



HAL
open science

An Effective Active Vision System for Gaze Control

Yann Ducrocq, Shahram Bahrami, Luc Duvieubourg, François Cabestaing

► **To cite this version:**

Yann Ducrocq, Shahram Bahrami, Luc Duvieubourg, François Cabestaing. An Effective Active Vision System for Gaze Control. 4th International Symposium on Advances in Visual Computing, ISVC'2008, Dec 2008, Las Vegas, United States. pp.267-277, 10.1007/978-3-540-89646-3_26 . hal-00521104

HAL Id: hal-00521104

<https://hal.science/hal-00521104>

Submitted on 25 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An Effective Active Vision System for Gaze Control

Yann Ducrocq¹, Shahram Bahrami¹
Luc Duvieubourg², and François Cabestaing²
yann.ducrocq@eipc.fr, shahram.bahrami@eipc.fr
luc.duvieubourg@univ-lille1.fr, francois.cabestaing@univ-lille1.fr

¹ École d'Ingénieurs du Pas-de-Calais, Département Automatique, Campus de la Malassise, 62967 Longuenesse Cedex - FRANCE,

² Laboratoire LAGIS - UMR CNRS 8146, Université des Sciences et Technologies de Lille, Cité Scientifique - Bâtiment P2, 59655 Villeneuve d'Ascq - FRANCE

Abstract. This paper presents the performances of an active vision system that mimic the human gaze control. A human can shift his gaze either by quickly moving his fixation point or by keeping a moving target in the fovea (high resolution). These two visual phenomena are called saccadic and smooth pursuit eye movements respectively. In order to mimic this human behavior, we have developed a novel active vision system based on a particular stereo-vision setup. It is composed with one camera, one prism and a set of mirrors. To point the field of view of the sensor at a target, the prism is rotated about its axis by a motorized stage. The system is designed for fast and accurate dynamical adjustments of gaze. To study the mechanical performances of our active vision system we have used three different but classical input signals. A step signal that simulates a change of target (saccadic eye movement), a velocity ramp and a sinusoidal signal that simulate a moving target (smooth pursuit). Whatever the input signal, the objective is to maintain the target in the middle of the image. The experiments demonstrate the efficiency of our vision sensor, in term of dynamical properties and measurement accuracy.

1 Introduction

Since the works of Aloimonos [1], Bajcsy [2], and Ballard [3] on active vision, most researchers attempt to design sensors for robotic heads that imitate as much as possible the human visual capacities. But, the first active vision systems built in the early nineties were relatively slow and limited in scope [4]. The reason is because these first generation of active vision systems were constructed with relatively large cameras and heavy mechanical setup. These drawbacks lead to poor dynamical characteristics. As cameras become smaller and lighter, sensors on robot heads can reach dynamics comparable to the human visual system. This has activated the research in the three different fields i) image processing algorithms, ii) control systems capable of exploiting the potential characteristics

of the mechanisms and iii) mechanical setup of the active vision sensors. This paper mainly addresses the latter issue.

Many projects in active vision relate to vision heads for mobile robot navigation. In the first part of their paper, Samson and co [5] describe many recent robotic heads with active vision systems. The main characteristics that a good active vision system has to have are i) the accuracy of the 3D information by stereopsis, ii) compactedness, iii) good dynamical performances for fast reaction to a change in the environment and iv) high speed tracking.

Robotic vision systems are often biologically inspired and are composed of two cameras mounted on a head, with the goal of coming close to the performances of the human. They have several degrees of freedom (DOF) which permit to adjust and control the pan (left/right), tilt (up/down) and vergence [4, 6] and sometimes the zoom, focus, aperture or baseline [4, 7, 5]. Pan and tilt permit gaze orientation and vergence leads to observe objects of the scene with a minimum disparity while keeping them near the fovea. Raising the number of DOF leads to system capabilities closer to the human ones, but that is at the expense of adjusting and controlling the axes, the accuracy of the data because mechanical clearance raises, the compactedness, the cost and the load of the system. Most authors [5, 8, 9] prefer a 4 DOF system with common pan and tilt axes and two independent vergence axes. That leads to a compromise between compact design, accuracy of data and flexibility.

