Automatic Assistant for Better Mobility and Improved Cognition of Partially Sighted Persons

Ruxandra TAPU[1,2], Bogdan MOCANU[1,2], Titus ZAHARIA[2]
[1]*Faculty of ETTI, University "Politehnica" of Bucharest, Romania*
[2]*Institut Mines-Télécom / Télécom SudParis, ARTEMIS Department, UMR CNRS MAP5 8145, France*
{ruxandra.tapu, bogdan.mocanu, titus.zaharia}@telecom-sudparis.eu

***Abstract*—In these paper we introduce a novel computer vision assistant for autonomous navigation of partially sighted people. We begin by detecting any type of static and dynamic obstacle present in the scene. Then, we introduce an adapted version of HOG (Histogram of Oriented Gradients) descriptor incorporated into the BoVW (Bag of Visual Words) retrieval framework and demonstrate how this combination can be used for obstacle classification. The design is completed with an acoustic feedback that alert user of potential hazards. The audio bone conduction is employed to allow the visually impaired to hear other sounds from the environment. At the hardware level, the system is totally integrated on a smartphone which makes it easy to wear, non-invasive and low-cost.**

***Index Terms*—visual impaired navigation assistant, obstacle detection and classification, audio feedback, smartphone device.**

I. Introduction

According to the National Federation for Blind the total number of visually impaired people worldwide is 160 million while 37 million are completely blind [1]. In this context the need of assistive devices for autonomous navigation was and will be constant. Nowadays the white cane and the walking dogs still represents the most popular tools used for obstacle detection. The cane is also the cheapest, the simplest and most reliable element used as navigation aid. However, it is not able to provide additional information elements such as: the speed and type of object the user is encountering, if the obstacle is static or dynamic, the distance and time to collisions. This information is gathered for normal users by their eyes and it is absolutely necessary to have it in order to percept and control the locomotion during the navigation [2].

In its absence the visual impaired users (VI) struggle to memorize all places they have been to in order to recognize them afterwards. In unknown setting, they feel insecure and depend on other humans to guide them and reach the desired destination [3]. The task of route planning in an unforeseen obstacle environment can severely impede the independent travel of VI and will reduce their willingness to travel, despite of having access to white canes or walking dogs [4].

In this context, in order to improve cognition and assist the navigation of VI users, it is absolutely necessary to develop a real time system that is able to recognize static and moving objects, in highly dynamic urban scenes. This technology will not replace the cane, but should complement it by alerting the user of obstacles in a few meters, provide guidance. Only one major constraint is imposed on the system: it should not interfere with the other senses as: acoustic or haptic [5].

Motivated by the above consideration in this paper we introduce a novel navigation assistant, developed using the computer vision techniques, designed to detect and classify obstacles encountered by the VI user in outdoor navigation. The system is completed with an acoustic feedback that will warn user by the presence of any type of obstacle situated in his/her near surrounding. At the hardware level the system is embedded on a regular smartphone running as a real time application. In this context the entire framework can be regarded as simple to use device, completely novel, non-intrusive and low-cost. The main steps involved in the proposed framework are illustrated in Fig. 1.

The rest of the paper is organized as follows: in Section II we review the technical literature. In Section III and IV we introduce a novel, real-time framework for obstacle detection and classification. In Section V the acoustic feedback for assisting the visual impaired is explained. Section VI gives the experimental results in which we consider challenging environments with various arbitrary moving object and real blind users. While at the end, Section VII concludes the paper and provides some perspectives for future work and implementation.

II. State of the Art Review

In the last couple of years various assistive technologies for blind or visually impaired people were introduced. Most of them are based on ultrasonic, infrared or laser sensors [6], [7]. Even though ETA (Electronic Travel Aids) have been widely used [8], [9], [10] they have had limited success because of an inadequate interface and usability. Such interface (*e.g.* acoustic, haptic) tries to provide a sensory substitution to vision. However, the information captured by the eyes cannot be entirely substitute by audition or touch. In addition, the overall hardware architecture of any existent ETA systems should be embedded, lightweight and should be comfortable to wear [11].

Because, normal humans acquire most of there navigation information through visual perception it is tempting to use artificial vision in the case of the partially sighted. Due mostly to the high limitations introduced by computational power and the lack of robustness of the vision algorithms most researchers have selected not to use the computer vision for the visually impaired people mobility assistants.

This work has been funded by the Sectorial Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/132395 (InnoRESEARCH).

Digital Object Identifier 10.4316/AECE.2015.03006

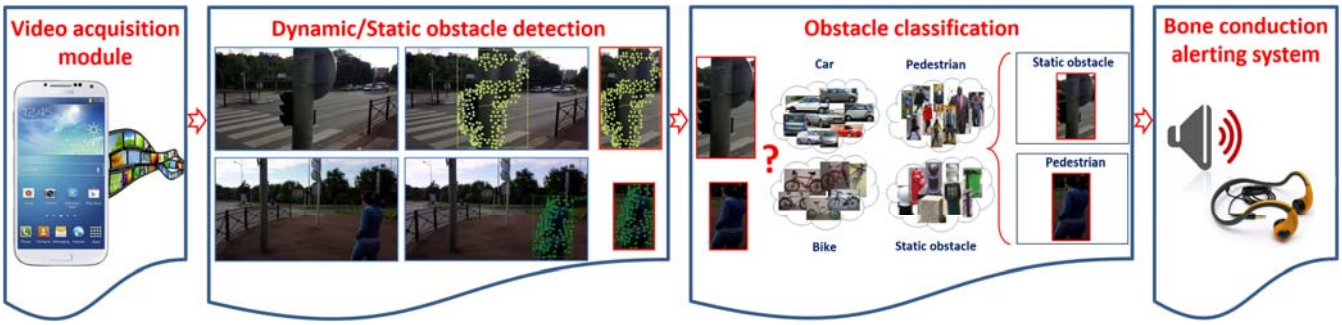

Figure 1. Obstacles detection and classification framework

Nevertheless, in the past years, extraordinary advances in computers and vision techniques have been made, and now it is possible to perform reliable real-time algorithms on embedded computers and even smartphones that run on powerful multi-core processors. The computer vision systems, unlike ultrasonic, infrared or laser technologies offer superior level of reality reproduction in exchange of processing complexity.

