Recognizing City Identity via Attribute Analysis of Geo-tagged Images

Bolei Zhou[1], Liu Liu[2], Aude Oliva[1], and Antonio Torralba[1]

[1] Computer Science and Artificial Intelligence Laboratory
[2] Department of Urban Studies and Planning
Massachusetts Institute of Technology
{bolei,lyons66,oliva,torralba}@mit.edu

Abstract. After hundreds of years of human settlement, each city has formed a distinct identity, distinguishing itself from other cities. In this work, we propose to characterize the identity of a city via an attribute analysis of 2 million geo-tagged images from 21 cities over 3 continents. First, we estimate the scene attributes of these images and use this representation to build a higher-level set of 7 city attributes, tailored to the form and function of cities. Then, we conduct the city identity recognition experiments on the geo-tagged images and identify images with salient city identity on each city attribute. Based on the misclassification rate of the city identity recognition, we analyze the visual similarity among different cities. Finally, we discuss the potential application of computer vision to urban planning.

Keywords: Geo-tagged image analysis, attribute, spatial analysis, city identity, urban planning

1 Introduction

In Kevin Lynch's work *The Image of The City*, a city is described as a form of temporal art in vast scale. Over hundreds of years of human settlement, different cities have formed distinctive identities. City identity is defined as the sense of a city that distinguishes itself from other cities [22]. It appears in every aspects of urban life. For instance, Fig.1 shows photos taken by people in different cities, organized by different urban dimensions. Although there are no symbolic landmarks in those images, people who have lived in these cities or even just visited there can tell which image come from which cities. Such a capability suggests that some images from a city might have unique identity information that different urban observers may share knowledge of.

Akin to objects and scenes, cities are visual entities that differ in their shape and function [16, 22]. As the growth of cities is highly dynamic, urban researchers and planners often describe cities through various attributes: they use the proportion of green space to evaluate living quality, take the land use to reflect transportation and social activity, or rely on different indicators to evaluate the urban development [26, 16]. Here, we propose to characterize city identity

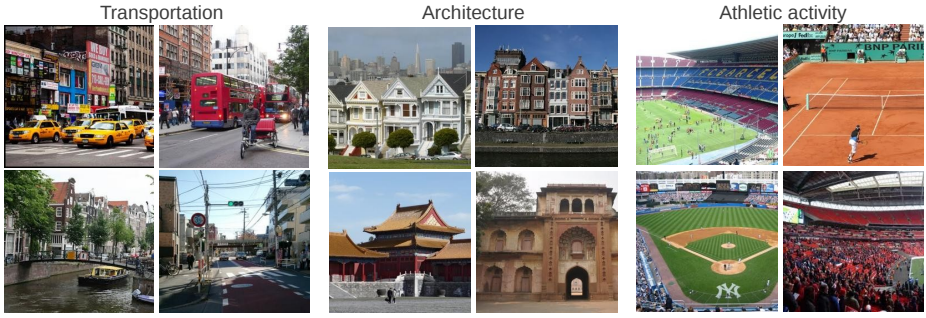

Fig. 1. City identity permeates every aspect of urban life. Can you guess from which cities these photos have been taken? Answer is below.[4]

via attribute analysis of geo-tagged images from photo-sharing websites. Photo-sharing websites like Instagram, Flickr, and Panoramio have amassed about 4 billion geo-tagged images, with over 2 million new images uploaded every day by users manually. These images contain a huge amount of information about the cities, which are not only used for landmark detection and reconstruction [12, 3], but are also used to monitor ecological phenomena [29] and human activity [9] occurring in the city.

In this work a set of 7 high-level attributes is used to describe the spatial form of a city (amount of vertical buildings, type of architecture, water coverage, and green space coverage) and its social functionality (transportation network, athletic activity, and social activity). These attributes characterize the specific identity of various cities across Asia, Europe, and North America. We first collect more than 2 million geo-tagged images from 21 cities and build a large scale geo-tagged image database: the City Perception Database. Then based on the SUN attribute database [20] and deep learning features [5], we train the state-of-the-art scene attribute classifiers. The estimated scene attributes of images are further merged into 7 city attributes to describe each city within related urban dimensions. We conduct both city identity recognition experiment ("is it New York or Prague?") and city similarity estimation ("how similar are New York and Prague?"). Moreover, we discuss the potential application of our study to urban planning.

1.1 Related Work

The work on the geo-tagged images has received lots of attention in recent years. Landmarks of cities and countries are discovered, recognized, and reconstructed from large image collections [2, 12, 30, 13, 3]. Meanwhile, the IM2GPS approach [7] is used to predict image geolocation by matching visual appearance with geo-tagged images in dataset. Cross-view image matching is also used to correlate

[4] New York, London, Armsterdam, Tokyo; San Francisco, Armsterdam, Beijing, New Delhi; Barcelona, Paris, New York, London.

satellite images with ground-level information to localize images [14]. Additionally, geo-tagged images uploaded to social networking websites are also used to predict ecological phenomena [29] and people activity occurring in a city [9]. Besides, recent work [8] utilizes the visual cues of Google Street images to navigate the environment.

