Wearable vision systems for personal guidance and enhanced assistance

G. López-Nicolás, A. Aladrén, and J. J. Guerrero
Instituto de Investigación en Ingeniería de Aragón - Universidad de Zaragoza, Spain

***Abstract*—The challenge addressed in this paper involves research of computer vision and robotic techniques to be part of a personal assistance system based on visual information. Main features envisioned include not only navigation assistance in known or unknown environments, but also enhanced capabilities in safety and human perception augmentation. Current systems in the context of assistive devices do not provide enough level of performance, and their capabilities are still quite limited. These are normally based on conventional sensors and, although perception can be greatly improved by sensor fusion, long term research should also consider unconventional vision devices to expand the system's potential. We refer as unconventional vision systems to those consisting in more than a classical or standard camera. So, the visual assistant will be wearable including conventional and unconventional camera systems. We aim for a human-centered system, which complements rather than replaces human abilities, allowing enhanced interaction and integration of the user in the environment. We present several complementary approaches based on different unconventional camera systems. Possible users of the investigated technologies will range from visually impaired people to users with normal visual capabilities performing specific tasks, passing also into humanoid robots.**

I. Introduction

The ability of navigating effectively in the environment is natural for people, but not easy to achieve under certain circumstances, such as the case of visually impaired people or unknown and intricate environments. In a similar way, the recognition of places, objects or signs is another fundamental ability in our daily life. Humans solve these problems mainly with vision and memory combined with our learning ability. According to the World Health Organization (WHO), in 2020 there will be 75 millions of blind people and more than 200 millions visually impaired. However, we should note that not only these people would benefit from a personal visual assistant, but also people with common visual capabilities performing specific tasks (such as firefighters, police or tourists). A personal assistant that helps with the localization and navigation in unknown, difficult or not frequently visited environments, with recognizing a building or with reading a sign in an unknown language would be of great interest.

Different wearable navigation systems have been proposed in the literature for visually impaired people, and a detailed survey is provided in [1]. Currently, there is an increasing interest in devising robotics systems for personal assistance. For instance, [2] presents a robotics proposal for guiding visual-impaired people and for people safety in rescue operations using portable laser-scanner systems on the head. Previous vision-based methods adapted to human navigation and guidance considered stereo vision [3], or wireless communication technology [4]. Another interesting related proposal is the vision based navigation assistant presented in [5], which builds a topological map using a system composed of four uncalibrated cameras mounted on the user's shoulders. However, the capabilities of these systems are still quite limited and they do not provide enough level of performance as would be required by the possible end users. Additionally, future applications envisioned by society are even more and more demanding in terms of the quality and quantity of the information gathered from the environment by the assistive system.

In this framework, usual approaches consider standard sensors to gather the input information, such as ultrasonic sensors, conventional cameras or compasses. The perception results are also greatly improved by sensor fusion, and relevant advances have been addressed in the last decades in this topic. Nevertheless, we think that long term research should also consider unconventional vision devices to expand the system's potential and to overcome the gap between the research results and the end user demands. Thus, the main goal is to investigate the possibilities of unconventional vision sensors, currently an open research issue, due to the great amount of information they provide and to their low cost and good miniaturization perspectives. A spectacular increase in low-cost computational power has also taken place, making possible the processing of massive sensorial information, opening applications that were inconceivable a few years ago.

The challenge of developing effective systems should be supported by the research of computer vision techniques as part of a personal assistant based on visual localization and scene understanding, which are complementary tasks that can help each other. This will bring new research opportunities but will also raise some challenges. The main difficulties basically consist of the usage of unconventional sensors, offering interesting advantages but also important issues, and the fact that it is a person (or humanoid robot) who wears and moves the sensor, introducing a source of uncertainty because of the unpredictability of human movement, different to the smooth trajectories that mobile robots usually follow.

In this paper, we present a brief survey of some approaches we are working on aiming at expanding the possibilities with the goal of wearable vision systems development for personal assistance. Some of the ideas presented are still in progress and are framed in long term research lines. In particular, we present and discuss three results of our work: (1) Omnidirectional vision for scene layout recovery [6], which is key issue for human navigation and scene understanding. (2) The use of range cameras is studied to improve the detection of obstacle-free paths. (3) The combination of cameras with visible laser for depth capture in flexible configuration.

