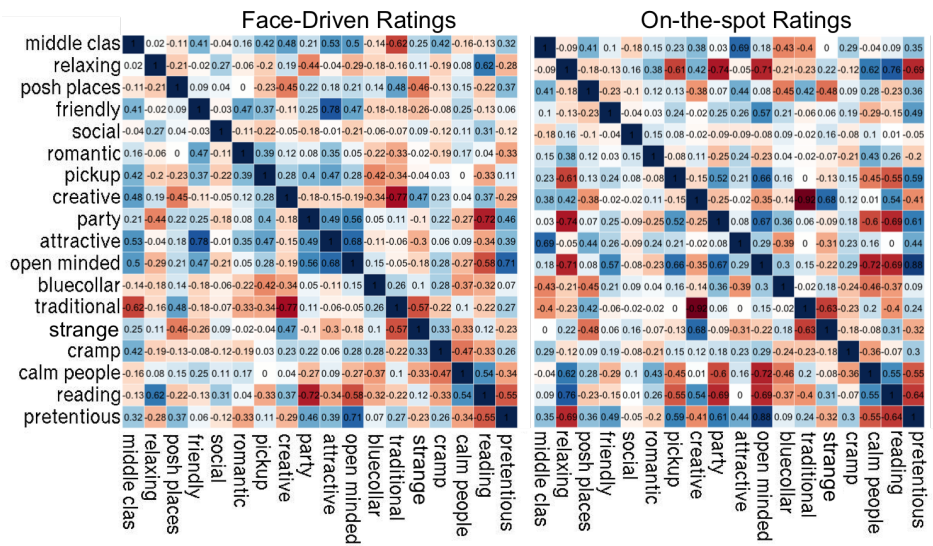


Figure 1: Correlation matrix of the target ambiance ratings.

- Demonstrating that the visual cues people appeared to have relied upon partly differ from those of the algorithm (Section 8). We find that people rely on emotional associations with colors in expected ways (e.g., green goes with calm places). By contrast, the algorithms rely on objective features such as those capturing basic compositional rules. We also find that, as opposed to the algorithms, people may be engaging in gender and racial stereotyping, but might be doing so only at times.

Section 9 concludes by discussing this work’s limitations, and theoretical and practical implications.

2 Related Work

In the context of location-based services, place reviews have been widely explored. Tips and descriptions have been explored to study the popularity of Foursquare places (Vasconcelos et al. 2014; 2012); to explain the demand of restaurants in Yelp; (Luca 2011); and to learn their relationship with check-ins and photos at Foursquare venues (Yu et al. 2014).

Face images have been studied in different disciplines. Computer vision researchers have analyzed faces for several decades. Researchers did so to automatically recognize face ovals (Viola and Jones 2004; Turk and Pentland 1991), identify face expressions (Fasel and Luetjens 2003), predict personality traits (Naumann et al. 2009; Biel et al. 2013), assess political competence (Olivola and Todorov 2010), infer visual persuasion (Joo et al. 2014), and score portraits for photographic beauty (Redi et al. 2015).

More recently, faces have been also studied in the context of social-networking sites. It has been found that, on Facebook, faces engage users more than other subjects (Bakhshi, Shamma, and Gilbert 2014), and that faces partly reflect personality traits (Haxby, Hoffman, and Gobbini 2000; Mehdizadeh 2010; Back et al. 2010).

There has not been any work on using profile pictures of

visitors to *automatically* infer the ambiance of commercial places.

3 Dataset

To study profile pictures and relate their features to ambiance ratings, we resorted to use the dataset introduced in (Graham and Gosling 2011), which we briefly summarize next.

Data for ambiance ratings. This dataset contains ratings from 49 places on Foursquare (24 bars and 25 cafes) located in Austin, Texas. These establishments were randomly selected and in order to be included into the sample contained at least 25 profile pictures of individuals who frequented the location. For each establishment, 72 ambiance dimensions were defined in a way that reflected a place’s ambiance (e.g., creepy), clientele (e.g., open-minded), and activities (e.g., pickup). The dataset comes with two sets of ambiance ratings:

(1) *Face-Driven Ambiance Ratings.* Each place’s 25 profile pictures were arranged on a survey sheet. Ten survey respondents were asked to rate the ambiance of the place (along the 72 dimensions) *only based* on the pictures of its visitors. The respondents were equally distributed across genders (average age: 20.5). The inter-annotator agreement was measured as intra-class agreement and was .32 for ambiance, .69 for clientele, and .33 for activity.

(2) *On-the-Spot Ambiance Ratings.* A separate team of 10 survey respondents were asked to visit the places and rate their ambiance on the same 72 dimensions. The participants were, again, equally distributed across genders, and their average age was 22.4 years. The inter-annotator agreement for on-the-spot ratings was higher: again, measured as intra-class agreement, it was .69 for ambiance, .79 for clientele, and .62 for activity.

The resulting dataset is limited in size because of two main reasons. First, the number of locations with at least 25 profile pictures is limited. Second, it takes time to have two groups of 5 raters each (research assistants above drinking age) at every location at two different times of the day/week. Despite that drawback, for the first time, this dataset thoroughly quantifies the multidimensional ambiance dimensions of real-world locations.