One of the first computer vision techniques, existent in the literature, is called vOICe [12] designed to augment the human hearing. By using the regular laptop and a video camera the proposed framework transforms the acquired images into a sound map of the environment. The sound is sent to a pair of headphones. The system is simple, small, light weighted and cheap. Its drawbacks are given by the use of regular headphone that blocks the user ears. Also, the visually impaired requires an extensive training phase to become familiar with the sound patterns.

The virtual acoustic space technique proposed in [13] creates a sound map of the environment. By using the stereoscopic vision the method develops depth maps and then generates sounds depending on the user position. The system is simple with two cameras embedded in eye-glasses and has low weight. As disadvantages the system was not testes in real environments and also blocks the user hearing.

The electron-neural vision system (ENVS) introduced in [14] aims to detect obstacles with the help of a head set stereo vision, GPS and electro-tactile stimulation. As major improvement compared with the above techniques, this system works in real time and doesn't block the users hearing. However, by using electrical nerve stimulation gloves the human hands are always occupied. Moreover the ground and overhead objects are not detected, while the walking path needs to be flat (*i.e.* no stairs).

The tactile vision system (TVS) proposed in [15] is a wearable device that detects obstacles and converts visual information into tactile signals.

The prototype is composed of stereo camera belt, a tractor belt and a portable computer. The system works in real time and gives user free hands without blocking the hearing. It major disadvantage is that it cannot differentiate overhead and ground obstacles. Also, no experiments on real visually impaired were made.

A complete framework for obstacle detection with the help of a stereo camera system, a jacket with 4x4 vibration array and a laptop computer is proposed in [16] called Tyflos. The technique can detect obstacles situated at various height levels and don't block the user hearing. However, the system is invasive (the vibration jacket needs to be situated near the skin) while some tests on real users need to be performed.

A people and obstacle detection system in proposed in [17] entitled Kindetect. The technique uses a Kinect, as a depth sensor, an acoustic feedback system and a computer in order to identify obstacles situated at head or foot level. In this case the system has limited operation area: in indoor environments.

The Vibratory Belt [11] architecture is composed of a Kinect sensor, a belt with vibration and an embedded computer for obstacle detection. The system can detect head level obstacles, doesn't block the hearing and it is easy to use. However the Vibratory Belt is limited to indoor environments.

To the best of our knowledge, the only paper addressing the problem of obstacle detection by using monocular cameras embedded on smartphones is introduced in [18]. The proposed approach is evaluated solely in indoor scenarios and only on normal humans. No acoustic feedback is sent to the user in order to infer information. In addition, the hands-free condition imposed by the visually impaired users [19] is violated.

III. Obstacle Detection System

For the obstacle detection we adopted our technique introduced in [20] which is further extended with a temporal consistency step and an object relevance establishment phase.

We start by selecting a set of interest points, regularly spread on a frame, based on an image grid. The points are tracked between successive frames using the multi-scale Lucas-Kanade algorithm. The camera and background motion are identified using a set of homographic transforms by applying the RANSAC algorithm on the interest points correspondence. Other types of movements are detected through an agglomerative clustering process.

Due to the foreground apparent movement even static obstacles situated in the user proximity act like dynamic objects.

One of the major constraints of the system presented above is given by the objects lack of temporal consistency between two successive frames (Fig. 2). In outdoor application it is very common that objects disappear, stop or are occluded for a period of time. In this context the tracking algorithm will lead to incomplete trajectories and will identify the same object as a new hazard existent in the scene.

Based on this observation we propose reinforcing the object detection process with a multi-frame fusion scheme.

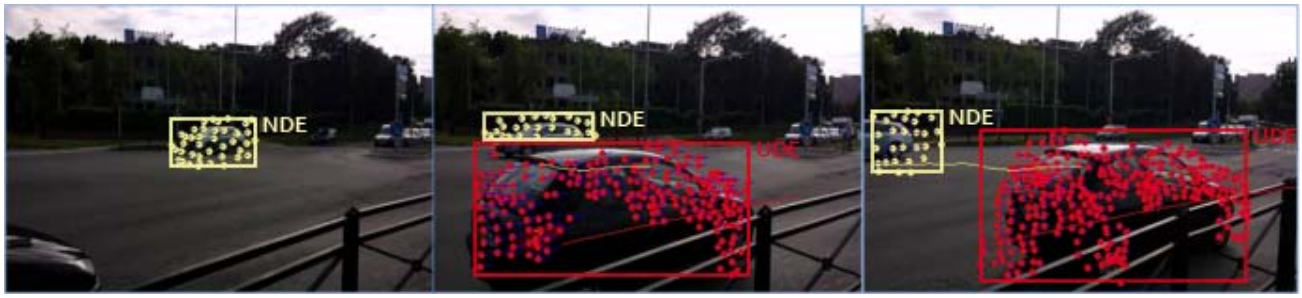

Figure 2. Motion classes temporal consistency

By saving the object location and its average velocity within a temporal sliding window of size *N*, we can predict its novel position using the following equation:

$$p_i(x_i,y_i,t_i)=p_i(x_i,y_i,t_{i-1})+\frac{\sum_{j=1}^{N} v_j(v_{jx},v_{jy},t_j)}{N}-p_i^{est}(x_i^{est},y_i^{est},t_i) \quad (1)$$

where $p_i(x_i, y_i, t_i)$ is the i^{th} interest point in the frame at the moment of time t_i, $v_i(v_{jx}, v_{jy}, t_j)$ the object's velocity and $p_i^{est}(x_i^{est}, y_i^{est}, t_i)$ the point location estimated using a set of homographic transforms. p_i^{est} helps us compensate the camera and background movement.

