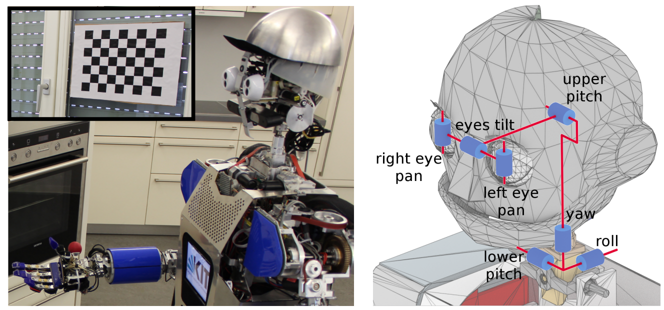
Multimodal Gaze Stabilization of a Humanoid Robot based on Reafferences

Timothée Habra[1], Markus Grotz[2], David Sippel[2], Tamim Asfour[2] and Renaud Ronsse[1]

***Abstract*— Gaze stabilization is fundamental for humanoid robots. By stabilizing vision, it enhances perception of the environment and keeps regions of interest inside the field of view. In this contribution, a multimodal gaze stabilization combining proprioceptive, inertial and visual cues is introduced. It integrates a classical inverse kinematic control with vestibulo-ocular and optokinetic reflexes. Inspired by neuroscience, our contribution implements a forward internal model that modulates the reflexes based on the reafference principle. This principle filters self-generated movements out of the reflexive feedback loop. The versatility and effectiveness of this method are experimentally validated on the ARMAR-III humanoid robot. We first demonstrate that all the stabilization mechanisms (inverse kinematics and reflexes) are complementary. Then, we show that our multimodal method, combining these three modalities with the reafference principle, provides a versatile gaze stabilizer able to handle a large panel of perturbations.**



(A) (B)

Fig. 1. The ARMAR-III humanoid robot. (A) The robot in its home environment used to validate the gaze stabilization. The robot's point of view is shown in the top left corner. (B) Kinematic chain of the head, with 4 degrees of freedom for the neck and 3 for the eyes.

I. Introduction

Vision plays a central role in our perception of the world. It allows to interpret our surrounding environment at a glance. Not surprisingly, humanoid robots heavily rely on visual perception.

However, the quality of visual information is severely degraded by uncontrolled movements of the cameras in space and by motions of the visual target. Points of interest can move out of the field of view and motion blur can appear.

In this context, gaze stabilization has emerged as a powerful solution to enhance robots visual perception. Notably, recent contributions showed that gaze stabilization for robots improves object localization with active stereo-vision [1] and 3D mapping of the environment [2].

Implementation of gaze stabilization for robots can be classified into two approaches, (i) bio-inspired approaches based on reflexes and (ii) classical robotic approaches exploiting inverse kinematics.

In humans and most animals with vision, gaze stabilization is governed by two reflexes: the vestibulo-ocular reflex (VOR) and the optokinetic reflex (OKR) [3]. These complementary reflexes trigger eye movements based on the head velocity and on the perceived image motion, respectively. Bio-inspired approaches implement gaze stabilization controllers emulating these reflexes.

Shibata and Schaal combined VOR and OKR using feedback error learning [4]. Their controller learns and adapts to the non-linear dynamics of the oculomotor system. Similarly, Lenz *et al.* introduced an adaptive VOR controller learning the oculomotor dynamics, but based on decorrelation control this time [5], [6]. In [7], both these methods (i.e. feedback error learning and decorrelation control) were compared. More recently, Vannucci *et al.* extended the adaptive gaze stabilization of [4] with a vestibulocollic reflex, stabilizing the head by means of the neck joints [8]. Interestingly, these bio-inspired approaches do not require an accurate model of the robot. Provided a sufficiently long and rich learning phase, they can adapt and be transferred to different robotic heads.

The classical robotic approaches rely on inverse kinematics (IK) models, linking a task space to the joint space, i.e. the neck and eye joints. Different representations of the task space were proposed for gaze control.

Milighetti *et al.* used the line of sight orientation (pan and tilt) [9]. Roncone *et al.* built a kinematic model of the fixation point described as the intersection of the lines of sight of both eyes [10]. In [11], Omerčen and Ude defined the fixation point as a virtual end-effector of a kinematic chain formed with the head extended with a virtual mechanism. We recently extended this virtual model method in order to solve the redundancy through a combined minimization of the optical flow and the head joint velocities [12]. Marturi *et al.* developed a gaze stabilizer based on visual servoing where the task is described as the pose of a visual target in the image space [13].

The main advantage of these kinematic-based approaches is the possibility to leverage the well established inverse kinematics control theory. Notably, the control theory of

[1] Center for Research in Mechatronics, Institute of Mechanics, Materials, and Civil Engineering, and "Louvain Bionics", Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium. timothee.habra@uclouvain.be
[2] Anthropomatics and Robotics, High Performance Humanoid Technologies Lab (H^2T), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

kinematically redundant manipulators was successfully applied in [9], [11], [12] and [14]. Moreover, the IK method offers to control the gaze direction, on top of stabilizing it, as presented in [14]. This gaze control is necessary either to catch up a target not centered in the image or to shift to a new visual target.

