Multimed Tools Appl
https://doi.org/10.1007/s11042-017-5472-5

Real-time pedestrian crossing lights detection algorithm for the visually impaired

Ruiqi Cheng[1] · **Kaiwei Wang**[1] · **Kailun Yang**[1] · **Ningbo Long**[1] · **Jian Bai**[1] · **Dong Liu**[1]

Received: 20 April 2017 / Revised: 7 October 2017 / Accepted: 27 November 2017

Abstract In defect of intelligent assistant approaches, the visually impaired feel hard to cross the roads in urban environments. Aiming to tackle the problem, a real-time Pedestrian Crossing Lights (PCL) detection algorithm for the visually impaired is proposed in this paper. Different from previous works which utilize analytic image processing to detect the PCL in ideal scenarios, the proposed algorithm detects PCL using machine learning scheme in the challenging scenarios, where PCL have arbitrary sizes and locations in acquired image and suffer from the shake and movement of camera. In order to achieve the robustness and efficiency in those scenarios, the detection algorithm is designed to include three procedures: candidate extraction, candidate recognition and temporal-spatial analysis. A public dataset of PCL, which includes manually labeled ground truth data, is established for tuning parameters, training samples and evaluating the performance. The algorithm is implemented on a portable PC with color camera. The experiments carried out in various practical scenarios prove that the precision and recall of detection are both close to 100%, meanwhile the frame rate is up to 21 frames per second (FPS).

Keywords Pedestrian crossing lights detection · Real-time video processing · Candidate extraction and recognition · Temporal-spatial analysis · Visually impaired people

1 Introduction

Lacking the capability to sense ambient environments effectively, the visually impaired feel inconvenient and always encounter various dangers, especially when crossing roads. Electronic Travel Aid (ETA) devices, since developed, have been taken as an efficient approach to assist the visually impaired in detecting accessible areas and obstacles [5, 6, 8, 12, 24–26]. However, the detection of Pedestrian Crossing Lights (PCL) is not assembled in most ETA systems. In urban areas, pedestrian crossing lights are ubiquitous, but not all of them are equipped with auxiliary devices.

✉ Kaiwei Wang
wangkaiwei@zju.edu.cn

[1] College of Optical Science and Engineering, Zhejiang University, Hangzhou, China

With the development of computer vision [22, 23], several related works have been dedicated to PCL detection. Shioyama et al. [20] made one of the first contributions to PCL detection algorithms, which is designed specifically for Japanese PCL. Using the primary image processing method, the robustness and efficiency of the algorithm are not guaranteed, so it cannot be applied to practical blind assistance. Designed for PCLs of the US, Ivanchenko et al. [10] proposed a real-time PCL detection algorithm which is based on mobile phone with accelerometer. Restricted by the limited computing power, the proposed algorithm only consider PCL detection in the middle of images. Roster et al. [17] proposed a mobile device based detection algorithm. The PCL are generated by color and shape analysis, and are verified by temporal analysis. To achieve high precision, complicated analytic image processing procedures, instead of machine learning schemes, are adopted in the algorithm. Although the parameters of classification procedures are optimized, the recall of the algorithm is not satisfactory. Moreover, the algorithm is designed for Germany PCL. Mascetti et al. [13, 14] also developed a complicated analytic image processing based algorithm on the mobile phone. PCL candidates are generated by color segmentation, pruned by geometrical properties and classified by template matching. Especially, the estimated distance and size of PCL are utilized in candidate classification. Exposure adjustment, which aims to resolve detection failures in dark environments, makes other objects invisible, thus the detection of other objectives, such as crosswalks or pedestrians, cannot be achieved. In addition, the algorithm is designed specifically for Italy PCL.

Almost all the previous works are applied to ideal scenarios, where a specified sort of PCL is captured with stable cameras at a moderate distance. However, when the users cross the roads with the camera, the PCL may suffer from challenging scenarios which are elaborated in Section 2.2. Furthermore, the PCL may have different styles in different countries. Almost all the previous works utilize complicated artificial classification strategies, which aims to improve precision in ideal scenarios indeed, but those strategies affects recall and frame rate and performs poorly in challenging scenarios. Many vehicle traffic lights detection algorithms, which are applied to autonomous vehicles navigation, deploy machine learning schemes to realize great accuracy in candidate classification. Among those algorithms, Histograms of Oriented Gradient (HOG) feature [7] is a commonly used descriptor [3, 19, 21], and Support Vector Machines (SVM) [3, 18, 19] is applied to train extracted features and classify candidates.

In this paper, we present a novel PCL detection algorithm for assisting the visually impaired to cross the roads. Compared with previous works, the superiority of our approach is apparent in the following respects.

(1) High precision and recall. High precision denotes few false alarms, which alleviates the misjudgment and warrants the user from hazards. High recall denotes few missing alarms, which reduces the omission of the PCL signal and makes the assistance efficient. The precision and recall are both above 90% on ordinary conditions.
(2) Real-time performance. The limited system resources of portable PC require an efficient algorithm to maintain a moderate frame rate and deliver instant feedback to the user. The proposed algorithm achieves the running speed of around 21 FPS on mobile devices.
(3) Robustness. In practical scenarios, the wearable or hand-hold cameras are unstable. Moreover, PCL may be located at arbitrary distances from the user in cluttered background. The proposed algorithm achieves satisfactory precision and recall when PCL suffer from severely shake in video.

(4) Low complexity. The proposed algorithm gets rid of complicated analytic image processing procedures and uses the concise machine learning scheme to achieve better performance compared with the previous works.

The remainder of this paper is outlined as follows. In section 2, the system architecture and dataset are presented. In section 3, we elaborate the pedestrian traffic lights detection algorithm, which is composed of candidate extraction, candidate recognition and temporal-spatial analysis. In section 4, experiment results in different types of urban scenarios are presented. In section 5, a brief conclusion is presented.

2 System architecture and dataset

In this section, the overall architecture of the wearable PCL aid system and its usage scenarios are presented. For assisting the visually impaired to cross the roads, the proposed real-time PCL detection algorithm has to be implemented on mobile devices. We achieve it by extending what was presented in [5, 24, 25], where we addressed the traversable area detection on a wearable navigation system.