Our present work is inspired from discovering visual styles of architectures and objects in images [4, 11], which use mid-level discriminative patches to characterize the identity of cities. Another relevant work [24] used Google street view images to estimate the inequality of urban perception with human's labeling. However, instead of detecting landmark images of cities and discovering local discriminative patches, our work aims at analyzing the city identity of the large geo-tagged image collection in the context of semantic attributes tailored to city form and function. *Attributes* are properties observable in images that have human-designated names (e.g. smooth, natural, vertical). Attribute-based representation has shown great potential for object recognition [1, 19] and scene recognition [18, 20]. Generally human-labeled attributes act as mid-level supervised information to describe and organize images. By leveraging attribute-based representations, we map images with a wide variety of image contents, from different cities, into the same semantic space with the common attribute dimension. Altogether, our approach presents an unified framework to measure the city identity and the similarity between cities. The proposed method not only automatically identifies landmarks and typical architectural styles of cities, but also detects unique albeit inconspicuous urban objects in cities. For instance, as shown in Fig.1 our results on the transportation attribute identify red double decker buses in London and yellow cabs in New York City as the objects with salient city identity value.

2 Describing City Perception by Attributes

In this section, we introduce a novel database of geo-tagged images[5] and its statistical properties. Then we propose a set of high-level city attributes from scene attributes to describe the city's spatial form (the amount of vertical buildings, type of architecture, water coverage, and green space coverage), as well as the city's social function (transportation network, athletic activity, and social activity). Attribute classifiers are trained using ground-truth from the SUN attribute database [20]. Furthermore, we analyze how the spatial distributions of city attributes vary across the urban regions and cities.

2.1 City Perception Database

Datasets of geo-tagged images can be either collected through cropping images from Google Street View as in [4] or downloading images from photo-sharing websites like Flickr and Panoramio as in [12, 13, 3]. These two data sources have

[5] Available at `http://cityimage.csail.mit.edu`.

different properties. Images from Google Street View are taken on roads where the Google vehicle can go, so the content of these images is limited, as a lot of content related to city perceptions, such as mountains and crowded indoor scenes are missing. Here we choose geo-tagged images from photo sharing websites. Interestingly, these images are power-law distributed on city maps (see Fig.3), given that people travel in a non-uniform way around a city, visiting more often the regions with historical, attractive tour sites as well as the regions with social events. Thus, these images represent *people's perception* of the city.

We build a new geo-tagged image dataset called City Perception Database. It consists of 2,034,980 geo-tagged images from 21 cities collected from Panoramio. To diversify the dataset, cities are selected from Europe, Asia, and North America. To get the geographical groundtruth for each city, we first outline the geographical area of the city, then segment the whole area into dense 500m×500m adjacent spatial cells. Geo-locations of these cells are further pushed to the API of Panoramio to query image URLs. Finally all the images lying within the city area are downloaded and the corrupted images are filtered out. The image numbers of the database along with their spatial statistics are listed in Fig.2. The negative Z-score of the Average Nearest Neighbor Index [23] indicates that the geo-locations of these images have the highly clustered pattern. Fig.3 shows the map plotting of all the images for two cities London and San Francisco.

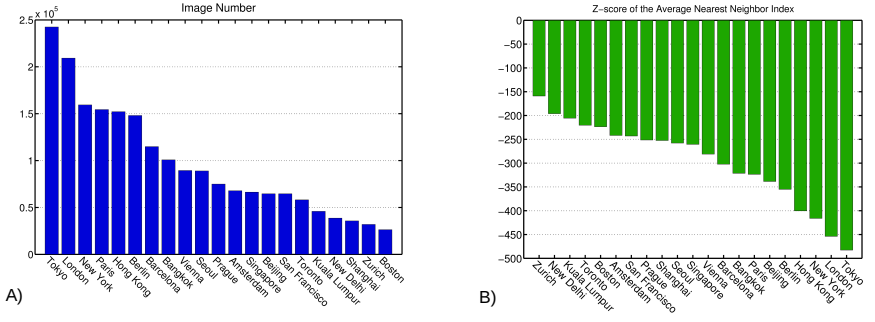

Fig. 2. A) The number of images obtained in each city. B) The Z-score of the Average Nearest Neighbor Index for each city. The more negative the value is, the more the geo-tagged images are spatially clustered. Thus images taken by people in a city are highly clustered.

2.2 From Scene Attributes to City Attributes

We propose to use attributes as a mid-level representation of images in our city perception database. Our approach is to train scene attribute classifiers, then to combine and calibrate various scene attribute classifiers into higher level city attribute classifiers.

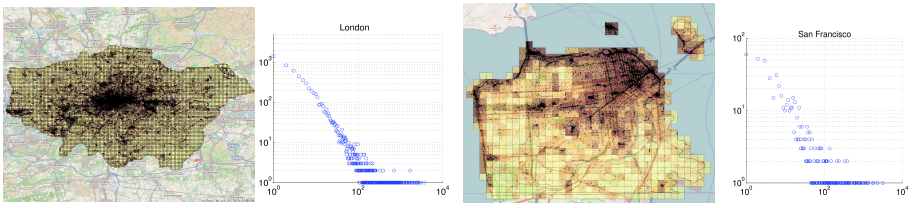

Fig. 3. The spatial plot of all the images, along with the spatial cells, on the city map of London and San Francisco. Each image is a black point on the map, while the color of the cell varies with the number of images it contains. Though these two cities have different areas and cell numbers, the distributions of image number per cell both follow a power law.