Once the obstacles are identified we must determine their degree of danger and classify them. This is done by using the object position and direction relative to the video camera (Fig. 3). An object is labeled as *approaching* (AP) to the subject if its associated direction points into the camera's focus of attention. Otherwise, the subject is considered as moving away from the obstacle or that the object is *departing* (DE).

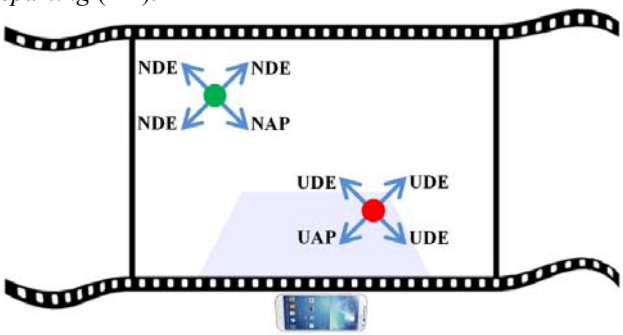

Figure 3. Obstacles relevance establishment based on their position and relative direction

Then, by using a trapezium projected on the image, an obstacle is marked as *urgent (U)* if it is situated in the proximity of the blind/visual impaired person (2 meters). Otherwise, if is located outside the trapezium, the obstacle is categorized as non-urgent or *normal (N)*. However, by employing two areas of proximity we can prevent the system to continuously warn the subject about any object existent in the scene. A warning can be launched just for objects situated in the urgent region (Fig. 3).

The downside of this assumption is given by the rejection of warnings for dynamic objects (*e.g.* vehicles) approaching the user very fast or for obstacles situated high at the head level, such as tree branches, arcades or small banner. To avoid these situations in the following part of the paper (Section IV) we introduced a novel classification method design to help us differentiate between various types of obstacles. Using this information we can generate warning outside the trapezium, when such action is required.

Nevertheless the size of the trapezium can be adjusted in a pre-calibration step by the user.

IV. Obstacle Classification System

In the content based representation, each frame of the video stream can be viewed in terms of a hierarchical structure with increasingly higher levels of abstraction oriented towards capturing the semantic meaning of the objects represented by it. In this framework we have considered a training dataset divided into the following four major categories depending on their relevance to a VI person: vehicles, bicycles, pedestrians and static obstacles (Fig. 4). The considered categories were selected according to the most important obstacles encountered in an outdoor navigation scenario.

The class entitled "static obstacles" contains a high variability of instances such as: overhanging branches, fences, garbage cans, bumps, trees, pylons, edge of pavements, traffic signs, steps…

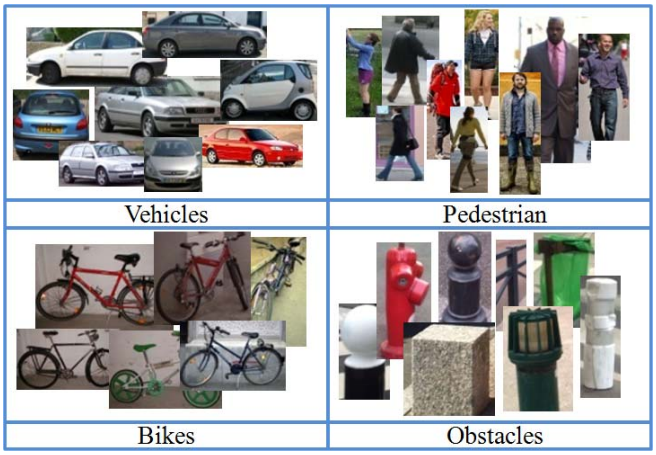

Figure 4. Outdoor obstacles

We started by exacting for each image in the training dataset an *adapted version of the HOG* (Histogram of Oriented Gradients) descriptor. The estimation of a traditional HOG [21] assume dividing an image *I(x,y)* into cells and computing for each cell its pixels gradient direction in order to form an one dimensional histogram. For better invariance to illumination changes and shadows, the local histogram energy of image block is used to normalize all cells included in that block. The HOG descriptor was originally developed for human recognition tasks.

In this context the authors in [21] propose using an analysis window of 64 x 128 pixels. In the case of our system this requirement implies constraining the resolution of the image patch (extracted using the obstacle detection method described in Section III) to a fixed size (Fig. 5).

