

Crossing the Road Without Traffic Lights: An Android-Based Safety Device

Adi Perry, Dor Verbin, and Nahum Kiryati^(✉)

School of Electrical Engineering, Tel Aviv University, 69978 Tel Aviv, Israel
`nk@eng.tau.ac.il`

Abstract. In the absence of pedestrian crossing lights, finding a safe moment to cross the road is often hazardous and challenging, especially for people with visual impairments. We present a reliable low-cost solution, an Android device attached to a traffic sign or lighting pole near the crossing, indicating whether it is safe to cross the road. The indication can be by sound, display, vibration, and various communication modalities provided by the Android device. The integral system camera is aimed at approaching traffic. Optical flow is computed from the incoming video stream, and projected onto an influx map, automatically acquired during a brief training period. The crossing safety is determined based on a 1-dimensional temporal signal derived from the projection. We implemented the complete system on a Samsung Galaxy K-zoom Android smartphone, and obtained real-time operation. Promising experimental results provide pedestrians with sufficiently early warning of approaching vehicles. The system can serve as a stand-alone safety device, that can be installed where pedestrian crossing lights are ruled out. Requiring no dedicated infrastructure, it can be powered by a solar panel and remotely maintained via the cellular network.

Keywords: Pedestrian crossing · Traffic analysis · Optical flow · Blind and visually impaired · Resource-limited computer vision · Android device

1 Introduction

Safe road crossing requires conflict-free sharing of road surface. Pedestrian crossing lights stop traffic to ensure secure crossing. Other right of way regulations, such as zebra-crossings, are less reliable since drivers do not consistently yield to pedestrians [19]. Thus, for safe crossing in the absence of crossing lights, pedestrians need to identify occasional traffic gaps [7].

The duration of a traffic gap allowing safe crossing is determined by the crossing time. US data [1] indicates a representative crossing speed of about 1.2 m/s and urban road widths of 5.5 m (one-way single lane) or 8.5 m (two-way or two-lane one-way) with shoulders. The corresponding crossing times are 4.5 s and 7 s respectively. The necessary traffic gap, equivalent to the needed advance warning of an approaching vehicle, must be at least that long.

Recognizing an approaching vehicle 7 s ahead of contact time is possible for people with normal or corrected vision and solid cognitive function. For blind and visually impaired people, as well as for elderly persons, children [15] and others, sufficiently early recognition of an approaching vehicle can be implausible. Blind people are trained to rely on hearing and follow the “cross when quiet” strategy [18]. Auditory-based performance is less than ideal [4,9] and further difficulties arise where background urban noise masks the low noise emitted by a modern car [6], or when the pedestrian has less than perfect hearing.

Image and video processing methods to assist visually handicapped people were reviewed in [13]. Many techniques convert visual data to auditory or haptic signals. These include systems to locate street crossings [11,16]. Converting traffic light status to an auditory or other indication was considered in [3]. Sensor-rich “smart canes” have been developed [5] for close-range obstacle detection and path planning. However, the problem of road crossing at unsignalized crosswalks has remained open. A noteworthy exception is the work of [2], aimed at providing a robotic system with street-crossing capability.

We recently presented [17] a system positioned near a road crossing, alerting pedestrians of approaching traffic and detecting traffic-gaps suitable for safe crossing. The experimental results in [17] were obtained using a powerful laptop computer connected to a good-quality external USB camera with an added narrow field of view lens. The feasibility of real-time implementation on a low-cost and power-frugal platform remained an open question. In this paper we meet this challenge using an off-the-shelf Android device, establishing the potential for widespread deployment and stand-alone solar-powered operation. To complete the feasibility proof, we demonstrate successful autonomous operation at actual road crossings.¹

2 Stationary Location-Specific Solution

Consider a vehicle approaching a crosswalk at 50 km/h. Detecting this vehicle 7 s in advance, implies detection at about 100 m from the crosswalk. At that distance, the frontal viewing angle of a typical car is about 1° horizontally and slightly less vertically. In many suburban and country roads, speeds up to 100 km/h are not unusual, corresponding to roughly 0.5° at the required detection distance.

Assuming a straight road towards the crosswalk, the tiny-looking approaching vehicle appears near the vanishing point in the picture. Reliable detection requires predetermination of the point of first appearance, and a narrow field of view (FOV) around that point. With some exaggeration, this can be illustrated as watching the scene through a straw. One might wish for a smartphone application, warning the holder of approaching vehicles. However, typical mobile phones have wide FOV lenses. Worse, a hand-held device would require careful pointing towards the point of first appearance, a challenge for anyone, and a formidable undertaking for people with visual, cognitive or motoric impairments.

