Real-Time Active Visual Surveillance by Integrating Peripheral Motion Detection with Foveated Tracking

Jorge Batista   Paulo Peixoto   Helder Araújo

ISR - Institute of Systems and Robotics

Dep. of Electrical Engineering

University of Coimbra - PORTUGAL

(batista,peixoto,helder)@isr.uc.pt

Abstract

In this paper we describe an active binocular tracking system integrating peripheral motion detection. The system is made up of a binocular active system used to track the objects and a fixed camera providing wide angle images of the environment. The system can cope with changes in lighting conditions by adjusting aperture and focus. Binocular flow enables tracking of non-rigid objects even when partial occlusion occurs.

1. Introduction

Surveillance applications are one of the most relevant applications of computer vision. The scenarios of surveillance applications are also extremely varied [12, 11, 13, 17, 18, 15]. Some applications are related to traffic monitoring and surveillance [13, 14], others are related to surveillance in large regions for human activity [16], and there are also applications (related to security) that may imply behavior modelling and analysis [21, 22, 23, 24, 25].

For security applications in man-made environments video images are the most important type of data. These applications are characterized by large amounts of data that require no specific action, while a single and specific event may require immediate intervention. Currently most comercial systems are not automated, and require human attention to interpret the data. Images of the environment are acquired either with static cameras with wide-angle lenses (to cover all the space), or with cameras mounted on pan and tilt devices (so that all the space is covered by using good resolution images).

Computer vision systems described in the literature are also based either on images from wide-angle static cameras, or on images acquired by active cameras. Wide-angle images have the advantage that each single image is usually enough to cover all the environment. Therefore any potential intrusion is more easily detected since no scanning is required. Systems based on active cameras usually employ longer focal length cameras and therefore provide better resolution images. Some of the systems are active and binocular [20]. These enable the recovery of 3D trajectories by tracking stereoscopically. Proprioceptive data from camera platform can be used to recover depth by triangulation. Trajectories in 3D can also be recovered monocularly by imposing the scene constraint that motion occurs in a plane, tipically the ground plane [19]. One of the advantages of an active system is that, in general, the tracked target is kept in the fovea. This implies a higher resolution image and a simpler geometry, thereby facilitating the higher level tasks of object recognition, motion analysis, non-rigid tracking and action understanding.

In this paper we describe a system made up of a static camera and a binocular active system. The static camera acquire wide-angle images of the environment enabling the early detection of intrusion. The system associated to the fixed camera can track several moving objects simultaneously using very simple techniques (image differencing and $\alpha - \beta$ trackers). This system behaves as a peripheral motion detector. If one of the intruders approaches a critical area the active system starts its tracking. For that purpose several visual routines were implemented [5, 6, 7]. By using optical flow the system is able to track binocularly non-rigid objects in real-time. Simultaneously motion segmentation is performed, which enables the extraction of high quality target images (since they are foveated images). These are useful for further analysis and action understanding. The 3D trajectory of the target is also recovered by using the proprioceptive data. This system does not require fixed lighting conditions since it adjusts its aperture and focus to the current lighting level.

Next we will describe the implementation of non-rigid motion tracking and segmentation.

2. Visual Routines

2.1. Peripheral Motion Detection

Peripheral motion detection is performed by a static camera with a overall view of the environment. For motion detection image differencing is performed. Previously to image differencing all the images are low-pass filtered.

An $\alpha - \beta$ filter is associated to each blob detected in the difference image.

The boundaries of the critical area are mapped into the image space. Therefore intrusion (in the critical area) is easily detected by checking the position of each blob/target.

In order that the locations of the targets can be communicated to the active binocular system, a mapping has to be performed during the set-up of the system. This mapping involves two procedures. The first procedure computes a simple homography transformation between the ground plane and the image plane of the static camera. The second procedure enables the definition of a common 2D world coordinate system between the static camera and the binocular active system.

Since targets typically have significant heights, considerable error may be involved in their estimated position (due to the ground plane mapping). To compensate for that error (to a certain extent), and considering the 3D position of the static camera, the target position in space is computed based on the bottom area of the corresponding blob. However this error is rarely relevant since the goal of the static camera is the detection of critical area intrusion. The tracking capabilities of the active system are, in most cases, good enough to lock on the target.

During tracking by the binocular system the target position is cross-checked with the ground plane position estimated based on the static camera images. Since both motions and positions have to be coherent, the probability of having the active system tracking a different object is small.

2.2. Binocular Fixation

The fixation process is realized using gross fixation solutions, defined as saccades, followed by a fine adjustments in both eyes in order to achieve vergence.

Because of the rapid movement of the MDOF[1] head joints during saccades, visual processing cannot be used. Therefore, saccades must be accomplished at the maximum speed or in the minimum possible time. Because the MDOF head has redundant degrees of freedom, motion planning is under-constrained. This means that the fixation can be achieved by an infinite number of possible motion trajectories. To guarantee a unique solution, an artificial constraint can be enforced:

1. The saccade is planned so that the head is at the best state for the next action;
2. Use only the motion of the lightest joints;
3. Use all the joint motions simultaneously to achieve minimum time saccades.

The first solution was the one we adopted, moving neck and eyes in order to achieve symmetric vergence and having the cyclopean eye looking forward to the target. Saccade motion is performed by means of position control of all degrees of freedom involved.