For analytical purposes, it was useful to reduce the 72 ambiance dimensions into a smaller group of dimensions. To this end, we used a semi-automated technique and managed to reduce those ambiance dimensions into 18 orthogonal ones. We did so upon the dataset of face-driven ambiance ratings (similar results were obtained with the on-the-spot ratings). Each place was expressed with a 72-element vector. We clustered those vectors using k -means, set k to one value in $k = \{5, 10, 15, 20, 25, 30\}$ at the time, and obtained different clusters for different values of k . To select the best cluster arrangement, we computed the average cluster silhouette (Rousseeuw 1987), and found $k = 25$ to return the best silhouette. By visual inspection, we determined that $k = 25$ indeed returned the best grouping yet this grouping showed some minor inconsistencies (e.g., “trendy” and “stylish” were assigned to different clusters despite being quite strongly related). To fix that, we run a small user study with 10 participants. We prepared a sheet containing 25 clusters, each of which was described with ambiance terms (in total, we had 72 terms). A participant had to label every cluster based on its associated terms. If the labeling turned out to be difficult because of spurious terms, the participant was free to move those terms across clusters. Also, the participant had to underline the most representative term for each cluster (which we call “target ambiance”). The implicit goal of this task was to improve both intra-cluster consistency and inter-cluster diversity. After analyzing the survey responses, we were left with 18 semantically-consistent ambiance clusters (Table 1) whose target ambiance scores are correlated in expected ways (Figure 1). Face-driven and on-the-spot ratings show very similar correlations.

4 Predictors

Our goal is to predict a place’s ambiance from the profile pictures of its visitors. Therefore, next, we need to extract some predictors out of each picture. We cannot use traditional computer vision features as predictors because these are mostly used to semantically analyze an image, i.e., to understand *which* objects are present in the image (Sivic and Zisserman 2003). That is not the task at hand simply because, for face-driven ambiance ratings, the main object of the image is known (it is a face). By contrast, stylistic characteristics appear more promising. These capture, for example, *how* a face is photographed, its aesthetic value, its dominant colors and textures, its affective content, and the self-presentation choices it entails. Stylistic features have been used previously to automatically assess images’ and videos’ aesthetic value (Datta et al. 2006; Birkhoff 1933), expressed emotions (Lu et al. 2012; Machajdik and Hanbury 2010), creativity (Redi et al. 2014), and interestingness (Gygli et al. 2013). In a similar way, we collect

definition	target	other ambiances
middle-class	trendy	stylish, modern, white-collar, impress
relaxing	relax	cozy, simple, clean, comfortable, pleasant, relaxed, homey
posh	formal	luxurious, upscale, sophisticated
friendly	cheerful	funny, friendly
social	drink /eat	meet new people, watch people, hangout
romantic	dating	cheesy, romantic
pickup	pickup	meat market
creative	artsy	quirk, imaginative, art, eclectic, edgy, unique, hipster, bohemian
party	music	energetic, loud, dancing, camp
attractive	attractive	
open-minded	open	open-minded, adventurous, extraverted
blue-collar	blue-collar	
traditional	bland	conservative, old-fashion, sterile, stuffy, traditional, politically conservative
strange	off path	strange
cramp	cramp	dark, dingy, creep
calm	agreeable	emotionally stable, concencious
reading	read	study, work, web
pretentious	douchy	pretentious, self centered

Table 1: Ambiance clusters and corresponding target ambiance.

here a set of image features (predictors) that reflect portrait-specific stylistic aspects of the image, of the face, and of its visual landmarks (eyes, nose, and mouth).

To infer face demographics, position and landmarks from our profile pictures, we use Face++, a face analysis software based on deep learning (Inc. 2013). Face++ has been found to be extremely accurate in both face recognition (Inc. 2013) and face landmark detection (Zhou et al. 2013). The information of whether a face is detected or not is encoded in a binary feature. If a face is not detected (that happened for 47% of the images), that feature is zero and all face-related features are set to be *missing* values. Next, we introduce our visual features used as predictors and, to ease illustration, we group them into five main types.

1) Aesthetics. An image style is partly captured by its beauty and quality. Psychologists have shown that, if the photo of a portrait is of quality, then the portrait is memorable, gives a feeling of familiarity, and better discloses the mood of the subject (Kennedy, Hope, and Raz 2009). Photo quality has also been related to its level of creativity (Redi et al. 2014), and of beauty and interestingness (Gygli et al. 2013). We implement computational aesthetics algorithms from (Datta et al. 2006; Birkhoff 1933) and score our profile pictures in terms of beauty and quality. More specifically, we compute:

Photographic Quality. The overall visual *photographic quality* reflects the extent to which an image is correct according to standard rules of good photography. To do capture this dimension, we compute the *camera shake amount* (Redi et al. 2014) (the quantity of blur generated by the accidental camera movements), the *face landmarks sharpness*, and *face focus* (which has been found to be correlated with beauty (Redi et al. 2015)). To see how those three dimensions translate in practice, consider Figure 2. This can be considered a good quality picture: there is no camera shake, the face is in focus compared to the background, and the facial landmarks (e.g., eyes, mouth)

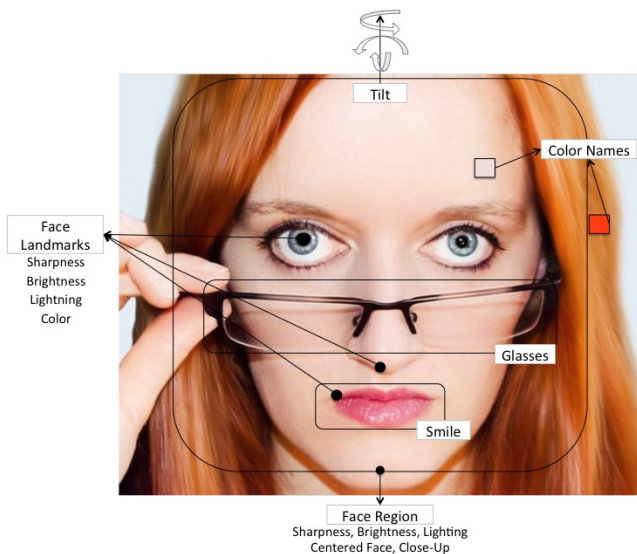


Figure 2: Running example of a profile picture.

are extremely sharp.