Figure 5. HOG descriptor extractor for a fixed resolution of the image patch

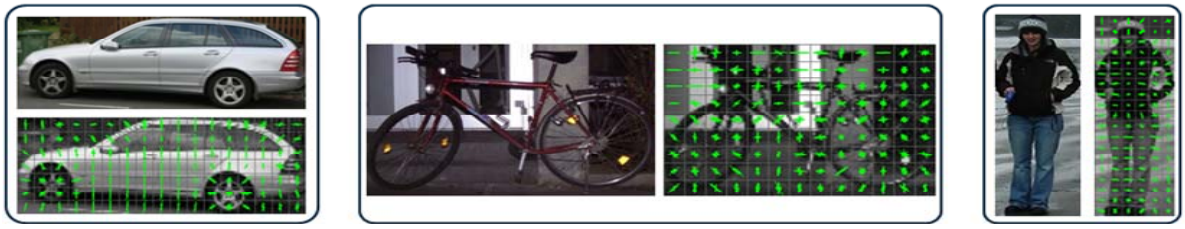

Figure 6. HOG descriptor extractor for a fixed number of cells

Because we are focused on classifying any type of obstacle the patch resizing to a fixed resolution will alter significantly its aspect ratio. In the case of other categories as pedestrians the associated HOG descriptor will have a reduce discriminative power. To overcome this shortcoming various authors [22] propose using different image sizes for each category (*e.g.* for bicycles 120 x 80 pixels, for cars 104 x 56 pixels…).

Even so, in our case the static obstacle class is characterize by a high variability of instances and it is impossible to find a specific resolution which can be adequate for any element.

Another aspect that needs to be highlighted is the real time application constraint imposed to our system, designed to assist the VI navigation. In our case, we cannot afford spending time to determine which resolution of the image patch is the most appropriate to a certain category and then make a decision. The user needs to be alerted as soon as possible so he could avoid collision.

In this context we propose overcoming the above limitation by introducing an extended version of histogram of oriented gradients descriptor denoted *adapted HOG (A-HOG)*. For extracting an A-HOG we conserve the original aspect ratio of the image patch and not modify (distort) it to match the requirements of a specific class. Then, we constrain the total number of cells for which we compute the descriptor. In order to satisfy this constraint we propose reducing the size of the patch in such a way to meet both requirements: conserving the initial aspect ratio and matching the fixed number of cells imposed (Fig. 6).

In our development we considered for the image patch a fixed number of cells equal to 128 and a cell dimension of 8 x 8 pixels with 18 orientation bins.

Next, we propose integrating the *Bag of Visual Words (BoVW)* model in our real-time obstacle classification task. In a typical BoVW framework a set of interest points (*e.g.* SIFT, SURF…) are extracted from each image [23]. Following, an unsupervised learning step is performed over the entire set of descriptors to determine k clusters. The centroid of each cluster represents a visual word, while the codebook gives the vocabulary.

In our case, each image $I(x,y)$ included in the dataset used for training can be represented using its associated descriptors:

$$I(x,y) = \{d_1, d_2, d_3, ..., d_n\} \quad (2)$$

where n is the total number of patches in the image and d_i is the A-HOG descriptor of an image patch. For each we considered as representative the cell central pixel.

However, developing visual words only by using cells of 18 bins is insufficient to capture the contextual information. In order to avoid this limitation we propose creating an off-line vocabulary $W = \{w_1, w_2, ..., w_M\}$ from larger visual words (*e.g.* blocks) obtained after concatenating adjacent cells. For clustering we used the k-means algorithm. Once the visual vocabulary is developed each image block is now mapped to the nearest word according to the following equation:

$$w(d_k) = \arg\min_{w \in W} Dist(w, d_k) \quad (3)$$

where $Dist(w, d_k)$ is the L_1 distance between the visual word w and the descriptor d_k and $w(d_k)$ denotes the visual word assigned to the k^{th} descriptor d_k.

At the end, the image is represented as a histogram of visual words. The number of bins in the histogram is equal to the total number of words in the vocabulary (*i.e.* M). If each bin b_i represents the occurrence of a visual word w_i in W then b_i = Card (D_i):

$$D_i = \{d_k, k \in \{1,...,n\} | w(d_k) = w_i\} \quad (4)$$

where D_i is the set of descriptors associated to a specific visual word w_i in the considered image. Card(D_i) is the cardinality of the set D_i. This process is applied recursively for every word in the vocabulary to form the final histogram.

After the feature vectors are quantized to clusters, the labeled data is sent to a *supported vector machine* (SVM). The SVM is used to adapt a statistical decision procedure that helps us distinguish between categories. We selected the SVM procedure introduced in [24] that tries to determine a separation hyperplane, between two classes while maximizing the margin:

$$\varphi(x) = sign\left(\sum_i y_i \alpha_i K(x, x_i) + b\right) \quad (5)$$

where b is the hyperplane free term, K is the SVM kernel, x_i are the training features from the data set, y_i the label of x_i while α_i is a parameter dependent on the kernel type. Because in our case a multi-category classification problem

is raised, we selected the one against all strategy.

The SVM training completes the offline process of our object classification framework. In the online phase, for each image patch extracted using our obstacle detection method presented in Section III, we develop a frequency histogram using the A-HOG features. Next, the histogram is match to its closest word in the vocabulary *W*. Our technique requires a reduce computation power because we are not performing an exhaustive sliding window search within the current frame in order to determine objects and their associated positions. In our case the obstacle classification receives as input query, the location and size of the object we want to label.

V. Acoustic Feedback

The acoustic feedback is responsible of informing the VI user about the presence of potential static/dynamic obstacles in his way. As a consequence, the subject will react accordingly to avoid collision. The audio interface need to satisfy one major requirement: "not to block the user's ears" [5]. In this work, two blind associations: Communication for Blinds and Disabled People and Union of the Blinds and Partially Sighted of Slovenia have collaborated with us to help design the human-machine interface (HMI) and the acoustic feedback. In order to determine the path to destination or insight potential dangerous situations the VI people use the sounds from their surroundings to infer information. For example, they use the sounds coming from vehicles to understand the orientation of the streets so they can avoid drifting or follow a straight trajectory.

