What Makes Photo Cultures Different?

Miriam Redi
Yahoo
London, UK

Damon Crockett
University of California
San Diego, CA, USA

Lev Manovich
CUNY
New York, NY, USA

Simon Osindero
Flickr
San Francisco, CA, USA

{miriam.redi, damoncrockett, manovich.lev} @ gmail.com   osindero@cs.toronto.edu

ABSTRACT

Billions of photos shared online today are created by people with different socio-economic characteristics living in different locations. We introduce a number of methods for quantifying the differences between such "photo cultures" and apply them to a large collection of Instagram images shared in five mega-cities around the world. First, we extract image content and style features and use them to design a new visualization technique for qualitative analysis of photo cultures. We then use supervised learning to automatically recognize and compare visual activity at different locations and expose surprising connections between geographically distant photo cultures. Finally, we perform a low-level quantitative analysis to understand what makes photo cultures different from each other.

CCS Concepts

•Human-centered computing → Social content sharing;

1. INTRODUCTION

In sociology and media history, the term "culture" is used to characterize behaviors, beliefs or artifacts of a group of individuals in a particular time period and location(s). For example, rather than thinking of "photography" as a single phenomenon, it is more precise to consider it as a collection of many different "photo cultures", each with its set of distinct aesthetic rulesand defining mechanisms. Examples of photo cultures include the "New Vision" European photography in late 20s [10], the socially conscious photography practiced in New York in the 30s [4], or the snapshot-style fashion photography popular in the 90s.

How can this perspective that combines sociology of culture and media history inform studies of today's online communities? Just as photography did during its first 160 years, online photo sharing platforms such as Instagram could include many photo cultures: many individuals and groups might use Instagram to define their cultural identities. However, previous work analyzing Instagram images [6, 1] often approaches this medium as a single global photo culture. Existing works tend to use large samples drawn from Instagram as a whole, generic dataset, without considering possible differences in how Instagram is used by people with different back-
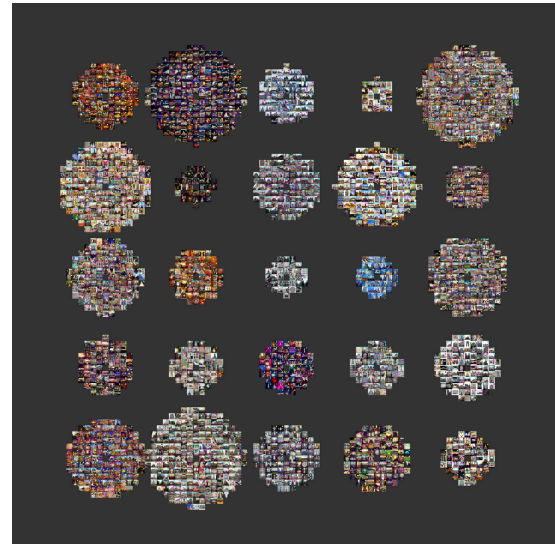
Figure 1: Stylistic clusters of Architecture images.

grounds or geographic areas. Instagram users show very diverse socio-demographic characteristics (location, gender), and our hypothesis is that different users adopt this medium in different ways, making Instagram a collection of photo cultures sharing pictures with different subjects and stylistic attributes.

Our study explores this hypothesis by comparing for the first time Instagram images along one important variable: geographic location. Specifically, we use deep-learning based object detectors and computational aesthetics tools to analyze subjects and styles of 100K images from five well-known global megacities. Using such features, we design a visualization technique based on unsupervised clustering that allows us to qualitatively explore the visual activity of users in different locations, and to expose the existence of different photo cultures. We then design a supervised-learning framework and classify the collected data to understand differences and similarities between the visual preferences of photo cultures. Finally, we dive deep into the visual activity of photo cultures, drawing a map of stereotypical photographic subjects and styles, to understand what makes photo cultures different from each other.

Rather than confirming existing stereotypes, our analysis reveals new, unexpected patterns. For example, we find that Bangkok has the most distinctive photo culture in terms of photographic styles (very bright, unique pictures), while pictures in Berlin and Tokyo show the most unique subjects. While previous work showed [14] that city similarity and proximity are strongly related, we find that geographically distant photo cultures such as Bangkok and São Paulo show the most similar photographic subject patterns.

