Aperture: Voice Annotations for Cognitive Assistance

Aditya Chanana, Shuning Lin, and Mihir Bala (Mentor)

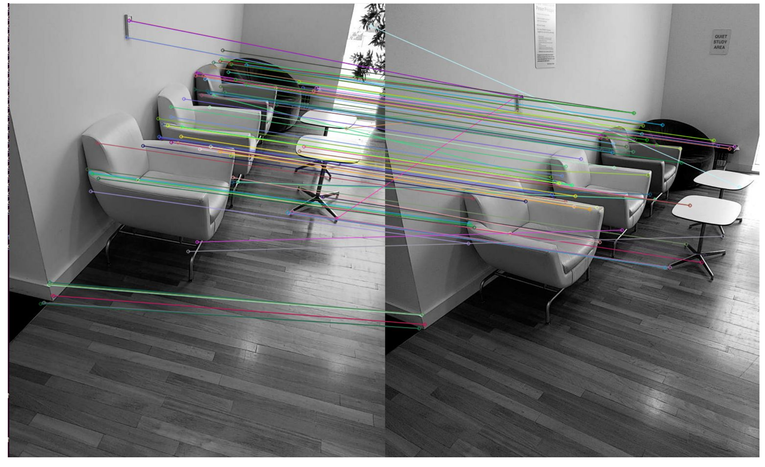
What is Aperture?

- **Aperture** serves as a memory aid for people with cognitive disabilities
- Allows caretakers to add annotations on objects in the surroundings by pointing their smartphone camera at them and typing out a message on an app
- Annotations could be reminders to take medication or warnings to stay away from hot objects
- Annotations are read out to users when they point their smartphone camera at an annotated object

Goal

Offload compute-intensive image matching to a nearby cloudlet to conserve constrained resources on the mobile device while preserving a seamless user experience

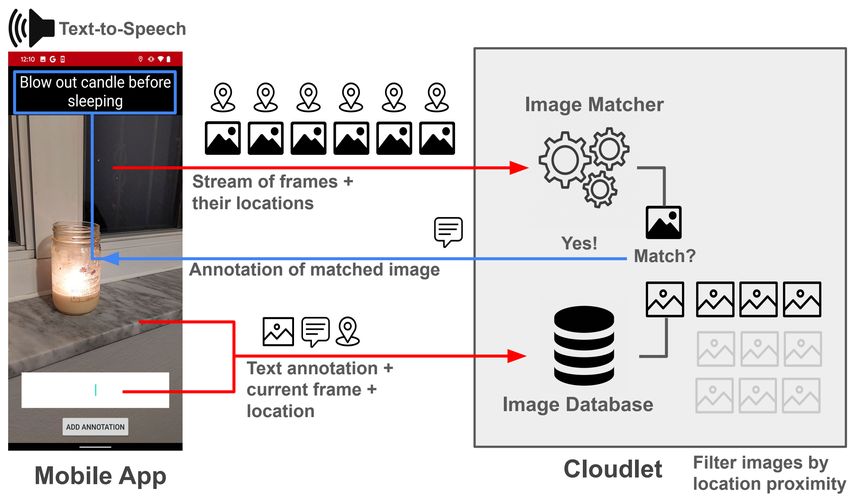
Background



- **Image matching**: matching different views of an object or scene
- Requires computation of distinctive image features that are invariant to image scale and rotation

[1] Kiryong Ha, Zhuo Chen, Wenlu Hu, Wolfgang Richter, Padmanabhan Pillai, and Mahadev Satyanarayanan. 2014. Towards wearable cognitive assistance. In Proceedings of the 12th annual international conference on Mobile systems, applications, and services (MobiSys '14). Association for Computing Machinery, New York, NY, USA, 68–81. https://doi.org/10.1145/2594368.2594383
[2] M. Muja and D. G. Lowe, "Scalable Nearest Neighbor Algorithms for High Dimensional Data," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 11, pp. 2227-2240, 1 Nov. 2014, doi: 10.1109/TPAMI.2014.2321376.
[3] Alcantarilla, P.F., Bartoli, A., Davison, A.J. (2012). KAZE Features. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds) Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7577. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33783-3_16