To train the scene attribute classifiers, we use the SUN attribute database [20], which consists of 102 scene attributes labeled on 14,340 images from 717 categories from the SUN database [28]. These scene attributes, such as 'natural', 'eating', and 'open area', are well tailored to represent the content of visual scenes. We use the deep convolutional network pre-trained on ImageNet [5] to extract features of images in the SUN attribute database, since deep learning features are shown to outperform other features in many large-scale visual recognition tasks [10, 5]. Every image is then represented by a 4096 dimensional vector from the output of the pre-trained network's last fully connected layer. These deep learning features are then used to train a linear SVM classifier for each of the scene attributes using Liblinear [6]. In Fig.4 we compare our approach to the methods of using single feature GIST, HoG, Self-Similarity, Geometric Color Histogram, and to the combined normalized kernel method in [20]. Our approach outperforms the current state-of-the-art attribute classifier with better accuracy and scalability.

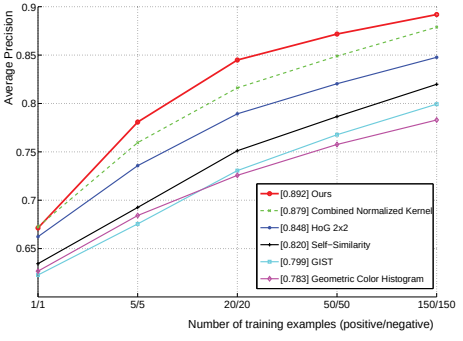

Fig. 4. Average precision (AP) averaged over all the scene attributes for different features. Our deep learning feature classifier is more accurate than the other individual feature classifiers and the combined kernel classifier used in [20].

For every image in the city perception database, we also use the pre-trained convolutional network to extract features. Fig. 5 shows images with 4 scene

attributes detected by our scene attribute classifiers from 3 cities Boston, Hong Kong, and Barcelona. Images are ranked according to the SVM confidence. We can see that these scene attributes sufficiently describe the semantics of the image content.

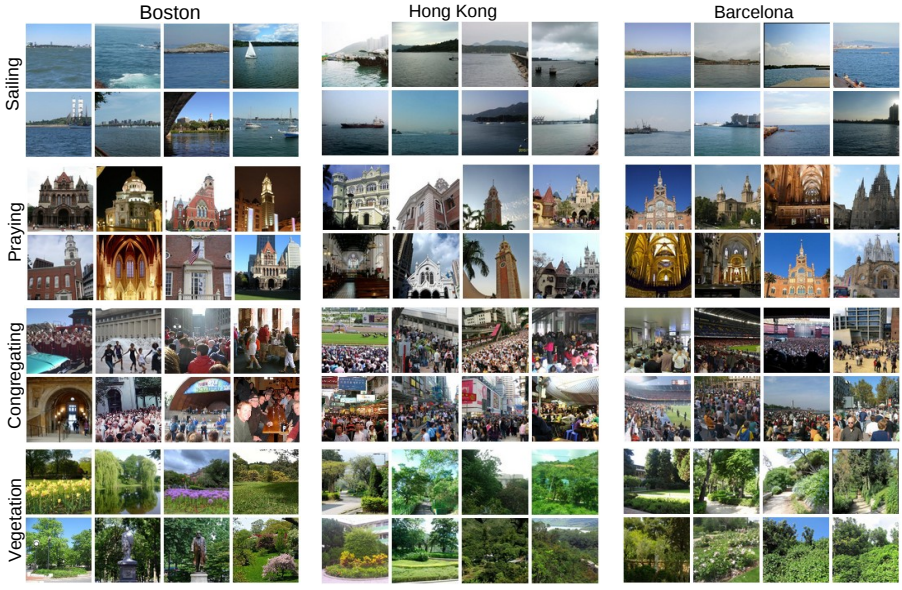

Fig. 5. Images detected with four scene attributes from Boston, Hong Kong, and Barcelona. Images are ranked according to their SVM confidences.

Table 1. The number of images detected with city attribute in 5 cities.

	Green	Water	Trans.	Arch.	Ver.	Ath.	Soc.	Total Images
London	53,306	15,865	25,072	12,662	38,253	6,311	11,405	209,264
Boston	5,856	2,735	3,059	1,488	6,291	618	1,142	26,288
Hong Kong	47,708	18,878	14,914	2,066	21,690	1,346	8,354	152,147
Shanghai	8,373	1,623	5,368	862	8,252	509	1,569	35,722
Barcelona	25,831	6,825	9,160	6,810	24,334	2,338	6,093	114,867

We further merge scene attributes into higher level city attributes. Given that some scene attributes are highly correlated with each other (like 'Vegetation' and 'Farming') and some other attributes like 'Medical activity' and 'Rubber' are not relevant to the city identity analysis, we choose a subset of 42 scene attributes that are most relevant to represent city form and function, and combine them into the 7 city attributes commonly used in urban study and city ranking [27, 26]: **Green space**, **Water coverage**, **Transportation**, **Architecture**, **Vertical**

building, **Athletic activity**, and **Social activity** (see the lists of selected scene attributes contained in each city attribute in the supplementary materials). Thus, each of the city attribute classifier is modeled as an ensemble of SVMs: One image is detected with that city attribute if any of the constituent scene attributes is detected, while the response of the city attribute is calibrated across the SVMs using logistic regression. We apply the city attribute classifiers to all the images in the City Perception Database. Table 1 shows the number of detected images on each city attribute across 5 cities. These numbers vary across city attributes due to the difference in the scenic spots, tourist places, or the urban characteristics of the cities. Note that one image might be detected with multiple attributes.

