Exploring the Geo-Dependence of Human Face Appearance

Mohammad T. Islam, Scott Workman
Computer Science
University of Kentucky
{tarik,scott}@cs.uky.edu

Hui Wu, Richard Souvenir
Computer Science
UNC Charlotte
{hwu13,souvenir}@uncc.edu

Nathan Jacobs
Computer Science
University of Kentucky
jacobs@cs.uky.edu

Abstract

The expected appearance of a human face depends strongly on age, ethnicity and gender. While these relationships are well-studied, our work explores the little-studied dependence of facial appearance on geographic location. To support this effort, we constructed GeoFaces, a large dataset of geotagged face images. We examine the geo-dependence of Eigenfaces and use two supervised methods for extracting geo-informative features. The first, canonical correlation analysis, is used to find location-dependent component images as well as the spatial direction of most significant face appearance change. The second, linear discriminant analysis, is used to find countries with relatively homogeneous, yet distinctive, facial appearance.

1. Introduction

What does the average person from Beijing look like and how do they differ from people in Mumbai or New York City? A computational model of such variations could be used in a range of applications, such as creating realistic content for games and animated movies, or studying patterns of human migration and travel. Our goal in this work is to explore and analyze the geo-spatial structure in publicly-available imagery. To support this effort, we construct a database of geotagged frontal face patches by gathering images and extracting aligned frontal face patches. This results in a dataset of approximately 248 000 geotagged faces, which, to our knowledge, is the largest publicly-available dataset of geotagged face images.

We use this dataset to explore the location-dependence of human face appearance using a variety of statistical models. We show, as one would expect, that locally weighted average images are highly dependent on geographic location. We further show that some of the coefficients of a principal component (PCA) decomposition of the dataset are location dependent, but some are not. We combine these coefficients in several ways to find the face image projections that are most related to location, using both canonical correlation analysis (CCA) and linear discriminant analysis (LDA).

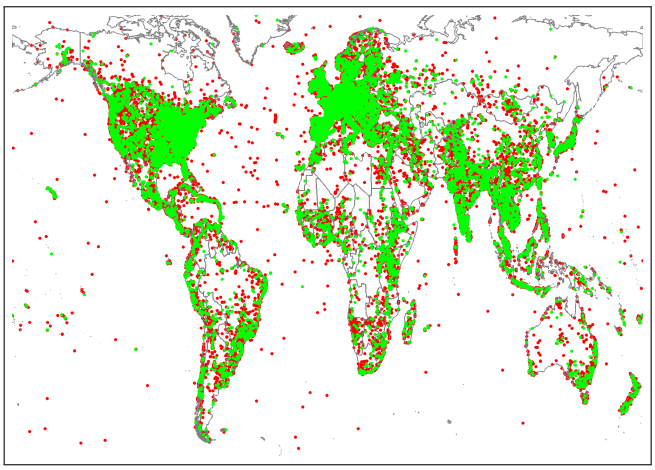


Figure 1: We construct a large database of geolocated face patches. The map depicts the geolocation of raw patches (red) and a filtered set of frontal face patches (green). These images serve as a foundation for our analysis of the geo-dependence of human face appearance.

1.1. Related Work

Despite limited research on the geo-dependence of face appearance, there is a significant amount of research in a number of related problems.

Large-Scale Image Datasets Many large-scale image datasets have been introduced recently to advance research in vision-related areas such as object detection [4, 19, 24], classification [8, 10], and outdoor scene analysis [13, 16]. Similarly, for the task of image geolocalization, datasets such as [1, 23, 26, 11] contain millions of geotagged images collected from the public domain. Most existing face image datasets are targeted at facial recognition and have size on the order of several thousands [17, 12, 14]. To our knowledge, the large scale geotagged face image dataset we have constructed is the first of its kind.

(a) Raw detections (b) After alignment (c) After filtering

Figure 2: Representative images at each stage of our dataset construction process.

Face Image Analysis The human face is one of the most intensely studied objects, with active research on a variety of subproblems, including: detection [22, 20], recognition and verification [28, 2] and pose normalization [18, 27, 22]. Our focus is on the higher-level question, "How does expected face appearance depend on location?" To our knowledge, this is the first work that addresses this question.

Geo-Dependence of Scene Appearance The relationship between scene appearance and location has been explored in great detail [11, 5, 3, 21, 9, 6, 15]. These methods attempt to automatically discover and interpret geo-informative features from geotagged imagery. We extend this line of research and examine the geo-dependence of facial appearance.