Brightness, Saturation, Contrast. The three aspects respectively correspond to the colors' lightness, colorfulness and discriminability in an entire image. They have all been found to be associated with picture aesthetic (Datta et al. 2006; Redi et al. 2015) and its affective value (Valdez and Mehrabian 1994). Darker colors evoke emotions such as anger, hostility and aggression, while increasing brightness evokes feelings of relaxation and is associated with creativity (Redi et al. 2014). For each of our pictures, we compute *brightness*, *lightning*, and *saturation* of the eyes, nose, mouth and the entire face oval. To do that, we add an overall *contrast* metric (Redi et al. 2015). To stick with our running example, Figure 2 has very bright colors: without being over-saturated (too colorful), the contrast is high enough to make bits of the face quite distinguishable.

Image Order. According to Birkhoff, the aesthetic value of a piece of (visual) information can be computed by the ratio between its order (number of regularities) and its complexity (number of regions in which it can be decomposed) (Birkhoff 1933). Order and complexity have been found to be associated with beauty, and to affect how fast humans process visual information (Snodgrass and Vanderwart 1980). We thus compute the *image order* and its *complexity* using a few information theory metrics (Redi et al. 2014), its *level of detail* (i.e., number of regions resulting after segmentation) (Machajdik and Hanbury 2010), and its overall *symmetry*. The picture in Figure 2 can be considered quite conventional: lines are symmetric, and regularities are introduced by the uniformity of its background and the smoothness of its textures.

Circles. The literature on affective image analysis suggests that the presence of circular shapes is registered when certain emotions (e.g., anger, sadness) are expressed (Lu et al. 2012). Therefore, we add the *presence of circular*

shapes to our list of predictors. We compute them by using Hough's transform (Redi et al. 2015). The face in Figure 2 has perfect round shapes in the eyes area only: 2 for the iris, and 2 for the eye pupils.

2) Colors. They have the power to drive our emotions, and are associated with certain abstract concepts (Mahnke 1996; Hemphill 1996; James and Domingos 1953): *red* is related to excitement (Wexner 1954); *yellow* is associated to cheerfulness (Wexner 1954); *blue* with comfort, wealth, trust, and security (Wexner 1954); and *green* is seen as cool, fresh, clear, and pleasing (Mahnke 1996). To capture colors from our pictures, we compute the *color name* features (Machajdik and Hanbury 2010) and *facial landmark colors* according to their hue values (Redi et al. 2015). In Figure 2, the dominant colors are white, red, and pinkish.

3) Emotions. Facial expressions give information not only about the personality of the subjects (Biel, Teijeiro-Mosquera, and Gatica-Perez 2012), but also about the communicative intent of an image (Joo et al. 2014). Faint changes in facial expressions are easily judged by people who often infer reliable social information from them (Iacoboni et al. 2001). That is also because specific areas of the brain are dedicated to the processing of emotional expressions in faces (Goldman and Sripada 2005). We therefore compute the probability that a face subject assumes one of these emotions: *anger*, *disgust*, *happy*, *neutral*, *sad*. We do so by resorting to Tanveer et al. (Tanveer et al. 2012)'s work based on eigenfaces (Turk and Pentland 1991). We also determine whether a face is *smiling* or not using the Face++ smile detector.

4) Demographics. The distribution of age and gender among visitors is expected to greatly impact the ambiance of the place. It is well-known that people geographically sort themselves (in terms of where they choose to live, which places they like) depending on their socio-demographic characteristics and end up clustering with others who are like-minded (Bishop 2009). We take *race* (caucasian, black, asian), *age*, and *sex* as our demographic features.

5) Self-presentation. The way people present themselves might also be related to what they like (Mehdizadeh 2010). To partly capture self-presentation characteristics, we determine whether *sunglasses* or *reading glasses* are used, whether a picture actually *shows a face* or not, and, if so, we determine three main facial characteristics: *face centrality*, whether there is a *tilted face*, and whether it is a *close-up*. Figure 2, for example, shows a close-up of a non-tilted and centered face. Our last self-presentation feature reflects whether the image composition is unique and memorable (we call it "*uniqueness*"). It indicates the extent to which the image is novel compared to the average profile picture (Redi and Merialdo 2012).

To sum up, for each profile picture, we have a total number of 64 features. To combine the features of a venue's faces together, we characterize each place with the *average* and

the *standard deviation* of the features across the 25 pictures. The diversity analysis arising from the standard deviation statistics is needed because “it seems likely that observers do more than simply averaging the individual impressions of the targets. If targets are too diverse, then the group is seen as diverse ...” (Graham and Gosling 2011). For each place, we therefore have a 128-dimensional feature vector, to which we add a value corresponding to the total number of faces present in the group of 25 pictures. Hence, we represent a place with a final feature vector of 129 elements.

5 People Associations

To determine which visual cues the respondents in our dataset may have relied upon to make their guesses, we study the extent to which a person’s visual features impacted respondents’ guesses. We resort to the face-driven ambiance ratings introduced in Section 3. For each place, we compute the pairwise correlations between each of the 129 visual features and each the 18 ambiance ratings. Of course, face-specific features are defined only for images that contain faces. To compute those 2,322 correlations, we use the Spearman Correlation as most of the visual features reflect the presence or absence of visual elements (e.g., glasses) and, as such, they are best interpreted in a comparative fashion rather than using raw numbers. To ease illustration, next, we will group the correlation results (Figure 3) by type of features.

Aesthetics Features. The most relevant aesthetic feature is brightness. Respondents associate eye brightness, mouth brightness, and nose brightness with people who like friendly, open-minded, romantic, and party places (with all correlations above $r = 0.4$). By contrast, dark pictures - in terms of face brightness ($r = -0.28$) and nose brightness ($r = -0.37$) - are associated with those who like cramp places.

Colors. The presence of pink in profile pictures is associated with those who like posh and calm places, while its absence is associated with those who like reading and creative places. The presence of yellow is associated with strange and creative people, and its absence with those who like traditional and posh places.

