# Comparison between low-cost passive and active vision for obstacle depth

**Vito F. Chiarella[1], Thiago Rateke[2,3], Karla A. Justen[3], Antonio C. Sobieranski[4], Sylvio L. M. Neto[3], Eros Comunello[5], Aldo von Wangenheim[2,3]**

[1]Graduate Program in Materials Science and Engineering (PGMAT)
Federal University of Santa Catarina (UFSC)
Florianópolis, SC, Brazil

[2]Graduate Program in Computer Science (PPGCC)
Federal University of Santa Catarina (UFSC)
Department of Informatics and Statistics
Florianópolis, SC, Brazil

[3]Image Processing and Computer Graphics Lab (LAPIX)
National Institute for Digital Convergence (INCoD)
Florianópolis, SC, Brazil

[4]Department of Computing
Federal University of Santa Catarina (UFSC)
Ararangua, SC, Brazil

[5]Master in Applied Computing (MCA)
University of Itajaí Valley (UNIVALI)
São José, SC, Brazil

{thiago,sylvio}@incod.ufsc.br, {justen.karla,vitochiarella}@gmail.com

{aldo.vw,a.sobieranski}@ufsc.br, eros@univali.br

***Abstract.*** *Obstacle detection is a key issue in many current applications, especially in applications that have been increasingly highlighted such as: advanced driver assistance systems (ADAS), simultaneous localization and mapping (SLAM) and autonomous navigation system. This can be achieved by active and passive acquisition vision systems, for example: laser and cameras respectively. In this paper we present a comparison between low-cost active and passive devices, more specifically LIDAR and two cameras. To this comparison a disparity map is created by stereo correspondence through two images and a point cloud map created by LIDAR data values (distances measures). The obtained results shown that passive vision can be as good as or even better than active vision in low cost scenarios.*

***Resumo.*** *A detecção de obstáculos é uma questão fundamental em muitos aplicativos atuais, especialmente em aplicativos que têm tido cada vez mais destaque, como: sistemas avançados de assistência ao motorista (ADAS), localização e mapeamento simultâneo (SLAM) e sistema de navegação autônoma. Isso pode ser alcançado por sistemas de visão ativa e passiva, por exemplo:*

*laser e câmeras, respectivamente. Neste artigo, apresentamos uma comparação entre dispositivos ativos e passivos de baixo custo, mais especificamente o LIDAR e duas câmeras. Para esta comparação, um mapa de disparidade é criado por correspondência estéreo através de duas imagens e também um mapa de nuvens de pontos criado pelos valores de dados do LIDAR (medidas de distâncias). Os resultados obtidos mostraram que a visão passiva pode ser tão boa quanto ou até melhor que a visão ativa em cenários de baixo custo.*

## 1. Introduction

Object detection together with the acquisition of depth information for each detected object is an important issue for robotic or vehicular navigation, providing information about potential obstacles and their location in relation to the vehicle or robot. A pair of cameras can be used to estimate the scene and the obstacles depth through Stereo Vision methods. Stereo Vision methods use the information captured from, at least, two cameras [Zureiki et al. 2007]. Current autonomous navigation approaches, however, do not employ only a camera, but pursue the integration of various types of sensors [Urmson et al. 2008], [Urmson 2014b], [Urmson 2014a] and [Fernandes et al. 2014]. One of the most widespread of these sensors is the LIDAR (*Light Detection and Ranging*), a laser that is used for remote sensing of properties of the reflected light in order to measure distances between the sensor and target objects such as obstacles, other vehicles and road irregularities [LIDAR 2015].

Most studies consider LIDAR as Class1 laser source, which is the lowest impact laser category, employing 905nm wave lengths and a low-powered laser source. According to material presented by [Commission 2001] and [STANDARD 2005] individual sources of Class1 lasers do not pose an immediate risk to the retina, as long as a direct contact with the human eye does not persist for a longer period. However, in [Commission 2001] there is a table with lasers categorization and their possible risks, when prolonged or repeated exposure occurs on each level. Lasers with a wavelength between 780nm and 1400nm, which is in the LIDAR coverage range, have been shown to cause cataracts and even burn the retina through repeated exposure in a small time window. If autonomous vehicles may become a reality in the day-to-day, a large number of vehicles will be performing LIDAR scans at the same location while provoking "lasersmog" situations such as in: traffic jams, intersections, reflections, and constant exposure of pedestrians on sidewalks during rush hours. Passive vision approaches can provide a less hazardous alternative to this future scenario.

