RESEARCH   **Open Access**

Low-power depth-based descending stair detection for smart assistive devices

Séverine Cloix[1,2*], Guido Bologna[2], Viviana Weiss[2], Thierry Pun[2] and David Hasler[1]

Abstract

Assistive technologies aim at improving personal mobility of individuals with disabilities, increasing their independence and their access to social life. They include mechanical mobility aids that are increasingly employed amongst the older people who rely on them. However, these devices might fail to prevent falls due to the under-estimation of approaching hazards. Stairs and curbs are among these potential dangers present in urban environments and living accommodations, which increase the risk of an accident. We present and evaluate a low-complexity algorithm to detect descending stairs and curbs of any shape, specifically designed for low-power real-time embedded platforms. Based on a passive stereo camera, as opposed to a 3D active sensor, we assessed the detection accuracy, processing time and power consumption. Our goal being to decide on three possible situations (safe, dangerous and potentially unsafe), we achieve to distinguish more than 94 % dangers from safe scenes within a 91 % overall recognition rate at very low resolution. This is accomplished in real-time with robustness to indoor/outdoor lighting conditions. We show that our method can run for a day on a smartphone battery.

Keywords: Stairs detection, Stereo vision, Elderly care, Rehabilitation, Visual impairment, Low-power cameras, Smart walkers

1 Introduction

In industrialized countries, the number of mobility impaired people increases especially among the aged individuals. To deal with the growth of the population of the over 65s, governments are asked to develop policies towards a range of accommodation in relation to the amount of support the seniors require. Such policies would indeed help postponing their move to a long-term nursing care facility. It includes support to individuals remaining at home, which starts with the access to assistive technologies such as the rollator, a walker equipped with wheels, widely spread among the elderly. These tools can, however, lead to falls especially in urban zones and buildings. They occur when the user misjudges the nature or the extent of some obstacles in any kind of familiar or unknown environments.

To answer these issues various prototypes of smart assistive devices are developed. These "intelligent walkers" shall meet a high level of requirements: (i) extended battery-life, (ii) ease of use, (iii) ability to operate in various lighting conditions and scenes, and (iv) affordability. Today, smart walkers are often motorized [1] and programmed to plan routes and to detect obstacles with several active and passive state-of-the-art sensors. Such aids are, however, complex and thus expensive even if produced in large quantities. As a result, most users may be reluctant to use them. In practice, their use is limited to indoor situations due to their heavy weight and their short battery life.

Unlike the current trend, our objective is to develop a low-cost, ultra-light computer vision-based device for rollator users. It is meant to be an independent accessory that can be easily fixed on any standard wheeled walker and with a daylong autonomy. Our device will warn users of potentially hazardous situations [2] and help locate particular items [3]. It has to operate in miscellaneous environments and under widely varying illumination conditions (indoors and outdoors). The users initially targeted are seniors that still live independently.

According to elderly care experts we interviewed, descending curbs and stairs are part of the most common

*Correspondence: severine.cloix@unige.ch
[1]CSEM SA, Jaquet Droz 1, Case Postale CH-2002 Neuchâtel, Switzerland
[2]Department of Computer Science, University of Geneva, 7 route de Drize, CH-1227 Carouge, Switzerland

Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 2 of 15

hazards. Thus, we aimed at developing a computer vision-based algorithm that predicts the presence of various stairs. Our goal is to find a low complexity algorithm that works with the lowest acceptable image resolution, the latter impacting both the power consumption of the sensor and of the processing. It is dedicated to a system that shall have a long battery-life for both outdoor and indoor usage. Figure 1 depicts examples of the usage of our algorithm.

In our previous work [4], we presented promising preliminary results of the evaluation of our stairs detection approach that employs depth information obtained from a stereo vision algorithm based on SAD (sum of absolute difference) methods that are known to be adapted to real-time. In this article, as an extension of our previous work [4], we focus on three types of experiments. Firstly, we assess our previous outcomes by cross-validation experiments including frontal and non-frontal stairs/curbs. Secondly, we experiment with a RGBD camera to determine whether this type of sensors can compete with stereo cameras in the context of our approach, i.e., in both indoor and outdoor environment. Finally, we benchmark our detector on embedded platforms to measure the execution time and the power consumption. These off-the-shelf platforms allow us to get a low-cost prototype in a short period of time.

This paper is organized as follows. Section 2 describes the state-of-the-art in computer vision for stairs detection. The main stereo vision approaches, which allow 3D information extraction, are recalled in Section 3 including our choices to evaluate our approach. Section 4 explains

Fig. 1 Desired embedded system with expected output alarm according to the input frame. The *first* row shows the case of a safe situation for which no alarm is raised. The *middle* row depicts a situation requiring a warning alarm while the *bottom* row represents a dangerous scene raising the appropriate alarm

how we detect descending stairs from 3D sensors. The experimental results, where we look for the lowest acceptable resolution and compare with the RGBD cameras, are detailed and discussed in Section 5 before concluding in Section 6.

2 Related work

From studies on visual accessibility to space for low-vision individuals [5] and the accidents encountered by mobility impaired people [6], constructors are urged to improve stairs and curbs accessibility according to laws and construction guidelines. But, there is still work to be done to facilitate the mobility of individuals with disabilities. In order to fill the gap, computer vision-based electronic technologies can be of great help.

The modelling of staircases has been a research topic in order to predict their presence in a broad range of applied research fields. Staircases concern robotics for the development of unmanned ground vehicles (UGV) and of (humanoid) bipedal robots. The domain of gerontology and low-sighted is also a motivation for developing devices, such as electronic travel aids (ETA), capable of detecting and warning dangerous situations where current aids might fail.

Among the literature, regardless the field of application, stairs detection research can be categorized into two main groups according to the data collected: (i) 2D or (ii) semi-dense to dense range data. The first category gathers works employing monocular cameras. The input data becomes a projection of the captured 3D scene. The captured images are pre-processed to benefit from the man-made structured attributes that stairs have and thus extract straight lines as predominant features. Humanoids equipped with a single camera and aiming at climbing stairs belong to this group [7] as well as tracked robots [8, 9]. Single view-based stairs detectors, however, encounter the issue of false positives raised by repetitive patterns such as zebra crossings [10].

In the second category we find the research based on range data, whether semi-dense from stereo cameras [11] or dense from RGBD cameras or lasers. A recent stereo vision based detector of ascending stairs [12] consists in a stereo camera and an inertial measurement unit fixed on a helmet. By taking advantage of the 3D geometric information, the authors modelled the stairs by extracting the surface orientations and the 3D edges. The model was evaluated by the measurement of the geometric dimensions of the steps (height, width and depth) on a single outdoor staircase.

As far as dense range data is concerned, [13] describes straight-line model-based methods, RGBD cameras being exploited only to extract distance information. Since these sensors can also provide with dense 3D point clouds, planar model-based approaches were also developed [14].

Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 3 of 15
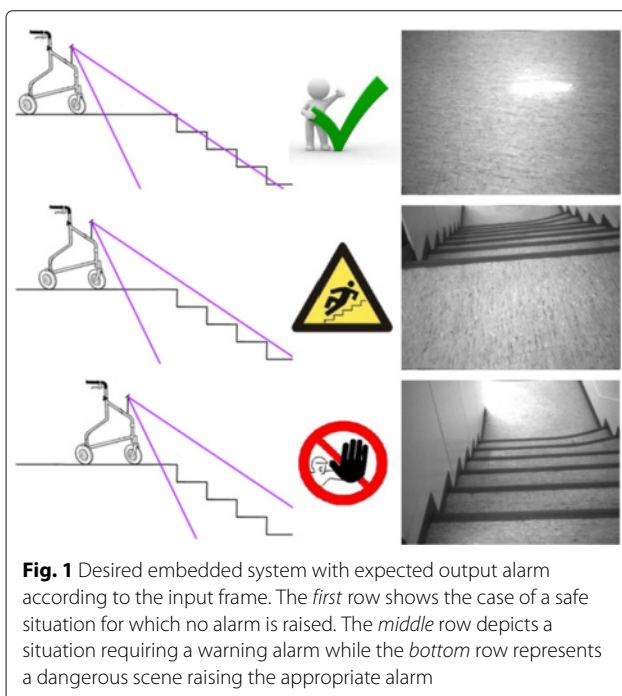
On the other hand, others implemented a piecewise planar algorithm [11] with the Point Grey stereo camera[1]. The authors of [15] presented a descending stair detector by range images captured with time-of-flight cameras. The approach, based on extracting step jumps below the ground level from height profiles, assessed a good recognition performance within 0.8 to 2 m. While the detection of staircases and climbing stairs/sidewalks is subject to research, rare are the studies on detecting descending stairs in the field of electronic travel aids. To our knowledge, authors of [11] are the only ones who proposed tackling the detection of the descending stairs using passive computer vision for mobility aids applications.

All the research works described above are dedicated to be ported on embedded systems both for autonomous robots and mobility aid devices. But, none of them actually gave the performance of their approach in terms of recognition rate and a few estimated the execution time of their algorithm. Among the ones who went through the exercise of measuring the computation time, the best result is reported in [16]: a stereo frame was processed in 30 ms on a MIPS R5000 at 400 MHz. To our knowledge, this processor is specified to dissipate 10 W at 200 MHz [17], which would run for less than 7 h with a professional portable laptop charger[2] in a best case scenario. Similarly, algorithms that run on recent standard laptops with a RGBD camera might not operate for more than half a day, the sensor alone consuming at best 2.5 W and a smartphone battery storing less than 12 Wh.