2.1 Prototype system

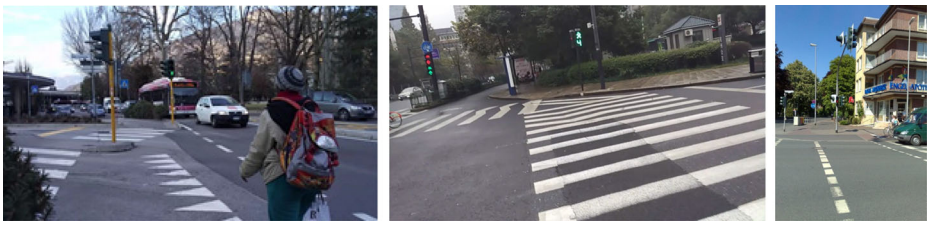
The wearable navigation system, as shown in Fig. 1(a), resembles a pair of glasses. It is constituted of Intel RealSense R200 camera [9], a pair of bone-conduction earphone and a portable PC. Based on the system, the PCL detection algorithm continuously receives video stream from the camera and gives feedback to the user through the earphone. The algorithm should detect PCL robustly when the user is standing at the opposite side of roads or walking across the roads. The PCL refers to the lamp part of a pedestrian crossing light device. Typical PCL have a human-like lamp (static or dynamic) with dark background, which are shown in Fig. 1(b).

2.2 Pedestrian crossing lights dataset

Common traffic lights databases, e.g. traffic lights recognition benchmark provided in [2], aim to serve vehicle traffic lights recognition. For the purpose of training and testing, the pedestrian

(a) (b)

Fig. 1 **a** Wearable blind navigation system. **b** Typical pedestrian crossing lights devices in urban areas, and the lamp parts are magnified



(a) (b) (c)

Fig. 2 Three video shots in the training set. The data are captured in (**a**) Italy, (**b**) China and (**c**) Germany

crossing lights dataset is established in this paper. The dataset is based on two sources for the generalization of algorithm. One source is the dataset collected in China and Italy by ourselves [4], the other source is the public dataset of Germany [16]. The data are captured by different types of cameras, e.g. Mi-4c mobile phone and Intel RealSense R200. The entire dataset is composed of two parts: training set and testing set.

Training dataset is established to tune parameters in the algorithm and to train the SVM models. Some frames of videos in the training dataset are presented in Fig. 2. PCL and other negative samples are extracted as described in section 3.1. For each generated sample, the ground truth is artificially labelled. The ground truth is referred to one of the three classes: red PCL (or yellow PCL), green PCL, and non-PCL. The statistics of labelled candidates are presented in Table 1.

The testing dataset is utilized to validate the performance of proposed algorithm comprehensively. In testing dataset, ground truths are labeled in every frame of video. The testing dataset is made of 22 videos, out of which 8 videos are captured by ourselves in the challenging scenarios [4] and 14 videos are obtained from [16]. Due to the movement or zooming of camera, the PCL have different sizes in images (see Fig. 3a, b). Other traffic lights, e.g. vehicle traffic lights or bicycle traffic lights, may appear in images (see Fig. 3c). Additionally, PCL are occluded occasionally by vehicles and pedestrians (see Fig. 3d). In some videos, the camera shakes violently, which causes blurry PCL (see Fig. 3 e).

The flicker effect is one of the main defects of acquired video, as presented in Fig. 4a. A continuously lighted PCL, whether red or green, shows diverse intensities among adjacent frames. When the brightness of PCL is very dark, the detection is prone to failures, which makes the successive recognition results of a video sequence unstable. Apart from the flicker effect, the dynamic PCL, as presented in Fig. 4b, also increase the difficulty of PCL recognition.

3 Pedestrian crossing lights detection

The proposed algorithm includes three steps: candidate extraction, candidate recognition and temporal-spatial analysis. In order to detect PCL in cluttered background, an effective extraction algorithm is required. Color segmentation based on binary thresholding is used to generate

Table 1 Details of training set samples

	Red PCL	Green PCL	Non-PCL
China set	130	153	7, 735
Italy set	575	244	15, 819
Germany set	377	100	9, 817
Total	1, 082	497	33, 371

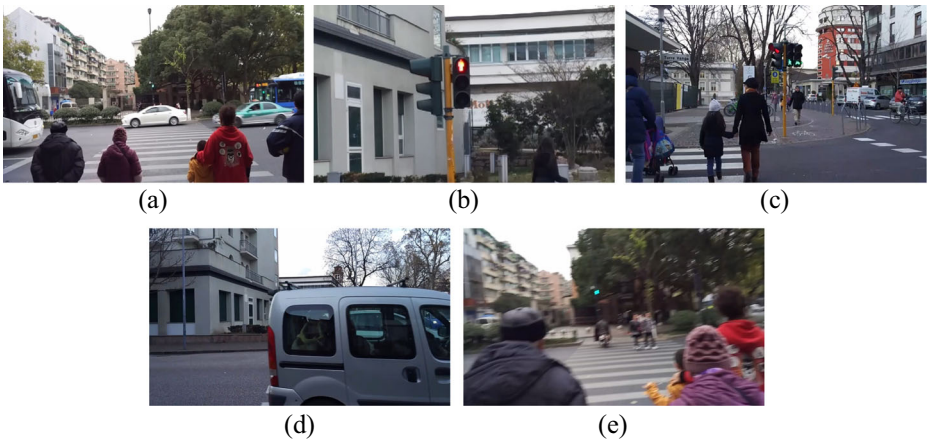


Fig. 3 The challenging testing dataset of PCL. **a** Small PCL. **b** Large PCL. **c** PCL with other types of traffic lights. **d** Occlusion. **e** Blurry PCL caused by movement

candidates. After candidate generation, the candidates need to be pruned by geometrical properties. In candidate recognition, the compounded HOG descriptor is extracted from a candidate, and SVM, as an efficient classifier on small scale database, is applied for the training and prediction of candidates. The detection based on single frame is not eligible for blind assistance due to the frequent detection failures. Temporal-spatial analysis guarantees stable detection results by comparing recognition results in current frame with those in former frames.

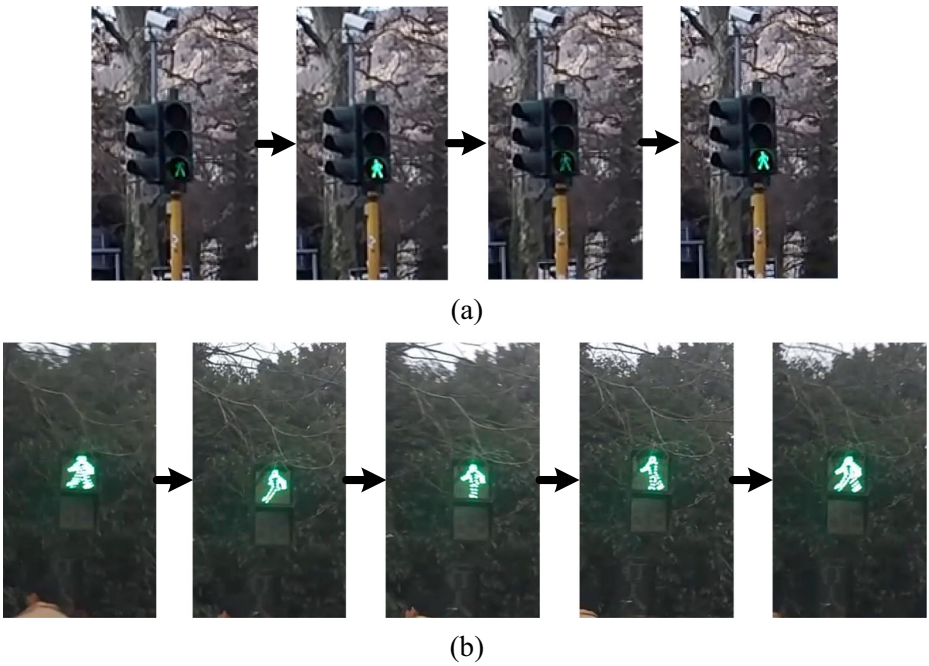


Fig. 4 Successive frames of continuously lighted green signals. **a** Diverse intensities between adjacent frames can be observed. **b** Different patterns of dynamic PCL are presented

