Descending Stairs Detection with Low-Power Sensors

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

[1] Centre Suisse d'Electronique et de Microtechnique, Neuchâtel, Switzerland
[2] Computer Science Department, University of Geneva, Switzerland

Abstract. With the increasing proportion of senior citizens, many mobility aid devices were developed such as the rollator. However among walker's users, 87% of their falls is attributed to rollators. The Eye-Walker project aims at developing a small device for rollators to protect elderly people from such dangers. Descending stairs are ones of the potential hazards rollator users have to daily face. We propose a method to detect them in real-time using a passive stereo camera. To meet the requirements of low-power consumption, we examined the performance of our stereo vision based detector with regard to the camera resolution. It succeeds in differentiating dangerously approaching stairs from safe situations at low resolutions. In the future, our detector will be ported on an embedded platform equipped with a pair of low-resolution and high dynamic range stereo camera for both indoor and outdoor usage with a battery-life of several days.

Keywords: Descending stair detection, stereo vision, elderly care, rehabilitation, visual impairment, low-power sensors.

1 Introduction

In industrialized countries the number of mobility impaired people increases especially among the elderly [21, 14]. Studies demonstrate that the population of the over 65s is growing and the governments' will is to improve the elderly's independence and home caring in order to postpone their move to an assisted living facility integration. The rollator, widely spread among elderly, aims at helping its users keep their independence and a safe mobility. However these tools can lead to falls especially in urban zones and buildings. They occur when the user misjudges the nature or the extent of some obstacles, which can happen in any kind of familiar or unknown environments. To answer these issues, various prototypes of "intelligent walkers" are motorized [17] and programmed to plan routes and to detect obstacles with active or passive sensors. However such aids are 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 weight and their short battery life.

Unlike the current trend, the EyeWalker project's objective is to develop a low-cost, ultra-light computer vision-based device for users with mobility problems. It is meant to be an independent accessory that can be easily fixed on a standard rollator and with a daylong autonomy. Our device will warn users of potentially hazardous situations and help locate particular items, such as everyday objects. It has to operate in miscellaneous environments and under widely varying illumination conditions. The users initially targeted by this project are elderly persons that still live independently. According to elderly care experts that we interviewed, descending sidewalks and stairs are among the most common hazards. Thus our system aims at detecting descending stairs. To meet the requirements of both low-power consumption and outdoor and indoor usage, we focused on employing methods based on 3D information obtained from a passive stereo camera.

The power consumption of a full system results from the power consumption of the hardware and the software. Since the image resolution impacts both the power consumption of the sensor and of the processing, our goal is to find a low complexity algorithm which works with the lowest acceptable image resolution. We propose a descending stair detector based on depth information obtained from a stereo vision algorithm adapted to real-time.

This paper is organized as follows: Section 2 describes relevant examples related to the state-of-the-art in stereo computer vision; Section 3 explains how to detect descending stairs from 3D sensors; The main stereo vision approaches, which allow 3D information extraction, are recalled in Section 4; Section 5 describes the hardware choices and setup built for the evaluation of our approach; The experimental results, where we look for the lowest acceptable resolution, are detailed and discussed in Section 6 before concluding on the future work in Section 7.