Developing a device for mobility aids raises several issues: (i) the environment where the device must operate is broad, both in familiar and unknown places; (ii) it requires robustness to any lighting conditions and scenes (indoor and outdoor); (iii) The size and light-weight requirements imply embedded and real-time capabilities; (iv) last and not least, the battery life shall meet an autonomy of a day. Finally, any detector shall avoid false alarms otherwise its users will turn away from the device that raises irrelevant alarms and worse, misses relevant ones. While assessing our approach as a classification problem, our work addresses these four challenges with an algorithm that is generic since not geometric model-based, robust to any illumination, fast and low-power.

3 Background
3.1 Stereo matching
The stereo matching approaches can be categorized into two groups: sparse or dense [18]. The first approach is also known as feature-based matching and results in a sparse output. The correspondence process is applied to features such as corners, edges or key points [19]. In order to compare the different key points, we shall measure their similarity. This similarity can either result from comparing the surroundings via patches or attributes commonly called descriptors [20]. Each descriptor of the left image points is compared to the list of descriptors of the right image points and matched to the most similar one. Feature descriptors tend to be robust against orientation and intensity variation while key points are robust to perspective changes. Thus, this method can be applied to real-time applications that require a very sparse depth map [3], for example in image registration applications.

The second stereo correspondence approach relies on comparing patches of images in order to minimize a cost function. This cost function can be local or global. In the case of local methods, the aim is to minimize the difference between the patches located on the epipolar lines in order to finally get the disparity for every pixel of the reference image. But, stereo matching algorithms can be time, memory and power consuming. Konolige proposed one based on sum of absolute difference (SAD) and implemented it on FPGA to run real-time [21]. From the matched points, we can extract a disparity map. The disparity, d, is the difference between the x-coordinates of the detected point in both pictures, i.e. $x_L - x_R$, x_L and x_R being the x-coordinates of the 3D point projected on the left and right imagers. Provided the correct matching, the depth map is built from the disparity map using image geometry triangulation [22]. Assuming the pin-hole camera model and the cameras having the same focal length f, separated by a baseline T, the distance of a detected point is

$$Z = \frac{f \times T}{d} \ , \tag{1}$$

where Z and T are expressed in meters and f and d in pixels.

3.2 Stereo cameras
Stereo correspondence is a challenging field of research in term of software and hardware implementation [18]. It has to respond to the high demand of real-time execution and frame rates in many domains like machine vision and navigation. Passive stereo vision also suffers from matching failure on low-textured regions and repetitive patterns [23]. Projecting a texture on the scene drastically improves the stereo matching. Projector-based systems became serious competitors to passive stereo cameras. However, the main drawback of such IR-projector-based sensors is their inability to work outdoors and their power consumption. Authors of [24] also showed the degradation of the 3D reconstruction at different times of the day. The stronger the illuminance, the poorer the quality of the resulting 3D map. Thus, passive stereo cameras keep on being employed for outdoor applications related to navigation [25] whereas active ones are leading the

Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 4 of 15

indoor application usage. Examples of commercially available active stereo cameras are the Microsoft Kinect[3] and the Asus Xtion[4].

3.3 Hardware and depth map acquisition requirements

According to [6] and [26], level changes are considered hazardous to mobility impaired people when sidewalks are 4 cm high on flat terrain and more than 3 cm high on a slope. As far as stairs are concerned, the step height is often between 15 and 18 cm. The latter constrains the acquisition system to have a corresponding depth resolution that can be deduced from Eq. (1). In other words, a disparity difference of a pixel must translate a height difference smaller than a step height.

From Table 1, the Bumblebee2 could capture steps as high as curbs with the assumption that the camera is laying parallel to the ground. The surfaces shall, however, be well textured to extract the optimal depth map.

To compute the depth map with a stereo camera, we apply Konolige's algorithm that consists in minimizing the difference between a patch from the left image and a patch from the right image located on the epipolar line (which corresponds to the horizontal axis in the case a horizontal stereo system). The comparison is made by computing the sum of absolute difference (SAD). The resulting values that are associated to a high matching score are kept [27]. We chose this stereo matching for two reasons. Firstly, SAD-based methods are the most appropriate to real-time implementation due to its low complexity (only summations and absolute values calculations) [18]. Secondly, as opposed to the Bumblebee library that is proprietary, we can easily port Konolige's implementation on our embedded platforms.

4 Approach

We are interested in falls related to the loss of balance caused by abrupt changes of the ground elevation in order to predict the dangerousness according to three classes (danger, warning and safe as defined below) when approaching such scenes. The ground topology gives the information of elevation variation. It can be measured from a depth map. From an acquisition of the scene with

tilted sensors towards the ground, the resulting captures will locate far stairs at the top of the depth map and close stairs at the bottom part, each pixel representing the distance to the camera. We want to keep this configuration while rectifying the depth value of each pixel so that a pixel value represents the height between the camera's horizontal plane and the ground (cf. Fig. 2). A dangerous situation is detected when the measured floor elevation close to the rollator is below the ground level. Given the acquisition of a semi-dense 3D map from a system as depicted in Fig. 3(a), each point (x, y) of the ground depth map can be expressed as follows. Let (X, Y, Z) be a 3D point in the world space and R the rotation matrix around the X-axis.

To get the corresponding pixel coordinates of (X, Y, Z) in the raw ground depth map, a point in 3D space is subject to a rotation around the X-axis:

$$R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \qquad (2)$$

with

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\theta & -sin\theta \\ 0 & sin\theta & cos\theta \end{pmatrix}, \qquad (3)$$

followed by a projection according the 3 x 3 camera projection matrix P:

$$P = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad (4)$$

where f is the focal length of the cameras, expressed in pixels. The resulting coordinates in the ground depth image are

Table 1 Bumblebee2 depth resolution in mm as a function of camera resolution and Z in mm, the distance to the camera. The stereo camera has a focal length of 6.3 mm, a pixel size of 7.6e-3 mm, and a baseline of 120 mm

Z	512×384	320×240	160×120	102×76
760	5.94	9.49	18.99	29.79
780	6.25	10.00	20.00	31.38
800	6.58	10.52	21.04	33.01
820	6.91	11.05	22.11	34.68

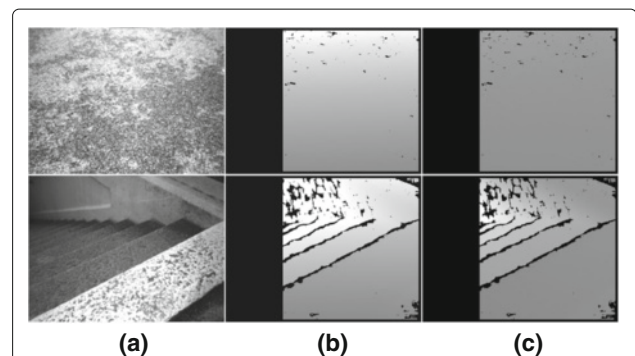
(a)   **(b)**   **(c)**

Fig. 2 Rectification of the depth values: The scenes in (**a**) represent the *left* captures of the stereo camera. The Konolige's algorithm allows the extraction of the raw depth map (**b**) where each pixel value is the raw distance between its corresponding 3D point and the sensor plane. Each pixel value of (**b**) is rectified according to the angle of the vision system on the rollator with respect to the floor. For a flat ground as depicted in the first row, the final depth map (**c**) is uniform
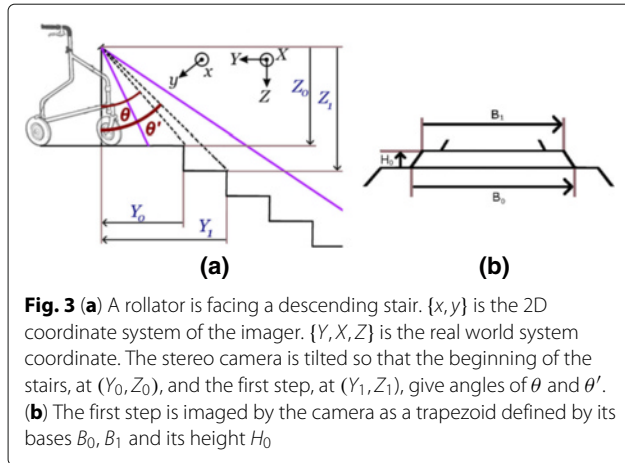
Fig. 3 (**a**) A rollator is facing a descending stair. {x, y} is the 2D coordinate system of the imager. {Y, X, Z} is the real world system coordinate. The stereo camera is tilted so that the beginning of the stairs, at (Y_0, Z_0), and the first step, at (Y_1, Z_1), give angles of θ and θ'. (**b**) The first step is imaged by the camera as a trapezoid defined by its bases B_0, B_1 and its height H_0

$$x = f \frac{X}{Y sin\theta + Z cos\theta}, \qquad (5)$$

$$y = f \frac{Y cos\theta - Z sin\theta}{Y sin\theta + Z cos\theta}. \qquad (6)$$

Note that we used capital letters for world coordinates and lower case for image coordinates. These equations allow defining the limits to detect the first stair step. Let the floor be located on plane $Z = Z_0$ and let the first step start at $(Y, Z) = (Y_0, Z_0)$ and end at $(Y, Z) = (Y_1, Z_1)$ in the original coordinate system. For the first step to be visible early enough, the camera must be tilted. To determine the minimal required angle, we assume that the stairs, of width L, are facing the stereo camera. The first step is then imaged by the camera as a trapezoid defined by its bases B_0, B_1 and its height H_0:

$$B_0 = f \frac{L}{Y_0 sin\theta + Z_0 cos\theta}, \qquad (7)$$

$$B_1 = f \frac{L}{Y_1 sin\theta + Z_1 cos\theta} \qquad (8)$$

$$H_0 = \frac{f (Y_1 Z_0 - Z_1 Y_0)}{(Y_0 sin\theta + Z_0 cos\theta)(Y_1 sin\theta + Z_1 cos\theta)}. \qquad (9)$$

The trapezoid's area of the projected step on the depth map is defined by

$$A = \frac{(B_0 + B_1) H_0}{2}, \qquad (10)$$

where the area A is constrained by its sign, i.e. by

$$A > 0 \iff \left(\frac{Y_1}{Z_1} - \frac{Y_0}{Z_0} \right) > 0, \qquad (11)$$

which corresponds to tilting the camera by an angle allowing the first step to be in the lower part of the image according to

$$\theta < \theta'. \qquad (12)$$

According to the stairs dimension standards [28] and from Eqs. 7 to 12, one can configure the vision system's angle, θ, knowing Y_0 and Z_0. With a vision system tilted with an angle of 52 degrees, the camera captures a flat floor located between 62 and 169 cm in front of the rollator. The dangerous scenes will be located between 62 and 100 cm, while scenes that shall be warned will be located between 100 and 169 cm.

