Does the use of geographic knowledge enhance multilingual text recognition?

Seong-Yun Hong (syhong@khu.ac.kr), Yeorim Kim (erin9906@khu.ac.kr), Yongchae Lee (yong03111@khu.ac.kr)
Department of Geography, Kyung Hee University, Seoul, Republic of Korea

Background

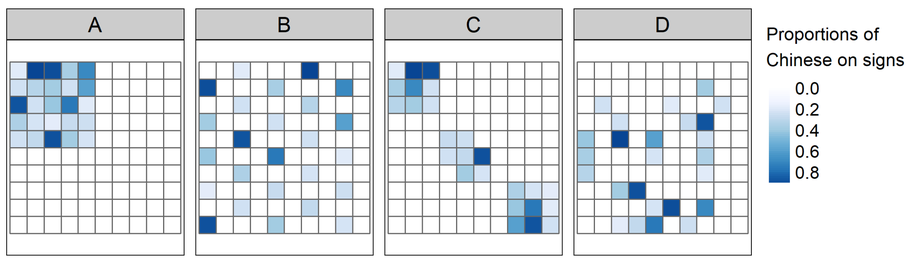
The salience and visibility of the languages in urban spaces reflect the ethnolinguistic composition of the population inhabiting the city. A careful examination of linguistic landscapes (LL) can reveal how different cultures, languages and identities occupy and consume places. The prevalence of minority languages in certain parts of the city implies the presence of ethnic neighbourhoods in those areas, and the widespread of foreign languages on commercial signs is typically observed in popular tourist destinations.

In this poster, we propose a method that uses geographic knowledge to more accurately recognise the languages on street-level images. While most existing text detection and recognition models rely only on the information on the image under consideration, the proposed approach takes the distribution of the languages in the study area into account. The primary purpose of this work is to examine whether using such geographic knowledge can enhance language recognition accuracy in multilingual environments. To this end, we will apply the proposed method to hypothetical data, as well as street-level images from Seoul.

Test data sets

The proposed method is applied to two data sets below:

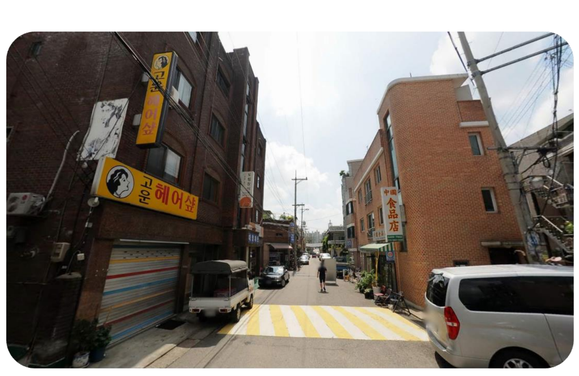
1. Hypothetical patterns with varying degrees of spatial autocorrelation

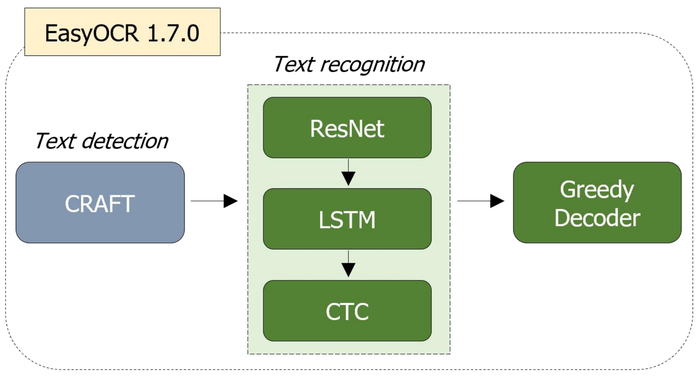


2. Street-level image data for Seoul, South Korea
   - A total of 1630 street-level images for the years between 2016 and 2018
   - Each text area on the images has been manually annotated as English, Korean, Chinese, numbers, or other languages.
   - The Moran's *I* indices for English, Korean, and Chinese were all close to 0, indicating no spatial autocorrelation.

Proposed method

In this work, we used a Python library called EasyOCR, which includes both text detection and recognition algorithms, to calculate initial confidence scores. Then, we adjusted the scores using a priori information about, for example, the proportions of the languages on signs in the study area. The diagram below illustrates the overall adjustment process.