3.1 Candidate extraction

Due to the small size of PCL in the background, candidate extraction is necessary to decrease time complexity and achieve real-time response. The flow chart of candidate extraction and process results are shown in Fig. 5. Extraction procedure includes color segmentation and geometrical properties analysis.

In order to extract PCL region with the specific color features, binary thresholding is a straightforward approach. Compared with the complex segmentation model in RGB color space [17], HSV color space with the dimensions of hue, saturation and value, makes setting thresholds convenient. The red and green PCL in training dataset are investigated, and their HSV color distributions are presented in Fig. 6. The PCL samples are randomly selected from training dataset, which are captured at different intersections under different illuminance conditions.

As shown in Fig. 6, the red and green PCL gather around specific values of Hue and Value. Hence, in order to extract PCL, setting thresholds in HSV space is feasible. Thresholds for color segmentation should be loose, so as not to eliminate possible PCL candidates. The thresholding rule for each pixel is

$$Binary(i,j) = \begin{cases} 1 & \left\{(i,j) : V\left(i,j\right) > 100 \; and \; H\left(i,j\right) < 200\right\} \\ 1 & \left\{(i,j) : V\left(i,j\right) > 100 \; and \; H\left(i,j\right) > 320\right\} \\ 0 & others \end{cases}. \tag{1}$$

Connected components are generated by clustering the qualified pixels. An example of connected component is presented as the solid area in Fig. 7. The geometrical properties of connected components, such as the width a, the height b and the area A of connected component, are analyzed to prune the unqualified ones. The qualified candidates are selected by three criteria:

$$Size = \frac{A}{Area_{\text{Image}}}, \tag{2}$$

$$AspectRatio = \frac{b}{a}, \tag{3}$$

$$FillingRatio = \frac{A}{ab}. \tag{4}$$

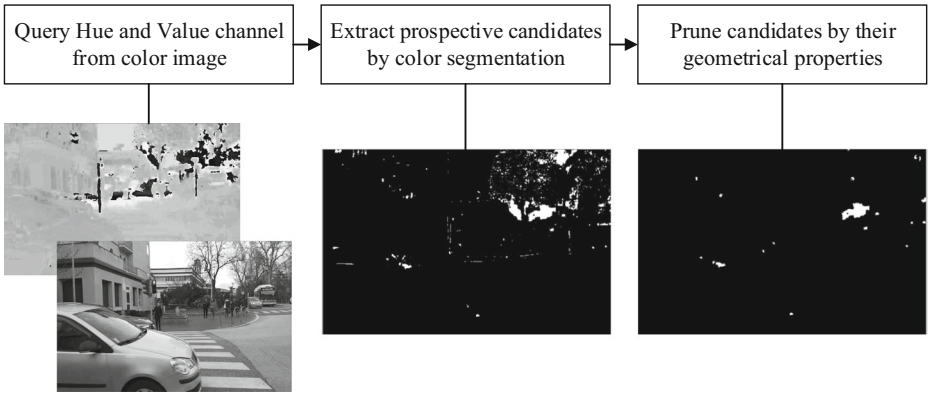


Fig. 5 Schematics of candidate extraction procedures and the results of every step. At the first step, the upper image is the hue channel of color image, and the lower one is the value channel of image

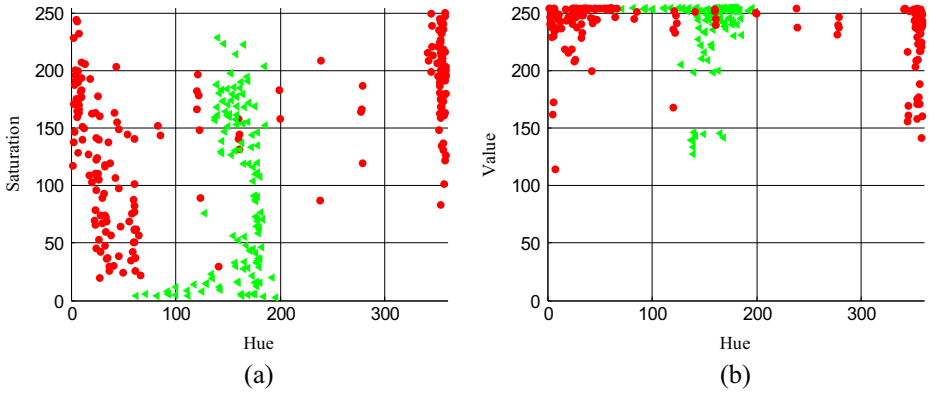


Fig. 6 The color distribution of sampled PCL in HSV color space. Red (green) points denote red (green) PCL. The color samples are presented in (**a**) Hue-Saturation view and (**b**) Hue-Value view

The thresholds of the three criteria are irrelevant to image size. Similar to color segmentation, in order not to eliminate inliers, the minimal bound and maximal bound of geometrical pruning are loose. Moreover, the pruning is not to select PCLs exactly, but to eliminate obvious outliers. Therefore, the parameters are determined by trial and error, and the values are shown as presented in Table 2.

Each qualified connected component generates a candidate of PCL. The enlarged candidate boundary (the solid line rectangle in Fig. 7) has the same center with the bounding box, but its edges have been enlarged proportionally as two times as that of the bounding box.

3.2 Candidate recognition by machine learning

Candidate classification is crucial to rule out outliers among candidates. In this paper, descriptor generation and machine learning scheme are deployed to classify the generated candidates into PCL or non-PCL.