Good quality, high performance cameras for stereo vision can be expensive. A high performance LIDAR sensor, on the other side, has also a much higher cost. In order o investigate if it is possible to reliably perform adequate object depth detection employing low-cost sensors, this paper presents a comparison between low-cost active and passive sensors to perform the obstacle detection task. Both acquisition devices are presented, a depth map is estimated from stereo vision methods, and a point cloud map is created from the active LIDAR scans (Figure 1).

The remainder of this paper is organized as follows: Section 2 we present and discuss the use of LIDAR and Stereo Vision for obstacle detection in the literature. Section 3 we describe briefly the materials and methods. In Section 4 we present the experiments,

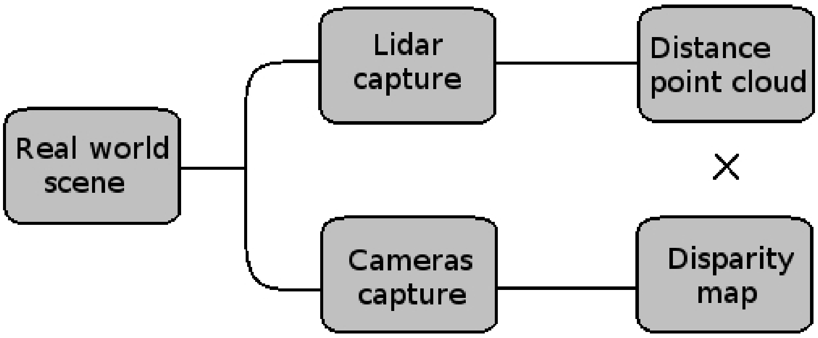


**Figure 1. Steps.**

the results and comparison. The paper is concluded in Section 5.

## 2. The use of LIDAR and Stereo Vision systems

Recent researches have been employing LIDAR to perform the obstacles detection task in autonomous navigation ([Zhao et al. 2014], [Young and Simic 2015] and [Weichselbaum et al. 2013]). In [Zhao et al. 2014] and in [Young and Simic 2015] authors present approaches which perform data fusion between LIDAR and camera data. In [Weichselbaum et al. 2013], a work about autonomous train navigation is presented, which deals with frontal images. In this work a lot of different sensors are used, such as: lasers scanners, single and stereo cameras working in the spectral ranges of visible light and infra-red, radar and ultra sonic sensors. The evaluation and fusion of their signals contributes to a useful recognition performance.

Many researches are also underway for passive vision obstacle detection (e.g.: [Wang et al. 2014], [Kim et al. 2015] and [Yoo et al. 2016]). Some works focuses on specific objects, such as pedestrian detection ([Keller et al. 2011] and [Llorca et al. 2012]) or vehicle detection ([Milanés et al. 2012] and [Chong et al. 2013]).

The majority of research works employ stereo vision, like [Wang et al. 2014] that proposes a disparity calculation algorithm based on multi-pass aggregation and local optimisation, enabling to detect an obstacle and a free space to navigate. In [Kim et al. 2015] the authors present a hierarchical census transform (HCT)-based stereo matching method, and proposes a real-time rear obstacle detection system using fish-eye stereo cameras. A real-time obstacle detection is presented in [Yoo et al. 2016], using three features such as: disparity, super-pixel segmentation and pixel-wise gradient.

A survey about pedestrian detection is presented in [Llorca et al. 2012] as well as a Region of Interest extraction approach through stereo vision. In [Keller et al. 2011] the stereo density information is used both to generate Regions of Interest and to do the pedestrians detection. The work from [Milanés et al. 2012] is based on a stereo vision system aiming the vehicle detection.

## 3. The Sensors

In this Section we present our experimental setup as well as steps aspects about Stereo Vision and 3D depth estimation using LIDAR. For the active vision process we used a low cost LIDAR sensor together with an Arduino. For passive vision we used two cameras to make a stereo pair. The sensors captured the same scene at the exact same moment.

### 3.1. LIDAR

LIDAR (Light Detection And Ranging) is a technology that measures distance by illuminating a target with a laser light. Ground-based LIDAR, which records "street scenes", has been around for several years [LIDAR 2015]. For the active vision process our work made use of the LIDAR Lite V2 (Figure 2), a compact high performance optical distance measurement sensor. LIDAR Lite V2 has a range of up to 40 meters and an accuracy of approximately 0.025m.

