

RESEARCH

Large-Scale Geo-Facial Image Analysis

Mohammad T Islam¹, Connor Greenwell¹, Richard Souvenir² and Nathan Jacobs^{1*}**Abstract**

While face analysis from images is a well-studied area, little work has explored the dependence of facial appearance on the geographic location from which the image was captured. To fill this gap, we constructed GeoFaces, a large dataset of geotagged face images, and used it to examine the geo-dependence of facial features and attributes, such as ethnicity, gender, or the presence of facial hair. Our analysis illuminates the relationship between raw facial appearance, facial attributes, and geographic location, both globally and in selected major urban areas. Some of our experiments, and the resulting visualizations, confirm prior expectations, such as the predominance of ethnically Asian faces in Asia, while others highlight novel information that can be obtained with this type of analysis, such as the major city with the highest percentage of people with a mustache.

Keywords: faces; geolocation; images; facial attributes

1 Introduction

What do people look like in southern Kenya? What does the average person from Tokyo look like and how do they differ from people in Jakarta or Los Angeles? Such questions are the focus of anthropological studies on human diversity [1], where the traditional approach relies on direct observation, which requires extensive manual effort. This severely limits the types of questions that can be addressed. A computational model of such variations could greatly expand our understanding of contemporary human diversity and enable applications in a wide range of disciplines, including: anthropology, sociology, fashion, security, and computer graphics. This avenue of analysis is enabled by the convergence of two phenomena. First, every day, a growing number of (geotagged) images are uploaded to social media sites. On one popular social media site^[1], geotagged photos are uploaded at a rate of around 500 per minute, or 260 million per year. Second, the state-of-the-art algorithms in computer vision have reached a level of accuracy and robustness that allows detailed scene information (e.g., people, objects, background) to be automatically extracted from images.

Our goal in this work is to explore and analyze the geospatial structure of facial appearance using publicly available imagery (Figure 1). To support this effort, we constructed a dataset by gathering geotagged images

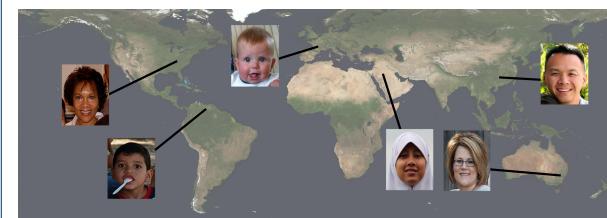


Figure 1 Human facial appearance differs for many reasons, including ethnicity, gender, and hair style. In this work, we explore the relationship between such visual attributes and geographic location.

and extracting aligned frontal face patches. This resulted in a dataset, GeoFaces, of approximately 0.8 million geotagged faces, which, to our knowledge, is the largest publicly available dataset of its kind. In addition, for each facial image patch, we also provide automatically extracted visual attributes such as gender, ethnicity, and facial hair.

We use the GeoFaces dataset to explore the location dependence of human face appearance and visual attributes using a variety of statistical models. This analysis highlights the strong underlying patterns hidden in the data. In addition to the dataset, the main contributions of this work are: (1) visualizations, constructed using techniques from machine learning, that highlight the geo-dependence of visual appearance and facial attributes; (2) quantitative results that further illuminate the dependence; and (3) an evaluation of

*Correspondence: jacobs@cs.uky.edu

¹Department of Computer Science, University of Kentucky, 329 Rose St., 40506 Lexington, KY, USA

Full list of author information is available at the end of the article

[1]<http://twitter.com>

several methods for estimating the location of a face at the continental, sub-continental, and country scale.

2 Related Work

The advent of inexpensive GPS-enabled cameras and persistent connectivity has led to a profusion of publicly available, geotagged imagery. Such images have been used to extract a wide variety of geospatial information, including 3D scene models [2], local weather conditions [3], land-cover type [4], and architectural styles [5]. To our knowledge, little work has used automated approaches for exploring the relationship between facial appearance and geographic location. Despite limited research on the geo-dependence of face appearance, there is a significant amount of research in a number of related problems.

2.1 Face Image Analysis

The human face is one of the most intensely studied object types in computer vision, with active research on a variety of subproblems. We give a brief overview of recent work in the areas of: detection [6, 7], pose normalization [8, 9, 6], attribute estimation [10, 11, 9], and recognition/verification [12, 13]. Modern approaches for face detection use machine learning techniques to automatically determine if an image patch contains a face. A variety of methods have been proposed, including the approach by Shen et al. which uses exemplar-based image retrieval [6] and the approach by Scherbaum et al. [7], which uses a traditional AdaBoost-based technique augmented with novel synthetic training imagery. Approaches for pose normalization [8, 9, 6] use either 2D or 3D warping and often rely on the output from the detector to guide the selection of warp parameters. For attribute estimation, Kumar et al. developed a method for pairwise face verification by comparing sets of human-describable features and visually descriptive similes [10]. Another approach built generative models for opposing facial attributes (smiling-to-frowning, etc.) [11]. Xiong et al. recently introduced IntraFace, a tool for identifying human facial features [9]. Recent work in face recognition has progressed along two fronts, developing methods for extracting more robust features [12] and using improved learning-based algorithms for classification [13].

We make use of recently developed commercial and academic tools for face detection, pose normalization, and attribute extraction. Our work uses these tools to address a higher-level question, “How does expected face appearance depend on location?”

2.2 Large-Scale Image Datasets

Many large-scale image datasets have been introduced recently to advance research in vision-related areas

such as object detection [14, 15, 16], classification [17, 18], and outdoor scene analysis [19, 20]. Similarly, for the task of image geolocation, datasets [21, 22, 23, 24], which contain millions of geotagged images have been collected from Internet search engines and photo sharing websites. We use a similar approach for constructing the GeoFaces dataset, except that the existing datasets have focused on scenes and our focus is faces. Many large facial image datasets have been developed, but most are targeted at facial recognition and have size on the order of several thousands [25, 26, 27]. To our knowledge, the large-scale geotagged face image dataset we have constructed is the first of its kind. It is significantly larger than existing face datasets, and is the only one that provides geotagged imagery.

2.3 Geo-Dependence of Scene Appearance

The relationship between scene appearance and location has been of significant recent research interest, with work focusing on image localization [24, 28, 29, 30, 31], detecting architectural styles [32], and extracting geo-informative features [31, 33]. These methods attempt to automatically discover and/or exploit the relationship between scene appearance and geographic location. We address many of the same issues, but extend this line of research and examine the geo-dependence of facial appearance and facial attributes.

