JustPoint: Identifying Colors with a Natural User Interface

Sergio Mascetti
Università degli Studi di Milano
and EveryWare Technologies
sergio.mascetti@unimi.it

Andrea Gerino
Università degli Studi di Milano
and EveryWare Technologies
andrea.gerino@unimi.it

Cristian Bernareggi
Università degli Studi di Milano
and EveryWare Technologies
cristian.bernareggi@unimi.it

Silvia D'Acquisto
Università degli Studi di Milano
silvia.dacquisto
@studenti.unimi.it

Mattia Ducci
Università degli Studi di Milano
mattia.ducci
@studenti.unimi.it

James M. Coughlan
Smith-Kettlewell Eye
Research Institute
coughlan@ski.org

ABSTRACT

People with severe visual impairments usually have no way of identifying the colors of objects in their environment. While existing smartphone apps can recognize colors and speak them aloud, they require the user to center the object of interest in the camera's field of view, which is challenging for many users. We developed a smartphone app to address this problem that reads aloud the color of the object pointed to by the user's fingertip, without confusion from background colors. We evaluated the app with nine people who are blind, demonstrating the app's effectiveness and suggesting directions for improvements in the future.

CCS Concepts

•Human-centered computing → Empirical studies in accessibility; Accessibility technologies; Accessibility systems and tools; *Mobile computing;* **•Computing methodologies →** *Object recognition;*

Keywords

Assistive Technology, Hand Recognition, Color Detection

1. INTRODUCTION

People with severe visual impairments usually have no way of identifying the colors of objects in their environment, such as clothing, food or other items. While a number of useful mobile smartphone apps are available to recognize colors and speak them aloud, they require the user to center the object of interest in the smartphone camera's field of view, which often poses significant difficulties for the user.

We describe *JustPoint*, a smartphone app that reads aloud the color of the object that is pointed to by the user's fingertip. The users can explore the object surface by moving their finger across it, obtaining the desired color information without confusion from background colors.

ASSETS '17 October 29-November 1, 2017, Baltimore, MD, USA

ACM ISBN 978-1-4503-4926-0/17/10.

DOI: https://doi.org/10.1145/3132525.3134802

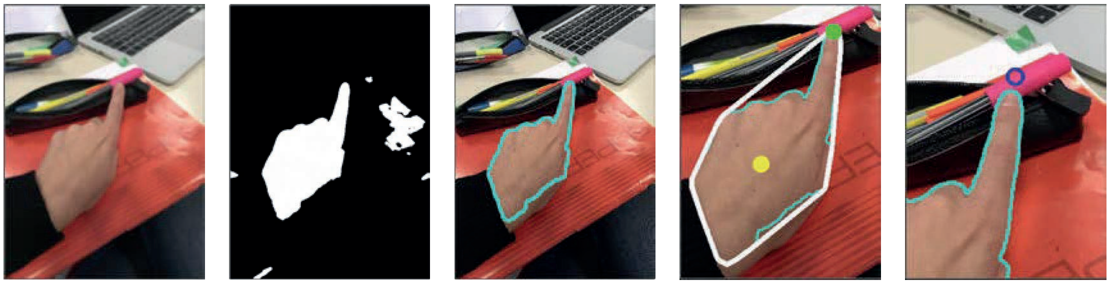
Our paper makes two main contributions: (i) an evaluation of a real-time fingertip detection technique, and (ii) the application of this technique to the task of color identification, which allows a blind user to point to a specific region to identify. We distinguish our contributions from the following past work. VizLens [1] performs real-time fingertip detection, but there is no explicit evaluation of fingertip detection, and it is not applied to the task of color identification. Access Lens [2] performs real-time fingertip detection and color identification, but it uses a laptop and a webcam rather than a smartphone, and there is only limited analysis of the system's ability to identify colors for visually impaired users. GIST [3] performs real-time fingertip detection and color identification, but it relies on a special depth camera rather than a standard smartphone camera. Finally, OrCam (http://www.orcam.com/) implements real-time fingertip detection using special-purpose hardware, but there are no publications describing the details of the algorithms used, and it does not perform color identification.

2. THE *JUSTPOINT* APP

JustPoint is an iOS prototype application whose primary purpose is to read aloud the color of the *target area*, i.e., a circular area located close to the fingertip in the direction it's pointing. This is obtained by acquiring frames from the device camera, detecting the target area with a computer vision technique, computing the name of the target area color and informing the user accordingly.

