Automatic bus line number localization and recognition on mobile phones—a computer vision aid for the visually impaired

Claudio Guida, Dario Comanducci, and Carlo Colombo

Dipartimento di Sistemi e Informatica,
Via Santa Marta 3, I-50139, Firenze, Italy
{guida, comandu, colombo}@dsi.unifi.it

Abstract. In this paper, machine learning and geometric computer vision are combined for the purpose of automatic reading bus line numbers with a smart phone. This can prove very useful to improve the autonomy of visually impaired people in urban scenarios. The problem is a challenging one, since standard geometric image matching methods fail due to the abundance of distractors, occlusions, illumination changes, highlights and specularities, shadows, and perspective distortions. The problem is solved by locating the main geometric entities of the bus façade through a cascade of classifiers, and then refining the matching with robust geometric matching. The method works in real time and, as experimental results show, has a good performance in terms of recognition rate and reliability.

Keywords: Visual machine learning, object recognition, geometric methods, accessibility software

1 Introduction

In the last few years, advanced technologies have greatly contributed to improve the mobility and autonomy of disabled people. For example, new interaction paradigms based on voice commands or eye movements have recently been developed for people with motor disabilities, and used to control wheeled chairs, or to communicate with the rest of the world through a computer.

Computer vision is one of the most important and useful technologies for the visually impaired (blind and low-sighted) people both in outdoor and indoor scenarios. The basic idea is that cameras can be used as additional eyes, whose images are automatically analyzed by the software so as to support the visually impaired in their everyday tasks. Most of the current work is focused on the development of specific methods for scene analysis/enhancement, and the construction of special devices that blind people can bring with themselves. In [5], a computer vision tool for producing automatic descriptions of video material is presented. Such descriptions can be useful to let blind users to follow better their favorite TV programs, especially for the parts containing little dialogue.

In [7], the authors describe a specific head-mounted device that can scan interesting parts of the scene and retrieve useful information for blind users. In [2], a special image contrast enhancement method is described, that dramatically improves the fruition of photos, text and other visual material by low-sighted people. The system described in [1] can be used to localize and recognize text in urban scenes; the system employs a portable computer placed in a back-sack, and one camera that is mounted on the user's shoulder.

In most urban scenarios, visually impaired people experiment everyday the difficulty of getting into the right bus. This is because the bus line number is typically provided only in a visual way to the people at the bus stop. In this paper, we propose an innovative method for localizing and recognizing in a completely automatic way the oncoming bus line number. Since the method uses only a standard smart phone (both for image grabbing and processing), no specific device is required. Image analysis is carried out by explicitly taking into account all the challenges arising in a real outdoor context, such as the illumination changes, highlights due to specular surfaces, the presence of occlusions, distractors, and perspective deformations. The method employs a careful combination of geometric and machine learning-based computer vision techniques, by which a satisfactory recognition performance is achieved.

The paper is organized as follows. In Section 2, an overview of the approach is provided, and then each of its main computational phases are described in detail. In Section 3 experimental results are presented. Finally, conclusions and future work are addressed in Section 4.

2 Our Method

Our method allows a visually impaired person standing at the bus stop with his mobile phone in hand to know the line number of the oncoming bus. As the bus is heard, the user directs the phone camera towards it, starting the acquisition from the mobile camera. The method can recover the line bus number from a single image of the bus. However, in order to attain a higher recognition reliability, several consecutive live images of the acquired sequence are processed separately, and the most frequent result is returned to the user via voice synthesis.

The method is split into several computational modules (Fig. 1).

First of all, the oncoming bus is localized inside the image and only the image region of interest including it ("Bus ROI") is considered for further processing ("Bus localization" module). As a result of this selection, unnecessary visual information is excluded from the next processing stages, thus speeding up computations. To further speed up the bus localization process, the original image (3 Mpixels) is reduced to a size of 0.75 Mpixels. The localization module exploits trained classifiers; it will be explained in detail in Section 2.1.

Once the bus ROI has been found, a further image cropping is performed ("Line number localization" module), in order to isolate the line number region ("Number ROI"). To this aim, a template description of the bus façade is matched with the image content of the Bus ROI (see Section 2.2).

