

Edge Based 3D Object Localization for Fast and Precise Industrial Machine Vision Applications

O. Demuynck, C. Pacheco Cedeño, A. L. Moore

Abstract—Machine Vision for industrial applications requires simple-to-use and robust solutions. Usually industrial machine vision solutions are reduced to a simple 2D application, for which algorithm robustness has already been fully demonstrated, and where a key part of all those visual inspection systems is the efficient illumination control systems that ensure the repeatability of captured image intensity. Nevertheless these projects sometimes require the 3D information for applications such as pick and place robotics, and this field of applications is not yet covered by most of the commercial industrial vision brands. We present in this paper a new fast, precise and robust 3D localization method based on object edge. This process offers an infinite number of possible cases of industrial applications when 3D localization is required, but could also be employed for classification since 3D localized object conserves its almost whole extracted edge. For example, in case of the food industry, fruits are selected depending on their size and could require a 3D localization process because of the random distribution on a conveyor belt. The presented industrial machine vision system is easy and fast to implement. As detailed in the final part of this paper, the computed statistics demonstrate the robustness and precision of this system, which results an inexpensive solution for industrial applications.

Keywords—3D edge localization, 3D edge classification, Industrial Machine Vision.

I. INTRODUCTION

DURING the last few decades, Machine Vision has progressively increased its quantity of industrial applications, by generalizing simpler previous solutions and developing new tools specially designed for the industrial market. Indeed, complex algorithms have demonstrated their robustness and emerged from the research field thanks to computer processor speed improvements. The beginning of such application was based on simple grayscale images treatments like binarized images being compared for defect

Manuscript received September 17, 2009; Revised version received September 17, 2009. This work was supported in part by the IMADE (Instituto Madrileño de Desarrollo), the FEDER (Fondo Europeo de Desarrollo Regional), and the PromoMadrid Desarrollo Internacional de Madrid.

O. Demuynck is with IDC Tecnología - Factory Automation, S.L. Company, TecnoAlcalá. Calle Punto Mobi, 10. Alcalá de Henares (Madrid). SPAIN (phone: +34 91 879 60 46; fax: +34 91 879 60 47; e-mail: odemuynck@idc.es).

C. Pacheco Cedeño is with the Electrical Department, UAH, Escuela Politécnica Superior, Campus Universitario. Ctra. Madrid-Barcelona, Km. 33,600. Alcalá de Henares (Madrid). SPAIN (cpachecocu@gmail.com).

A.L. Moore is with the MIT (Massachusetts Institute of technology), 77 Massachusetts Ave. Cambridge (Massachusetts), USA (almoore@mit.edu).

detection, and generally involved an analog camera, a frame grabber and a computer for its implementation. Nowadays, “smart cameras” include all of those devices and communicate information directly with a Programmable Logic Controller (PLC) or a robot through discrete output and/or a serial port. Industrial applications, with such hardware improvements are also much more powerful than the previously sited ones. Industrial machine vision processes are not only efficient enough to robustly check the printing, the measurements, and any other inspection or quality control parameters of a very fast production line, but also some higher level processes like pick and place robotic arm assistance. This last application is usually also easily solved with a simple 2D camera system. Nevertheless it could require 3D localization when the presentation of multiple objects is randomly distributed. For such 3D information extraction, most of the current literature, such as [1] and [2], converges on the use of a stereoscopic camera systems for a suitable solution. Precision improvement by using multiple cameras [3] or process acceleration methods have also been introduced. Processing time acceleration is usually associated with a precision loss as in [4] where the term “3D reconstruction” is actually called “3D perception”. In order to achieve a global 3D reconstruction of a scene, it is also possible to compute the disparity map of stereoscopic images [5] that offer a fast global scene understanding but still return a poor localization precision for industrial applications. Others studies improve the global execution speed by centering on the matching of line segments to partially accelerate the stereoscopic point-correspondences task [6], or present 3D reconstruction based on a multiple perspectives view where direct correspondences are arranged [7]. Precision in 3D reconstruction is also intensively studied and usually focus the topic improvement on stereoscopic points correspondences [8] [9] where most of error are introduced. Epipolar geometry for direct point correspondence is also refined in [10], of course considerably increasing the global processing time. Finally, in the industrial field of applications such a stereoscopic system for fast and precise localization could be very useful when 3D localization is required. The precise and fast compromise is robustly achieved by the stereoscopic system developed in this work. Many techniques can be employed for 3D localization. Epipolar geometry, as previously sited, is commonly used for precise stereoscopic image point correspondence that finally ensures a precise 3D reconstruction of the simultaneously observed point and is mandatory when stereoscopic scenes