**Figure 2. LIDAR Lite V2.**

We used two Micro Servo SG90 to makes LIDAR rotations, which can rotate approximately 180 degrees, 90 in each direction One servo makes horizontal rotation and another makes vertical rotation, this enable LIDAR to measures distances in all environment ahead (Figure 3).

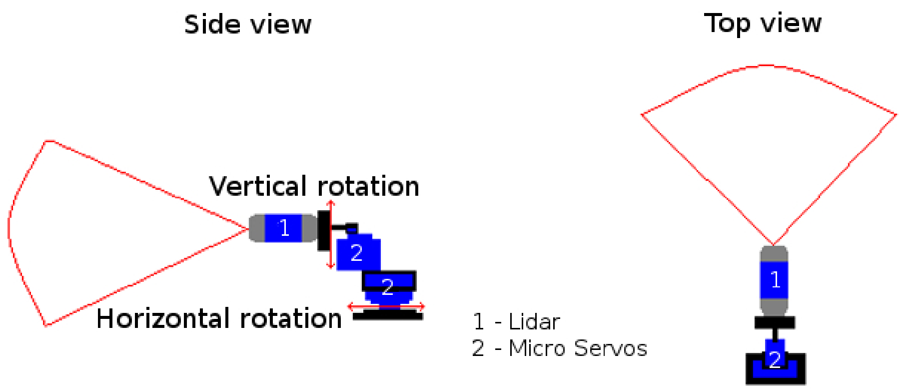


**Figure 3. LIDAR rotation.**

An Arduino Uno is used to receive data values from LIDAR and to control the Micro Servo. This values enable us to create a point cloud visual image in gray scale, similar to a disparity map, where darker gray value is farther (Figure 4). This is possible because Arduino wrights the data values, distances that LIDAR measures, in a serial file. Therefore, by reading this serial file, we can create a point cloud map.

 ISSN 2447-7028

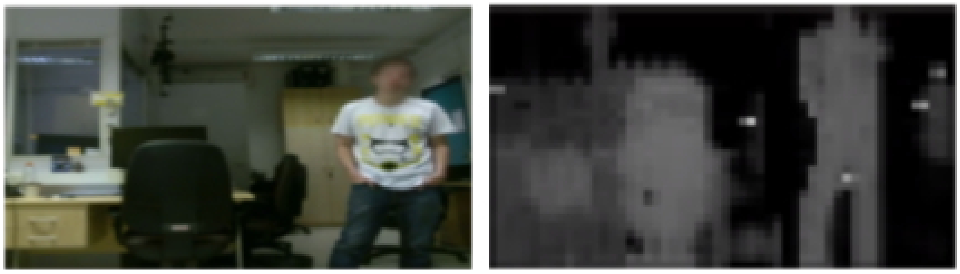

**Figure 4. Original image (left). LIDAR Point cloud (right).**

### 3.2. Stereo Vision

For the passive vision process our work used two HP Webcam HD-4110 (Figure 5) to build our stereo pair structure. The stereo vision is basically composed of two capture sources, cameras and lenses, parallel or not, and spaced apart, which acquire two-dimensional images, right and left, containing slightly different content due to the positioning of the cameras and the light incidence. An algorithm makes the correlation between these images, making possible to calculate the disparity map. [Bleyer 2013].

**Figure 5. HP Webcam HD-4110.**

Preliminary, a calibration step is required to estimate the different parameters of a stereo equipment and spatial relations between the two cameras. This knowledge allows the estimation of the 3D coordinates of a point and its projections in the two images through a simple triangulation [Zureiki et al. 2007].

In cameras calibration, we get the data, as the points projection in two dimensions of the captured scene, with this data we obtain the cameras internal values (focal point and center of the image), also the camera external values (rotation and translation). These data will be used later for rectification and disparities calculations. To do the calibration step, we used an object with a known geometric pattern like a chess board that allows us to easily identify the points in the scene. Next, the images rectifying step is performed by placing the two images on the same plane.