Another domain where active vision systems are widely used is driver assistance. Clady [7] developed a vision system which used a PTZ (Pan Tilt Zoom) camera coupled with a fixed large focal camera. His tracking method allowed him to follow a moving object, "the most dangerous car", in a sequence of road traffic images. In order to follow it, he controlled the tilt, pan and zoom of the camera to obtain a constant resolution of the car. The position of the optical axis could move $\pm 30^\circ$ horizontally and $\pm 15^\circ$ vertically with an angular speed of $150^\circ/s$. In [10], Pellkofer and Dickmanns developed an approach of an optimal gaze control system for autonomous vehicles. Their goal was to determine the direction to observe and to predict the future situation. In [11], they proposed a vision sensor composed of four cameras with different focal lengths. Their system was mounted on a pan-tilt camera head and was designed in order to perform smooth pursuit and saccadic movements. More recently, the Team AnnieWAY in the DARPA Urban Challenge [12] also developed an active vision system for gaze direction to important areas of the scene. In table 1 we have summarized some major works, including our active vision system, and their principal characteristics.

However, most of the active vision systems mentioned before are mechanically complex, big and heavy. In this case, inertia does not permit movement of the whole system sufficiently fast and precise. In order to overcome this drawback, we have developed an active stereo vision system able to quickly change its gaze, thanks to a very compact and light motorized part.

Our project deals with the difficult problem of road safety and specially in driver assistance. Our primary goal with our active vision system is to implement

| Author | Year | Degree of freedom | Adjustable features | Speed (<i>degree/s</i>) | Application |
|--------------------|------|-------------------|----------------------------|---------------------------------------|---------------------------------|
| Pahlavan [4] | 1992 | 13 | gaze, baseline zoom, focus | 180 | robotic head KTH |
| Wavering [9] | 1993 | 4 | gaze, vergence | pan:660 tilt:1000 vergence:1830 | robotic head TRICLOPS |
| Santos-Victor [13] | 1994 | 4 | gaze, vergence | 180 | Medusa |
| Sharkey [8] | 1998 | 4 | gaze, vergence | pan:425 tilt:680 vergence:560 | robotic head |
| Samson [5] | 2006 | 4 | gaze, vergence, baseline | pan : 1950 tilt : 1350 | Not any but several possibility |
| Kuhnlenz [14] | 2006 | 2-7 | gaze, vergence | 8400 | robotic head |
| Clady [7] | 2001 | 2 | gaze, zoom | 150 | driver assistance |
| Pellkofer [11] | 2003 | 2 | gaze | - | driver assistance |
| Duvieubourg [15] | 2005 | 1 | gaze | 70 | driver assistance |

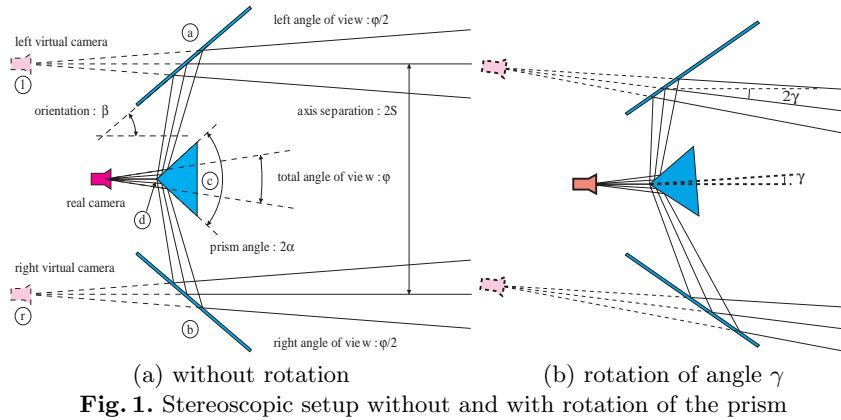
Table 1. Some active vision systems

a long distance, 100 meters and above, vision-based lane tracker in order to maintain the lane always in the center of the image, even in winding road. The final goal is to develop a strategy of observation of a road traffic scene, as human does when he drives.

The next paragraphs of the paper are organized as follows. Section 2 presents an overview of our active stereo vision system. Section 3 describes the operations needed for controlling the gaze of the active vision system. In section 4, the performances are demonstrated with three classical experiments in control command. Finally, conclusion and outlooks are given in section 5.