2. RELATED WORK

Our work closely relates to research that combines computational social science and computer vision to study the impact of social media images. Researchers have analyzed image diffusion [12], relations between image quality and popularity [11], the importance of faces for user engagement in Instagram[1], the content of Instagram posts [6]. These studies reveal many important patterns in Instagram usage, but, unlike our study, they do not consider possible differences due to the image geo-location thus possibly missing many local patterns. Also very related to ours, recent research works have explored the relation between visual features and image location. [9, 8, 3, 14]. The work in this paper differs from previous works on city identity recognition [14, 3] in that we do not focus on pure city elements: we use instead location information to analyze characteristics of photos shared on Instagram within the same geographic area, and to compare these characteristics between a number of photo cultures. We characterize users' pictures though *photographic* and *stylistic* attributes rather than *architectonical* or *urban* elements. As a matter of fact, as we shall see in Sec. 6, our findings differ from the ones in [14].

A set of online projects has used visualization tools to discover patterns in Instagram images from different cities[1] Although our study is inspired by these works, their main method for comparison between urban areas is based on visualizations arising from few basic visual features or face characteristics.

3. METHODOLOGY

To discover photo cultures and their activity, we first draw a sample of Instagram images at different locations, and then analyze their content and style through computer vision techniques.

3.1 Dataset

To explore the similarities and differences in photo cultures at different locations, we selected five cities located on 3 continents: Europe, Asia, and South America. We carefully selected cities that are very different in material and objective ways, as they are situated in different climates, they have different colors and architecture, different fashions, etc. These cities are Bangkok, Berlin, Moscow, Sao Paulo, and Tokyo. For each city, we identified the point coordinates corresponding to the official city center. We then used the Instagram API to collect images and their metadata in a 5km x 5km area around each point, thus capturing a significant part of a city[2]. In order to capture images and data from the same days of the week and hours from all cities, we used a single full week: Dec 4–Dec 12, 2013. The collection process resulted in 656K images, divided between the cities as follows. Bangkok: 162K, Berlin: 24K, Moscow: 140K, Sao Paulo: 123K, and Tokyo: 207K. We then randomly sample 20K images per city (100K in total).

3.2 Features

To understand photo cultures, we characterize various aspects of an image: we extract a group of *Subject* features that describe the image objects and scenes, and a group of *Stylistic* features that describe the image photo techniques and styles.

Subject Features. To describe picture subjects, we compute, for all images, the Flickr *machine tags*:[3] we run the Flickr deep learning-based object detectors, and obtain a set of object tags with their corresponding confidence scores (e.g. "flower, 0.6").
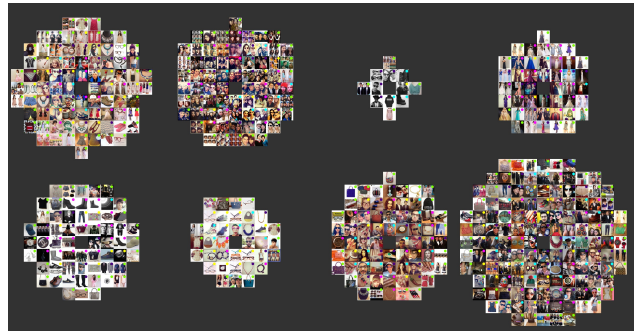

Figure 2: Details of clothing image clusters.

However, machine tag scores can be very sparse: an image contains only few of the possibly detectable objects. Moreover, object-level tag semantics are very fine-grained: for example, among the detectable objects, we can find a variety of different flower species. Such degree of specificity might not be useful to determine higher-level differences among photo cultures. We therefore manually organize object-level tag names into a smaller set of 14 groups, according to their semantics: *architecture, artifacts, fashion, furniture, tools, vehicles, animal, natural, plants, humans, food, activities, concepts, other.*. We associate each subject group with the maximum confidence score among the object-level tags falling inside a group: if an image had object tags "dog, 0.6" and "cat, 0.7", the subject group *animal* will get the score 0.7. The resulting subject feature vector is 14-dimensional.

