Supporting Pedestrians with Visual Impairment During Road Crossing: a Mobile Application for Traffic Lights Detection.

Sergio Mascetti[1,2], Dragan Ahmetovic[1], Andrea Gerino[1,2], Cristian Bernareggi[1,2], Mario Busso[1], and Alessandro Rizzi[1]

[1] Università degli Studi di Milano, Dept. of Computer Science, EveryWare Lab
[2] EveryWare Technologies, Milano, Italy

Abstract. Many traffic lights are still not equipped with acoustic signals. It is possible to recognize the traffic light color from a mobile device, but this requires a technique that is stable under different illumination conditions. This contribution presents *TL-recognizer*, an application that recognizes traffic lights from a mobile device camera. The proposed solution includes a robust setup for image capture as well as an image processing technique. Experimental results give evidence that the proposed solution is practical.

Keywords: Blind people, visual impairments, mobile device, smartphones, traffic lights, computer vision.

1 Introduction

Independent mobility involves a number of challenges for people with visual impairment or blindness, including being aware of the presence and the current color of traffic lights. This is particularly challenging when traffic lights are not equipped with acoustic signals. A number of solutions have been proposed in the scientific literature to recognize traffic lights (among others, [1–3]). Existing solutions share a common problem: they use images acquired through the device camera with automatic exposure. With this approach, in conditions of low ambient light (e.g., at night) traffic lights result overexposed while in conditions of high ambient light (e.g., direct sunlight) they are underexposed (see Figure 1(a)).

This contribution, extracted from our previous work [4], presents *TL-recognizer*, a software module that addresses the above problem with an effective solution: besides image processing and recognition, it proposes a robust setup for image capture that allows to acquire clearly visible traffic light images regardless of ambient light variability due to time and weather. The proposed recognition technique adopting this approach is reliable (full precision and high recall), robust (works in different illumination conditions) and efficient (it can run several times a second on commercial smartphones). The experimental evaluation conducted with visual impaired subjects shows that the technique is indeed practical in supporting road crossing.

(a) Acquisition with automatic exposure: high ambient light (left) and low ambient light (right).

(b) Acquisition with our technique: high ambient light (left) and low ambient light (right).

Fig. 1. Image acquisition with automatic exposure and with our technique in different illumination conditions.

2 The technique to recognize traffic lights

The recognition process is organized in two main phases: ‘input-acquisition’ and ‘image-processing’. During input-acquisition a frame is captured by the device camera using specifically designed exposure parameters. The overall idea is that, while light conditions during day and night are extremely variable, luminance coming from traffic lights is pretty stable. For this reason, instead of relying on smartphone camera automatic exposure, which balances the mean luminance of every point in the entire image hence possibly resulting in underexposed or overexposed traffic lights, our solution sets a fixed exposition value (EV) chosen among a small group of EVs pre-computed to encompass the luminance variations.

Figure 1(b) shows details of two pictures, each representing a green light in a different illumination conditions. The pictures were acquired with the technique described above and with parameters defined during the experiments (see Section 3). From left to right, the two light intensities are: very high (i.e, sunny day at noon), and low (night). These results are examples of the stable acquisition that is guaranteed by our technique, in contrast with what is obtained with automatic exposure (see Figure 1(a)).

The image-processing phase is aimed at identifying the active optical units (i.e., AOU - the active light in the traffic light), that appear in the image. The overall computation can be logically divided into three steps: extraction, pruning and validation.

During extraction of candidate AOUs for each traffic light color c (i.e., green, yellow and red), *TL-recognizer* identifies a set of image portions, each one representing a candidate AOU. This is achieved by first applying a filter that excludes the pixels with low luminosity values and incompatible hue. The result is a binary image for each color c. Then, white regions in binary images are segmented into blocks of contiguous pixels with the technique proposed by Suzuki and Abe [6]. The result is a list of contours, each one composed of a set of points.

In the pruning step the algorithm removes the contours that are too small or too big to represent an AOU. Technically, each contour is assumed to be an

AOU (whose size is known) and consequently its distance along the horizontal and vertical axes from the device camera is computed. The AOU is discarded if it is too far (i.e., contour is too small) or too close (i.e., contour is too big) from the user.

Finally, in the validation step, the image portion (the “patch”) corresponding to each contour is extracted, rotated and resized to the same size as a template (a different one for each AOU color). The similarity between the two figures (patch and template) is evaluated with the fast normalized cross-correlation technique [7]. The patch is considered to be an active optical unit if the result of the comparison is larger than a given threshold T whose value is defined through empirical evaluation (Section 3).