With the images rectifying, left and right images plans are placed in a common plan. The corresponding pixels now have the same $y$ coordinate. Then, to find the corresponding pixel we only need to search along the horizontal axis, known as “scan line” [Hartley and Zisserman 2004], [Bleyer 2013]. These correspondences will be used in stereo matching and correlation between pixels, building the disparity map. The disparity is the difference that the same pixel has between two images, that is, how far this pixel is shifted between the images (Equation 1).

$$D = x_l - x_r \tag{1}$$

Where $xl$ is the specific pixel coordinate in left image, $xr$ is the coordinate of the same specific pixel in the right image and $D$ is the disparity value between these points.

The disparity of each pixel is coded by intensity values, where high intensity values represent high disparities and low intensity values represent lower disparities (Figure 6). The disparity of each pixel is inversely proportional to its depth in the scene, so disparity is commonly used as a synonym for depth [Bleyer 2013].

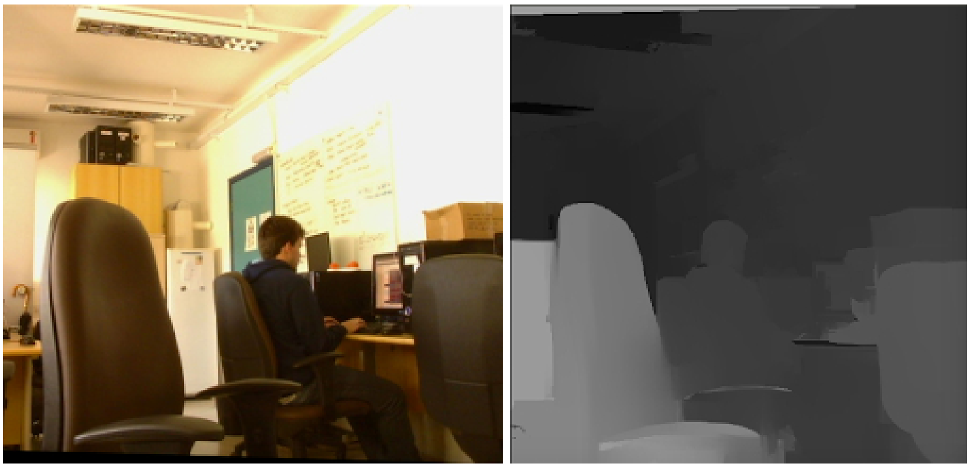

**Figure 6. Original image (left). Disparity Map (right)**

## 4. Comparisons

We put LIDAR in the center of our structure, between the cameras (Figure 7), allowing both LIDAR and cameras to capture the scene at the same time and also enabling the comparison from obtained results. Our comparisons are qualitatively evaluated, based on visual analysis.

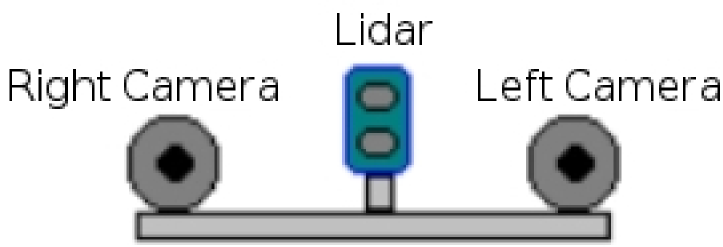


**Figure 7. Structure**

We perform our tests and evaluations in an indoor environment using the aforementioned sensors. We evaluated the capture of different objects at different distances on each sensor. Our focus here is to compare these two sensors with a frontal image, and analyze if there are differences between active and passive sensors.

### 4.1. LIDAR X Stereo Vision

We can compare the results from a disparity map, by passive vision, with a distance point cloud, by active vision, in Figure 8. Where each line shows a person at different distances from the sensors. In (a) are the original images, (b) shows disparity map by passive vision, cameras capture and stereo correspondence, and finally in (c) we present a point cloud map by LIDAR distance measures.