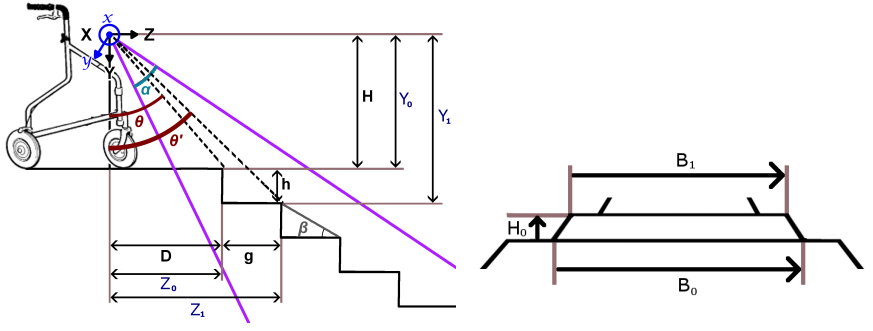
2 Related Works

Sidewalks and stairs are among the obstacles mobility impaired people have to daily face [27]. Even though laws and new constructions are made to improve their accessibility, there is still work to be done. Several institutions propose guidelines both for constructors [3] and users [9]. Works on stair detection have started in the robotic domain for Unmanned Ground Vehicles (UGV) [26]. The UGVs are built to navigate in buildings where security is not assured, for example for searching victims in buildings. While detecting staircases and climbing stairs/sidewalks [7, 19, 2, 8] are subject to research, rare are the studies on detecting descending stairs in the computer vision domain. In the field of electronic travel aids (ETA), PAM-AID is the only prototype demonstrating its ability to detect descending stairs [18]. However this prototype uses an active IR sensor. To our knowledge, authors of [13] are the only ones who proposed to tackle the detection of the descending stairs using passive computer vision. In their method, once the staircase candidates are detected with texture energy measurement, one is randomly selected and used to validate the presence of descending stairs. Other

computer vision based stair detectors use monocular cameras and encounter the issue of false positives raised by repetitive patterns such as zebra crossings [25].

3 Method

We are interested in the falls related to the loss of balance caused by the change of the ground elevation. We aim at measuring this change from a ground depth map. Given the acquisition of a semi-dense 3D map from a system as depicted in 1, each point (x, y) of the ground depth map can be expressed by: let (X, Y, Z) be a 3D point in the world space and R the rotation matrix that positions the depth axis vertically. Thus the new coordinate system has its Z-axis orthogonal to the floor, X-axis becomes normal to the rollator's motion and Y-axis parallel to said motion. The centre of both coordinate systems is the centre of the stereo rig, both cameras having the same focal length f, expressed in pixels. To get the



(a) (b)

Fig. 1. (a) A rollator facing a descending stair. The stereo camera is tilt so that the beginning of the stairs and its first step give angles of θ and θ'. The latter are respectively located at a distance of Z_0 and Z_1 and a height of Y_0 and Y_1 from the camera. (b) The first step imaged by the camera as a trapezoid defined by its bases B_0, B_1 and its height H_0.

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

$$R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{1}$$

with

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\theta & -sin\theta \\ 0 & sin\theta & cos\theta \end{pmatrix} \quad , \tag{2}$$

followed by a projection according the 3x3 camera projection matrix P:

$$P = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{3}$$

The resulting coordinates in the ground depth image are

$$x = f\frac{X}{Ysin\theta + Zcos\theta} \quad , \tag{4}$$

$$y = f\frac{Ycos\theta - Zsin\theta}{Ysin\theta + Zcos\theta} \quad . \tag{5}$$

Note that we used capital letters for world coordinates and lower case for image coordinates. These equations define the limits to detect the stair first. Let the floor be located at depth $Z = Z_0$ 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. We assume the stairs are centred in front of the stereo camera. The step is imaged by the camera as a trapezoid defined by its bases B_0, B_1 and its height H_0:

$$B_0 = \frac{f(Z_1 - Z_0)}{Y_0 sin\theta + Z_0 cos\theta} \quad , \tag{6}$$

$$B_1 = \frac{f(Z_1 - Z_0)}{Y_1 sin\theta + Z_1 cos\theta} \tag{7}$$

$$H_0 = f\frac{Y_0 cos\theta - Z_0 sin\theta}{Y_0 sin\theta + Z_0 cos\theta} - f\frac{Y_1 cos\theta - Z_1 sin\theta}{Y_1 sin\theta + Z_1 cos\theta} \tag{8}$$

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

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

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

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

$$A > 0 \iff (\frac{Y_0}{Z_0} - \frac{Y_1}{Z_1}) > 0 \quad , \tag{11}$$

which corresponds to tilting the camera by an angle big enough to see the first step in the lower part of the image according to

$$\theta < \theta' \quad . \tag{12}$$

This area also defines the proportion of pixels located at a deeper level than the ground if we consider the projection of several steps. This proportion of pixels is then compared to a threshold T_D: The stair presence is predicted when the

ratio of pixels located under a ground is greater than the proportion T_D. To classify each capture into one of the three classes (danger, warning or safe), the decision making strategy follows the flowchart depicted in Fig. 2 and summarized in Table 1. By defining two zones in the picture we impose the condition that the angle θ defined by the focal axis of the system is less than the angle θ' defined by the location of the first step that must be detected as a danger.