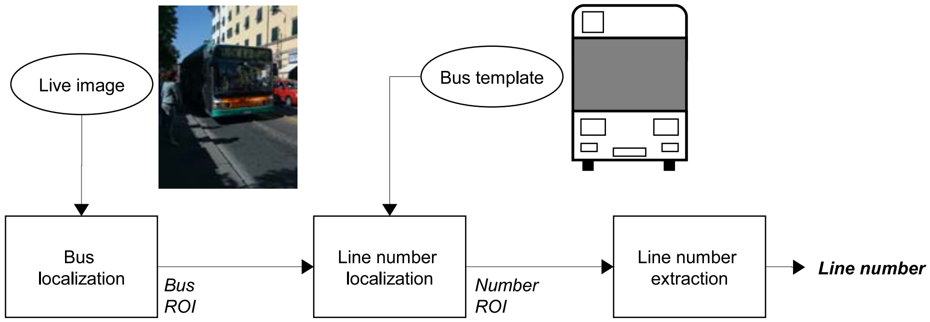


Fig. 1. Flow-chart of the proposed method.

The last stage of our method is dedicated to reading the number of the bus line ("Line number extraction" module), discussed in Section 2.3. For this task, we first binarize the Number ROI, and then use trained classifiers to read the bus line number from the binarized subimage.

2.1 Bus localization

To detect the bus presence and to localize it in the image we use the machine learning algorithm proposed by Viola and Jones in [10]. Originally proposed for face recognition, the algorithm exploits a cascade of weak classifiers and a boosting training method [4]. In our case the classifiers were trained for bus recognition.

Fig. 2 shows three examples of bus localization, related to photos taken from a mobile phone under different viewing/lighting conditions and for different bus models. The rectangle delimits the detected bus ROI. Notice that the bus is correctly localized even in the presence of occlusions (Fig. 2(a)), and in the presence of distractors as the van in Fig. 2(c). Distinguishing between real buses and van or trucks is a very hard task for blind people, that rely only on acoustic cues.

To train the classifier, we took several photos at bus stops as positive examples to feed the boosting training algorithm. To enforce robustness to pose and light variation, we also created 100 virtual views from each original image. We trained the model using about 1500 positive examples and 2100 negatives.

2.2 Line number localization

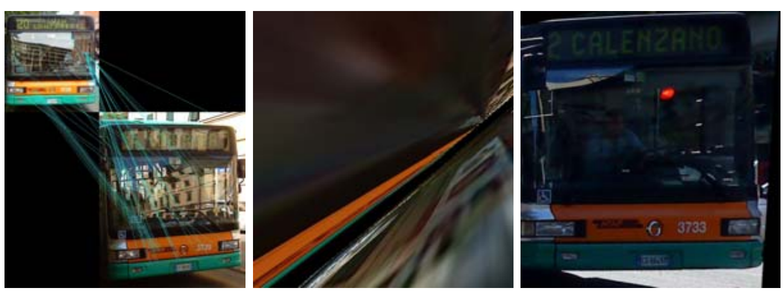
Once the bus is localized, all subsequent processing is limited only to the bus ROI. In order to make the method robust to pose variations, we use a template bus façade description to geometrically rectify the projective distortions arising as the result of image projection. Other image distortions that make line number localization challenging arise with variations in light and reflection phenomena.

(a) (b) (c)

Fig. 2. Viola & Jones bus localization: The classifier is run on three images of an urban scene. (a): Notwithstanding people occlusion, the bus is correctly localized. (b): Correct localization under different lighting conditions. (c): Correct localization of a different bus model. **Best viewed in color.**

Indeed, classic matching techniques with a frontal image of the bus façade are not useful in this context, due to the presence of important specularities, that make standard rectification methods prohibitive, due to the difficulty to extract robust features and match their descriptors. Figs. 3(a) shows an example where several wrong SIFT matches occur between a frontal view of a bus facade and a bus ROI image, thus making the rectification totally incorrect (Fig. 3(b)) even with the use of robust estimation techniques such as RANSAC [6] or LMedS [11]. This is because the number of outliers in the matching set is simply too high.

The template description of the bus façade allows us to overcome most difficulties: In Fig. 3(c) a good rectification of the façade, by exploiting our template, is shown.

