

Vision-based monitoring of pedestrian crossings

Alessandra Fascioli, Rean Isabella Fedriga and Stefano Ghidoni
VisLab - Dip. Ingegneria dell'Informazione - Università di Parma

Abstract

This paper presents a vision-based intelligent pedestrian crossing able to derive information on the occupancy status of its zebra crossing. The detection of pedestrians can help enhancing safety for road users who need longer to cross the road. At the same time, it produces savings in traffic delay. A stereo vision approach is adopted: the difference between two bird's eye views of the crossing is analyzed to extract objects rising up from the road plane. Background subtraction is applied to get rid of noise due to small drifts in the calibration and lens distortion. The discrimination between pedestrians and vehicles is performed taking into account their direction of movement.

1 Introduction

This work has been developed as part of the L.A.I.C.A. project¹, a research initiative aimed at introducing Ambient Intelligence solutions in the urban environment. The idea is to deploy throughout the city an extended network of intelligent sensors capable of acquiring, processing, and transmitting data. Sensors can be queried via distributed software, and a central unit collects and handles the data. The Ambient Intelligence system is able to interact with private users or public entities by supplying information and services on demand, and produces improvements such as increased safety and better mobility.

The system here presented is an intelligent pedestrian crossing able to derive information on the occupancy status of the zebra crossing thanks to computer vision.

Statistics assert that a significant proportion of accidents involving pedestrians occur at signalized intersec-

tions, where the conflict between vehicles and pedestrians is high [5]. The detection of pedestrians at zebra crossings can help enhancing safety for pedestrians who need longer to cross the road, e. g. disabled or elderly people, by increasing the clearance period when needed, and sending warnings to drivers. At the same time, it produces savings in traffic delay, as green time could be reduced when the crossing ends before the green man disappears, avoiding unnecessary vehicle delay.

Previous research on traffic signal control was mainly focused on vehicle monitoring, while little research was conducted on the pedestrian side. Most of the systems developed to detect pedestrians at intersections rely on a single fixed camera, and the observation point is quite far from the monitored area, for example on a building [8, 2, 4, 7]. These systems generally use motion detection algorithms to segment objects, based on the construction of a continuously updated reference image of the background. In some cases a statistical background model [8] or a reference image composed of the stationary edges in the scene [7] are used to gain robustness with respect to the changes in illumination that typically characterize outdoor environments. One of the major concerns about background subtraction is that shadows are also detected as part of the moving objects, and this is critical if the following object classification bases on size or shape. For this reason, specific techniques for shadow removal are often adopted [2, 6].

Conversely, the proposed system consists of a pair of cameras placed right over the zebra crossing. This setup suits every pedestrian crossing, even if far from buildings or hidden by trees. A stereo vision approach is adopted to segment objects: perspective is removed from the images acquired, and their difference is analyzed to extract objects rising up from the road plane. Background subtraction is then applied to the stereo difference to get rid

¹The L.A.I.C.A. (Laboratorio di Ambient Intelligence per una Città Amica) project has been funded by the Regione Emilia-Romagna, Italy.

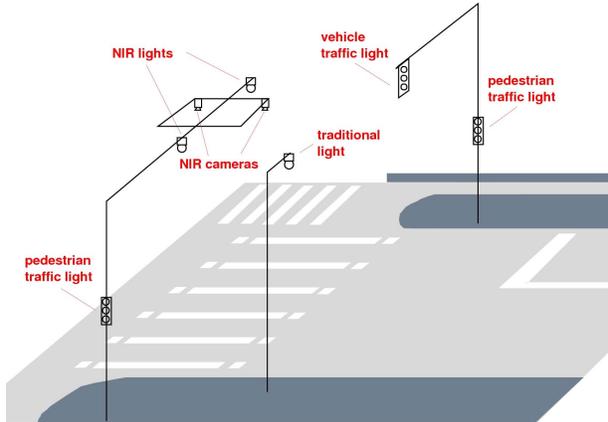


Figure 1: System configuration with two NIR cameras and two NIR lights overlooking the crossing.

of noise due to small drifts in the calibration and lens distortion. With this technique shadows are not detected as objects, as they lie on the road surface.