This above theory can be generalize to the case of approaching stairs from the side as long as the camera parameters (focal length and sensor's dimension) permit. Equations 5 and 6 become

$$x = f \frac{X cos\varphi - Y sin\varphi}{(X sin\varphi + Y cos\varphi) sin\theta + Z cos\theta}, \qquad (13)$$

$$y = f \frac{(X sin\varphi + Y cos\varphi) cos\theta - Z sin\theta}{(X sin\varphi + Y cos\varphi) sin\theta + Z cos\theta}, \qquad (14)$$

where φ is the rotation angle around the Z-axis.

The minimal width advised for stairs is 70 cm. With the optic characteristics of the BumbleBee2, if such stairs are located at the warning distance and captured from aside still on the rollator's path, they will still be detected but as a danger. If the stairs are in the danger distance, an angle above to ($\pm$)70° will lead the stairs to be out of the camera's field of view.

The captures result in the image like depicted in the bottom row of Fig. 2.

The area A also defines the minimal proportion of pixels located at a deeper level than the ground, since it represents the projection of the first descending step. This proportion of pixels is then compared to a threshold T_R. Specifically, the stair presence is predicted when the ratio of pixels located under the ground is greater than T_R.

To classify each capture into one of the three classes, the decision making strategy follows the flowchart depicted in Fig. 4. In other words, our three-bin classifier works as follows: the ground depth map is extracted from the stereo pictures and divided into the upper and lower sub-images
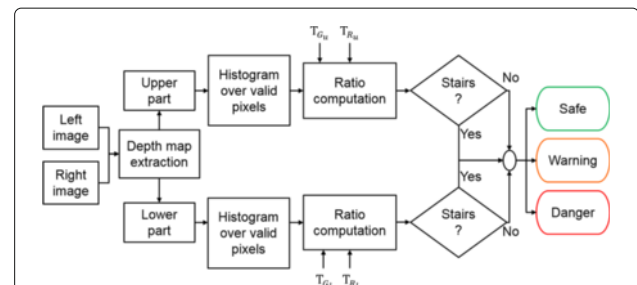

Fig. 4 Flowchart of our approach to detect descending stairs. T_{R_u}, T_{R_l}, T_{G_u} and T_{G_l} are the thresholds on the pixel ratio and on the ground for the *upper* and *lower* sub-images, respectively

of same size. For each sub-image we compute the histogram over valid depth values. The ratio of pixels located below a ground level $T_{G_i}(i = \{u, l\})$ is then compared to a threshold T_{R_i}. If this ratio is greater than T_{R_i} then the sub-image is classified as a stair (positive). The algorithm is detailed in Algorithm 1. The final decision is made from the binary classification of the two sub-images (cf. Algorithm 2): (i) the fronting scene is safe if both sub-images are negative; (ii) it is a warning if the upper sub-image is positive and the lower one is negative; (iii) it is a danger if the lower sub-image is positive, no matter what the prediction is for the upper sub-image.

The evaluation requires a rectified ground depth map as input, the depth being defined as the distance from the camera to the ground. A 3D active camera directly gives the raw depth information on which we compute the ground level at each pixel according to the rotation around the X-axis. On the other hand, a passive stereo camera captures a pair of raw images. In order to proceed to the stereo matching that produces the disparity map followed by the depth map (raw then rectified depth), the raw images have to be undistorted and rectified. This calibration process is of uttermost importance [29]. The Bumblebee2 being already calibrated, we recorded the rectified pairs of images.

Algorithm 1 Binary prediction algorithm. D_p is the upper or lower part of the full depth map extracted from the stereo pictures. Each pixel value, if valid, is the distance between the camera and the corresponding point of the ground. x and y are the coordinates of pixel $D(x, y)$. T_G and T_R are, respectively, the threshold of the ground (in distance unit) and the threshold on the ratio of valid pixels with values greater than T_G

Require: The partial depth map $Dp[x][y]$ (x from 1 to W, y from 1 to H), the ground threshold T_G and the ratio threshold T_R.
Ensure: the binary prediction.
1: **constants:**
2:   $W \leftarrow$ width of partial depth map D_p
3:   $H \leftarrow$ height of partial depth map D_p
4:   $N \leftarrow$ number of bins of histograms
5: **procedure** SUBPREDICT($D_p[H][W], T_G, T_R$)
6:     Compute N-bin histogram H with the valid depth values of D_p
7:     Compute the N-bin cumulative histogram C from H
8:     $R \leftarrow \frac{C[T_G]}{C[N]}$, compute the ratio of valid pixels greater than T_G
9:     **if** $R > T_R$ **then return** true
10:    **else return** false
11: **end procedure**

Algorithm 2 Three-class prediction algorithm. D is the full depth map extracted from the stereo pictures. It is divided into two sub-depth maps. Each pixel value, if valid, is the distance between the camera and the corresponding point of the ground. T_{G_i} and T_{R_i} are, respectively, the threshold of the ground (in distance unit) and the threshold on the ratio of valid pixels with values greater than T_{G_i} for each binary classifier

Require: The full depth map $D[x][y]$ (x from 1 to W_F, y from 1 to H_F).
Ensure: the 3-bin prediction.
1: **constants:**
2:   $W_F \leftarrow$ width of full depth map D
3:   $H_F \leftarrow$ height of full depth map D
4:   $T_{G_u} \leftarrow$ ground threshold for upper sub-image
5:   $T_{R_u} \leftarrow$ ratio threshold for upper sub-image
6:   $T_{G_l} \leftarrow$ ground threshold for lower sub-image
7:   $T_{R_l} \leftarrow$ ratio threshold for lower sub-image
8: **procedure** PREDICT($D[H_F][W_F]$)
9:     $H_U \leftarrow H_F/2$
10:    $H_L \leftarrow H_F - H_U$
11:    $upperPredict \leftarrow subPredict(D[H_U][W_F], T_{G_u}, T_{R_u})$
12:    $lowerPredict \leftarrow subPredict(D[H_L][W_F], T_{G_l}, T_{R_l})$
13:    **if** $lowerPredict = true$ **then return** danger
14:    **else if** $upperPredict = true$ **then return** warning
15:    **else return** safe
16: **end procedure**

Finally, we look for $T_{G_u}, T_{G_l}, T_{R_u}$ and T_{R_l} that minimize false positives and false negatives, i.e., that they maximize the accuracy on a training set of the data collected and labelled "Danger", "Warning", or "Safe". T_{G_u} and T_{G_l} could directly be set to $Y_0 = 78$ cm. We, however, chose to estimate them by training while we expect these thresholds to be similar to Y_0. For each experiment presented below, the optimal values of the four thresholds are thus determined for each binary classifier before the evaluation on a test set. The generalisation performance is evaluated by cross-validation.

5 Experiments and results
The goal of the experiments is fourfold: (i) validate our previous preliminary results about the performance with the resolution with extension to curbs, (ii) measure the processing time and (iii) the power consumption, and (iv) assess the feasibility to port our algorithm on a lightweight embedded platform. This section is structured as follows. We first describe the data collected and the experimental protocol in Section 5.1. A resolution study carried out on the stereo frames is analysed in Section 5.2. A comparison of the performance of the approach according to the sensor employed is detailed in Section 5.3. Before

discussing the overall outcomes, we present a benchmarking of our algorithm on embedded platforms, specifically with regard to the resolution.

5.1 Collected data

The approach, being dedicated to assistive devices, its performance has to be evaluated under any varying illumination and on a wide range of stairs. The evaluation was performed off-line on frames captured with the Asus Xtion and the Bumblebee2 (Fig. 5) according to the requirements detailed in sections above, the cameras being located at 78 cm height with the tilted angle of 35°. The images were captured at 512×384 and 640×480 pixel resolution, respectively, with the stereo camera and the RGBD sensor.

The assessment of our approach was carried out using thirteen scenes of descending stairs and curbs described in Table 2. Among a total of 8939 stereo frames the database includes 56 % of stairs or curbs captures, 52 % of them being non-frontal. Within 6469 RGBD images, there is 75 % of unsafe situations. Figure 6 is a sample of our database. The experiments were run in a customized cross-validation framework inside each group to evaluate

Fig. 5 Our experimental setup mounted with the Bumblebee2 stereo camera

the performance of our method under specific conditions. In a group of k scenes, $k - 1$ scenes were employed for training, each scene being left out once for testing. At the end we calculate the average performance for each group.