present strong disparity between two points of view as in [11]. This process basically requires a point-extracting step in one of the stereoscopic images, and a fine texture correspondence along the computed epipolar line in the other image. Thus, one point observed simultaneously in two images can be localized precisely in 3D. This process for one-point localization is very fast but usually it is interesting to reconstruct the whole shape of an object using edge extraction with a Canny filter to allow its recognition or classification. In this case, time processing can increase drastically and make the treatment not viable in real time. The combination of localization and classification treatment has already been introduced in some studies [12]. Even if in the cited case it is done by 2D treatment completed by a 3D geometrical interpretation, this final classification, or identification is commonly needed in vision artificial process once an object localized. In the cited case, the topic is the access control of vehicles, where a detected even must be localized precisely to be geometrically interpreted as a known vehicle by its extremities points. This method could be perfectly generalized by our work since the localization offer a cloud of 3D points that would robustly allow its identification among a list of candidates.

In this paper we present a fast algorithm to process the shape 3D localization of an object in a previously defined industrial context. We will depict in the first section the context and the designed industrial artificial vision system. The second section is dedicated to explaining the treatment and program parameters. Finally, the last section shows the results obtained, where we observe the 3D edge reconstruction of an object and compare the precision obtained with the simple center-of-mass 3D object localization. As it is done in [13], we present an evaluation of accuracy and repeatability of 3D measurement in industrial situation.

Finally, results presented in some images clearly demonstrate the possible use of this treatment for an additional classification step.

II. INDUSTRIAL CONTEXT AND VISION SYSTEM DESCRIPTIONS

Industrial contexts are usually much more favorable compared to most artificial vision contexts. Indeed, outdoor scenes can switch from bright to almost dark conditions due to partly cloudy weather that can saturate the imaging sensor. In a fixed position camera, the background scene can also change quickly because of external conditions such as wind, rain, etc. Finally a lot of other parameters would be affected by uncontrolled events. Thus, in order to ensure the robustness of the system, we first have to control the global illumination of the system. Basically, in the indoor situation, it is important to conserve a constant global intensity of light. Such problems of illumination are usually solved by isolating the area of observation and the whole visual inspection system. The main illumination device we use to precisely get the shape of the observed object is a backlight system. These systems are composed of a board of equally distributed LEDs placed in front of a light diffusing material. This way, the cameras are

not saturated by the illumination system and will observe a precise, sharpened and contrasted edge of any object obscuring the sensor when passing between the cameras and the backlight system.

The stereoscopic camera system is composed of two 1280x1024 CMOS cameras, and, to correspond with the objects' range of size in this study (see figure 1), each camera was equipped with a 6 millimeter focal distance lens to allow a wide field of view. The calibration is performed using a typical chessboard pattern and the intrinsic and extrinsic camera parameters are obtained as depicted in [14] and [15], so the stereoscopic system is ready to correct the lens distortions and compute triangulation for 3D reconstruction.

The list of objects we used for this study is shown in figure 1. A trophy cup, a paddle, a dark glass bottle, and a pneumatic piston appear respectively in (a), (b), (c), and (d) of figure 1 at a similar resizing scale factor.



(a)



(b)



(c)



(d)

Fig. 1: Object used for this study observed at a similar resizing scale factor. (a) a trophy cup. (b) a paddle. (c) a dark glass bottle. (d) a pneumatic piston.

The way we ultimately proceed is to place these objects on a conveyor belt until they are detected by the cameras and then treated as described in section 3. In figure 2, we present a 3D scheme of the system.