This work was supported by project VINEA DPI2012-31781 and FEDER.

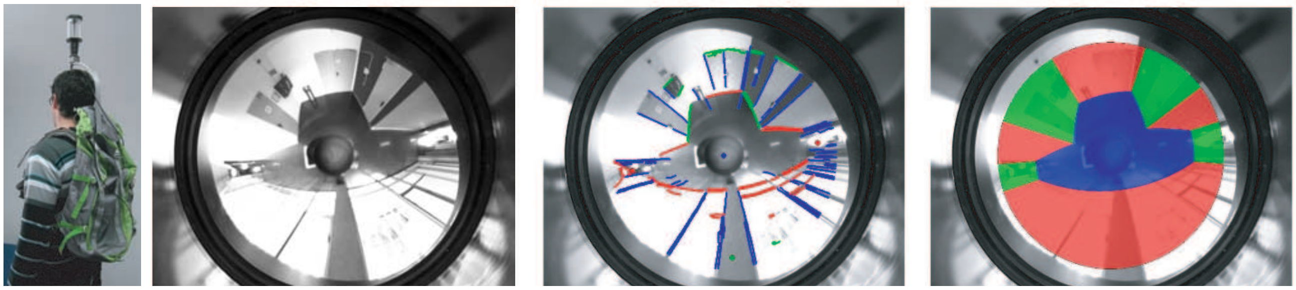

Fig. 1. From left to right: Backpack used for omnidirectional image acquisition. Example of input image fed into our algorithm. Extracted lines and estimated vanishing points classified according to the 3 main directions. Last image shows the floor and vertical walls segmentation obtained with our method.

II. Scene layout from an omnidirectional view

Conventional cameras are poor imitations of the human eye, which make them restrictive for many applications. Omnidirectional cameras are becoming increasingly popular in computer vision and robotics because their relevant properties. A wide field of view naturally enables a system to *see* in all directions, to detect obstacles or landmarks for navigation, and allows much more robust ego-motion estimation. Moreover, omnidirectional systems capture complete environment information with a reduced number of viewpoints, and therefore a reduced size of the visual memory is enough. But on the other hand, they also raise important geometric and imaging issues which are still open.

Obtaining structural distribution of a scene from an image is an easy task for anyone. However, it is not simple for visually impaired people. Knowing structural limits of the environment is the first step of any autonomous navigation system. Hence the goal is to recover the spatial layout of indoor environments from omnidirectional images assuming a Manhattan world structure. We propose a method for scene structure recovery from a single image. This method is based on line extraction for omnidirectional images, line classification, and vanishing points estimation combined with a hierarchical expansion procedure for detecting floor and wall boundaries.

The works in the literature addressing the problem of spatial layout recovery have been proposed for images acquired by conventional cameras, and most of them work under the Manhattan-world assumption [7]. In general, a conservative spatial layout is enough for defining a navigability map of the environment, and this can be a powerful tool that provides very useful information for performing tasks such as navigation or obstacle detection. Therefore, rather than a precise and detailed map of the scene, we focus on providing a conservative map in which the distribution of the different elements of the scene are classified as floor or walls.

A. Single view spatial layout recovery

In this section, we briefly describe the proposed algorithm to come up with the spatial layout of the scene from a single omnidirectional image. We start extracting lines from the image, which are then classified according to their orientation in order to carry out the estimation of vanishing points [8]. Combining this information with a set of geometrical constraints we generate hypothesis about the floor contour. From the classified lines, which are conics in this kind of images, a set of points is selected. These points are used to fit conic lines which represent plausible wall-floor boundaries. Then, a conservative four walls-room hypothesis is generated by selecting the four most voted conic lines. Finally, the initial hypothesis is expanded, according to the image data distribution, to obtain a representative hypothesis. This is carried out by replacing the initial floor contours with a set of appropriate conic lines so that successive hypotheses approximate better the actual shape of the scene. Finally, the output of the proposed algorithm is the spatial layout.

B. Matching-free sequential hypothesis propagation

This previous single image based algorithm shows good performance and it is robust to occlusions of the scene contours. However, depending on the complexity of the scene, misclassifications can occur. Each single omnidirectional image independently provides a useful hypothesis of the 3D scene structure. In order to enhance robustness and accuracy of this single image-based hypothesis, we extend this estimation with a new matching-free homography-based procedure applied to the various hypotheses obtained along the sequence images.