System Architecture



- Image matching algorithm running on the cloudlet as a cognitive engine in the Gabriel[1] framework
- Camera frames sent to the cloudlet, annotations received back
- Matches found using FLANN[2], which performs an approximate nearest neighbor search

Benchmarks

- Dataset of 60 images constructed with our use case in mind
- Various feature extraction algorithms compared on their precision and recall characteristics (see figure 1)
- KAZE[3] features selected for their high precision and recall, as well as low computation time (see figure 2)

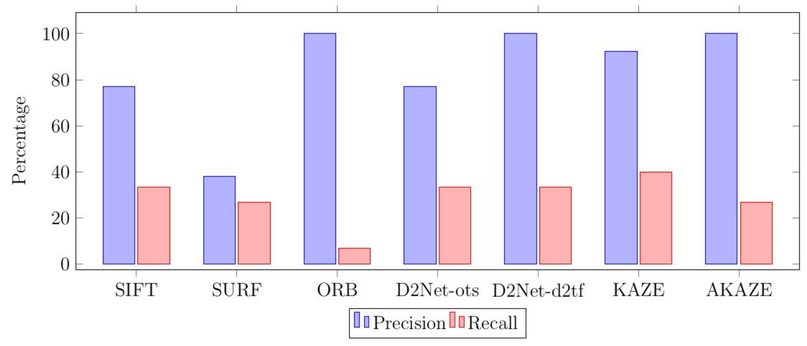


Figure 1: Performance of various feature extraction algorithms tested

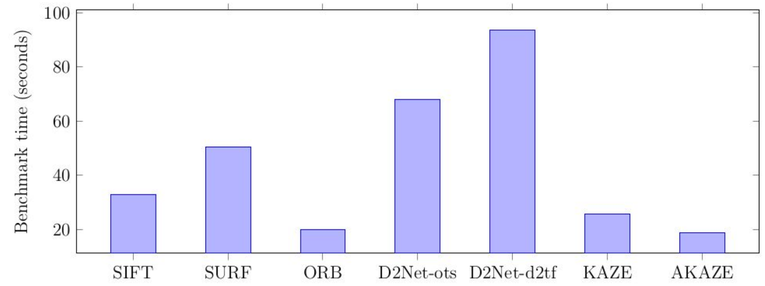


Figure 2: Benchmark time for different feature extraction algorithms

Scalability

- Image matching time increases with the number of annotations added (see figure 3)
- Incorporate location awareness to limit annotation search space

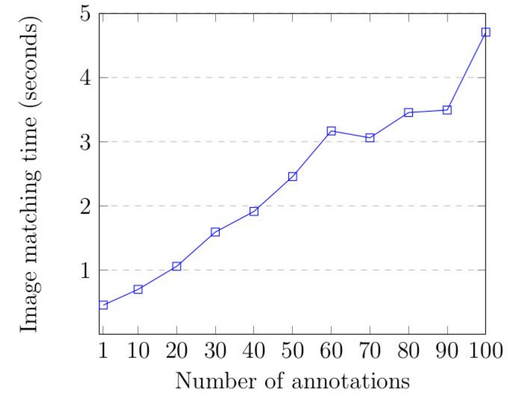


Figure 3: Time taken to find a matching annotation for a given number of annotations. By incorporating location awareness, the number of annotations we need to search may be much smaller than the total number of annotations added.

Challenges

- Many feature extraction algorithms with good performance reported in literature do not have implementations ready to use
- Existing image matching work focuses on different scenarios than our use case
- Adapting code that has not been maintained for years involves resolving dependency issues and rewriting code

Future Directions

- Permit more than one annotation to match current camera frame
- Ability to browse through all annotations in the current location
- Specify a "time-to-live" parameter to automatically remove stale annotations
- Enhance privacy by limiting access to annotations to specified users