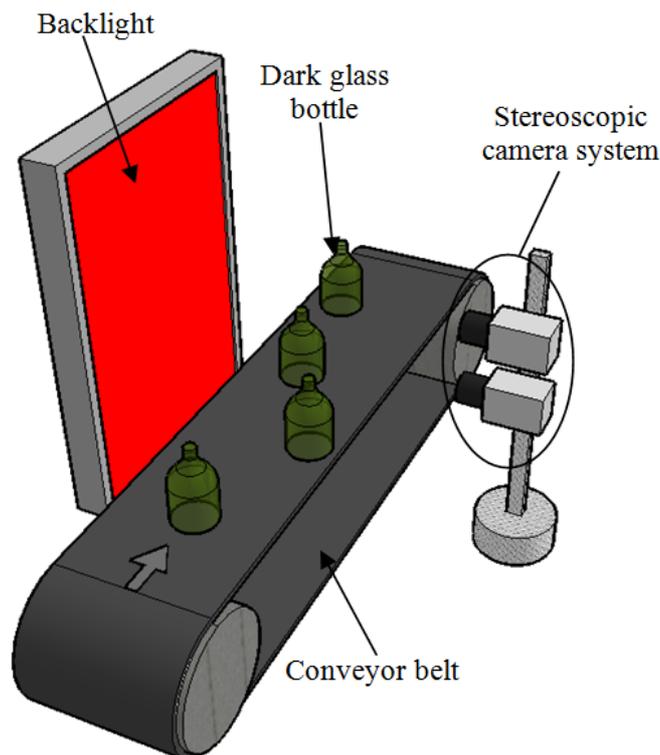


Fig. 2: Scheme of the complete system design.

III. OPERATIONAL PRINCIPLES

The system implementation is quite direct and fast. The stereoscopic system, once calibrated using the calibration method depicted in [14] and [15], is placed in front of a backlight. Lens focal distance and backlight dimensions are defined by the range of objects' sizes with which the system has to operate. The region of interest (ROI) for each camera is defined by the user directly on the PC screen and the process is launched.

The process starts by capturing the first image as the reference image and waiting until an object is detected by background subtraction when it generates a sharpened shadow by passing in front of the backlight. The binarization of the images is direct and precise due to the high contrast, and depending on the possible illumination variations the threshold can be a constant value or an adaptive value computed by algorithms such as Isodata or Kapur threshold [16] [17]. The object detection and image segmentation for foreground for both up and down cameras is shown in figure 3 referring to the situation presented in figure 2. Images (a) y (b) in figure 3 correspond respectively to the up (reference) and down simultaneously captured images. Images (c) and (d) in figure 3 present the foreground extraction result from the images (a) and (b) of figure 3 respectively.

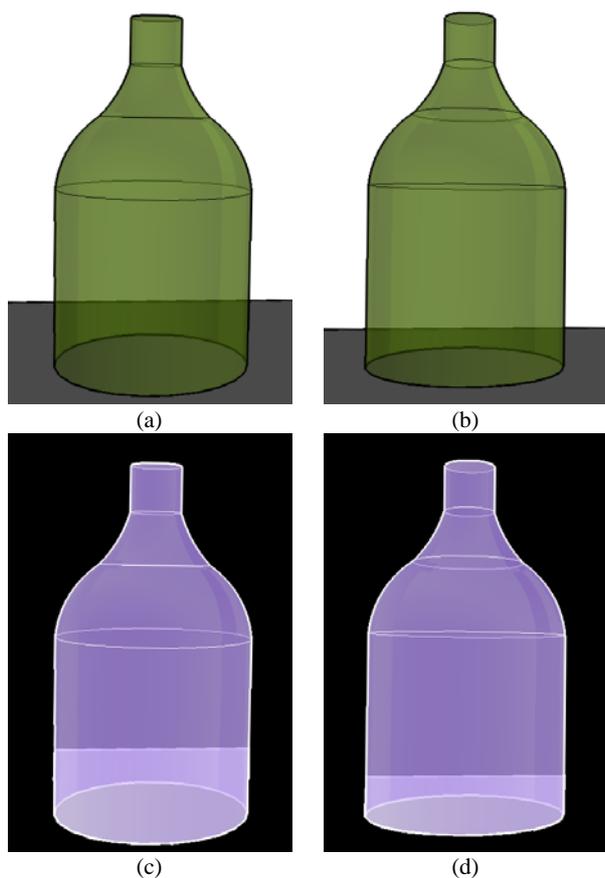


Fig. 3: Detail of the Image Treatment for fast 3D localization. (a) and (b) corresponds respectively to the up (reference) and down simultaneously captured images. (c) and (d) present the foreground extraction result from the images (a) and (b).