The rest of this paper is organized as follows. Section 2 describes the system setup. In section 3 the algorithm structure is discussed. Results obtained are presented in section 4. Section 5 contains some final considerations and ideas for future developments.

2 System setup

A very general setup has been studied to fit different pedestrian crossings. A pair of cameras are mounted on an L-shaped post overlooking the zebra crossing (see figure 1). Near infrared cameras have been selected and near infrared illumination has been provided to allow vision during night-time, however this capability has not been exploited in this first stage of the project.

An intersection including a bicycle crossing has been selected for the prototype system envisaging future extensions, nevertheless bicycle markings have not been considered in this stage, not to loose in generality.

Since the cameras are only six meters high on the ground, fish-eye optics have been chosen to be able to frame the complete zebra crossing. As it can be seen in figure 2.a, the images acquired by these cameras suffer from lens distortion. The preliminary operations of the

algorithm pursue the correction of this effect.

3 The algorithm

In this system a stereo matching approach is adopted to segment objects. Since the two stereo images have to be compared, the distortion introduced by lenses should be compensated using a proper model of the lenses. This is actually the first step to be performed.

The main idea underlying the algorithm is the application of Stereo Inverse Perspective Mapping to left and right images and the analysis of the difference be-

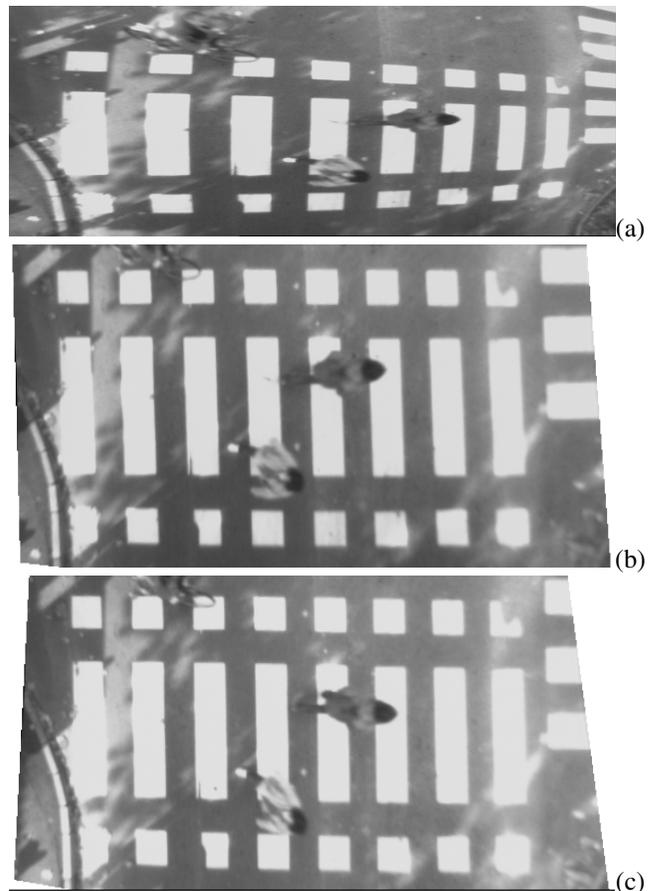


Figure 2: (a) Image acquired by left camera. (b) Left and (c) right images after distortion and perspective removal.

tween the two obtained bird's eye views to extract objects emerging from the road plane.

The stereo difference can be affected by noise due to small errors in the calibration of the stereo system and imperfect removal of lens distortion. To cope with such noise a reference image of the background is subtracted from the stereo difference.

The image of an object in the stereo difference is fragmented and incomplete because it derives from the diversity between the object and the ground, the latter being made of grey asphalt and white stripes. In fact, objects appear as groups of bright areas. A search for bright blobs is performed and each one is enclosed into a bounding box. Adjacent and overlapping bounding boxes likely to represent the same object are fused into a single bounding box. Bounding boxes are then expanded to better adhere to the object shape by using specific points on the ground as sensors of presence.

