# ACTIVE VISION SYSTEM FOR TRACKING TWO TARGETS IN 3D

**Luis Perdigoto, Diogo Marques, João P. Barreto, Helder Araújo**

*Institute of Systems and Robotics*
*Dept. of Electrical Engineering–Polo II*
*University of Coimbra*
*3030 Coimbra, Portugal*

Abstract
This paper describes a new active vision system designed to implement an algorithm for simultaneously tracking two targets while reconstructing their 3D trajectory. The tracking algorithm is summarized. The system design and construction are then described and the main performances and characteristics of the system are presented. The paper ends with the description of the use of the system in a previously developed application for the binocular tracking of one target.

Keywords: active vision, tracking

## 1. INTRODUCTION

Tracking of people and events is an application of computer vision systems. In these applications, often attention must be focused in one or more targets (Eklundh *et al.*, 1995). Active vision systems accomplish this by providing the camera with enough degrees of freedom to keep the target in the centre of the image(Murray *et al.*, 1993). When two cameras are used to track the same object, it becomes possible to reconstruct its 3D trajectory (Batista *et al.*, 1998), based on image information and on the system's position, usually measured with motors' encoders.

Two independent cameras, with independent degrees of freedom, can be used to track two different targets. If one additional fixed wide-angle camera is used, two targets can be selected among other moving objects in the scene. If both objects are visible in the images of both cameras, 3D reconstruction is possible.

Problems arise when target occlusion occurs. In a previous paper (Barreto *et al.*, 1999) an active vision system was proposed to track two independent targets and to deal with the possibility of occlusion.

This paper describes the construction of a system to implement that idea.

## 2. TRACKING TWO TARGETS IN 3D

In (Barreto *et al.*, 1999) it is assumed that each of two cameras will track one target while keeping the second target visible. For that, each camera must be supported by a mechanism with two degrees of freedom: rotations in pan and in tilt. It can be shown (Barreto *et al.*, 2000) that geometry simplifications can be used in the tracking algorithms if the motion range of the dependent degree of freedom is reduced. Thus, for tracking objects moving in the horizontal plane, a pan-tilt configuration (Fick's model), where the tilt axis is dependent on the pan rotation, is preferable to a tilt-pan configuration.

An obvious approach is to use the two pan-tilt units at two different fixed locations. Each of the cameras focuses its attention on one of the targets (keeping it on the centre of its image). If the other target is also visible in the periphery of the image, 3D trajectory reconstruction is possible for that target. When occlusion occurs the disparity between the target in the image center and the peripheral target becomes null, seriously affecting tracking robustness and 3D recovery. Thus, the system must try to maximize the disparity of the targets in the image and guarantee that 3D reconstruction is always possible.

The use of multiple cameras, selecting at each instant the two cameras with best viewing angles, could be a solution, but would imply a greater complexity of the system and might not be an economical solution. The proposed active vision system introduces two additional degrees of freedom that permit the positioning of the pan-tilt units during operation, in order to control the disparity in the images. These degrees of freedom are achieved by mounting the pan-tilt units on the extremities of two concentric rotating arms (fig. 1), thus permitting positioning along a circular path.

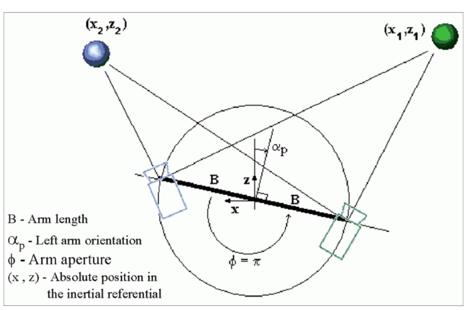


Figure 1. Vision System Configuration

Any of the pan-tilt units can be placed, by controlling the arm rotation angle, in a position where the disparity between the central and the peripheral target is maximized. However, this position is equal for both cameras, i. e., their optical centers would have to be coincident. Although the disparity can not be simultaneously maximized it can be kept equal for both cameras. This symmetry is useful to increase the robustness of the visual processing and guarantees that whenever the two targets are seen by one of the cameras the same happens with the other. Consider figure***, where $\alpha_p$ denotes the angular position of camera 1's arm and $\phi$ the angle between them. Equation 1 yields the relation between $\alpha_p$ and $\phi$ that has to be satisfied in order to achieve equal disparities in both cameras.