Due to the similarity between pedestrians and PCL, HOG descriptor, which performs well in pedestrian detection [7], is chosen as the image descriptor of candidates. The compounded HOG descriptor is generated as Fig. 8. Each candidate is scaled into the same size (e.g. 32 by 32), and is split into red, green and blue channels. Two

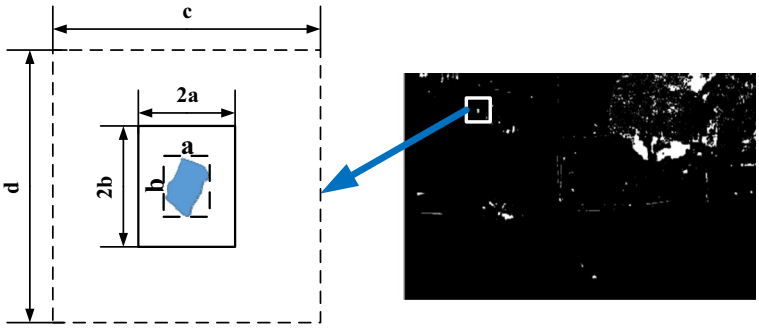


Fig. 7 The schematics of a candidate (bounded by solid line rectangle). The connected component is bounded by the inner dased rectangle

Table 2 Parameter list for the analysis of candidates geometrical properties

	Minimal bound	Maximal bound
Size	5×10^{-5}	0.01
Aspect ratio	0.5	2.5
Filling ratio	0.5	–

HOG descriptors are generated from the red and green channels of the candidate by the method proposed in [7]. The compounded HOG descriptor is obtained by stitching green and red HOG descriptors.

SVM is employed as the machine learning architecture in this paper in that SVM is sufficient to tackle the small-scale samples problem. The descriptors of red and green PCL samples with corresponding class labels are fed into SVM to train one-vs-all classifier models. Due to large dimensions of HOG, e.g. 864 dimensions in this paper, linear kernel function is adopted in the SVM models. Moreover, 10-fold cross validation is carried out during training course to get the optimized parameters. With the trained SVM models, the generated candidate is predicted as one of three labels: green PCL, red PCL or non-PCL.

3.3 Temporal-spatial analysis

Despite that SVM has been utilized to classify the candidates, occasional false classifications are still inevitable. In order to decrease false alarms as much as possible, the final detections of PCL state are synthesized by single-frame results. In this part, we present the temporal-spatial analysis, so as to further improve the precision and recall of detection.

3.3.1 Prospective region

As a container to record the results of former frames, prospective region (PR) is designed to achieve fast tracking. The prospective region denotes that the detected PCL may occur in that region. Multiple PRs are possible to coexist in one frame. For example, $B = \left\{ pr_k^{i-1} (1 \le k \le n) \right\}$ is the set of the PRs of the frame i-1, and the number of PRs is n. As shown in Fig. 9, PR has

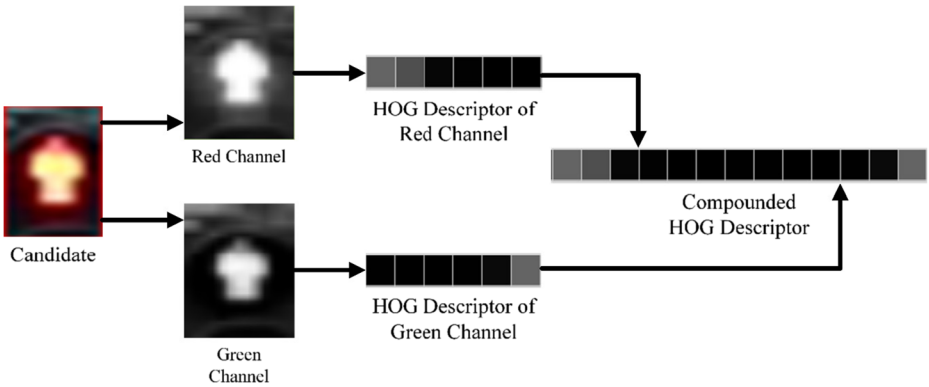


Fig. 8 The schematics of the generation of a compounded HOG descriptor

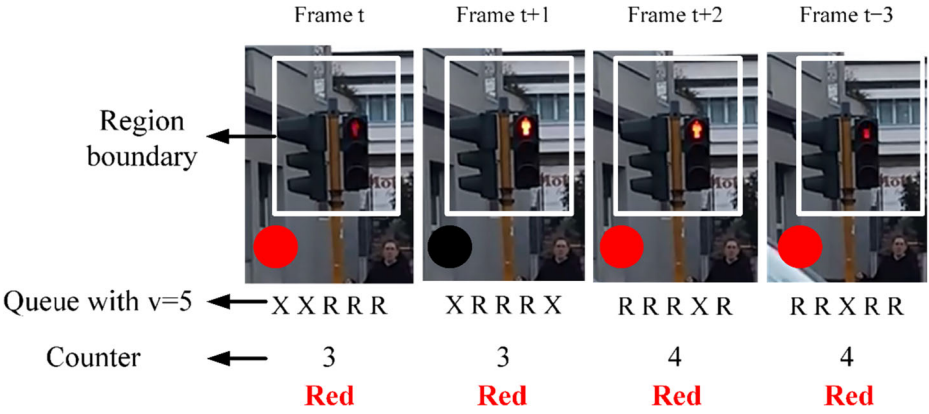


Fig. 9 A prospective region in successive frames. The colored circle at the bottom-left of each frame is the PCL recognition result of that frame. The red denotes that a red PCL is recognized, and the black denotes that no PCL is detected. The words at the bottom are the final results after temporal-spatial analysis

three properties: *region boundary*, *queue of detected PCL* and *counter of positive PCL*. The region boundary is the region of interest where the PCL are tracked. The queue with limited volume v accumulates detected PCLs within the region boundaries of former v frames. The counter counts the number of the positive PCL in queue, which reflects the possibility that the region contains PCL.

For tracking positive PCL (red or green PCL classified by SVM models) in the new frame, the features of PCL are matched with those of PCL in PR. In this paper, SURF (Speeded Up Robust Features) descriptors are chosen as the feature, in view of its fast speed and robustness against scale and rotation. Since PCL candidate (the solid line rectangle in Fig. 7) contains limited cues, we extracted SURF feature in the region between the two dashed rectangles of Fig. 7, where the edge c and d of outer boundaries are determined by

$$\begin{aligned} c &= \min(2l_x a, \alpha e) \\ \mathrm{d} &= \min(2l_y b, \alpha e) \end{aligned}. \tag{5}$$

In Eq. 5, 2a and 2b are the width and height of a PCL candidate, l_x and l_y are coefficients which expand the candidate. Besides, α is the bound factor, and e is the shorter edge of image, hence αe controls the area of expanded region to prevent much time consumed in descriptor extraction. Herein, α is set to 0.5 by trial and error. As the algorithm presented in [1], around the key points obtained by Hessian matrix in different scales, SURF descriptors are established by Haar wavelet responses.

3.3.2 Combination of new frame and former frames

The PRs are renewed when the PCL detection results of a new frame come. The renewal algorithm of prospective regions is presented in Algorithm 1. Firstly, as the PRs of the last frame are sorted by the counter value in descending order, the PR with more PCL detections has more priority to match PCL. Then, according to priority, the sorted PRs are matched with detected PCL within their boundary (also defined as S_k). In other words, the SURF descriptors of newly detected PCL are matched with the former descriptors saved in PR, and the matching

method is FLANN (Fast Library for Approximate Nearest Neighbors) [15]. For the k-th PR of frame i-1 (pr_k^{i-1}), the optimal matched PCL of frame i (pcl_k^*) is the one with the largest number of matching pairs among S_k, meanwhile its matching number should be above the threshold *th*.