¹ Video: www.eng.tau.ac.il/~nk/crosswalk-safety.html.



Fig. 1. The system, implemented on a Samsung Galaxy K-zoom smartphone, attached to a lighting pole near a road crossing. The camera faces the approaching traffic. Note the extended position of the zoom lens.

We suggest a stationary and location-specific solution, where the system is attached, for example, to a traffic sign or a lighting pole, facing the approaching traffic. The absence of substantial ego-motion improves robustness, facilitates training and adaptation, and reduces the computational requirements, thus minimizing cost and power consumption. Beyond its technical advantages, the location-specific approach does not require a personal device, and is thus accessible to all members of society.

The essential system elements are a video camera with a narrow field of view lens, a processing unit, and a user interface. Optional elements include a compact solar panel and a rechargeable battery, enabling off-grid operation, and a communication unit (cellular, WiFi, etc.), allowing remote maintenance as well as cooperation with nearby instances of the system, e.g. in two-way multi-lane road crossings. As discussed in the sequel, we successfully implemented the system on the Samsung Galaxy K-Zoom smartphone, see Fig. 1.

Indication of incoming traffic, or lack thereof, can be delivered to the user as an audio, tactile or visual signal, or via a local networking interface (such as Bluetooth or WLAN) to a smartphone or a dedicated receiver.

3 Principles of Operation

Vehicles approaching a crosswalk emerge as tiny blobs at the point of first appearance. A key design choice has been to provide the pedestrian with the earliest possible warning of an approaching vehicle, rather than the latest warning that would allow crossing. This implies that the task at hand is early detection, rather than precise estimation of the time to contact (TTC) [10, 12].

We transform the incoming video to a scalar temporal signal, referred to as *Activity*, quantifying the overall risk-inducing motion in the scene. Vehicles approaching the camera lead to pulses in the Activity signal, so that substantial activity ascent reflects an approaching vehicle. Timely detection of incoming traffic, with sparse false alarms, calls for reliable separation between meaningful Activity ascent and random-like fluctuations.



Fig. 2. Influx map, computed by averaging the optical flow field over a training period. Left: arrow-style vector display. Right: color coded vector representation, corresponding to the color-key shown top-right. The motion direction in each pixel is represented by hue, its norm by saturation. (Color figure online)

3.1 Activity Estimation

Motion in the scene leads to optical flow. Parts of the optical flow field may correspond to vehicles approaching the camera. Other optical flow elements are due to harmless phenomena, such as traffic in other directions, crossing pedestrians, animals on the road, moving leaves and slight camera vibrations. The suggested activity signal A_n , where n is the temporal index, is computed as the projection of the optical flow field $\mathbf{u}_n(x, y)$ onto an *influx map*. The influx map $\mathbf{m}(x, y)$ is a vector field, representing the location-dependent typical course of traffic approaching the camera. The effective support of the influx map largely corresponds to incoming traffic-lanes. Mathematically,

$$A_n = \sum_{x,y} \mathbf{m}(x, y) \cdot \mathbf{u}_n(x, y). \quad (1)$$

The scalar activity signal quantifies the overall hazardous traffic in the scene.

Assuming a one-way street, the influx map can be obtained by temporal averaging the optical flow field over a training period, as illustrated in Fig. 2. Temporal averaging cancels randomly oriented motion elements due to vibrations, wind and similar effects, while maintaining the salient hazardous traffic motion patterns in the road area.

The influx-map vectors can be intensified to emphasize motion near the point of first appearance, thus increasing the advance warning time. Note that the point of first appearance can be identified during training, e.g. by back-tracking the influx-map vectors to the source. Minor adjustment of the influx map computation procedure is necessary for handing two-way roads, where consistent outbound traffic is also expected.

We have tacitly presented optical flow and influx map computation in *dense* form, meaning that their spatial resolution is uniform and commensurate with the resolution of the video frame. However, *sparse* computation of the optical

flow and influx map, at spatially variant density, is beneficial with regard to both performance and computational cost. Foveated computation of Eq. 1, dense near the point of first appearance and sparse elsewhere, has the desirable effect of emphasizing motion in that critical region.

3.2 Detecting Approaching Vehicles

Based on the raw Activity signal, we wish to generate a binary indication, suggesting whether risk-inducing traffic is approaching ('TRAFFIC' state), or a sufficient traffic gap allows safe crossing ('GAP' state).

