Instant camera translation and voicing for signs

Kaifeng Chen[1], Shanshan Xu[2], Yao Zhou[3]

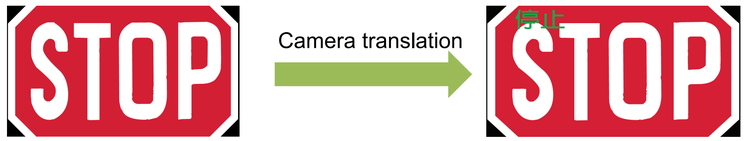
[1]Department of Applied Physics, Stanford University

[2]Department of Physics, Stanford University

[3]Department of Material Science Engineering, Stanford University

Objective

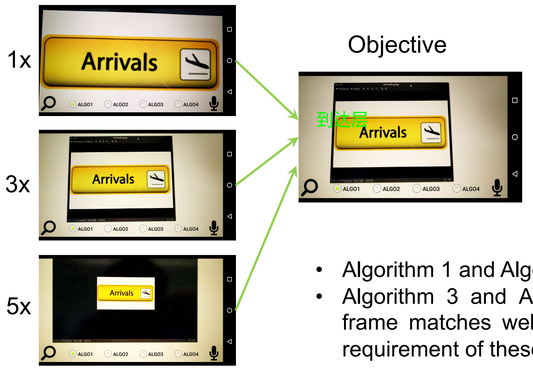
For Chinese visitors who are not familiar with English, especially the elderly, to facilitate their trip, it is of great importance to have a smart phone application that can handle the real time sign translation for them. For example the translation of a STOP sign:



This problem is one example of Optical Character Recognition (OCR) problems, where the underlying algorithms are closely related to image processing techniques. In this project, we build a simple but useful Android application for English to Chinese translation with voicing. We explore several algorithms that process either the text block in the image or the whole image itself. In addition, we also compare these algorithms and point out the advantages and disadvantages.

Results

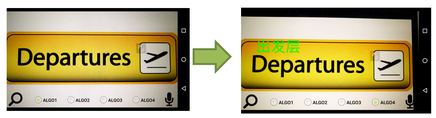
Comparison of accuracy



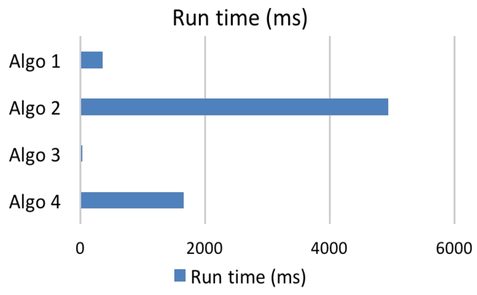
Factor	1x	3x	5x
Algo 1	✘	✔	✔
Algo 2	✘	✔	✔
Algo 3	✔	✘	✘
Algo 4	✔	✘	✘

- Algorithm 1 and Algorithm 2 work better with large distance.
- Algorithm 3 and Algorithm 4 work better when the input frame matches well with the database images due to the requirement of these algorithms.

Comparison of speed



- Algorithm 3 is the fastest as a result of the significantly reduced dimension of the image.
- Algorithm 2 is noticeably much slower than the others due to the time consuming MSER feature detector.

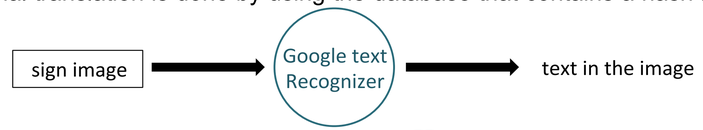


Algorithms

We provide the following four algorithms for sign recognition. The final translation is done by using the database that contains a hash map between English text and the corresponding Chinese translation.

OCR detection

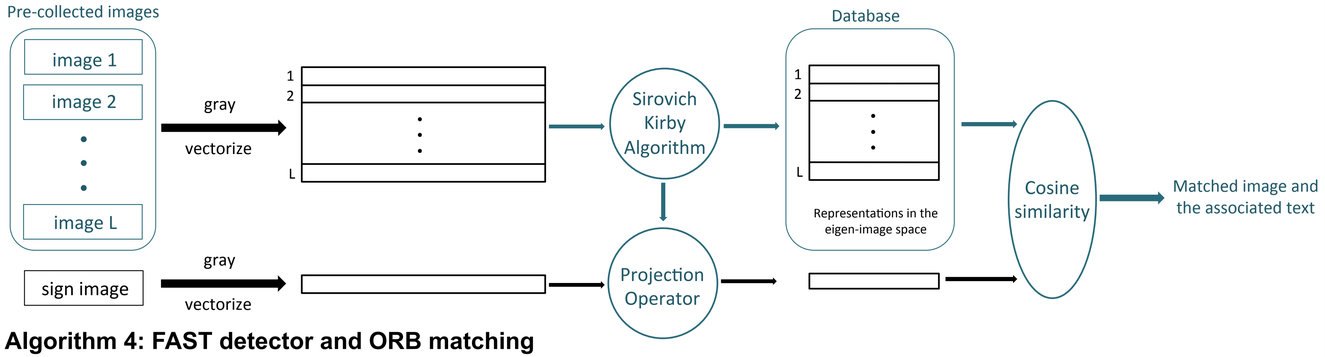
- **Algorithm 1: Google Mobile Services (GMS)**



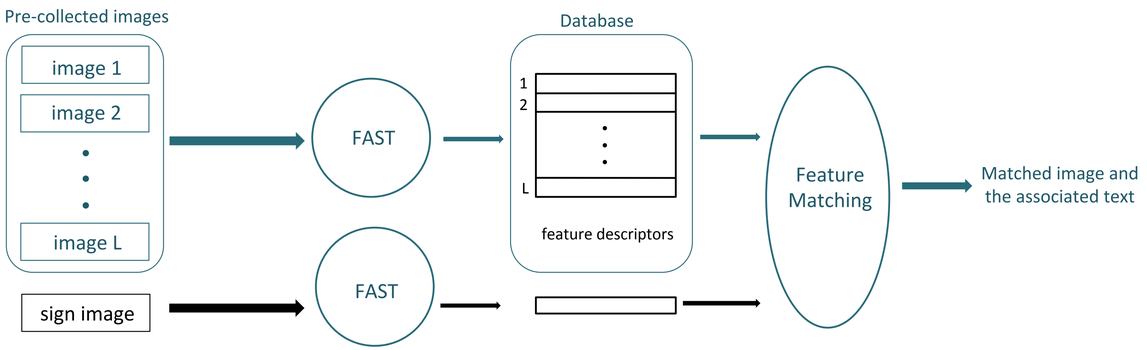
- **Algorithm 2: MSER and GMS**

Image Detection

- **Algorithm 3: Sirovich and Kirby Algorithm.**



- **Algorithm 4: FAST detector and ORB matching**



1. J. Matas, O. Chum, M. Urban, and T. Pajdla. "*Robust wide baseline stereo from maximally stable extremal regions*." Proc. of British Machine Vision Conference, pages 384-396, 2002.
2. L. Sirovich; M. Kirby (1987). "*Low-dimensional procedure for the characterization of human faces*". Journal of the Optical Society of America A. 4 (3): 519–524
3. Edward Rosten and Tom Drummond, "*Machine learning for high speed corner detection*" in 9th European Conference on Computer Vision, vol. 1, 2006, pp. 430–443.
4. Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary R. Bradski: ORB: An efficient alternative to SIFT or SURF. ICCV 2011: 2564-2571.