All the stabilization methods reviewed above are based on three sources of information, namely visual from the camera video stream, vestibular from the head inertial sensing and proprioceptive from the joint kinematic measurements. Although one can expect that combining these three types of cues could provide a better gaze stabilization than with a limited subset of these modalities, very few studies addressed this topic. More precisely, bio-inspired methods are usually limited to visual and inertial cues, whereas the classical robotic approaches typically rely on a single source of information, usually proprioception (i.e. joint encoders) as in [9], [11], [12]. Recently, an inverse kinematics method based on vision was reported in [13][1]. Roncone *et al.* also introduced two other approaches, one triggered by inertial measurements and one by copies of motor commands [10]. Later, Roncone *et al.* combined the methods based on inertial and proprioceptive information, but no visual feedback was involved [14].

Interestingly, in [15], it is shown that adding an inverse kinematics head stabilization to VOR and OKR effectively improves the gaze stabilization. This approach nicely decoupled a kinematic and a reflex-based approach, the former controlling the neck joints and the latter the eye joints. However, the proprioceptive information was not fully exploited for the gaze stabilization, since it was only used for head stabilization, thus indirectly supporting gaze stabilization.

In this contribution, a novel gaze stabilization method combining proprioception measurements (i.e. joint kinematic) with inertial and visual cues is introduced. It associates an inverse kinematic model with bio-inspired reflexes (VOR and OKR). Drawing inspiration from neuroscience, it implements the reafference principle [16] by means of a forward model [17]. This gaze stabilization is validated with the ARMAR-III humanoid robot [18] in a home environment (Fig. 1). It is important to stress that the main contribution of this study lies in the proposed reafference method. Rather than demonstrating the high performance of a new gaze stabilizer, we demonstrate that the versatility of gaze stabilization can be improved by combining different stabilization mechanisms using proprioceptive, inertial and visual cues.

It is first shown that each isolated method, namely IK, VOR and OKR, is well suited for a specific type of perturbation. Indeed, each individual cue captures a tradeoff between reactivity and versatility. Inverse kinematics methods based on proprioception are the most reactive, but also the least versatile (being only able to compensate for self-induced perturbations). On the other hand, the optokinetic reflex

[1]Note that each of the inverse kinematics approach uses proprioception for computing the task Jacobians, but we only consider here the modality at the origin of the stabilization commands.

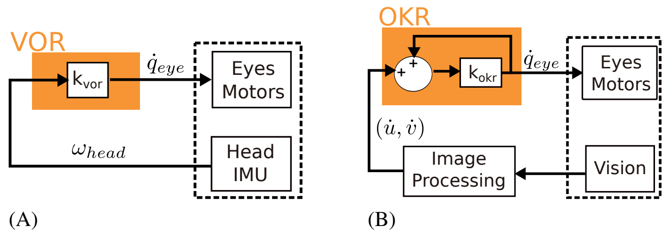


Fig. 2. Block diagrams of the gaze stabilization reflexes, (A) the vestibulo-ocular reflex (VOR) and (B) the optokinetic reflex (OKR).

can theoretically compensate for any disturbance but suffers from long latency due to the inherent image processing. In between, inertial measurements are rather fast but can only detect (and thus stabilize) motions of the robot, e.g. external pushes or joint movements, but not those of the visual target.

Most importantly, we show that, by combining IK, VOR and OKR, the proposed reafference method is both reactive and versatile. It can handle any kind of perturbation as the OKR and can be as fast as the IK. The control automatically adapts to the situation leveraging the best of each method. It is worth noting that no parameter tuning is necessary for achieving this multimodal combination.

This contribution begins by introducing the individual gaze stabilization controls of the proposed approach. In section III, the method to combine these three principles based on reafferences is detailed. Then, the experiments and their results are discussed in sections IV and V. Finally, future work and conclusion are reported in section VI.

II. Isolated Gaze Stabilization Methods

This section introduces the isolated gaze stabilization mechanisms used in this contribution, namely the vestibulo-ocular reflex (VOR), the optokinetic reflex (OKR) and the inverse kinematics (IK). Importantly, this contribution is not aiming at implementing the most advanced stabilization for each of these three mechanisms, but rather at introducing a new method to combine them. Therefore, the principles reported in this section should be viewed as conceptual building blocks used to illustrate the combination by reafference.

A. Vestibulo-ocular reflex

The VOR stabilizes the gaze by producing eye movements counteracting head movements [3]. As displayed in Fig. 2A, this reflex is triggered by a measurement of the head rotational velocity ω_{head}, e.g. provided by the gyroscopes of an *Inertial Measurement Unit* (IMU) located in the head.