2. GeoFaces: A Dataset of Geolocated Faces

We use publicly available images to construct a large dataset of geolocated face patches. Starting with geotagged images labeled with face-related strings, the data is processed to obtain a large set of frontal face image patches. This section describes, in detail, the dataset collection and construction process.

Collecting and Aligning Face Patches We downloaded geotagged imagery from Flickr[1] with face-related tags (*e.g.*, face, portrait, men, family, friends). For each image, a commercial face detector[2] is used to identify faces and fiducial points. The detector is tuned to find frontal (or nearly frontal) faces. This resulted in a set of 2.65 million geolocated face patches. Each face patch is automatically aligned to a common reference frame using a similarity transform, with eye centers as control points.

Dataset Validation and Cleanup To eliminate false positives and non-frontal faces, we apply a filtering approach that follows the method of [7] where the similarity score of a point is the sum of similarities to neighboring points in feature space. Unlike [7], where the feature representation is used both to define the neighborhood and measure similarity, we manually define the neighborhood. The *frontalness score*, s_i, of a putative face patch, $\{f_i\}$ represents the similarity of $\{f_i\}$ to a set of representative frontal face patches, $\{F_j\}$, manually selected from the database. The score, s_i, is defined as follows:

$$s_i = \sum_{j \in \{F\}} \kappa(f_i, F_j), \tag{1}$$

where κ is a similarity measure between image patches. To reduce the effects of image illumination and contrast change, each image patch is represented by its gradient image, and κ is the linear correlation between the gradient images. The frontalness score for each image is used to filter out non-frontal faces by keeping only the top 10% of face patches.

GeoFaces Summary Figure 2 demonstrates the effect of image processing, showing representative patches in the dataset after alignment and filtering. The resulting dataset, called GeoFaces, contains approximately $248\,000$ geolocated face patches from around the world. This dataset will evolve as we collect more images and improve the methods for detecting, aligning and filtering. The full dataset, including detected, aligned and filtered face patches, is freely available online[3].

3. Geo-Dependence of Eigenfaces

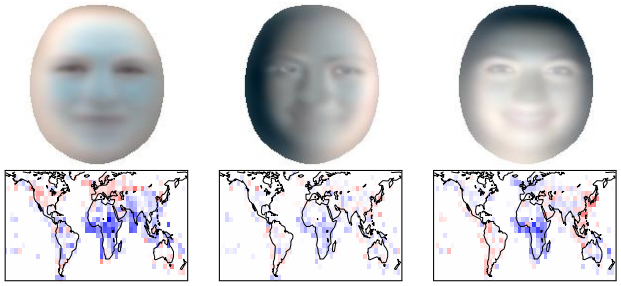
Eigenfaces [25] are a frequently used statistical model in facial image analysis. We explore the relationship between the output Eigenface coefficients and geographic location using the GeoFaces dataset.

For computational efficiency and to remove background clutter, each face patch is resized to 200×200 and pixels outside of a manually specified elliptical region are ig-

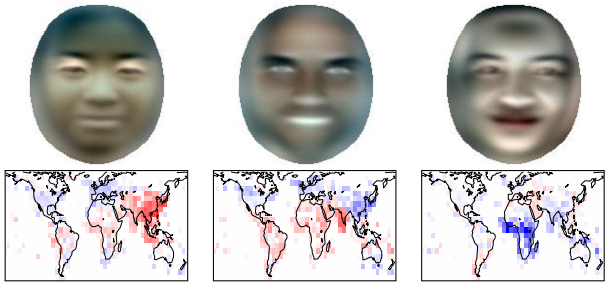
[1] http://flickr.com

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

[3] Available at http://geofaces.csr.uky.edu



(a) PCA 1 (b) PCA 2 (c) PCA 3



(d) CCA 1 (e) CCA 2 (f) CCA 3

Figure 3: The top three PCA component images and location-dependent CCA component images from our global indicator variable experiment. 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.

nored. The shape of this region can be seen from the faces in Figure 3. Using PCA, we obtain the top Eigenfaces for the database. Figure 3(a-c) show 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; it appears to encode the direction of lighting on the face.

We estimate 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 compute 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 $\sigma = 5°$) and reconstructing the corresponding image. The average images for a set of locations around the globe are shown in Figure 4. Locations with a low number of images are omitted. From these images, the location-dependence of face appearance can be seen. To further demonstrate the utility of image pre-processing, Figure 5 shows a set of locally-weighted average images before and after our alignment and filtering steps.