Algorithm 1: the renewal of PRs (prospective regions)

Input:

$A_0 = \{pcl_q^i (1 \le q \le m)\}$: The set of the positive PCL detected in the frame i, and the number of PCL is m.

$B = \{pr_k^{i-1} (1 \le k \le n)\}$: The set of the PRs of the frame i-1, and the number of PRs is n.

$C_0 = \varnothing$: The set of the PRs of the frame i.

Method:

Sort B by the value of counter in descending order.

For k = 1 to n

S_k= {pcl| pcl is within the boundary of pr_k^{i-1}, $pcl \in A_{k-1}$}.

Find pcl_k^* from S_k by matching descriptor.

$pr_k^i \leftarrow$ combine pcl_k^* with pr_k^{i-1}

$A_k \leftarrow A_{k-1} \setminus \{pcl_k^*\}$.

$C_k \leftarrow C_{k-1} \cup \{pr_k^i\}$.

End for

t = #A_n. t is the number of elements in set A_n.

For q = 1 to t

Generate pr_q^i for pcl_q^i in A_n.

$C_{n+q} \leftarrow C_{n+q-1} \cup \{pr_q^i\}$.

End for

$C = C_{n+t} \setminus \{pr_k^{i*}\}$, where $\{pr_k^{i*}\}$ is the set of empty PRs in C_{n+t}.

Output:

$C = \{pr_k^i (1 \le k \le r)\}$: The set of the PRs of the frame i.

The prospective region pr_k^{i-1} is combined with the optimal matched pcl_k^* to generate pr_k^i, which includes:

(1) The region boundary of pr_k^i is in proportion to that of pcl_k^*, and shares the same center (the expansion rule is as Eq. 5).
(2) The prediction state and SURF descriptors of pcl_k^* are pushed to the queue, and the counter plus one.
(3) If there is no solution of pcl_k^* for pr_k^{i-1}, the empty state is pushed into the queue, the boundary and the counter remains unchanged.

The new prospective region (pr_q^i) is generated based on the non-matched PCL (pcl_q^i) in A_n, which includes:

(1) The region boundary of pr_q^i is in proportion to that of pcl_q^i, shares the same center (the expansion rule is as Eq. 5).
(2) The prediction state and SURF descriptor of pcl_q^i are pushed to the queue. The counter plus one.

If none of positive PCL (red and green PCL) exists in the queue of a PR, the corresponding PR is regarded as empty (noted as pr_k^{i*}). Obviously, pr_k^{i*} is an invalid region, and should be deleted from the set of PRs.

3.3.3 Confident PCL result

The existing PRs in set C are utilized to synthesize the most reasonable PCL state of the new frame. The PCL in the queue of PR are sorted by the order of frame, e.g. index i = *v* denotes the latest PCL for the PRs with the volume *v*. We define the confidence of red or green PCL as

$$Confidence_s = \frac{1}{v}\sum_{i=1}^{v} iPCL_{\mathrm{i,s}}, s\in\{red, green\}. \tag{6}$$

If the i-th detection result of the PR is red, then $PCL_{\mathrm{i,\ red}} = 1$ and $PCL_{\mathrm{i,\ green}} = 0$, and vice versa. Larger weights are given to the recent detected PCL, by using i multiplied by PCL. If the size of detection results in the PR is smaller than the predefined volume *v*, the PR is not considered until it is full.

If none of red and green PCL's confidence value is larger than the threshold *conf*, the corresponding PR is taken as unqualified and is neglected. If more than one qualified PRs exist, the PR with the largest counter value is chosen. The largest confidence of the chosen PR is selected and the corresponding color is the final detection conclusion of the frame. If none of prospective regions in the frame is qualified, no existence of PCL is the final detection conclusion of the frame.

3.3.4 Optimization of the parameters

For improving the precision and recall of PCL, the optimization of parameters is crucial. As presented above, the parameters include the volume of queue *v*, the threshold of matching pairs *th*, the confidence threshold *conf*, the expansion coefficients l_x and l_y for SURF extraction and

the expansion coefficients l_x and l_y for PR boundary. The values of these parameters are not related closely to the certain dataset, because image features are not needed during parameter tuning. Based on the training dataset, the different parameter combinations are utilized to find the optimized parameters, where both precision and recall are high.

In view of the large number of the parameters to be tuned, grid search is not suitable. An initial parameter combination is assigned to start the tuning procedure, and only one parameter is changing during a tuning course. After tuning a parameter, the optimized value is updated into the parameter combination, which is prepared for the next tuning. Over 2600 parameter combinations are tried, and the precision and recall (defined in Section 4) of different parameter combinations are shown in Fig. 10, where the red (green) points denote red (green) PCL.

There is a trade-off between precision and recall, but in our case we attach more importance to precision, since false alarms are more hazardous than poor sensitivity [17]. Hence, we choose the parameter combination listed in Table 3 as the optimum. The precision and recall of optimal parameters are shown as blue points in Fig. 10. Using the optimal parameters, the algorithm achieves a precision of 0.97 and a recall of 0.90 for green PCL detection, a precision of 0.99 and a recall of 0.90 for red PCL detection.

4 Experiment results

In order to verify the effectiveness of proposed algorithm, a series of experiments are carried out. In this part, we present the experimental statistics and further discussions.

Precision and recall are usually used as the criterions of detection performance. If the detected positive PCL is consistent with ground truth, it is defined as true positive, otherwise the positive PCL is defined as false positive. If nothing is detected and the ground truth is non-PCL, it is defined as true negative. Comparatively, if the ground truth is positive, it is defined as false negative. For multiple detection results (e.g. the results of a video), we count the total number of true positives, false positives and true negatives, and define them respectively as TP, FP and FN. Herein, for red or green PCL, we define the precision and recall of multiple detection results as Eqs. 7 and 8, respectively.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

The performance of proposed algorithm is validated on the captured testing set [4]. The detailed statistics of detection results are listed in Table 4, where the performance of the proposed algorithm is compared with that of the algorithm without temporal-spatial (TS) analysis. Obviously, the proposed algorithm achieves extremely high precision and recall. As shown in Table 4, it is evident that temporal-spatial analysis improves the robustness of detection. Especially, the recall values of red and green PCL are largely promoted, which enhances that the sensitivity of PCL detection.