The performance on the test sets was assessed from the analysis of the true positive rate TPR (also called recall), the false positive rate FPR, the missed rate FNR (false negative rate), the true negative rate TNR, the accuracy ACC (also called recognition rate) and the precision PPV (also called positive predictive value). The recall is the ratio of true stairs correctly predicted. The false positive rate is the ratio of safe cases predicted as stairs. The missed rate is the ratio of true stairs predicted as safe situations. The accuracy is the ratio of good predictions out of all the samples. The true negative rate is the ratio of safe cases correctly predicted among all predictions of safe cases. Finally, the precision is the ratio of correctly predicted stairs out of stairs prediction.

From the cross-validation within each of the five groups of scenes, the optimal values for the ground thresholds were proved to correspond to the distance between the floor and the first step of the stairs for the lower sub-image classifier. The upper sub-image classifier required an optimised thresholding at deeper distances since the top of the image does not belong to the first step but to deeper ones. The thresholds on pixel proportions were low at high resolutions and increasing with the decreasing resolution. The following performance results were obtained on the test sets according to these criteria.

5.2 Resolution study

Each main stage of the algorithm, i.e., the stereo matching and the histograms computation, is a succession of loops across the pixels, which inherently has an impact on the processing time and the power consumption: the larger the number of pixels, the longer the processing. We thus aim at determining the lowest resolution that still gives good performance. From the collected stereo images, we generated lower resolution images with a pixel area relation-based algorithm to assess the impact of the camera resolution. Eleven resolutions were experimented, from 512×384 pixels to 51×38 pixels. The SAD window size and the disparity range required by the stereo matching algorithm were adapted to the resolution. Figure 7 shows the resulting depth map according to the resolution.

Figures 8, 9, and 10 depict the performance on three of the five groups, respectively, outdoor stairs, indoor stairs and curbs. With a varying resolution, the performance is evaluated from five measures (accuracy, precision, recall, FPR and FNR) (sub-figures (a), (b), and (c)) and the recognition rate among each class (sub-figure (d)). The error bars represent the standard error of the mean.

Table 2 Collected data. The thirteen scenes are described along with their illumination

Description	Illuminance (lux)	G1	G2	G3	G4	G5
Outdoor stairs in a city center	9920	●		●		●
Outdoor stairs in a train station	9870	●		●		●
Outdoor painted stairs	716	●		●		●
Outdoor backyard stairs	657	●		●		●
Outdoor stairs under a canopy cover	1655	●		●		●
Indoor stairs in a house	37.8	●	●			●
Indoor single stair in a house	21.5	●	●			●
Indoor linoleum stairs	62.8	●	●			●
Indoor scene close to a glass door	57.3	●	●			●
Curb in a city centre	2000				●	●
Curb ramp in a city centre	4680				●	●
Outdoor tram platform	8460				●	●
Second Curb scene in a village	6240				●	●

Each group represents (*G1*) all the stairs, (*G2*) all the indoor stairs, (*G3*) all the outdoor stairs, (*G4*) all the curbs and (*G5*) all the scenes including both stairs and curbs. The bullet (●) indicates to which group the scene was included

The accuracy in distinguishing dangerous situations from the others is fairly stable from the medium to high resolution variation within the three groups presented in Figs. 8, 9, and 10. A common trend is noticeable in all the groups of stairs scenes (sub-figures 8, 9 (a) and (b)): the overall performance degrades at very low resolution. The results on the groups of stairs present an improvement of the performance when expected warnings cases are ignored ((sub-figures 8, 9 (b))) unlike in curbs scenes. It highlights the ambiguity in annotating some warning cases. When the stairs start to appear in the upper part of the full frame, the detector predicts this scene as safe. The same behaviour happens when the stairs start to appear

Fig. 6 Samples of the scenes. The images were captured with both the Asus Xtion and the Bumblebee2: outdoor stairs (*first* row), indoor stairs (*middle* row), and curb scenes (*bottom* row)
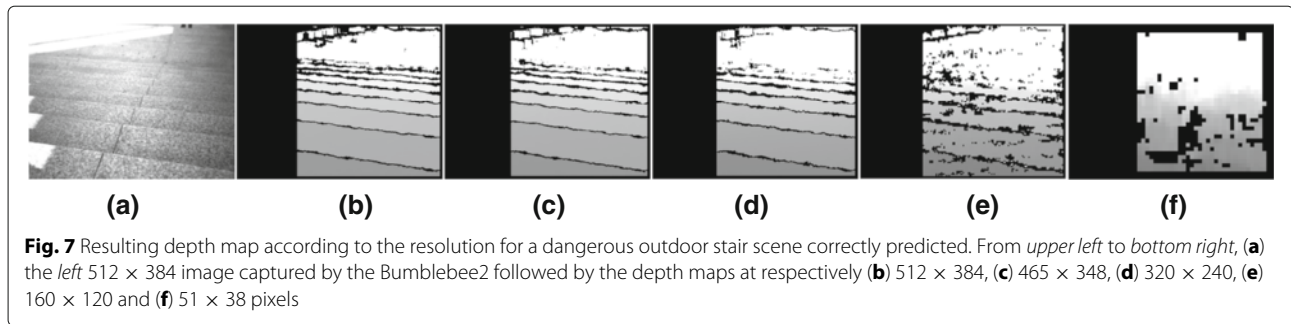
in the upper part of lower sub-image. These cases can either be considered as a danger or still a warning. If we compare indoor stairs scenes to outdoor stairs situations, the impact of the texture (higher in outdoor places) leads to a better quality in the depth map. It results in a larger false positive rate in indoor places.

While looking at the recognition rate for each individual class in Figs. 8, 9, and 10 (c), all danger prediction rates are stable from medium to high resolution, except for outdoor stairs at the highest resolution (Fig. 8 (c)). In outdoor stairs at high resolution, some dangers can still be classified as warnings when the stairs start to appear in the lower sub-image. All warning prediction rates decrease with increasing resolution. The quality of the depth map affects the classification of warnings that tend to be predicted as dangers or safe. However, the safe prediction rate in indoor scenes gets higher with increasing resolution while it tends to decrease in outdoor scenes (Figs. 8, 10 (c)). This behaviour indicates the influence of the illumination and the texture. The outdoor reflection glare has a negative impact on the depth map extraction. The texture of indoor scenes is improved with increasing resolution. In any of the three groups, the classification between critical scenes and harmless places presents a very low false positive rate despite a slight increase at the lowest resolution.

While dangerous scenes were mainly correctly predicted (cf. Fig. 11), a few of them were always misclassified as safe at any resolution. These cases presented similarities, namely the presence of the stairs on the left hand side of the images or a highly sparse depth map due to either a lack of texture or motion blur (Fig. 12).

The experiments validate our preliminary results regarding the ambiguity raised by annotating warning

Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 9 of 15

Fig. 7 Resulting depth map according to the resolution for a dangerous outdoor stair scene correctly predicted. From *upper left* to *bottom right*, (**a**) the *left* 512 × 384 image captured by the Bumblebee2 followed by the depth maps at respectively (**b**) 512 × 384, (**c**) 465 × 348, (**d**) 320 × 240, (**e**) 160 × 120 and (**f**) 51 × 38 pixels

cases: these situations can be tagged as warning or safe by two different experts. As a consequence, they are easily predicted as safe by the detector. These samples present descending stairs that are appearing at the very top of the frame. In practice they can be considered as a safe situation since they are far enough from the user. Nevertheless, in this problem, safe situations are clearly distinct from dangerous ones and most importantly dangerous situations are distinguished from the other cases with an accuracy greater than 85 % at low resolution (102 × 76). For the following sections, the warning cases will not be taken into account unless mentioned otherwise. For the following sections, we will focus on classifying problems of dangerous situations versus the others and dangerous situations versus safe.

Our previous study assessed that the scenes illuminance affects the quality of the resulting depth maps [4]. While RGBD cameras are unable to work under bright sunlight [24], passive stereo cameras' performance drastically drops under low or bright illumination unless the sensors have a high dynamic range. Our approach works on depth maps which are not completely dense. This density depends on illumination and on the camera resolution. Under normal illumination conditions for walker users and a sufficient camera resolution, the depth map density shall prevent the risk of missing relevant alarms.

At last, the stereo matching algorithm reduces the width of the exploitable depth map. Due to the disparity range given to the matching algorithm as a parameter, there is a vertical margin of invalid pixels that contains no data on the left hand side of the disparity map. The annotation did not take into account that the depth information is only available on a cropped part of the images. Thus, scenes with dangerous situations on the left-hand side are predicted as safe.

5.3 RGBD camera versus stereo camera

3D active cameras are expected to excel only in indoor places. To assess this specification, we compared results obtained with similar obstacles both indoor and outdoor. Unlike curbs, stairs belong to the category that is encountered both inside buildings as well as in open-air places. As a consequence, curbs were not included in the comparison.