2 Sensor description

A schematic top view of the stereoscopic sensor is presented in figure 1(a). The stereoscopic sensor is composed of a single camera, of two lateral plane mirrors, marked as \textcircled{a} and \textcircled{b} in figure 1, and of a central prism with two planar reflective surfaces, marked as \textcircled{c} . Thanks to the prism and the lateral mirrors \textcircled{a} and \textcircled{b} we obtain the left and right images of the stereo pair onto both halves of the imaging surface of the camera. For a specific pose of the lateral mirrors and for a rotation angle γ equal to zero, we have shown that the optical axis of a virtual camera is parallel to the optical axis of the real camera [15]. When the prism is in rotation, the optical axes of both virtual cameras remain parallel, but the baseline is slightly modified [16]. Figure 1(b) shows that for a rotation γ of the prism the orientation of the field of view is twice the angle γ , but in the opposite direction. In fact this orientation of the stereoscopic setup is similar to what happens when we change the direction of our gaze without rotating the head. The camera sensor resolution used for these experiments is 1024 x 1280 pixels.



However, the top and lower parts of the image are removed - because we are not interested in those parts - and we get a 440 x 1280 pixels image. An example of stereoscopic images (two half images) is shown in figure 2.

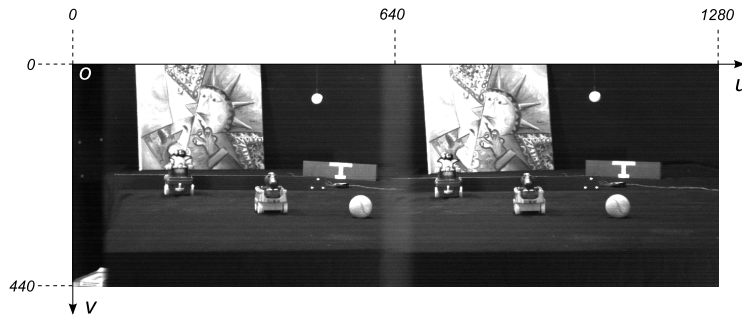


Fig. 2. Stereoscopic image

A photograph of our laboratory prototype is shown in figure 3

3 Gaze Control

3.1 Different Gaze Movements

When a driver observes the scene in front of his car, his eyes make various and rapid movements. To get precise informations of the observed objects, human always shifts his gaze in the center of the retina, because in this region of the eyes the resolution is higher. This center part is called the fovea, that is also the center of his field of view. These gaze shifts are the combination of three basic movements [17, 18] :

- Saccades : quick and simultaneous movements of both eyes in the same direction,

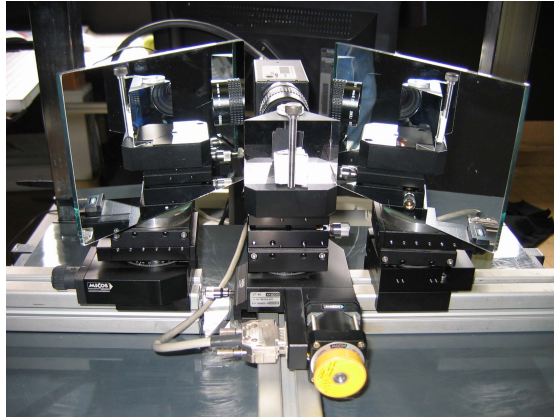


Fig. 3. Photo of our laboratory prototype

- Smooth pursuit : this is the ability of the eyes to smoothly follow a moving object,
- Vergence : when we observe objects at different distances, eyes move in opposite directions such as sighting axes converge to the objects.

We remember that our final goal is to imitate, with our sensor, the strategy of observation of a driver in a real road traffic scene. Before developing this gaze control strategy, we have to evaluate the performances of our sensor. The gaze control strategy is out of the scope of this paper and will be published in a near future. With our system, we have shown that the rotation of the prism allows a vision similar to the human gaze shift, except the vergence which can not be implemented on the sensor.

3.2 The Gaze Control Platform

Our active vision host system is a personal computer, equipped with a piccolo frame grabber. The stepper motor which is coupled with a reducer (288000 steps per revolution) is controlled by a GCD 93-70 Phytron module. The maximum angular speed is near $70^\circ/s$ and the accuracy near 0.001° with this material.