2.3 Spatial Analysis of City Attributes

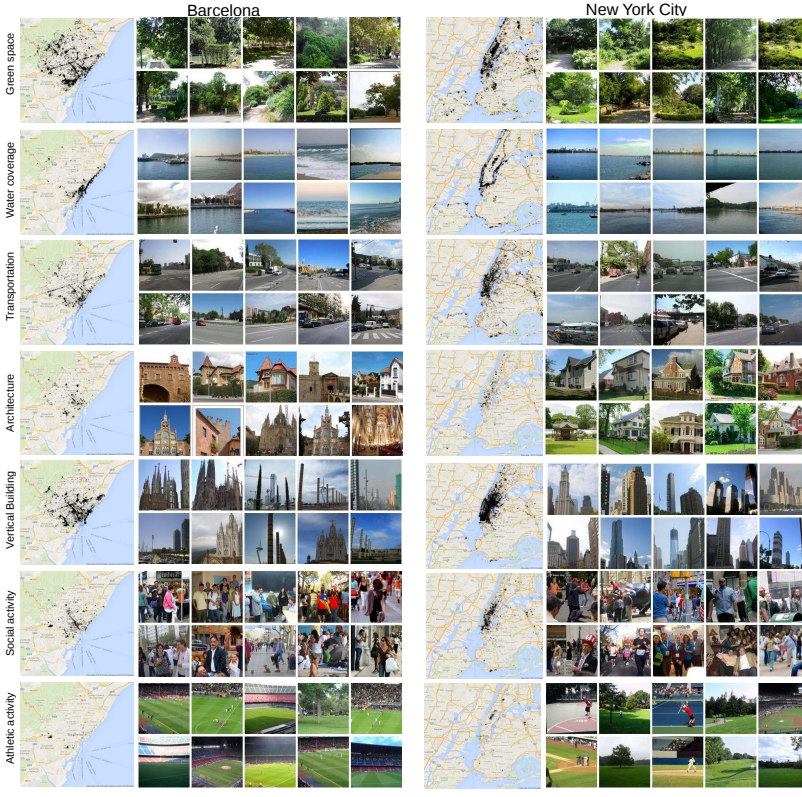

Fig. 6. Spatial distribution of city attributes and the top ranked images classified with each city attribute in Barcelona and New York.

Fig.6 shows the images detected with each of the 7 city attributes on a map. We can see that different city attributes are unequally distributed on map. This

makes sense, given that cities vary in structure and location of popular regions. For example, images with water coverage lie close to the coast line, rivers, or canals of the city, and images with social activities lie in the downtown areas of the city. Note that the images detected by city attribute classifiers have more visual variations to the result of the scene attribute classifiers.

Fig.7 shows the city perception maps for Barcelona, New York City, Amsterdam, and Bangkok, which visualize the spatial distribution of the 7 city attributes in different colors. The city perception map exhibits the visitors' and inhabitants' own experience and perception of the cities, while it reflects the spatial popularity of places in the city across attributes.

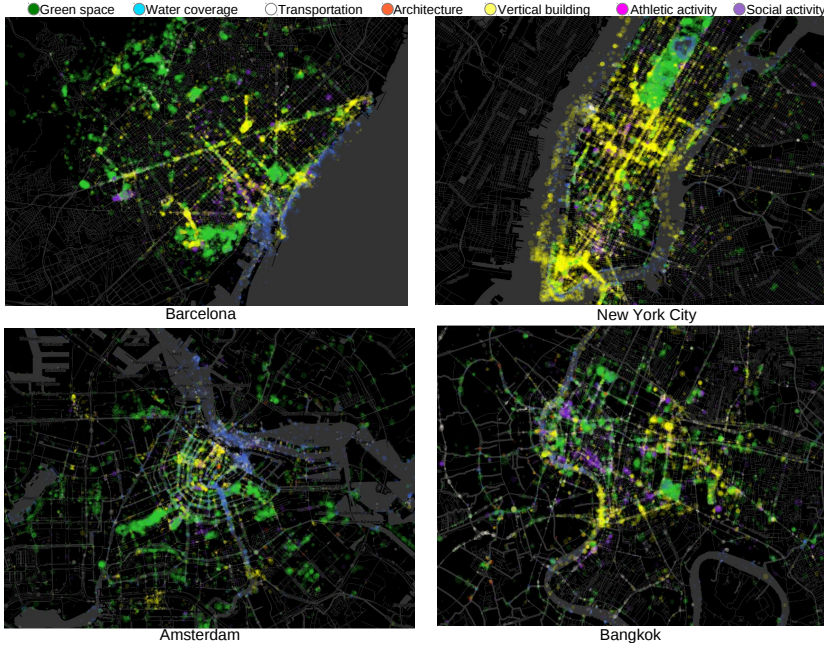

Fig. 7. City perception map of Barcelona, New York, Amsterdam, and Bangkok. Each colored dot represents a geo-tagged image detected with one city attribute.