Vehicles and other objects moving along the lanes are differentiated from pedestrians thanks to considerations on their direction of movement in subsequent frames.

A complete tracking of objects is not performed, however temporal correlation is taken into account to get rid of non persistent false positives and to compensate for occasional false negatives.

The steps of the algorithm are described in detail in the following sections.

3.1 Removal of distortion and perspective

Distortion and perspective are removed from the acquired images through a quick pixel mapping based on a look-up table. The mapping to be applied depends on the specific lenses and calibration parameters. An automatic procedure to characterize the lenses and calibrate the cameras has been developed [3]. It is based on the observation of how the zebra signs appear in the images, assuming they are parallel and of the same size in the world.

This calibration procedure is of basic importance to make the comparison between the two stereo images feasible. In fact, this setup makes the precise alignment and orientation of the camera pair difficult. In some cases it may even be impossible to place the stereo rig right in the center of the road. Moreover, it allows the system to work with different cameras and lenses, whose choice depends on the crossing characteristics.

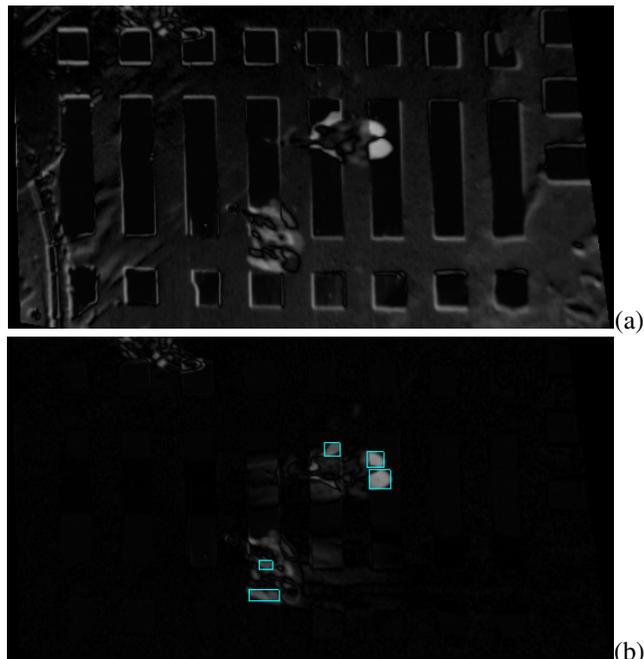


Figure 3: Stereo difference image (a) before and (b) after background removal. Bright blobs are surrounded by cyan bounding boxes.

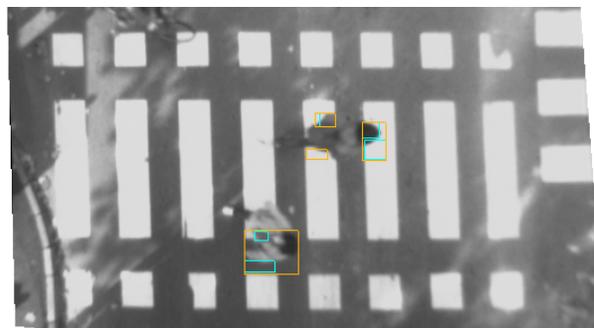


Figure 4: Adjacent or overlapped cyan boxes are fused together: the result is the orange box. This box can be larger than the smallest box enclosing the cyan boxes (as happens for the bottom pedestrian) because the boxes deriving from the previous frame (not shown in this image) are also involved in the fusion process.

The automatic recalibration does not require human intervention and can be performed on a regular basis to

compensate the drifts in the camera orientation and position that can occur during the system operational time due to vibrations, temperature variations, and atmospheric phenomena.

3.2 Stereo difference and background subtraction

The two bird's eye views obtained by Stereo Inverse Perspective Mapping (see figure 2.b and 2.c) significantly differ only in correspondence to objects rising up from the road plane, therefore their difference is computed and analyzed. This operation erases objects belonging to the ground, such as shadows, leaves, road signs. Actually, small drifts in the calibration of the stereo pair and unperfect removal of lens distortion make the two remapped images differ slightly so that their difference also contains elements belonging to the background, as shown in figure 3.a. To cope with such noise, a reference image of the background is subtracted from the stereo difference. This reference image is computed as the weighted average of a certain number of the previous stereo difference images. The most recent values that contribute to the average are given a small weight compared to the oldest values. In such way the reference image mostly contains static elements, while occasional elements have little influence.