Compared to human capacities whose eyes movements angular velocity can reach $900^\circ/s$ [19], the dynamic performance of our sensor is low. However, in the particular application to driver assistance, our sensor is able to shift the field of view quickly enough and track vehicles precisely. It has one degree of freedom that permits rotation about the vertical axis but it does not allow rotation about any horizontal axis. However, as for our application the gaze control is about pan axes, it is not necessary to have another degree of freedom. The vergence and the baseline are adjusted manually by a fine tuning of mirrors locations and orientations. The optical axes of the cameras are parallel [16]. Finally, the baseline has a fixed value, computed with respect to our application to obtain a good compromise between compactedness and precision of the 3D map.

The photography of our stereo active vision rig used for the experiments is shown in figure 3 and the block diagram of the closed loop presented in figure 4. In this control architecture, the loop deals with determining the angular position of the prism using visual information as feedback. The angular position of the prism, is controlled in three steps :

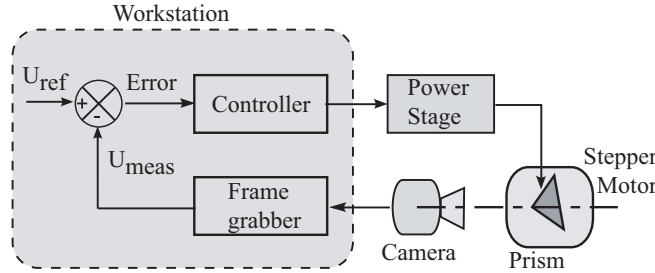


Fig. 4. Active vision rig

- Image acquisition and processing : template matching yields the center of mass of the target U_{meas} , as well as its speed and its distance to the camera,
- Error signal, that is the computed distance between image center U_{ref} and the center of mass of the target U_{meas} ,
- Control law, the error signal is used by a PID controller to compute the command signal sent to the stepper motor.

To evaluate the performance of our active vision rig we perform numerous tests similar to the gaze movement of a human. Then, to evaluate the saccadic movement we change instantaneously the position of the target in the scene, this corresponds to a step input signal. The smooth pursuit movement can be evaluated by using a known movement of the target, this is similar to a ramp or an harmonic input signal. The parameters of the PID controller were designed experimentally, regardful to obtain a good compromise between position and velocity control. Position control is appropriate for a saccadic movement, that is generally a fast movement, and velocity control concerns the smooth tracking of a moving target.

4 Performance evaluation

To test the validity of the described approach and the performances of the sensor, many experiments were carried out in which the prism orientation is controlled to change the gaze toward the target. The desired value U_{ref} is the center of the right-half image, and is represented by the vertical dotted line at u -coordinate = 960 in figure 5.

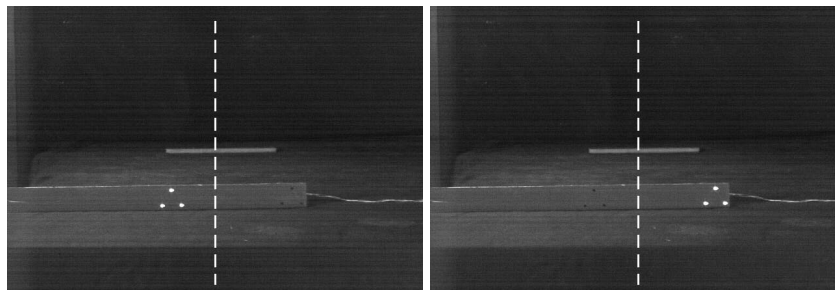
4.1 Step signal

To simulate a step signal a target swaps between two locations. The target is in fact two sets of three light emitting diodes (LED). The two sets of LEDs are arranged in a same triangle shape in order to make easier the matching process. At the beginning of the sequence, the three LEDs on the left are switched on and the three LEDs on the right are switched off (figure 5a). Then, we switch on the LEDs on the right (figure 5b) and switch off the left ones, and so on alternately. On figure 6a (resp. 6b), we can see the step response of our active vision rig. In this experiment, the gaze is shifted to bring the image of the left LEDs (resp. right LEDs) in the center of our chosen dominant half image.