As a vehicle approaches the camera, the peak of the corresponding Activity pulse is usually evident. However, *early* warning requires detection at an early stage of ascent, where the signal to noise ratio is poor.

Reliable, timely detection cannot be achieved by direct thresholding of the Activity signal. The signal is therefore evaluated within a sliding temporal window of N samples. This improves the effective signal to noise ratio, consequently increasing the detection probability and reducing the false alarm rate. However, sliding window computation inherently introduces detection latency. The extent N of the sliding window must therefore be limited, to maintain the advance warning period required for safe crossing.

Suppose that the sliding window falls either on a traffic gap ('GAP' hypothesis θ_0), modelling the Activity signal as $A_n = \nu_n$, where ν_n is zero-mean additive white Gaussian noise, or at the precise moment a vehicle emerges ('TRAFFIC' hypothesis θ_1), such that $A_n = s_n + \nu_n$, where s_n is the clean Activity pulse shape for the specific vehicle and viewing conditions.

We discriminate between the two hypotheses using the Likelihood Ratio Test

$$\text{LRT} = \frac{\frac{1}{(2\pi)^{N/2}\sigma^N} \prod_{n=0}^{N-1} \exp\left[-\frac{(A_n - s_n)^2}{2\sigma^2}\right]}{\frac{1}{(2\pi)^{N/2}\sigma^N} \prod_{n=0}^{N-1} \exp\left[-\frac{A_n^2}{2\sigma^2}\right]} \geq \lambda, \quad (2)$$

where λ is a discriminative threshold. This leads to the detection rule:

$$\tilde{y} \equiv \sum_{n=0}^{N-1} A_n s_n \geq \frac{1}{2} \left(2\sigma^2 \ln \lambda + \sum_{n=0}^{N-1} s_n^2 \right) \equiv \gamma, \quad (3)$$

where \tilde{y} denotes the correlation of the Activity signal with the pulse template. The pulse shape depends on the particular installation, and can be learned during training. For a given installation, specific vehicle characteristics lead to some variability. The right hand side is an application-dependent threshold denoted γ . We determine the threshold using the Neyman-Pearson criterion, tuned to obtain a specified false-alarm probability.

Given either hypothesis, the random variable \tilde{y} is Gaussian with variance $\sigma^2 \sum_{n=0}^{N-1} s_n^2$. The probability of false alarm is determined by the tail distribution of the 'GAP' hypothesis θ_0 :

$$P_{FA} = Pr(\tilde{y} > \gamma | \theta_0) = 1 - \phi\left(\frac{\gamma}{\sigma \sqrt{\sum s_n^2}}\right), \quad (4)$$

where $\phi(x)$ is the cumulative distribution of the standard normal distribution. Let $Q(x)$ denote its tail probability $Q(x) = 1 - \Phi(x) = 1 - \phi(x)$. The threshold γ can thus be expressed in terms of the admissible false-alarm rate, the noise and the Activity pulse template. The decision rule can be reformulated as

$$\frac{\sum_{n=0}^{N-1} A_n s_n}{\sqrt{\sum_{n=0}^{N-1} s_n^2}} \geq Q^{-1}(P_{FA}) \sigma, \quad (5)$$

which is the familiar matched filter result, correlating the input sequence with the Activity pulse template we wish to detect.

So far, we have discussed the most demanding detection condition, the earliest detection of an incoming vehicle. The Activity pulse increases as the vehicle approaches; thus, once the correlator output \tilde{y} crosses the threshold, it generally exceeds the threshold as long as the vehicle has not reached the crossing. When a vehicle appears late (e.g. leaving a parking space near the crossing), the Activity pulse rapidly ascends, and the correlator promptly crosses the threshold.

3.3 Setting Detector Parameters

The detector is tuned according to the noise variance and the Activity pulse template. During training, after the influx map had been computed, the system produces Activity measurements. These measurements are used to automatically estimate the Activity pulse template and the noise variance.

The system locates the salient local maxima of the Activity signal during training, and sets a temporal window containing each maximum, where most of the window precedes the maximum. We set the Activity pulse template as the median of the Activity signal windows, see Fig. 3.