Then the edge of the object in the images is processed by a Canny filter followed by dilatation effect. Those steps are shown in figure 4, offering a continuous vision of the global image treatment succession. Images (a) y (b) in figure 4 present the edge extraction result using a Canny filter from the images (c) and (d) of figure 3 respectively. Images (c) and (d) in figure 4 present the edge dilatation result from the images (a) and (b) of figure 4 respectively.

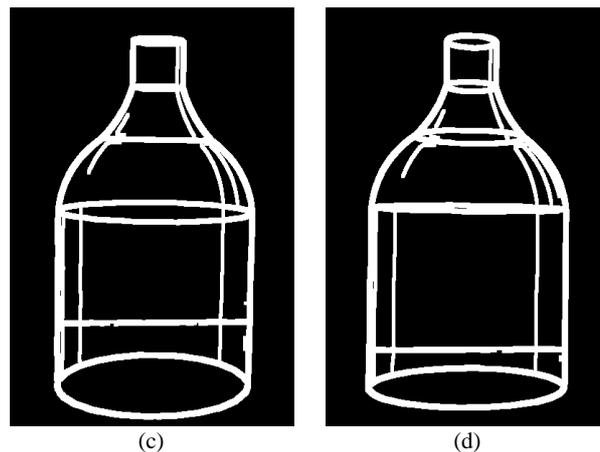
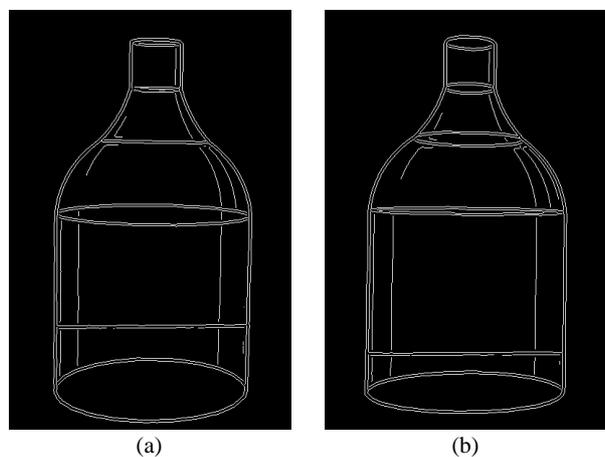


Fig. 4: Detail of the Image Treatment for fast 3D localization. (a) and (b) present the edge extraction result using a Canny filter from the images (c) and (d) of figure 3 respectively. (c) and (d) present the edge dilatation result from the images (a) and (b) respectively.

This dilatation is useful to proceed with the next correlation step, which finds the best matching between both shapes. But once positions are encountered in the stereoscopic images, dilatation is even more useful to fuse the two dilated shapes in a unique correspondences image and apply the distance transform to just conserve the “stereoscopic” skeleton of the shape. Actually this skeleton is obtained by applying a local maximum filter on the distance transformed image. The resulting skeleton obtained in case of the presented situation in figure 2 is shown in figure 5.

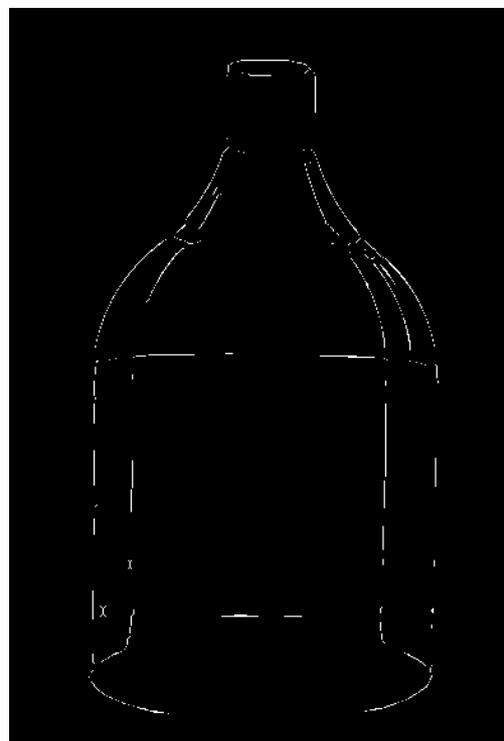


Fig. 5: Stereoscopic skeleton obtained in case of the presented situation in figure 2.