On figure 7, the response of the system to a sequence of step signal inputs is shown with :

- the rotation angle of the prism (a),
- the u-coordinate of the target (b).

The controller's parameters are empirically adjusted to achieve a slightly damped response, which presents an overshoot of about 25% for this saccadic movement. The rising time is near 80 *ms* and the settling time is less than 1 second. As the steady state error is null the final position U_{meas} reaches the desired position U_{ref} , so the image of the target is exactly at the center of the half-right image.



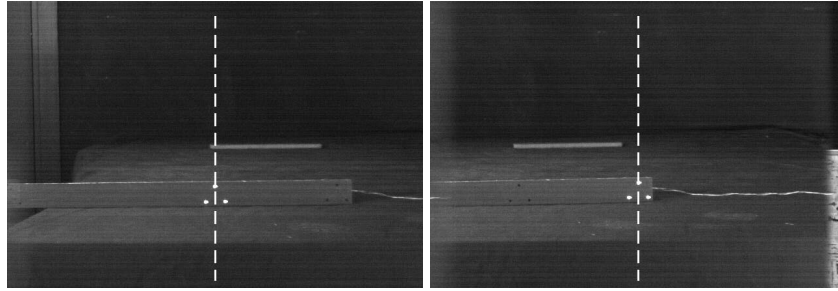
(a) Left leds

(b) Right leds

Fig. 5. Left and right LEDs for step signal. The vertical dotted line, at u-coordinate=960, is the target value.

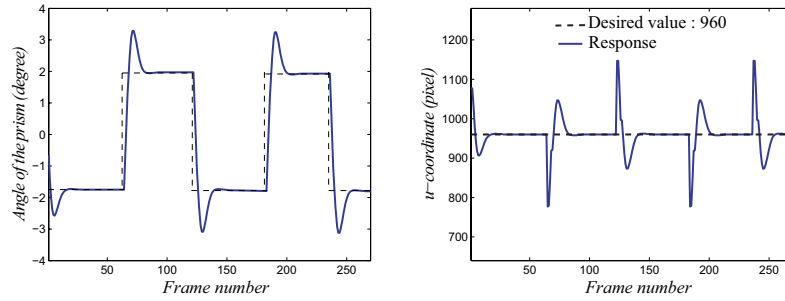
4.2 Velocity ramp signal

In this experiment, the target is a ball placed on a pitched handrail. First the ball is motionless and the prism angle is around 2° in order to obtain the image of the target, i.e. the ball, at the center of the right-half image (fig. 8, frames 1-17). Then, the ball rolls on its rail and the gaze control tracks it during its descent (fig. 8, frames 17-60).



(a) Gaze pointed on the left LEDs (b) Gaze pointed on the right LEDs

Fig. 6. Gaze controlled (The vertical dotted line at u-coordinate=960 is the target value).



(a)

(b)

Fig. 7. Sequence of step responses of our active vision rig. The gaze swaps between approximately -2° and $+2^\circ$ (a). Figure (b) shows the position of the target over frame, i.e. time

When the ball begins to roll, it has a constant acceleration $a(t) = \ddot{x}(t)$, with $x(t)$ the vertical coordinate of the position over time of the ball in the 3D space. This coordinate is measured by the v-coordinate of the ball in the images. The velocity is $v(t) = v(t_0) + a(t).t = a(t)t$ since $v(t_0)$ is null. The vertical coordinate is $x(t) = x(t_0) + v(t_0).t + \frac{1}{2}a(t).t^2$. In this experiment, as the ball velocity grows the command and the prism angle do as well. The system tries to keep the u-coordinate of the ball at the desired value : 960 (cf. image 8b, frames 17-60).

On figure 8b, the velocity ramp response of the active vision rig illustrates that the sensor is able to track the ball finely but a little steady-state error persists.

4.3 Sinusoidal signal

This experiment consists in observing a ball hanged from the ceiling as a pendulum. The pendulum is placed at 3.5 m from the sensor and the amplitude of the free oscillations is 0.7 m at the beginning and of course decreases over time.