This approach relies on the homography computation of the floor across the views. It can be demonstrated that considering planar motion and taking into account the estimated vanishing points reduces the minimal set of lines required to compute the homography from 4 to 1. Then the exhaustive search in the number of samples, instead of a random search, is feasible in practice, without requiring any prior matching. This homography parametrization allows the design of a matching-free method for spatial layout propagation along a sequence of images. The last step of our method compares hypotheses obtained for the sequence in order to obtain an averaged hypothesis which best fits the set of floor contours.

C. Experimental results

The proposed method has been tested with a vision system composed by a hypercatadioptric camera attached to a backpack carried by the user (Fig. 1). An example of the results obtained is presented in Fig. 1. It corresponds to a long hall section with an enlargement in one of its sides. Despite the high level of luminosity coming through the window, our method is able to identify every element of the scene.

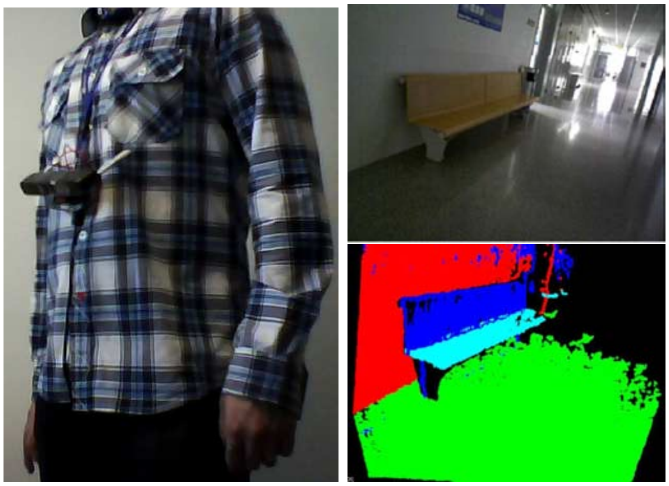

Fig. 2. Example of experimental setup and a resultant planar segmentation.

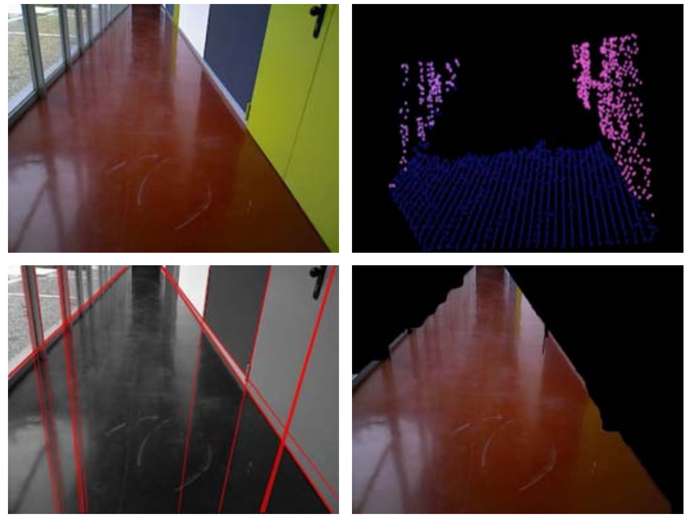

Fig. 3. Experimental results with range camera. Top-left: the original image. Top-right: The point-cloud segmentation. Below-left: Lines extracted. Below-right: The final result after the floor expansion.

III. Range cameras for navigation aid

Most mobile robots trust on range data for obstacle detection. Popular sensors based on range data are ultrasonic sensors, radar, stereo vision and infrared range sensors. These sensors are able to measure the distance from the sensor to all surrounding obstacles. Obviously, none of these sensors provides perfect results. Ultrasonic sensors have poor angular resolution. Radar works better than ultrasonic sensors but they are more complex and they may have interference problems indoors. Stereo vision needs textured environment and infrared range sensors fail with solar lighting.