We then consider this skeleton as the stereoscopic point correspondences along the edge in both images, and localize it in 3D. This method is explicitly depicted in the chart in figure 6.

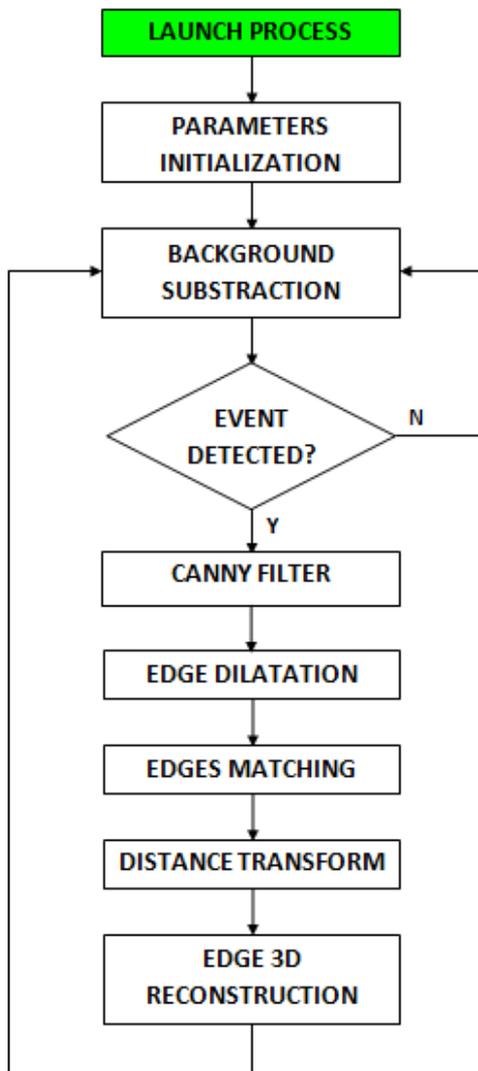


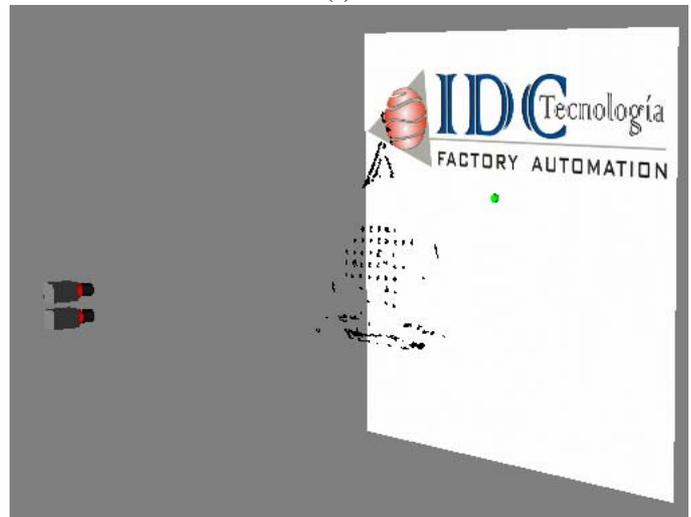
Fig. 6: Chart of the fast 3D localization method.

IV. RESULTS AND CONCLUSION

To observe the error produced by a simple center-of-mass 3D localization, we also compute this stereoscopic point on the object once binarized. The images in figure 7 show the 3D representation of the results obtained for each object on the list presented in section 2. The group of black points is extracted from the stereoscopic edges, and the small green sphere represents the center-of-mass 3D localization.



(a)



(b)



(c)



Fig. 7: Positioning error of the center of mass, compared with the edge localization of the different objects.

We checked the distances between the reference camera and object, and our first conclusion is that there is fairly good system precision for 3D shape localization. We also note in those four 3D representations that the center-of-mass is strongly separated from the localized edge, which indicates a quite strong positioning error of the first one. Indeed, due to object segmentation imprecision and the stereoscopic images differences caused by the different angles of view, the 3D localization of the center-of-mass is quite imprecise, at least at short distances. In addition, we repeated the method a few times in the same object position and noted how this center-of-mass can strongly shift due to precision error in each stereoscopic image. On the other hand, the use of the whole shape for 3D localization ensures the repeatability of the measurement. The images in figure 8 show this center of mass imprecision with the same object: a trophy cup.