$$\begin{aligned} K_w * \sin(\phi) = (K_x(1-\cos(\phi)) + \\ + K_z \sin(\phi)) \sin(\alpha_p) + (K_z(1-\cos(\phi)) - \\ - K_x \sin(\phi)) \cos(\alpha_p) \end{aligned} \quad (1)$$

$$\begin{aligned} K_x &= (\rho_1^2 - B^2)x_2 - (\rho_2^2 - B^2)x_1 \\ K_z &= (\rho_1^2 - B^2)z_2 - (\rho_2^2 - B^2)z_1 \\ K_w &= B(\rho_1^2 - \rho_2^2) \\ \rho_1 &= \sqrt{x_1^2 + z_1^2} \\ \rho_2 &= \sqrt{x_2^2 + z_2^2} \end{aligned}$$

Thus,the equal disparity restriction yields one equation for two variables. The remaining degree of freedom can be used to control the range of the disparity. So, the system has to continuously estimate the most adequate values for $\alpha_p$ and $\phi$, establishing a trade-off between maximizing the disparity and keeping the baseline (distance between the two cameras) not too small, since that would affect the 3D reconstruction precision.

One possible way to simplify the control strategy is to always maximize the baseline, making $\phi = \pi$. In this case, equation 1 is simplified yielding equation 2.

$$\tan(\alpha_p) = -\frac{(x_1^2 + z_1^2 - B^2)z_2 - (x_2^2 + z_2^2 - B^2)z_1}{(x_1^2 + z_1^2 - B^2)x_2 - (x_2^2 + z_2^2 - B^2)x_1} \quad (2)$$

Also as shown in (Barreto *et al.*, 1999), in this configuration, occlusion only occurs when both targets are aligned with the center of rotation of the arms. When this happens, it is necessary to exchange the positions of the arms, making a 180°

saccade, so that the left camera always tracks the left target and the right camera the right target.

The system will, thus, be constituted of two pan-tilt units and a central platform with independent control of the position of the two arms.

## 3. THE PAN-TILT UNITS

The cameras used are two JAI CV-M536 micro-head monochromatic cameras. These are small cameras and the the pan-tilt units only have to support the remote heads, which are cylinders of $1,7cm$ diameter by $7cm$ of length, weighing only $15g$ (lens included). The cameras outputs are monochromatic CCIR video signals that are genlocked.

DC motors are used as actuators. This type of motor presents a smooth motion, which makes it appropriate for pursuit applications, low cost and very good velocity and acceleration specifications. Their main disadvantage lays in their small torque, which can be overcome by the use of gearheads. The gearheads will, however, introduce backlash and friction. To provide position feedback, encoders are used on the backshaft of the motors. The resolution of these encoders is thus multiplied by the reduction of the gearheads.

The design of the pan-tilt units was made considering the following goals:

- Rotation around the optical centre;
- High performances;
- Small size and weight;
- Modularity;

For visual processing purposes it is important that the pan and tilt rotation axes intercept the optical center of the camera, thus avoiding any translation component in the camera motion.

The desired values for the dynamic performances of the pan-tilt units were based on those of the human eye and those presented by similar systems (Brooks *et al.*, 1997; Scassellati, 1998). In order to obtain these performances, small size and weight are required for the mobile parts of the units. Furthermore, since the units will be moved by the rotating arms platform, the overall weight of these units should be minimized.

It was also desirable that this units can be easily used in other applications or systems, thus requiring a modular design. The overall dimensions of the units are crucial for this requirement. The units were designed to have a solid base that could be easily applied on any surface and to contain all the movements of its mechanical parts in a small volume.

Several design configurations were considered and their dynamic performances predicted by calculating the approximate moments of inertia of the moving parts. The final design led to a $300\,g$ unit with overall dimensions of $9.5 \times 6.5 \times 4.5\,cm$, which is shown in fig. 2.

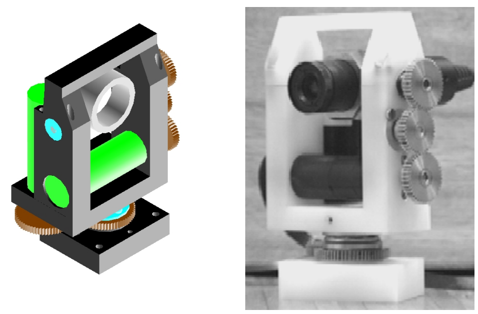

Figure 2. Pan-Tilt unit.