3 GeoFaces

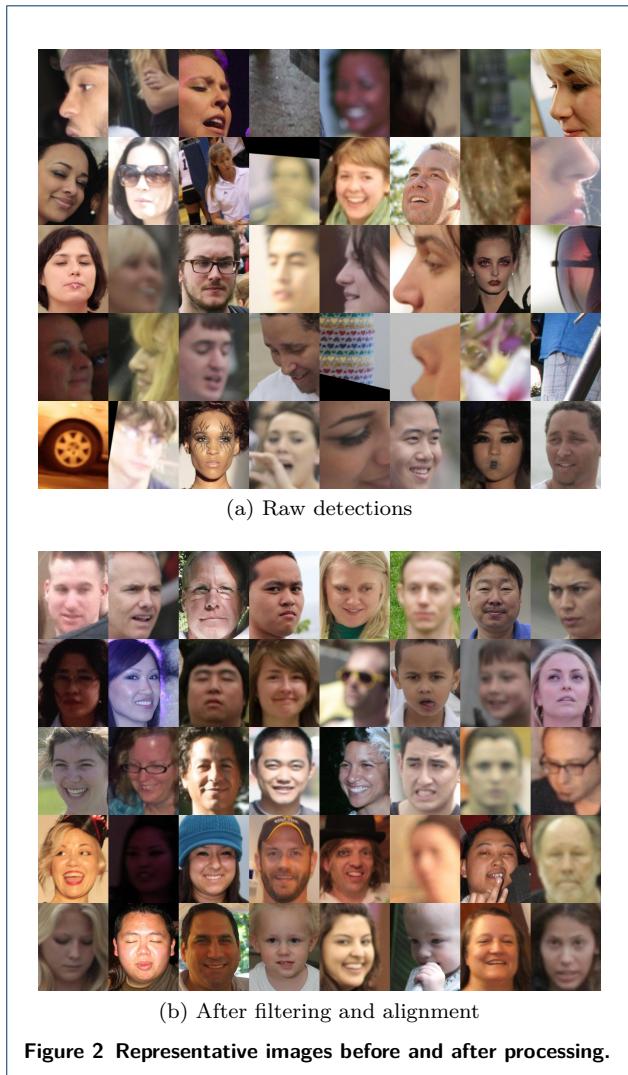
To build GeoFaces, a large dataset of geolocated face patches, we downloaded geotagged imagery from Flickr^[2] with face-related tags (e.g., face, portrait, men, family, friends). For each image, a commercial face detector^[3] was used to detect faces and fiducial points. The detector was tuned to find frontal (or nearly frontal) faces. From 3.14 million images, 3.8 million face patches were extracted. For each face patch, the detector reported the estimated pose direction and detection confidence and also the locations and confidences of predefined fiducial control points (e.g., eyes, nose, mouth). Each face patch was automatically aligned to a common reference frame using a similarity transform, with eye centers as control points.

3.1 Dataset Validation and Cleanup

To eliminate false positives and non-frontal faces, we filtered the original set of facial image patches. Initially, we relied on the confidence values provided by the face detection software. Specifically, we retained

^[2]<http://flickr.com>

^[3]http://www.omron.com/r_d/coretech/vision/okao.html



images with an estimated pose of zero degrees (directly facing the camera), and we empirically determined that a detection confidence greater than 600 (the face detector assigns a confidence value between 0 to 1000 to each of the detected faces) filtered most of the non-face false positive detections. This simple thresholding preserved roughly 30% of the face patches and eliminated most of the non-frontal patches or face-like patterns that were initially detected.

While the detector was, in general, quite reliable, we observed that the detection confidence values and pose estimates were often unreliable for small image patches (i.e., inter-pupillary distance of ≤ 10 pixels). For additional filtering, we trained a classifier using the detected pose and the correlation of the intensity gradient of the image patch with a set of reference faces as features. Using roughly 100 examples (split evenly between positive and negative front-facing patches), we trained a C-SVM classifier with

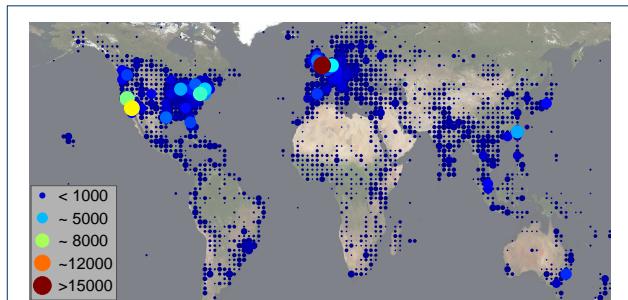


Figure 3 Distribution of images in the GeoFaces dataset.

linear kernel ($c = 1$) [34]. Of the 3.8 million original face patches, this process resulted in 0.8 million face patches in GeoFaces. Visual inspection of the resulting images suggests more accurate filtering than relying solely on the confidence estimates of the face detector. Figure 2 shows representative initial detections and final aligned patches from the dataset. The geographic locations of these patches are shown in Figure 3.

3.2 Extracting Facial Attributes

For each face in our dataset, we extracted facial attributes using IntraFace [9]. The software computes five facial attributes: beard, mustache, gender, glasses, and race. Except for glasses (“Eye-glasses”, “Sunglasses”, “No glasses”) and race (“Asian”, “Black”, “Indian”, “White”), the attributes are binary. The output is a real value for each attribute that reflects the degree of confidence in the selected label. Currently, we make a hard assignment to the particular binary/categorical label and discard the confidence values. While it may be preferable in some cases to use the confidence values, such as when an image is labeled as “No beard” and “No mustache” but the person has visible stubble, we find that the final labels are quite accurate.

3.3 GeoFaces Summary

The image processing pipeline (face detection, alignment, attribute detection) took roughly 5 seconds per image, with most of the computation spent on attribute detection. GeoFaces will evolve as we collect more images and improve the methods for detecting, aligning and filtering. The full dataset, including face patches and visual attribute values, is freely available online^[4]. The remainder of this work we describe various ways of using this dataset to better understand the relationship between human appearance and geographic location. See the Appendix for additional non-geographic analysis of the dataset, such as how the expected size of a face relates to the textual tags of the enclosing image.

^[4]Available at <http://geofaces.cs.ruky.edu>.

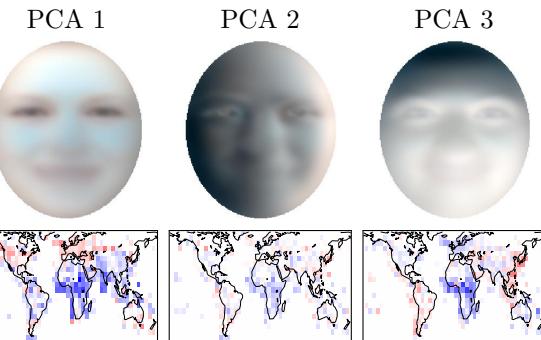


Figure 4 Geographic distribution of Eigenface coefficients. Each map shows the expected value of the coefficient of the corresponding Eigenface image across the globe with blue and red indicating lower and higher values, respectively.



Figure 6 The mean images of clusters found using dictionary learning.

4 Geo-Dependence of Facial Appearance and Attributes

Using GeoFaces, we visualized trends in facial appearance and attributes both across the globe and within smaller regions.

4.1 Geo-Facial Appearance

We explored the relationship between facial appearance and geographic location using Principal Component Analysis (PCA), a statistical model that is commonly called Eigenfaces [35] when applied to facial image analysis. For computational efficiency and to remove background clutter, each face patch was resized to 200×200 and pixels outside of a pre-defined elliptical region were ignored. This elliptical mask has major and minor axes of 157 pixels and 130 pixels respectively, and has center at the middle of the face patch. We construct a vector from all RGB pixel values under the mask and use a standard approach, based on the Singular Value Decomposition (SVD), to estimate the global average face, μ , the PCA components (Eigenfaces), E , and the coefficients for each image, C_i . Figure 4 shows the top three Eigenfaces and corresponding distribution of PCA coefficients mapped by image location. Based on observing multiple regions of smoothly varying coefficients, the first and third Eigenfaces appear to be related to geographic location. The effect is less visible with the second Eigenface, which appears to encode the direction of lighting on the face.