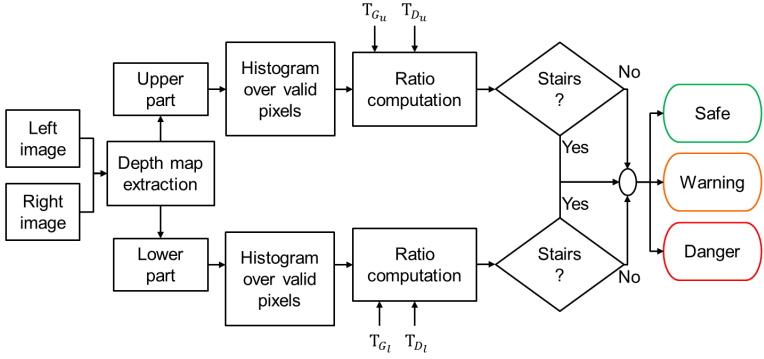


Fig. 2. Flowchart of our approach to detect descending stairs, T_{D_u}, T_{D_l}, T_{G_u} and T_{G_l} being the pixel ratio and ground depth thresholds for the upper and lower sub-images respectively.

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 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_{Gi} (i={u,l}) is then compared to a threshold T_{Di}. If this ratio is greater than T_{Di} then the sub-image is classified as a stair (positive). The final decision is made from the binary classification of the two sub-images: (i) it 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; (ii) it is a danger if the lower sub-image is positive, no matter what the prediction is for the upper sub-image.

4 Stereo Vision

4.1 Stereo Matching

Stereo correspondence is a challenging field of research [20] in term of software and hardware implementation. It has to respond to the high demand of real-time execution and frame rates. From the matched points we can extract a disparity map. The disparity is the difference between the x-coordinates of the detected point in both pictures. Provided the correct matching, the depth map is built

Table 1. Three-bin classification rules according to the dangerousness of the fronting floor, T_{Du} and T_{Dl} being pixel ratio thresholds

Upper sub-image	Lower sub-image	Class prediction	Class definition
$Ratio < T_{Du}$	$Ratio < T_{Dl}$	Safe	No stairs in the whole image
$Ratio > T_{Du}$	$Ratio < T_{Dl}$	Warning	Stairs appear in the upper sub-image only i.e. far from the user
Any	$Ratio > T_{Dl}$	Danger	Stairs appear at least in the lower sub-image i.e. close to the user

from the disparity map using image geometry triangulation [12]. Assuming the pin-hole camera model [12] and the cameras having the same focal length f, separated by a baseline T, the distance of a detected point is

$$Z = \frac{f \times T}{d} \quad , \tag{13}$$

where Z is expressed in meters, f in pixels, T in meters and d the disparity in pixels.

The stereo matching approaches can be categorized into two groups: sparse or dense [20]. 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 [1]. 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 [5]. 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 for real-time applications that require a very sparse depth map [6], for example in image registration applications. Besides it does not require precise calibration.

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 [20]. As far as local methods are concerned, 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. The taxonomy of stereo matching [22] is the reference in the domain of global stereo correspondence. However the algorithms can be time, memory and power consuming. Konolige proposed a real-time stereo matching algorithm based on sum of absolute difference (SAD) and implemented on FPGA [15]. The patch centred on each pixel of the reference image is compared to a patch centred on a pixel in the other image

located on the epipolar line and within a disparity range to reduce the processing time, especially when the desired depth range is known. The size of the SAD window is also a heuristic defined before processing. The algorithm used by the commercial stereo camera Bumblebee2[3] dedicated for real-time applications and used for our experiments is also SAD-based. Our problem can be solved using depth maps that are only partially dense. To compute the depth map, we apply Konolige's algorithm and keep only the values that are associated to a high matching score. Thus some of the locations of the depth map are undefined, which leads to a semi-dense depth map. The remaining valid pixels form a set of reliable depth values.

4.2 Stereo Cameras

As explained in [16], passive stereo vision suffers from matching failure on low-textured regions and repetitive patterns. 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. Authors of [11] 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 [24] whereas active ones are leading the indoor application usage. Commercially available stereo cameras are the Microsoft Kinect[4] and the Asus Xtion[5] for the active ones.

