

# Factors Affecting the Accuracy of an Active Vision Head

Antonios Gasteratos<sup>1</sup> and Giulio Sandini<sup>2</sup>

<sup>1</sup> Laboratory of Electronics  
Section of Electronics and Information Systems Technology  
Department of Electrical and Computer Engineering  
Democritus University of Thrace  
GR-671 00 Xanthi, Hellas  
agaster@ee.duth.gr

<sup>2</sup> Laboratory for Integrated Advanced Robotics  
Department of Communication, Computer and System Sciences  
University of Genoa  
Viale Causa 13, I-16145 Genoa, Italy  
sandini@dist.unige.it  
<http://www.lira.dist.unige.it>

**Abstract.** In any measuring system the categorization of the error generation factors leads to simplification of complex error problems and to higher suppression of the error. In this paper we categorize, quantify and analyze the errors that affect a binocular active vision head. Simulations have been made and experimental results on a high resolution pan-tilt-vergence mechanism are also proposed. As a conclusion it can be said that the system performs optimal when it is initialized so that the two cameras are perfectly aligned and perpendicular to the baseline. Small variations in the vergence angle or small horizontal deviations of the principal point alters the measurement dramatically. On the other hand, variations in pan and tilt and vertical deviations of the principal point, affect the measurement insignificantly.

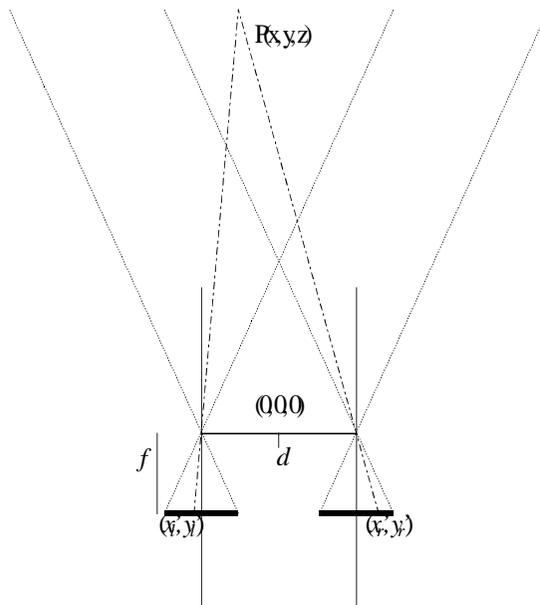
## 1 Introduction

Stereo is used widely in artificial vision to extract 3D information such as depth, surface normal and the exact position of a point [1-3]. The most widely applied method is disparity, i.e. the difference of the projection of the same point on the left and the right image. Using this information, the focal length of the cameras and the baseline, the 3D coordinates of any point can be determined. Let us consider the simple case of Figure 1: the 3D coordinates of the point P are given by:

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

where:  $d$  is the baseline,  $f$  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.

However, several problems arise when attempting to realize the above formula. The first is the *correspondence problem* [4]. It is obvious that in order to be able to apply eq. (1), one should know which are the  $x'_l$ , and  $x'_r$  that correspond to the same point P in space. An analysis of the errors due to false correspondence has been carried out in [5]. The second problem is that of *camera calibration*. As is obvious in Figure 1 the distortion introduced by the lenses has not been taken into consideration and, moreover, the focal lengths of the two cameras were considered as equal. The lens distorted images are usually rectified using a of the camera calibration method (e.g. [6, 7]). An equally important issue is the *stereo setup alignment*. Eq. (1) demands that the two cameras are aligned exactly and also that they are perpendicular to the baseline. The camera calibration/alignment issue is studied in [8, 9], where a taxonomy and quantification of the errors due to misalignment of the stereo pair and due to misalignment of the sensor on the camera has been done.



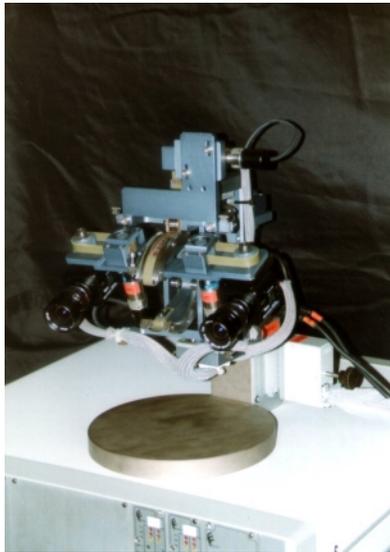
**Fig. 1.** Simple stereo geometry for 3D estimation

In [10] a method based on the measure of the vergence angle, was presented. In this method the cameras are not calibrated, because they are moved in a closed loop, so that the point under measure is put on the principal points of the two cameras. In this paper we attempt a complete analysis of the errors that may affect the method in [10]. These can basically be discriminated into visual errors, misalignment errors and mechanical errors. We also provide a theoretical analysis, supported by simulations for the above errors. Experiments have been carried out. These show that for the given setup the most critical errors are the visual and the misalignment ones, as the mechanical ones tend to be zeroed. The critical parameters for a high accuracy measurement are: the principal point identification; the initial alignment of the pair of

cameras on the head and an image processing method eliminating the errors due to false correspondence.