3.3 Object extraction

The stereo difference image after background removal is shown in figure 3.b. It features few bright areas on a black background. These bright blobs are identified by an iterative procedure [1] that alternatively computes histograms on image rows and columns. These histograms evidence interesting horizontal and vertical stripes of the image. Repeatedly intersecting those stripes and recomputing the histograms in the resulting areas leads to find bounding boxes enclosing bright regions (see figure 3.b). Each box is considered only if sufficiently large.

Since the bright blobs of an object can be quite small and fragmented, bounding boxes close to each other or overlapped are iteratively fused together into a single box enclosing them all. Moreover, to cope with the high variability from frame to frame of the blobs corresponding to an object, the history of the bounding boxes is taken into account. More precisely, the bounding boxes labelled as

pedestrians in the previous frame are added to the group of boxes to be fused in the current frame. In this manner more stable results can be obtained (see figure 4).

The next phase of the algorithm is aimed at a further and more precise adjustment of the bounding boxes' borders. More specifically, bounding boxes are expanded to better adhere to the object shape by using specific points on the ground as *sensors of presence*. The sensors are regularly spaced points lying on the edges of the zebra stripes (see red and green points in figure 5). These points are featured by a neighborhood made of two uniform areas, one grey (the asphalt) and the other one white (the stripe). An object (pedestrian or vehicle) that passes over the sensor (or better hides it) is very likely to change this particular pattern. This modification is interpreted as an off to on status change of the sensor. Bounding boxes are expanded on each side until the closest external sensors whose status got changed from off to on.

After this adaptation the iterative procedure that fuses close and overlapped boxes is repeated (see figure 5).

3.4 Pedestrians and vehicles classification

Vehicles cannot be differentiated from pedestrians based on the size of bounding boxes because when only a portion of the vehicle is present in the image the size of the bounding box can be similar to that of pedestrians. Moreover, adjacent pedestrians could generate a single large box and then be misclassified.

The discrimination between pedestrians and vehicles is performed considering the direction of movement. It is assumed that the objects in the observed scene are pedestrians if their movement in the image is predominantly horizontal, otherwise they are classified as vehicles.

The positions of the bounding boxes (i. e. their centers of mass) in a certain number of previous frames are stored in a buffer. When the direction of a box of the current frame has to be evaluated, two triangular areas are examined (see figure 6). The idea is to check if the box has left a predominantly horizontal track on its left or right. If a good quality set of footprints is found in the left (right) triangle, the box is classified as a right (left) bound pedestrian. Otherwise, it is considered as a non-pedestrian (i. e. a vehicle). To evaluate the quality of a track in the left triangle, footprints left by right bound pedestrians and non-pedestrians are considered, the former ones with a

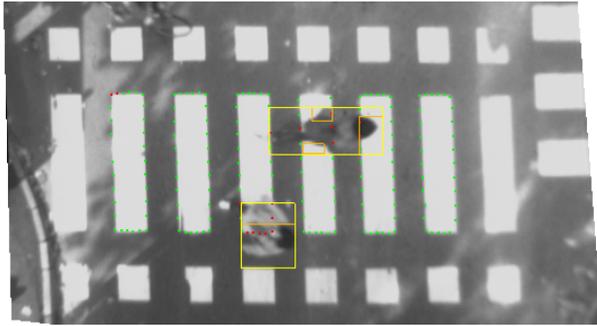


Figure 5: Green points represent off sensors (no objects on it), while red points represent on sensors (hidden by an object). Orange bounding boxes are expanded until the closest external red sensors and then fused if close or overlapped. Resulting boxes are displayed in yellow.