Stylistic Features. We describe image style in two different ways.
Computational Aesthetics Features: To help classifiers tell which images users will find more beautiful, computational aesthetic features are designed to describe how much an image follows standard rules for good photography.[2, 11]. In this work, we utilize them as a set of individual stylistic features. We first extract information regarding the *Color* distribution in the HSV colorspace (following Itten's Color wheel), together with three HSV-based indicators of emotional responses, namely *Pleasure, Arousal and Dominance*[7]. We then gather information regarding the homogeneity of the *Texture* using Haralick's features from the Gray-Level Co-occurrence Matrices [5]. We describe the overall image *Layout* by computing symmetry and object distribution features [11]. Finally, we collect features reflecting the image *Basic Photographic Quality*: the amount of balance in contrast and exposure, the amount of compression artifacts, and the sharpness of the foreground objects.
Stylistic Machine Tags: To enrich the pool of features describing the stylistic patterns of photo cultures, we also include those machine tags that specifically refer to technical terms related to photography (e.g. black and white, monochrome, lens flare). The stylistic feature vector is 95-dimensional.

4. VISUALIZING PHOTO CULTURES

We present here a new visualization method specifically tailored for photo culture studies.We want to provide a visualization tool to explore subjects and style of large collections of images and spot visual patterns of photo cultures in different cities on a 2D canvas.

Pre-processing: Subject-Specific Photo Style Clustering After feature computation, each image is characterized over 110 (95 stylistic +14 subject + location) dimensions. To allow visualization on a 2D canvas, we must further compress the data characterizing each image, while preserving subject, style, and location information as much as possible. To preserve information regarding the **content**

of Instagram images, we partition the whole image collection according to the subject of the pictures. For each of the 14 subject groups, we create subject-based image subsets by selecting the images whose corresponding subject confidence score is above 0.8, thus ensuring that the image contains a certain subject. To preserve information about the image **style**, we cluster the stylistic features of the images in each subject group. This allows us to group images according to the dominant stylistic choices that photographers make when representing their subjects. We use hierarchical clustering and choose the number of clusters $k = 25$ by manually inspecting the quality of the visualizations resulting from varying k 5 to 50 in steps of 5. After these steps, each image is situated inside one of 14 subject groups, characterized by one of the 25 style clusters computed for each group, and labeled with its city location.

Data Visualization The visualization task is to present the computed clusters on a 2D canvas while incorporating geographic information. To do this, we generate a separate visualization for each of 14 subject groups. Each visualization shows all images in the group divided into 25 style clusters using the "growing entourage" plot, our own visualization technique designed to map high-dimensional image clusters onto a 2D canvas.

To "grow entourages", we first cluster centroids in the 95-d feature space, and then compute, for each image in a cluster, its Euclidean distance from the centroid. We then project the cluster centroids to 2D using t-distributed stochastic neighbor embedding (t-SNE), giving each cluster ("entourage") a location on a 2D map [13]. We plot 2D centroids on a 5x5 grid, and let the algorithm add members to each cluster, starting with those closest to its centroid. Each image added occupies the open grid square nearest its centroid. The images closest to the centroid on the 2D grid are therefore the images closest to the centroid in 95-dimensional space. Finally, we tag each image, in its upper-right corner, with a colored dot indicating its city of origin. Bangkok is green, Berlin is blue, Moscow red, Sao Paulo purple, and Tokyo yellow.

Visualization Properties The resulting plots (see Fig. 2) allow us to understand the dominant photographic styles (clusters) that different photo cultures (geolocation tag) tend to use when representing subjects (image subsets), as well as the extent to which different subjects are associated with each city by Instagram users. Geographic locations and subject groupings are fully preserved. Because images remain bound by their original cluster memberships, the plot preserves a significant amount of intra-cluster stylistic information between single images. Finally, because the clusters themselves are not arranged randomly on the canvas but instead are projected from the original feature space, some measure of inter-cluster similarity is likewise preserved. Our full set of 14 high-resolution visualizations showing all subject groups and style clusters can be found here: https://www.flickr.com/photos/137574408@N03/ .

