Learning High-level Judgments of Urban Perception

Vicente Ordonez · Tamara L. Berg
University of North Carolina at Chapel Hill

http://www.cs.unc.edu/~vicente/urban
European Conference on Computer Vision (ECCV) 2014. Zurich, Switzerland

Introduction

- We apply computer vision to predict human perceptions of place to potentially uncover *the image of the city* [1]
- We analyze the generalization of vision models for urban perception to other regions, cities or points in time.
- We propose a collective prediction based on geographical smoothing.

Place Pulse 1.0 Dataset [2]

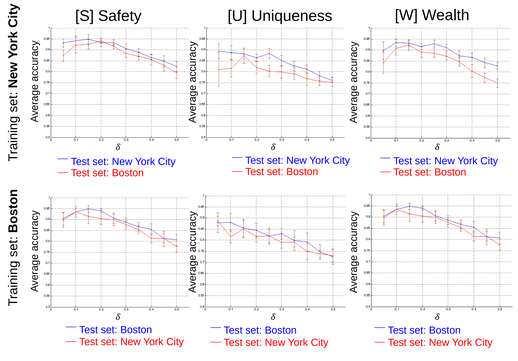
Which place looks *safer*? *wealthier*? more *unique*?

$$q_{i,k} = \frac{10}{3}\left(W_{i,k} + \frac{1}{w_{i,k}}\sum_{j_1=1}^{w_{i,k}} W_{j_1,k} - \frac{1}{l_{i,k}}\sum_{j_2=1}^{l_{i,k}} L_{j_2,k}\right)$$

Score images based on the proportion (*W, L*) and number (*w, l*) of times they were chosen ("won") or disregarded ("lost") over other images and the scores of such images.

Classification

Feature representation: Fisher Vectors + SIFT (FV)

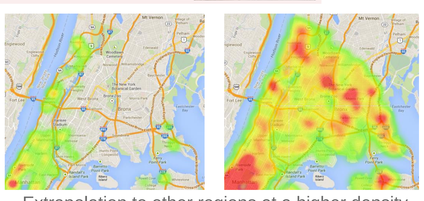


Regression

Training data	Metric	Test on New York: Gist	SIFT+ FV	DeCAF [3]	Test on Boston: Gist	SIFT+ FV	DeCAF [3]
New York	Safety	0.64	**0.69**	0.68	0.64	0.65	**0.70**
	Unique	0.53	0.52	**0.55**	0.50	0.44	**0.52**
	Wealth	0.61	0.65	**0.65**	0.57	0.60	**0.66**
Boston	Safety	0.60	0.62	**0.64**	0.67	0.67	**0.72**
	Unique	0.45	0.38	**0.46**	0.52	0.49	**0.55**
	Wealth	0.56	0.56	**0.60**	0.59	0.64	**0.68**

City: New York City

Sampled predicted safety scores using regression



Extrapolation to other regions at a higher density

Generalization Across Time

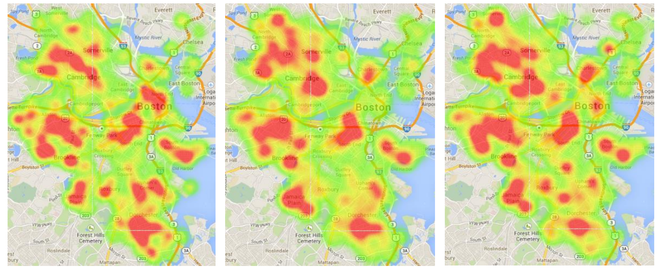
The same place in 2011 and 2013

Training using label annotations from 2011 but image data from 2013

Training data	Metric	Test on New York: Gist	SIFT+ FV	DeCAF [3]	Test on Boston: Gist	SIFT+ FV	DeCAF [3]
New York	Safety	0.54	**0.59**	0.56	0.52	0.53	**0.56**
	Unique	0.44	**0.45**	0.44	0.41	0.36	**0.44**
	Wealth	0.53	**0.57**	0.55	0.47	0.49	**0.56**
Boston	Safety	0.51	0.49	**0.52**	0.55	**0.58**	0.58
	Unique	0.40	0.35	**0.42**	0.42	0.37	**0.45**
	Wealth	0.50	0.48	**0.52**	0.52	0.54	**0.59**

Safety Heatmaps

Predictions on Boston



Wealth scores | Predicted wealth scores | Predicted wealth scores with a model trained on images of New York City.

Collective Prediction

Objective function

$$\hat{Y} = \underset{Y}{\operatorname{argmax}} \prod_i \phi_1(y_i|x_i, w_s) \prod_{i,j\in E} \phi_2(y_i, y_j|x_i, x_j, p_i, p_j, \alpha_1, \alpha_2)$$

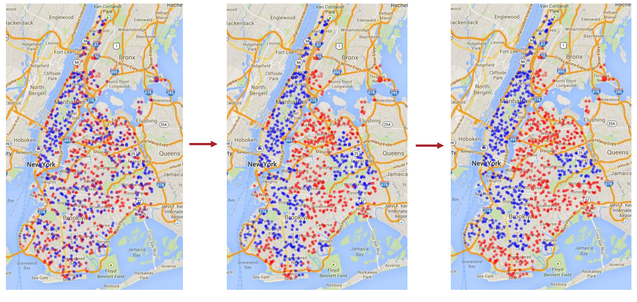
Regression-based potential

$$-\ln \phi_1 = y_i w_S^T x_i$$

Smoothing term

$$-\ln \phi_2 = 1[y_i = y_j] \cdot \left(\frac{\alpha_1}{\|x_i - x_j\|} + \frac{\alpha_2}{\|p_i - p_j\|}\right)$$

Image similarity | Geographic affinity

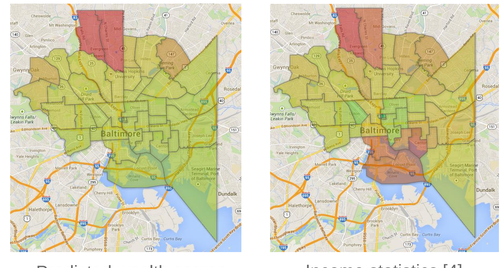


	¬Safe	¬Unique	¬Wealth
Isolated prediction (SIFT + FV)	0.6077	0.4420	0.5755
Isolated prediction (DeCAF [2])	0.5929	0.4652	0.5613
Collective prediction (SIFT + FV)	0.6069	0.4457	0.5700
Collective prediction (DeCAF [2])	0.6089	0.4777	0.5545

Generalization to other cities

City: Chicago

Geographical Aggregation



Predicted wealth scores | Income statistics [4]

References

[1] Lynch K. The image of the city. Vol. 11. MIT Press (1960)

[2] Salesses P., Schechtner K., Hidalgo C.A. The Collaborative Image of the City. PLOS ONE (2013)

[3] Donahue J., Jia Y., Vinyals O., Hoffman J., Zhang N., Tzeng E., Darrell T. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. Arxiv e-prints (Oct 2013).

[4] Baltimore City Health Department (2011)