Emotions. The most important emotion feature is smiling. Profile pictures with smiling faces are associated with those who like posh, attractive, and friendly places (smiling faces are associated with friendly places with a correlation as high as $r = 0.57$), while strange people are thought not to smile.

Demographics. Old people are associated with reading places but, of course, not with pickup places. Race is also associated with ambiance: Asian are associated with social places, while Caucasian with romantic places. The presence of female among a place’s visitors results into considering the place to be good for pickup, to be middle-class, open-minded, romantic, and catered to attractive people.

Self-Presentation. Those who wear glasses are associated with relaxing and reading places, while those who do not with pretentious places. Those who use profile pictures that deviate from the conventional one (i.e., they tend to be *unique*) are associated with middle-class and creative places, while those who are conventional are associated with posh places.

6 Algorithmic Predictions

We have just seen that survey respondents made systematic associations between place ambiance and visual features. Now one might wonder whether an algorithm could make associations as well to automatically predict place ambiance ratings.

To address that question, we use the on-the-spot ratings (Section 3). Hence, for each place, we have 72 ambiance ratings (which might well differ from the face-induced analyzed in the previous section). Again, those ambiance ratings are summarized into 18.

Having this data at hand, we could train a regression framework on part of the 49 places (each of which is represented by the usual 129 features), and we could then test the extent to which the framework is able to predict the 18 ambiance dimensions on the remaining places. The problem is that we have too few observations. To avoid overfitting, the standard rule for regression is to have at least 10 observations per variable (Peduzzi et al. 1996). Unfortunately, in our case, we have 49 places (observations) and 129 features (variables).

To fix this problem, we train the regression framework not on all the 129 features but on the 5 most correlated ones. Since our dataset consists of only 49 observations, we need to carefully maximize the number of training samples. Even a 90 %-10% train-test partition might be restrictive as it might remove important outliers from the training data. To tackle this issue, we resort to the widely-used leave-one-out validation (Salakhtudinov, Tenenbaum, and Torralba 2013). At each iteration, we leave one sample out as test, and train the framework on the remaining samples; the job of the framework is to predict the test sample. The difference between the predicted value and the actual one (that was left out) is the error, which we summarize as the percentage Mean Squared Error (MSE).

Figure 4 shows that our framework accurately predicts all the ambiance dimensions (the error is always below 10%) despite having only five visual features as input. The two ambiance dimensions of friendly and social are particularly easy to predict (the error is close to 0). That is because the pictures in those places tend to be distinctive: they are similar to each other but differ from the pictures of other ambiance dimensions. Party places tend to be associated with relatively more diverse pictures, yet the error is quite reasonable (12%).

7 Algorithmic Associations

So the framework makes accurate predictions of place ambiance, suggesting that visual features are not only likely impact people’s perceptions (as one would expect from the lit-

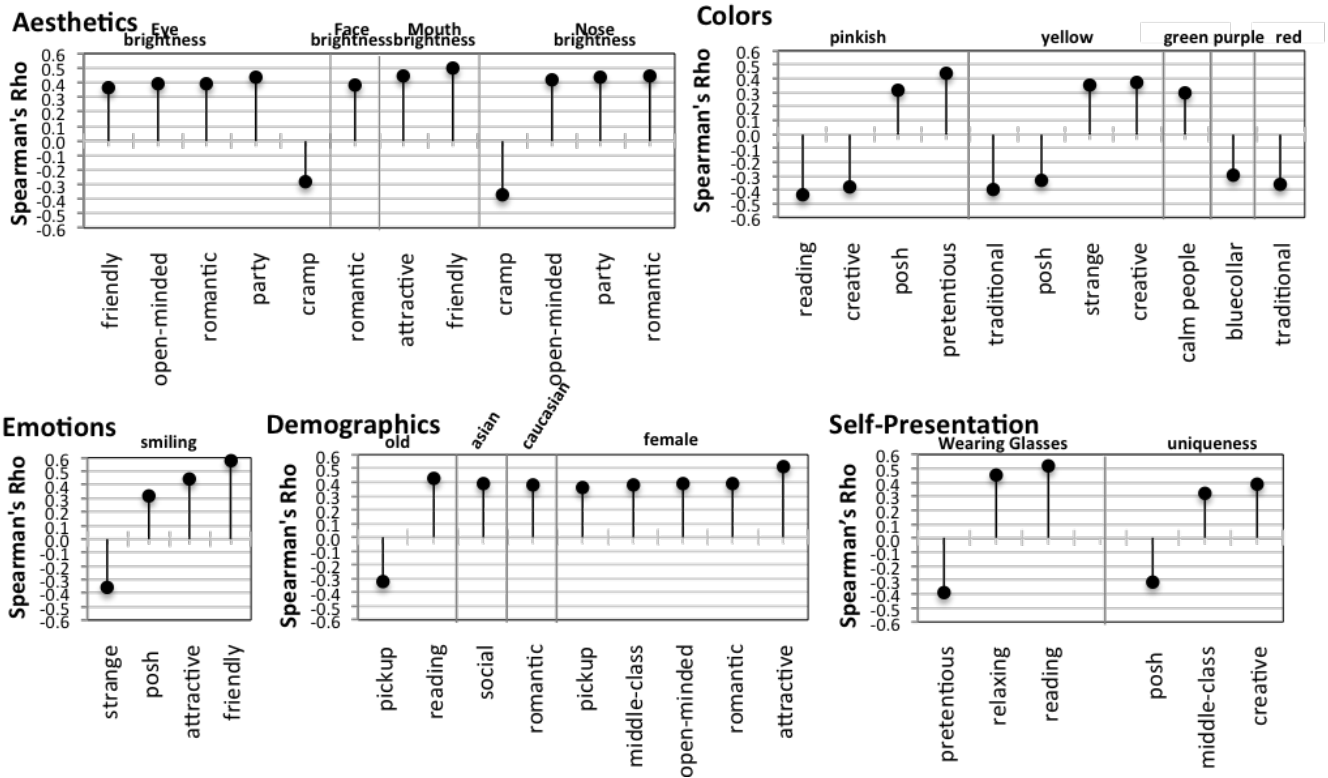


Figure 3: Spearman correlations between the patrons' visual features and judgments of where those patrons might like to go. The visual features are grouped into five types: aesthetics, colors, emotions, demographics, and self-presentation features. All the correlations are statistically significant at $p < 0.05$.

