PRECISE 3D MEASUREMENTS WITH A HIGH RESOLUTION STEREO HEAD

A. GASTERATOS, R. MARTINOTTI, G. METTA AND G. SANDINI

LABORATORY FOR INTEGRATED ADVANCED ROBOTICS

DEPARTMENT OF COMMUNICATION, COMPUTER AND SYSTEM SCIENCES

UNIVERSITY OF GENOA

VIALE CAUSA 13, I-16145 GENOA, ITALY

E-MAIL: {antonis, koga, pasa, giulio}@lira.dist.unige.it

ABSTRACT

In this paper, we investigate a method to obtain accurate 3D measurements and we compare it with a standard photogrammetry method. The method utilizes a closed-loop to bring the point under measure in the middle of the two images of a stereo pair. It uses poor calibrated cameras and the accuracy of the measurements relies on the high resolution of the stereo-head encoders and on its calibration as well. Experimental results have shown that the proposed method is more accurate than standard photogrammetry method, as well as it is more robust to small variations of the initial position of the cameras.

KEYWORDS: *stereo vision, calibration, measurements*

I. INTRODUCTION

In human vision system, the differences between the left and right images are used to recover 3D properties of a scene. Similarly on an artificial vision system, the differences between the two images can be used for the extraction of many useful 3D characteristics such as the depth, the surface normal and the exact position of a point (Horn 1986). The disparity is the difference of the projection of the same point on the left and the right image. Using this information along with the focal length of the cameras and the distance between the two cameras the 3D coordinates of any point can be determined. If we consider the simple case of Fig. 1, then the 3D coordinates of the point P are given by:

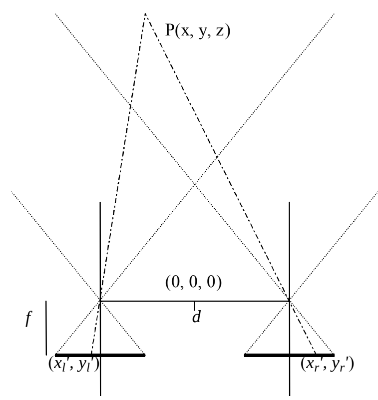


Figure 1: Simple stereo geometry for 3D estimation

$$x = d\frac{x_l^{'} + x_r^{'}}{2\left(x_l^{'} - x_r^{'}\right)},\ y = d\frac{y_l^{'} + y_r^{'}}{2\left(x_l^{'} - x_r^{'}\right)},\ z = d\frac{f}{2\left(x_l^{'} - x_r^{'}\right)} \quad (1)$$

where: *d* is the base-line, *f* is the focal length of the two cameras (it is supposed that both cameras have the same *f*) and $x_l^{'} - x_r^{'}$ is the stereo disparity. The subscripts are referred to the left and the right images respectively.

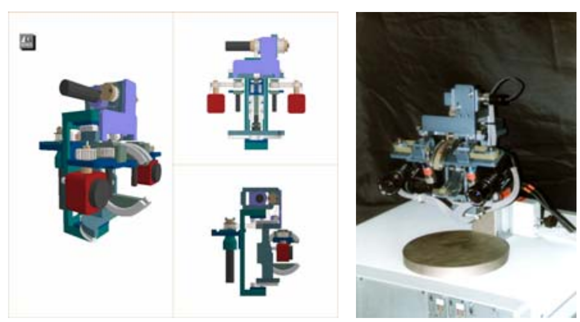
However, in practice several problems arise when attempting to implement the above formulae. The first is the *correspondence problem* (Barnard and Thompson 1980). It is obvious that in order to be able to apply equations (1), one must know which are the x_l' and x_r' that correspond to the same point P in space. The second problem is that of *camera calibration.* As it is obvious in Fig. 1, no distortion introduced by the lenses has been taken into consideration. In order to solve this problem several camera models have been proposed (Tsai 1987), (Heikkila and Silven 1996), (Heikkila and Silven 1997). Finally, a very important issue is the stereo setup calibration. The equations (1) demand that the two cameras are aligned exactly and also that they are both perpendicular to the base-line. Any small variation on the alignment of the cameras reflects dramatically on the accuracy of the measure (see Fig. 1).