5 Hardware Setup and Depth Map Acquisition requirements

According to [27] and [23] level changes are considered hazardous to mobility impaired people when sidewalks are 4 cm high on flat terrain and more than 3 centimetres high on a slope. As far as stairs are concerned, the step height is often between 15 and 18 centimetres. The latter constrains the acquisition system to have a corresponding depth resolution that can be deduced from (13) as

$$\Delta Z = \Delta d \frac{f Z^2}{f T} \quad . \tag{14}$$

In other words a disparity difference of one pixel (δd) must translate in a height difference smaller than a step height. Equation (14) demonstrates that for a camera located at 80 cm from the ground, the theoretical depth resolution is 6.58 mm, 10.52 mm and 21.05 mm at respectively 512 x 384, 320 x 240 and 160

[3] `http://ww2.ptgrey.com/stereo-vision/bumblebee-2`
[4] `http://www.xbox.com/en-US/kinect`
[5] `http://www.asus.com/Multimedia/Xtion_PRO`

x 120 pixel resolution. The bumblebee2 could capture steps as high as sidewalks with the assumption that the camera is laying parallel to the ground. However the surfaces shall be well textured to extract the optimal depth map.

Fig. 3. Our experimental setup mounted with the Bumblebee2 stereo camera.

To evaluate our approach the stereo camera is attached on a standard three-wheel rollator (Fig. 3). The Bumblebee2, our passive stereo camera, is fixed at H_0=76 cm from the ground. For the sake of comparison, we also use the Microsoft Kinect fixed at 78 cm from the ground. Both systems are tilted with an angle of $\theta = 35$ degrees. Knowing the standard dimensions of a stair step [3], our set-up should detect the first step as a danger at Z_0=56 centimetres. The acquisition is carried out with a laptop computer. The Bumblebee images are captured at the highest resolution (512 x 384 pixels). The impact of the camera resolution is evaluated by resizing the captured frames for different intermediate resolutions from 512 x 384 pixels to 160 x 120 pixels with a pixel area relation based algorithm.

6 Experiments and Results

The evaluation is performed offline on frames captured with the Microsoft Kinect and the Bumblebee2.

6.1 Required Data

The evaluation of our approach requires a ground depth map as input, the depth being defined as the distance from the camera plan to the ground. The active camera directly gives the depth information on which we computed the ground level at each pixel according to the rotation around the X-axis. 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, the raw images have to be undistorted and rectified. This calibration process is of uttermost importance [10]. The Bumblebee2 being already calibrated, we record the rectified pairs of images. The algorithm employed to extract the depth map from the rectified pictures is the one of Konolige [15, 4]. We choose this stereo matching as opposed to the Bumblebee library in order to have a manageable code to

(#1) (#2) (#3) (#4) (#5)

Fig. 4. Scenes where the images were captured with both the Kinect and the Bumblebee2: indoor stair case (#1), entrance indoor stairs (#2), outdoor stairs to enter a building (#3), outdoor stairs (#4), indoor emergency stair case (#5).

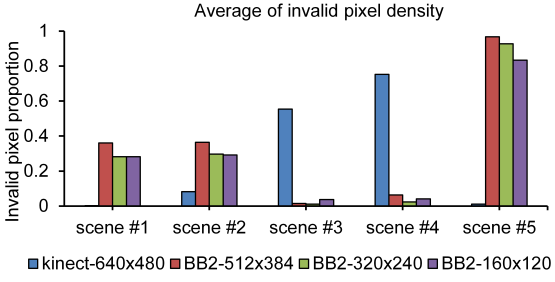


Fig. 5. Average of the proportion of pixels with unknown depth from the Kinect and the Bumblebee2 (BB2) captures at three different resolutions.

further port on an embedded platform. Moreover the detection algorithm does not require accurate depth, nor does it require a fully dense map. The resulting depth map is processed according to the transformation demonstrated in Section 3. Each sub-image of a frame is annotated according to the presence of stairs.