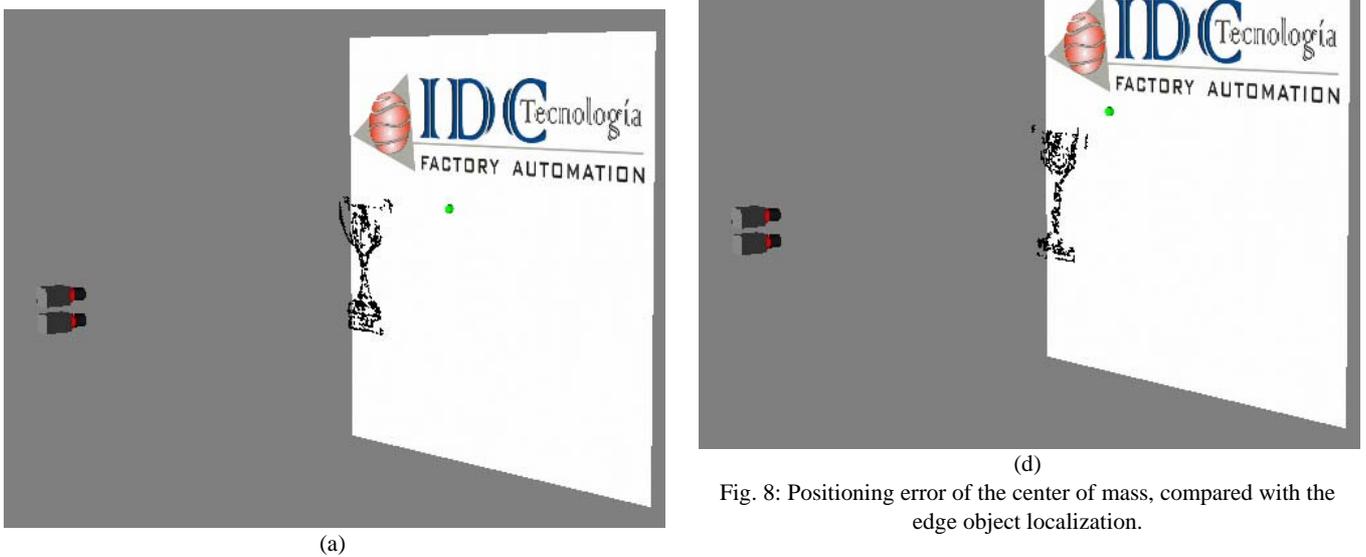


Fig. 8: Positioning error of the center of mass, compared with the edge object localization.

It is interesting for some industrial applications to note that

the object is still easily recognizable and thus can offer information on a particular point that is still identifiable and associable to a part of the real object.

In order to evaluate the positioning error, repeatability and to compare edge base 3D localization with center-of-mass 3D localization, we placed the object, a trophy cup, at a fixed distance from the upper (reference) camera. Actually, the distance between the object and the upper camera along the Z axe is fixed to 1 meter, and we smoothly shifted the position of the trophy cup along the Y axe to observe the measurement error in the whole image. We processed almost 50 times the fast 3D localization and it comparative center of mass 3D localization.

In figure 9, we observe the edge based 3D localization result in yellow color. The result situates the object at about 97 centimeters from the upper camera which is a fairly good precision.

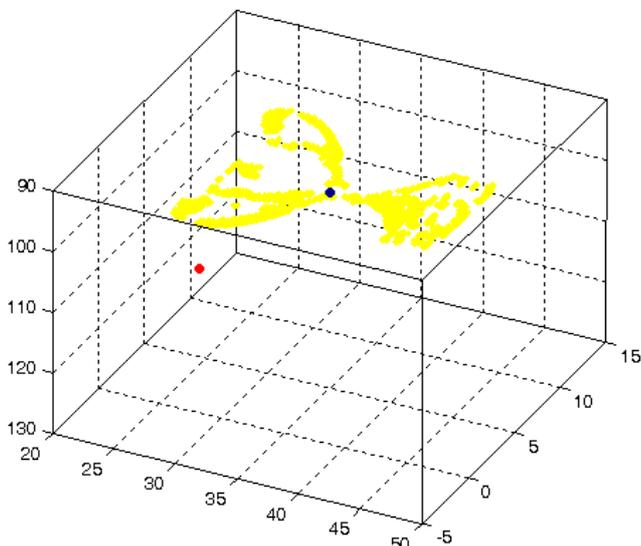


Fig. 9: Positioning error of the center of mass, compared with the edge object localization.