In this paper, we present a method for calibration of a stereo setup, using visual information. Furthermore we investigate by means of accuracy and robustness, a methodology to obtain accurate 3D measurements. This methodology is based on a recursive closed-loop approach. The closed-loop is used in order to bring the point under measure into the middle of the two images. The accuracy of the measurements with this method relies on the high resolution of the encoders of the motors of the head and its calibration as well. The method is compared with the photogrammetry method of formulae (1). In both cases a priori knowledge of a CAD model is used to localize the target under measurement and, thus, we solve the correspodence problem. The end goal is to find and measure the location of 3D structures with respect to a CAD-model and determine the pose of a moving legged robot. Experimental results have shown that the recursive method is more accurate than the photogrammetry method, as well as it is more robust in the sense that a small variation at the initial state does not affect the accuracy of the first method as much as it affects the accuracy of the second.

II. The Stereo-Head

A. Characteristics

The stereo head that is used in our application is shown in Fig. 2. It has been designed and implemented to be an accurate vision-based measuring device. For the control of the pan, tilt and vergence, four harmonic drive actuators are used. These actuators have been chosen according to their mechanical characteristics, which, due to their harmonic drive gearing, provide high reduction ratios in a single stage, zero backlash and high precision. For the movement transmission from the actuator to the joints teeth belts have been used, which gives better results in term of accuracy, than usual gearing transmission. The specifications of the head are summarized in Table 1. The head was leisurely designed in order to be compact, portable and low weight. Its dimensions are 209mm x 222mm x 185.16 mm and its weight is about 3kg.



(a) (b)

Figure 2: The Euro-Head: (a) Mechanical design and (b) implementation

Table 1: Synopsis of the stereo head characteristics.

	Range	Velocity	Acceleration	Resolution
Pan	$\pm 45^{o}$	$\geq 73^{o}/s$	$\geq 1600^{o}/s^2$	0.007^{o}
Tilt	$\pm 60^{o}$	$\geq 73^{o}/s$	$\geq 2100^{o}/s^2$	0.007^{o}
Vergence	$\pm 45^{o}$	$\geq 330^{o}/s$	$\geq 5100^{o}/s^2$	0.03^{o}

B. Calibration

The calibration of the head is divided into two parts. The first is the computation of the starting point for the pan and the tilt. This part is done purely mechanically. The head is smoothly moved along the pan axis firstly towards the hard stop on the left and then towards the right side of the axis. The encoders are read at each hard stop and the middle of the pan axis is determined. Similarly, the tilt axis is initialized.

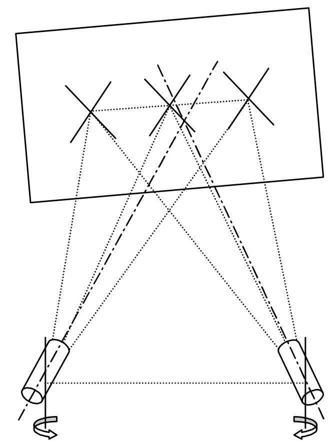
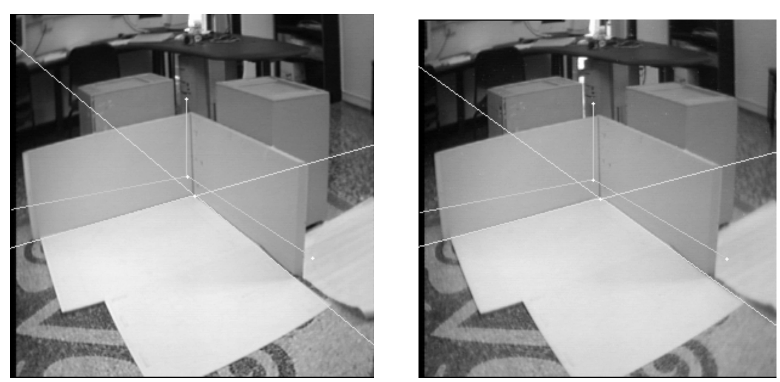


Figure 3: Vergence calibration: A plate with three aligned crosses is roughly posed in front of the head

The second part of the calibration procedure is the initialization of the vergence. The goal is to align the two optical axes perfectly in parallel and perpendicular to the tilt axis. The method followed allows us to correct the vergence error of the cameras using visual information. The idea is to create a virtual triangle between the cameras and a fixed target on the horizontal plane of the cameras. The vergence calibration setup is shown in Fig. 3. We aim with the cameras towards the target and the distances of the target from each camera are measured. The next step is to move the two cameras so that the two distances will be equal and, therefore, the corresponding angles will be equal as well. The target is the central one of three aligned crosses put on a plane plate and equally distanced from each other (Fig. 3). For each camera we use the crosses and simple trigonometric laws to estimate the distance from the central point.