In this section, we present a novel method that combines the short-range 3D point-cloud processing and long-range basic image processing for obtaining the free space on the floor. The combination of these processes yields a robust system that works in the presence of severe shadows and reflections, which are common in practice. The sensor chosen is an infrared camera combining monocular vision with range information. This camera provides a point-cloud which contains RGB and 3D information of the scene. The main steps of the proposed algorithm are described next.

A. Point-cloud analysis

The point-cloud of a scene provided by the range sensor contains a huge amount of 3D information. In order to process the point-cloud as fast as possible, we downsample the point-cloud before the segmentation step. At this point we identify the most representative planes of the scene via a RANSAC procedure. Once we have detected the planes, we classify them by analysing its normal vector and obtain the floor. An example of this classification is shown in Fig. 2. The method works properly indoors and it is robust to lighting changes. On the other hand, it has some limitations: It is susceptible to sunlight and the maximum distance is around 3.5 meters.

B. Monocular vision analysis

The infrared sensor of the range camera has several limitations as pointed out above and we use the monocular camera to improve the results. In particular, the ground plane detected from the infrared sensor is used for obtaining the whole plane of the ground in the RGB image. A significant amount of research has focused upon image segmentation problem [9] [10]. Here, we combine the depth information with the image using RGB and HSI color spaces and geometry image features. First, we perform a seeded region growing algorithm where the seed belongs to the 3D floor's plane of the depth point-cloud. In order to reduce the shadows and reflections influence, we apply a mean-shift filter to a pyramid image.

The next step is to compare the lighting and hue channel of the homogenized image with the seed. The floor is not homogeneous so the seed will have a variety of hue values. So, we compare each part of the image satisfying the first criterion, with each hue value of the seed. We carry this task out evaluating how well the pixels fit a histogram model. Pixels which satisfy this criterion will become seeds. Finally, we propose a polygon-based region growing step. We use the Probabilistic Hough Line Transform in order to extract lines and extend them till the image's border. Once we have the image segmented by polygons and the seeds distributed along the putative floor, the last step is to grow those seeds. Each polygon which has one or more seeds inside, will be labelled as floor.

C. Experimental evaluation

The hardware used is an Asus Xtion Pro live camera that hangs from the user's neck as shown in Fig. 2. The range camera will be slightly tilted towards the ground in order to detect the closest obstacles. To evaluate the performance of our algorithm, we have tested it in different kind of corridors exhibiting a wide variety of different visual characteristics and lighting conditions. Fig. 3 presents an example of a typical corridor image, notice the floor reflectivity and lighting conditions.

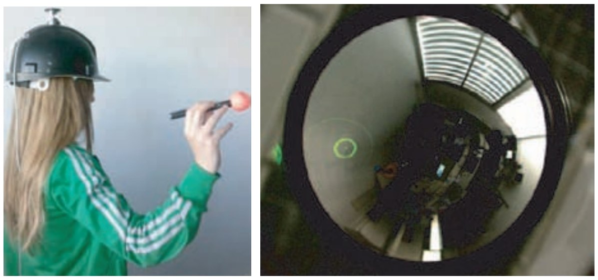

Fig. 4. Wearable vision system with laser in hand (left). Omnidirectional view with the projected conic pattern light for image processing (right).

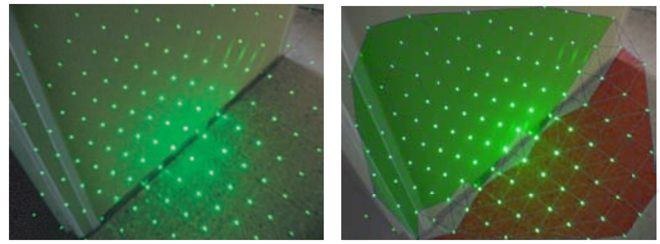

Fig. 5. Image with dotted pattern laser projected indoors (left). Resultant floor and wall homography-based segmentation (right).

IV. VISION-BASED PERCEPTION WITH VISIBLE LASER

Structured light is an active method that provides 3D information from images of the scene by projecting synthetic features with a low-cost light emitter. In the previous section, infrared system was considered, in order to develop a more flexible system, we study the use of visible laser in the context of wearable personal assistance systems.