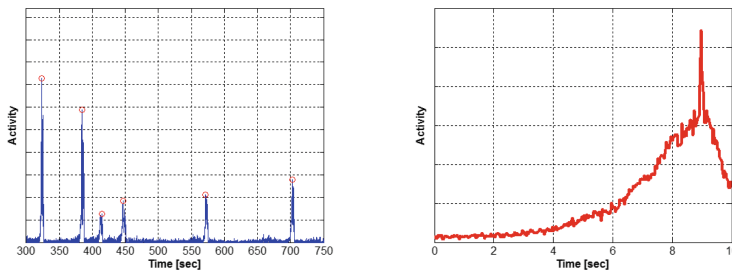


Fig. 3. Generating the activity pulse template. Left: The Activity signal over a small part of the training period, with the salient local maxima detected. Right: The Activity pulse template, generated as the median of the Activity signal windows

4 Examples and Evaluation

Figure 4 shows the activity signal computed during six minutes at a test site. The color represents the prediction issued to the pedestrian, red signifying TRAFFIC and blue meaning GAP. TRAFFIC is typically indicated about 8–10 s before the Activity signal peak, i.e. before the car reaches the camera, allowing secure crossing with sufficient margin. Multimodal activity pulses, e.g. at $t \approx 160$ s and at $t \approx 250$ s are each due to several cars arriving in sequence. Corresponding snapshots are shown in Fig. 5. The false alarms at $t \approx 80$ s and $t \approx 205$ s are not due to random noise: the first was caused by a person on the road, the second

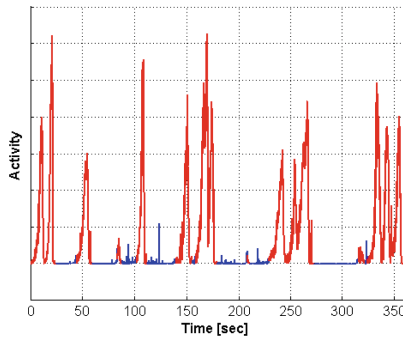


Fig. 4. The activity signal computed over six minutes, including the events shown in Fig. 5. TRAFFIC and GAP indications are represented by red and blue respectively. (Color figure online)



Fig. 5. Video frames corresponding to two multiple car events.

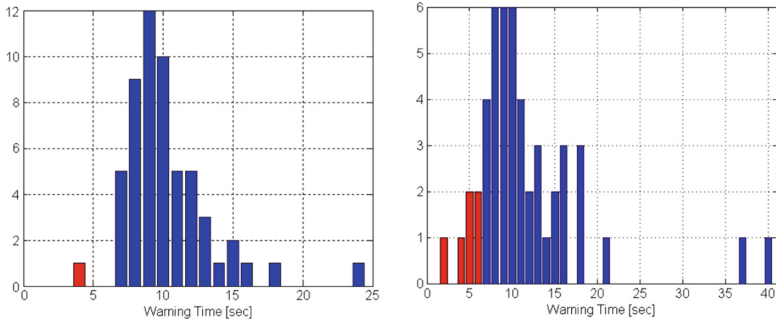


Fig. 6. Left: histogram of advance warning times provided by the real-time implementation during two hours at the test site shown in Fig. 5. Right: histogram of advance warning times obtained during one hour at the test site shown in Fig. 7.

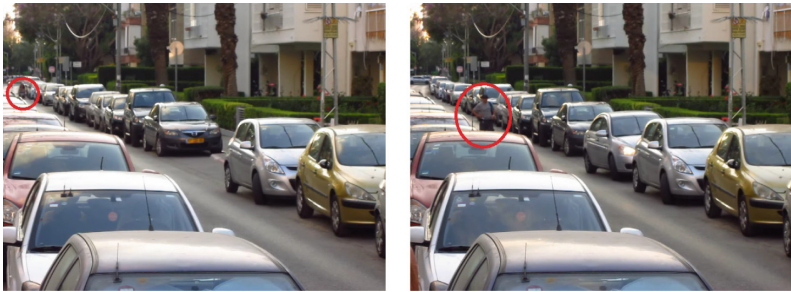


Fig. 7. Irregular events on the road. Left: a slow, distant bicyclist. Right: a person suddenly appearing on the road.

by a cat. The random noise level is rather low, as seen circa $t = 300s$ in Fig. 4. Figure 6(left) is the histogram of advance warning times provided by the system during two hours of testing.