Using the Eigenface representation, we estimated average faces for different parts of the world and observe, perhaps unsurprisingly, that the expected appearance of a face depends on geographic location. For a given location, l , we computed a location-dependent average face, $\hat{f}_l = \mathbb{E}[f|l]$, by estimating the weighted average of nearby Eigenface coefficients (with a Gaussian weight

function centered at l with $\sigma = 5^\circ$) and reconstructing the corresponding image from the Eigenface coefficients. Specifically, when computing the average face for a given location, l , the weight for a face located at x_i is $w_{li} = e^{-(x_i - l)^2/\sigma^2}$. We compute the Eigenface coefficients C_l of the average image located at l as follows: $C_l = \sum \frac{w_{li} C_i}{w_{li}}$. Finally, we reconstruct the average face \hat{f}_l from Eigenface coefficients C_l and global average face μ using the following formula: $\hat{f}_l = EC_l + \mu$. The average images for a set of locations around the globe are shown in Figure 5. Locations with a low number of images are omitted by filtering based on the total weight ($\sum w_{li} < 50$). While it is clear that average facial appearance depends on geographic location, these images do not capture the wide variety of facial appearances that can be seen in a particular place.

4.2 Geo-Facial Dictionary Learning

The Eigenface representation used in the previous section implicitly assumes that the distribution of face appearance in a particular location is well represented by a Gaussian distribution. In this section, we explore a model, based on dictionary learning, that can capture multi-modal structure in the data.

For a subset of faces ($n \approx 250k$), we used sparse dictionary learning [36] to find $k = 300$ representative cluster centers and assigned each face to one of the centers. For computational reasons we used the top 50 PCA coefficients as our feature representation. Figure 6 shows the 32 cluster centers with the largest membership, from largest to smallest along rows. These average images capture variations in gender, pose, lighting, and ethnicity.

We used a kernel density estimate (KDE) (using a Gaussian kernel with $\sigma = 5^\circ$) to approximate the geographic distribution of faces assigned to a particular cluster. Figure 7 shows the difference between this conditional distribution and the distribution of all faces,

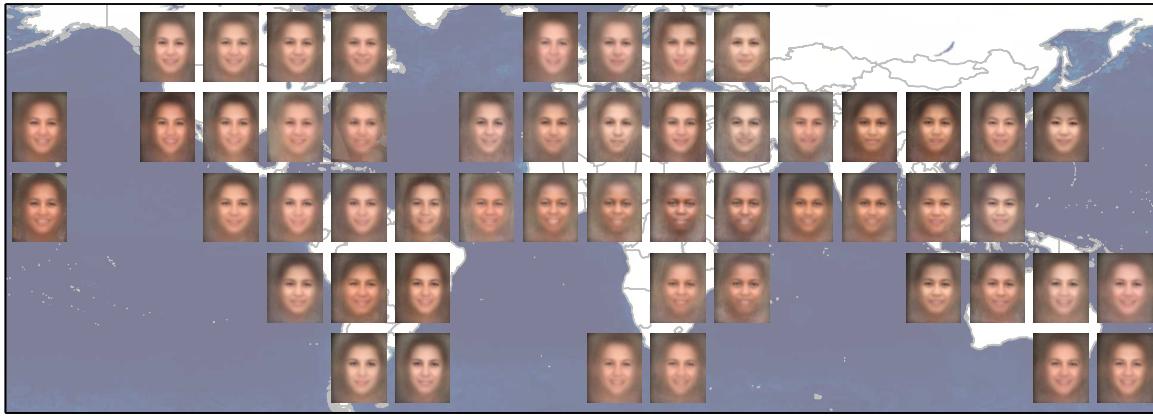


Figure 5 A map of locally weighted average images. The spatial similarities demonstrate the geo-dependence of facial appearance.

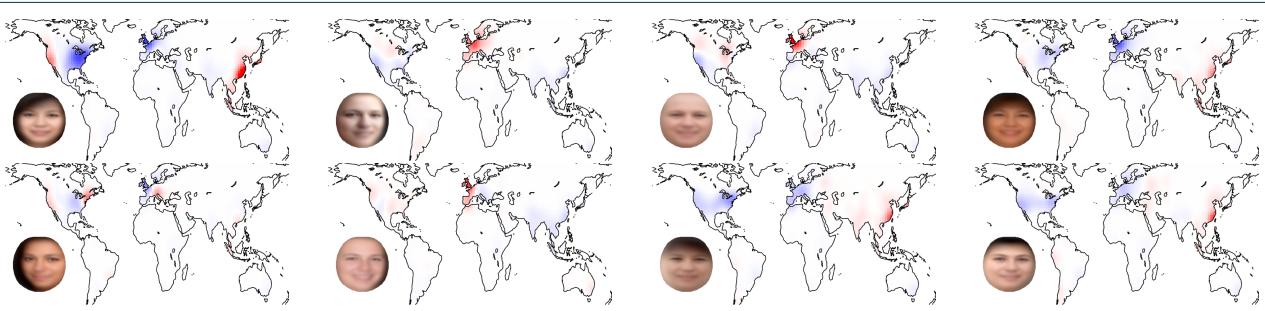


Figure 7 The conditional distribution of geo-facial cluster members. False color images representing the geographic distributions (red=more common, blue=less common) of images assigned to a given geo-facial cluster, together with the cluster center image (inset).

regardless of cluster membership. For example, members of the top-most cluster, which appear similar to an Asian female, are more likely to be found in China and Japan than in Europe or the eastern United States. We also found that many cluster centers shared similar spatial distributions. The primary visual difference between clusters with similar spatial distributions appears to depend on gender, viewpoint, and lighting differences. This motivated us to group face clusters in a manner not based solely on image appearance.

We form super-clusters by grouping clusters based on the conditional geographic distribution of their members. For each conditional KDE, we sampled it on a grid and vectorized the resulting matrix. We then applied non-negative matrix factorization (NMF) to the resulting matrix and extracted three components, each of which can be reshaped and visualized as a false-color map. Figure 8 shows the three resulting geospatial components, and representative faces for the cluster centers that are best described by the given distribution. Super-clusters appear to differ primarily in ethnicity and skin tone, which are known to depend on geographic location [1]. Within a super-cluster the