3 Recognizing City Identity of Images

City identity emerges in every aspect of daily life and implicitly exists in the people's perception of the city. As shown in Fig. 1, people can easily recognize the city identity of these photos based on their former experience and knowledge of the cities. This raises the interesting questions: 1) can we train classifiers to recognize the city identity of images? 2) what are the images with high city identity values, *i.e.*, the representative images of the city?

In this section, we formulate the city identity recognition as a discriminative classification task: Given some images randomly sampled from different cities, we hope to train a classifier that could predict which city the newly given images come from. The challenge of the task lies in the wide variety of the image contents across cities. Here we show that city identity actually could be recognized on different city attributes, while the misclassification rate in the city identity recognition experiment could be used to measure the similarity between cities.

3.1 Attribute-based city identity recognition

As shown in Table 1 and Figure 6, images of each cities with different city attribute are detected. We are more curious about which images are unique in one city as well as discriminative across other cities on some city attribute. Thus we conduct the discriminative classification of all the 21 cities: For each of the 7 city attribute, 500 images with that city attribute are randomly sampled from each city as the train set, while all the remaining images are included in the test set. A linear SVM classifier is trained and tested for each of the 7 city attributes respectively. Here the train set size 500 is empirically determined, as we assume such a number of images contain enough information about the city identity.

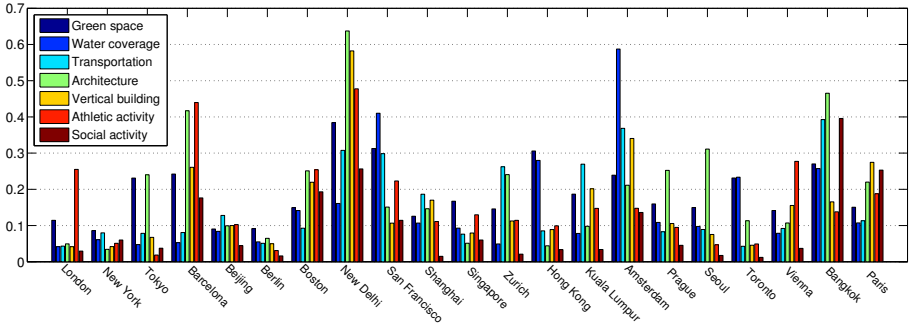

Fig. 8. The accuracies of city identity recognition for each city attribute.

Figure 8 plots the accuracies of city identity recognition for each city attribute. Figure 9 illustrates the confusion matrices of city identity recognition for architecture and green space. The performance of city identity recognition is not very high due to the large variety of image contents, but the trained linear SVM classifier actually has good enough discriminative ability compared to the chance. Meanwhile, we can see that the recognition accuracy varies across both cities and city attributes. It is related to the uniqueness of one city for that city attribute. For example, New Delhi and Bangkok have high accuracy in architecture attribute, since they have unique architectures compared to all the other cities selected in the City Perception Database. Interestingly, the misclassification rate in the city identity recognition actually reflects the similarity of two cities, since there are a high number of indistinguishable images from the

two cities. In our case, Paris, Vienna, and Prague are all similar to Barcelona in architecture attribute. This observation leads to our data-driven similarity of cities in Section 3.2.

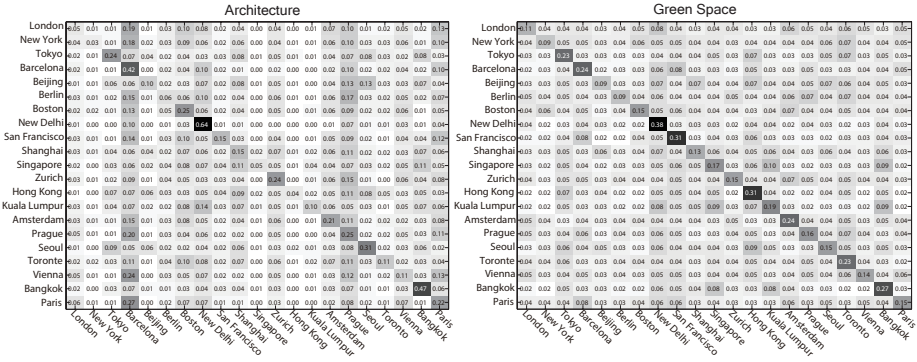

Fig. 9. Confusion matrices of city identity recognition for architecture attribute and green space attribute.

The SVM confidence of an image actually indicates the city identity value of that image. We rank the correctly classified images to discover the images representing salient city identity. Fig.10 shows the images with salient city identity on 5 city attributes respectively. For example, in transportation attribute, there are lots of canal cruises in Amsterdam since it has more than one hundred kilometers of canals across the whole city; Tokyo has narrow streets since it is pretty crowded; Red double decker buses and yellow cabs are everywhere on the street of London and New York respectively, while tram cars are unique in San Francisco. Images with salient city identity in architecture attribute show the representative construction styles, while images with salient city identity in athletic activity attribute indicate the most popular sports in these cities.

3.2 Data-driven visual similarity between cities

How similar or different are two cities? Intuitively we feel that Paris is more similar to London than to Singapore, while Tokyo is more similar to Beijing than to Boston. Measuring the similarity of cities is still an open question [25, 21].