In these conditions we propose using the audio bone conduction technology which is easy to wear and ears-free. After our discussion with the association of VI user we determine that it would be appropriate to use voice messages rather than beeping. The beeping strategy can only warn user about the proximity or relative position of an obstacle. In our opinion is quite important to distinguish between different types of hazards the subject is facing. So, let consider an outdoor navigation scenario in which the VI comes across a static obstacle or even a pedestrian. In this case the degree of danger is reduced and the user can continue his/her displacement. However, if a bicycle of vehicle is approaching it is mandatory that the system alert the VI so he/she can change direction, stop, walk away from the imminent menace and prevent collision.

In the case were various objects are presented in the scene, in order to not confuse the partially sighted, only one warning at a time, should be generated by the system. Table I gives the set of alarms our framework launches in descendent order of importance (Relevance). In the first phase of the prototype we wanted to limit the alerts only to those objects situated within the trapezium of proximity (Fig. 3). But, due to the high impact an approaching vehicle/bicycle has on the VI we extended the warning list.

In order to inform the visual impaired by the relative orientation of each obstacle the messages is encoded in stereo: the right, the left or both channels. If the warning is transmitted from the left channel the obstacle is situated on the left side of the subject. An analogous process can be made for obstacles situated on the right. In the case of frontal objects the message can come from both channels.

TABLE I. WARNING MESSAGES GENERATED BY OUR SYSTEM

Rel.	**Warning**	Rel.	**Warning**
1	Car urgent approaching	7	Obstacle normal approaching
2	Car normal approaching	8	Car urgent departing
3	Bike urgent approaching	9	Bike urgent departing
4	Bike normal approaching	10	People urgent departing
5	People urgent approaching	11	Obstacle urgent departing
6	Obstacle urgent approaching	12	People normal approaching

VI. Experimental Results

The system architecture is composed of a regular smartphone (Samsung Galaxy S4) attached to a chest mounted harness (Harness-GoPro) and bone conduction headphones (Aftershokz AS 301). The harness has two major roles: it helps satisfy the hands-free requirement imposed by the VI and improves the video acquisition process. Even though the partially sighted have learned not to swing much their bodies during movement if the recording camera will be attached only by strings the resulted video flow will be unstable in time leading to cyclic pan and tilt oscillation.

In this context our system can be described as: a simple device, ready to use by the VI without any training, low-costs because it does not require any expensive hardware architecture since all components can be independently found on the market and it is also non-intrusive satisfying the hands-free and ears-free requirements imposed by users.

We tested our system in multiple complex outdoor urban environments in Paris, France with the help of visual impaired users. The videos were also recorded and used to develop a testing database with 30 elements. The average duration of each video is around 10 minutes acquired at 30 fps in color format at an image resolution of 320 x 240 pixels.

The image sequences are very challenging because they contain in the same scene multiple static and dynamic obstacles such as: vehicles, pedestrian or bicycles. Also, because the recording process is done by VI users, the videos are trembled, noisy, include dark, clutter and dynamic scenes. In addition, different types of camera and background motion are present.

In Fig. 7 we give some experimental results. For each video we present five frames that are very close in time. Different colors are used to represent various moving obstacles existent in the scene. Due to our temporal consistency step, the color associated with the object remains unchanged between successive frames.

From *Videos 1 – 6* we can observe that our system can correctly detect static obstacles (*e.g.* pillars, road signs and bushes) situated either at the head level or down on the foot area at around two meters distance from the user. In all case the recoding camera has important motion caused mostly by the subject own displacement.

Regarding the dynamic obstacles (*e.g* pedestrian, bikes or vehicles) these are detected at larger distances from subjects (ten meters). However, because in some cases, only parts of the obstacle are detected, the classification phase can be penalized by this behavior.