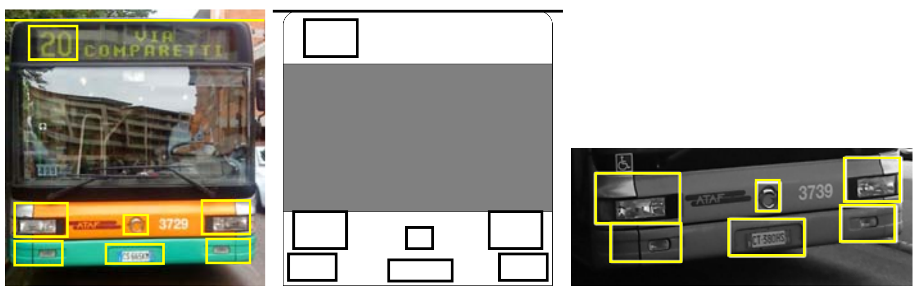


(a) (b) (c)

Fig. 3. Example of wrong rectification. (a): Fake matches. The upper-left part of the picture shows the image with bus façade frontal view, while in the bottom-right part the bus ROI to be matched is reported. (b): Wrong rectification of the bus ROI. (c): A correct rectification of the bus ROI by using the template. **Best viewed in color.**

The template was built by selecting some geometrically distinctive features of the façade (see Fig. 4(a)): Some elements of the bottom part (the four lights, the plate and the central logo) and the top line of the bus façade. The template description contains also the line number ROI (the rectangle in the top left region of the bus façade), to be used in the next step. For each of the template features the corresponding position within a normalized frontal view of the façade was annotated (Fig. 4(b)), and a separate classifier $\mathcal{C}_i$ was trained (with the exception of the top line and bus number area). The annotated positions are exploited to recover the rectifying homography of the bus façade.

Each classifier $\mathcal{C}_i$ applied to the bus ROI of the live image returns the minimum enclosing rectangle of the detected object from which its center $\mathbf{x}_i$ is calculated, and associated with its corresponding point $\mathbf{x}'_i$ in the template. Fig. 4(c) shows a result of the six classifiers for the bottom part of the template.



(a) (b) (c)

Fig. 4. (a): Annotation of the distinctive features of the bus façade. (b): Template model; the annotated elements with their position are shown with bold lines. (c): The distinctive elements found by the classifiers $\mathcal{C}_i$.s **Best viewed in color.**

To retrieve the upper border line in the bus ROI we use classic image analysis methods such as Canny edge detection and Hough transform [9]. High noise present in the upper side of the image, i.e. sky, trees, buildings, etc. makes line retrieval quite difficult. To mitigate this problem, we apply the graph-based segmentation discussed in [3]. Fig. 5 shows an example of intermediate image processing which leads to upper border line retrieval. The retrieved line $\mathbf{l}$ is put into correspondence with the top border line $\mathbf{l}'$ of the bus template.

The rectifying homography $\mathtt{H}$ of the bus façade is estimated by combining into a linear Least Squares problem the constraints [6]

$$\mathbf{l}' = \mathtt{H}^{-\top}\,\mathbf{l} \tag{1}$$

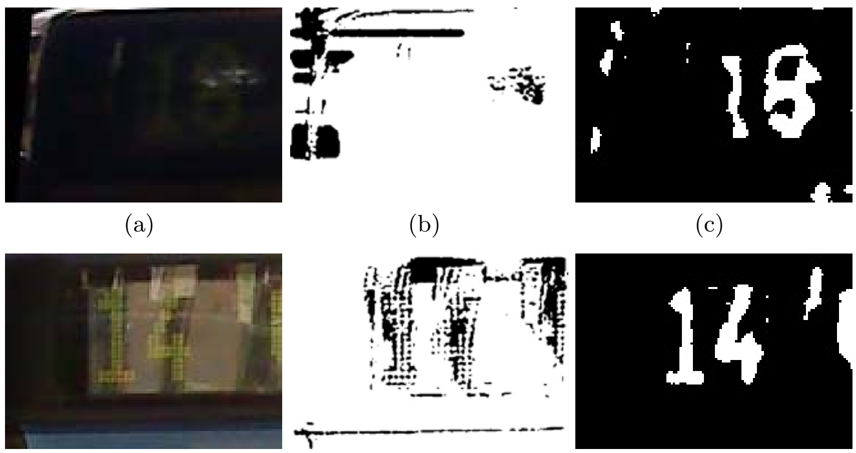
$$\mathbf{x}'_i = \mathtt{H}\,\mathbf{x}_i \quad i = 1 \dots 6. \tag{2}$$

(a) (b) (c) (d)

Fig. 5. Retrieving the upper line. (a): Result of graph-based segmentation. (b): Result of Canny edge detector. (c): Filtering of shorter lines. (d): Approximate upper border retrieved. **Best viewed in color.**

2.3 Reading the number

Once the rectified bus ROI is computed, we can safely localize the number ROI by simply selecting the corresponding area from the template, and mapping it onto the rectified ROI. Figs. 6(a) and 6(d) illustrate two examples of number ROI localization.