In figure 9, we also can observe a blue dot that marks the center of mass of the yellow cloud of points, and a red dot that marks the 3D localized center of mass, originally extracted from stereoscopic images.

From those almost 50 measurements, we compare those red and blue dots and computed some statistical information we observe in figure 10.

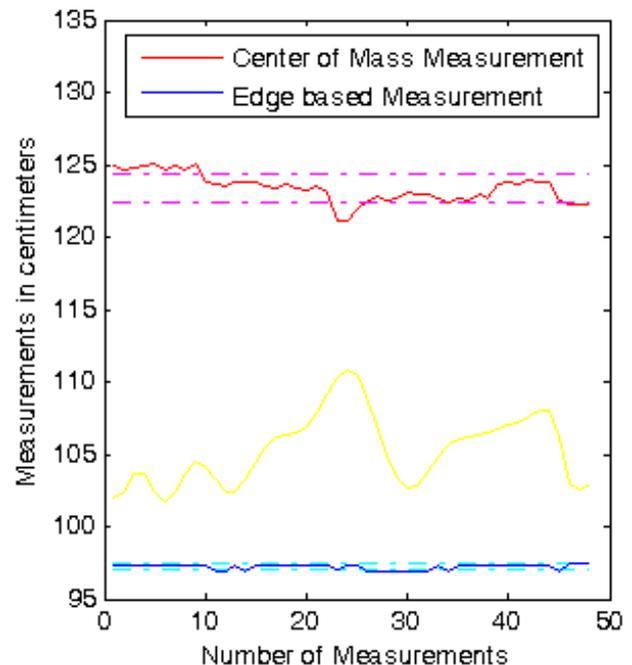


Fig. 10: Positioning error of the center-of-mass (red line) compared with the edge based 3D localization center-of-mass (blue line). The yellow line present some positioning variation along the Y axes to cover the whole image. The magenta dash dot line present the standard deviation centered on average value of red line. The cyan dash dot line present the standard deviation centered on average value of blue line.

The standard deviation in case of the fast edge based 3D measurement is represented in figure 10 by a cyan dash dot line and it value is equal to 0,2126 centimeters. The average computed distance of the trophy cup from upper (reference) camera among the Z axe is equal to 97,21 centimeters.

The standard deviation in case of the fast center of mass based 3D measurement comparing method is represented in figure 10 by a magenta dash dot line and it value is equal to 1,2275 centimeters. The average computed distance of the trophy cup from upper (reference) camera among the Z axe is equal to 123,24 centimeters.

Those numerical result definitively present the robustness of this fast 3D localization method for industrial application. The additional advantage of this result is that we still can recognize and so could easily classify the 3D localized object in a multiple object application case.

ACKNOWLEDGMENT

We would like to thank the UAH (Universidad de Alcala de Henares), our main partner for this investigation. Additionally we would like to thank the Massachusetts Institute of Technology (MIT) and the International Science and Technology Initiatives (MISTI) program.

REFERENCES

[1] Richard Hartley, Andrew Zisserman, "Multiple View Geometry in computer vision", Cambridge editions.