III. MEASUREMENT STRATEGY



(a) (b)

Figure 4: The junction is found in both images (right (a) and left (b)) and marked with two crossing lines. The junction should have been found, according to the CAD model to where it is marked with three crossed lines

The junctions are the intersections of two or more edges in the image. They are tracked in both images by intersecting two different lines. The different lines are found in the image plane by applying first SUSAN edge detector (Smith and Brady 1997), followed by a Hough transform (Duda and Hart 1972). An example of junction tracking using this sequence is shown in Fig. 4. The two crossing lines represents the two lines tracked by the Hough transform, whilst the three crossing lines represents the position of the junction, according to the CAD model. Fig. 4 demonstrates that small displacements of our target do not affect our algorithm. From the CAD model we get the 3D position of the target as well as number of the lines in each sight (left and right). The first is used in order to aim roughly with the head towards the target and the second as a threshold in the Hough accumulation array. The lines laying around the diagonal of the image have a much bigger value in the accumulation array than the lines near

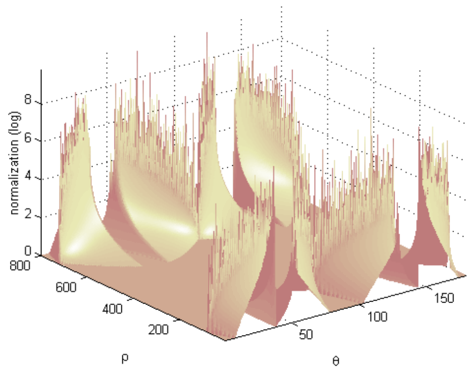


Figure 5: Normalization mask applied to the Hough accumulation array of a 640x480-pixels image (r=800).

the corners. For this reason the Hough accumulation array is normalized, by multiplying it with a mask as the one of Fig. 5. In this manner faulty lines that have greater values at the accumulation array than existing lines, are eliminated.

The above described junction tracking technique is the first step that is applied on both left and right images in order to take measurements. For the measurement of the distance of a junction from the binocular head according to this technique the junction should be at the center of the image in both images (right and left). When this is the case the two cameras are verging on the certain junction and the direct kinematics of the head is applied, in order to determine the position of the junction relatively to the head, i.e. $(x, y, z) = K(q_1, q_2, q_3)$ (see also Fig. 6).

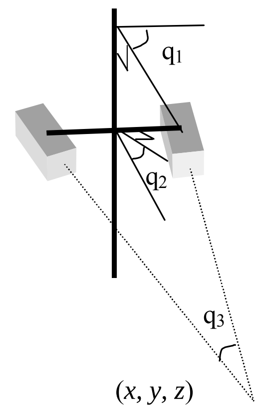


Figure 6: The 3 d.o.f. stereo head, verging on point (x, y, z)

In order to put the junction in the center of both the images the following closed loop procedure is followed:

The system is initialized, by constructing a behavior table M. M contains the parameters according to which the error is corrected in each step of this procedure. By error here we define the vector from the image center to the mapping of the junction on the image plane.

In each step the differences in the head angles are calculated according to:

1. $[\Delta q_1, \Delta q_2, \Delta q_3]^T = M\,[\Delta x_l, \Delta y_l, \Delta x_r, \Delta y_r]^T$
2. If $M\,[\Delta x_l, \Delta y_l, \Delta x_r, \Delta y_r]^T \neq [0, 0, 0, 0]^T$ goto 1

   else, make the measurement according to:

   $(x, y, z) = K(q_1, q_2, q_3)$
3. end

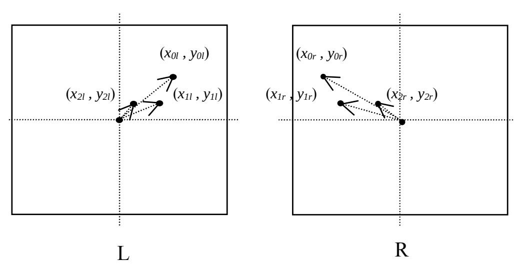


Figure 7: Graphical illustration of the closed loop procedure with sequential error correction

The procedure described above is graphically illustrated in Fig. 7. Because of the fact that a closed loop control is followed, it is obvious that no measurement is taken until the junction is put in the center of both the images. Therefore the accuracy measurement is dependent only on the accuracy of the motors, which control the joints of the head, whilst the camera calibration is redundant. In fact the only parameter needed from the camera calibration table is the image center, i.e. the center of the radical distortion.

IV. Experiments

In order to measure the efficiency of this poor calibrated camera method we use the standard photogrammetry method as a benchmark. In order to increase the accuracy of the photogrammetry method we used well-calibrated cameras (Heikkila and Silven 1997).

Calibration is performed on each camera separately and the calibration table of each camera is obtained. The camera model allows us to calculate the actual position of (x_l, y_l) and (x_r, y_r) on both image planes. These are applied on formulae (1) and, thus, the 3D position (x, y, z) of the target point is measured. A theoretical study of the expected errors has been performed for both cases. The errors are maximized in the proposed method when the two optical axes are in parallel and, therefore, no measurement can be performed (see Fig. 6). In the case of the photogrammetry method, the errors are maximized when $x_l^{'} - x_r^{'}$ approaches 0. In Fig. 8 three cases of simulating the error caused by initial misalignment in only one of the two cameras are presented. Fig. 8 (a) presents the expected error in the X direction with the photogrammetry method and with an initial misalignment of the left camera by 0.3 degrees. An area of 80 pixel around the center of the image has been selected to perform the simulation. The error is maximized, as it is expected, when $x_l^{'} \approx x_r^{'}$. At the rest of the plot the error is practically 0 (65% in [0mm,5mm]). Fig. 8 (b) depicts the expected error also with the photogrammetry method for different initial misalignments of the left camera (from –1.5 to 1.5 degrees). Again here at the major part of the plot the error is practically 0 (65% in [0mm,6mm]). Finally Fig. 8 (c) shows the expected error in distance with the proposed method also for different initial misalignments of the left camera. The two cameras have symmetrical vergence angle, which is in the range [0,89] degrees. Here at the major part of the plot the error is comparable to the two previous cases (65% in [0mm, 4mm]), but the maximum error is much lower that before. Moreover, in these simulations no misalignments of the cameras with respects to the horizontal plane have been considered, i.e. the two sensors of the cameras were considered as being perfectly perpendicular to the horizontal plane. Misalignments of this kind affects only the photogrammetry method (see Fig. 1), but not the proposed one. Finally, formulae (1) assume the same focal length in both cameras, which is rather difficult to be achieved. Hence, the photogrammetry method is ill against small variations of the focal lengths of the two cameras, which does not stand for our method.