To oppose the RGBD camera to stereo, we present the best and the worst results obtained with the stereo camera according to the precision. The precision indeed defines the ratio of positives correctly predicted among all the expected positives. As in our study positives are the
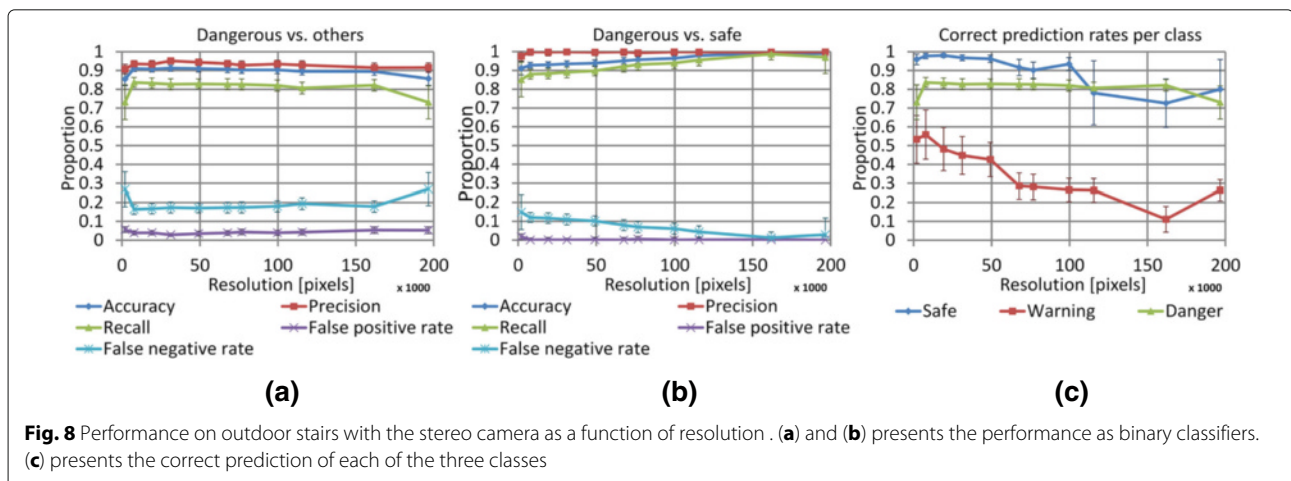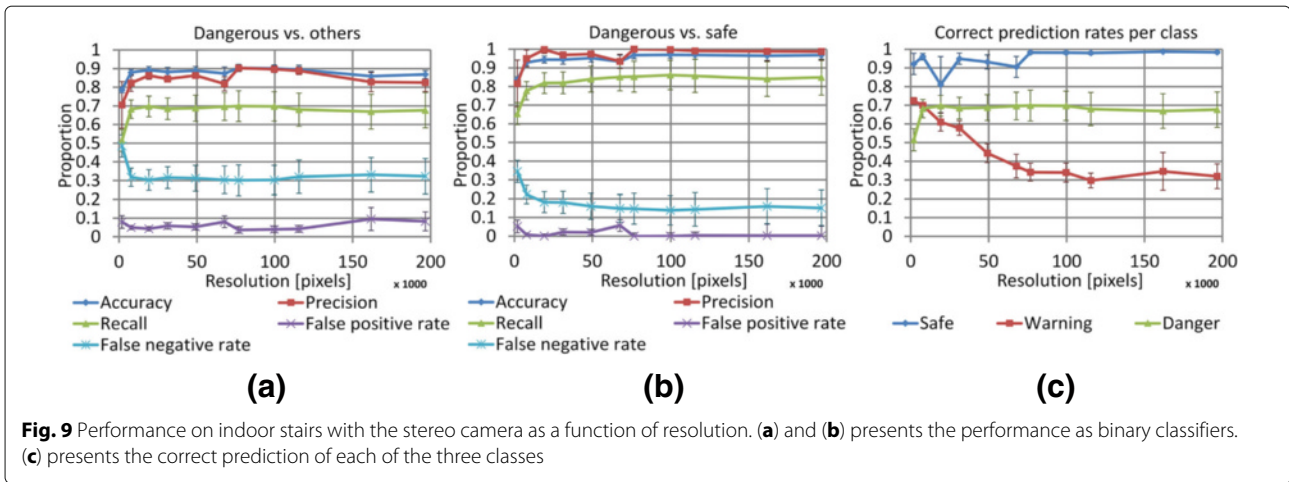

Fig. 8 Performance on outdoor stairs with the stereo camera as a function of resolution . (**a**) and (**b**) presents the performance as binary classifiers. (**c**) presents the correct prediction of each of the three classes

Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 10 of 15

Fig. 9 Performance on indoor stairs with the stereo camera as a function of resolution. (**a**) and (**b**) presents the performance as binary classifiers. (**c**) presents the correct prediction of each of the three classes

dangers, the better they are correctly predicted, the better the detector is in terms of user requirements.

Our approach needs enough pixels with valid depth estimation in order to produce a correct prediction. The 3D active camera, meant to produce dense depth map, reaches its goal in indoor places (Fig. 13). When operating outdoor, the density of the disparity map decreases with the illumination intensity. Only areas in the shadow allow the projector to be visible by the camera for valid depth estimation.

The depth maps of the outdoor stairs are depicted in Fig. 14. The ones of bright places explain why the ground thresholds obtained at cross-validation do not correspond to the expected value (between the ground and the first step). Still, while the pattern of camera's projector can be visible, in cloudy conditions for instance, the resulting sparse depth map is exploitable and allows good predictions.

Our approach relies on the binary classifiers that predict the presence or the absence of stairs in the upper and the lower sub-images of each frame. A frame is a

danger when the stairs are present in the lower sub-image. The evaluation of this binary classifier assesses the binary problem of classifying dangerous situations versus the others. The Figs. 15 and 16 illustrates this study.

In indoor scenes, the Asus Xtion allows our approach to perfectly differentiate secure scenes from dangers. The dropping performance of the RGBD camera in outdoor places is also assessed in Fig. 16 where the recall is capped at 64 %. In outdoor stairs scenes, the RGBD camera allows to distinguish the two classes with a good accuracy Fig. 16 bottom row). However, the rate of missed alarms[5] remains greater than the one obtained with the stereo camera (Fig. 16 first row).

The missed alarms (the detector predicting a danger as safe) come from the depth map having no valid depth values, these scenes being encountered bright illumination with no shadow for the pattern projector to compete with the sun light (scene "Outdoor stairs in city centre"). This specific scene had 78 % of its dangerous cases predicted as safe. The others scenes being captured under cloudy
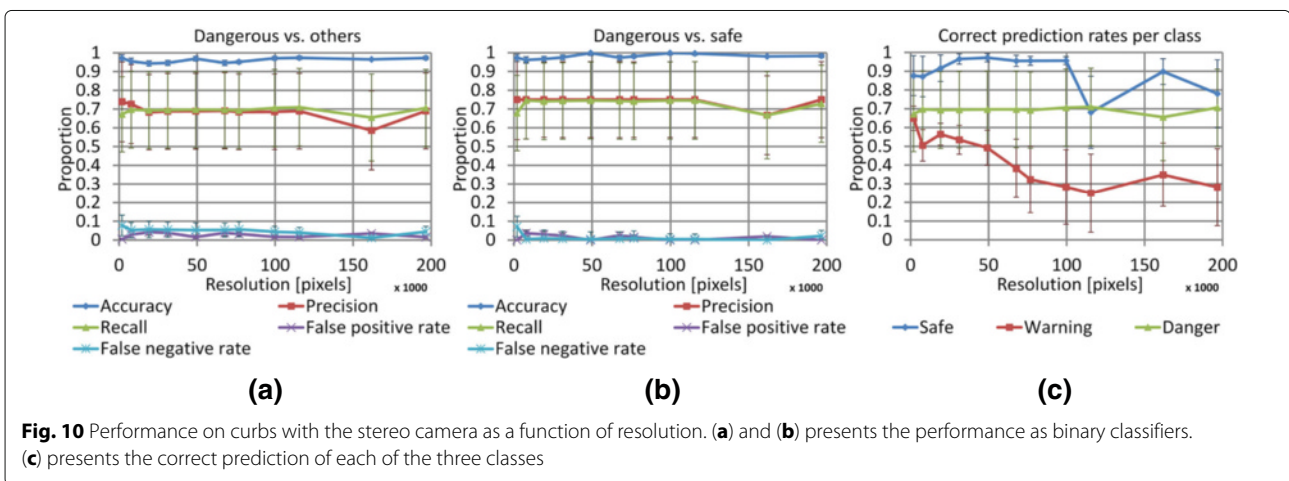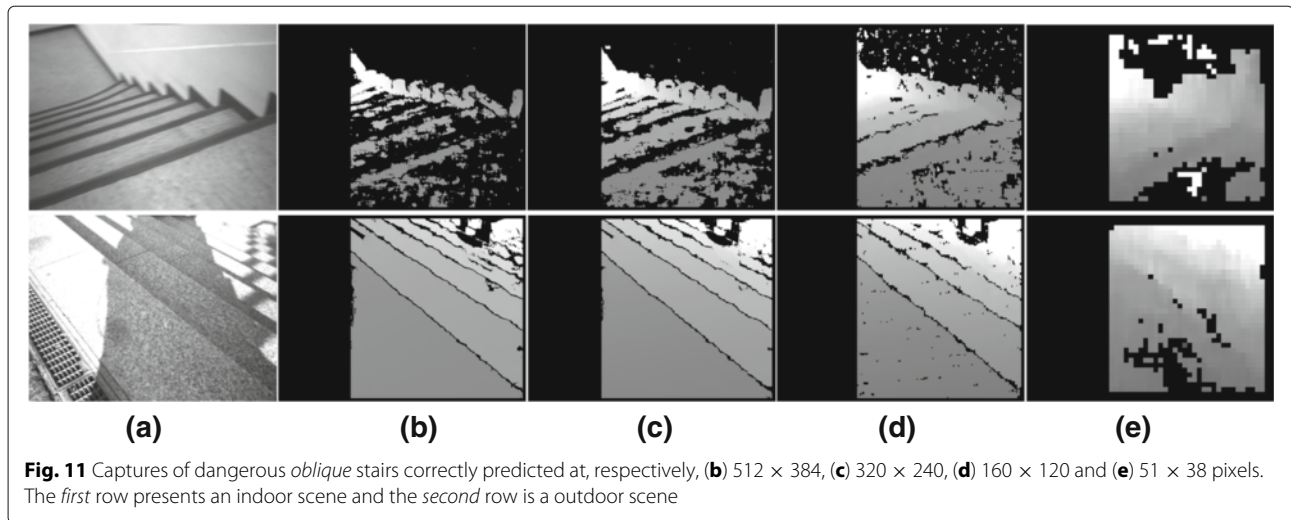