In order to achieve perfect gaze of both eyes in the moving target, and since the center of mass is probably not the same in both retinas for non rigid objects, after the saccade a fine fixation adjustment is performed. A grey level cross-correlation tracker is used to achieve perfect fixation of both eyes. With this procedure the image of the target on the retinas is near the center of the fovea.

2.3. Smooth Pursuit Using Optical Flow

During this process, the motion of the head must satisfy two basic requirements:

1. stabilize the images of the target on both retinas;
2. maintain fixation on the target.

A prerequisite to use pursuit planning is that the target is not far from the fixation point of the head (the images of the target on the retinas must not be far from the center of the foveal window). Otherwise, a saccade must be started prior to the smooth-pursuit process. This means that saccades have higher priority than pursuit.

[1]MDOF:Multi-Degrees-of-Freedom

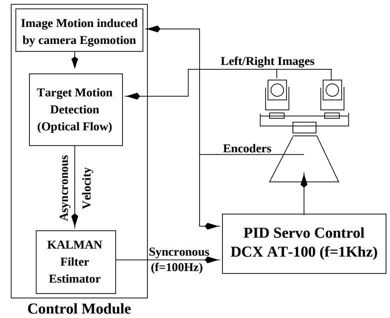

Figure 1. The MDOF High Level Gaze Controller

After fixating on the target the pursuit process is started by computing the optical flow. During the pursuit process velocity control of the degrees of freedom is used instead of position control as in the case of the saccade.

Assuming that the moving object is inside the fovea after a saccade, the smooth pursuit process starts a Kalman filter estimator, which takes the estimated image motion velocity of the target as an input.

Using a **constant acceleration model**, the system state equation is $X_{k+1} = F_k \cdot X_k + V_k$ where V_k is noise, and

$$F_k = \begin{bmatrix} 1 & \delta_{t_k} & \delta_{t_k}^2 \\ 0 & 1 & \delta_{t_k} \\ 0 & 0 & 1 \end{bmatrix}$$

where $\delta_{t_k} = t_{k+1} - t_k$ and F_k is called the *state transfer matrix*. The state of the plant is $X_k = [x_k, v_k, a_k]^T$ at instante t_k, where x_k, v_k and a_k are position, velocity and acceleration respectively.

The real output of the plant (observation) at sample instant $t+1$ is $Z_{k+1} = H \cdot X_{k+1} + W_{k+1}$ where $H = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$ is the measurement matrix, and W_{k+1} is the measurement noise at $(k+1)^{th}$ sample instant. V and W are independent and temporally uncorrelated.

Using the state transfer function of the Kalman filter, the current state of the plant can be predicted if the sample interval is not too large with respect to the motion of the target, simply using linear prediction, $X_{m|k} = F_k \cdot X_{k|k}$, where $\delta_{t_k} = t_m - t_k$ and t_k is the latest sample instant and t_m is current time.

Using this approach, the smooth pursuit controller generates a new prediction of the current image target velocity, and this information is sent to the motion servo controller every 10ms (see fig. 1). The smooth-pursuit controller assumes that the moving target is always located on the horopter and the cyclopean eye is pointing straight to the target. With this assumption, the motion induced on the retina by the moving target is almost the same on both eyes (see fig. 2). Two kalman filters were used to filter the estimated image motion velocities and the velocity used to control the neck pan and tilt joints was considered as the average value of both velocities (cyclopean eye velocity). Maintaining the target on the horopter is acomplished by the vergence process. To maintain tracking, the desired velocity of each of the neck joints should be $V_{des} = V + K_p \cdot \Delta_i$ where K_p is a gain matrix and Δ_i is the position error of the tracking point. The primary function of the first term is to estabilize the images of the target on the retinas. The second term is used to enforce the positional constraint for

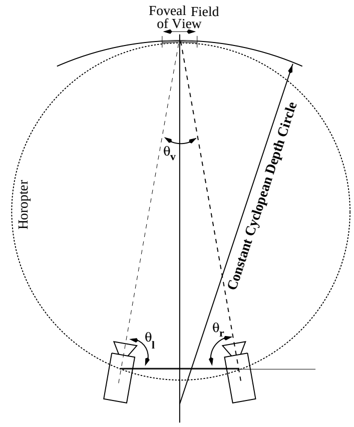

Figure 2. Fixation geometry for vergence and smooth-pursuit.

dynamic fixation.

The Vergence Process

Considering the constraint that both eyes must have equal vergence angles, and the positions of the target on the image plane must be ${}^{l/r}P_i = (0,0)$, the neck and eyes joints rotation angles can be computed based on ${}^{l/r}P_i = {}^{l/r}T_b \cdot P_b$, where P_b represents the target position in the head base coordinate system, and ${}^{l/r}T_b$ represents the transformation matrix between the head base coordinate system and each one of the retina coordinate systems. Due to the particular geometry of the MDOF robot head, computing the joint rotation angles is not an easy task. To simplify this problem, some assumptions must be made, neglecting some of the translation components of the head kinematics, and considering that the target is not moving too much close of the cyclopean eye. Based on the fact that the baseline distance is relatively small compared to the target distance, and that the target is not far from the center of the fovea, if the target is moving along the horopter its location is almost the same in both retinas (see fig 3). For this particular situation, within the field of view of the foveal area, the horopter and the circle of equal cyclopean depth are almost coincident (see fig. 2), which means that targets moving along the horopter mantain a constant cyclopean depth and no vergence control is required.

