Smartphone Based Image Captioning for Visually Impaired People

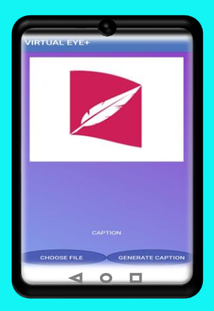
Özkan ÇAYLI, Volkan KILIÇ
caylio64@gmail.com
İzmir Katip Çelebi University, Dept. Of Electrical-Electronics Engineering

ABSTRACT

Generating captions and text descriptions of images will enable visually impaired extended accessibility to the real world, thus reducing their social isolation, improving their well-being, employability and education experience. In this study, a captioning system based on a smartphone, which offers ultra-low-cost, portable and user-friendly platform for visually impaired is examined. Image captioning aims to understand and describe a visual scene in terms of natural language expressions, which requires an advanced level of image understanding. The captioning system is run on the remote server integrated to the custom-designed Android application via a cloud system after trained on MSCOCO 2017 captioning dataset. The user can choose an image from the gallery or capture a new image to transfer the image to the remote server. The caption is generated by the captioning system and transferred back to the Android application to be displayed on the screen. The caption can be read out loudly using the narrator option. Experimental results show that the proposed system offers great advantages to the visually impaired in terms of portability, simple operation and rapid response.

Smartphone

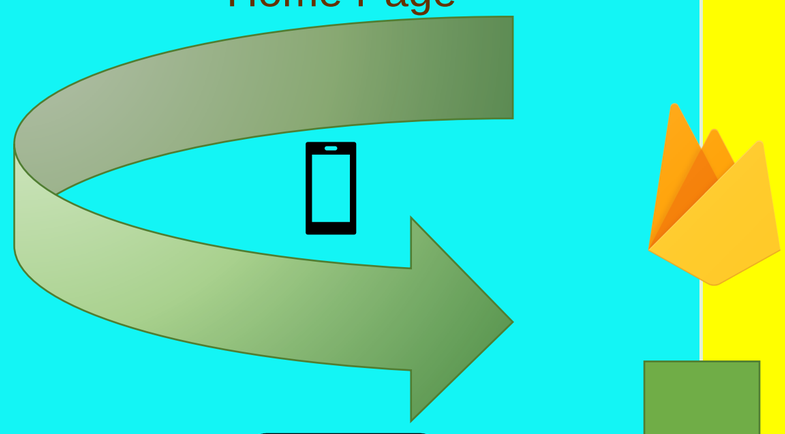
Accessing Camera



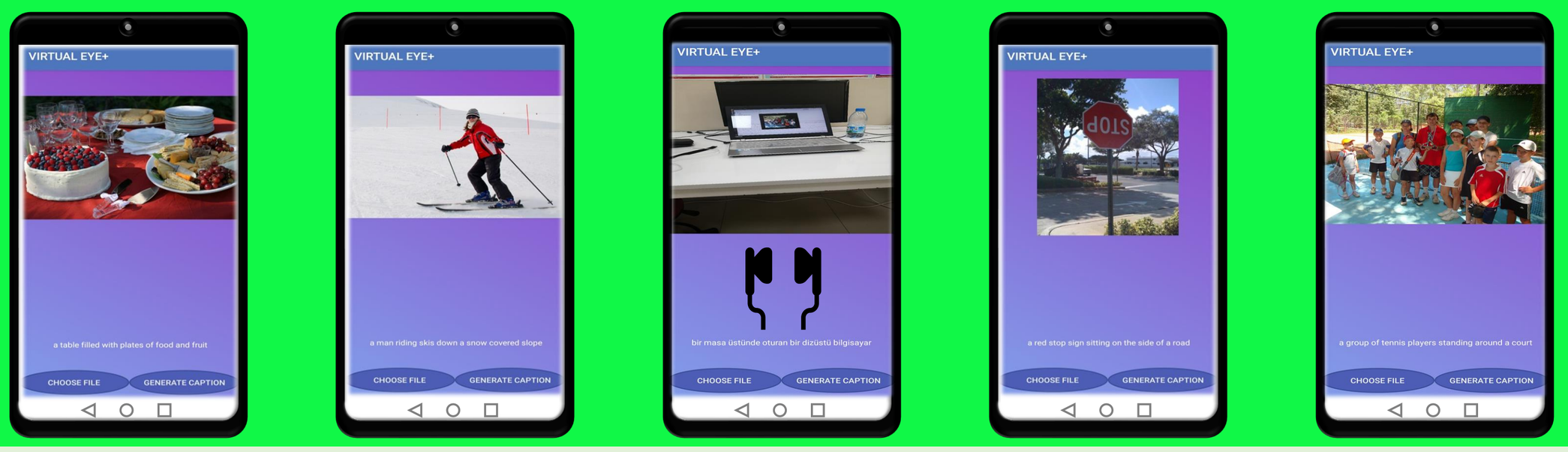
Home Page

Image Capturing

Image Transferring to the Server



RESULTS



PUBLICATIONS

Çaylı, Ö., Makav, B., Kılıç, V., & Onan, A. (2020, July). Mobile Application Based Automatic Caption Generation for Visually Impaired. In *International Conference on Intelligent and Fuzzy Systems* (pp. 1532-1539). Springer, Cham.

Çaylı, Ö., Makav, B., Kılıç, V., & Onan, A. (2020, October). Natural Language Description of Images Using a Smartphone Application. In 2nd International Eurasian Conference on Science, Engineering and Technology (EurasianSciEnTech 2020).

ACKNOWLEDGEMENT

This research was supported by the Scientific and Technological Research Council of Turkey (grant no. 1139B411900533).