The assessment of our approach was carried out on five scenes of descending stairs (Fig. 4): Scene #1 and scene #5 are indoor scenes; scene #2 is an indoor scene close to a glass door; scene #3 is an outdoor stair under a canopy cover; scene #4 is a scene completely outdoor close to a building that created shadow during the experiment. The scenes' illuminance is respectively 90, 430, 5000, 2200 and 8 lux, which explains the quality of the resulting depth maps. Since the active cameras are unable to work outdoor due to the powerful infrared wavelengths from sunlight, it results in very sparse depth maps (Fig. 5). The passive stereo camera gives denser depth maps, especially outdoors where the scenes are highly textured. However its performance drastically drops in poor lighting conditions, which is rarely a problem since we assume rollator users do not wander in such dark places.

In normal lighting conditions the Bumblebee2 (512 x 384) and the Kinect (640 x 480) have similar performance in term of valid depth values computed. The Kinect has a non-negligible advantage when the indoor lighting drops because of the infrared illumination it projects. The illuminance impacts the quality of the required data. This analysis confirms our choice of evaluating our approach only with a passive stereo camera.

After a training phase done with 70% of data randomly chosen from the first three scenes (cf. 4), the final detectors described in the following sections ran on two test sets: (i) the remaining 30% of scenes #1 to #3, (ii) all captures of scene #4. The results were similar for both sets. We thus present results for the second set as it is completely different from the training set.

6.2 Classification for High Accuracy

As depicted in Fig. 2 our detector needs four parameters, T_{G_u}, T_{G_l}, T_{D_u} and T_{D_l}. In order to determine their optimal values, we proceed to a training phase with 70% of data randomly chosen from the first three scenes. T_{G_i} and T_{D_i} vary from 88 to 130 (cm) and 0 to 1 respectively. The SAD window size and the disparity range were adapted to the resolution, starting from 15 pixels and 144 pixels respectively for the highest resolution. The training set performance is measured 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.

According to the training, the best accuracy scores on each detector are obtained for T_{G_l} between 88 and 92 (cm) and T_{G_u} is located around 104 cm. T_{G_l} optimal value corresponds to the distance between the floor and the first step of the stairs whereas the optimal T_{G_u} comes from the fact that the deepest pixels appearing in the top of the image do not belong to the first step but to deeper ones. The resulting optimal parameters for best accuracy values are listed in Table 4. To produce relevant feedback to the rollator users, we opted for the decision making strategy described in Section 3. The resulting three-bin classification on the second test set with the parameterisation for best accuracy scores is depicted in Fig.6. Safe and dangerous situations are rarely mixed up. At any resolutions, no safe cases are predicted as dangers and up to 0.5% of dangers are classified as safe situations.

Table 2. Performance of the descending stair detector on the lower sub-images of the training set at the best accuracy for various resolutions. These results are used to tune TG_l and TD_l for the final detector run on the test sets

Resolution	T_{D_l}	T_{G_l}	FPR	TPR	PPV	ACC	FNR	TNR
512 x 384	0.015	92	0.001	0.944	0.997	0.980	0.056	0.999
465 x 349	0.02	92	0.000	0.941	0.999	0.979	0.059	1.000
393 x 295	0.015	98	0.002	0.944	0.996	0.979	0.056	0.998
365 x 274	0.02	92	0.001	0.942	0.997	0.979	0.058	0.999
320 x 240	0.02	92	0.003	0.940	0.995	0.977	0.060	0.997
301 x 225	0.03	92	0.000	0.929	1.000	0.975	0.071	1.000
256 x 192	0.03	90	0.002	0.927	0.996	0.973	0.073	0.998
204 x 153	0.045	90	0.007	0.918	0.986	0.967	0.082	0.993
160 x 120	0.095	88	0.009	0.875	0.981	0.950	0.125	0.991

6.3 Classification to Minimize False Alarms and Misses

To help rollator users it is important: (i) to avoid false alarms otherwise they will turn away from the device that raises irrelevant alarms; (ii) not to miss relevant alarms. Thus we looked for T_{G_u}, T_{G_l}, T_{D_u} and T_{D_l} that minimize false positives and false negatives. In other words, we looked for the best true positive rate at full precision. From the training results, we got the desired parameters listed in Table 5. The resulting performance on the test data (scene #4) is summarized in Fig. 7. Again, at any resolutions, all safe situations are never predicted as dangers. Warnings are more misclassified, mainly as safe situations than in the previous experiment. For resolutions higher than 204 x153, less than 0.8% of dangers are missed because predicted as safe and goes up to 2.6% at the lowest resolution.