Quantitative Cluster Analysis The data pre-processing allows not only for qualitative but also for quantitative analysis. By computing subject distributions across images, style clusters and locations, we can get an initial understanding of the trending patterns in photo cultures. We start observing that localized photo cultures actually exist. For example, around 50% of the photos containing food have been taken in Tokyo, 12% in Sao Paulo, 8% in Moscow, 12% in Berlin and 18% in Bangkok. This suggests that Tokyo's users tend to take pictures of food more than users in other cities. Similarly, we find that subjects such as architecture are more popular in Berlin. Bangkok's users tend to prefer fashion, and in Sao Paulo, we can find many pictures of people and activities.

We can also find some information about the style uniqueness of photo cultures. To do so, we look at city-specific clusters, namely stylistic clusters where at least 3/4 of the images come from the
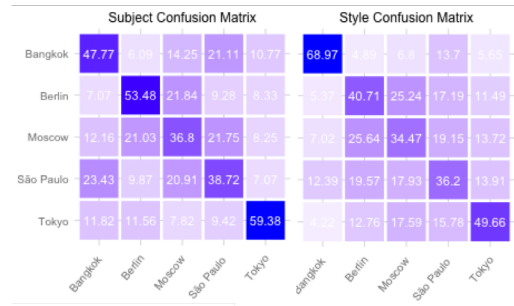

Figure 3: Prediction Task: classification rates for subject and stylistic features. The numbers inside the squares show percentages of photos from cities indicated on x-axis classified as belonging to cities indicated on y-axis.

same location. Photo cultures with high number of city-specific stylistic clusters will likely be more unique in terms of style. We find that around 5% of clusters are city-specific, out of which 42 % belongs to Tokyo, 42% to Bangkok, thus suggesting photo styles in Tokyo and Bangkok are more unique than in other cities.

5. ANALYZING PHOTO CULTURES

In the previous Section, we used clustering to visually compare photo cultures of different cities. In this Section, we use supervised learning to quantitatively compare the photo cultures in our dataset.

5.1 Detecting and Comparing Photo Cultures

To support our comparative analysis, we design a multi-class classification problem, similar to previous work [14]. Given groups of images randomly sampled from the same city, we want to train a classifier able to predict the location where such images originated. The intuition behind this is the following. A group of images sampled from a location should preserve the typical visual patterns of the photo culture from that location. If visual patterns of a city photo culture are clearly distinguishable from others, it will be easy for the classifier to identify the correct location of the image groups drawn from that city. On the other hand, the classifier will misclassify the image groups of cities with similar visual patterns.

Experimental Setup: We formulate the image group location detection problem as follows. We randomly split the images from each city into 50% train and 50% test sets. For both sets, we then randomly sample distinct groups of 10 images[4]. We then compute mean and standard deviation for both the stylistic and the subject features in each group, thus characterizing each group with a 28-d subject feature vector and a 190-d style vector. Next, we label each image group with a category corresponding to the city they have been sampled from, and train a multi-class 10-tree random forest classifier with the resulting data. We report in Fig. 3 the results in a confusion matrix: the number in each circle shows the percentage of image groups from the location indicated by the column label that were classified as belonging to the city indicated by the row label. The higher the percentage of correctly classified image groups for a given city, the more unique and distinctive are the visual patterns of that city's photo culture. Conversely, the higher the misclassification rate between two cities, the higher is the visual similarity of their photo cultures.

Experimental Results: Given that we use a balanced test set, the

---

[4]Although absolute accuracy numbers change when choosing a different group size, we found similar relational patterns between cities for all the group sizes considered
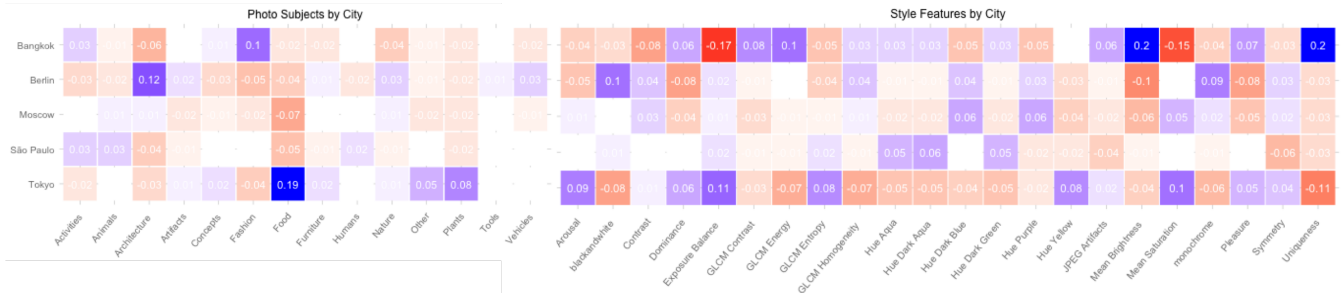
Figure 4: Stereotypical patterns of photo cultures. The squares show correlations between photo subjects and their locations. Correlation values range from -1 to +1. Red color indicates a negative value, white indicates 0, and blue indicates a positive value.