The corresponding detection results of the eight testing videos are presented in Fig. 11, where the PCL results are drawn at top-left corner of images. The different PCL in Italy and

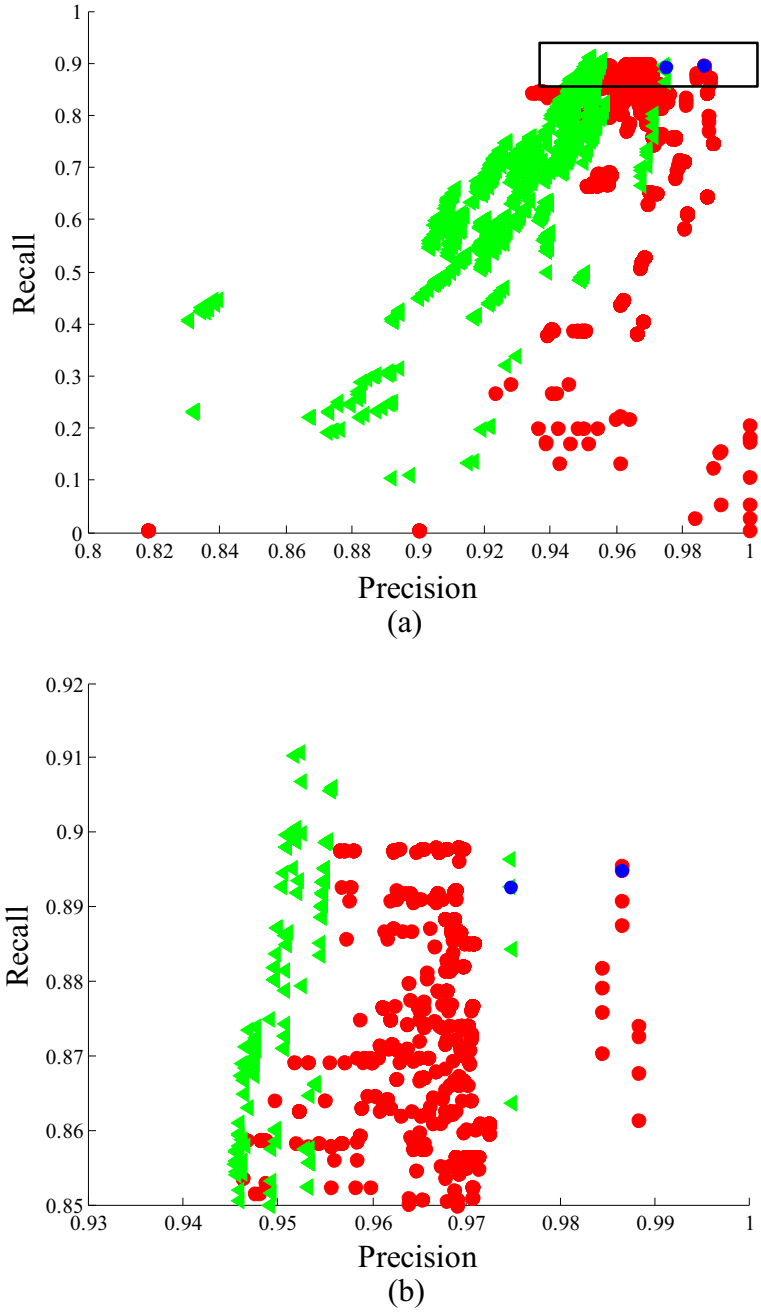


Fig. 10 Precision and recall of (**a**) all combinations and (**b**) partial combinations. The left blue point denotes green PCL, and the right blue point denotes red PCL

China are detected by the proposed algorithm. The algorithm performs well when the camera zooms in (see Fig. 11a, c) and the user walks close to PCL (see Fig. 11d, g), so the size of PCL does not affect the detection performance. Especially, vehicles traffic lights are not recognized

Table 3 The optimal parameter combination for the proposed algorithm

Parameter	v	$conf$	th	SURF extraction		PR boundary	
				l_x	l_y	l_x	l_y
Value	10	0.5	3	7	5	11	11

Table 4 Detection statistics of testing videos

Video			I	II	III	IV	V	VI	VII	VIII	Mean
Proposed algorithm	Red	Recall	97.8%	92.6%	99.5%	99.8%	88.4%	72.5%	–	98.4%	92.71%
		Precision	99.3%	99.8%	99.3%	99.5%	95.8%	94.8%	–	99.4%	98.27%
	Green	Recall	93.8%	100%	98.5%	99.9%	–	90.5%	88.9%	94.5%	95.16%
		Precision	99.5%	98.3%	99.0%	99.5%	–	99.8%	100%	94.3%	98.63%
Without TS analysis	Red	Recall	86.2%	79.8%	84.3%	60.3%	88.3%	40.4%	–	80.7%	74.29%
		Precision	100%	100%	100%	99.7%	96.5%	85.2%	–	97.2%	96.94%
	Green	Recall	85.9%	97.8%	99.7%	71.1%	–	81.6%	80.4%	78.2%	84.96%
		Precision	100%	96.9%	98.4%	99.1%	–	100%	100%	95.6%	98.57%

as PCL results (see Fig. 11b, h). When multiple PCL appear in images (see Fig. 11b, c, h), the algorithm correctly detects the frontal or the nearest PCL. The robustness against those challenging scenarios is derived from the trained SVM model and temporal-spatial analysis. The trained SVM model provides highly precise PCL recognition, which filter out extraneous

Fig. 11 Some typical PCL detection results of videos in test dataset. **a**–**d** represents detection results in video I-IV (Italy dataset). **e**–**h** represents detection results in video V-VIII (China dataset). The blue points denote the extracted SURF descriptors

Table 5 Recall and precision of our algorithm and Roters et al.'s algorithm

		Our algorithm	Roters et al.'s algorithm [17]
Red	Recall	90.3%	52.4%
	Precision	98.3%	100%
Green	Recall	57.3%	55.3%
	Precision	97.6%	100%

objects with similar color. Furthermore, temporal-spatial analysis salvages missed recognitions and filter out unstably recognized results among successive frames.

Furthermore, in order to compare our algorithm with Roters et al.'s algorithm presented in [17], we test our algorithm on their testing dataset [16]. As shown in Table 5, for both red and green PCL, our algorithm enhances the recall, meanwhile the precision remains high. In the dataset, the green PCL are darker than red one, which results in the unsatisfied recall of green PCL.

In order to validate the performance under challenging weather conditions, we run the algorithm on the dataset of the rainy and snowy scenarios [4, 16], and some detection results are shown as Fig. 12. The different PCL in Germany and China are detected by the proposed algorithm. As shown in Fig. 12, even if the PCL are located at the edge of images, the algorithm is able to detect them (see Fig. 12b). Despite video suffers shaking (see Fig. 12b), PCL are detected correctly.

As presented in Table 6, the algorithm achieves high precisions, but the recalls are not ideal under rainy and snowy scenarios. Due to the imperfect exposure of camera under poor illuminance, the blurry PCL affect candidate extraction, which results in low recall.