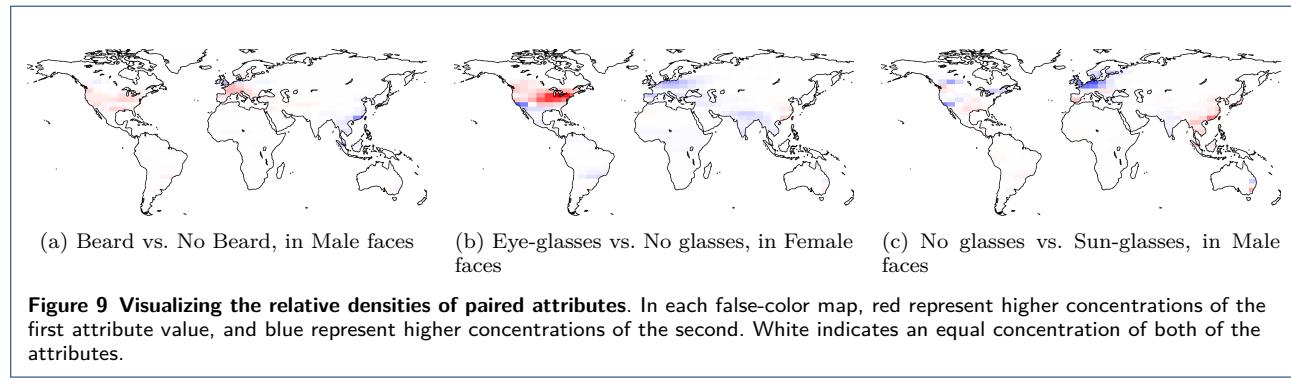
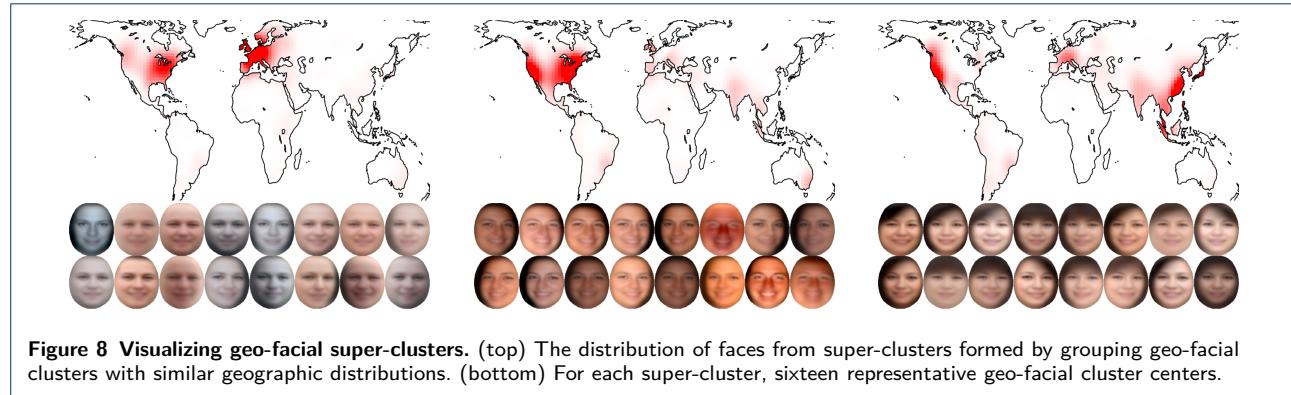
dominant variations appear to be hair style, pose, gender, and lighting conditions, which are less dependent on geographic location.

4.3 Geo-Facial Attributes

Moving beyond appearance-based features, we also explored the geo-dependence of the facial attributes extracted using IntraFace [9]. We computed the relative distribution of each pair of attribute values (e.g., “Male” versus “Female”, “Beard” versus “No beard”) and plotted the relative density histogram (Figure 9), which shows the relative density of pairs of attribute values. For fixed-sized bins, the intensity represents the relative frequency, λ , of a pair of labels:

$$\lambda(n_1, n_2) = \frac{(n_1 - n_2)}{(n_1 + n_2 + p)} \quad (1)$$

where n_1 and n_2 are the number of facial images in a region of interest with a particular attribute value and p is a pseudo-count (for this work, $p=20$), which serves as a prior that reduces noise in the visualization caused by regions with few faces.



At the global scale, we first group images into 5° square bins, an area roughly 340 000 square kilometers (about the size of Germany). For the set of images in each bin, the color of the tile represents the relative frequency of the attribute pair at that location. Many of the global attribute distributions are unsurprising and follow expected geographic distributions. For example, comparing the relative frequency of the “Asian” and “White” attribute values for race reveals distinct, and opposite, modes in Southeast Asia and Europe. Other patterns are perhaps more unexpected. Our dataset contains a higher proportion of “Female” images in eastern Asia and western United States and a higher proportion of “Male” images in the Middle East.

For regions of high image density, we were able to observe finer-grained patterns. Figure 10 shows the relative density histograms, with a bin size was about 6.5 square kilometers, for several attribute pairs for London, Los Angeles, New York, and the whole world. The red “hot spots” in LA and NYC correspond to regions of these cities with large Asian populations (e.g., Chinatown in NYC). Also, there is a high proportion of “Beard” faces (compared to “No Beard”) in the heart of downtown London. On first glance, it is not always clear if these types of observations are due to particular biases in the dataset or cultural norms.

For the regions surrounding 10 major cities (New York, London, Paris, Mumbai, Rio de Janeiro, Hong Kong, Sydney, Beijing, Los Angeles, Tokyo), the stacked bar charts in Figure 11 show the ratios of certain attribute values in each city. Despite inherent biases in the data due to source (e.g., geotagged images uploaded to social media sites), we observed that the ethnicity attributes distribution seem to follow expected patterns. Also, we observed that facial hair is much more common in some cities (Paris and Mumbai) than in others (Hong Kong and Tokyo).

To further explore the relationship between global location and facial attributes, we sought to find groups of faces that are nearby spatially as well as similar in terms of attributes. Many well-known unsupervised clustering techniques could be applied to this problem. We employed the normalized cuts algorithm [37] on a graph-based representation of the dataset, where nodes were facial images and edges encode attribute similarity. Each node was connected to its five nearest (spatially) neighbors and the weight encodes the similarity between attribute values and the pixel intensity of the corresponding facial regions. The clustering process identified groups of faces from similar regions that share similar attributes.

Interestingly, although the clustering process was unsupervised, the discovered clusters tend to align with