## 2 The Stereo Head

Our stereo head is shown in Figure 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. Teeth belts have been used for the movement transmission. This gives better results in term of accuracy than usual gearing transmission.



**Fig. 2.** The head. The head has been utilized as an accurate measuring device for the purpose of ROVISION project [11, 12]. More specifically, it has been mounted on a walking/climbing and the measurements obtained by it have been used for self-localization of the robot.

## 3 Error Generation Factors

As has been stated above, the method in [10] utilizes a recursive method, so that a junction under measure is put on the principal points of both the images. The junctions are tracked by intersecting two different lines. The measurement is taken as:

$$l = d / (2 \sin \nu) \tag{2}$$

where  $\nu$  is the vergence angle.

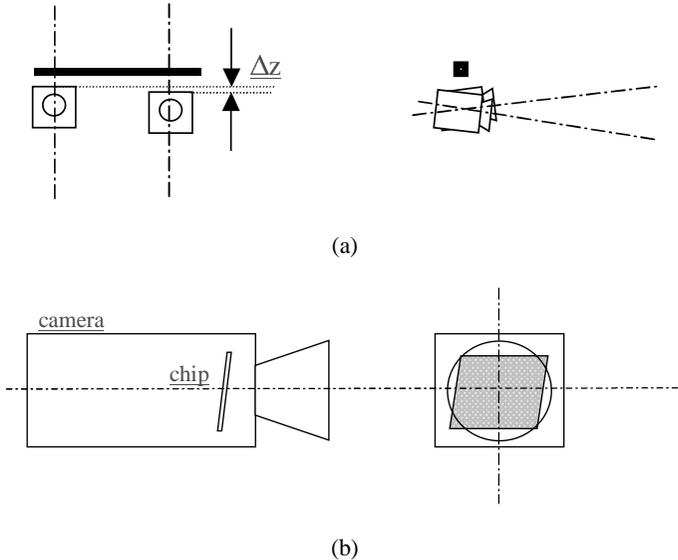
Because of the fact that the baseline  $d$  of the head is comparatively small (190mm), the measured distance is very sensitive to small variations of the vergence angle. The measurement is obviously more sensitive in long distances, where  $\nu \rightarrow 0$  and, consequently,  $\sin \nu \rightarrow 0$ . The 3D position of a point given the pan ( $p$ ), tilt ( $t$ ) and vergence angles is given by:

$$[x, y, z]^T = \mathbf{T}(p, t, \nu, d) \cdot [l, 0, 0]^T \tag{3}$$

with  $\mathbf{T}$  being the 4x4 matrix of the kinematics of the head.

The factors that affect the accuracy of this measurement can be summarized as:

- Visual errors
- Alignments errors
- Mechanical errors



**Fig. 3.** (a) Misalignments of two cameras on a stereo head and (b) misalignment of the chip on a single camera

If the junction is not on the principal point of each camera, then we call that error visual error. This error is almost eliminated by the closed loop procedure to approach the junction. We say almost, because we have two degrees of freedom (dof) to move each camera separately in the horizontal direction but none on the vertical one. Therefore it is easy to compensate horizontal disparities using the vergence dof, but we are not able to compensate vertical ones. It may be expected that since the cameras are mounted on the same horizontal plane, with respect to the head, no vertical disparities should exist. However, in a real system no matter how well are mounted the cameras on the head or the sensor on camera, there are always displacements as

these shown in Figures 3a and 3b. Moreover, due to the fact that the lenses are not identical they introduce different distortions around different principal points. In other words the different lenses may decline the optical axis from the perpendicular to the baseline position, though the bodies of the two cameras might be perfectly parallel.