However, if the target is moving outside the horopter, then its location on the retinas is no longer the same (see fig 3). This displacement is used to control the vergence angles and redirect the head gaze.

Consider the existence of a point P_c with coordinates (X_c, Y_c, Z_c) in the ciclopean eye coordinate system, moving with velocity $V_c = -\Omega_c \times P_c - t_c$, being $\Omega_c = [\Omega_1, \Omega_2, \Omega_3]^T$ the angular velocity and $t_c = [t_x, t_y, t_z]^T$ the translational velocity. This velocity, V_c, can be expressed in each one of the retina coordinate systems, $V_{left/right}$, by

$$V_{l/r} = -R_{l/r}\left[\Omega_c \times P_c + t_c\right] \qquad (1)$$

representing $R_{l/r}$ the rotation matrix between the ciclopean and the left/right retinal coordinate systems.

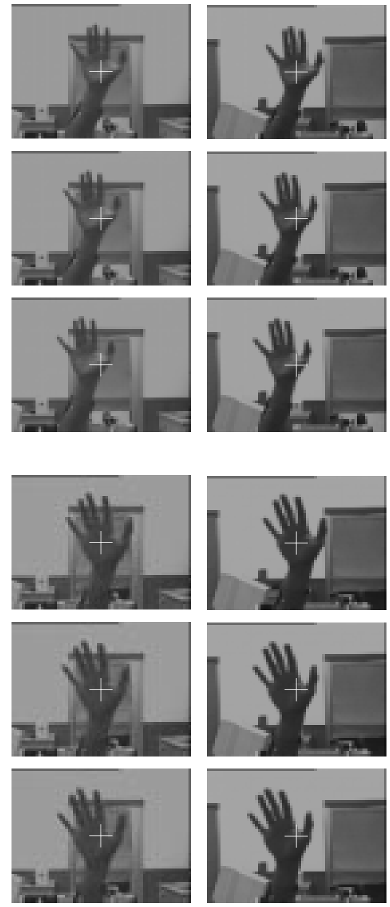

Figure 3. Retinal image disparity. top: targets moving along the horopter. bottom: targets moving outside the horopter.

Defining the retinal image velocity by

$$v_{l/r} = \frac{f}{P_{l/r} \cdot \hat{z}_{l/r}} - \frac{f}{(P_{l/r} \cdot \hat{z}_{l/r})^2} P_{l/r}(V_{l/r} \cdot \hat{z}) \qquad (2)$$

and representing $P_{l/r}$ by $P_{l/r} = R_{l/r}P_c + T_{l/r}$, the retinal image motion flow disparity is given by

$$\begin{aligned} \Delta_v &= v_l - v_r \qquad (3) \\ &= \left[\frac{f}{P_l \cdot \hat{z}_l} - \frac{f}{(P_l \cdot \hat{z}_l)^2} P_l(V_l \cdot \hat{z})\right] - \\ &\quad \left[\frac{f}{P_r \cdot \hat{z}_r} - \frac{f}{(P_r \cdot \hat{z}_r)^2} P_r(V_r \cdot \hat{z})\right] \end{aligned}$$

After some mathematical manipulation, and considering that the target is verged with equal vergence angles ($\theta = \theta_l = \theta_r$), which means that its coordinates are $P_c = [0, 0, Z_c]$ and its image projection are $[0,0]_{l/r}$, the retinal image motion flow disparity is given by

$$\Delta_v = \begin{bmatrix} \Delta v_x \\ \Delta v_y \end{bmatrix} = \begin{bmatrix} \frac{t_z f \sin 2\theta}{Z_c} \\ 0.0 \end{bmatrix} \qquad (4)$$

being f the focal length of both lenses.

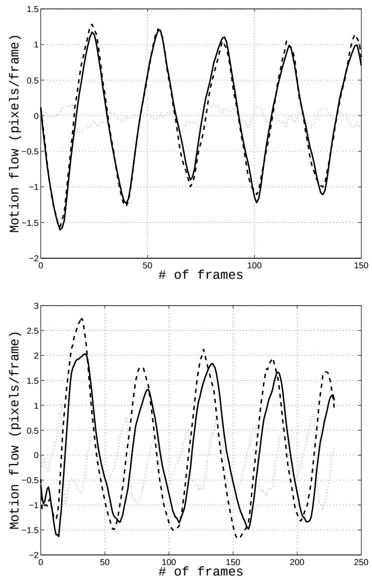

Figure 4. Motion flow and retinal disparity. top) targets moving along the horopter. bottom) targets moving outside the horopter.