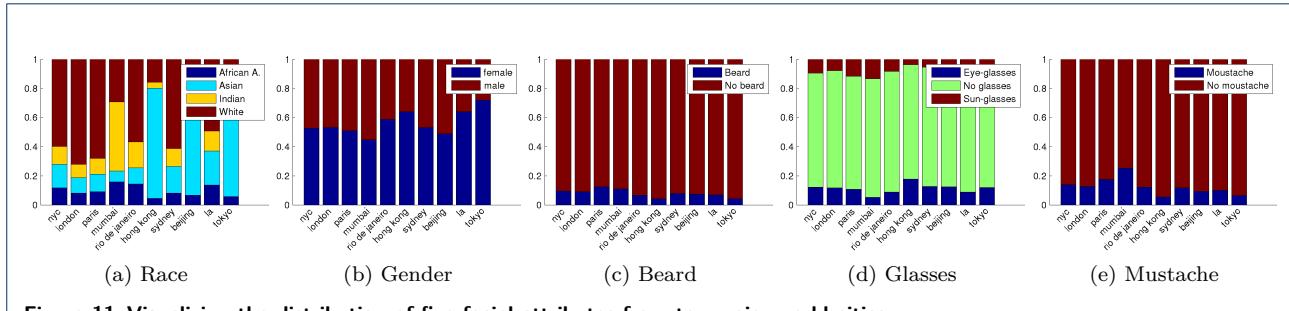
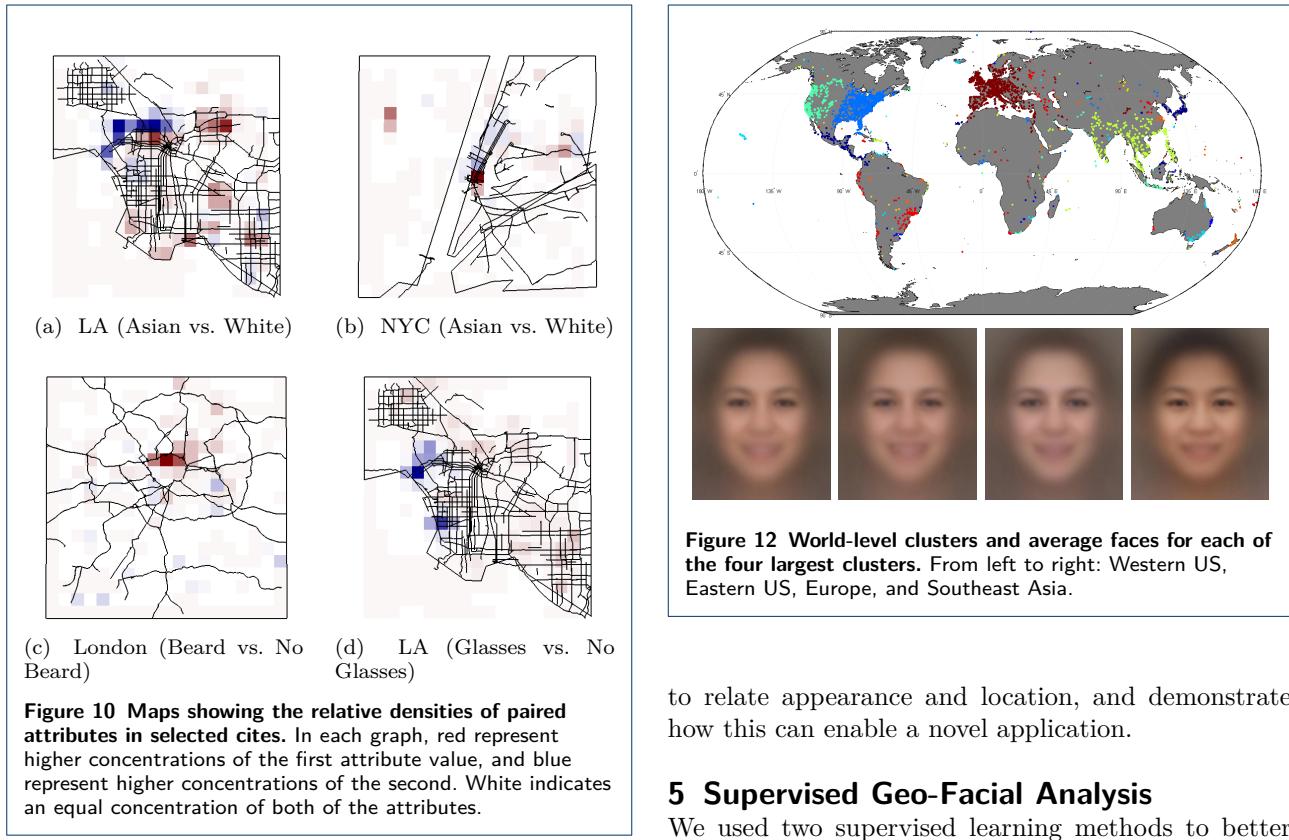


Figure 11 Visualizing the distribution of five facial attributes from ten major world cities.



geographically meaningful regions. Figure 12, shows the results of an experiment with 15 clusters. Regions such as Eastern United States, Central America, and Europe are evident in the clustering output. Average face images from several groups illustrate the correspondence between clusters and geographic regions.

These results demonstrate that facial appearance and attributes are strongly dependent on the geographic location where the image was captured. In this section, our focus was on using unsupervised methods and visualization, which is useful for human understanding of the data. In the following section, we describe several experiments that use supervised learning

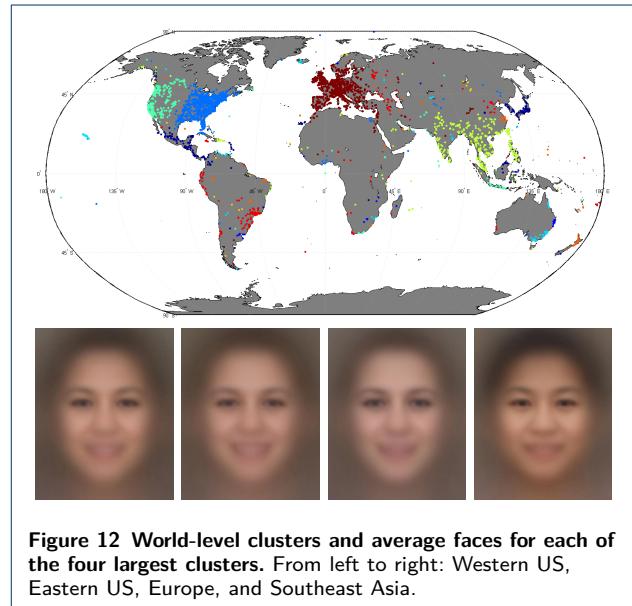


Figure 12 World-level clusters and average faces for each of the four largest clusters. From left to right: Western US, Eastern US, Europe, and Southeast Asia.

to relate appearance and location, and demonstrate how this can enable a novel application.

5 Supervised Geo-Facial Analysis

We used two supervised learning methods to better understand the relationship between geographic location and face appearance: canonical correlation analysis (CCA) and linear discriminant analysis (LDA). We use both methods to extract location dependent features that support observations about facial appearance in various world regions, and, for LDA, enable the novel application of estimating the geographic location of a face.

5.1 Location-Dependent Component Images

CCA is a multivariate statistical tool for exploring relationships between paired sets of variables. Given two datasets $A \in \mathcal{R}^{m \times n}$ and $B \in \mathcal{R}^{p \times n}$ containing paired observations, CCA finds sets of projection vectors (u_1, u_2, \dots) and (v_1, v_2, \dots) such that the random variables $(u_1^\top A, v_1^\top B)$ are maximally correlated. That

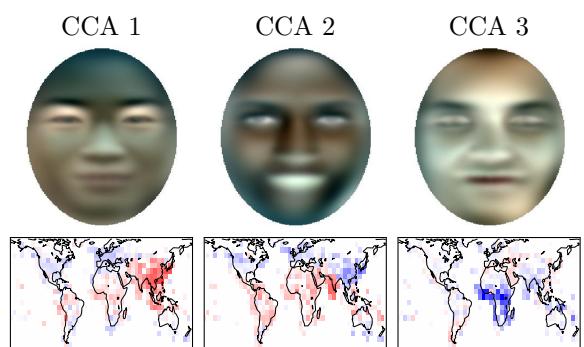


Figure 13 The top three location-dependent CCA component images. Each map shows the expected value of the coefficient of the corresponding image across the globe. The distributions of the CCA coefficients are more strongly related to geographic location than the PCA coefficients (Figure 4). The color blue indicates lower values of CCA coefficients while red indicates higher values.