To apply to the practical blind assistance, the algorithm runs on the prototype system. The resolution of inquired color images need to be moderate, since sufficient processing speed should be guaranteed on portable PC. However, images with limited resolution reduce information volume and PCL may not be noticeable. Empirically, the resolution is set to 960 by 540, which is enough for most scenarios where PCL are located within the range of 15 m. For the PCL which are located at over 20 m away, the blur imaging caused by the optical aberration of camera makes recognition hard, meanwhile higher resolution does not improve the performance. In the prototype system, the portable PC is deployed with Intel Atom x5-

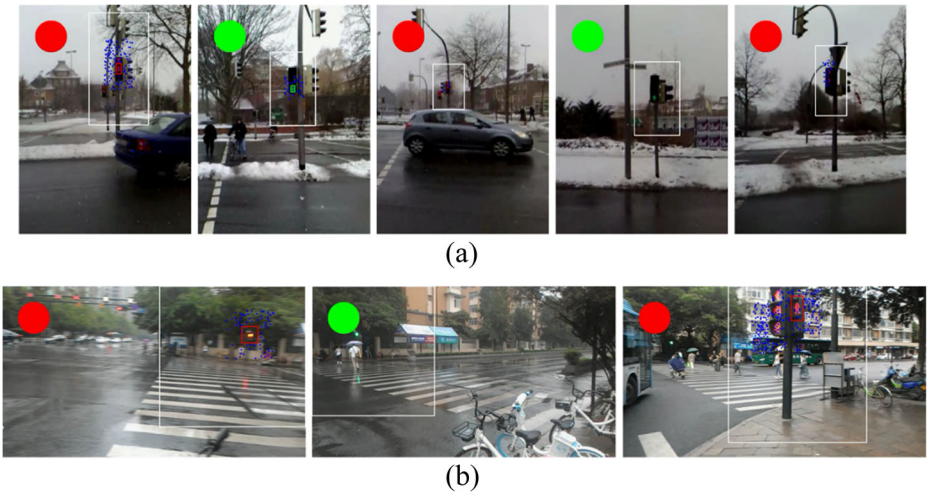


Fig. 12 Some typical PCL detection results of videos in test dataset

Table 6 Recall and precision of the proposed algorithm in the rainy and snowy scenarios

Scenarios		Snow	Rain
Red	Recall	84.4%	41.0%
	Precision	97.7%	95.5%
Green	Recall	38.4%	30.4%
	Precision	98.7%	100%

Z8500 processor and 2GB memory [11]. Under these conditions, the mean time consuming results is presented in Table 7.

Stage 1 is referred to candidate extraction, and stage 2 is referred to candidate recognition and TS analysis. The majority of time is consumed during candidate extraction, because of pixel-level operations in candidate generation. Compared with the frame rate of 0.6 FPS reported in [17], the overall frame rate of 21 FPS is outstanding for blind assistance. Although the visually impaired user does not need so frequent responses, a sufficient frame rate is still needed. Firstly, our prototype system not only runs PCL detection algorithm, but also runs traversable area detection algorithm and stereophonic interface [25]. Besides, temporal-spatial analysis requires that the interval of two successive frames is not long. Otherwise, camera movement may result in the large displacement of PCL detection in two frames, which causes tracking failed.

5 Conclusion

In this paper, a real-time PCL detection algorithm for the visually impaired is proposed. The proposed detection algorithm includes three procedures: candidate extraction, candidate recognition and temporal-spatial analysis. HSV based segmentation is utilized to extract PCL candidates, which are pruned by the geometrical features. The compounded HOG descriptor is used to represent candidates, and SVM model is trained to recognize PCL. In order to decrease false alarms and improve the performance in challenging scenarios, temporal-spatial analysis is applied to track the detected PCL. Moreover, a dataset constituted of the PCL of China, Italy and Germany is established in the paper.

The experiments carried out on the prototype system prove that the algorithm has real-time response, extremely low false alarms, and environmental robustness. On the captured testing dataset, for red PCL, the precision is higher than 98% along with the recall higher than 92%; for green PCL, the precision is higher than 98% along with the recall higher than 95%. Compared with Roters et al.'s algorithm presented in [17], the proposed algorithm achieves higher recall for both red and green PCL and much faster processing speed. On the prototype system, the frame rate of the proposed algorithm is up to 21 FPS. The proposed algorithm achieves the satisfactory performance in the challenging scenarios, such as various distance, occasional occlusion and shake, bad weather.

In the future, the proposed algorithm will be improved to achieve PCL detection in more complicated environments. In the dark environments, due to the overexposure of camera, PCL present severe blur in images, and are not detected by the proposed algorithm. Under rainy and snowy weather conditions, the recall of PCL detection is not ideal, which may lower the sensitivity

Table 7 Time consuming results on the prototype system

Stage 1 (ms)	Stage 2 (ms)	Frame rate (fps)
31	16	21

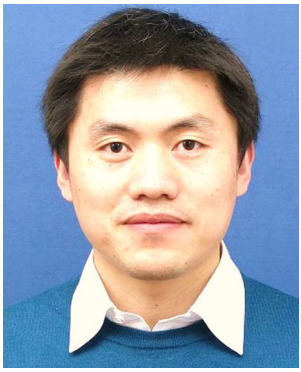
of blind assistance. Hence, adaptive candidate extraction procedures should be developed to extract the PCL in various challenging environments. Furthermore, crosswalks detection algorithm will be developed to provide comprehensive aid for visually impaired users when crossing roads.