Fig. 10 Performance on curbs with the stereo camera as a function of resolution. (**a**) and (**b**) presents the performance as binary classifiers. (**c**) presents the correct prediction of each of the three classes

Cloix *et al. EURASIP Journal on Image and Video Processing*   (2016) 2016:33

Page 11 of 15

Fig. 11 Captures of dangerous *oblique* stairs correctly predicted at, respectively, (**b**) 512 × 384, (**c**) 320 × 240, (**d**) 160 × 120 and (**e**) 51 × 38 pixels. The *first* row presents an indoor scene and the *second* row is a outdoor scene

conditions in autumn, the projector allowed the sensor to produce useful depth maps.

In both outdoor and indoor scenes the ROC and precision-recall curves are improved when the expected warning cases are not taken into account (Figs. 15, 16 (b)). It underlines the ambiguity coming from the misclassified warning cases.

5.4   Porting on embedded platforms

The detector is aimed at running on a light device clipped on a mobility aid. So, our final objective is to prove the feasibility of integrating our algorithm on an embedded device. We want to assess the execution time and the power consumption for a real-time processing and a battery-life of at least 8 h. For a user walking at 1.4 m/s the appropriate frame rate would be at least of 5 captures per second. It corresponds to a processing time of 200 ms per frame. Concerning the batteries, typical smartphone batteries have a capacity of 3200 mAh and weight 45 g while portable laptop chargers can reach 21,000 mAh but weight 600 g.

We ported our algorithm into a light C-code and ran it on three platforms: (i) a standard Windows laptop (Intel Core Duo at 2.4 GHz), (ii) on embedded Linux board (ARM Cortex-A8 at 800 MHz, the algorithm loaded on a 4 GB microSD flash memory) and (iii) on a customized board equipped with an ARM Cortex-M4 (180 MHz, the algorithm loaded on an external RAM). The results of the benchmarking are gathered in Fig. 17.

We obtained that a stereo frame required 13 s (±0.99) to be processed on our Cortex-M4 platform at full resolution, 4 s (±0.6) at 320 × 240 and 227 ms at 102 × 76. On the Cortex-A8 board, with a customized light version of Linux, we reached 1.88 s at full resolution, 556 ms at 320 × 240 and 32 ms at 102 × 76. Both platforms made the whole process run respectively 100 and 10 times slower than on a standard Windows laptop. The main demanding part is the disparity computation, up to 86 %
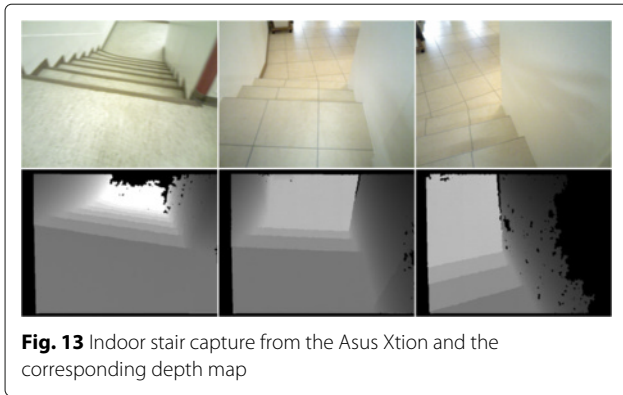

Fig. 12 Missed alarms: Captures of dangerous indoor stairs predicted as safe at, respectively, (**b**) 512 × 384, (**c**) 320 × 240, (**d**) 160 × 120, and (**e**) 51 × 38 pixels

Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 12 of 15
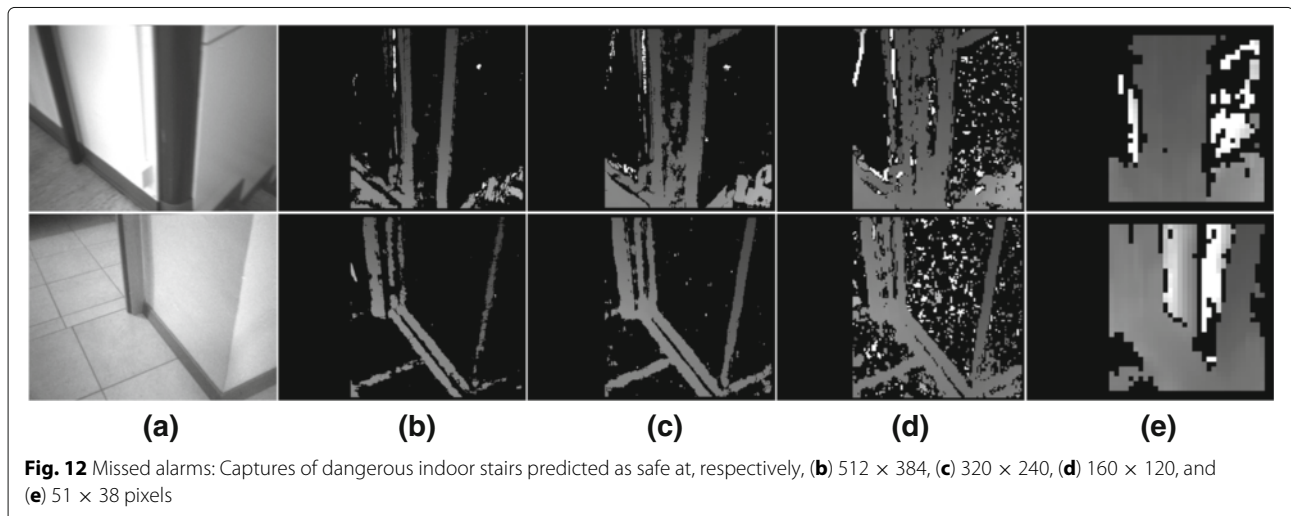
Fig. 13 Indoor stair capture from the Asus Xtion and the corresponding depth map

of the time being dedicated to this task (cf. Fig. 17). In terms of power consumption, the Cortex-M4-based platform needed 300 mW while the Cortex-A8 board drained up to 400 mA at 4.5 V, i.e. 1.8 W.

5.5 Discussion

Our goal was to evaluate the performance of our algorithm from computer vision-based depth maps and to assess its usability on an embedded platform, i.e., its processing time and its power consumption for a battery-life of at least a day. With a low-resolution stereo camera, our detector shall avoid mixing up safe situations with dangers and vice versa, to get the best recognition rate (accuracy) and to minimize the false positives and false negatives.

Despite the ambiguity of the annotation of warning cases, safe situations are clearly distinct from dangerous ones even at low resolution. With no surprise, the performance is affected by the resolution. The higher the resolution, the better the classification performance, but the higher the processing time. With regard to safe and dangerous situations considered as a two-way classification problem, the recognition rate is not less than 91.14 % ($\pm$2.5) in open-air areas and 83.56 % ($\pm$2.27) indoors on 51×38 pixel stereo frames. The performance indoor are worse than outdoor due to the lack of texture and motion blur that affect the stereo matching. Under indoor
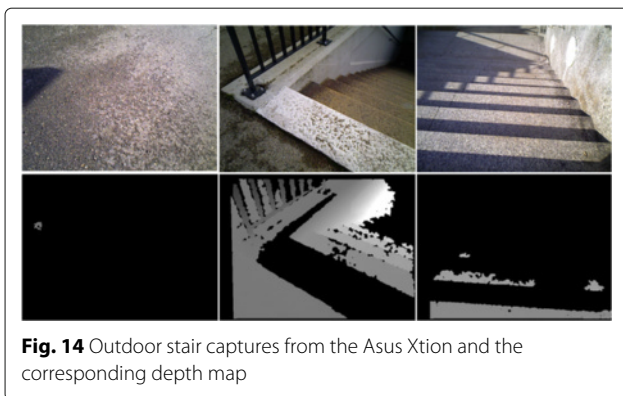