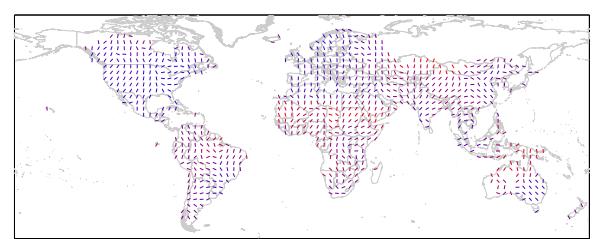


Figure 14 A direction field estimated using CCA between face appearance and geographic location in local neighborhoods. The lines show the cardinal direction that is most correlated with facial appearance change. The lines are color coded by the correlation coefficients, from blue (low) to red (high).

is, it finds u_1, v_1 such that $\rho = \text{corr}(u_1^T A, v_1^T B)$ is maximized. The pair of vectors (u_1, v_1) is called the first canonical pair, and subsequent canonical pairs are defined similarly.

For geo-facial analysis, we sought to find a set of component images and corresponding coefficients that are strongly location dependent. Let A be our set of Eigenface coefficients, one for each image. Let B be an indicator variable encoding image location. The non-zero entry corresponds to the latitude/longitude bin where the image was captured (we use 6° square spatial bins). Performing CCA on this paired data resulted in a projection of our PCA basis and a projection of our locations, as represented by our bin structure. The results of this method applied to our full dataset are shown in Figure 13. The top three components, and their corresponding geographic distribution show a strong location dependence. Based on the distribution maps, it appears that the first three components correspond to the extent to which the face is East Asian, African or Indian, respectively.

5.2 Directions of Facial Appearance Variation

In the previous section, CCA was used to convert weakly location-dependent Eigenfaces into strongly location-dependent components. This analysis was dominated by global population patterns and obscured some of the local structure in facial appearance variation. Here we focus our analysis on local facial appearance variations. Instead of using a global indicator variable to represent geolocation, we performed CCA on the images from small spatial areas and used a linear model for geolocation.

For a given location, l , we created a filtered dataset, A_l , of all faces within 10° . We used the latitude and

longitude of these faces as the paired dataset, B_l . From CCA, the element of the first canonical pair that corresponds to B_l is a vector, v_l , which represents the direction with the most significant face appearance change. Figure 14 shows the resulting direction field for this analysis over multiple locations overlaid on a world map. Visually, the computed gradients match our intuition. For example, in the area around the Mediterranean the directions are mostly vertical because of the strong differences in appearance between Africa and Europe. We see similar patterns between the United States and Mexico, and between India and East Asia.

5.3 Location-Dependent Face Classification

We address the novel application of estimating the geographic location of an image using only a face it captures. We compared three commonly used classifiers linear discriminant analysis (LDA) [38], random decision forests (RF) [39] (150 trees), and a support vector machine (SVM) [34] (C-SVM, linear kernel with $c = 1$) using three feature representations: PCA, one based on Histogram-of-Oriented Gradients (HOG) [40, 41], and another based on Local-Binary Pattern (LBP) [42, 43] features. For the PCA feature vector, we used the top 50 features. For the HOG-based feature, we used a cell size of 8×8 pixels and a block size of 2×2 and computed the features from a 210×168 region from the center of patch. The final, concatenated feature vector is 17360 dimensional. In order to reduce the dimensionality of the feature, we selected the top 300 PCA coefficients and use that as our HOG-based feature representation. We used a similar approach to compute the LBP-based descriptors with 21×21 cells.

We trained a one-vs-all (OVA) classifier for each classification method and each feature representation. For each class, the training set for each consisted of 1500 positive examples and 1500 negative examples sampled from the rest of the world. The remainder of the dataset was used for testing. Table 1 shows the average

Table 1 The average accuracy for continental-level one-vs-all classifiers for different combinations of features and classifiers.

	PCA	HOG	LBP
LDA	62%	62%	63%
RF	61%	61%	61%
SVM	61%	54%	55%

Table 2 The accuracy for all one-vs-all continental-level classifiers.

Continent	Acc. (%)
Asia	73%
Africa	70%
Europe	60%
Americas	56%
Oceania	56%

accuracy, across all continents, for all methods. The results show that LDA+LBP outperformed all others. In the remainder of this section, we analyze the performance of this method in greater detail.

Table 2 shows the overall accuracy of each OVA classifier. The respective classifiers for Africa and Asia were significantly more accurate than the others. Figure 15 shows the proportion of faces from different continents (columns) that were classified as being from a particular continent (rows). This affinity matrix shows that, for example, fewer people from Asia were misclassified as being from Europe (25%) than were misclassified as being from Africa (41%). There is clearly a pattern, but overall the accuracy is fairly low. We speculate that one significant source of error is the large spatial area and diversity of a continent.

We also trained classifiers on 23 sub-continental regions^[5]. Table 3 lists the sub-continent regions and classification accuracies of the classifiers with the best performance. Figure 16 shows the affinities (proportion of positives) between the sub-continental regions. The block diagonal structure of the matrix shows four distinct clusters. For example, Australia, New Zealand, Europe, North America, and Central and West Asia form a cluster indicating that faces from these regions look similar. Finally, we trained OVA classifiers at the country-level, mainly a political, rather than geographic, partitioning. Table 4 shows the ten countries

^[5]<http://unstats.un.org/unsd/methods/m49/m49regin.htm>

Table 3 The sub-continental regions with the highest accuracy classifiers.

Sub-Continent	Acc. (%)
Middle Africa	86%
Western Africa	83%
Eastern Africa	80%
Eastern Asia	78%
Southern Asia	76%
South-Eastern Asia	74%



Figure 15 Continental-level affinity matrix. The proportion of faces classified as positive for different continents (columns) when trained one-vs-all for a particular continent (rows). For example, for a classifier trained to distinguish between European and non-European faces (row 3) we found that 46% of the faces in the Americas (column 2) were labeled as being European.

Table 4 Countries (identified by ISO country code) for which the one-vs-all classifier had the highest (a) and lowest (b) localization accuracy.

Country	Acc. (%)
GHA	82%
ETH	81%
TWN	77%
NGA	75%
KHM	74%
MMR	74%
HKG	73%
BGD	73%
GTM	73%
LAO	72%