4. Supervised Feature Analysis

In contrast to the unsupervised exploration with Eigenfaces, in this section, we use two supervised learning methods, canonical correlation analysis (CCA) and linear discriminant analysis (LDA), to extract location dependent features and support observations about facial appearance in various world regions.

4.1. Location-Dependent Component Images

CCA is a multivariate statistical tool for exploring relationships between paired sets of variables that can be used to find a set of component images and corresponding coefficients that are strongly location-dependent. 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 $(u_1^\mathsf{T} A, v_1^\mathsf{T} B)$ are maximally correlated. That is, it finds u_1, v_1 such that $\rho = corr(u_1^\mathsf{T} A, v_1^\mathsf{T} B)$ is maximized. The pair of vectors (u_1, v_1) is called the first canonical pair, and subsequent canonical pairs are defined similarly.

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 results 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 3(d-f). 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.

4.2. 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 perform CCA on the images from small spatial areas and use a linear model for geolocation.

For a given location, l, we create a filtered dataset, A_l, of all faces within $10°$. We use 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 6 shows the resulting direction field for this analysis over multiple locations overlaid on a world map. Visually, the computed gradients

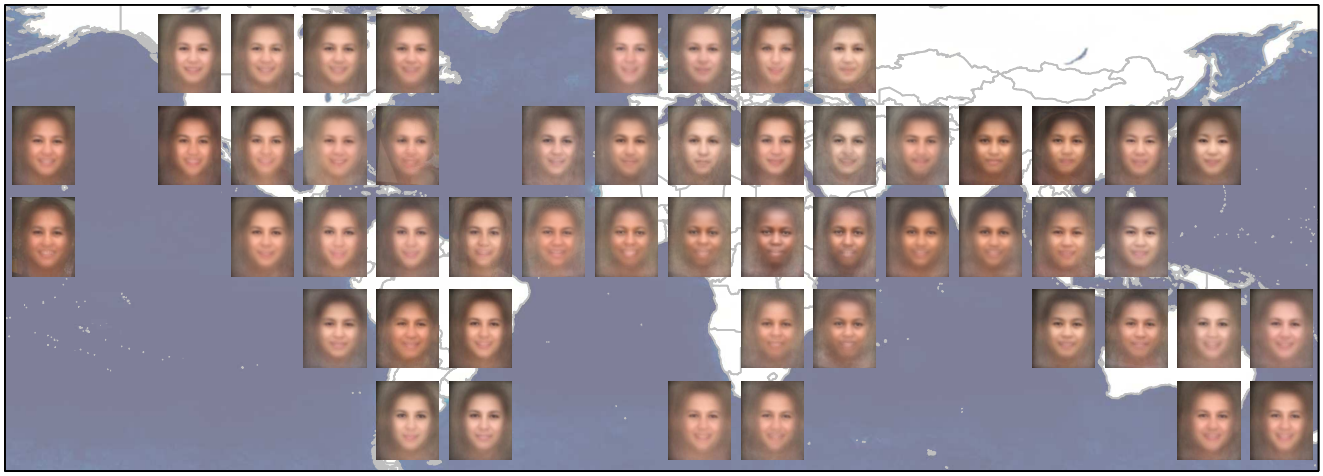


Figure 4: A map of locally-weighted average images. The spatial similarities demonstrate the geo-dependence of facial appearance.