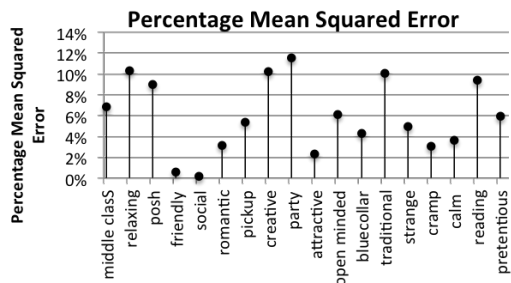


Figure 4: Accuracy Errors of the Algorithmic Predictions.

erature) but are indeed related to the places one likes: one's face partly reveals one's likes. To determine which visual cues our algorithm has used to make its predictions, for each place, we compute the correlation between each of the 129 visual features and each the 18 actual ambiance ratings. To compute those 2,322 correlations, we use, again, the Spearman Correlation. By grouping the correlation results by feature type (Figure 5), we see that:

Aesthetics Features. Dark pictures (those lacking brightness) are indeed used by people who go to cramped places. That is in line with the associations made by the respondents. Our framework finds circular shapes in the

profile pictures of people who go to open-minded, blue-collar, and strange places, while they are absent in the profiles of those who go to posh places. Unlike those going to creative, romantic and middle-class places, those going to traditional places tend to use pictures with more complex backgrounds. People going to relaxing and creative places use quality pictures in their profiles. By contrast, people going to party, middle-class, attractive, friendly, and open-minded places are less prone to quality images: in any of those places, not only a few pictures are of low quality but, since the group of pictures as a whole shows low variability (see panel 'variable photo quality' in Figure 5), all of the place's pictures are systematically of low quality.

Colors. The profile pictures making use of yellow are of those going to relaxing and strange places; white of those going to social, attractive, and cramped places; red of those going to romantic places; purple of those going to cramped places. Instead, pinkish pictures are avoided by those going to cramped places, blue ones by those going to blue-collar places, and black ones by those going to posh places.

Emotions. The most important emotion feature is, again, smiling. In line with the respondents' assessments, our framework learned that those going to strange places do not smile, while those going to places catered to attractive

people do so.

Demographics. Old people do not go to party and blue-collar places, but they are found in cramped, calm, middle-class, and relaxing places. Men seem to avoid pretentious places, while a balanced woman-man ratio tends to be enjoyed by places where attractive people are thought to go.

Self-Presentation. Interestingly, the self-presentation features that matter the most boil down to only two: the use of glasses and face position. Those who wear glasses go to relaxing places, while those who do not wear them go to party, pickup, and open-minded places. People wearing sunglasses go to friendly places. As for face position, those going to blue-collar and party places tend to have their faces in a similar position, while a variety of positions is experimented by those going to relaxing, strange, creative and posh places. Those going to friendly places tilt their heads. By contrast, those going to traditional places do not do so: their faces are centered, and they avoid close-ups. Instead, those going to creative and pretentious places indulge in close-ups. In addition to not smiling, strange people seems to have a tendency to not always show their faces. Finally, the uniqueness feature also matters: those who use profile pictures that deviate from the conventional one go to reading places, while those who have conventional pictures go to traditional places.

8 People vs. Algorithmic Associations

We have seen that both the algorithmic framework and the group of people are able to estimate the ambiance of a place given the profile pictures of its visitors. One might now wonder who is more accurate: the algorithms or the group of people? The answer is ‘depends’. To see why, consider Table 7. A row in it refers to a specific ambiance dimension and reports the predictive accuracy (Spearman correlation) for our group of respondents (2nd column) and that for our algorithmic framework (3rd column). The highest accuracy between the two is marked in bold. The remaining three columns report the top5 features (if available) that turn out to be most important for people, for the algorithm, and for both. The features are placed in two rows depending on whether they are correlated positively (↑ row) or negatively (↓ row).

The algorithm performs better in half of the cases. It generally does so for ambiance dimensions that are well-defined (e.g., posh, friendly, social, romantic). For more multi-dimensional and complex dimensions (e.g., creative, open-minded), the group of people outperforms, but not to a great extent (the predictive accuracies are quite comparable).

Algorithm wins. As opposed to the algorithm, people find it difficult to correctly guess these five ambiance dimensions from profile pictures (the correlations are below 0.20):

Posh places. People appear to rely on the presence of pinkish and yellow, and that of smile; the algorithm, instead, relies on the presence of black and of circular shapes, and on the face position;

Friendly places. People are likely to rely on smiles, while

the algorithm relies on sunglasses (and achieves higher accuracy in so doing).

Social places. People seems to engage in race profiling, while the algorithm goes more for picture brightness and presence of the white color.

Romantic places. People look at the presence of women and of bright pictures, while the algorithm look for red elements and for warmer colors in the face landmark.

Blue-collar places. People unsuccessfully go for the absence of purple, while the algorithm successfully goes for the absence of blue.

People win. As opposed to people, the algorithm finds it difficult to correctly guess one ambiance dimension from profile pictures: that of *open-minded places*. The algorithm relies on the presence of circular shapes and on the absence of reading glasses. Instead, people correctly infer that very bright pictures are typical for open-minded places.