Here we use a data-driven similarity metric between two cities based on the misclassification rate of the city identity recognition. We assume that if two cities are visually similar to each other, the misclassification rate in the city identity recognition task should be high across all the city attributes. Thus we can use the pairwise misclassification rates averaged over all 7 city attributes as a similarity measure. The misclassification rate on each city attribute is computed from the city identity recognition repeated on every pairs of cities, which is the sum of the

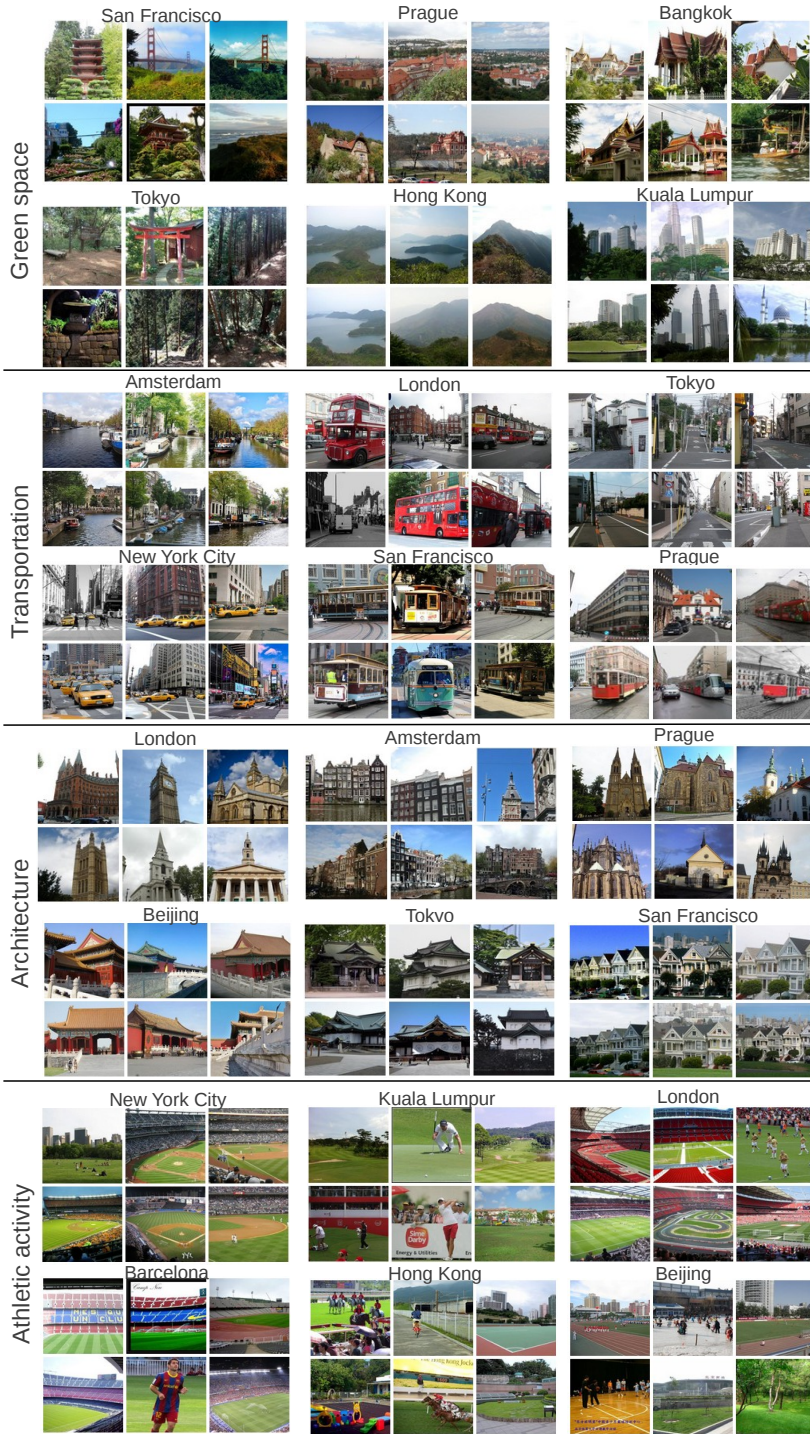

Fig. 10. Images with high city identity values on 4 city attributes.

rate of misclassifying images of city A as city B and the rate of misclassifying images of city B as city A.

Fig.11A plots the similarity network of all the 21 cities in our database. The thickness of the edges indicates the averaged pairwise misclassification rate (for better visualization we remove weak edges). Nodes with the same color are those that belong to the same group, clustered by the community detection algorithm [17]. Cities in one group are closely connected with each others: they are grouped based on their located continents-Europe, North America, or Asia. In Fig.11B we scatter our data-driven similarity with the geodesic distances between the geographical centers of the cities. The correlation coefficient r is -0.57 ($p < 0.01$).

Our result validates that the geographical distance between cities plays an important role in determining the similarity of cities. Indeed, historically there were more trade and cultural exchanges between spatially-neighboring cities, so they would share similar elements in their form and culture. Besides, there are some outliers detected in our result, such as New Delhi and New York. This is because the culture of New Delhi is quite different from other cities in our database, while New York is a metropolis which mixes the cultures from all over the world. Our data-driven similarity metric between cities is useful for the study of urban sociology and city history.