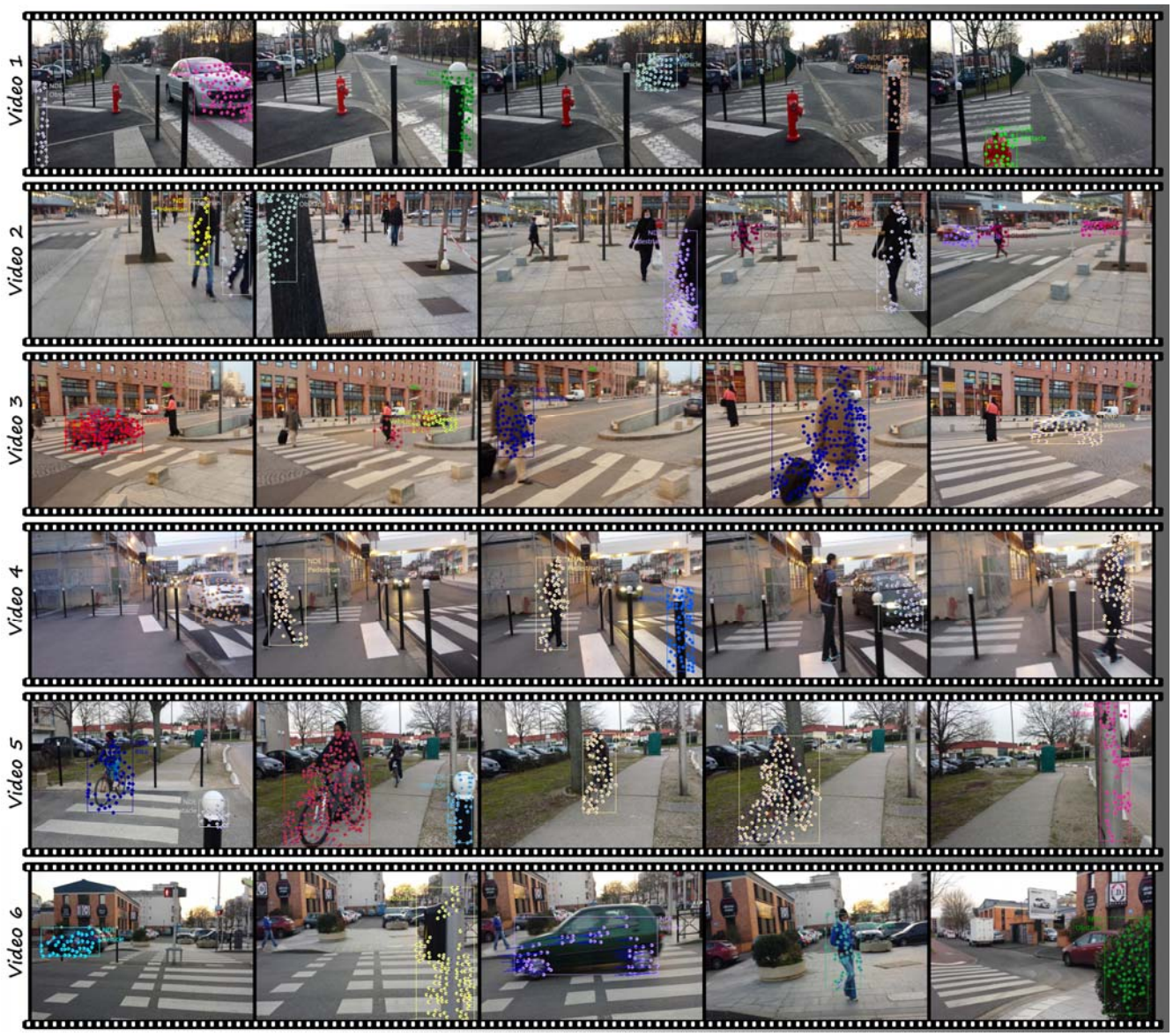

Figure 7. Experimental results of our obstacle detection and classification framework

In the case of *Video 3* the pedestrian in the second frame is labeled as obstacle because only the body of the subject is given as input to the classification method.

The system is robust to important changes in the illumination conditions, as for the *Video 4* where the recording is done at sunset. In this case our technique can correctly detect and classify the objects presented in the scene: vehicle, pedestrian or pillar.

In the following part we present a comprehensive evaluation of the obstacle classification module when modifying the various parameters involved. We conducted multiple tests on a set of 2432 image patches that were extracted from the video database using our obstacle detection method introduced in Section IV.

Concerning the performance evaluation metrics, two types of classification errors are commonly used:

- The first one corresponds to the so-called missed detection/classification (or false negatives) and relates to the case where the detection/classification failed to identify some static/dynamic object existent in the scene or the object was not classified to any category (*e.g.,* car, bike, pedestrian, obstacle or beep).

- The second error corresponds to the situation where some false elements (*e.g.,* shadows) have been erroneous classified as static/dynamic obstacles. In this case a false alarm or false positives is given by an object that is assigned to the wrong category.

By using the ground truth test data set, such classification errors can be globally described with the help of two error parameters, denoted *MC* and *FC*, representing the number of missed classified objects and false classified obstacles. Let us denote by C the total number of correctly classified obstacles. Based on these entities, the most often popular evaluation metrics encountered in the technical literature are the so called, precision (*P*) and recall (*R*) rates, respectively defined as described bellow:

$$P = \frac{C}{C + FC}; \tag{6}$$

$$R = \frac{C}{C + MC}; \tag{7}$$

Precision can be interpreted as a measure of exactness or fidelity, whereas recall is a measure of completeness.

TABLE II. OBSTACLE CLASSIFICATION MODULE PERFORMANCE EVALUATION

	Cars	Bikes	People	Static Obs.	Outlier	Ground Truth (GT)	Missed Classified (MC)	False Classified (FC)	Precision (P)	Recall (R)	F1score (F1)
Cars	762	8	0	14	10	794	32	40	0.95	0.96	0.96
Bikes	11	214	58	9	17	309	95	51	0.81	0.69	0.74
People	18	49	826	6	8	907	81	52	0.94	0.91	0.92
Static Obs.	15	13	22	334	38	422	88	29	0.92	0.79	0.85