Again, we want to keep the image of the target, i.e. the ball, in the center of the image. We have done many experiments for different lengths of the pendu-

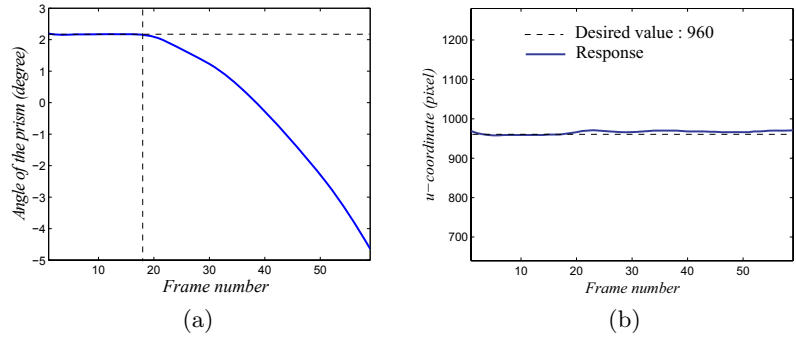


Fig. 8. Velocity ramp response of the active vision rig

lum, that correspond to different oscillations frequencies. In the small-angle case, the motion of a pendulum is approximated by a simple sinusoidal function. In this case the angular acceleration is directly proportional to the angle between the string and the vertical axis.

The active vision sensor tracks the ball with a delay which is practically constant whatever the oscillation frequencies and is independent of the angular acceleration (table 2 and figure 9). This delay is due to image processing and grabbing. The implemented control is not sufficient to track efficiently this kind of trajectory. In order to overcome this delay, we will improve the control by computing the error from a predicted value and not a measured value.

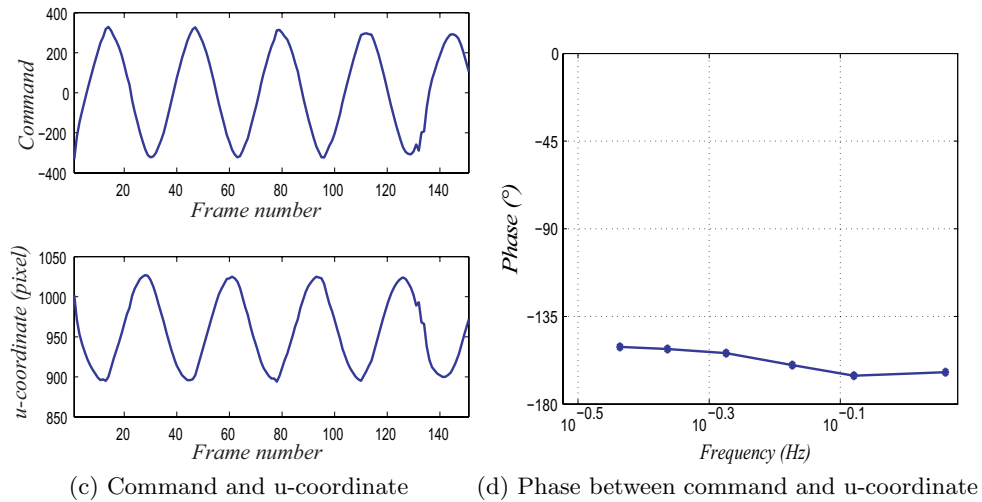


Fig. 9. Sinusoidal response of the active vision rig

| | | | | | | |
|----------------------|------|------|------|------|------|------|
| Period (s) | 0.9 | 1.2 | 1.5 | 1.9 | 2.3 | 2. |
| Phase ($^{\circ}$) | -164 | -165 | -160 | -154 | -152 | -151 |

Table 2. Sinusoidal signal period and corresponding phase

5 Conclusion

This paper presents a basic approach to gaze control with a particular active vision system. In the first part of the paper we have described our active vision system. It is built to be fast and precise, thanks to the low inertia of the mobile prism. Experiments seem to validate our approach. The step experiment shows that the system reacts enough fast and finely for a saccadic movement and the error position is null. Nevertheless, we need to improve the control settings in the case of smooth pursuit situation. In the future, we will characterize more precisely the process in order to derive a more robust control law for increasing the gaze control speed. The gaze control is one part of a more sophisticated long distance stereovision system for driver assistance. Stereovision and strategy of observation are others aspects we are studying in parallel.