They both agree. More interestingly, from the last column of Table 7, we see that the algorithm partly relies on people’s stereotyping: both agree that the profile pictures for middle-class places suffer from the lack of details (as opposed to those for traditional places); those for relaxing and reading places portray elderly people with reading glasses; those for places catered to attractive people have smiling faces; those for cramped places show darker lightning; and those for strange places show high variability in the use of yellow.

9 Discussion

We have shown that a state-of-the-art algorithmic framework is able to predict the ambiance of a place based only on 25 profile pictures of the place’s visitors. By then looking at the ambiance-visuals associations that the framework has made, we have tested to which extent those associations reflect people’s stereotyping.

Limitations and Generalizability. Some of our results reflect what one expects to find (e.g., the color green is associated with calm places), and that speaks to the validity of those results (concurrent validity). However, one of the limitations of this work is still that we cannot establish how our insights generalize beyond restaurants and cafes in Austin. It is well-known that ambiance and people perceptions change across countries. To conduct such a study, new sets of data need to be collected, perhaps by combining the use of Mechanical Turk (to get the face-driven ratings) and that of TaskRabbit (to get the on-the-spot ratings). Also, researchers in the area of social networks might complement our study and analyze the relationship between activity features (e.g., reviews, ratings) and ambiance.

Theoretical Implications. We have also contributed to the literature of computational aesthetics. Some of our visual features have been indeed used in previous studies that automatically scored pictures for beauty and emotions (Datta et al. 2006; Machajdik and Hanbury 2010). We have now

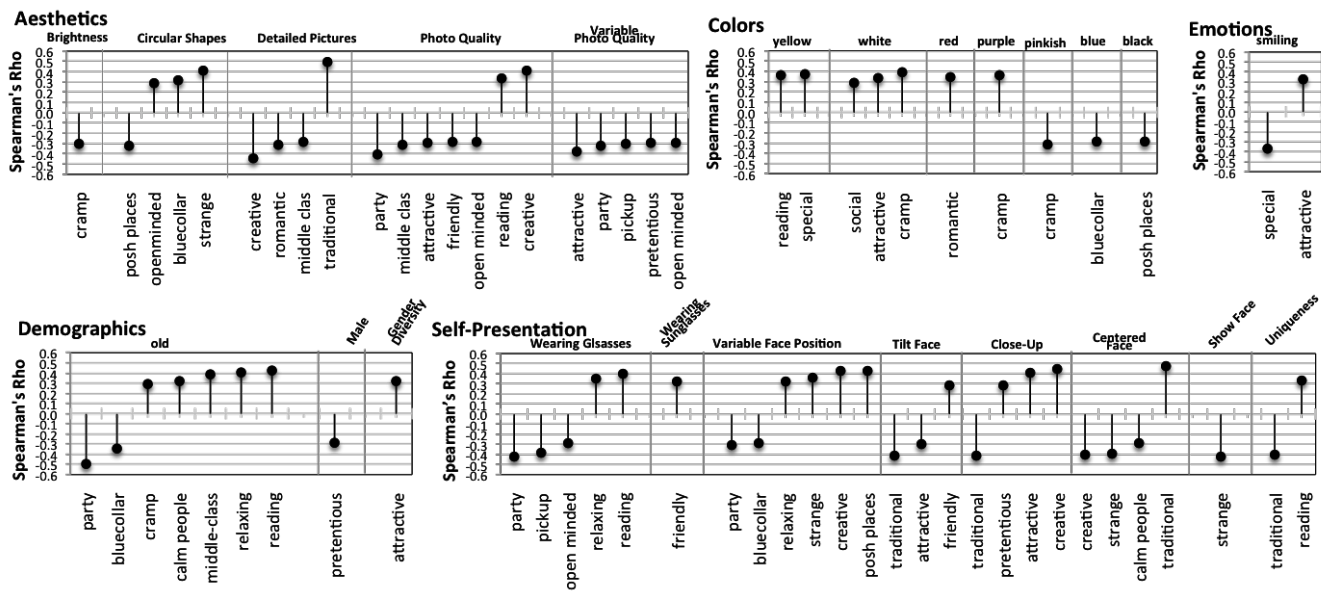


Figure 5: Spearman correlations between visual features of a place’s visitors and the place’s actual ambiance. The visual features are grouped into five types: aesthetics, colors, emotions, demographics, and self-presentation features. All the correlations are statistically significant at $p < 0.05$.

learned that the very same features can help infer place ambiance. More broadly, if one thinks about those places within the context of neighborhoods, then our framework could be used to infer ambiance of not only individual places but also of entire cities. Financial and social capitals of neighborhoods have been widely explored in economics, and aesthetic capital (i.e., the extent to which neighborhoods are perceived to be beautiful and make people happy) has also received attention (Quercia, O’Hare, and Cramer 2014). Adding the concept of ambiance capital to current urban studies might open up an entirely new stream of research in which human perceptions take center stage.

Practical Implications. Our results might enable a number of applications. The easiest one is to recommend places to users based on ambiance. People increasingly manage their private parties by creating Facebook invitation pages. Such a page contains the profile pictures of those who are invited, and of those who are definitely going. To that group of people, it might now be possible to recommend the party venue that best matches those people’s ambiance preferences. It is also possible to do the opposite: to recommend the best faces for a business (e.g., the faces to show on the web page of a coffee shop). To see why the business world might be interested in that, consider the work done by the sociologist Yasemin Besen-Cassino. She found that “chains target affluent young people, marketing their jobs as cool, fashionable, and desirable ... Soon, their workers match their desired customers.” (Besen-Cassino 2013) Indeed, according to the 2000 Department of Labor’s Report on Youth Labor Force, youth from higher socio-economic status are more likely to work than their less affluent counterparts. As Besen-Cassino

puts it: “young [affluent] people see low-paid chain stores as places to socialize with friends away from watchful parental eyes. They can try on adult roles and be associated with their favorite brands.” (Besen-Cassino 2013)

Legible Algorithms. We wish to stress that this descriptive study was possible only because we have used visual features that are explainable. Increasingly, the area of computer vision is moving away from descriptive features, and researchers are putting their effort into high-accuracy deep-learning algorithms (Bengio 2009). Yet, here we have shown that, despite the small quantity of input data, considerable accuracy has been enjoyed by state-of-the-art vision approaches that do not rely on black-box features. Those approaches might be beneficial to a broad audience of researchers (e.g., computational social scientists) who need interpretable insights.