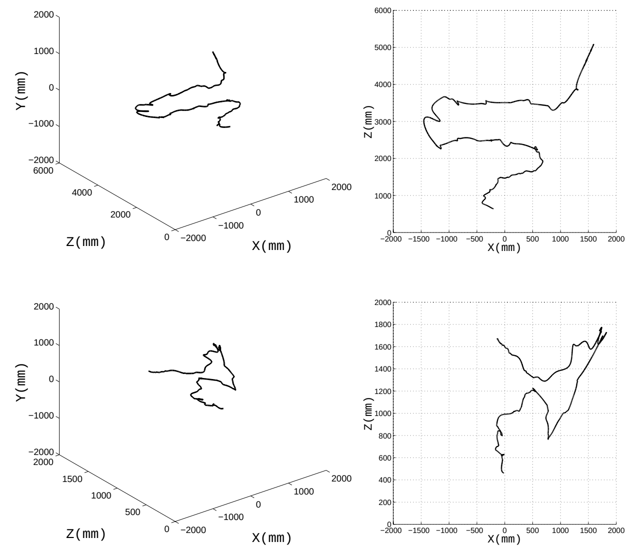

Figure 5. 3D target motion obtained with the proposed vergence and smooth-pursuit process. The right plots represent the top-view of the volume plot, showing the depth movement of the target.

For this particular geometry, the horizontal retinal motion disparity allows the computation of *time-to-contact*

$$\frac{Z_c}{t_z} = \frac{f \sin 2\theta}{\Delta v_x}. \tag{5}$$

Assuming that the target is verged with equal vergence angles on both retinas, $Z_c = \frac{B}{2} \tan\theta$, results for the horizontal retinal motion disparity

$$\Delta v_x = 4 \frac{t_z f \cos^2\theta}{B}, \tag{6}$$

that allows the computation of the Z component of the translational velocity perform by the target, t_z.

Differentiating Z_c with respect to time, results

$$\frac{\partial Z_c}{\partial t} = \frac{B}{2\cos^2\theta} \frac{\partial \theta}{\partial t}. \tag{7}$$

Considering that

$$\frac{\partial Z_c}{\partial t} = t_z = \frac{B \Delta v_x}{4 f \cos^2\theta} \tag{8}$$

and replacing $\frac{\partial Z_c}{\partial t}$ in equation 7 results

$$\frac{\partial \theta}{\partial t} = \frac{\Delta v_x}{2f} \tag{9}$$

that represents the angular velocity of the vergence joints to mantain vergence on the moving target. For this particular geometry for vergence, only the horizontal motion flow disparity is required to control the joints vergence velocity of both retinas.

Figures 4 show the filtered motion vector on both retinas and the retinal disparity, respectively for the case of a target moving along and outside the horopter[2].

The solution adopted to control the vergence of the MDOF robot head is based on two procedures running at different frequencies. The main process is optical flow based and the target motion detected between both retinas (retinal disparity) is used to control the eyes vergence. Since this process only garantees that the target center of mass is located in the center of the fovea, we use a grey level cross correlation to adjust vergence and adjust the target depth, at a sample rate 10 times smaller (every 10th frame). The target depth is used to control the auto-focusing of both eyes, taking advantage of the pre-calibration of the focusing depth [5]. Figure 5 shows the target depth obtained by triangulation using the proposed binocular vergence process.

Target Motion on the Retinas

Unlike the motion of the target in 3D space, which is independent of the head joint motions, the target motions on the retinas are related to the motion of the head joints. Taking into account that the target motions on the retinas caused by the joint motion are generally faster than that caused by the target motion in space, we can not ignore the effects of joint motions on the target motions on the retinas.

We considered the analysis of motion described by the two-component model proposed in [27]. Two different motions must be considered to exist in the scene: one caused by the motion being undertaken by the head and the second one coming from the object. Since the first one is known

[2] On both figures, full line and dashed line correspond respectively to left and right retina motion, and dotted line represents the retinal disparity

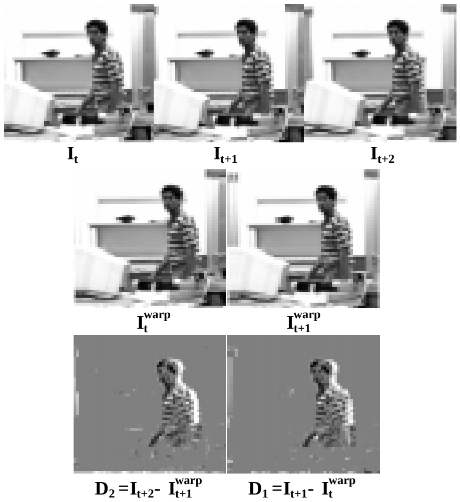

Figure 6. Target motion detection subtracting the image motion induced by camera movement.

Figure 7. Pursuit sequence. Stack of images obtained during the smooth-pursuit process. The δt between consecutive images is $200ms$ (5frames).