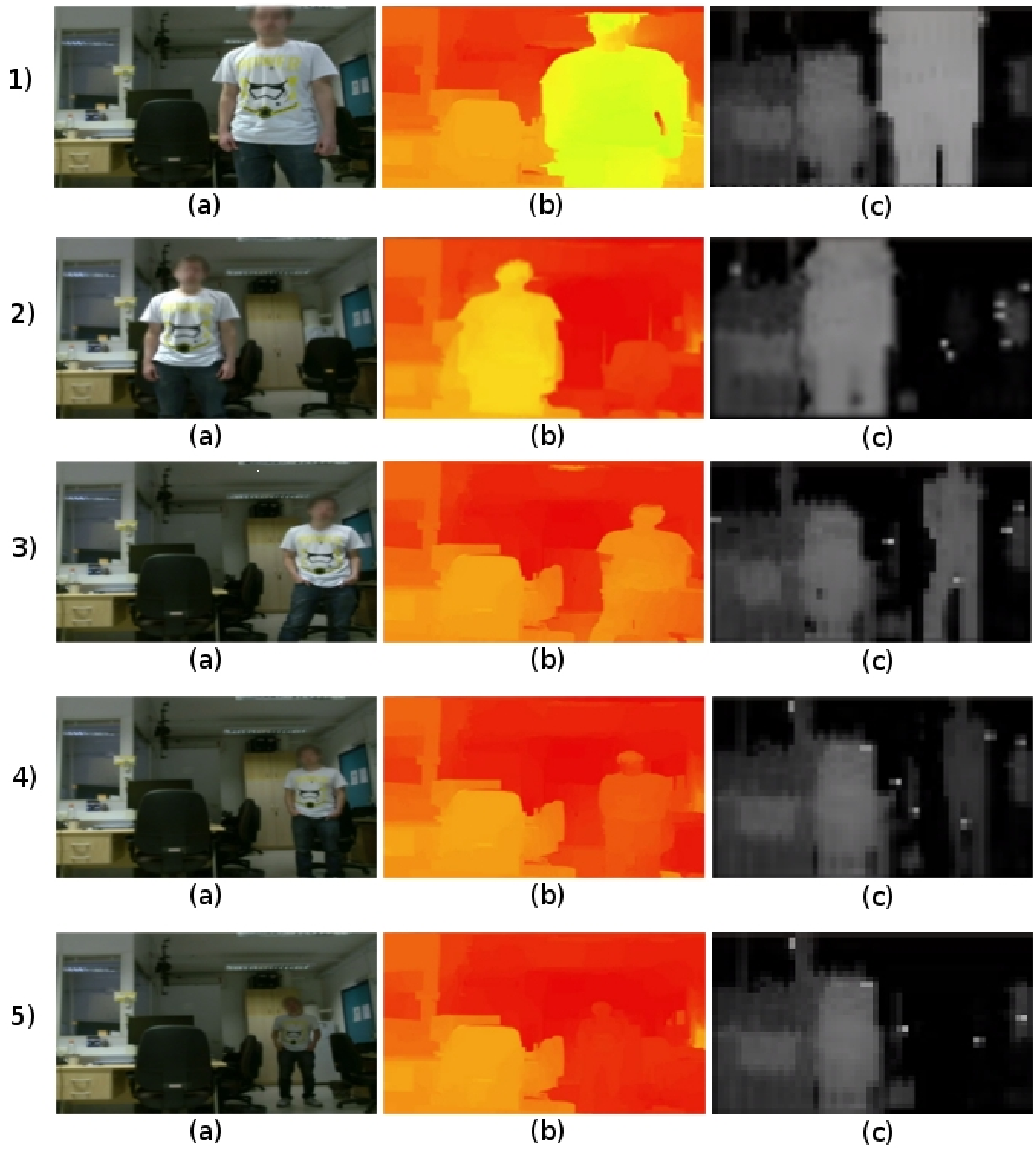


**Figure 8. Results. (a) original image, (b) disparity map (Passive Vision), (c) point cloud map (Active Vision).**

It is important to say that the head "cut" in the point cloud map is not a failure of the LIDAR, but related to scanning and more specifically, how much each Micro-Servo moved (in this case the Servo responsible for vertical rotation), but that does not disturb the results and still allows us to make a series of comparisons in detail.

Line 1 shows the person considerably nearest to the sensors (Figure 8 (a)). And the disparity map has a good result, person has a high intensity, meaning a high disparity value and a low depth from the sensor, also the chair on the left side of the image is further than the person (Figure 8 (b)).

Finally, the point cloud map, created based on the LIDAR measure distances data. This point cloud map show, like the disparity map, the person more closer to the sensors with light grey color and farther objects with dark grey values (Figure 8 (c)). With these two maps we can see the depth based in intensity values (person is more closer to the sensors than the table and chair in left of the image).