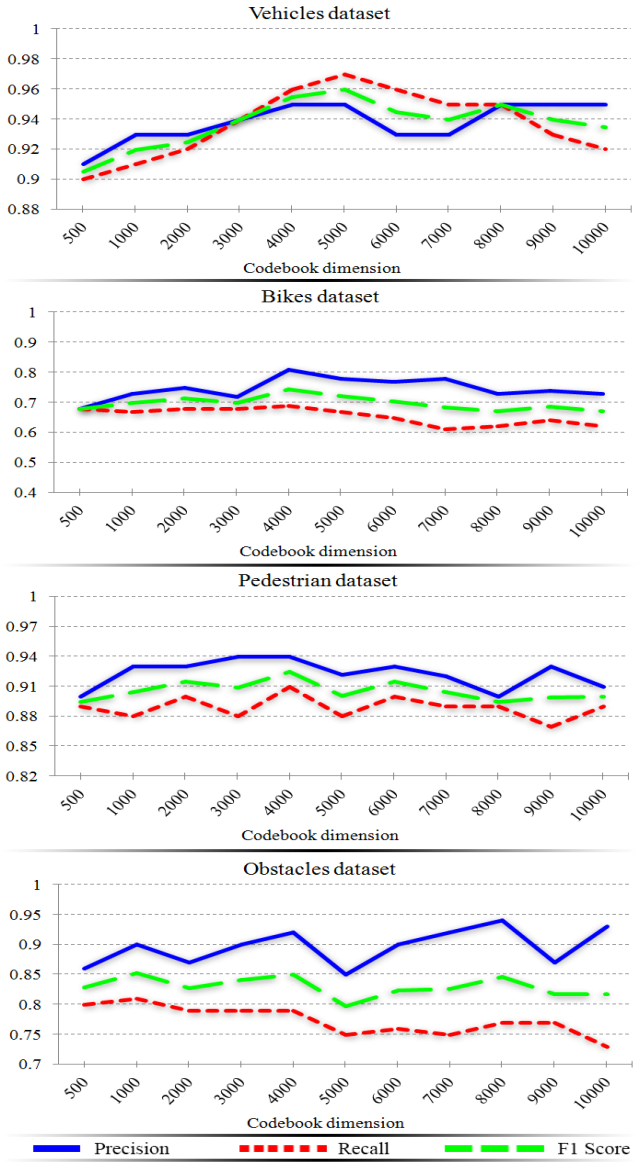

Figure 8. Precision, recall and F1 score variation with the increase of the codebook size

The recall and precision can be combined within a unique evaluation metric, denoted by F1 score defined as the harmonic mean of precision and recall rates:

$$F1score = \frac{2 \cdot P \cdot R}{P + R}; \tag{8}$$

Ideally, the precision, recall and F1 score should be equal to 1, which corresponds to the case where all detected objects are correctly classified, without neither false alarms nor missed classifications.

In Table II we give, for a vocabulary size of 4000 words, the system performance for each considered category and the associated confusion matrix.

As it can be observed from Table II we extended the total number of categories on object can be classified with one denoted *Outliers*. We adopted this approach to make sure that our system classifies a patch to a class due to its high resemblance with a word in the vocabulary and not just because it is required to make a decision. For all the objects included in the *Outlier* class a beep signal will be generated in order to alert the user about its presence.

We studied next the impact the vocabulary size has on the overall system performance. We present in Fig. 8 the experimental results obtained in terms of precision, recall and F1 score. As it can be noticed, a vocabulary with 4000 words returns the best results.

However, we have to consider that our framework is designed to work as a real time-application for which the classification speed is a crucial parameter. With the increase of the vocabulary size, the computational complexity will multiply exponentially. So, due to this constraint we selected for the vocabulary a size of 1000 words.

We studied next the impact of the density of obstacles over the detection system performance. As explained above, from the visually impaired point of view it is not important to detect all the obstacles (dynamic or static) existent in the scene but, just the ones situated in his near surrounding. So, we focused our attention only on the obstacles captured inside the trapezium of proximity (Fig. 9). As it can be noticed, when increasing the total number of objects, the F1 score is decreasing. This behavior can be motivated by the fact that when multiple objects have the same movement and are if they are close together in space they could be identified as one single object. However, even if multiple obstacles are present, in order to not confuse the partially sighted, only one warning at a time is generated by the system, at an interval of two seconds.

Regarding the computational complexity, the average processing time of the entire framework (obstacle detection and classification) when it is run on a Samsung S4 smartphone is around 140 ms per frame which leads to a processing speed around 7 frames per second.

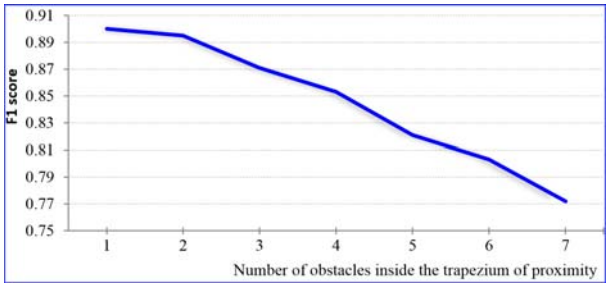

Figure 9. Detection performance with the variation of obstacle density

VII. CONCLUSIONS AND PERSPECTIVES

In this paper we have introduced a novel assistive device simple and portable satisfying both the hands and ears-free constraints to facilitate the partially sighted person navigation in outdoor scenarios. Without any *a priori* information about the obstacle type, size, position or location the proposed framework is able to detect and

classify, in real time, both static and dynamic obstacles. Then, through an audio feedback a set of warning is launched to the VI. The methods works as a real-time application, embedded on a regular smartphone that is attached to the user with the help of a chest mounted harness.

We tested our method on different outdoor scenarios with visually impaired participants. The system shows robustness and consistency even for important camera and background movement or for crowded scenes with multiple obstacles.

For further work we propose integrating our system in a much more developed assistant which includes: navigation tools (that offer guidance to reach a desired destination), stairs and crossings detectors and people recognizer (that helps identify the familiar persons).

In the presence of sudden movements or when the illumination conditions change drastically the overall system performance may decrease. Based on this observation we intend to develop a safety mechanism that will automatically put on hold the obstacle detection and classification module when such cases are identified by using the smartphone sensorial information. We intend develop to a safety mechanism that will automatically put on hold the obstacle detection module when very sharp movements are identified. On the other hand a much more elaborated study on VI users will offer us some further guidelines.

REFERENCES

[1] D. Pascolini, S. P. Mariotti, "Global data on visual impairments 2010," World Health Organization, Geneva, 2012.