In the archetypal case, an approaching vehicle emerges close to the point of first appearance, thus maximizing the achievable warning time. However, in various urban or suburban settings, a vehicle can exit a parking place close to the camera, or turn into the monitored road from an adjacent driveway. Thus, while the typical visibility period of an incoming vehicle is often greater than 10s, in unusual cases it can be as short as 1s, providing neither the system nor an alert human with sufficient warning. Figure 6(right) is the histogram of the advance warning times obtained during one hour at the challenging installation shown in Fig. 7, characterized by substantial occlusion near the point of first appearance and a residential afternoon activity pattern, including parking, biking and strolling. Short warning times usually correspond to bikes and very slow vehicles, that impose minimal hazard. Irregular events acquired by the system, leading to unusually short or long advance warnings, are exemplified in Fig. 7. Additional results, including ROC curves, can be found in [17].

5 Real-Time Implementation

To show that low-cost realization and wide deployment are possible, we implemented the system using an off-the-shelf Android device. Since the solution requires a narrow field-of-view lens, we selected the Samsung Galaxy K-zoom smartphone, that has a built-in variable zoom lens. Note that the variability of the optical zoom lens is not necessary, and the system can be readily built and operated using a fixed-zoom lens. Fixed-zoom lenses are widely available as low-cost smartphone accessories. Furthermore, unless remote management and distributed cooperative operation are sought, the cellular communication parts of the smartphone are redundant. In addition, although the K-zoom has a GPU capable of reducing computation time and increasing the frame rate, our implementation only utilized the smartphone's CPU. Thus, the system can be implemented on extremely low-cost Android tablet-like devices. In many applications, the screen of the Android device is not in use, and can be shut down or removed altogether to further reduce the power consumption and cost.

Other than its variable zoom lens, the Samsung Galaxy K-zoom, introduced in 2014, is similar to smartphones of its generation, with Android v4.4.2 (KitKat) as its native operating system. Figure 1 shows an installation of the system at a test site, attached to a lighting pole, with its zoom lens extended.

The target system software was developed using OpenCV4Android SDK version 2.4.10, FFmpeg and JavaCPP. The most demanding computational task is optical flow estimation. During training, dense optical flow estimation is required for generating the influx map. For that purpose, we used the OpenCV4Android implementation of Farnebäck's algorithm [8], skipping frames to improve the estimation of the tiny optical flow vectors near the point of first appearance. Following the identification of that point, a subset of 2000 image points was sampled using a Gaussian density function centered at the point of first appearance. In the online phase, sparse optical flow is evaluated only at those points, using the OpenCV4Android implementation of the Lucas-Kanade [14] algorithm. The Activity signal is computed by projecting the sparse optical flow field onto the corresponding sparse subset of the influx map. A rate of 8 frames per second is easily obtained on the K-zoom smartphone. To reduce power consumption, the rate can be decreased to two frames per second without significant performance degradation. The template matching process (3) is carried out at 30 frames per second using temporal interpolation. The results shown in Figs. 4 and 6(left) were obtained using the real-time smartphone-based implementation.

Two way streets, as encountered for example in the test site shown in Fig. 1, raise two issues. First, warning about traffic approaching the crossing from the opposite direction can be obtained by installing an additional instance of the system with its camera pointing in that direction. Cellular, WiFi or Bluetooth communication between the two systems allows issuing a mutually agreed indication to the pedestrian. Second, in a two way road, as the camera monitors incoming traffic, distancing vehicles may be seen in the field of view. To distinguish incoming from outgoing traffic based on optical flow, one might compute vector field measures such as divergence. In practice, assuming that the camera

is installed higher than typical vehicles, the distinction can be based on the sign of the vertical component of the optical flow vector: incoming vehicles are associated with descending optical flow vectors. In the Android implementation, upwards-pointing vectors in the influx map are simply nullified. Indeed, the system works well at the test site shown in Fig. 1 using that method.

6 Discussion

The global number of visually impaired people worldwide has been estimated to be 285 millions [20], and many more have age-related disabilities limiting their road-crossing skills. Reliable mobile personal assistive devices for road crossing without traffic lights are yet to be invented, but in any case are likely to be beyond the means of needy people worldwide. The goal of this research has been the development of an effective, easy to use, low cost solution. It is facilitated by the stationary, location-specific approach, avoiding system ego-motion altogether. System installation is simple, as the location-dependent influx map and the Activity pulse template are automatically obtained by training.

Compared to our previous work [17], the breakthrough is real-time stand-alone implementation using an off-the-shelf Android device. The retail price of a suitable Android tablet, including all necessary components, is less than US\$40 at the time of writing, and the marginal cost of a fixed narrow field lens is negligible. The mass-production manufacturing bill of materials, especially for a screen-less version of the system, is therefore likely to be widely affordable.