In Figure 9 a difference between these two sensors is highlighted. In this case, the person is further away. With Disparity Map, from camera capture, it is still possible to see the difference in intensity level between background and the person indicating an object. But on the point cloud map, generated by LIDAR, we can't see any difference.

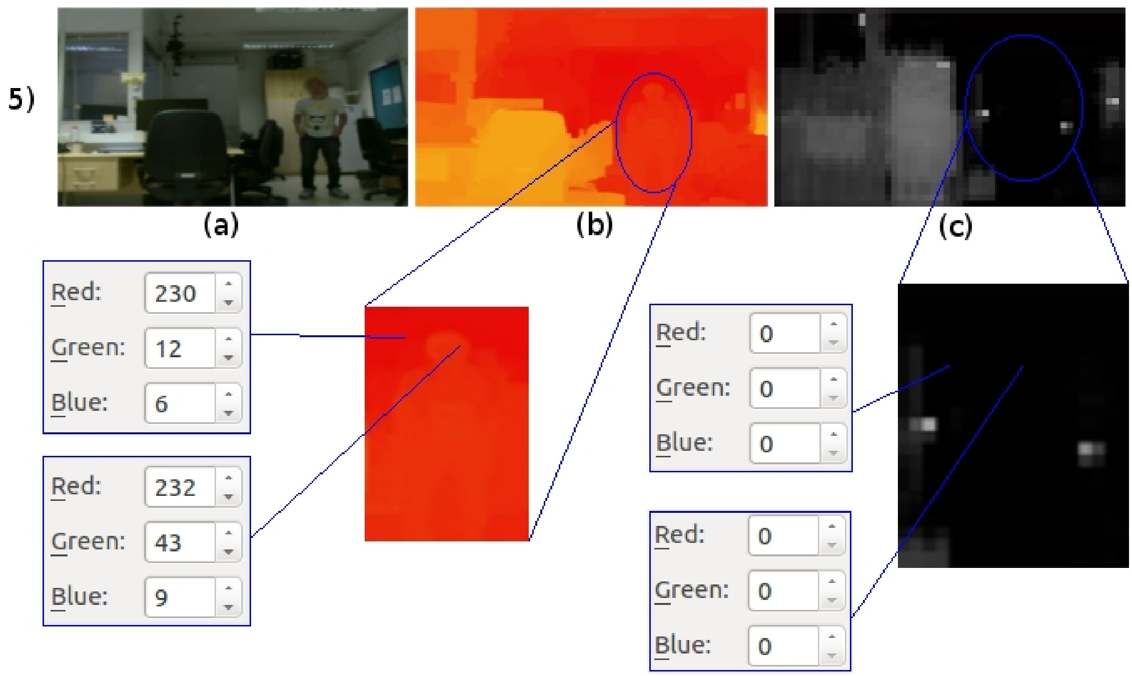


**Figure 9. Differences at greater depths.**

In Figure 10 we present the results when there are objects with open spaces and also with transparent glass objects. Line 1 contains objects closest to the sensors, line 2 with intermediate distance and line 3 has the most distant objects.

The LIDAR did not have a good result, showing a point cloud map that does not resemble to the visualized object, whereas the Stereo Vision results was better and it is possible to identify the object. In line 3, even the Stereo Vision did not have a good result with the glass vase.

In Figure 11 we present the detection of a thin object. Again, the Stereo Vision had a better result. Some defects are visible in the Disparity Maps results in Figures 10