performances of the system (see Fig. 3) for this task are pretty high: while the accuracy of a random classifier would be around 20% for this task, the lowest performance that we get is 34% of correctly classified image groups. Overall, we can see that, in general, subjects are more discriminative than photographic styles for this task. The only exception is the style-based classification of Bangkok image groups, the most accurate among all, showing how unique Bangkok photo culture is.

Subject Features: As we can see from the confusion matrix, Tokyo has the most distinctive photo culture among the cities considered in terms of subjects represented. This is mainly due to the fact that, as we shall see in the next Section, food-related image abound in Tokyo's Instagram pictures, while architecture images common in other cities are missing. On the other hand, Moscow has the least distinguishable photo culture, often misclassified with Berlin or Sao Paulo. Moreover, we can see that, in the subjects depicted, Bangkok and Sao Paulo are the most similar photo cultures, unlike Bangkok and Berlin, whose photo cultures are rarely confused by the classifier. These findings are somewhat different from the ones in [14]: while that study found a high correlation between city identities and geographic proximities, we find here that photo cultures share similarities even though they belong to geographically distant cities, such as Bangkok and Sao Paulo.

Stylistic Features: As mentioned, Bangkok's photo culture has the most unique photo styles, followed by Tokyo: the classifier is able to correctly classify Bangkok's image groups (69%). Again, photo cultures from Moscow seem to be the least unique in terms of photographic styles used, and highly similar to Berlin's photo culture: around 25% of Berlin's image groups are classified as Moscow. We can also see that Bangkok and Berlin have the least similar stylistic patterns, similar to what we observed in the case of subjects.

5.2 What Makes Photo Cultures Different?

After looking at overall cross-photocultural patterns, we dive deep into each photo culture, looking at what makes localized photo cultures unique in terms of the subjects represented and styles used.

To discover the most *stereotypical* visual patterns for each photo culture, we look at how much visual features correlate with the location of the images in our collection. The higher this correlation between feature and city, the higher the unique presence of a given subject or style in the photo culture of the city. To do so, we correlate each feature with 5 binary city vectors, one for each location in our dataset. For a city c, for each image I in our data, the binary city vector will be 1 if the location of image I corresponds to c, and 0 otherwise. In Figures 4 we report the Pearson's ρ correlation coefficients (p-value < 0.05) of the subjects and styles that are statistically significantly related to each photo culture. In many cases, our findings confirm the intuitions suggested by our visualizations in Section 3. We observe the following.

Subject Features: As noticed in Section 4, Tokyo and Berlin have the most clearly distinctive characterizing subjects (see Fig. 4). People in Tokyo tend to take photos of food: by looking at the correlations of the individual object tags and the city vectors, we can see that objects such as *food, meal, meat, soup* are positively correlated ($\rho > 0.1$) with the Tokyo location. On the other hand, users in Berlin tend to depict the architectural aspect of the city (e.g. *building, house, palace*). Bangkok's images are characterized by the presence of clothes and fashion, for example *clothing, dress*, while Sao Paulo's photographers tend to focus on images of people and activities, showing high correlation with object tags such as *face, friends, people*. Among the 5 cities, Moscow shows the least prominent subject patterns, although at an individual object level, we found that the most related tags are *snow* and *night*.