It is only in the case where the junction under measure appears on the principal point when we know that the two optical axes form a proper triangle, from which we can obtain the measurement. However, this is a necessary but not sufficient condition to form the proper triangle. In order that the measurement according to eq. (2) is correct the triangle must be isosceles and the vergence angle precisely known. This brings us to the second point of the error classification, i.e. the head calibration method (cameras alignment). The head calibration method should ensure the following two conditions: (i) Initially the two cameras are perfectly aligned on the horizontal plane of the head; (ii) they are perpendicular to the baseline ( $\nu = 0$ ). If these conditions are fulfilled, then equally we can derive the previously described statement.

## 4 Experiments

In this section we describe several experiments carried out to measure the accuracy of the overall system. Cameras with nominative focal length 6mm lenses were used. For the measurement a perfect metallic cube of dimension 100mm was used. Several experiments were carried out in order to measure the sensitivity and the tolerance of the system to several kinds of errors, as described below. The experiments were carried out with human supervision, to avoid problems of choosing a wrong pixel due to image processing. However, an unsupervised experiment is also included at the end of this section to test the overall efficiency of the system. Analytical presentation of the experimental results can be found in [13].

### a. Visual Errors

Let us consider first the visual errors due to horizontal displacement of the principal point. This case is graphical illustrated in Figure 4a. In this figure we have considered a displacement on the left camera. Taking into consideration the pin-hole model we can directly reduce the displacement in pixel into an angle as:

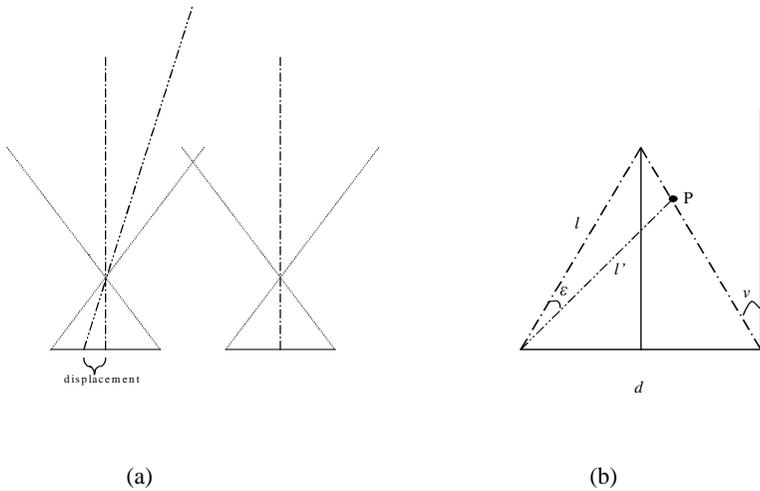
$$\varepsilon = \arctan(p_d CL / p_t f) \tag{4}$$

where  $\varepsilon$  is the reduced angle of the displacement,  $p_d$  is the displacement in pixels,  $p_t$  the total pixel number in one line and  $CL$  the chip length.

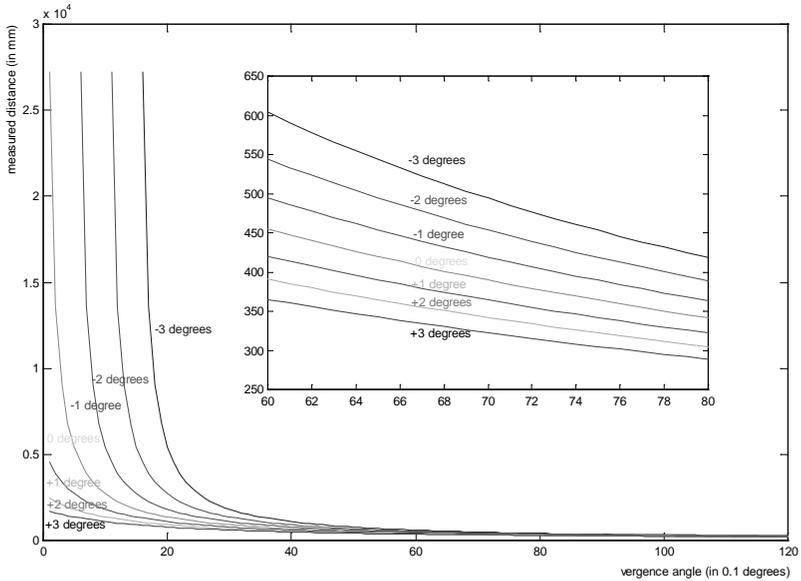
In Figure 4b the resulting error is shown. Due to the fact that  $(p_d CL) / p_t$  and  $f$  are negligible compared to the distance  $l'$  of the point to the left camera and the measured distance  $l$ , we can say that Figure 4b is a very good approximation. It is easily derived from that figure that:

$$l' = d \cos \nu / \sin(2\nu + \varepsilon) \tag{5}$$

which for  $\varepsilon=0$  it is reduced to eq. (2).