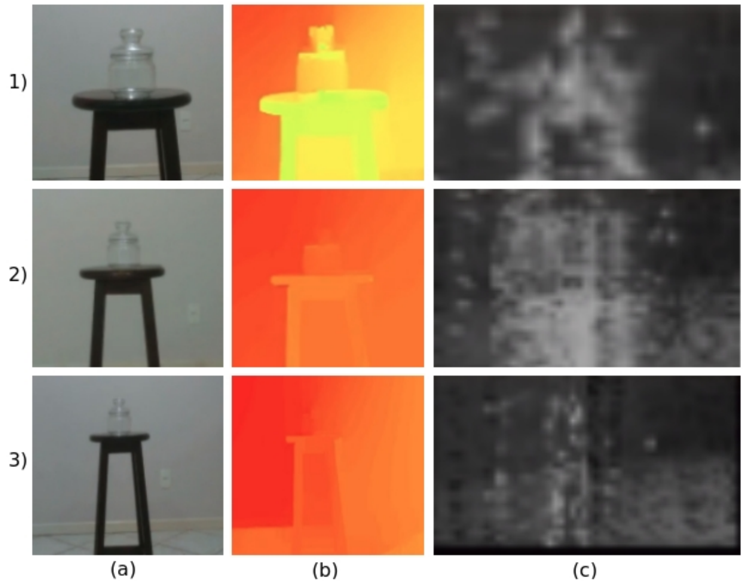


**Figure 10. Difference in object with open space and in transparent glass.**

and 11, which has open spaces inside the obstacles. Also, the algorithm trying to correct occlusion problems, can consider areas where there is a space inside the object, as part of the object. In any case, this "defect" is mitigated if we consider that we want to avoid the whole object, no matter the internal open spaces.

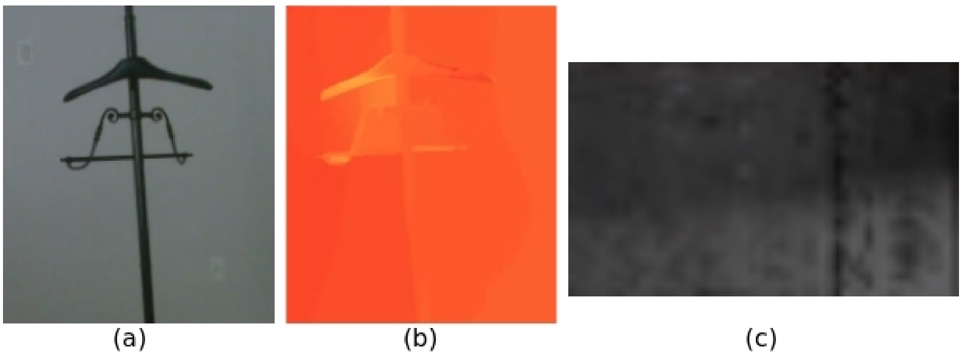


**Figure 11. Difference with thin object.**

## 5. Conclusion and Discussions

Based on the results we obtained, it is possible to show that in a low-cost scenario, passive vision (eg.: common cameras with computer vision methods and digital image processing) is as good as the active vision (eg.: scanning with lasers) or even better.

However, the experiments were done indoors and under controlled lighting. In darker environments, such as outdoors and at night, LIDAR will have an advantage over cameras. When there are objects with colors very similar to the background (walls), the stereo vision will have difficulty identifying the corresponding pixels, due the lack of color and texture differences.

When there is a lot of movement, the quality of stereo vision does not works well because these low-cost cameras do the image capture process with Rolling Shutter type lenses. Frames sequence capture by a Rolling Shutter lens can cause distortions such as horizontal inclination of the scene and a "gelatinous" effect on objects.

To avoid this kind of problems it is necessary to employ cameras with Global Shutter lenses that allows the camera to capture all the movement at the same time, avoiding unwanted effects and distortions. Even good, expensive, performance-enhancing and Global Shutter lenses cost much less than high-performance LIDAR sensors. Recently, we have seen LIDAR sensors with a cost closer to those of good Global Shutter cameras. But even so, we encountered the eye health situation, where despite LIDAR being Class1, the least aggressive, there is the possibility that, in traffic situations when "lasersmog" is generated, it will still pose a risk to the retina of the people around.

Either way, these low cost sensors show that they can be useful in more controlled scenarios, with lower movement, as a factory environment, or even for Simultaneous Localization and Mapping (SLAM), where the vehicle can navigate slowly through the environment performing scenario mapping, usually with a static background.

## Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. CAPES (Brazilian Federal Agency for Support and Evaluation of Graduate Education). It was also supported by the Brazilian National Institute for Digital Convergence (INCoD), a research unit of the Brazilian National Institutes for Science and Technology Program (INCT) of the Brazilian National Council for Science and Technology (CNPq).

## Competing interests

The authors have declared that no competing interests exist.

## References

Bleyer, M. (2013). *VU Stereo Vision*. Karlsplatz, Vienna, Austria.

Chong, Y., Chen, W., Li, Z., Lam, W. H., Zheng, C., and Li, Q. (2013). Integrated real-time vision-based preceding vehicle detection in urban roads. *Neurocomputing*, 116:144 – 149. Advanced Theory and Methodology in Intelligent ComputingSelected Papers from the Seventh International Conference on Intelligent Computing (ICIC 2011).