Table 3. Performance of the descending stair detector on the upper sub-images of the training set at the best accuracy for various resolutions. These results are used to tune TG_u and TD_u for the final detector run on the test sets

Resolution	T_{D_u}	T_{G_u}	FPR	TPR	PPV	ACC	FNR	TNR
512 x 384	0.015	104	0.013	0.919	0.987	0.951	0.081	0.987
465 x 349	0.015	106	0.016	0.919	0.984	0.949	0.081	0.984
393 x 295	0.020	110	0.015	0.915	0.986	0.948	0.085	0.985
365 x 274	0.030	100	0.023	0.910	0.978	0.941	0.090	0.977
320 x 240	0.030	104	0.027	0.910	0.974	0.940	0.090	0.973
301 x 225	0.050	104	0.018	0.898	0.983	0.938	0.102	0.982
256 x 192	0.055	108	0.019	0.894	0.982	0.935	0.106	0.981
204 x 153	0.070	110	0.040	0.895	0.962	0.926	0.105	0.960
160 x 120	0.095	110	0.124	0.898	0.891	0.887	0.102	0.876

6.4 Discussion

Our goal was to assess the performance of a stereo system that fits the requirements of a battery lifetime of at least day. Our objective is to develop a

Table 4. Chosen parameters for the descending stair detector for various resolutions according to the best accuracies obtained from the training phase (70% of data from scenes #1 to #3)

Resolution	T_{D_u}	T_{G_u}	T_{D_l}	T_{G_l}
512 x 384	0.015	104	0.015	88
465 x 349	0.020	104	0.020	88
393 x 295	0.020	104	0.020	88
365 x 274	0.030	104	0.020	88
320 x 240	0.030	104	0.025	88
301 x 225	0.050	104	0.030	88
256 x 192	0.090	104	0.030	88
204 x 153	0.190	104	0.045	88
160 x 120	0.275	104	0.095	88

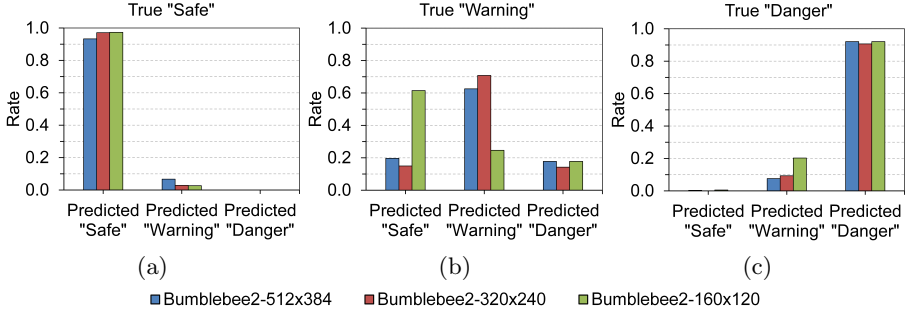


Fig. 6. Prediction on the test data (scene #4) with the parameters set according to the best accuracies obtained on training for three different resolutions.

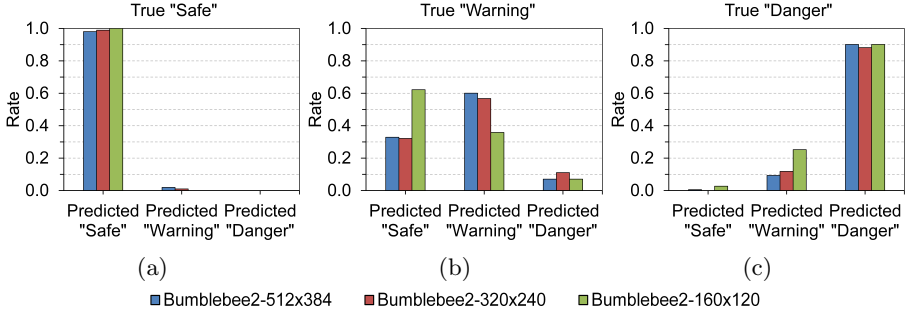


Fig. 7. Prediction on the test data (scene #4) with the parameters set according to the best minimization of false positives and false negatives obtained on training for three different resolutions.