**Fig. 4.** (a) Horizontal displacement of the principal point and (b) the resulting error due to this displacement



**Fig. 5.** The distance  $l'$  as a function of the vergence angle, for several displacement angles  $\epsilon$

A graphical illustration of the distance error, for several displacements, is provided in Figure 5. As shown a displacement of 1 degree, which corresponds to a displacement of about 16 pixels, may result in an error of several meters in the distance

measurement, according to eq. (4). From eq. (3) it is obvious that this affects directly to the determination of the 3D coordinates of the point.

We evaluated this simulation by real measurements. The cube was positioned at a distance such that its center of mass was approximately 800mm from the center of the head. We inserted intentionally a displacement of  $-20, -10, 10$  and  $20$  pixels at the principal point of the left camera. When the displacement is  $0$  the distance  $l$  was optimal and near to the correct one  $l'$ . However, for displacement different to  $0$   $l$  was varying accordingly.

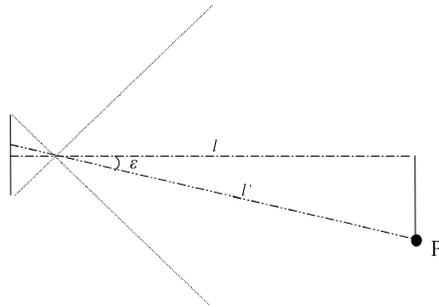


Fig. 6. Vertical displacement of the principal point and the resulting error.

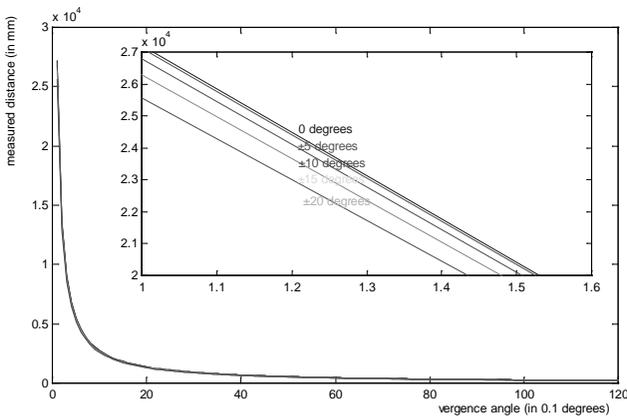


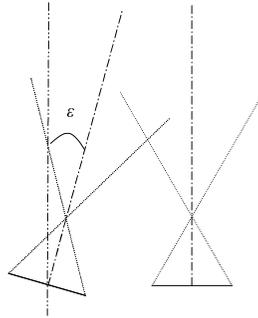
Fig. 7. The distance  $l'$  as a function of distance  $l$  and the displacement angle  $\epsilon$

Let us now consider a vertical displacement of the principal points. Because of the nature of the measurement method in [10], a triangular cannot be formed and, therefore, no measurement can be taken, unless the vertical displacement has taken place simultaneously in both cameras. In Figure 6 we present the profile of one of the two cameras with a vertical displacement corresponding to angle  $\epsilon$  and for the other camera we consider exactly the same case. As one can see in this figure the actual

distance  $l'$  is related with the measured distance  $l$ , as:  $l' = l / \cos \varepsilon$ . Intuitively it is apparent that the distance  $l'$  should not vary much with respect to  $l$ . This is also shown in Figure 7, where the curves for  $\varepsilon$  equal to 0,  $\pm 5$ ,  $\pm 10$ ,  $\pm 15$  and  $\pm 20$  degrees, respectively, are almost overlapping. In the top-right corner snapshot it is shown that for vergence angles from 0.1 to 0.16 degrees (about 54m to 34m distance) the error is not greater than 1.5m for displacement of  $\pm 20$ , which in our setup corresponds to about 380 pixels. Therefore, we can deduce that vertical displacements do not affect the accuracy of the system as much as the horizontal ones. The same effect was evaluating by our experimental setup.

### b. Alignment Errors

Taking into consideration eqs. (2) and (3) we can easily derive that from the three head angles the most dominant in affecting the accuracy of the measurements is the vergence. The other two affect the absolute 3D coordinates of a point, but do not distort the shape of the measured object. Let us consider the case of Figure 8, where one of the two cameras declines from the perpendicular position. In that case, when the two cameras are verging the case is identical to the one presented in Figure 4b and studied with eq. (5) and Figure 5. Indeed, in our experiments we observed same effect as in the case of the horizontal displacement of the principal point, as was expected.