Acknowledgements

The authors want to acknowledge the financial support of the French Nord-Pas de Calais Regional Council and EU FEDER under grant OBJ2-2005/3-4.1-253-7820, since the equipment have been obtained thanks to these funds.

References

1. Aloimonos, J.Y., Weiss, I., Bandopadhyay, A.: Active vision. *International Journal of Computer Vision* **1** (1988) 333–356
2. Bajcsy, R.: Active perception. *Proceedings of the IEEE, Special issue on Computer Vision* **76** (1988) 996–1005
3. Ballard, D.H.: Animate vision. *Artificial Intelligence* **48** (1991) 57–86
4. Pahlavan, K., Eklundh, J.: A head-eye system—analysis and design. *CVGIP: Image Understanding* **56** (1992) 41–56
5. Samson, E., Laurendeau, D., Parizeau, M., Comtois, S., Allan, J.F., Gosselin, C.: The agile stereo pair for active vision. *Machine Vision and Applications* **17** (2006) 32–50
6. Milios, E., Jenkin, M., Tsotsos, J.: Design and performance of TRISH, a binocular robot head with torsional eye movements. *International journal of pattern recognition and artificial intelligence* **7** (1993) 51–68
7. Clady, X., Collange, F., Jurie, F., Martinet, P.: Object tracking with a pan tilt zoom camera : application to car driving assistance. In: *Proceedings of the IEEE International Conference on Robotics and Automation, ICRA'01, Seoul, Korea, May 23-25th. Volume 2.* (2001) 1653–1658

8. Sharkey, P.M., Murray, D.W., Mclauchlan, P.F., Brooker, J.P.: Hardware development of the yorick series of active vision systems. Special Issue on Mobile Robots. *Microprocessors and Micro Systems* **21** (1998) 363–375
9. Wavering, A.J., Fiala, J.C., Roberts, K.J., Lumia, R.: TRICLOPS: a high-performance trinocular active vision system. In: *Proceeding of the IEEE International Conference on Robotics and Automation*. Volume 3., Atlanta, USA (1993) 410–417
10. Pellkofer, M., Dickmanns, E.D.: EMS-vision: gaze control in autonomous vehicles. In: *Proceeding of the IEEE Intelligent Vehicles Symposium IV'2000*. (2000) 296–301
11. Pellkofer, M., Lutzeler, M., Dickmanns, E.D. In: *Vertebrate-type Perception and Gaze Control for Road Vehicles*. Volume 6 of *Robotics Research, Springer Tracts in Advanced Robotics*. Springer-Verlag Berlin Heidelberg (2003) 271–288
12. Kammel, S., Pitzer, B., Vacek, S., Schroeder, J., Frese, C., Werling, M., Goebel, M.: Technical system description. DARPA Urban Challenge 2007, Team AnnieWAY, <http://annieway.mrt.uni-karlsruhe.de/> (2007)
13. Santos-Victor, J., van Trigt, F., Sentieiro, J.: Medusa - a stereo head for active vision. In: *International Symposium on Intelligent Robotic Systems*, Grenoble, France (1994)
14. Kuhlntz, K., Bachmayer, M., Buss, M.: A multi-focal high-performance vision system. In: *Proceedings of the 2006 IEEE International Conference on Robotics and Automation, ICRA 2006*, May 15-19, Orlando, Florida, USA. (2006) 150–155
15. Duvieubourg, L., Ambellouis, S., Cabestaing, F.: Single-camera stereovision setup with orientable optical axes. *Computational Imaging and Vision* (2005) 173–178
16. Duvieubourg, L., Cabestaing, F., Ambellouis, S., Bonnet, P.: Long distance vision sensor for driver assistance. In: *Proceedings of 6th IFAC Symposium on Intelligent Autonomous Vehicles (IAV'07)*. (2007)
17. Robinson, D.: The oculomotor control system : A review. *Proceedings of the IEEE* **56** (1968) 1032–1049
18. Bernardino, A., Silva, C., Santos-Victor, J., Pinto-Ferreira, C.: Behaviour based oculomotor control architecture for stereo heads. In: *International Symposium on Intelligent Robotic Systems*, Pisa, Italy (1995)
19. Kandel, E.R., Schwartz, J.H., Jessel, T.M.: *Principles of neural science*. Elsevier, New York (1991)