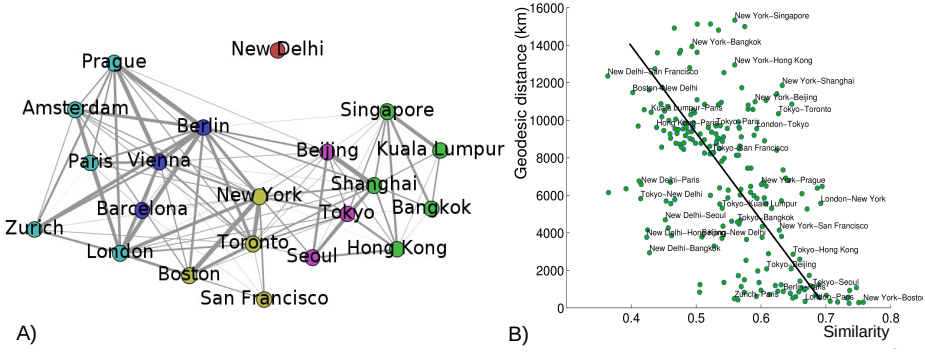

Fig. 11. A) Similarity graph of cities. The nodes indicate cities, the thickness of the edges indicates the similarity computed from the misclassification rate of the city identity recognition. Nodes with the same color belong to the same cluster. B) The scatter of the data-driven similarity over geodesic distance of cities. Each point indicates one pair of cities. Some representative labels of points are plotted. There is high negative correlation between the geodesic distance of a pair of cities and their visual similarity.

4 Further Application in Urban Planning

Estimating people's perceptions of a city from a massive collection of geo-tagged images offers a new method for urban studies and planning. Our data-driven approach could help assess and guide urban form construction.

Following the seminal work of Kevin Lynch [15], our data-driven analysis of images taken by people in a city relates the subjective perception of people with the urban construction. City identity has been described along various urban dimensions, which additionally allows us to think about urban design in various terms. Here we propose to quantify the correlation between our geo-tagged image analysis and urban construction, using a set of quantitative indicators provided by MassGIS and Boston(Mass.) Department of Innovation and Technology [6] :building height and density, openness rate, sidewalk density, block size, and the ratio of street height and width, available for the city of Boston.

First, we segment the city into 50m×50m connected but non-overlapping spatial cells, so that city regions are quantified by hundreds of cells. For each cell, we represent the people's perception of this cell as the number of images inside, the attribute diversity (the number of the different city attributes in the cell), and the dominant city identity value. Then we calculate the Pearson correlation to see how people's perceptions are relevant to those urban design measurements.

The result of the statistical analysis can be found in Table 2. Most of the urban design indicators are significantly correlated with people's perception of the city. For example, sidewalk density is positively correlated with the attribute diversity (0.296). This means that a place with more space of sidewalk is likely to provide more views of people while it allows people to move between areas more easily. Similar results are found for building height and the ratio of height and width of streets. Regarding the negative correlation, it shows that when block size grows the place becomes less friendly to the pedestrians. Therefore it has a negative impact on the people's perception of the city. Openness rate, which refers to the percentage of open space in a cell, is a more complex indicator. On the one hand, the positive coefficient between openness rate and city identity value suggests that a larger open space would have higher city identity value. On the other hand, the negative coefficient of the image number and attribute diversity indicates that such a place is less diverse and crowded. For example, Boston common is one of the Boston's well-known public center with a salient city identity, and it also has a very high openness rate. However, because visitors are scattered across the large park, the geo-tagged images won't look quite spatially dense. Fig.12 plots the attribute diversity value and the sidewalk density of Boston in the city maps. This comparison between urban form and city identity provides valuable reference and instructions for urban planners because they incorporate people's subjective preference toward the built environment.

[6] `http://www.cityofboston.gov/DoIT/`
`http://www.mass.gov/`

Table 2. The correlation between the results of geo-tagged image analysis and the urban design indicators for the city of Boston. **: P-value<0.01

Indicators	Building height	Openness rate	Sidewalk density	Block size	H/W
Image number	0.156**	-0.101**	0.241**	-0.076**	0.112**
Attribute diversity	0.098**	-0.061**	0.296**	-0.092**	0.080**
City identity value	0.070**	0.029	0.235**	-0.092**	0.065**

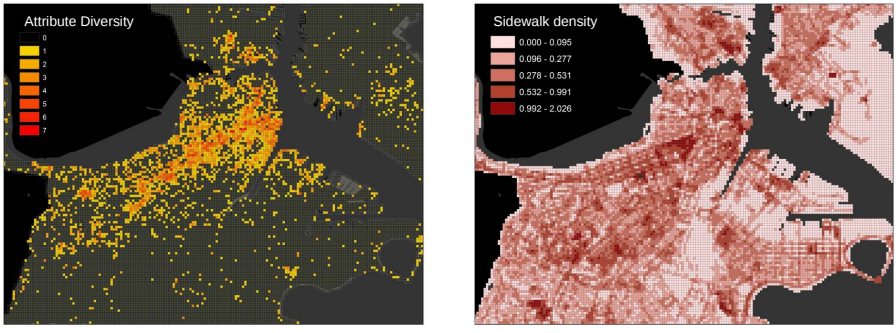

Fig. 12. The map of attribute diversity and the map of sidewalk density in Boston. Color indicates the value intensity.