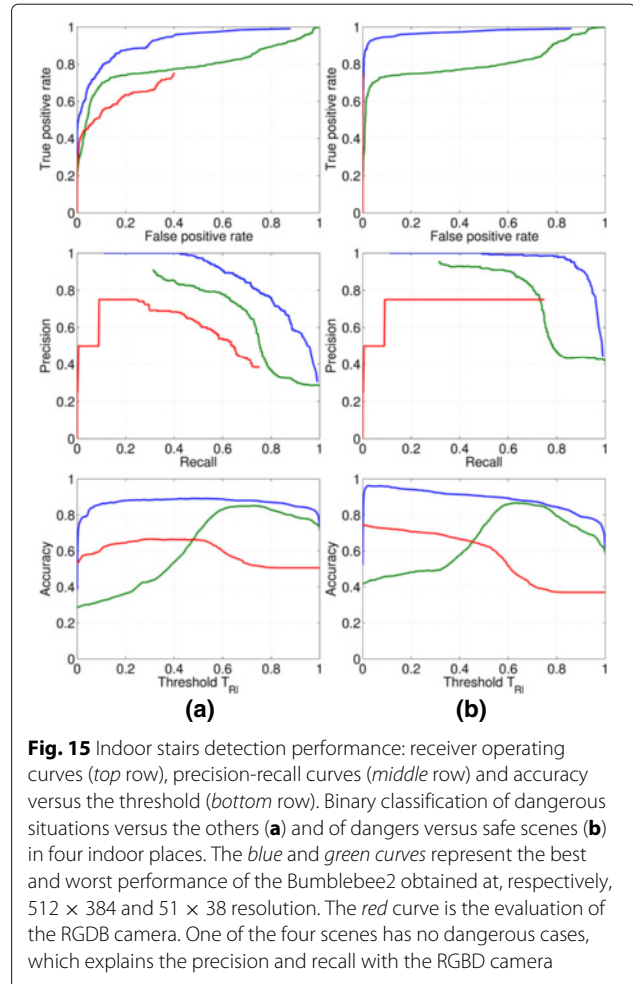**Fig. 14** Outdoor stair captures from the Asus Xtion and the corresponding depth map

Fig. 15 Indoor stairs detection performance: receiver operating curves (*top* row), precision-recall curves (*middle* row) and accuracy versus the threshold (*bottom* row). Binary classification of dangerous situations versus the others (**a**) and of dangers versus safe scenes (**b**) in four indoor places. The *blue* and *green curves* represent the best and worst performance of the Bumblebee2 obtained at, respectively, 512×384 and 51×38 resolution. The *red* curve is the evaluation of the RGDB camera. One of the four scenes has no dangerous cases, which explains the precision and recall with the RGBD camera

and cloudy outdoor lighting conditions, the Bumblebee2 (512×384) and the Asus Xtion (640×480) return depth maps of similar quality in terms of sparsity. A 3D active camera has a non-negligible advantage when the indoor lighting drops because of the infrared illumination it projects.

The detection accuracy relies heavily on depth data. The presence of a small hole will not disrupt the detector as long as it is not deeper than T_Gi and the pixel proportion remains under the thresholds T_Ri. If the hole is larger and deeper than T_Gi then the detector will raise an alarm, which is not inconsistent with fall prevention in terms of project requirements. In the same way, if there is an obstacle on the stairs and it is completely located deeper than T_Gi, the detector will respond as expected. However, any obstacle located above the ground will hide the stairs according to the obstacle dimensions. But, in terms of user interaction, the presence of an obstacle can prevent the risk of falls.

Regarding the execution time, our actual algorithm was embedded as a stripped down pure C99 implementation without any optimisation for the target boards. Using the
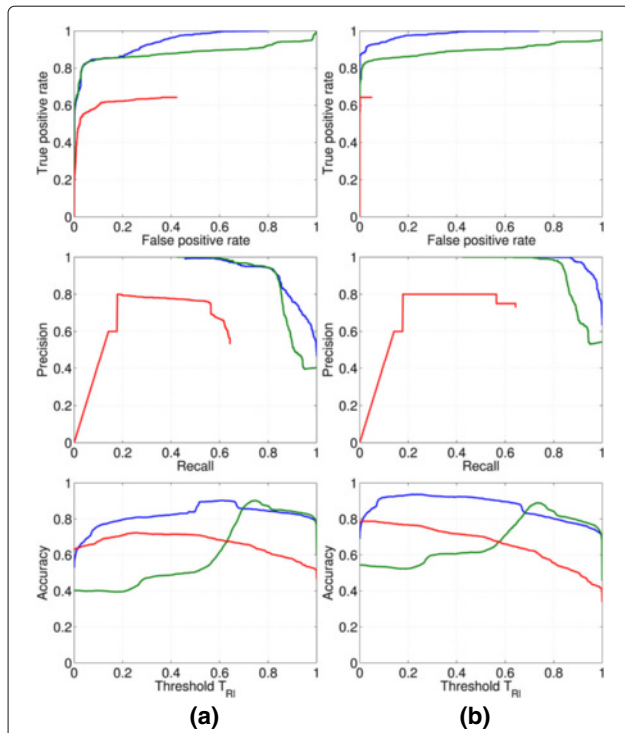
Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 13 of 15

Fig. 16 Outdoor stairs detection performance: receiver operating curves (*top* row), precision-recall curves (*second* row) and accuracy versus the threshold (*bottom* row). Binary classification of dangerous situations versus the others (**a**) and of dangers versus safe scenes (**b**) in five outdoor places. The *blue* and *green* curves represent the best and worst performance of the Bumblebee2 obtained at, respectively, 512 × 384 and 51 × 38 resolution. The *red* curve is the evaluation of the RGDB camera

smallest acceptable resolution (102 × 76) the algorithm runs approximately at the desired speed (5 fps) on the smallest processor (Cortex-M4). Nevertheless, there is a large room for speed improvement in order to use higher resolution and thus achieve more accurate results. The reasons of the slow processing are both related to the hardware and the software: (i) hardware-wise, the algorithm and the input data were stored on an external memory, because of a lack of space in internal processor memory; (ii) software-wise the disparity computation is the bottleneck in the overall processing time. There are several ways to improve runtime. Transferring one byte of data from an external RAM to the processor takes several clock cycles compared to one clock cycle for any data located in the processor internal RAM. Luckily, embedded platforms are equipped with digital signal processors (DSPs), external flash memory, external and internal RAM and direct memory access features (DMA). DSPs are dedicated to process routines that are highly regular—which is the case of disparity computation—while DMA is a process that can transfer data from external to internal RAM. Both DMA and DSP run in parallel to the processor. The computation of disparity is a process that requires only few lines of the image to be present in memory at a given time. Thus, the DMA can be used to transfer these few lines from external to internal memory, while the processor computes the disparity on the few lines transferred previously. By doing so, we get an acceleration factor approximately equal to the number of clock cycles needed to access external memory, which in most cases ranges from 4 to 10. Disparity computation can be transferred on the DSPs to further reduce its processing time by at least a factor of two.
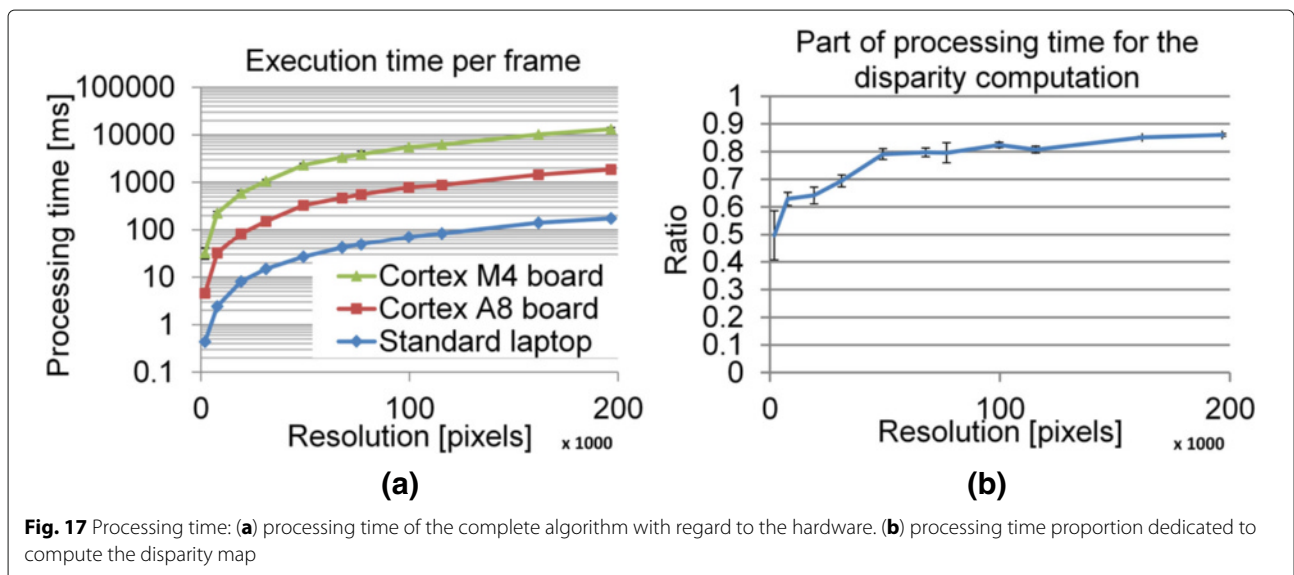

Fig. 17 Processing time: (**a**) processing time of the complete algorithm with regard to the hardware. (**b**) processing time proportion dedicated to compute the disparity map

Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 14 of 15

After all the above optimisation stages, we expect the processing time to go down to less than 200 ms and 28 ms on respectively the Cortex-M4 board and the Gumstix[6] platform for an input resolution of (320 × 240). From our experience on projects for industrial clients, an optimisation using the DMA, some programming techniques and data organization in the different memories, without even using the DSP, resulted in a speed increased by a factor of 40 compared the non-optimised C99 implementation on the same platform with a similarly regular algorithm. In our case, it would lead to processing times of 100 and 14 ms, respectively. In our previous work, we estimated the power consumption according to the specifications given by the manufacturer of the Cortex-M4. The estimation did not take into account the power required by peripheral components of a standalone board. We went through the exercise of measuring the actual power required by our Cortex-M4-based platform to run our algorithm and obtained 589.5 mW. Once the processor is integrated on a functional board among other pieces of hardware, the power consumption is multiplied by 10 compared to processor specifications. We were first surprised, but it turns out that this is a realistic figure. At last, let us estimate the autonomy of our detector by referring to off-the-shelf batteries dedicated to embedded systems. Today smartphone batteries are light (45 g) and can provide with 3200 mAh at 3.7 V, which represents 11840 mWh. With low power cameras, consuming 30 mW each (83 times less than the Microsoft Kinect), our algorithm could run for 24 h on the Cortex-M4-based board compared to 6.3 h on the Cortex-A8-based Gumstix.