The computer vision technique processes video frames captured by the device's camera in five main steps (see Figure 1). First, the image is pre-processed with a blur filter. Then, skin detection is performed using the threshold-based technique proposed in [4]. The third step is to decide which skin-colored region(if any) is actually a hand. We start by considering each contiguous skin region as a candidate, then we prune this set according to two criteria: size and shape (i.e., template match). In the next step the coordinates of the pointing fingertip have to be found. This computation is based on the idea that the fingertip is generally the vertex of the hand's convex hull that has the largest distance from the center of mass. The center of the target area is determined in the last step by extending the line segment that begins at the hand's center of mass and passes through the fingertip.

While this recognition technique is generally stable, it sometimes reports false positives (i.e., shapes that are erroneously detected as a hand) or false negatives (i.e., the hand



(a) Pre-processed (b) Skin detection (c) Hand detection (d) Fingertip detection (e) Target area

Figure 1: Five main steps of the single frame processing procedure.

is visible but is not detected). During app testing we observed that, in some cases, the app has a sequence of correct recognitions interwoven with incorrect ones. This results in a confusing sequence of readings for the user. In order to make the output more stable, *JustPoint* aggregates results among successive frames.

3. EVALUATION

We conducted both 'computational-based' experiments - aimed at assessing the effectiveness of the computer vision technique - and a 'human-based' evaluation - aimed at evaluating *JustPoint*'s effectiveness and ease of use.

Computational-based experiments use a test-set of 100 images collected from about 30 sighted users. Half of these images contain a hand in the "pointing position" (as in Figure 1(a)), while the others do not contain a hand. Each of these image was manually annotated indicating the target area. Overall, precision is 0.87 and recall is 0.89 (Table 1).

Image set	**TP**	**TN**	**FP**	**FN**
Images with hand	42	0	3	5
Images without hand	0	47	3	0
All images	42	47	6	5

Table 1: Computational-based evaluation results

'Human-based' evaluation was conducted with 9 blind participants who, after a short training, were asked to complete 3 tasks, each consisting of recognizing, by using *JustPoint*, an object in a set of 3, indistinguishable by touch. For example, in the first task participants had to identify a box of strawberry candies, which is mainly red, distinguishing it from other two boxes. Credit cards were used in the second task and bottles in the third.

All the participants were able to complete the first and the second task by providing the correct answer in less than 2 minutes for each task (on average 79 sec. for task 1 and 82 sec. for task 2). In the third task 3 participants gave the wrong answer. This task also turned out to be more time demanding than the others (123 sec. on average). Based on the comments and on the report by the observer, two reasons account for the difficulties encountered in the third task. First, it was challenging for the participants to point the finger on the bottle and hold it steadily in one location; second, the participants failed to hold the bottle with one hand, while maintaining, with that hand, the correct pointing configuration; consequently in some cases *JustPoint* failed to recognize the correct target area.

4. CONCLUSIONS AND FUTURE WORK

We have described *JustPoint*, which was designed to solve the problem of how people with severe visual impairments can identify the colors of objects, including objects that are small or contain multiple colors in a small area. The evaluation shows promising results in terms of both hand detection reliability and app effectiveness judged by end users.

While the hand-detection technique is generally effective, most errors are due to errors in skin identification. So, we intend to experiment with alternative techniques for skin detection. Alternatively, existing solutions based on machine learning can be adapted for skin detection or for the entire process of target area identification. From the point of view of human-computer interaction, we wish to explore the use of a wearable (e.g., eyeglass-mounted or head-mounted) camera. We will also investigate the possibility of augmenting the text-to-speech interface with sonification techniques.

ACKNOWLEDGMENTS

James M. Coughlan was supported by NIH grant 5R01EY025332 and NIDILRR grant 90RE5024-01-00. Sergio Mascetti was supported by grant "Fondo Supporto alla Ricerca" under the project "Assistive Technologies on Mobile Devices".

5. REFERENCES

[1] A. Guo, X. Chen, H. Qi, S. White, S. Ghosh, C. Asakawa, and J. P. Bigham. Vizlens: A robust and interactive screen reader for interfaces in the real world. In *Proc. of the 29th Annual Symp. on User Interface Software and Technology*. ACM, 2016.

[2] S. K. Kane, B. Frey, and J. O. Wobbrock. Access lens: a gesture-based screen reader for real-world documents. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013.

[3] V. Khambadkar and E. Folmer. Gist: a gestural interface for remote nonvisual spatial perception. In *Proc. of the 26th annual ACM symposium on User interface software and technology*. ACM, 2013.

[4] N. A. A. Rahman, C. W. Kit, and J. See. Rgb-h-cbcr skin colour model for human face detection. In *MMU Int. Symp. on Inf. & Comm. Tech.*, 2006.