5 Conclusion

This work introduced a novel data-driven approach to understand what makes the identity of a city. From the attribute analysis of many geo-tagged images of cities, we found that the identity of a city could be represented and expressed in various urban dimensions, such as green space, architecture, transportation, and activity. Using the attribute representation of images tailored to describe the form and function of cities, we identified images with salient city identity and measured the similarity between cities. Further applications in the urban planning were discussed.

Our work suggests a fruitful direction of computer vision research in the domain of urban computation. In the future work, we plan to study how a city identity develops over time based on the time-stamps of the geo-tagged images. Meanwhile, by collecting additional urban measurements, such as the distribution of the crime rates and the cleanliness of neighborhoods, we hope to more faithfully identify what makes a city distinctive and suggest how to optimize the urban planning of a specific city, or a cluster of cities, to take into account the people's needs and expectations.

Acknowledgments. We thank Zoya Bylinskii, Christopher Dean, Scott Hu, and the reviewers for insightful suggestions. This work is partly funded by NSF grant 1016862 to A.O, and Google Research Awards to A.O and A.T.

References

1. Branson, S., Wah, C., Schroff, F., Babenko, B., Welinder, P., Perona, P., Belongie, S.: Visual recognition with humans in the loop. In: Proc. ECCV (2010)
2. Chen, D.M., Baatz, G., Koser, K., Tsai, S.S., Vedantham, R., Pylvanainen, T., Roimela, K., Chen, X., Bach, J., Pollefeys, M., et al.: City-scale landmark identification on mobile devices. In: Proc. CVPR (2011)
3. Crandall, D.J., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: Proceedings of the 18th international conference on World wide web (2009)
4. Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A.A.: What makes paris look like paris? ACM Transactions on Graphics (TOG) (2012)
5. 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 preprint arXiv:1310.1531 (2013)
6. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: A library for large linear classification. The Journal of Machine Learning Research (2008)
7. Hays, J., Efros, A.A.: Im2gps: estimating geographic information from a single image. In: Proc. CVPR (2008)
8. Khosla, A., An, B., Lim, J.J., Torralba, A.: Looking beyond the visible scene. In: Proc. CVPR (2014)
9. Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., Andrienko, G.: Event-based analysis of people's activities and behavior using flickr and panoramio geotagged photo collections. In: Information Visualisation (IV), 2010 14th International Conference (2010)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: In Advances in Neural Information Processing Systems (2012)
11. Lee, Y.J., Efros, A.A., Hebert, M.: Style-aware mid-level representation for discovering visual connections in space and time. In: Proc. ICCV (2013)
12. Li, X., Wu, C., Zach, C., Lazebnik, S., Frahm, J.M.: Modeling and recognition of landmark image collections using iconic scene graphs. In: Proc. ECCV (2008)
13. Li, Y., Crandall, D.J., Huttenlocher, D.P.: Landmark classification in large-scale image collections. In: Proc. ICCV (2009)
14. Lin, T.Y., Belongie, S., Hays, J.: Cross-view image geolocalization. In: Proc. CVPR (2013)
15. Lynch, K.: The image of the city. MIT press (1960)
16. Nasar, J.L.: The evaluative image of the city. Sage Publications Thousand Oaks, CA (1998)
17. Newman, M.E.: Modularity and community structure in networks. Proceedings of the National Academy of Sciences (2006)
18. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. Int'l Journal of Computer Vision (2001)
19. Parikh, D., Grauman, K.: Relative attributes. In: Proc. ICCV (2011)
20. Patterson, G., Hays, J.: Sun attribute database: Discovering, annotating, and recognizing scene attributes. In: Proc. CVPR (2012)
21. Preoţiuc-Pietro, D., Cranshaw, J., Yano, T.: Exploring venue-based city-to-city similarity measures. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing (2013)

22. Proshansky, H.M., Fabian, A.K., Kaminoff, R.: Place-identity: Physical world socialization of the self. Journal of environmental psychology (1983)
23. Ripley, B.D.: Spatial statistics, vol. 575. John Wiley & Sons (2005)
24. Salesses, P., Schechtner, K., Hidalgo, C.A.: The collaborative image of the city: mapping the inequality of urban perception. PloS one (2013)
25. Seth, R., Covell, M., Ravichandran, D., Sivakumar, D., Baluja, S.: A tale of two (similar) cities: Inferring city similarity through geo-spatial query log analysis. In: Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (2011)
26. Unit, E.I.: Best cities ranking and report. In: The Economist (2012)
27. Unit, E.I.: Global liveability ranking and report. In: The Economist (2013)
28. Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A.: Sun database: Large-scale scene recognition from abbey to zoo. In: Proc. CVPR (2010)
29. Zhang, H., Korayem, M., Crandall, D.J., LeBuhn, G.: Mining photo-sharing websites to study ecological phenomena. In: Proceedings of the 21st international conference on World Wide Web (2012)
30. Zheng, Y.T., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.S., Neven, H.: Tour the world: building a web-scale landmark recognition engine. In: Proc. CVPR (2009)