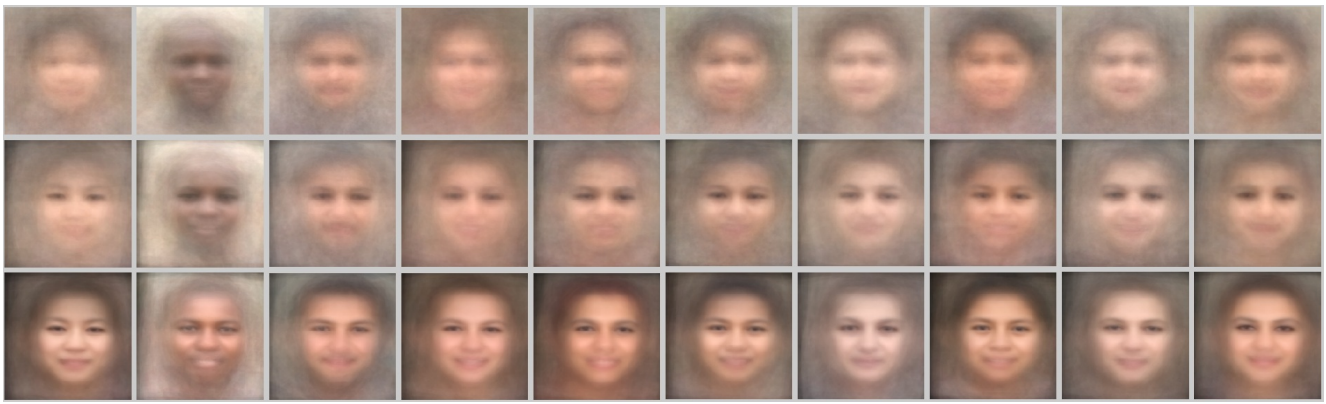


Figure 5: Visualization of the effect of alignment and filtering on country-level average images. The original set of face patches (top), after alignment (middle) and after filtering out non-frontal faces (bottom).

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.

4.3. Location-Dependent Face Classification

We use linear discriminant analysis (LDA) to classify face patches into geospatial regions and use the performance of the resulting models to better understand facial appearance variation. We perform our analysis with partitions at the level of continents, sub-continents and countries.

We train a one-vs-all (OVA) classifier for each continent with the top 50 PCA coefficients as features. For each class, the training set for each consists of 500 positive examples and 500 negative examples sampled from the rest of the world. The remainder of the dataset is used for testing.

Continent	Asia	Africa	Europe	Americas	Oceania
Accuracy	0.73	0.68	0.59	0.55	0.49

Table 1: Accuracy of continental LDA classifiers. The respective classifiers for Africa and Asia are significantly better at predicting the location of a face image.

Table 1 shows the overall accuracy of the classifiers. The respective classifiers for Africa and Asia are significantly more accurate at predicting the geotag of a face image. Figure 7 shows the conditional accuracies of these classifiers. While there are many sources of error, we speculate that one significant source is the large spatial area and diversity of a continent.

We trained classifiers on 23 sub-continental regions[4]. As

[4] http://unstats.un.org/unsd/methods/m49/m49regin.htm

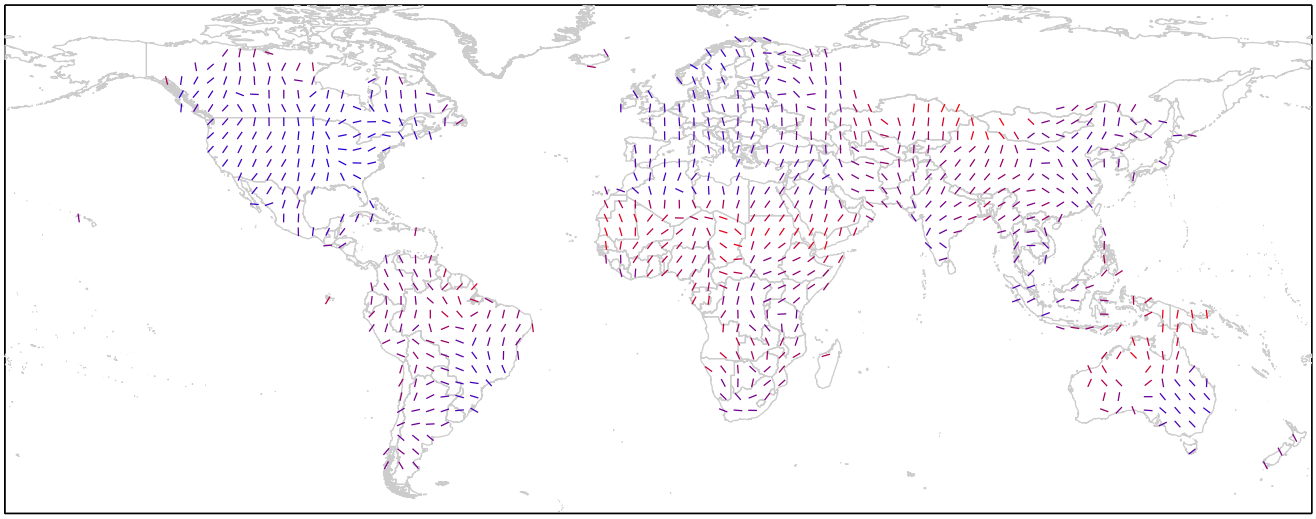


Figure 6: A direction field estimated using CCA between face appearance and geographic location in local neighborhoods. The lines show the cardinal direction that is most correlated with facial appearance change. The lines are color coded by the correlation coefficients, where blue (red) indicate low (high) correlation.