The camera is supported by a ring located in the middle of the tilt axis. The tilt motor is located below this axis, and parallel to it. The motion of the motor's shaft is transmitted to the tilt axis through 3 spur gears: one attached to the motor shaft, one attached to the axis and one connecting the other two. Alternatively only two gears could have been used but with much larger diameter. The axis/ring piece is made out of steel but the frame structure that supports this axis, the tilt motor, and the gears is constructed from a lightweight yet resistant plastic.

All this structure lays on top of the fixed base and rotates along the pan axis. This axis is aligned with the center of the camera supporting ring. The Pan motor is coupled to the tilt frame structure and is mounted upside down, with a gear on its shaft. This gear (and thus the motor and the entire structure) rotates around another gear coupled to the fixed base, achieving the pan rotation.

In order to minimize friction ball bearings are used to support all the rotation axes.

Two Minimotor 1524T 024SR DC motors with a gear reduction of 102:1 are used in each of the units.

## 4. ROTATING ARMS PLATFORM

In the design of the rotating arms platform, factors like modularity, size or weight are not as important as they were in the design of the pan-tilt units. Also, the expected dynamic performances can not be too high, since this mechanism has to rotate a load located far away from the rotation centre (each pan-tilt unit is mounted at a distance of $30\,cm$ from the centre). In normal pursuit, when following a target with a smooth movement in the scene, the required velocity and acceleration will not be very large. Higher performances are only required when there is an alignment of the targets with the centre of rotation. In that situation, as previously described, it is necessary to make a $180°$ saccade, exchanging the positions of the arms. In order not to compromise the tracking, this saccade must be performed as quickly as possible.

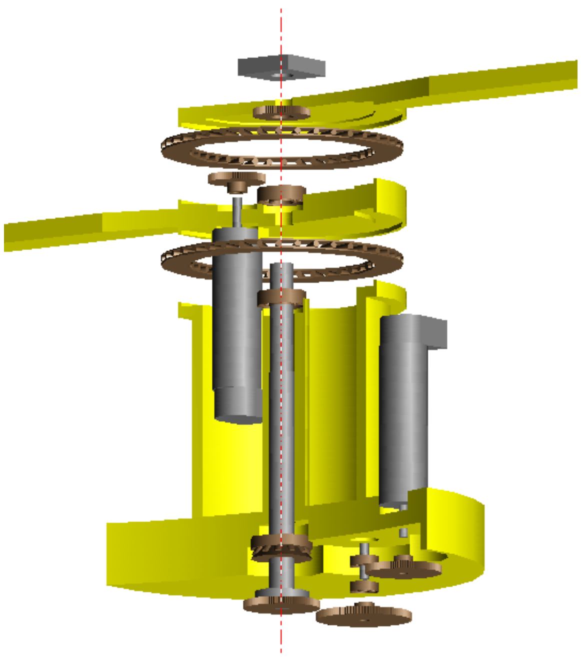

Figure 3. Rotating arms platform assembly

Figure 3 depicts a scheme of the assembly of the rotating arms platform.

It can be seen that both arms rotate around a common central axis. The top arm is attached to the axis, which is rotated by the external motor (orientation motor). Fixed to the bottom arm is a motor with a gear coupled to its shaft. This gear rotates around the central axis in a manner similar to the pan rotation of the pan-tilt units. Thus, the inner motor controls the angle between the arms. Once again, the axis is supported by ball bearings.

The motors used in this platform are two Maxon A-Max ∅32mm, 20W DC motors with a 66:1 reduction.

Figure 4 shows a photo of the actual system.

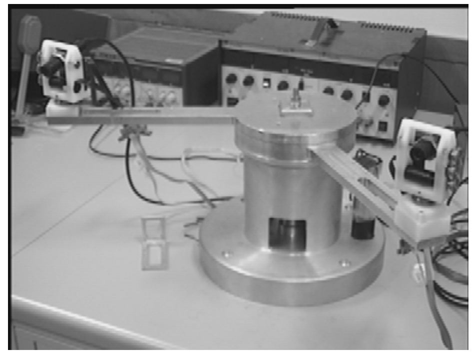

Figure 4. The System

## 5. SYSTEM INTEGRATION

A PC (with a Pentium III 450MHz processor) is used to control the system. Installed in the PC is a Matrox Meteor/RGB frame grabber board and a PMC DCX-AT200 Motion Control Motherboard.

The Meteor/RGB board simultaneously adquires the monochromatic images from both cameras (which are genlocked) using two of the RGB component inputs. The frame acquisition rate is 25Hz, or 50Hz if even and odd fields are processed sequentially.