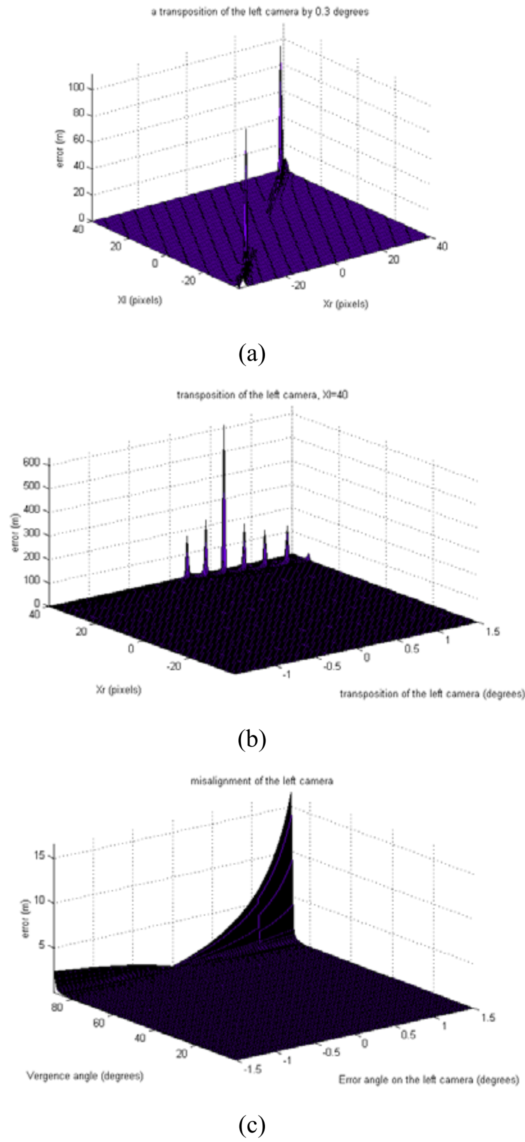


Figure 8: Theoretical cases of errors: (a, b) photogrametry method, (c) proposed method

The experiments on accuracy of the measurements with the above described methods where done using the experimental setup of Fig. 9. A 3m x 4m x 2.5m metal mockup was used to take measurements with the binocular head. The position of the head was measured on each pose using a teodolite-like instrument of high accuracy (monmos). At the same time the position of the target on the mockup was measured. A series of measures has shown that the proposed method is more accurate than the standard photogrammetry method. More specifically, for medium and big distances (>=3m) the deviation with the proposed method was around 5%, whilst the photogrammetry method was around 20%. For smaller distances (~1m) the accuracy is even bigger, i.e. the deviation with our method was around 10%, whilst the photogrammetry method was around 20%.

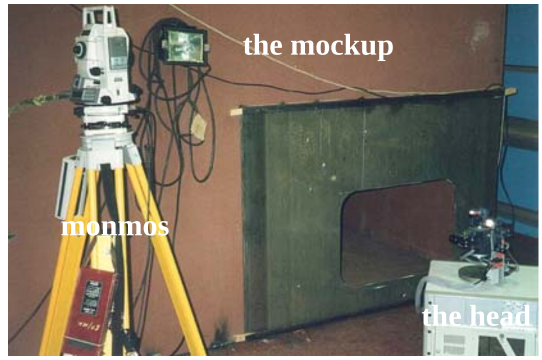


Figure 9: The experimental setup

V. Conclusions

In this paper we have presented a closed-loop method for obtaining 3D measurements. The proposed method is intended to be used for the localization of a legged robot in a known, by its CAD model, environment. The proposed method does not need highly calibrated cameras, but only the center of the radial distortion of the two images. For the evaluation of its efficiency, our method has been compared to standard stereo photogrammetry method. It has been shown that the proposed method performs better in terms of accuracy, as well as it is affected less by misalignments of the initial position of the two cameras. However, due to the utilization of the closed-loop to put the point under measure in the center of the images, the proposed method is slower. Nevertheless, this is not expected to cause problems to the overall system, since (i) the accuracy is more critical than the time and (ii) the legged robot which will carry the head is reasonably slow, i.e. 6m/min, which allow us to obtain the measurements and perform the pose estimation.

Acknowledgments

The research presented in this paper has been supported mainly by the Esprit project ROBVISION (EP-28867) and partially by VIRGO-TMR project.

References

Barnard, S. T., and Thompson, W. B. (1980). “Disparity Analysis of Images.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, pp. 333-340.

Duda, R. D., and Hart, P. E. (1972). “Use of the Hough Transform to Detect Lines and Curves in Pictures.” *Commun. ACM*, Vol. 15, pp. 11-15.

Heikkila, J., and Silven, O. (1996). “Calibration Procedure for Short Focal Length Off-the-Shelf CCD Cameras.” *Proceedings of ICPR '96*, pp. 166-170.

Heikkila, J., and Silven, O. (1997). “A Four-step Camera Calibration Procedure with Implicit Image Correction.” *CVPR '97*, San Juan, Puerto Rico, pp. 1106-1112.

Horn, B. K. P. (1986). *Robot Vision*, MIT Press, Cambridge, Massachusetts.

Smith, S. M., and Brady, J. M. (1997). “SUSAN - a new approach to low level image processing.” *International Journal of Computer Vision*, Vol. 23, pp. 45-78.

Tsai, R. Y. (1987). “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using off-the Shelf TV Cameras and Lenses.” *IEEE J. Robotics and Automation*, Vol. 5.