	S. Am.	C. Am.	S. Af.	N. Am.	S. Eu.	Aus,NZ	W. As.	E. Eu.	N. Eu.	W. Eu.	E. As.	SE. As.	C. As.	W. Af.	E. Af.	N. Af.	M. Af.	Carib.	S. As.	Mal.
S. Am.	0.57	0.54	0.36	0.43	0.50	0.38	0.48	0.37	0.32	0.36	0.28	0.42	0.40	0.32	0.27	0.39	0.22	0.49	0.50	0.42
C. Am.	0.54	0.59	0.44	0.47	0.44	0.41	0.47	0.33	0.35	0.32	0.28	0.43	0.39	0.43	0.34	0.45	0.29	0.53	0.50	0.41
S. Af.	0.36	0.44	0.63	0.57	0.41	0.56	0.38	0.39	0.55	0.43	0.16	0.28	0.32	0.50	0.51	0.47	0.49	0.57	0.35	0.54
N. Am.	0.43	0.47	0.57	0.58	0.44	0.55	0.39	0.39	0.52	0.43	0.26	0.31	0.34	0.33	0.31	0.35	0.25	0.49	0.30	0.42
S. Eu.	0.50	0.44	0.41	0.44	0.61	0.44	0.51	0.48	0.45	0.45	0.19	0.27	0.34	0.24	0.22	0.39	0.16	0.44	0.32	0.36
Aus,NZ	0.38	0.41	0.56	0.55	0.44	0.57	0.37	0.41	0.54	0.46	0.30	0.33	0.37	0.43	0.43	0.41	0.43	0.47	0.29	0.46
W. As.	0.48	0.47	0.38	0.39	0.51	0.37	0.58	0.43	0.39	0.42	0.25	0.35	0.46	0.25	0.22	0.46	0.16	0.39	0.42	0.29
E. Eu.	0.37	0.33	0.39	0.39	0.48	0.41	0.43	0.62	0.50	0.55	0.19	0.21	0.38	0.17	0.23	0.33	0.20	0.29	0.25	0.29
N. Eu.	0.32	0.35	0.55	0.52	0.45	0.54	0.39	0.50	0.63	0.55	0.15	0.18	0.31	0.22	0.25	0.34	0.20	0.37	0.20	0.32
W. Eu.	0.36	0.32	0.43	0.43	0.45	0.46	0.42	0.55	0.55	0.60	0.17	0.22	0.33	0.25	0.31	0.37	0.30	0.33	0.27	0.34
E. As.	0.28	0.28	0.16	0.26	0.19	0.30	0.25	0.19	0.15	0.17	0.74	0.62	0.50	0.20	0.14	0.26	0.18	0.18	0.32	0.17
SE. As.	0.42	0.43	0.28	0.31	0.27	0.33	0.35	0.21	0.18	0.22	0.62	0.68	0.53	0.43	0.35	0.44	0.39	0.36	0.54	0.41
C. As.	0.40	0.39	0.32	0.34	0.34	0.37	0.46	0.38	0.31	0.33	0.50	0.53	0.62	0.35	0.32	0.45	0.38	0.29	0.46	0.35
W. Af.	0.32	0.43	0.50	0.33	0.24	0.43	0.25	0.17	0.22	0.25	0.20	0.43	0.35	0.76	0.70	0.59	0.73	0.59	0.60	0.65
E. Af.	0.27	0.34	0.51	0.31	0.22	0.43	0.22	0.23	0.25	0.31	0.14	0.35	0.32	0.70	0.74	0.58	0.75	0.53	0.57	0.66
N. Af.	0.39	0.45	0.47	0.35	0.39	0.41	0.46	0.33	0.34	0.37	0.26	0.44	0.45	0.59	0.58	0.65	0.58	0.52	0.62	0.54
M. Af.	0.22	0.29	0.49	0.25	0.16	0.43	0.16	0.20	0.20	0.30	0.18	0.39	0.38	0.73	0.75	0.58	0.79	0.53	0.61	0.61
Carib.	0.49	0.53	0.57	0.49	0.44	0.47	0.39	0.29	0.37	0.33	0.18	0.36	0.29	0.59	0.53	0.52	0.53	0.64	0.51	0.56
S. As.	0.50	0.50	0.35	0.30	0.32	0.29	0.42	0.25	0.20	0.27	0.32	0.54	0.46	0.60	0.57	0.62	0.61	0.51	0.73	0.54
Mal.	0.42	0.41	0.54	0.42	0.36	0.46	0.29	0.29	0.32	0.34	0.17	0.41	0.35	0.65	0.66	0.54	0.61	0.56	0.54	0.71

Figure 8: The conditional accuracies of sub-continental LDA classifiers. The numbers in a given row indicate how similar different regions are to the training region. The block diagonal structure of the table shows four distinct clusters corresponding to the most dominant ethnic groups in the world.