The application outputs the color of the detected traffic light through a multimodal interface: the color is read aloud using speech synthesis, it is visualized by coloring the entire screen of the device, and it is conveyed haptically through specific vibration patterns.

3 Parameters tuning and experimental evaluation

The empirical evaluation can be divided into three main sets of experiments: the first and the second are ‘computational-based’, the third one is ‘human-based’. The two ‘computational-based’ experiments are conducted with the following methodology: images of urban scenarios were recorded, divided into the “tuning” and “evaluation” datasets (see Table 1) and manually annotated with the position and the color of AOUs (if any). Then, *TL-recognizer* is run off-line, and its results are compared with the expected ones.

The first set of experiments is aimed at tuning the parameters. We omit the description due to page limit, details can be found in [4].

The second set of experiments is aimed at assessing the performance of *TL-recognizer* in terms of computation time, precision and recall. The average values are reported in Table 1. We can observe that the results are very similar between the “tuning” and the “evaluation” datasets. This means that, after the parameters tuning process, system performances are stable under different conditions. In both datasets the precision is 1, which means that no traffic light is erroneously recognized. Still, there is a high value for recall, which means that most traffic lights are correctly identified. Finally, the computation time (computed on a Nexus 5 device with Android 5) is in the order of 100ms, which means that about 10 frames can be processed each second.

Testset	**Images**	**Precision**	**Recall**	**Computation time (ms)**
Tuning	501	1	0.85	113
Evaluation	1252	1	0.81	107

Table 1. Performances of *TL-recognizer*

Finally, the last set of experiments involved 2 blind subjects and 2 low-visioned subjects (unable to see traffic lights). All subjects have been trained for about one minute on how to use the mobile application implementing *TL-recognizer*. Then, in a real urban intersection, subjects were asked to walk towards a crossroad and to determine when it was safe to start crossing in a given direction (straight, left or right) i.e., when a green traffic light appears right after a red one. For each attempt, a supervisor recorded whether the task was successfully completed and took note of any problem or delay in the process. Each subject repeated this task five times.

Overall, all subjects have been able to successfully complete the assigned tasks. The only exception was with the first attempt made by the first subject: since he was pointing the camera too high up and almost towards the sky, the traffic light was always out of the camera field of view. The problem was solved in the following experiments by simply explaining how to correctly point the camera. Note that this problem could also be solved by monitoring the pitch angle of the device and by warning the user if the he/she is pointing too high or too low.

Acknowledgments. The work of Andrea Gerino, Cristian Bernareggi and Sergio Mascetti was partially supported by grant “fondo supporto alla ricerca 2015” under project “Assistive technologies on mobile devices”. The work of Alessandro Rizzi was partially supported by grant “fondo supporto alla ricerca 2015” under project “Discovering Patterns in Multi-Dimensional Data”.

References

1. Kim, Y., Kim, K., Yang, X.: Real time traffic light recognition system for color vision deficiencies. In: Proceedings of the Fourth International Conference of Mechatronics and Automation, IEEE Computer Society (2007) 76–81
2. Ivanchenko, V., Coughlan, J., Shen, H.: Real-time walk light detection with a mobile phone. In: Proceedings of the Twelfth International Conference on Computers Helping People with Special Needs, Springer-Verlag (2010) 229–234
3. Roters, J., Jiang, X., Rothaus, K.: Recognition of traffic lights in live video streams on mobile devices. Transactions on Circuits and Systems for Video Technology **21**(10) (2011) 1497–1511
4. Mascetti, S., Ahmetovic, D., Gerino, A., Bernareggi, C., Busso, M., Rizzi, A.: Robust traffic lights detection on mobile devices for pedestrians with visual impairment. Computer Vision and Image Understanding (2015), http://dx.doi.org/10.1016/j.cviu.2015.11.017
5. Technical Committee CEN/TC 226 “Road equipment”: European Standard EN 12368:2006 on “traffic control equipment - signal head”. (2006)
6. Suzuki, S., Abe, K.: Topological structural analysis of digitized binary images by border following. International Journal on Computer Vision, Graphics, and Image Processing **30**(1) (1985) 32–46
7. Lewis, J.P.: Fast normalized cross-correlation. International Journal on Vision interface **10**(1) (1995) 120–123