References

1. Bay H, Tuytelaars T, Gool LV (2006) SURF: speeded up robust features. In: The 9th European conference on computer vision, Graz, Austria. Springer-Verlag, 2094476, pp 404–417. https://doi.org/10.1007/11744023_32
2. Charette Rd Traffic Lights Recognition (TLR) public benchmarks (2010) http://www.lara.prd.fr/benchmarks/trafficlightsrecognition. Accessed 7 Dec 2016
3. Chen Q, Shi Z, Zou Z (2014) Robust and real-time traffic light recognition based on hierarchical vision architecture. In: 7th International Congress on Image and Signal Processing (CISP), 14–16 Oct 2014, pp 114–119. https://doi.org/10.1109/CISP.2014.7003760
4. Cheng R (2016) Pedestrian traffic light recognition (PTLR) public database. http://www.wangkaiwei.org/file/PTLR%20dataset.rar. Accessed 11 Dec 2016
5. Cheng R, Wang K, Yang K, Zhao X (2015) A ground and obstacle detection algorithm for the visually impaired. In: IET International Conference on Biomedical Image and Signal Processing, 19 Nov. 2015, pp 1–6. https://doi.org/10.1049/cp.2015.0777
6. Chia-Hsiang L, Yu-Chi S, Liang-Gee C (2012) An intelligent depth-based obstacle detection system for visually-impaired aid applications. In: 13th international workshop on image analysis for multimedia interactive services, 23-25 May 2012, pp 1–4. https://doi.org/10.1109/WIAMIS.2012.6226753
7. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 I.E. computer society conference on computer vision and pattern recognition (CVPR'05), 25-25 June 2005, pp 886-893 vol. 881. https://doi.org/10.1109/CVPR.2005.177
8. Filipe V, Fernandes F, Fernandes H, Sousa A, Paredes H, Barroso J (2012) Blind navigation support system based on Microsoft Kinect. Procedia Comput Sci 14:94–101. https://doi.org/10.1016/j.procs.2012.10.011
9. Intel RealSense R200 (2016) https://software.intel.com/en-us/realsense/r200camera. Accessed 10 Apr 2017
10. Ivanchenko V, Coughlan J, Shen H (2010) Real-time walk light detection with a mobile phone. In: the 12th international conference on computers helping people with special needs, Vienna, Austria. Springer-Verlag, 1880791, pp 229–234
11. Kangaroo Kangaroo Mobile Desktop Pro (2016) http://www.kangaroo.cc/kangaroo-mobile-desktop-pro/. Accessed 18 Dec 2016
12. Leung TS, Medioni G (2014) Visual navigation aid for the blind in dynamic environments. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 23-28 June 2014, pp 579–586. https://doi.org/10.1109/CVPRW.2014.89
13. Mascetti S, Ahmetovic D, Gerino A, Bernareggi C, Busso M, Rizzi A (2016) Robust traffic lights detection on mobile devices for pedestrians with visual impairment. Comput Vis Image Underst 148:123–135. https://doi.org/10.1016/j.cviu.2015.11.017
14. Mascetti S, Ahmetovic D, Gerino A, Bernareggi C, Busso M, Rizzi A (2016) Supporting pedestrians with visual impairment during road crossing: a mobile application for traffic lights detection. In: the 15th International Conference on Computers Helping People with Special Needs, Cham, 13–15 July 2016. Springer International Publishing, pp 198–201. https://doi.org/10.1007/978-3-319-41267-2_27
15. Muja M, Lowe DG (2009) Fast approximate nearest neighbors with automatic algorithm configuration. In: International conference on computer vision theory and application (VISAPP'09). pp 331–340
16. Roters J (2011) Pedestrian lights database. http://www.uni-muenster.de/PRIA/en/forschung/index.shtml. Accessed 28 Mar 2017
17. Roters J, Jiang X, Rothaus K (2011) Recognition of traffic lights in live video streams on mobile devices. IEEE Trans Circuits Syst Video Technol 21(10):1497–1511. https://doi.org/10.1109/TCSVT.2011.2163452
18. Salarian M, Manavella A, Ansari R (2015) A vision based system for traffic lights recognition. In: SAI Intelligent Systems Conference (IntelliSys), 10–11 Nov 2015, pp 747–753. https://doi.org/10.1109/IntelliSys.2015.7361224
19. Shi X, Zhao N, Xia Y (2016) Detection and classification of traffic lights for automated setup of road surveillance systems. Multimed Tools Appl 75(20):12547–12562. https://doi.org/10.1007/s11042-014-2343-1

20. Tadayoshi S, Haiyuan W, Naoki N, Suguru K (2002) Measurement of the length of pedestrian crossings and detection of traffic lights from image data. Meas Sci Technol 13(9):1450. https://doi.org/10.1088/0957-0233/13/9/311
21. Wei Y, Kou X, Lee MC (2014) A new vision and navigation research for a guide-dog robot system in urban system. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 8-11 July 2014, pp 1290–1295. https://doi.org/10.1109/AIM.2014.6878260
22. Yan C, Xie H, Yang D, Yin J, Zhang Y, Dai Q (2017) Supervised hash coding with deep neural network for environment perception of intelligent vehicles. IEEE Trans Intell Transp Syst
23. Yan C, Xie H, Liu S, Yin J, Zhang Y, Dai Q (2017) Effective Uyghur language text detection in complex background images for traffic prompt identification. IEEE Trans Intell Transp Syst
24. Yang K, Wang K, Cheng R, Zhu X (2015) A new approach of point cloud processing and scene segmentation for guiding the visually impaired. In: IET international conference on biomedical image and signal processing, 19 Nov. 2015. Pp 1-6. https://doi.org/10.1049/cp.2015.0778
25. Yang K, Wang K, Hu W, Bai J (2016) Expanding the detection of traversable area with RealSense for the visually impaired. Sensors 16(11):1954. https://doi.org/10.3390/s16111954
26. Yang K, Wang K, Cheng R, Hu W, Huang X, Bai J (2017) Detecting traversable area and water hazards for the visually impaired with a pRGB-D sensor. Sensors 17(8):1890. https://doi.org/10.3390/s17081890

Ruiqi Cheng was born in China in 1992. He received the Bachelor degree from Zhejiang University in 2015, and is currently a Master student at Zhejiang University, China. His current research interests are image processing and machine learning on blind assisting technology.

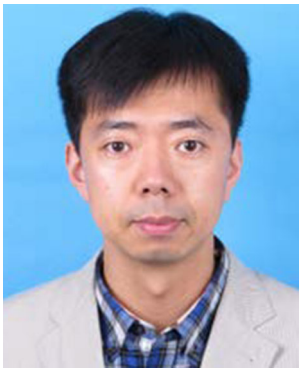


Kaiwei Wang received his BSc and PhD degree in 2001 and 2005 respectively, both at Tsinghua University, Beijing, China. He joined the Centre for Precision Technologies, University of Huddersfield, in October 2005 as a postdoctoral Research Fellow under the support of International Incoming Fellowship awarded by the Royal Society and then by EPSRC of UK. From 2009, he has been working with Zhejiang University as an associate professor. To date his research has been primarily concerned on intelligent guide for the visually impaired.

Kailun Yang was born in China in 1991. He received a bachelor's degree from School of Optoelectronics, Beijing Institute of Technology, in 2014, and is currently a PhD candidate at College of Optical Science and Engineering, Zhejiang University. His current research interests include stereo vision.

Ningbo Long was born in China in 1989. He received the M.S. degree from Tianjin University in 2015, and is currently a PhD candidate at the College of Optical Science and Engineering, Zhejiang University, China. His current research interests are the small and short range radar systems.

Jian Bai received his Master and PhD degree in 1992 and 1995 respectively, both at Zhejiang University, China. Since 1995, he has been working with Zhejiang University. His current researches focus on optical system and optical measurement.



Dong Liu received his BSc and PhD degree in 2005 and 2010 respectively, both at Zhejiang University, China. He joined National Aeronautics and Space Administration (NASA) in 2010 as a postdoctoral research fellow. Since September 2012, he has been working with Zhejiang University. His current researches focus on optical measurement and remote sensing.