Sub-continent	Middle Africa	Eastern Asia	Eastern Africa	Western Africa	Southern Asia	South-Eastern Asia
Accuracy	0.8140	0.8081	0.7626	0.7559	0.7254	0.7015

Table 2: Sub-continental regions sorted by overall classifier accuracy.

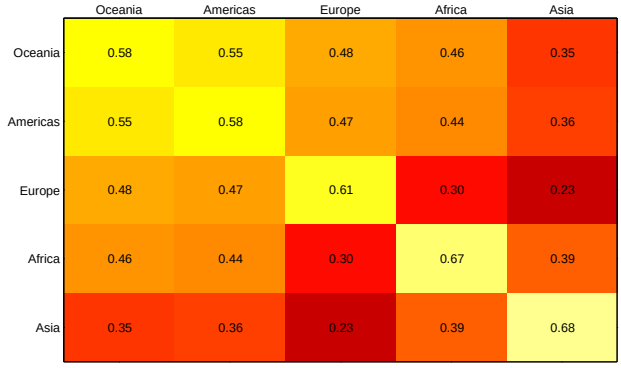


Figure 7: The conditional accuracies of continental LDA classifiers. The numbers in a given row indicate the false positive rates of the corresponding classifier in different continents. For example, the number 0.47 inside the cell located at (row 5, column 4) indicates that the classifier trained on faces from Europe classified 47% of the faces from the Americas as Europeans.

with the continent level classifiers, there is a wide spread in the classification accuracy across the regions. Table 2 lists the sub-continent regions and classification accuracies of the classifiers with the best performance. In Figure 8 we show the conditional accuracies of the sub-continental regions. The block diagonal structure of the matrix shows four distinct clusters corresponding to the most dominant ethnic groups in the world. For example, Australia, New Zealand, South Africa, Europe, North America and West Asia form a cluster indicating that faces from these regions look similar. Finally, we designed OVA classifiers at the country-level, mainly a political, rather than geographic, partition. Table 3 shows the ten countries with the highest classifier accuracy. The top 30 classifiers in terms of classification accuracy represent countries from Africa or Asia. Figure 9 shows maps for eight target countries color-coded based on the percentage of faces detected by the country-specific classifier.

5. Conclusion

We constructed the largest publicly available dataset of geotagged face patches, which we hope will spur further research into understanding how human facial appearance depends on geographic location. In this initial investigation, 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 the future, we plan to improve the facial feature extraction process to grow the dataset and also apply our model to inference problems, such as geo-locating untagged images containing faces.

References

[1] Y. Avrithis, Y. Kalantidis, G. Tolias, and E. Spyrou. Retrieving landmark and non-landmark images from community photo collections. In *ACM-MM*, 2010. 1

[2] Xudong Cao, David Wipf, Fang Wen, and Genquan Duan. A practical transfer learning algorithm for face verification. In *IEEE International Conference on Computer Vision*, 2013. 2

[3] David J Crandall, Lars Backstrom, Daniel Huttenlocher, and Jon Kleinberg. Mapping the world's photos. In *International World Wide Web Conference*, 2009. 2

[4] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 1

[5] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A Efros. What makes paris look like paris? *ACM Transactions on Graphics (TOG)*, 2012. 2

[6] Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing. A latent variable model for geographic lexical variation. In *EMNLP*, 2010. 2

[7] Eleazar Eskin, Andrew Arnold, Michael Prerau, Leonid Portnoy, and Sal Stolfo. A geometric framework for unsupervised anomaly detection. In *Applications of data mining in computer security*. 2002. 2

[8] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 2010. 1

[9] Quan Fang, Jitao Sang, and Changsheng Xu. Giant: Geo-informative attributes for location recognition and exploration. In *ACM-MM*, 2013. 2

Country ISO Code	ETH	MMR	BEN	AGO	BGD	SDN	RWA	PRK	JPN	VNM
Accuracy	0.8231	0.8034	0.7918	0.7918	0.7769	0.7763	0.7659	0.7644	0.7640	0.7547

(a) Most distinctive countries.