10 Conclusion

Faces are telling, and researchers have known that for decades. We have re-framed the importance of faces within the context of social-networking sites, and we have shown surprisingly strong associations between faces and ambiance. Beyond the practical implications of our insights, the theoretical ones are of interest: we have seen that what appears to be people’s stereotyping matches objective statistical insights at times (e.g., reading glasses go with reading and relaxing places), while it is unfounded at other times (when, e.g., dealing with race and gender).

We are currently exploring a variety of ways to collect more sets of data at scale. The goal behind this effort is to rephrase this ambiance work within the wider urban context.

Ambiance	People	Algorithm	Top-5 Features		
			People	Algorithm	Both
middle class	0.26	0.45**	↑ uniqueness, female ↓ male	old	detail, photo quality
relaxing	0.45**	0.45**	↑ show face ↓	old variable position	glasses, glasses (variability)
posh	0.15	0.30*	↑ smile, pinkish ↓ yellow, quality, unique	variable position black (variability), centered face, number of circles	
friendly	-0.02	0.42**	↑ smile, brightness (mouth,eyes) ↓ photo quality (variability)	sunglasses, sunglasses (variability) brightness (mouth, nose) (variability)	
social	0.15	0.30*	↑ asian, asian (variability), caucasian (variability) ↓	white, white(std), sad (std), brightness (face) (variability)	
romantic	0.15	0.29*	↑ brightness (nose, mouth, eyes), female, caucasian ↓	color (nose, eye), color (nose, eye) (variability), red	
pickup	0.43**	0.34*	↑ female, photo quality ↓ male, old	glasses, glasses (variability), photo quality	
creative	0.60***	0.45**	↑ uniqueness (variability), uniqueness, yellow, white (variability) ↓ pinkish	close-up centered face, detail	
party	0.58***	0.44**	↑ female ↓	glasses, old	photo quality (variability), glasses (variability)
attractive	0.39***	0.57***	↑ female, brightness (mouth, face) ↓ male	closeup, white, white (variability), gender diversity photo quality	smile
open-minded	0.50***	0.18	↑ brightness (mouth, face, eyes) ↓ male	presence of circles glasses, photo quality	
bluecollar	0.10	0.41**	↑ sad, sad (variability), photo quality, purple ↓	presence of circles old, blue, face position (variability)	
traditional	0.59***	0.46***	↑ yellow, red, yellow (variability), uniqueness (variability) ↓	centered face	detail
strange	0.41**	0.56***	↑ yellow (variability), centered face, close-up (variability) ↓	black (variability), presence of circles shows face, centered face	yellow smile
cramp	0.31*	0.46***	↑ disgusted (variability) ↓	white, white (variability), purple, purple (variability)	brightness (nose, eyes)
calm	0.41**	0.24	↑ centered face, green, green (variability) ↓	old centered face (variability)	
reading	0.78***	0.59***	↑ glasses (variability), photo quality, photo quality (variability) ↓ pinkish	yellow	glasses, old
pretentious	0.33*	0.28	↑ pinkish ↓ glasses (variability), glasses	presence of circles photo quality(variability), male	

Table 2: Predictive accuracy of the ambiance dimensions for the group of people vs. the algorithm. (Note: *** = $p < .01$; ** = $p < .01$; * = $p < .05$)

References

- Back, M. D.; Stopfer, J. M.; Vazire, S.; Gaddis, S.; Schmukle, S. C.; Egloff, B.; and Gosling, S. D. 2010. Face-book profiles reflect actual personality, not self-idealization. *Psychological science*.
- Bakhshi, S.; Shamma, D. A.; and Gilbert, E. 2014. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the 32nd ACM Conference on Human Factors in Computing Systems*, 965–974.
- Bengio, Y. 2009. Learning deep architectures for ai. *Foundations and trends in Machine Learning* 2(1):1–127.
- Besen-Cassino, Y. 2013. Cool Stores Bad Jobs. *Contexts* 12(4):42–47.
- Biel, J.-I.; Tsiminaki, V.; Dines, J.; and Gatica-Perez, D. 2013. Hi youtube!: personality impressions and verbal content in social video. In *Proceedings of the 15th ACM Conference on International Conference on Multimodal Interaction*, 119–126.
- Biel, J.-I.; Teijeiro-Mosquera, L.; and Gatica-Perez, D. 2012. Facetube: predicting personality from facial expressions of emotion in online conversational video. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, 53–56.
- Birkhoff, G. D. 1933. *Aesthetic measure*. Harvard University Press.
- Bishop, B. 2009. *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt.
- Datta, R.; Joshi, D.; Li, J.; and Wang, J. Z. 2006. Studying aesthetics in photographic images using a computational approach. In *Proceedings of the 9th IEEE European Conference on Computer Vision*, 288–301.
- Fasel, B., and Luetttin, J. 2003. Automatic facial expression analysis: a survey. *Pattern recognition* 36(1):259–275.
- Goldman, A. I., and Sripada, C. S. 2005. Simulationist models of face-based emotion recognition. *Cognition* 94(3):193–213.
- Graham, L. T., and Gosling, S. D. 2011. Can the Ambiance of a Place be Determined by the User Profiles of the People Who Visit It? In *Proceedings of the 5th International Conference on Weblogs and Social Media*.
- Gygli, M.; Grabner, H.; Riemenschneider, H.; Nater, F.; and Gool, L. V. 2013. The interestingness of images. In *Proceedings of the 14th IEEE International Conference on Computer Vision*, 1633–1640.
- Haxby, J. V.; Hoffman, E. A.; and Gobbini, M. I. 2000. The distributed human neural system for face perception. *Trends in cognitive sciences* 4(6):223–233.
- Hemphill, M. 1996. A note on adults’ color–emotion associations. *The Journal of genetic Psychology* 157(3):275–280.
- Iacoboni, M.; Koski, L.; Brass, M.; Bekkering, H.; Woods, R.; Dubeau, M.; Mazziotta, J.; and Rizzolatti, G. 2001. Reafferent copies of imitated actions in the right superior