(through inverse kinematics of the head), we only have to compute the latter. Let $I(x,y,t)$ be the observed gray scale image at time t. We are assuming that the image is formed by the combination of two distinct image patterns, P(the background motion) and Q (the target motion), having independent motions of p and q,

$$I(x,y,0) = P(x,y) \odot Q(x,y) \quad I(x,y,t) = P^{tp} \odot Q^{tq}$$

where the symbol $\odot$ represents an operator such as addition, multiplication or even a more complex operator to combine the two patterns. In configurations where a foreground object moves in front of a moving background, the combination operator must be more complex, representing the asymmetric relationship between foreground and background. However, since the motion q of the object in the image is in fact $q = p + q'$ where q' is the real image object motion, we can consider that the $\odot$ operator can be approximately additive. Since motion p is known, only q must be determined. The pattern component P moving at velocity p can be removed from the image sequence by shifting each image frame by p and subtracting it from the following frame. The resulting sequence will contain only patterns moving with velocity q (see fig. 6). Let D_1 and D_2 be the first two frames of this difference sequence, obtained from three original frames.

$$\begin{aligned}
&D_1 \equiv I(x,y,t+1) - I^p(x,y,t)\\
&= (P^{2p} + Q^{2q}) - (P^{2p} + Q^{q+p}) = Q^{2q} - Q^{q+p}\\
&= (Q^q - Q^p)^q\\
&D_2 \equiv I(x,y,t+2) - I^p(x,y,t+1)\\
&= (P^{3p} + Q^{3q}) - (P^{3p} + Q^{2q+p}) = Q^{3q} - Q^{2q+p}\\
&= (Q^q - Q^p)^{2q}
\end{aligned} \tag{10}$$

As we can see the sequence of difference images is undergoing a single motion q. That means that it can be computed using any single-motion estimation technique.

In our case we model image formation by means of the scaled orthographic projection. Even if we model image formation as a perspective projection this is a reasonable assumption since motion will be computed near the origin of the image coordinate system (in a small area around the center x and y are close to zero). We can therefore assume that the optical flow vector is approximately constant throughout all the image, i.e., $u = p_x$ and $v = p_y$.

To compute the optical flow vector we minimize

$$\sum_i (I_{x_i} p_x + I_{y_i} p_y + I_{t_i})^2 = 0 \tag{11}$$

Taking the partial derivatives on p_x and on p_y and making them equal to zero we obtain:

$$(\sum_i I_{x_i}^2) p_x + (\sum_i I_{y_i} I_{x_i}) p_y + (\sum_i I_{x_i} I_{t_i}) = 0 \tag{12}$$
$$(\sum_i I_{y_i} I_{x_i}) p_x + (\sum_i I_{y_i}^2) p_y + (\sum_i I_{y_i} I_{t_i}) = 0 \tag{13}$$

The flow is computed on a multiresolution structure. Four different resolutions are used: $16*16, 32*32, 64*64$. These are sub-sampled images. A first estimate is obtained at the lowest resolution level $(16*16)$, and this estimate is propagated to the next resolution level, where a new estimate is computed and so on. The optical flow computed this way is used to control the angular velocity of the motors. The Fig. 7 shows images of the pursuit sequence (25 frames/sec were processed).

Background Image Motion Estimation

The image motion induced by the camera egomotion can be computed using the following equations

$$v_u = \frac{v_x \cdot f_x}{z} - \frac{v_z \cdot f_x \cdot x}{z^2} \qquad v_v = \frac{v_y \cdot f_y}{z} - \frac{v_z \cdot f_y \cdot y}{z^2} \tag{14}$$

where $(f_x,\ f_y)$ represents the focal length of the lens in pixels and $V = \left[\begin{array}{ccc} v_x & v_y & v_z \end{array}\right]^T$ represent the velocity of the point $P = \left[\begin{array}{ccc} x & y & z \end{array}\right]^T$ in the camera coordinate system due to the egomotion of the head.

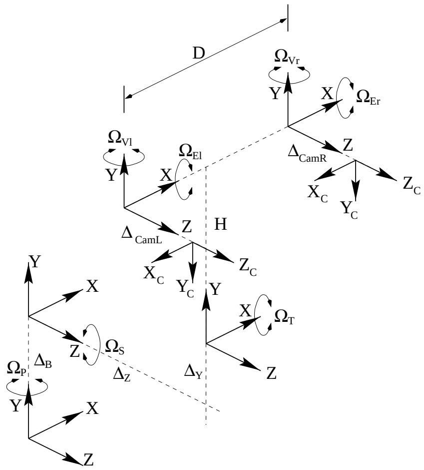

Figure 8. Joints coordinate systems of the MDOF head.

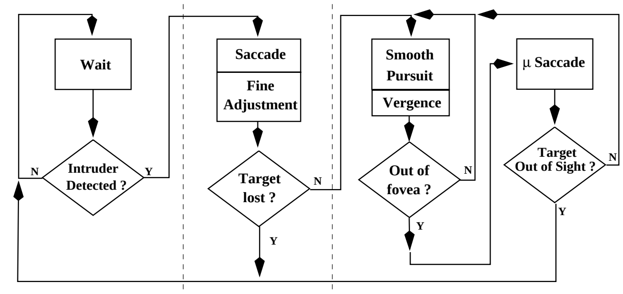

Figure 9. State Transition System.

	Precision	Range	Velocity
Zoom	$Range/90000$	$[12.5 \ldots 75]mm$	$\sim 1.2 * range/s$
Aperture	$Range/50000$	$[1.2 \ldots 16]$	$\sim 2.2 * range/s$
Focus	$Range/90000$	$[1 \ldots \infty]m$	$\sim 1.2 * range/s$

Table 1. Optical structure characteristics of the MDOF active vision system

This velocity V results from the combination of several joints rotation (eye, tilt, swing and pan) and is defined as: $V = V_{eye} + V_{tilt} + V_{swing} + V_{pan}$.