Country ISO Code	LBR	GMB	LUX	ARM	PSE	BIH	UZB	VAT	GLP	KGZ
Accuracy	0.5058	0.4867	0.4817	0.4632	0.4589	0.4504	0.4500	0.4401	0.4367	0.4201

(b) Least distinctive countries.

Table 3: Countries sorted by overall classifier accuracy.

[10] Gregory Griffin, Alex Holub, and Pietro Perona. Caltech-256 object category dataset. Technical report, California Institute of Technology, 2007. 1

[11] James Hays and Alexei A. Efros. im2gps: estimating geographic information from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 1, 2

[12] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst, 2007. 1

[13] Nathan Jacobs, Nathaniel Roman, and Robert Pless. Consistent temporal variations in many outdoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007. 1

[14] Vidit Jain and Erik Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical report, University of Massachusetts, Amherst, 2010. 1

[15] Evangelos Kalogerakis, Olga Vesselova, James Hays, Alexei A Efros, and Aaron Hertzmann. Image sequence geolocation with human travel priors. In *IEEE International Conference on Computer Vision*, 2009. 2

[16] SrinivasaG. Narasimhan, Chi Wang, and ShreeK. Nayar. All the images of an outdoor scene. In *European Conference on Computer Vision*. 2002. 1

[17] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000. 1

[18] O. Rudovic and M. Pantic. Shape-constrained gaussian process regression for facial-point-based head-pose normalization. In *IEEE International Conference on Computer Vision*, 2011. 2

[19] Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. Labelme: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 2008. 1

[20] Kristina Scherbaum and James Petterson. Fast face detector training using tailored views. In *IEEE International Conference on Computer Vision*, 2013. 2

[21] Pavel Serdyukov, Vanessa Murdock, and Roelof Van Zwol. Placing flickr photos on a map. In *SIGIR*, 2009. 2

[22] Xiaohui Shen, Zhe Lin, Jonathan Brandt, and Ying Wu. Detecting and aligning faces by image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 2

[23] G. Tolias and Y. Avrithis. Speeded-up, relaxed spatial matching. In *IEEE International Conference on Computer Vision*, 2011. 1

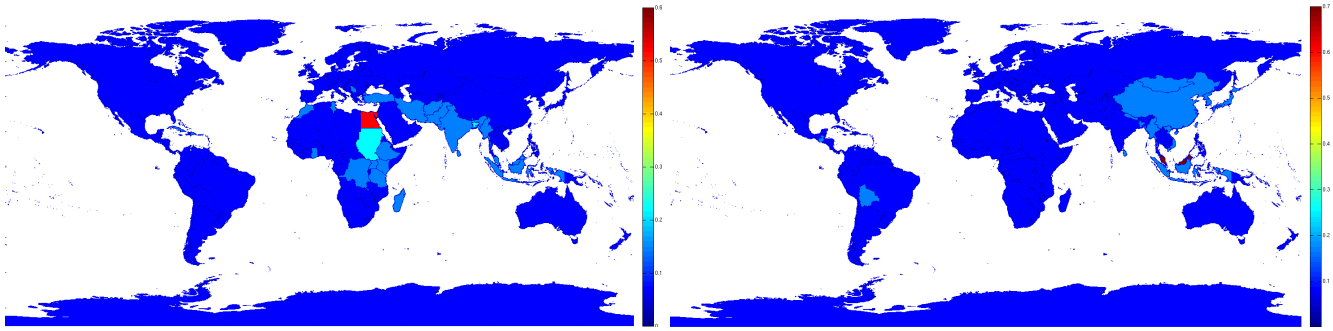
[24] Antonio Torralba, Robert Fergus, and William T Freeman. 80 million tiny images: A large data set for non-parametric object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008. 1

[25] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, pages 71–86, 1991. 2

[26] Tobias Weyand, Jan Hosang, and Bastian Leibe. An evaluation of two automatic landmark building discovery algorithms for city reconstruction. In *European conference on Trends and Topics in Computer Vision*, 2010. 1