(a) Most distinctive

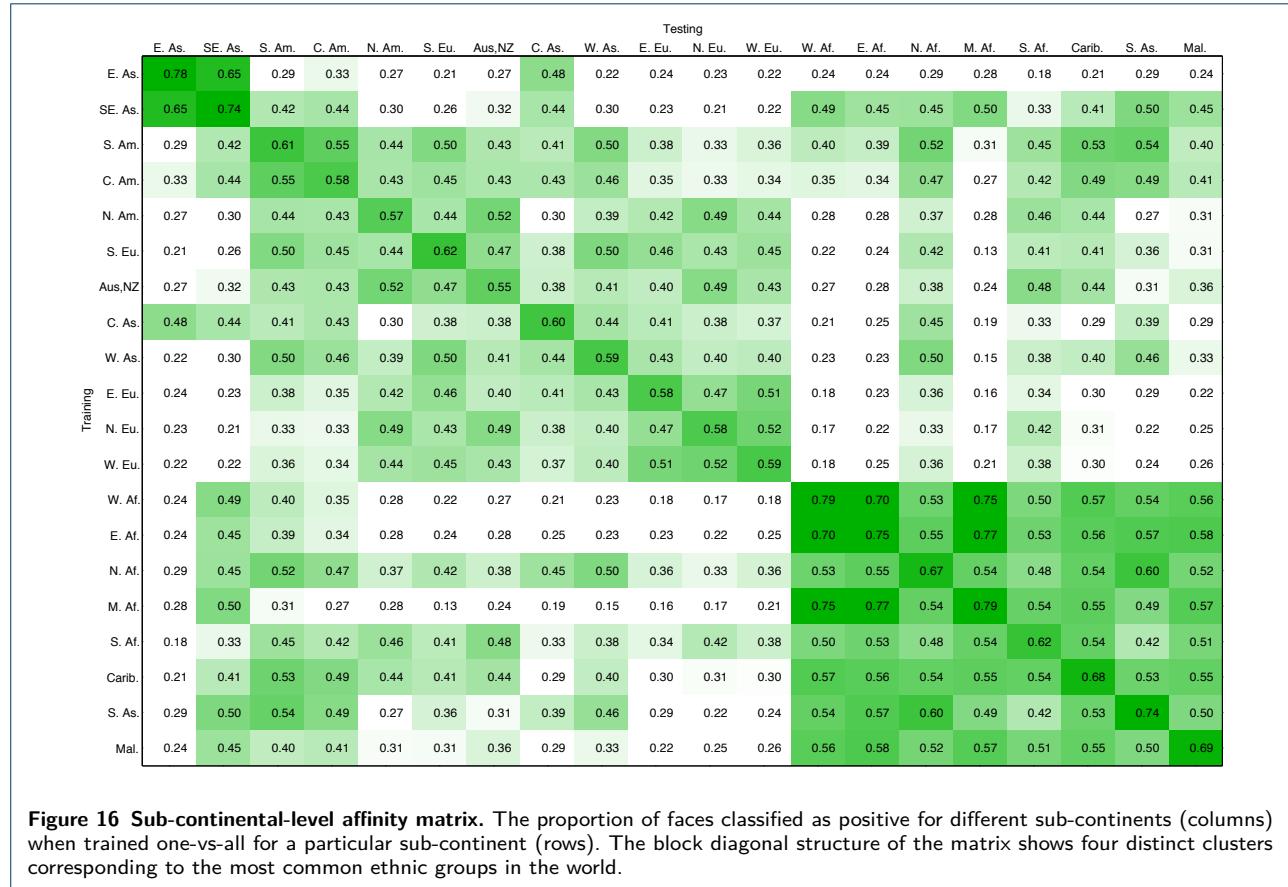
Country	Acc. (%)
SWE	55%
FRA	55%
ARG	55%
BEL	54%
NZL	54%
GBR	54%
USA	54%
DEU	53%
AUS	51%
CAN	49%

(b) Least distinctive

with the highest classifier accuracy. Figure 17 shows maps for ten target countries color-coded based on the percentage of faces labeled as positive by the given country's OVA classifier.

6 Discussion

We used a large dataset of geotagged face patches, collected from the Internet, to explore the geo-dependence of human facial appearance. We applied statistical techniques to explore this geo-dependence and found that there is rich structure in this relationship that is not fully explained by differences in the distribution of ethnic or racial groups. In our analysis, we rely on existing techniques for face detection, pose estimation, appearance normalization, and attribute estimation. Our work with the GeoFaces dataset highlights the need for continued improvements to these core algorithms. Such improvements will increase the number



and quality of facial image patches we can use, which will increase the accuracy of our higher-level analysis.

There are many potential future applications of this type of analysis, both within computer vision and in other domains. In computer vision, we envision the learned models supporting the creation of algorithms for facial image detection, recognition, and alignment that are tailored for particular geographic locations. Outside of computer vision, we envision a wide-range of users posing questions to our system. Such questions could come from sociologists (e.g., “What are current trends in facial hair?” [44]), security officers (e.g., “Where is this person probably from?” [45]), or school children (e.g., “What do people look like in Bangladesh?”).

This work was made possible by the availability of large repositories of geotagged images and the maturity of facial image analysis algorithms. In addition to the immediate applications to facial imagery, we envision that this work will motivate similar work for other types of objects, both natural and man-made.

We continue to expand the GeoFaces dataset, improve the underlying computer vision tools, refine our analysis techniques, and attempt to reduce dataset

Table 5 The distribution of number of frontal-faces per image for different tags.

Tags	0	1	2	3	≥ 4
Face	85.9%	13.6%	0.3%	0.1%	0.1%
Family	78.6%	18.4%	1.5%	0.7%	0.8%
Friends	75.7%	21.1%	1.7%	0.8%	0.8%
Group	84.8%	11.8%	1.3%	0.8%	1.3%
Party	74.0%	22.9%	1.7%	0.7%	0.7%
Portrait	78.9%	20.0%	0.6%	0.3%	0.3%
Wedding	68.7%	25.6%	2.6%	1.4%	1.7%

bias by finding alternative sources of geotagged face imagery. For future work, we plan to investigate the time-varying aspects of facial appearance by collecting geotagged images with known capture time. As a long-term goal we plan to enable interactive analysis, at finer geographic scales and over a wider variety of attributes.

Appendix: Non-Geographic Data Analysis

In addition to analyzing the geo-dependence of face appearance, the GeoFaces dataset enables us to explore non-geographic dependence. Here we focus on the relationship between facial appearance and the textual

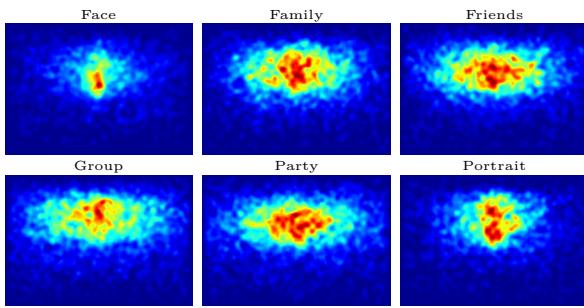


Figure 18 Visualizing the expected face location in images with different tags. Red (blue) pixels are more (less) likely to be the center of a face.

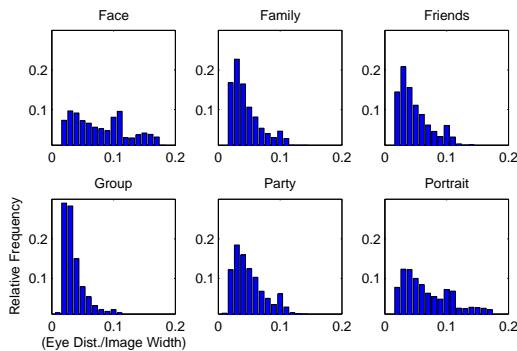


Figure 19 Visualizing the distribution of relative face size in images with different tags. The horizontal axis corresponds to inter-ocular distance relative to the image size, and the vertical axis shows relative frequency.

tags applied to images, which are often added by the individual that uploaded the image.

Table 5 shows, for various textual tags, the distribution of the number of faces found in an image. For example, for images tagged “Wedding”, 25.6% of the images contained a single face, and 5.7% contained two or more faces. For images tagged “Portrait” 20.0% of images contained a single face, but only 0.6% contained two or more faces. For all tags a large percentage of images (generally 70–80%) contained no images due to our aggressive filtering. This type of analysis could be used in the future to guide the selection of search keywords that are likely to result in a large number of high-quality faces.

Figure 18 shows the differences in the expected location of detected faces for different tags. Each heatmap was formed by counting the number of images that contained a face centered at each pixel and then convolving the resulting 2D histogram with a Gaussian kernel. For example, in images tagged “Face” most of the faces are centrally located in the upper half of the image but for images tagged “Family”, “Friends”,

“Group”, or “Party” faces are more frequently found on the left and right sides of the image. In images tagged “Portrait” the typical horizontal position of a face is much more constrained than the vertical position.