Table 5. Chosen parameters for the descending stair detector for various resolutions according to the best true positive rate at 100% precision. obtained from the training phase (70% of data from scenes #1 to #3)

Resolution	T_{D_u}	T_{G_u}	T_{D_l}	T_{G_l}
512 x 384	0.025	104	0.030	92
465 x 349	0.030	106	0.030	92
393 x 295	0.035	112	0.030	92
365 x 274	0.040	112	0.030	92
320 x 240	0.045	114	0.030	92
301 x 225	0.060	116	0.030	92
256 x 192	0.070	118	0.035	92
204 x 153	0.140	118	0.065	92
160 x 120	0.220	118	0.175	92

descending stair detector with low resolutions cameras. The detector shall avoid mixing up safe situations with dangers and vice versa. From the training phase, we could choose two optimal sets of parameters: (i) to get the best recognition rate (accuracy); (ii) to minimize the false positives and false negatives. Both tests highlighted the ambiguity of the annotation of warning cases which 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.

The two parameterisations gave similar results. As expected the decrease of the resolution alters the performance. However when we look at the detection of safe and dangerous situations as a binary classification, the accuracy keeps being greater than 98.9%. We have to focus on improving the confusion between warnings and safe cases both on the annotation and detection side. In terms of time and power consumption, the Konolige's algorithm is dedicated to real-time applications and ran at 6 fps on FGPA in 1997[15]. We expect our algorithm to run on an ARM cortex-M4 at 10 fps. An Arm cortex-M4 runs at 180MHz and consumes 157 μW/MHz. With two cameras that consume up to 60 mW, we should expect our detector to run for about 160 hours on a mobile phone battery (700mAh at 3.7V).

7 Conclusions and Future Work

We proposed a reliable descending stair detector based on stereo vision. We demonstrated the robustness of our approach at low resolution. The classifier is capable of recognizing more than 98.9% of safe and dangerous situations at very low resolution and up to 99.8% at higher resolutions. Our results enable us to be highly confident in integrating our algorithm on an embedded platform

equipped with low resolution sensors to reach the project's user requirements of low power consumption (several days) and real-time feedback (about 10 fps). As future steps, we will extend our experiments to sidewalks. Our detector will be embedded on an off-the-shelf hardware board connected to a pair of low resolution and high dynamic range cameras for both indoor and outdoor usage for elderly safety and self-confidence.

8 Acknowledgements

This project is supported by the Swiss Hasler Foundation SmartWorld Program, grant Nr. 11083. We thank our end-user partners: the FSASD, "Fondation des Services d'Aide et de Soins à Domicile", Geneva, Switzerland; EMS-Charmilles, Geneva, Switzerland; and Foundation "Tulita", Bogotá, Colombia.

References

1. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Computer Vision and Image Understanding 110(3), 346–359 (2008)
2. Bouhamed, S.A., Kallel, I.K., Masmoudi, D.S.: New electronic white cane for stair case detection and recognition using ultrasonic sensor. International Journal of Advanced Computer Science & Applications 4(6) (2013)
3. BPA: "brochure technique escaliers" and "garde-corps base: norme sia 358". `http://www.bfu.ch/sites/assets/Shop/bfu_2.007.02_Escaliers.pdf` and `http://www.inoxconcept.ch/images/normes_sia_358.pdf` (2009), bureau de prévention des accidents, [Online, accessed 20-June-2014]
4. Bradski, G., Kaehler, A.: Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, 1st ed edn. (Sep 2008)
5. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: computing a local binary descriptor very fast. Pattern Analysis and Machine Intelligence, IEEE Transactions on 34(7), 1281–1298 (Jul 2012)
6. Cloix, S., Weiss, V., Bologna, G., Pun, T., Hasler, D.: Obstacle and planar object detection using sparse 3D information for a smart walker. In: 9th International Conference on Computer Vision Theory and Applications. vol. 2, pp. 305–312. Lisbon, Portugal (Jan 2014)
7. Cong, Y., Li, X., Liu, J., Tang, Y.: A stairway detection algorithm based on vision for ugv stair climbing. In: Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on. pp. 1806–1811. IEEE (2008)
8. Delmerico, J.A., Baran, D., David, P., Ryde, J., Corso, J.J.: Ascending stairway modeling from dense depth imagery for traversability analysis. In: Robotics and Automation (ICRA), 2013 IEEE International Conference on. pp. 2283–2290. IEEE (2013)
9. Equiterre: "rampes, marches et escaliers" and "trottoirs" guidelines. `http://www.mobilitepourtous.ch/pdf/fiche_7.pdf` and `http://www.mobilitepourtous.ch/pdf/fiche_1.pdf`, [Online, accessed 20-June-2014]
10. Furukawa, Y., Ponce, J.: Accurate camera calibration from multi-view stereo and bundle adjustment. International Journal of Computer Vision 84(3), 257–268 (2009)