(a) (b) (c)

(d) (e) (f)

Fig. 6. (a, d): Shadows and specularities make reading the line number a difficult task. (b, e): standard OCR binarization techniques fail the extraction of the bus line number. (c, f): The binarization obtained with our method. **Best viewed in color.**

To perform line number reading, it is important that the contents of the number ROI are properly binarized. Binarization is also a common and crucial procedure in OCR techniques. As one can see from Figs. 6(b) and 6(e), the severe shadowing and mirroring phenomena make classic OCR binarization methods as [8] unuseful for the bus line number reading task.

We propose here an alternative and original binarization method based on complementary colors, that exploits the knowledge of the bus number color. Working in the HSV color space, we can easily find the complementary of the bus number color, and make an adaptive thresholding operation to filter all colors that don't have the same hue. Figs. 6(c) and 6(f) show the results of this new binarization method on the example illustrated in Figs. 6(a) and 6(d). Notice the dramatic improvement in image quality after binarization.

To make the process invariant to light variations, we also developed an adaptive thresholding method. The process tries to find the optimal thresholding value τ by iteratively binarizing the original image, until the number of white points in the entire image is inside a predefined interval. Fig. 7 shows an example of some iterations of our binarization algorithm. We can see that there

(a) (b) (c) (d)

Fig. 7. Some results of our binarization method obtained by varying the threshold τ. (a) $\tau = 20$, (b) $\tau = 50$, (c) $\tau = 70$ and (d) $\tau = 90$. Best thresholding value τ is 50.

is a particular thresholding τ value for which the number is quite clear. In this particular case $\tau = 50$.

Once the binarization operation is completed, the bus line number can be extracted. Once again machine learning is exploited for this task. In particular, we trained a classifier for each line number (not digit) to be recognized. The process of classification is as follows. Classifiers are put in a chain of ascending precision of classification, i.e., the first classifiers in the chain have better performance. We use then a first-win-takes-all schema to recognize the number.

3 Experimental results

In this section we will show experimental results obtained by applying our method to a dataset collected on the road.

The performance of the bus localization module is first addressed, and then results on the recognition of the bus line number are provided.

3.1 Bus detection and localization

To test the classifier for the bus localization task, we used various video sequences captured from mobile phones. Fig. 8 shows some results from three particular sequences. As one can see from Fig. 8(c), not all the frames of the sequence return good results. This is not a critical problem, because our method works on multiple images and the line number detection is activated as soon as a bus is recognized. For the same reason, false positive results are not a severe problem.

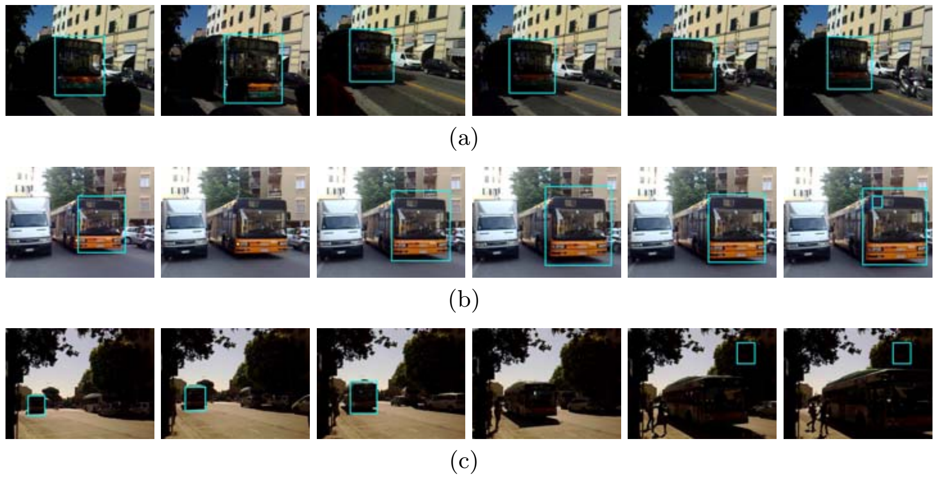


Fig. 8. Example of bus localization in three video sequences. **Best viewed in color.**

Table 1 illustrates detailed statistics concerning all the frames of the sequences. Bus detection rate is quite high in most of the sequences, with the exception of the first, which has just 6 frames over 32 with correct bus localization. Some frames of this sequence are indeed reported in Fig. 8(c): the very dark appearance of the bus façade, due to the sun in front of the camera, is the main reason of the low detection rate for that sequence. Although the low bus localization rate, our method successfully recovers the correct line number in all the 6 frames where the bus is detected.