Interestingly, in their future work section, [2] imagined a stationary solution for the unsignalized road crossing problem. Our approach is substantially more robust, computationally lighter and better exploits stationarity when compared to the tracking-based method applied by [2] in their robotic system.

We evaluated the system at several test locations, obtaining promising results. Additional work is needed to ensure reliable operation at night or in the presence of rain or snow.

References

1. American Association of State Highway and Transportation Officials. A Policy on Geometric Design of Highways and Streets. Washington, DC (2004)
2. Baker, M., Yanco, H.: Automated street crossing for assistive robots. In: Proceedings of the 9th International Conference on Rehabilitation Robotics. IEEE (2005)
3. Barlow, J., Bentzen, B., Tabor, L., Signals, A.P.: Synthesis and Guide to Best Practice. National Cooperative Highway Research Program (2003)
4. Barton, B., Ulrich, A., Lew, R.: Auditory detection and localization of approaching vehicles. *Accid. Anal. Prev.* **49**, 347–353 (2012)
5. Buchs, G., Maidenbaum, S., Amedi, A.: Obstacle identification and avoidance using the ‘EyeCane’: a tactile sensory substitution device for blind individuals. In: Auvray, M., Duriez, C. (eds.) *EUROHAPTICS 2014*. LNCS, vol. 8619, pp. 96–103. Springer, Heidelberg (2014). doi:[10.1007/978-3-662-44196-1_13](https://doi.org/10.1007/978-3-662-44196-1_13)

6. Emerson, R., Naghshineh, K., Hapeman, J., Wiener, W.: A pilot study of pedestrians with visual impairments detecting traffic gaps and surges containing hybrid vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* **14**(2), 117–127 (2012)
7. Emerson, R., Sauerburger, D.: Detecting approaching vehicles at streets with no traffic control. *J. Vis. Impair. Blindness* **102**(12), 747–760 (2008)
8. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Bigun, J., Gustavsson, T. (eds.) SCIA 2003. LNCS, vol. 2749, pp. 363–370. Springer, Heidelberg (2003). doi:[10.1007/3-540-45103-X_50](https://doi.org/10.1007/3-540-45103-X_50)
9. Hassan, S.: Are normally sighted, visually impaired, and blind pedestrians accurate and reliable at making street crossing decisions? *Invest. Ophthalmol. Vis. Sci.* **53**(6), 2593–2600 (2012)
10. Horn, B., Fang, Y., Masaki, I.: Time to contact relative to a planar surface. In: *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 68–74 (2007)
11. Ivanchenko, V., Coughlan, J., Shen, H.: Crosswatch: a camera phone system for orienting visually impaired pedestrians at traffic intersections. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1122–1128. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-70540-6_168](https://doi.org/10.1007/978-3-540-70540-6_168)
12. Lee, D.: A theory of visual control of braking based on information about time-to-collision. *Perception* **5**, 437–459 (1976)
13. Leo, M., Medion, G., Trivedi, M., Kanade, T., Farinella, G.: Computer vision for assistive technologies. In: *Computer Vision and Image Understanding* (2016, in press)
14. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *Proceedings of the 7th IJCAI*, vol. 2, pp. 674–679 (1981)
15. Morrongiello, B., Corbett, M., Malinovic, M., Beer, J.: Using a virtual environment to examine how children cross streets: advancing our understanding of how injury risk arises. *J. Pediatr. Psychol.* **41**(2), 265–275 (2016)
16. Murali, V., Coughlan, J.: Smartphone-based crosswalk detection and localization for visually impaired pedestrians. In: *Proceedings of the International Conference Multimedia Expo Workshops*, pp. 1–7. IEEE (2013)
17. Perry, A., Kiryati, N.: Road-crossing assistance by traffic flow analysis. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8927, pp. 361–374. Springer, Cham (2015). doi:[10.1007/978-3-319-16199-0_26](https://doi.org/10.1007/978-3-319-16199-0_26)
18. Sauerburger, D.: Developing criteria and judgment of safety for crossing streets with gaps in traffic. *J. Vis. Impair. Blindness* **93**(7), 447–450 (1999)
19. Schroeder, B., Roupail, N.: Event-based modeling of driver yielding behavior at unsignalized crosswalks. *J. Transp. Eng.* **137**(7), 455–465 (2011)
20. World Health Organization. Global data on visual impairments (2010). <http://www.who.int/blindness/GLOBALDATAFINALforweb.pdf>