The DCX-AT200 board has 6 DCX-MC200 servo modules. When installed in the board, each of these servo modules can be used to control one of the 6 DC motors of the system. The board and modules implement, for each motor, a low-level 4KHz position control, based on a PID compensation filter with velocity and acceleration feedforward, whose reference is supplied by a trajectory generator. Communication between the PC and controller board is made every 4ms.

The control of the system can thus be divided in 3 loops: a inner 4KHz loop that performs low-level motor control, a 250Hz loop that performs high-level motor control and a top level 25Hz (or 50Hz) loop that performs visual processing.

The command signal from the servo modules must be amplified to be applied to motor. For the pan-tilt motors linear amplification is used. For the central platform motors, due to their larger power, PWM amplifiers are used.

Figure 5 summarizes the control scheme of the system.

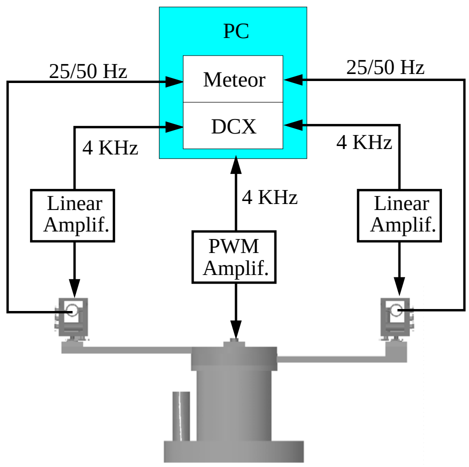


Figure 5. Control scheme

## 6. SYSTEM CHARACTERIZATION

In order to evaluate the system several tests were made. Table 1 summarizes the main characteristics of the system.

| **Pan-Tilt units** | Pan | Tilt |
|---|---|---|
| Max. vel. | $480°/s$ | $520°/s$ |
| Max. Accel. | $650\,rad/s^2$ | $600\,rad/s^2$ |
| Resolution ($CPT$) | 208896 | 208896 |
| Rotation amplitude | — | 102° |
| Average backlash | 4° | 5° |

| **Arms Platform** | Orientation | Angle |
|---|---|---|
| Max. vel. | $580°/s$ | — |
| Max. Accel. | $268\,rad/s^2$ | — |
| Resolution ($CPT$) | 132000 | 132000 |
| Rotation amplitude | — | $\sim 330°$ |
| Average backlash | 1.7° | 1.5° |
| 180° saccade time | $0.43\,sec$ | — |

Table 1. Summary of the system's performances

Maximum velocities and accelerations were obtained applying the nominal voltage to each of the motors in open loop. The obtained values for the pan-tilt units are very good although they are slightly below the initially intended performances. As for the central platform the performances, the are very satisfactory when considering the dimensions and mechanical loads involved (the pan-tilt units were mounted on the extremities of the arms). No open-loop dynamic characterization was made for the angle rotation, since its movement amplitude is very restricted, when considering the inertial loads involved.

In terms of ranges, the pan and orientation rotations have no physical limitation except for the length of the cables. The tilt and angle rotations were designed to be suitable for their purposes.

The main problem with this system is the backlash, which can essentially be attributed to the motors' gearheads.

The encoders have extremely good resolution, which might seem unnecessary when considering the backlash. Although a good precision positioning can not be achieved, this order of resolution is useful for velocity control.

The duration of a 180° saccade is also an important indicator for the orientation rotation (for reasons already pointed out) and was also measured, with very good results.

## 7. BINOCULAR TRACKING OF ONE TARGET

The implementation of the tracking of two targets algorithm is still in progress, but the system has already been used in other applications. As a first test, an existing algorithm for binocular tracking of one target, already used in another system (Batista *et al.*, 1997), was adapted to this system. In order to enable the implementation of the algorithm, a few simple adaptations had to be made in the configuration of the system.

The used algorithm assumed 3 degrees of freedom in the system: simultaneous pan and tilt of both eyes and symmetric vergence. If tilt-pan units were used, instead of pan-tilt, it was possible to implement the algorithm using the platform orientation as Pan, keeping the angle as 180°, using equal tilt in both units and symmetric pan as vergence.

To get the tilt-pan units, the pan-tilt units were mounted on L-shaped adaptors, making a right angle with their usual position. Another change in the configuration is that the L-shaped adaptors are not mounted on the extremities of the arms, reducing the baseline to $40\,cm$. The new geometry is shown in figure 6.