Table 1. Results of bus detection for each video sequence.

Sequence	Frames	Detection	False Positive	Rate
No. 1	32	6	2	18%
No. 2	40	40	n/a	100%
No. 3	17	11	1	64.7%
No. 4	24	23	n/a	95.8%
No. 5	46	33	3	72%
No. 6	48	27	1	56.2%

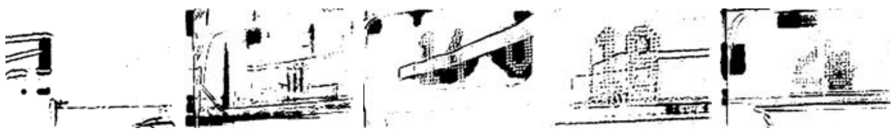
3.2 Recognition of bus line number

In this section the performance of our method for the bus line number recognition task is presented. A dataset composed by 50 images taken from frames not previously used in the training session (both for the bus localization and for the detection of the objects of interest) was exploited; in particular 10 images for 5 bus lines (2, 8, 14, 18 and 28) were used.

For each bus line considered in the dataset, an example of the rectified number ROI is provided in Fig. 9(a), while in Fig. 9(b) the corresponding binarized images are shown. Fig. 9(c) shows the binarization results obtained with the binarization technique of [8]. Working on the complementary color space makes our binarization algorithm invariant to specularities, shadows and changes in illuminations, while [8] fails.

(a)

(b)



(c)

Fig. 9. Some results obtained applying our proposed binarization method. (a): The rectified number ROI. (b): our binarization results. (c): the binarized image obtained with [8]. **Best viewed in color.**

Table 2 reports the detection rates on our dataset and shows that, with the exception of classifier relative to the line number 8, every bus line number is correctly recognized with 100% rate. The classifiers in the first-takes-all chain are ordered as 14 (the strongest), 18, 28, 8, 2 (the weakest). We found that several mutual false positives occur between the classifiers for the bus line numbers 2 and 8, because of the particular font used by the transport company, and probably this is the reason of the weak performance of the classifier for the bus line number 8.

4 Conclusions and future work

In this paper, an auxilium for visually impaired people aimed at locating and recognizing an incoming bus line number was presented. The method combines geometric computer vision with machine learning so as to achieve robustness with respect to highlights, specularities, shadows, occlusions, and so on. Experimental results show that the method has a higher reliability with respect to traditional

Table 2. Results of our rectification and binarization algorithm to the entire dataset.

Line Number	Detection rate
2	100%
8	75%
14	100%
18	100%
28	100%

geometric matching methods and standard OCR techniques. Future work will address using multiple bus templates, and dealing with several oncoming buses simultaneously. To achieve these goals we need to modify our detection and localization algorithm, in order to speed up the overall process. Bus localization will be performed within a pyramidal framework, and faster alternatives to the Viola and Jones approach will be investigated for the detection of template elements and the localization of the number region.

References

1. Chen, X., Yuille, A.: A time-efficient cascade for real-time object detection: with applications for the visually impaired. In: Proc. Conf. Computer Vision and Pattern Recognition (2005)
2. Choudhury, A., Medioni, G.: Color contrast enhancement for visually impaired people. In: Proc. 3rd Computer Vision Applications for the Visually Impaired (2010)
3. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. International Journal of Computer Vision 59, 167–181 (2004)
4. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Proc. 2nd European Conference on Computational Learning Theory. pp. 23–37 (1995)
5. Gagnon, L., Chapdelaine, C., Byrns, D., Foucher, S., Hritier, M., Gupta, V.: Computer-vision-assisted system for videodescription scripting. In: Proc. 3rd Computer Vision Applications for the Visually Impaired (2010)
6. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2004)
7. Pradeep, V., Medioni, G., Weiland, J.: Robot vision for the visually impaired. In: Proc. 3rd Computer Vision Applications for the Visually Impaired (2010)
8. Seeger, M., Dance, C.: Binarizing camera images for ocr. In: Proc. 6th International Conference on Document Analysis and Recognition. pp. 54–58 (2001)
9. Trucco, E., Verri, A.: Introductory Techniques for 3-D Computer Vision. Prentice Hall (1998)
10. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
11. Zhang, Z.: Parameter estimation techniques: A tutorial with application to conic fitting. Image and Vision Computing 15(1), 59–76 (1997)