Representing the velocity induced by each joint by $V_{ref} = -V_{Trans} - \Omega \wedge P$ and since each joint only performs rotations ($V_{Trans} = 0$), the velocity induced by each joint can be expressed as $V_{ref} = -\Omega \wedge P$.

Assuming the following angular velocities for each of the rotation joints $\Omega_{eye} = \begin{bmatrix} \Omega_e & \Omega_v & 0 \end{bmatrix}^T$, $\Omega_{tilt} = \begin{bmatrix} \Omega_t & 0 & 0 \end{bmatrix}^T$, $\Omega_{swing} = \begin{bmatrix} 0 & 0 & \Omega_s \end{bmatrix}^T$ and $\Omega_{pan} = \begin{bmatrix} 0 & \Omega_p & 0 \end{bmatrix}^T$ the velocities induced by the rotation of each joint are

$$\begin{aligned} V_{eye} &= -{}^{cam}T_{eye} \cdot (-\Omega_{eye} \wedge P_{eye}) && (15)\\ V_{tilt} &= -{}^{cam}T_{tilt} \cdot (-\Omega_{tilt} \wedge P_{tilt}) && (16)\\ V_{swing} &= -{}^{cam}T_{swing} \cdot (-\Omega_{swing} \wedge P_{swing}) && (17)\\ V_{pan} &= -{}^{cam}T_{pan} \cdot (-\Omega_{pan} \wedge P_{pan}) && (18) \end{aligned}$$

where P_{eye}, P_{tilt}, P_{swing} and P_{pan} represents the coordinates of the point P in each one of the joints coordinate systems (see fig. 8).

2.4. The Global Gaze Controller Strategy

The strategy adopted by the gaze controller to combine saccade, smooth pursuit and vergence to track moving objects using the MDOF robot head was based on a *State Transition System*. This controller defines five different states : *Waiting, Fixation, Pursuit, Vergence and μSaccade*. Each one of these states receives control commands from the previous state, and triggers the next state in a closed loop transition system. The global configuration of this state transition system is show in figure 9.

The distances of the target to the center of the foveal windows on the retinas are well described by the view direction difference between head fixation point and the target in space. Suppose the minimum view angle difference is α_{min} to guarantee that foveal image processing can still yield reliable results. If $\alpha - \alpha_{min} > 0$ then a μsaccade motion planning is activated. Otherwise, smooth-pursuit is used. During the μsaccade motion planning, no visual information is processed, and we use the Kalman filter estimator to predict the position and velocity of the target after a μsaccade.

3. Optical Accomodation

Optical accomodation has not been one of the major research topics in the active vision field, mainly because of the complexity involved in modelling the off-the-shelf motorized lens commonly used in active vision robot systems, and also because the constraints and limitations imposed by these lenses.

New motorized lenses have been developed to enable this head to accommodate the optical system in real time (25 images per second), with very good precision. These lenses have controllable zoom, focus and iris and they use small harmonic drive DC motors with encoder feedback information.

With such performances (see table 1), the lens is able to make continuous, small optical adjustments required by many algorithms in near real time with excellent precision. Qualitative improvements in lens performances increase the advantages of active vision techniques that rely on controlled variations of intrinsic parameters.

3.1. The Motorized Lens Aperture Control

The aperture control of the motorized lens was designed to work as a background process, running in parallel with the visual behaviors developed. The main purpose of this sub-section is to describe what criteria should be used to control the aperture of the lens. By optimum lens aperture we mean the aperture that enable sharp image acquired by the retinas. Several statistical measures have been made with gray level images such has mean, variance and median. Simultaneously we measure the focus parameter in order to confirm the correct behavior with the aperture control.

The behavior of these statistical measures were obtained using several lens aperture and focus positions, and we observed that the grey level image variance was the best parameter to be used for the aperture control (see fig.10). Maximum grey level variance occurs simultaneously with

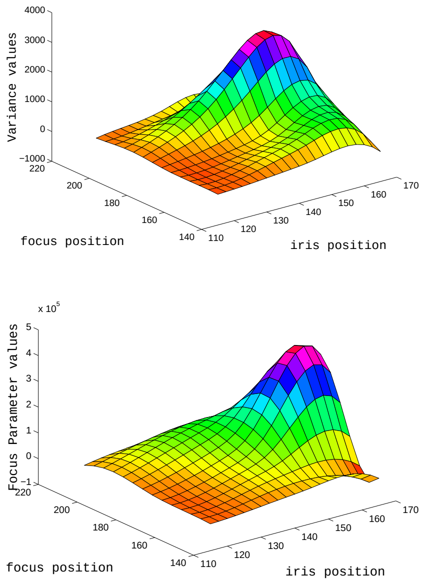

Figure 10. Variance and Focusing behavior as function of the aperture and focus lens position. The axes of the graphics were scaled to [0..256] discrete positions.

the maximum of the focus parameter, and in fact sharp images were obtained for that situation. This behavior was shown to be independent of the focus and aperture of the lens.

After selecting the grey level variance as the aperture control parameter, we developed a background process that always try to achieve the best aperture position using an *hill climbing* grey level variance maximum search.