11. Gupta, M., Yin, Q., Nayar, S.K.: Structured light in sunlight. In: IEEE International Conference on Computer Vision (ICCV) (2013)
12. Hartley, R., Zisserman, A.: Multiple View Geometry in compter vision. Cambridge University Press, second edn. (2003)
13. Hesch, J.A., Mariottini, G.L., Roumeliotis, S.I.: Descending-stair detection, approach, and traversal with an autonomous tracked vehicle. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. pp. 5525–5531. IEEE (2010)
14. Kinsella, K., He, W.: An aging world: 2008. Tech. rep. (Jun 2009), http://www.aicpa.org/research/cpahorizons2025/globalforces/socialandhumanresource/downloadabledocuments/agingpopulation.pdf
15. Konolige, K.: Small vision systems: hardware and implementation. In: Eighth International Symposium on Robotics Research. pp. 111–116. Hayama, Japan (1997)
16. Konolige, K.: Projected texture stereo. In: Robotics and Automation (ICRA), 2010 IEEE International Conference on. pp. 148–155. IEEE (2010)
17. Lacey, G., Rodriguez-Losada, D.: The evolution of guido. Robotics & Automation Magazine, IEEE 15(4), 75–83 (2008)
18. Lacey, G., Dawson-Howe, K.M., Vernon, D.: Personal adaptive mobility aid for the infirm and elderly blind. Lecture Notes in Computer Science (1458), 211–220 (1998)
19. Lee, Y.H., Leung, T.S., Medioni, G.: Real-time staircase detection from a wearable stereo system. In: Pattern Recognition (ICPR), 2012 21st International Conference On. pp. 3770–3773. IEEE (2012)
20. Nalpantidis, L., Sirakoulis, G., Gasteratos, A.: Review of stereo vision algorithms: From software to hardware. International Journal of Optomechatronics 2(4), 435–462 (Nov 2008)
21. OECD: Society at a glance. OECD Social Indicators, 2006 Ed. (2006), organization for Economic Co-operation and Development, 2009 data are available on the Web
22. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision 47(1-3), 7–42 (Apr 2002)
23. Schmidt, E., Manser, J.A.: Directives "voies piétonnes adaptées aux handicapés" rues - chemins - places. http://www.mobilitepietonne.ch/fileadmin/redaktion/publikationen/Strassen_Wege_Plaetze__Richtlinien_fuer_behindertengerechte_Fusswegnetze_f.pdf, [Online, accessed 20-June-2014]
24. Serro, M., Shahrabadi, S., Moreno, M., Jos, J.T., Rodrigues, J.I., Rodrigues, J.M.F., Buf, J.M.H.: Computer vision and GIS for the navigation of blind persons in buildings. Universal Access in the Information Society (Feb 2014)
25. Shahrabadi, S., Rodrigues, J.M., Du Buf, J.H.: Detection of indoor and outdoor stairs. In: Pattern Recognition and Image Analysis, pp. 847–854. Springer (2013)
26. Tseng, C.K., Li, I., Chien, Y.H., Chen, M.C., Wang, W.Y.: Autonomous stair detection and climbing systems for a tracked robot. In: System Science and Engineering (ICSSE), 2013 International Conference on. pp. 201–204. IEEE (2013)
27. Walter, E., Cavegn, M., Scaramuzza, G., Niemann, S., Allenbach, R.: Fussverkehr unfallgeschehen, risikofaktoren und prävention. http://mobilitepourtous.ch/pdf/Bpa_Pietons_accidentes_2007.pdf (2007), [Online, accessed 20-June-2014]