Stylistic Features: By looking at Fig. 4, we can clearly see the distinctiveness of Bangkok's photo culture in terms of stylistic attributes. Images from Bangkok tend to be brighter, more unique, highly unbalanced in terms of exposure, more homogeneous (high GLCM Energy), and with more pleasant and dominant emotions. The second most unique photo culture in terms of style is Tokyo: less unique and highly balanced in terms of exposure, Tokyo's images show higher saturation compared to the others, and they tend to be more colorful (negative correlation *with black and white* and *monochrome* styles, high correlation with the *yellow* hue element). On the other hand, Berlin's images' distinctive pattern is monochromaticity (in particular, black and white). We find less prominent stylistic patterns for the remaining two cities, although, in terms of dominant colors, Sao Paulo's images tend to show more green/aqua subjects, while Moscow's tend to show dark blue and purple.

6. DISCUSSION AND CONCLUSIONS

How does a digital medium such as Instagram reflect cultural differences around the world? Our study, for the first time, presents a comprehensive comparison of Instagram photo cultures of different cities. Using a sample of 100K photos from 5 cities, we employ computer vision techniques to compare these photo cultures in terms of subjects and visual styles. We find significant and often unexpected differences between the cities. For example, despite being on different continents, Bangkok and Sao Paulo are most similar in terms of subjects shown. We also found that Bangkok and Tokyo have the most unique photo styles.

By equating the use of the Instagram medium and its reception with the "average" and the "most" (most frequently used, most popular), some of the previous research treated "Instagram" as a monoculture. The work presented in this paper is part of our effort to show that Instagram and other online media sharing platforms support many distinct photo cultures, and that we need to discover and describe more such different "Instagrams".

7. REFERENCES

[1] BAKHSHI, S., SHAMMA, D. A., AND GILBERT, E. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (2014), CHI, ACM, pp. 965–974.

[2] DATTA, R., JOSHI, D., LI, J., AND WANG, J. Z. Studying aesthetics in photographic images using a computational approach. In *Proceedings of the 9th IEEE European Conference on Computer Vision* (2006), IEEE, pp. 288–301.

[3] DOERSCH, C., SINGH, S., GUPTA, A., SIVIC, J., AND EFROS, A. A. What makes paris look like paris? *ACM Transactions on Graphics (SIGGRAPH) 31*, 4 (2012), 101:1–101:9.

[4] DOHERTY, R. J. *Social-documentary Photography in the USA*. Amphoto, 1976.

[5] HARALICK, R. M. Statistical and structural approaches to texture. *Proceedings of the IEEE 67*, 5 (1979), 786–804.

[6] HU, Y., MANIKONDA, L., KAMBHAMPATI, S., ET AL. What we instagram: A first analysis of instagram photo content and user types. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (2014), AAAI.

[7] MACHAJDIK, J., AND HANBURY, A. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM International Conference on Multimedia* (2010), ACM, pp. 83–92.

[8] PORZI, L., ROTA BULÒ, S., LEPRI, B., AND RICCI, E. Predicting and understanding urban perception with convolutional neural networks. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference* (2015), ACM, pp. 139–148.

[9] QUERCIA, D., O'HARE, N. K., AND CRAMER, H. Aesthetic capital: what makes london look beautiful, quiet, and happy? In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (2014), ACM, pp. 945–955.

[10] SAHLI, J. *Filmische Sinneserweiterung: László Moholy-Nagys Filmwerk und Theorie*. Schüren Marburg, 2006.

[11] SCHIFANELLA, R., REDI, M., AND AIELLO, L. An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures. In *Proceedings of the 9th International AAAI Conference on Web and Social Media* (2015), AAAI.

[12] TOTTI, L. C., COSTA, F. A., AVILA, S., VALLE, E., MEIRA, JR., W., AND ALMEIDA, V. The impact of visual attributes on online image diffusion. In *Proceedings of the 2014 ACM Conference on Web Science* (New York, NY, USA, 2014), WebSci '14, ACM, pp. 42–51.

[13] VAN DER MAATEN, L., AND HINTON, G. Visualizing data using t-sne. *Journal of Machine Learning Research 9*, 2579-2605 (2008), 85.

[14] ZHOU, B., LIU, L., OLIVA, A., AND TORRALBA, A. Recognizing city identity via attribute analysis of geo-tagged images. In *Proceedings of the 14th European Conference on Computer Vision*. Springer, 2014, pp. 519–534.