A. Conic laser pattern

Traditionally, structured light methods consider a rigid configuration, where the position and orientation of the light emitter with respect to the camera are known beforehand. We have developed a new omnidirectional flexible structured light system which overcomes the rigidness of the traditional structured light systems. We propose the use of an omnidirectional camera combined with a conic pattern light emitter in hand (Fig. 4). Since the light emitter is visible in the omnidirectional image, the computation of its location is possible. With this information and the projected conic in the omnidirectional image, we are able to compute the 3D conic reconstruction.

Our approach [11] combines the omnidirectional image and a virtual image generated from the light emitter. With our method, we obtain the depth and orientation of the scene surface where the conic pattern is projected. The long term application of this structured light system in flexible configuration is a wearable omnicamera with a laser in hand for visual impaired personal guidance.

B. Dotted laser pattern

Alternatively, we study the use of dotted laser patterns avoiding the need of system calibration and 3D reconstruction. The setup consists of a green laser in hand, projecting light points following a squared pattern, and a camera mounted on the person. We have designed a homography-based approach that uses the image of the dotted pattern projected in the scene to obtain plane segmentation. This is an efficient procedure that provides fast and robust obstacles detection without constraining to a fixed configuration between camera-laser. Fig. 5 shows an example of the results obtained. The method works correctly with changes in illumination conditions and shadows. Therefore, we consider this setup as an interesting choice in the framework of personal guidance systems.

V. DISCUSSION AND CONCLUSION

Several unconventional vision-based systems have been studied for the particular task of scene analysis for human navigation guidance. Omnidirectional views provide a wide field of view and we presented a method for spatial layout recovery from a single omnidirectional image based on conic lines classification and its propagation along a sequence of images skipping the prone to error and computational expensive matching algorithms. Range infrared cameras are a powerful tool providing 3D information at close range, robust to lighting changes, glows and reflections. We combine this information with vision to extend the obtained information along all the visible scene. Finally, we study the use of visible laser in the framework of obstacle detection.

These are just some examples of the possibilities of wearable vision systems for personal assistance. These systems can be easily carried by a person and are able to guide a visually impaired people by, for instance, audio instructions. However, none of them will guarantee total absence of failures, and many issues will emerge, opening novel research lines in order to address this challenge.

REFERENCES

[1] D. Dakopoulos and N. Bourbakis, “Wearable obstacle avoidance electronic travel aids for blind: A survey,” *IEEE Trans. on Systems, Man, and Cybernetics, Part C*, vol. 40, no. 1, pp. 25–35, 2010.

[2] M. Baglietto, A. Sgorbissa, D. Verda, and R. Zaccaria, “Human navigation and mapping with a 6 DOF IMU and a laser scanner,” *Robotics and Autonomous Systems*, vol. 59, no. 12, pp. 1060 – 1069, 2011.

[3] J. M. Saez Martinez and F. Escolano Ruiz, “Stereo-based Aerial Obstacle Detection for the Visually Impaired,” in *Workshop on Computer Vision Applications for the Visually Impaired, in ECCV*, Marseille, 2008.

[4] R. Öktem, E. Aydin, and N. Cagiltay, “An indoor navigation aid designed for visually impaired people,” *IECON 2008*, pp. 2982–2987, 2008.

[5] O. Koch and S. Teller, “Body-relative navigation guidance using uncalibrated cameras,” in *ICCV*, 2009, pp. 1242–1249.

[6] J. Omedes, G. López-Nicolás, and J. J. Guerrero, “Omnidirectional vision for indoor spatial layout recovery,” in *Frontiers of Intelligent Autonomous Systems*. Springer, 2013, pp. 95–104.

[7] D. Lee, M. Hebert, and T. Kanade, “Geometric reasoning for single image structure recovery,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2009.

[8] J. Bermudez-Cameo, L. Puig, and J. J. Guerrero, “Hypercatadioptric line images for 3D orientation and image rectification,” *Robotics and Autonomous Systems*, vol. 60, pp. 755–768, 2012.

[9] Y. Li and S. Birchfield, “Image-based segmentation of indoor corridor floors for a mobile robot,” *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 837–843, 2010.

[10] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. Bradski, “Self-supervised monocular road detection in desert terrain,” *Proceedings of Robotics: Science and Systems*, August 2006.

[11] C. Paniagua, L. Puig, and J. J. Guerrero, “Omnidirectional structured light in flexible configuration,” *Technical Report, I3A*, 2013.