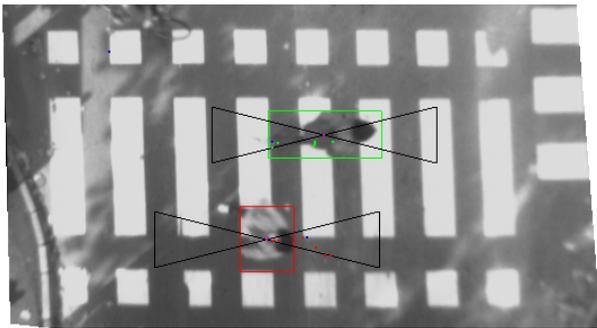


Figure 6: Pedestrians are indicated by a green box when right bound, by a red one when left bound. Their centers of mass is shown with a pink point, while the two triangular areas are shown in black. Footprints are displayed as green points when left by a right bound pedestrian, red points when left by a left bound pedestrian.

larger contribution. Analogous considerations are applied for the right triangle. Obviously, this step introduces a small delay in the recognition of pedestrians when they first appear in the image. Since the system runs at 23 Hz, this delay is neglectable.

3.5 Use of temporal correlation

Being the boxes very variable from frame to frame, tracking objects is not trivial. In fact, due to perspective, the

size of the pedestrian picture varies in the different regions of the image. Moreover, its shape changes while walking, particularly when it is observed from the top.

Temporal correlation is taken into account to stabilize the results. For each previously detected pedestrian, it

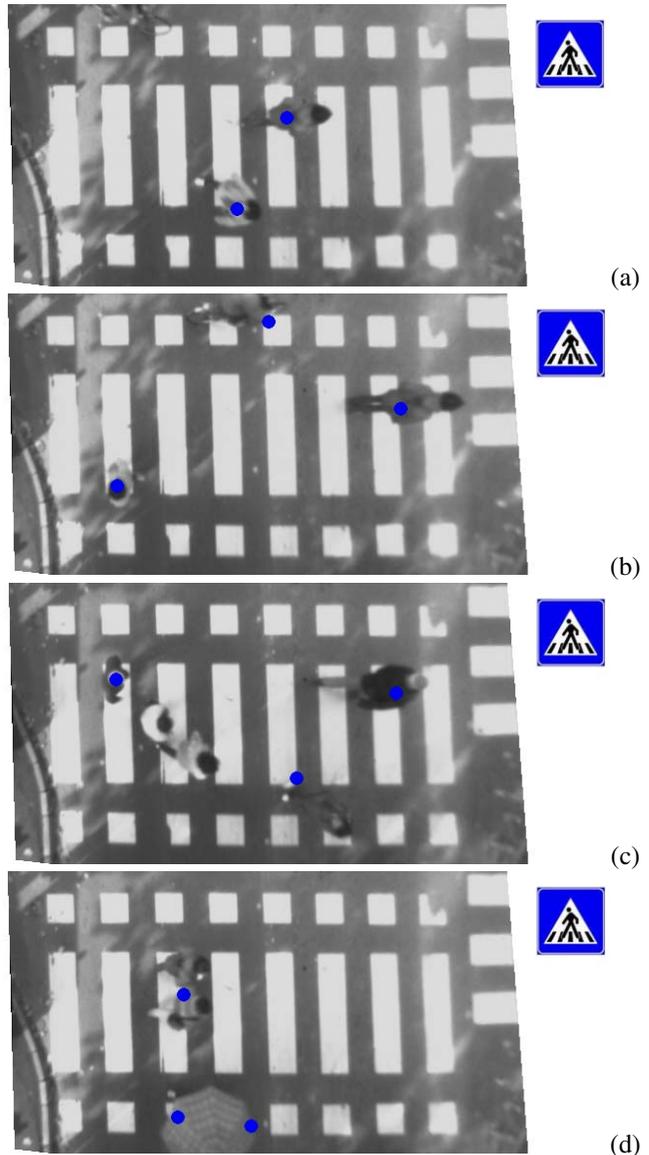


Figure 7: Results in different situations.

is ensured that a new pedestrian with the same direction is present in the current image in a position compatible with its movement. Otherwise, a false negative occurred that has to be compensated. A fake pedestrian with the same direction is added in a position estimated from its recent trajectory. In case the misdetection repeats in subsequent frames, this correction can be performed for a certain number of times until the pedestrian is hooked again or irreparably lost.