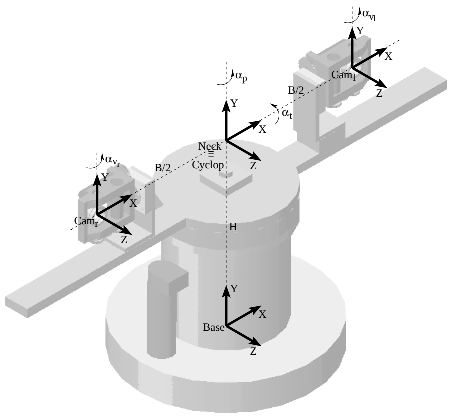


Figure 6. Binocular tracking configuration

The pan and tilt degrees of freedom are used so that an imaginary cyclopean eye, placed in the mid point between the two cameras, can track the target, while the Vergence degree of freedom is used to place the target in the centre of both images.

Target velocity in the image is estimated using optical flow. The flow is calculated on the points of the image where there has been movement of the target, considering a constant flow model on those points. To segment these points, the present image is subtracted to the previous one and a mask is created with the pixels where that difference is bigger than a given threshold.

However, since this is an active vision system, there is motion of the cameras (egomotion), which generates a velocity field in the image that must be eliminated before the mask is calculated.

Since the geometry of the system is known, it is possible, based on the information of the encoders,to warp the pixels of the previous image in order to compensate for the flow induced by the egomotion. However, when there is a translation component in the movement of the cameras (which happens in this case), parallax effects have to be accounted for. To be rigorous (and have no approximations),it is necessary to know the depth of every point in the image to calculate the required displacement. That is not feasible, so, a constant depth is assumed (which has given acceptable results, when the depth range in the scene is limited).

After the target velocity in both images is estimated, the velocities of the three degrees of rotation (fig. 6) will be:

$$\alpha_p = \frac{v_{x_l} + v_{x_r}}{2F}$$

$$\alpha_t = \frac{v_{y_l} + v_{y_r}}{2F}$$

$$\alpha_{v_l} = -\alpha_{v_r} = \frac{v_{x_l} - v_{x_r}}{2F}$$

$v_{x_l}$, $v_{x_r}$, $v_{y_l}$ and $v_{y_r}$ are the horizontal and vertical components of the velocities in left and right images. $F$ is the focal length of the cameras measured in pixels (in this case, the vertical and horizontal pixel scale factor is the same).

The results obtained with this application were not as good as the ones achieved in the previously mentioned system, for which the algorithm had been developed. The backlash of the system is particularly harmful to the estimation of the egomotion, deteriorating the algorithm at its essence and compromising its results.

## 8. CONCLUSIONS

In this paper we describe a binocular active vision system with special characteristics, namely that the angle between the arms supporting the cameras can be changed. This system was specifically developed to track binocularly two targets. The full development of the application is ongoing. New tracking algorithms are being developed to operate in this system, in an effort to take advantage of its properties and to be less sensitive to its problems. One of the main challenges is to deal with both the backlash and the effects of the parallax.

## 9. REFERENCES


Barreto, J., J. Batista and H. Araujo (2000). Mathematical analysis for visual tracking assuming perspective projection. In: *Proceedings of SIRS'2000-8th International Symposium on Intelligent Robotic Systems*. University of Reading, U.K.. pp. 319–328.

Barreto, J., P. Peixoto, J. Batista and H. Araujo (1999). Tracking multiple objects in 3d. In: *IROS'99-IEEE/RSJ International Conference on Intelligent Robots and Systems*. Kyongju, Korea.

Batista, J., P. Peixoto and H. Araújo (1997). Real-time vergence and binocular gaze control. In: *IROS97–IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. Grenoble, France.

Batista, J., P. Peixoto and H. Araujo (1998). Real-time active visual surveillance by integrating peripheral motion detection with foveated tracking. In: *Proc. of the IEEE Workshop on Visual Surveillance*. pp. 18–25.

Brooks, A., G. Dickins, A. Zelinsky, J. Kieffer and S. Abdallah (1997). A high performance platform for real time active vision. In: *Proceedings of the First International Conference on Field and Service Robotics*.

Eklundh, J. O., K. Pahlavan and T. Uhlin (1995). The kth head-eye system. In: *Vision as a Process*. Chap. 15, pp. 237–259.

Murray, D., P. McLauchlan, I. Reid and P. Sharkey (1993). Reactions to peripheral image motion using a head/eye plataform. In: *Proc. of IEEE International Conference on Computer Vision*. pp. 403–411.

Scassellati, Brian (1998). A binocular, foveated active vision system. TechReport 1628. MIT AI Lab. Cambridge, MA.