- temporal cortex. *Proceedings of the National Academy of Science* 98(24):13995–9.
- Inc., M. 2013. Face++ research toolkit. www.faceplusplus.com.
- James, W., and Domingos, W. R. 1953. The effect of color shock on motor performance and tremor. *The Journal of General Psychology* 48(2):187–193.
- Joo, J.; Li, W.; Steen, F. F.; and Zhu, S.-C. 2014. Visual persuasion: Inferring communicative intents of images. In *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition*.
- Kennedy, K. M.; Hope, K.; and Raz, N. 2009. Life span adult faces: Norms for age, familiarity, memorability, mood, and picture quality. *Experimental Aging Research* 35(2):268–275.
- Lu, X.; Suryanarayan, P.; Adams Jr, R. B.; Li, J.; Newman, M. G.; and Wang, J. Z. 2012. On shape and the computability of emotions. In *Proceedings of the 20th ACM International Conference on Multimedia*, 229–238.
- Luca, M. 2011. Reviews, reputation, and revenue: The case of yelp. com. Technical report.
- Machajdik, J., and Hanbury, A. 2010. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM International Conference on Multimedia*, 83–92.
- Mahnke, F. H. 1996. *Color, environment, and human response: an interdisciplinary understanding of color and its use as a beneficial element in the design of the architectural environment*. John Wiley & Sons.
- Mehdizadeh, S. 2010. Self-presentation 2.0: Narcissism and self-esteem on facebook. *CyberPsychology, Behavior, and Social Networking* 13(4):357–364.
- Naumann, L. P.; Vazire, S.; Rentfrow, P. J.; and Gosling, S. D. 2009. Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*.
- Olivola, C. Y., and Todorov, A. 2010. Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior* 34(2):83–110.
- Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T. R.; and Feinstein, A. R. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology* 49(12):1373–1379.
- Quercia, D.; O’Hare, N. K.; and 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*, 945–955.
- Redi, M., and Merialdo, B. 2012. Where is the interestingness? retrieving appealing videoscenes by learning flickr-based graded judgments. In *Proceedings of the 20th ACM International Conference on Multimedia*, 1363–1364.
- Redi, M.; O’Hare, N.; Schifanella, R.; Trevisiol, M.; and Jaimes, A. 2014. 6 seconds of sound and vision: Creativity in micro-videos. In *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society.
- Redi, M.; Rasiwasia, N.; Aggarwal, G.; and Jaimes, A. 2015. The beauty of capturing faces: Rating the quality of digital portraits. In *Proceedings of the 11th IEEE International Conference on Automatic Face and Gesture Recognition*.
- Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20:53–65.
- Salakhutdinov, R.; Tenenbaum, J. B.; and Torralba, A. 2013. Learning with hierarchical-deep models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35(8):1958–1971.
- Sivic, J., and Zisserman, A. 2003. Video google: A text retrieval approach to object matching in videos. In *Proceedings of the 9th IEEE International Conference on Computer Vision*, 1470–1477.
- Snodgrass, J. G., and Vanderwart, M. 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental Psychology: Human learning and memory* 6(2):174.
- Tanveer, M. I.; Anam, A. I.; Rahman, A. M.; Ghosh, S.; and Yeasin, M. 2012. Feps: A sensory substitution system for the blind to perceive facial expressions. In *Proceedings of the 14th International ACM Conference on Computers and Accessibility*, 207–208.
- Turk, M. A., and Pentland, A. P. 1991. Face recognition using eigenfaces. In *Proceedings of the 6th IEEE Conference on Computer Vision and Pattern Recognition*, 586–591.
- Valdez, P., and Mehrabian, A. 1994. Effects of color on emotions. *Journal of Experimental Psychology: General* 123(4):394.
- Vasconcelos, M. A.; Ricci, S.; Almeida, J.; Benevenuto, F.; and Almeida, V. 2012. Tips, dones and todos: uncovering user profiles in foursquare. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, 653–662.
- Vasconcelos, M.; Almeida, J.; Gonçalves, M.; Souza, D.; and Gomes, G. 2014. Popularity dynamics of foursquare micro-reviews. In *Proceedings of the 2nd ACM Conference on Online Social Networks*, 119–130.
- Viola, P., and Jones, M. J. 2004. Robust real-time face detection. *International journal of Computer Vision* 57(2):137–154.
- Wexner, L. B. 1954. The degree to which colors (hues) are associated with mood-tones. *Journal of applied Psychology* 38(6):432.
- Yu, Y.; Tang, S.; Zimmermann, R.; and Aizawa, K. 2014. Empirical observation of user activities: Check-ins, venue photos and tips in foursquare. In *Proceedings of the First International Workshop on Internet-Scale Multimedia Management*, 31–34.
- Zhou, E.; Fan, H.; Cao, Z.; Jiang, Y.; and Yin, Q. 2013. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In *Proceedings of the 14th IEEE International Conference on Computer Vision (Workshops)*, 386–391.