Commission, I. E. (2001). Safety of laser products. part 1: Equipment classification, requirements and user's guide.

Fernandes, L. C., Souza, J. R., Pessin, G., Shinzato, P. Y., Sales, D., Mendes, C., Prado, M., Klaser, R., Magalhães, A. C., Hata, A., Pigatto, D., Branco, K. C., Jr., V. G., Osorio, F. S., and Wolf, D. F. (2014). Carina intelligent robotic car: Architectural design and applications. *Journal of Systems Architecture*, 60(4):372 – 392.

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Keller, C. G., Enzweiler, M., Rohrbach, M., Llorca, D. F., Schnorr, C., and Gavrila, D. M. (2011). The benefits of dense stereo for pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1096–1106.

Kim, D., Choi, J., Yoo, H., Yang, U., and Sohn, K. (2015). Rear obstacle detection system with fisheye stereo camera using {HCT}. *Expert Systems with Applications*, 42(17–18):6295 – 6305.

LIDAR (2015). *LIDAR - http://www.lidar-uk.com/*. UK.

Llorca, D., Sotelo, M., Hellín, A., Orellana, A., Gavilán, M., Daza, I., and Lorente, A. (2012). Stereo regions-of-interest selection for pedestrian protection: A survey. *Transportation Research Part C: Emerging Technologies*, 25:226 – 237.

Milanés, V., Llorca, D. F., Villagrá, J., Pérez, J., Fernández, C., Parra, I., González, C., and Sotelo, M. A. (2012). Intelligent automatic overtaking system using vision for vehicle detection. *Expert Systems with Applications*, 39(3):3362 – 3373.

STANDARD, A. N. (2005). *American National Standard for Safe use of Lasers Outdoors*. Orlando, FL.

Urmson, C. (2014a). Just press go: designing a self-driving vehicle.

Urmson, C. (2014b). The latest chapter for the self-driving car: mastering city street driving.

Urmson, C., Anhalt, J., Bae, H., Bagnell, J. A. D., Baker, C. R., Bittner, R. E., Brown, T., Clark, M. N., Darms, M., Demitrish, D., Dolan, J. M., Duggins, D., Ferguson , D., Galatali, T., Geyer, C. M., Gittleman, M., Harbaugh, S., Hebert, M., Howard, T., Kolski, S., Likhachev , M., Litkouhi, B., Kelly , A., McNaughton, M., Miller, N., Nickolaou, J., Peterson, K., Pilnick, B., Rajkumar, R., Rybski, P., Sadekar, V., Salesky, B., Seo, Y.-W., Singh, S., Snider, J. M., Struble, J. C., Stentz, A. T., Taylor , M., Whittaker, W. R. L., Wolkowicki, Z., Zhang, W., and Ziglar, J. (2008). Autonomous driving in urban environments: Boss and the urban challenge. *Journal of Field Robotics Special Issue on the 2007 DARPA Urban Challenge, Part I*, 25(8):425–466.

Wang, Y., Gao, Y., Achim, A., and Dahnoun, N. (2014). Robust obstacle detection based on a novel disparity calculation method and g-disparity. *Computer Vision and Image Understanding*, 123:23 – 40.

Weichselbaum, J., Zinner, C., Gebauer, O., and Pree, W. (2013). Accurate 3d-vision-based obstacle detection for an autonomous train. *Computers in Industry*, 64(9):1209 – 1220. Special Issue: 3D Imaging in Industry.

Yoo, H., Son, J., Ham, B., and Sohn, K. (2016). Real-time rear obstacle detection using reliable disparity for driver assistance. *Expert Systems with Applications*, 56:186 – 196.

Young, J. and Simic, M. (2015). {LIDAR} and monocular based overhanging obstacle detection. *Procedia Computer Science*, 60:1423 – 1432. Knowledge-Based and Intelligent Information & Engineering Systems 19th Annual Conference, KES-2015, Singapore, September 2015 Proceedings.

Zhao, G., Xiao, X., Yuan, J., and Ng, G. W. (2014). Fusion of 3d-lidar and camera data for scene parsing. *Journal of Visual Communication and Image Representation*, 25(1):165 – 183. Visual Understanding and Applications with RGB-D Cameras.

Zureiki, A., Devy, M., and Chatila, R. (2007). Stereo matching using reduced-graph cuts. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 1, pages I – 237–I – 240.