In addition, to get rid of occasional false positives a certain amount of persistence in consecutive frames is required before a newly appeared pedestrian is included in the final result. The small latency introduced in the detection is neglectable.

4 Experimental results

To test the efficacy of the system several image sequences have been acquired in daylight conditions, at different times of the day, weather and traffic situations. The system has proven to be robust: on a 7000 image sequence it has not given any error on the presence of pedestrians on the crossing. The entire computation runs at 23 Hz on a Pentium IV @ 2.8 GHz processor system.

Figure 7 reports some results in different situations. Detected pedestrians are highlighted by a blue bullet. When the system finds the zebra crossing occupied by at least one person the signal indicating a pedestrian crossing is also shown in the image. Figure 7.a shows a simple situation with two pedestrians walking on the stripes. Figure 7.b reports the correct detection of a cyclist passing on the stripes, even if very lateral. In figure 7.c the presence of many people, even very close to each other, makes detection of every single pedestrian very difficult. Nonetheless, the algorithm is able to detect the presence of some of the pedestrians on the stripes, which is enough to indicate the crossing as occupied by pedestrians, and thus not clear for vehicles. Figure 7.d shows another complex situation: a pedestrian is carrying an umbrella and is crossing out of the stripes. In this case, due to the lack of the sensor adaptation phase, the bounding boxes belonging to the umbrella are not grouped together and a multiple detection occurs.

5 Conclusions

The proposed system can be used to improve traffic signal control: it can enhance safety for vulnerable road users by increasing pedestrian compliance at signal controlled crossings, while maintaining high levels of traffic flow. Future work will be focused on making the system more robust with respect to the distinction of close by pedestrians. The system will be extended to estimate the average crossing time for pedestrians and count the number of pedestrians walking over the crossing during certain periods of the day for statistical purposes. Future extensions also include the experimentation of the system at night.

References

- [1] M. Bertozzi, A. Broggi, M. Del Rose, and A. Lasagni. Infrared Stereo Vision-based Human Shape Detection. In *Procs. IEEE Intelligent Vehicles Symposium 2005*, pages 23–28, Las Vegas, USA, June 2005.
- [2] V. Bhuvaneshwar and P. B. Mirchandani. Real-time detection of crossing pedestrians for traffic-adaptive signal control. In *Procs. 7th IEEE Intl. Conf. on Intelligent Transportation Systems*, pages 309–313, Oct. 2004.
- [3] A. Broggi, A. Fascioli, R. I. Fedriga, and S. Ghidoni. Automatic Calibration of a Urban Video Surveillance System through the Observation of Zebra Crossings. In *Procs. 5th Intl. Conf. on Computer Vision Systems*, Bielefeld, Germany, Mar. 2007.
- [4] C. Conde, Á. Serrano, L. J. Rodríguez-Aragón, J. Pérez, and E. Cabello. An Experimental Approach to a Real-Time Controlled Traffic Light Multi-Agent Application. In *Procs. 3rd Intl. Conf. on Autonomous Agents and Multi Agent Systems*, July 2004.
- [5] F. N. McLeod, N. B. Hounsell, and B. Rajbhandari. Improving Traffic Signal Control for Pedestrians. In *Procs. 12th IEEE Intl. Conf. on Road Transport Information and Control*, pages 268–277, Apr. 2004.
- [6] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(7):918–923, July 2003.
- [7] P. Vannoorenberghe, C. Motamed, J. M. Blosseville, and J. G. Postaire. Monitoring pedestrians in a uncontrolled urban environment by matching low-level features. In *Procs. IEEE Intl. Conf. on Systems, Man, and Cybernetics*, volume 3, pages 2259–2264, Oct. 1996.
- [8] H. Vceraraghavan, O. Masoud, and N. Papanikolopoulos. Vision-based monitoring of intersections. In *Procs. 5th IEEE Intl. Conf. on Intelligent Transportation Systems*, pages 7–12, 2002.