Figure 19 shows the expected size, in terms of the fraction of the image width between the eyes, of a face for different tags. This shows, for example, that images tagged “Face” are more likely to occupy a large part of the image than in images tagged “Family”. There is also a pronounced increase in probability at around 0.1 for each tag, likely due to the popular posing of full-body portraits.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was partially supported by the NSF (CNS-1156822) and DARPA (D11AP00255). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor. We gratefully acknowledge flickr users ennuislife, hotlantavoyer, rocketboom, besighawn, and scubabix for allowing us to use their images via Creative Commons licenses. See <http://cs.uky.edu/~tarik/papers/geofacial/> for additional information regarding the original images.

Author details

¹Department of Computer Science, University of Kentucky, 329 Rose St., 40506 Lexington, KY, USA. ²Department of Computer Science, University of North Carolina at Charlotte, 9201 University City Blvd., 28213 Charlotte, NC, USA.

References

- Lewontin, R.C., Freeman, W.: Human Diversity. Scientific American Library, New York (1982) [1](#), [5](#)
- Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S.M., Szeliski, R.: Building Rome in a Day. Communications of the ACM **54**(10), 105–112 (2011) [2](#)
- Zhang, H., Korayem, M., Crandall, D.J., LeBuhn, G.: Mining Photo-Sharing Websites to Study Ecological Phenomena. In: International World Wide Web Conference, pp. 749–758 (2012) [2](#)
- Leung, D., Newsam, S.: Proximate Sensing: Inferring What-is-Where from Georeferenced Photo Collections. In: IEEE Conference on Computer Vision and Pattern Recognition (2010) [2](#)
- Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A.A.: What Makes Paris Look like Paris? ACM Transactions on Graphics (SIGGRAPH) **31**(4) (2012) [2](#)
- Shen, X., Lin, Z., Brandt, J., Wu, Y.: Detecting and aligning faces by image retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition (2013) [2](#)
- Scherbaum, K., Petterson, J.: Fast face detector training using tailored views. In: IEEE International Conference on Computer Vision (2013) [2](#)
- Rudovic, O., Pantic, M.: Shape-constrained gaussian process regression for facial-point-based head-pose normalization. In: IEEE International Conference on Computer Vision (2011) [2](#)
- Xiong, X., De la Torre, F.: Supervised descent method and its applications to face alignment. In: IEEE Conference on Computer Vision and Pattern Recognition (2013) [2](#), [3](#), [5](#)
- Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and Simile Classifiers for Face Verification. In: IEEE International Conference on Computer Vision (2009) [2](#)
- Parikh, D., Grauman, K.: Relative Attributes. In: IEEE International Conference on Computer Vision (2011) [2](#)
- Yi, D., Lei, Z., Li, S.Z.: Towards pose robust face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (2013) [2](#)

13. Cao, X., Wipf, D., Wen, F., Duan, G.: A practical transfer learning algorithm for face verification. In: IEEE International Conference on Computer Vision (2013) [2](#)
14. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (2009) [2](#)
15. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. International Journal of Computer Vision (2008) [2](#)
16. Torralba, A., Fergus, R., Freeman, W.T.: 80 million tiny images: A large data set for nonparametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2008) [2](#)
17. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision (2010) [2](#)
18. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. Technical report, California Institute of Technology (2007) [2](#)
19. Jacobs, N., Roman, N., Pless, R.: Consistent temporal variations in many outdoor scenes. In: IEEE Conference on Computer Vision and Pattern Recognition (2007) [2](#)
20. Narasimhan, S., Wang, C., Nayar, S.: All the images of an outdoor scene. In: European Conference on Computer Vision, (2002) [2](#)
21. Avrithis, Y., Kalantidis, Y., Tolias, G., Spyrou, E.: Retrieving landmark and non-landmark images from community photo collections. In: ACM-MM (2010) [2](#)
22. Tolias, G., Avrithis, Y.: Speeded-up, relaxed spatial matching. In: IEEE International Conference on Computer Vision (2011) [2](#)
23. Weyand, T., Hosang, J., Leibe, B.: An evaluation of two automatic landmark building discovery algorithms for city reconstruction. In: European Conference on Trends and Topics in Computer Vision (2010) [2](#)
24. Hays, J., Efros, A.A.: im2gps: estimating geographic information from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition (2008) [2](#)
25. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence (2000) [2](#)
26. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst (2007) [2](#)
27. Jain, V., Learned-Miller, E.: Fddb: A benchmark for face detection in unconstrained settings. Technical report, University of Massachusetts, Amherst (2010) [2](#)
28. Crandall, D.J., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: International World Wide Web Conference (2009) [2](#)
29. Serdyukov, P., Murdock, V., Van Zwol, R.: Placing flickr photos on a map. In: SIGIR (2009) [2](#)
30. Fang, Q., Sang, J., Xu, C.: Giant: Geo-informative attributes for location recognition and exploration. In: ACM-MM (2013) [2](#)
31. Kalogerakis, E., Vesselova, O., Hays, J., Efros, A.A., Hertzmann, A.: Image sequence geolocation with human travel priors. In: IEEE International Conference on Computer Vision (2009) [2](#)
32. Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A.A.: What makes paris look like paris? ACM Transactions on Graphics (TOG) (2012) [2](#)
33. Lee, S., Zhang, H., Crandall, D.J.: Predicting geo-informative attributes in large-scale image collections using convolutional neural networks. In: IEEE Winter Conference on Applications of Computer Vision (2015) [2](#)
34. Chang, C.-C., Lin, C.-J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27–12727 (2011). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm> [3](#), [8](#)
35. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of cognitive neuroscience, 71–86 (1991) [4](#)
36. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. The Journal of Machine Learning Research 11, 19–60 (2010) [4](#)
37. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2000) [6](#)
38. Etemad, K., Chellappa, R.: Discriminant analysis for recognition of human face images. JOSA A 14(8), 1724–1733 (1997) [8](#)
39. Fanelli, G., Dantone, M., Van Gool, L.: Real time 3d face alignment with random forests-based active appearance models. In: Automatic Face and Gesture Recognition (FG), International Conference and Workshops On (2013). IEEE [8](#)
40. Albiol, A., Monzo, D., Martin, A., Sastre, J., Albiol, A.: Face recognition using hog–ebgm. Pattern Recognition Letters 29(10), 1537–1543 (2008) [8](#)
41. Déniz, O., Bueno, G., Salido, J., De la Torre, F.: Face recognition using histograms of oriented gradients. Pattern Recognition Letters 32, 1598–1603 (2011) [8](#)
42. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. PAMI 28(12), 2037–2041 (2006) [8](#)
43. Maturana, D., Mery, D., Soto, A.: Face recognition with local binary patterns, spatial pyramid histograms and naive bayes nearest neighbor classification. In: International Conference of the Chilean Computer Science Society (2009) [8](#)
44. Robinson, D.E.: Fashions in shaving and trimming of the beard: The men of the illustrated london news, 1842–1972. American Journal of Sociology, 1133–1141 (1976) [10](#)
45. Fu, S., He, H., Hou, Z.: Learning race from face: A survey (2014) [10](#)