**Fig. 8.** Declination of the left camera from the perpendicular position.

### c. Mechanical Errors

These are the errors due to transmission of the belts of the head. In our experiments the motor of each separate joint was moved by a constant angle, the encoders were read and the angle difference on the encoder was calculated. Images were taken before and after the movement. Any movement of any joint is directly measurable on the rectified image, with the value of the angle being:

$$\phi = \arctan(\text{disp} / f) \quad (6)$$

where  $\text{disp}$  is the disparity (horizontal or vertical, depending on the case) from one view to the other.

In order to extract eq. (6) we considered that the two lines, formed by the sensor and the observed point in the two consequent views, are parallel. This assumption can be made due to the fact that the sensor size and the focal length are insignificant compared to the distance of the observed point (about 4m). We measured the mechanical errors according to eq. (6) in our setup and we find them almost zero. Due to the fact that we did not apply a method to estimate the disparity with sub-pixel accuracy, it can be concluded that the error is mainly due to image quantization rather than the motors.

#### d. Overall System Efficiency

In order to test the accuracy of the whole system (the head employing image processing) we performed an unsupervised junction tracking. The cube was placed in several positions. The relative distances from one cube corner to the other were estimated as in the previous experiments. The tolerance that was set in order to take a measurement was  $\pm 1$  pixel. The experiments resulted a maximum error of about 1.5cm at a distance of 1.1m. The main error factor in this case was false correspondence, as was observed during the experiment. The other crucial factor is the lighting conditions, due to which the tracked point may be several pixels away from the exact junction.

## 5 Conclusions

The main factors that affects the measurement with such an active vision stereo head are small variations in the vergence angle or small horizontal deviations of the principal point. On the other hand variations in pan and tilt and vertical deviations of the principal point do not significantly affect the measurement. When the image processing is used, i.e. unsupervised junction determination, other factors such as false correspondence or lighting conditions are also important. From the experiments carried out it can be deduced that the overall system can be very accurate in the absence of false correspondences, which may lead to large errors.

**Acknowledgments.** The work presented in this paper has been supported by the Esprit project ROBVISION (EP-28867).

## References

1. Horn, B. K. P.: Robot Vision, MIT Press, Cambridge MA (1986)
2. Haggren H., Mattila, S.: 3-D Indoor Modeling from Videography, in Proc. SPIE, Vol. 3174, San Diego (1997) 14-20
3. Lim, H. S., Binford, T. O.: Curved Surface Reconstruction using Stereo Correspondence, in DARPA88, Cambridge, MA (1988) 809-819

4. Barnard, S. T., Thompson, W. B.: Disparity Analysis of Images. *IEEE Trans. PAMI* **2** (1980), 333-340
5. Mohan, R., Medioni, G., Nevatia, R.: Stereo Error Detection, Correction and Evaluation. *IEEE Trans. PAMI* **11** (1989) 13-120
6. Tsai, R. Y.: 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* **5** (1987)
7. Heikkila, J., Silven, O.: A Four-step Camera Calibration Procedure with Implicit Image Correction, in *CVPR '97, San Juan, Puerto Rico* (1997) 1106-1112
8. Zhao, W., Nandhakumar, N: Relative Influence of Camera Alignment Errors on 3D Stereoscopic Measurements, in *ACCV95, Singapore* (1995)
9. Zhao W., Nandhakumar, N: Effects of Camera Alignment Errors on Stereoscopic Depth Estimates. *PR* **29** (1996) 2115-2126
10. Gasteratos, A., Martinotti, R., Metta, G., Sandini, G.: Precise 3D Measurements with a High Resolution Stereo Head, in *IWISPA 2000, Pula, Croatia* (2000) 171-176
11. Vincze, M., Ayromloy, M., Beltran, C., Gasteratos, A., Hoffgaard, S., Madsen, O., Ponweiser W., Zillich, M.: A System to Navigate a Robot into a Ship Structure in *Lecture Notes in Computer Science*, Schiele, B., Sagerer, G., (Eds.), Vol. 2095, Springer-Verlag Berlin-Heidelberg (2001) 268-283
12. Gasteratos, A., Beltran, C., Metta, G., Sandini, G.: PRONTO: A System for Mobile Robot Navigation via CAD-Model Guidance, to appear in *Microprocessors and Microsystems*.
13. Gasteratos, A, Sandini, G.: On the Accuracy of the Eurohead. TR-2/00, LIRA-Lab, DIST, University of Genova, Genova, (2000)