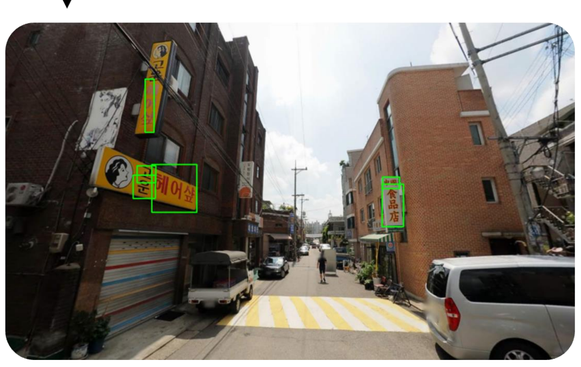


English		Korean		Chinese Simplified		…
Token	Score	Token	Score	Token	Score	
#0	0.4931	무터	0.0726	中圃	0.0531	…
8	0.9714	운	0.9996	8	0.5426	
0	0.2033	헤어석	0.2399	刮圜	0.0009	
…						

Adjust the confidence scores by multiplying additional weights, π_i and τ_{ip} (see the box on the right)

English		Korean		Chinese Simplified		…
Token	Score	Token	Score	Token	Score	
#0	0.0429	무터	0.0583	中圃	0.0722	…
8	0.0844	운	0.8028	8	0.7374	
0	0.0177	헤어석	0.1927	刮圜	0.0012	
…						

Choose the language that satisfies: $\underset{i}{\operatorname{argmax}}\left[\left(\prod_i x_{ij}\right)^{2/\sqrt{n_i}} \times \pi_i \times \tau_{ip}\right]$

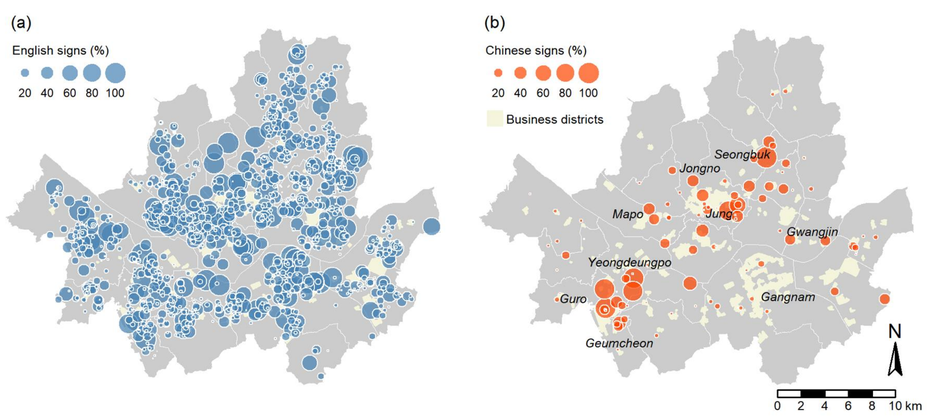


Proposed weighting method

Sample proportions of the languages on used on signs in Seoul (n = 1630)

Language	Korean, π_1	English, π_2	Chinese, π_3	Numbers	Others, π_4
Proportion	0.8031	0.0869	0.0151	0.0874	0.0076

π_i is global weight of the language. If the proportions of the languages on LL are known, these can be used as the weights.



τ_{ip} is the local weight of the language i at location p.

$\tau_{ip} = 0$ when the language is not spatially autocorrelated as in (a).

If the Chinese signs show moderate positive spatial autocorrelation as shown in (b), the local weight is applied to the adjustment.

Source: Hong, Kim, and Lee (2023, Figure 3 on p. 9)

Preliminary results

The table below shows the Pearson correlation coefficients, ρ, between the true proportion of each language and that estimated by different methods, namely, the existing OCR model as a baseline, the proposed model with global weights only applied, and the full model (i.e., both global and local weights applied).

In all settings, the 'global weights only' model performed better than the baseline. The full model was more effective when strong spatial autocorrelation was present (i.e., Pattern A).

Method	Pattern	Korean	English	Chinese
Existing OCR framework	All	0.6685	0.3412	0.1941
Global weights only	All	0.6779	0.3439	0.3425
Proposed model (using both global and local weights)	A	0.6428	0.3490	0.4244
	B	0.6685	0.3386	0.1881
	C	0.6799	0.3378	0.1689
	D	0.6664	0.3437	0.2927

For the LL data of Seoul, the 'global weights only' model worked better than the model with the local adjustment, presumably because there was no strong spatial autocorrelation in the distribution of the languages in Seoul.

References

Landry, R., & Bourhis, R. Y. (1997). Linguistic landscape and ethnolinguistic vitality: An empirical study. *Journal of Language and Social Psychology*, **16**(1), 23-49. https://doi.org/10.1177/0261927x970161002

Hong, S.-Y. (2020). Linguistic landscapes on street-level images. *ISPRS International Journal of Geo-Information*, **9**(1), 57. https://doi.org/10.3390/ijgi9010057

Hong, S.-Y., Kim, Y., & Lee, Y. (2023). Ethnic segregation on linguistic landscapes. *Environment and Planning B: Urban Analytics and City Science*, OnlineFirst, 23998083221150240. https://doi.org/10.1177/23998083221150240

Jaided AI (2023). EasyOCR (software). https://github.com/JaidedAI/EasyOCR

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (NRF-2021R1C1C1009849.