Another hardware improvement to have in mind is the choice of the sensor. High dynamic range (HDR) cameras perform in representing a greater range of illumination than standard sensors, resulting to successfully capturing scenes under very bright direct sunlight. These sensors can be employed to overcome the possible saturation from a bright illumination that makes the stereo matching fail. Finally, DSPs can also be engaged to run a motion blur removal algorithm before computing the disparity in order to improve the SAD stereo matching.

6 Conclusions
Through this research work, we proposed a universal descending stairs detection algorithm based on passive stereo-vision that is robust to a wide range of conditions of use, either indoor or outdoor. Its reliability covers any types of stairs since our approach does not rely on the constraint of man-made geometric structures but on the ground elevation. In addition, the same approach is used to detect curbs. One of the requirements was to propose a low-power system. But, low-power processors tend to be slow thus we needed to reduce the processing time of our algorithm. The image resolution being one of the

main parameters affecting the execution time and thus the power consumption, we studied the system performance as a function of the resolution. Our study showed that the detection stayed reliable at very low resolution, distinguishing dangerously approaching stairs or curbs from safe scenes on 102×76 pixel captures with at least 94.94 % accuracy. The latter resolution makes our detector already portable on an off-the-shelf embedded system running at 30 fps for 6.3 h with a 45 g smartphone battery. Once the algorithm optimised for a Cortex-M4-based board, we can expect with confidence that it will run for 24 h on a ultralight device. As future steps, we aim at achieving better performance in terms of recognition rate with HDR stereo sensors.

Endnotes
[1] http://www.ptgrey.com//bumblebee2-firewire-stereo-vision-camera-systems

[2] https://www.powertraveller.com/en/shop/portable-chargers/professional/powergorilla/

[3] http://www.xbox.com/en-US/xbox-one/accessories/kinect-for-xbox-one

[4] http://www.asus.com/3D-Sensor/Xtion_PRO/

[5] The rate of missed alarms is deduced from the recall i.e. FNR = 1 - TPR

[6] https://store.gumstix.com/index.php/overo-waterstorm-com.html.

Authors' contributions
SC participated in the specifications of the project requirements, carried out the descending stairs detector studies and analysis, and drafted the manuscript. GB participated in the specifications of the project requirements and the study analysis and helped to draft the manuscript. VW participated in the specifications of the project requirements, the experimental setup for this study, and helped to draft the manuscript. TP initiated the project and participated in the specifications of the project requirements, the design of the study, and helped to draft the manuscript. DH participated in the specifications of the project requirements, conceived of the study, participated in its design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

References
1. G Lacey, D Rodriguez-Losada, The evolution of guido. IEEE Robot. Autom. Mag. **15**(4), 75–83 (2008)
2. V Weiss, S Cloix, G Bologna, D Hasler, T Pun, in *9th International Conference on Computer Vision Theory and Applications*. A robust, real-time ground

Cloix *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:33

Page 15 of 15

change detector for a "smart" walker, vol. 2 (SCITEPRESS, Lisbon, 2014), pp. 292–298

3. S Cloix, V Weiss, G Bologna, T Pun, D Hasler, in *9th International Conference on Computer Vision Theory and Applications*. Obstacle and planar object detection using sparse 3D information for a smart walker, vol. 2 (SCITEPRESS, Lisbon, 2014), pp. 305–312

4. S Cloix, G Bologna, V Weiss, T Pun, D Hasler, in *Computer Vision - ECCV 2014 Workshops*. Descending stairs detection with low-power sensors (Springer International Publishing Switzerland, Zürich, 2014). Second Workshop on Assistive Computer Vision and Robotics (ACVR) held with ECCV2014

5. GE Legge, D Yu, CS Kallie, TM Bochsler, R Gage, Visual accessibility of ramps and steps. J. Vis. **10**(11) (2010). doi:10.1167/10.11.8. Accessed 20 June 2014

6. E Walter, M Cavegn, G Scaramuzza, S Niemann, R Allenbach, Fussverkehr Unfallgeschehen, Risikofaktoren und Prävention (2007). http://mobilitepourtous.ch/pdf/Bpa_Pietons_accidentes_2007.pdf. Accessed 28 Sept 2015

7. S Oßwald, A Hornung, M Bennewitz, in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Improved proposals for highly accurate localization using range and vision data (IEEE Computer Society, 2012)

8. DC Hernandez, K-H Jo, in *Frontiers of Computer Vision (FCV), 2011 17th Korea-Japan Joint Workshop on*. Stairway tracking based on automatic target selection using directional filters, (IEEE Computer Society, 2011), pp. 1–6. doi:10.1109/FCV.2011.5739752

9. S Wang, H Wang, in *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, ed. by IEEE. 2d staircase detection using real adaboost (IEEE Computer Society, 2009), pp. 1–5. doi:10.1109/ICICS.2009.5397508

10. S Shahrabadi, JM Rodrigues, JH Du Buf, in *Pattern Recognition and Image Analysis*, ed. by Springer. Detection of indoor and outdoor stairs (Springer-Verlag, Berlin, 2013), pp. 847–854

11. V Pradeep, G Medioni, J Weiland, in *Workshop on Computer Vision Applications for the Visually Impaired*. Piecewise planar modeling for step detection using stereo vision (Springer International Publishing, Switzerland, 2008)

12. H Harms, E Rehder, T Schwarze, M Lauer, in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference On*. Detection of ascending stairs using stereo vision (IEEE Computer Society, 2015), pp. 2496–2502. doi:10.1109/IROS.2015.7353716

13. JA Delmerico, D Baran, P David, J Ryde, JJ Corso, in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, ed. by IEEE. Ascending stairway modeling from dense depth imagery for traversability analysis (IEEE Computer Society, 2013), pp. 2283–2290

14. A Pérez-Yus, L-N Gonzalo, GJ J., in *Computer Vision - ECCV 2014 Workshops*. Second Workshop on Assistive Computer Vision and Robotics (ACVR) held with ECCV2014 (Springer International Publishing, Switzerland, 2014)

15. C Stahlschmidt, S von Camen, A Gavriilidis, A Kummert, in *Intelligent Vehicles Symposium (IV), 2015 IEEE*. Descending step classification using time-of-flight sensor data (IEEE Computer Society, 2015), pp. 362–367. doi:10.1109/IVS.2015.7225712

16. J Gutmann, M Fukuchi, M Fujita, in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, September 28 - October 2, 2004*. Stair climbing for humanoid robots using stereo vision (IEEE Computer Society, 2004), pp. 1407–1413. doi:10.1109/IROS.2004.1389593

17. M open RISC Technology, MIPS R5000 Microprocessor technical backgrounder. http://www.futuretech.blinkenlights.nl/depot/R5000_Pr_Ov.pdf. Accessed 28 Sept 2015

18. L Nalpantidis, G Sirakoulis, A Gasteratos, Review of stereo vision algorithms: from software to hardware. Int. J. Optomechatronics. **2**(4), 435–462 (2008)

19. H Bay, A Ess, T Tuytelaars, L Van Gool, Speeded-up robust features (SURF). Comp. Vision Image Underst. **110**(3), 346–359 (2008)

20. M Calonder, V Lepetit, M Ozuysal, T Trzcinski, C Strecha, P Fua, BRIEF: computing a local binary descriptor very fast. Pattern Anal. Mach. Intell. IEEE Trans. **34**(7), 1281–1298 (2012)

21. K Konolige, in *Eighth International Symposium on Robotics Research*. Small vision systems: hardware and implementation (Springer London England, Hayama, 1997), pp. 111–116

22. R Hartley, A Zisserman, *Multiple view geometry in computer vision*, 2nd edn. (Cambridge University Press, Cambridge, 2003)

23. K Konolige, in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, ed. by IEEE. Projected texture stereo (IEEE Computer Society, 2010), pp. 148–155

24. M Gupta, Q Yin, SK Nayar, in *IEEE International Conference on Computer Vision (ICCV)*. Structured light in sunlight (IEEE Computer Society, 2013)

25. M Serrão, S Shahrabadi, M Moreno, José, JI Rodrigues, JMF Rodrigues, JMH Buf, Computer vision and GIS for the navigation of blind persons in buildings. Univ. Access Inf. Soc. **14**(1), 67–80 (2014)

26. PA Williams, in *Basic Geriatric Nursing*, ed. by Sciences, EH. 9 Meeting safety needs of older adults (Elsevier, St. Louis, 2015), pp. 167–179

27. G Bradski, A Kaehler, *Learning OpenCV: computer vision with the OpenCV Library*, 1st edn. (O'Reilly Media, Inc., Sebastopol, 2008)

28. BPA, Garde-corps Base: norme sia 358. Bureau de prévention des accidents. http://www.inoxconcept.ch/images/normes_sia_358.pdf. Accessed 28 Sept 2015

29. Y Furukawa, J Ponce, Accurate camera calibration from multi-view stereo and bundle adjustment. Int. J. Comput. Vis. **84**(3), 257–268 (2009)