3.2. Motorized Zoom Lens Modeling

Having determined the parameters of the camera model for a range of lens settings, no matter the calibration procedure adopted, we should characterize how they vary with the lens settings. For that purpose, we must obtain some generic functional relationships between a dependent variable s (the lens parameters) and an independent variable l (the lens settings). The MDOF motorized lens have the capability to adjust focus, zoom and aperture.

Since only focus and zoom are going to be considered in this modeling process, the functional relationship used to model the lens behavior has two independent variables, the focus motor setting foc_p and the zoom motor setting $zoom_p$. We used to describe these functional relationships bivariate polynomials. The general formula for a n^{th} order bivariate polynomial is

$$BP(foc_p, zoom_p) = \sum_{i=0}^{n} \sum_{j=0}^{n-i} a_{ij} foc_p^i zoom_p^j \qquad (19)$$

and the number of coefficients required by the polynomial is $NC = (n+1)(n+2)/2$.

Using these functionals two main relationships were modeled with direct influence in the auto-focusing mechanism implemented. They modeled the relationship between

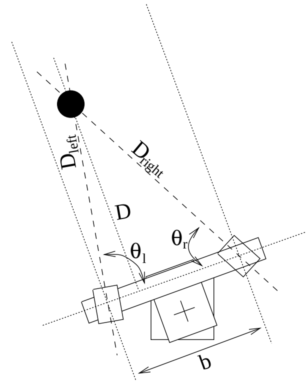

Figure 11. Vergence fixation process

the focus and zoom motor settings and the effective focal length of the lens f, and the relationship between focus and zoom motor settings and focused target depth D:

$$f = BP(foc_p, zoom_p) \qquad D = BP(foc_p, zoom_p). \qquad (20)$$

The focused target depth is measured using stereo triangulation from verged foveated images. The effective focal length of both eye lenses is computed and both images were focused using the Fibonacci maximum search procedure of Tennengrad focus criterion.

3.3. Auto-Focusing Mechanism

The auto-focusing mechanism is mainly based on a previous calibration of the focused target depth D. During the smooth-pursuit process both eyes are verged on the moving target. Taking advantage of the precise information provided as feedback by all MDOF robot head degrees of freedom, a rough estimate of the target depth related to the robot head can be obtained through triangulation of fixated foveated images.

Observing figure 11 and since the vergence angles (θ_l, θ_r) and baseline distance b are known with accuracy, the target depth D can be computed by

$$D = b \cdot \frac{\tan\theta_l \cdot \tan\theta_r}{\tan\theta_l + \tan\theta_r} \qquad (21)$$

Since the target distance D is obtained using equation 21 making use of the vergence angles, the lens-target distance can now be computed using equations:

$$D_{left} = \frac{D}{\sin\theta_l} \qquad D_{right} = \frac{D}{\sin\theta_r} \qquad (22)$$

Taking the lens-target depth (D_{left}, D_{right}) computed using equations 22 and the zoom motor setting $zoom_p$, that was considered fixed during the smooth-pursuit process, the focus motor position foc_p can now be computed using the bivariate polynomial that modeled the relationship between the focus and zoom motors settings and the target depth D. Since this mechanism for auto-focusing requires simple mathematical computation and taking advantage of the MDOF motorized zoom lens speed and accuracy, a real-time auto-focusing process was performed with good results. Figure 12 presents the behavior of the computed target depth and focus motor setting during a smooth-pursuit process.

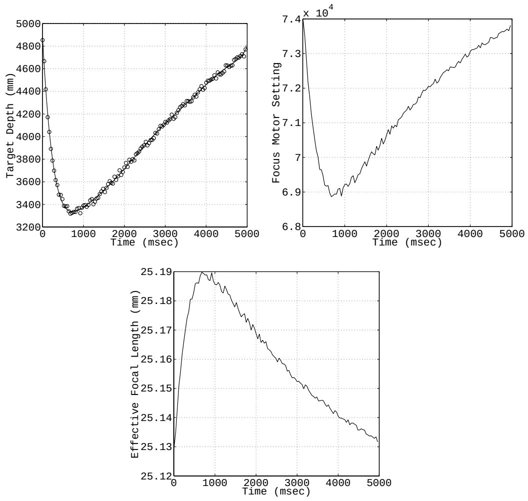

Figure 12. Target Depth and auto-focusing focus motor setting during a $5sec$ smooth-pursuit process.

4. Conclusions

In this paper we described a system that can track non-rigid objects in real-time by using differential flow. Since no shape information is used to track the objects, non-rigidity is not a problem. Also the system can cope with occlusion. The degree of occlusion tolerance depends upon the distance at which the target is located. The closer to the system, the higher the degree of occlusion that can be handled. Since the target is always foveated and focused, good quality images can be extracted for further processing. The 3D trajectories are reconstructed easily by using the proprioceptive data. Behavior modelling can advantageously use the reconstructed 3D trajectories.