- [2] Olivier Faugeras and Quang-Tuan Luong, "The Geometry of Multiple Images", The MIT Press.
- [3] Mehmet Celenk, Rafic A. Bachnak, "Multiple Stereo Vision System for 3D Object Reconstruction", IEEE, 1990.
- [4] S. Y. Chen, Y. F. Li, and Jianwei Zhang, "Fast 3D Perception by a Color-Coded Vision System", Proceeding of the 2006 IEEE International conference on Mechatronics and Automation, June 25-28, 2006, Luoyang, China.
- [5] Hyejin Kim, Jaewan Choi, Howook Chang, Kiyun Yu, "Generation of Disparity Map Using Piecewise Linear Transformation", 5th WSEAS International Computational Intelligence, Man-Machine Systems and Cybernetics, 2006.
- [6] Humberto Loaíza, Jean Triboulet, Sylvie Lelandais, and Christian Barat, "Matching Segments in Stereoscopic Vision", IEEE Instrumentation & Measurement Magazine, 2001.
- [7] Mike Jones and John P Oakley, "A New Shape Representation for Fast 3D Reconstruction from Multiple 2D Perspective Images", IEEE.
- [8] Jose-Carlos Martinez, Eloina Coll, Jose Herraéz, "Software for Videogrammetry Image Matching", WSEAS MIV'01, Computer and Communications Systems, 2001.
- [9] Issa Hazem, Ruichek Yassine, Postaire Jack-Gérard, "A new genetic encoding for edge point stereo matching", 2nd WSEAS International Conference on Signal, Speech and Image Processing (ICOSSIP 2002), 2002.
- [10] Yong-In Yoon, Jong-Soo Choi, Ho-Keun Song, Jong-Ho Choi, "A Robust Corresponding Points for Uncalibrated Stereo Images", 7th WSEAS International Conference on Multimedia Systems and Signal Processing, 2007.
- [11] Val Petran and Frank Merat, "Stereoscopic Correspondence without Continuity Assumptions", IEEE, 2006.
- [12] Nauman Iftikhar, Saba Iqbal, Naveed Sarfraz Khattak "Location and Classification of Moving Objects Using Critical Area Analysis and Template Matching", 6th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, 2007.
- [13] Drago Torkar, Gregor Papa, "Accuracy of a 3D reconstruction system", 6th WSEAS International Conference on Signal Processing, Robotics and Automation, 2007.
- [14] T.A. Clarke and J.G. Fryer, "The Development of Camera Calibration Methods and Models", Photogrammetric Records, 16(91): pp 51-66, 1998.
- [15] Janne Heikkilä and Olli Silvén, "A Four-step Camera Calibration Procedure with Implicit Image Correlation", 1997.
- [16] N.Otsu. "A threshold selection method from gray-level histograms". IEEE Transactions on Systems, Man and Cybernetics, 9(1):62-66, 1979.
- [17] T.W. Ridler and S. Calvard. Picture thresholding using an iterative selection method. IEEE Transactions on Systems, Man and Cybernetics, SMC-8:630-632, 1978.

O. Demuynek was born in Colombes (France) in 1977. He obtained the Computer Science Master degree from the ISIMA (Institut Supérieur en Informatique Modélisation et ses Applications), in Clermont-Ferrand (France), in 2001. He is currently working on a PhD in Artificial Vision at the ETSI (Escuela Técnica Superior de Ingenieros de Telecomunicación) of the UPM (Universidad Politécnica de Madrid), in Madrid (Spain).

He worked as Computer Science Engineer for the Defense and Security field of applications investigating and developing Artificial Vision algorithms at NTGS S.L. (New Technology Global Systems) Company, in Las Rozas de Madrid, Madrid (Spain). He is currently working as R&D Engineer in the Industrial Artificial Vision field of applications at IDC Tecnología S.L. Company, in Alcalá de Henares, Madrid (Spain). Current research interests focus on robotic, stereoscopic vision systems for 3D localization, Artificial Vision for hazardous areas, object classification, and Image Treatment.

C. Pacheco Cedeño was born in Granma (Cuba) in 1978. He graduated in Computer Science at the University of Oriente, Santiago de Cuba (Cuba), in 2002. He is currently studying a Master degree in advanced electronic systems at the UAH, Madrid, Spain

He is professor in the Department of Computer Science at the University of Oriente, Santiago de Cuba (Cuba). He is also associated to IDC Technology Company. He is mainly experienced in computer vision, data base topics, and high level programming. Current research interests are centred on Computer Vision and Detection behavior on Stereoscopic Images.

A. L. Moore was born in Milwaukee, Wisconsin (USA) in 1988. She is currently an undergraduate student in economics and mathematics with computer science at the Massachusetts Institute of Technology in Cambridge, MA.

She has worked as a Summer Intern at IDC Tecnología in Madrid, Spain in the summer of 2009 and as a Summer Intern at Artisan Partners, Ltd. in Milwaukee, Wisconsin in the summer of 2008. During the academic year she currently works as a Lab Assistant for the MIT Department of Electrical Engineering and Computer Science and has worked in the past as a Student Assistant for the MIT Enterprise Forum of Cambridge.