[2] B. B. Blasch, W. R. Wiener, and R. L. Welsh, "Foundations of Orientation and Mobility", 2nd New York: American Foundation for the Blind AFB Press, pp. 42-55, 1997.

[3] C. Shah, M. Bouzit, M. Youssef, and L. Vasquez, "Evaluation of RU-Netra Tactile Feedback Navigation System for the Visually Impaired," International Workshop on Virtual Rehabilitation, pp. 72-77, 2006. [Online]. Available: http://dx.doi.org/10.1109/IWVR.2006.1707530.

[4] R. G. Golledge, J. R. Marston, and C. M. Costanzo, "Attitudes of visually impaired persons towards the use of public transportation," Journal of Visually Impairment Blindness, vol. 90, pp. 446–459, 1997.

[5] A. Rodriguez, J. J. Yebes, P. F. Alcantarilla, L. M. Bergasa, J. Almazan, and A. Cela, "Assisting the visually impaired: obstacle detection and warning system by acoustic feedback," Sensors, vol. 12, pp. 17476-17496, 2012. [Online]. Available: http://dx.doi.org /10.3390/s121217476.

[6] D. Dakopoulos, N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: a survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol.40, no.1, pp.25-35, Jan. 2010. [Online]. Available: http://dx.doi.org/10.1109/TSMCC.2009.2021255.

[7] J. A. Hesch, S. I. Roumeliotis, "Design and analysis of a portable indoor localization aid for the visually impaired," Journal of Robotics Research, pp. 1400–1415, 2010. [Online]. Available: http://dx.doi.org/10.1177/0278364910373160.

[8] J. M. Sáez, F. Escolano, and A. Peñalver, "First steps towards stereo based 6DOF SLAM for the visually impaired," IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, pp.23-23, June 2005. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2005.461.

[9] V. Pradeep, G. Medioni, J. Weiland, "Robot Vision for the Visually Impaired," IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.15-22, June 2010. [Online] Available: http://dx.doi.org/ 10.1109/CVPRW.2010.5543579

[10] J. M. Saez, F. Escolano, "Stereo-Based Aerial Obstacle Detection for the Visually Impaired", In ECCV workshop. on Computer Vision Applications for the Visually Impaired, France 2008.

[11] J. M. Loomis, R. L. Klatzky, N. A. Giudice, "Sensory substitution of vision: importance of perceptual and cognitive processing", in Assistive Technology for Blindness and Low Vision, Eds. Boca Raton, pp. 161-192, 2013.

[12] P. B. L. Meijer, "An experimental system for auditory image representations", IEEE Trans. Biomedical Engineering, vol. 39(2), pp. 112–121, 1992. [Online]. Available: http://dx.doi.org/10.1109/ 10.121642.

[13] J. L. Gonzalez-Mora, A. Rodrıguez-Hernandez, L. F. Rodrıguez-Ramos, L. Dıaz-Saco, N. Sosa, "Development of a new space perception system for blind people, based on the creation of a virtual acoustic space", Lecture Notes in Computer Science Engineering Applications of Bio-Inspired Artificial Neural Networks, pp. 321-330, 1999. [Online]. Available: http://dx.doi.org/10.1007/BFb0100499

[14] S. Meers and K. Ward, "A substitute vision system for providing 3D perception and GPS navigation via electro-tactile stimulation," 1st Int. Conf. Sens. Technol., New Zealand, Nov. pp. 21–23, 2005.

[15] L. A. Johnson and C. M. Higgins, "A navigation aid for the blind using tactile-visual sensory substitution," in Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 6298–6292, 2006. [Online]. Available: http://dx.doi.org/10.1109/IEMBS.2006.259473.

[16] D. Dakopoulos and N. Bourbakis, "Preserving visual information in low resolution images during navigation for visually impaired," Proceedings of the 1st International Conference on PErvasive Technologies Related to Assistive Environments, Athens, Greece, pp. 1-6, 2008. [Online].Available:

[17] http://dx.doi.org/10.1145/1389586.1389619

[18] A. Khan, F. Moideen, J. Lopez, W. L. Khoo, Z. Zhu, "KinDetect: Kinect detecting objects", in Computer Helping people with special needs, vol. LNCS7382, pp. 588-595, 2012.

[19] E. Peng, P. Peursum, L. Li, S. Venkatesh, "A smartphone-based obstacle sensor for the visually impaired", Lecture Notes in Computer Science, Ubiquitous Intelligence and Computing, pp. 590–604, 2010. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-16355-5_45

[20] R. Manduchi, "Mobile vision as assistive technology for the blind: An experimental study", Proceedings of the 13th International Conference on Computers Helping People with Special Needs, volume 2, pp. 9–16, Austria, 2012. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-31534-3_2

[21] R. Tapu, T. Zaharia, "Salient object detection based on spatiotemporal attention models," IEEE International Conference on Consumer Electronics (ICCE), pp.39-42, Jan. 2013. [Online]. Available: http://dx.doi.org/10.1109/ICCE.2013.6486786.

[22] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.886-893, June 2005. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2005.177.

[23] N. Dalal, B. Triggs, "Object detection using histograms of oriented gradients", in European Conference on Computer Vision, vol. 1, pp 886-893, 2006.

[24] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, C. Bray, "Visual categorization with bags of keypoints," In Workshop on Statistical Learning in Computer Vision, ECCV, pp. 1-22, 2004.

[25] S. Tong, E. Chang, "Support Vector Machine Active Learning for Image Retrieval," Proceedings of the Ninth ACM International Conference on Multimedia., pp. 107-118, 2001. [Online]. Available: http://dx.doi.org/10.1145/500141.500159.