References

[1] McFarland, D., Bosser, T.: *Intelligent Behavior in Animals and Robots.* MIT Press (1993)

[2] Ho, Y.: Dynamics of Discrete Event Systems. *Proc. IEEE* **77** (1989) 3–7

[3] Sloman, A.: On Designing a Visual System. *J. Exper. and Theor. Artif. Intell.* **1** (1989) 289–337

[4] Horridge, G.: The evolution of visual processing and the construction of seeing systems. *Proc. Roy. Soc. London B* **230** (1987) 279–292

[5] Batista, J., Peixoto, P. Araújo, H., Real-Time Visual Behaviors with a Binocular Active Vision System. *ICRA97 - IEEE Int. Conf. on Robotics and Automation.* Albuquerque, New Mexico, USA, April, (1997)

[6] Batista, J., Peixoto, P. Araújo, H., Real-Time Vergence and Binocular Gaze Control. *IROS97 - IEEE/RSJ Int. Conf. on Intelligent Robots and Systems.* Grenoble, France, September, (1997)

[7] Batista, J., Peixoto, P. Araújo, H., A Tracking System Integrating Dynamic Accomodation, *SIRS97 - 5th Symposium on Intelligent Robotics Systems*, Stockholm, Sweden, July, (1997)

[8] Pahlavan, K.: Active Robot Vision and Primary Ocular Processes. *PhD Thesis*, CVAP, KTH, (1993), Sweden.

[9] Dias, J., Paredes, C., Fonseca, I., Batista, J., Araújo, H., Almeida, A.: Simulating Pursuit with Machines: Experiments with Robots and Artificial Vision. *Proc. IEEE Int. Conf. on Rob. and Auto.* 472–477 May 21–27, (1995), Nagoya, Japan

[10] Brown,C.: Gaze controls with interaction and delays. *IEEE Transactions on Systems, Man and Cybernetics* **20**, May, (1990), 518–527.

[11] Kanade, T., Collins, R.T., Lipton, A.J., Anandan, P., Burt, P., Wixson, L.: Cooperative Miltisensor Video Surveillance, In *DARPA97*, (1997), 3-10.

[12] Crowley, J.: Coordination of Action and Perception in a Surveillance Robot, In *IEEE Expert (2)*, November, (1987).

[13] Howarth, R., Buxton, H.: Visual Surveillance Monitoring and Watching, In *ECCV96-II*, (1996), 321-334.

[14] Davis, L., Chellapa, R., Yaccob, Y., Zheng, Q.: Visual Surveillance and Monitoring of Human and Vehicular Activity, In *DARPA97*, (1997), 19-27.

[15] Yedanapudi, M., Bar-Shalom, Y., Pattipati, K.R.: IMM Estimation for Multitarget-Multisensor Air-Trafic Surveillance, In *Proc. IEEE*, no.1, January, (1997).

[16] Rao, K.: Shape Description of Curved 3D Objects for Aerial Surveillance, In *ARPA96*, (1996), 1065-1076.

[17] Toal, A.F., Buxton, H.: Spatio-Temporal Reasoning within a Traffic Surveillance System, In *ECCV92*, (1992), 884-892.

[18] Carlotto, M.J., Detection and Analysis of Change in Remotely-Sensed Imagery with Application to Wide Area Surveillance, In *IP (6)*, no. 1, January, (1997).

[19] Bradshaw, K.J., Reid, I., Murray, D.: The Active Recovery of 3D Motion Trajectories and Their Ise in Prediction, In *IEEE Trans. PAMI*, vol.19, no.3, March, (1997), 219-234.

[20] Fairley, S., Reid, I., Murray, D.: Transfer the Fixation for an Active Stereo Platform via Affine Structure Recovery, In *Proc. ICCV95-I*, (1995), 100-105.

[21] Shen, X., Hogg, D.: Shape Models from Image Sequences, In *Proc. ECCV94*, (1994), 225-230.

[22] Baumberg, A., Hogg, D.: Learning Flexible Models From Image Sequences, In *Proc. ECCV94*, (1994).

[23] Metta, G., Hogg, D.: Control of a Steerable Camera with Variable Zoom, In *SMART Workshop*, Lisbon, (1995).

[24] Kollnig, H., Nagel, H-H., Otte, M.: Association of Motion Verbs with Vehicle Movements Extracted from Dense Optical Flow Fields, In *Proc. ECCV94*, (1994), 338-350.

[25] Koller, D., Daniilidis, K., Thórhallsson, T., Nagel, H-H.: Model-Based Object Tracking in Traffic Scenes, In *Proc. ECCV92*, (1992), 437-452.

[26] Murray D., Bradshaw K., MacLauchlan P., Reid I., Sharkey P.: Driving Saccade to Pursuit Using Image Motion. *Intern. Journal of Computer Vision* **16**, No.3, November, (1995), 205–228.

[27] Bergen, J., Burt, P., Hingorani, R., Peleg. S.: Computing Two Motion from Three Frames In *David Sarnoff Research Center* April, (1990).

[28] Brown,C., Coombs D.: Real-Time Binocular Smooth Pursuit *Intern. Journal of Computer Vision* **11**, No.2, October, (1993), 147–165.

[29] Burt, P., Bergen, J., Hingorani, R., Kolczynski, R., Lee, W., Leung, A., Lubin, J., Shvaytser, H.: Object tracking with a moving camera. *Proc. IEEE Workshop Visual Motion*, Irvine, (1989).

[30] Willson, R.: Modelling and Calibration of Automated Zoom Lenses. *CMU-RI-TR-94-03*, Carnegie Mellon University, (1994).