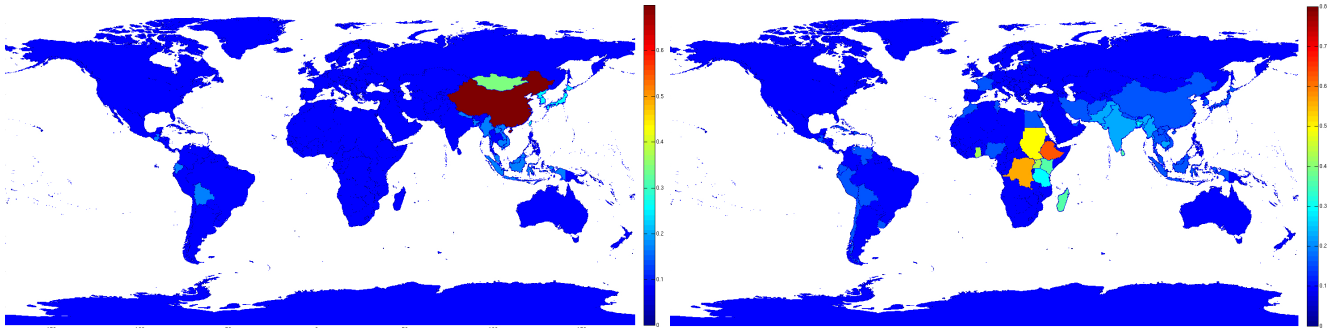
[27] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 2

[28] Dong Yi, Zhen Lei, and Stan Z Li. Towards pose robust face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 2



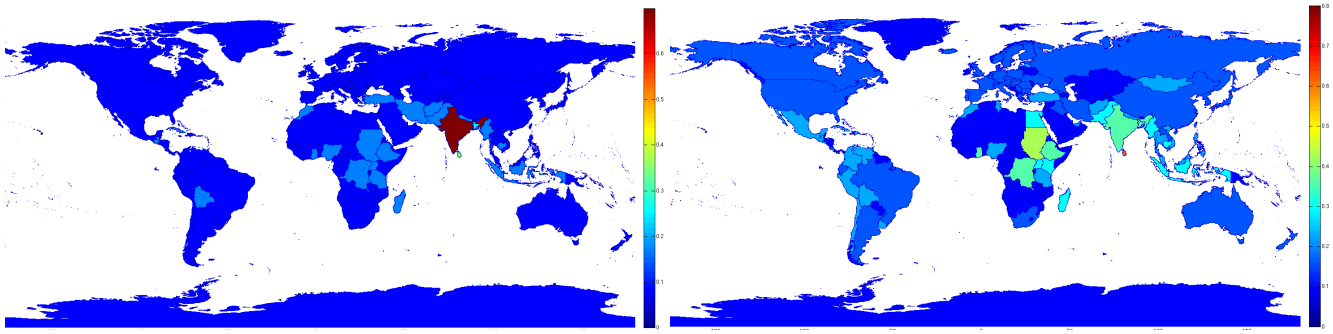
(a) Egypt

(b) Malaysia



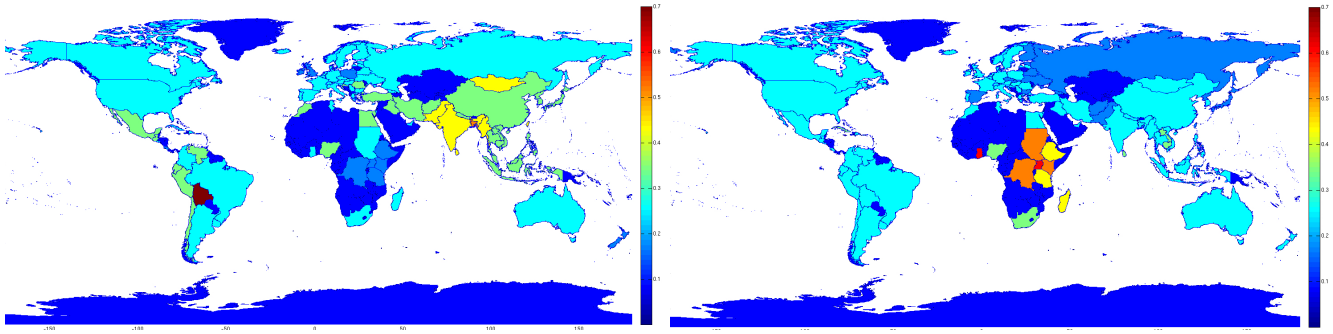
(c) China

(d) Ethiopia



(e) India

(f) Sri Lanka



(g) Bolivia

(h) Ghana

Figure 9: The maps depict the country-level affinity of facial appearance for selected target countries.