In this context, compensatory eye movements can be computed as:

$$\dot{q}_{eye} = -k_{vor} \cdot \left[\omega_{yaw} \quad \omega_{pitch}\right]^T , \qquad (1)$$

with ω_{yaw} and ω_{pitch} being the yaw and pitch rotational velocity expressed in the head frame. The control output $\dot{q}_{eye} = [\dot{q}_{yaw} \; \dot{q}_{pitch}]^T$ is the desired velocity for the eye motors (around yaw and pitch angles respectively)[2]. The gain

[2]These velocities are used as references for a low level joint controller, not represented in this article for the sake of brevity.

k_{vor} should be close to 1 in order to fully compensate for the head rotations.

This reflex benefits from a reliable information provided at a high sampling rate and requires little computation. It is thus very robust, although, it can only compensate perturbations due to robot motions (externally or self-induced). In contrast, motions of the visual target would not be detected and thus not compensated. Note that the present implementation does not compensate for head translations which would require additional sensory input like head translational velocity and distance to target.

B. Optokinetic reflex

The OKR stabilizes the gaze by producing eye movements cancelling the retinal slip, i.e. the perceived target motion within the image. Retinal slip in the horizontal axis of the image elicits yaw rotations while vertical retinal slip elicits pitch rotations. The retinal slip $(\dot{u},\ \dot{v})$, is typically obtained from image processing, i.e. by computation of the optical flow [19]. An implementation of the OKR can be achieved by computing the eye velocities as:

$$\dot{q}_{eye} = k_{okr} \cdot \begin{bmatrix} \dot{u} & \dot{v} \end{bmatrix}^T \tag{2}$$

Knowing the opening angles and the frame rate of the camera, it is possible to express the retinal slip $(\dot{u}, \dot{v})$ in rad/s. In this case, the gain k_{okr} should also be close to 1.

A more efficient implementation can be obtained by taking into account the eye velocity in the control loop as proposed in [20]:

$$\dot{q}_{eye} = k_{okr} \cdot \begin{bmatrix} \dot{u} + \dot{q}_{yaw} & \dot{v} + \dot{q}_{pitch} \end{bmatrix}^T \tag{3}$$

This cancels the static error in the case of a perturbation of constant velocity. A block diagram of this implementation of the OKR is shown in Fig. 2B.

In contrast to the VOR, the input of the OKR is usually noisy and available at a lower frequency (e.g. 30 Hz for standard cameras). This inherent drawback of image processing makes this reflex less accurate and less reactive. On the other hand, vision provides the only direct feedback about the task, i.e. cancelling a potential retinal slip. Therefore, this is the sole source of feedback that can stabilize the image in a dynamic environment (with unpredictably moving objects).

C. Inverse kinematics control

The IK method relies on a task space representation of the control problem. It applies to the gaze stabilization task, the classical inverse kinematics control scheme of the canonical form:

$$\dot{x}_{des} = K_p(x_{des} - x) + \dot{x}_{pred} \tag{4}$$

$$\dot{q}_{des} = J^{\dagger}(q)\dot{x}_{des} \tag{5}$$

Where x_{des} is the desired state, x is the current state, K_p is a proportional gain and $J^{\dagger}$ is a pseudo inverse of the task Jacobian. $\dot{x}_{pred}$ is a predictive command.

This classic control scheme is executed in two steps. First, a corrective velocity $\dot{x}_{des}$ is computed in the task space, e.g. Cartesian (4). Then, the desired joint velocities $\dot{q}_{des}$ are obtained by projecting these desired task velocities in the joint space, using differential inverse kinematics (5).

Various task space representations were proposed in the literature for gaze control (e.g. [9]–[13]). In this contribution, a method based on a virtual linkage model (as first proposed by Omerčen and Ude [11]) similar to the one developed in [12] is used. In this method, the kinematic model of the robot is extended with a virtual spherical arm connected to the robot camera. The state x is the Cartesian pose of the virtual end effector of this new kinematic chain. Controlling x to match the pose of the visual target is equivalent to impose that the target remains centered and aligned in the image frame [12]. This virtual linkage model thus rephrases the gaze stabilization problem as the classic control of a serial manipulator. This stabilizer controls all the head joints (neck and eye joints), i.e. $\dot{q}_{des} = \dot{q}_{head} = [\dot{q}_{neck}, \dot{q}_{eye}]$.

The desired state x_{des} is thus the pose of the visual target. This target is typically provided by a higher level active vision module selecting the area of interest to be further inspected by the robot. An example of the integration of such an active vision module with this gaze stabilizer is presented in [1]. For gaze stabilization, x_{des} is thus kept constant.

The predictive term $\dot{x}_{pred}$ is chosen to compensate for the motion induced by the body's own movements as proposed by Roncone *et al.* [10]. It is computed as the opposite of the velocity of the virtual end-effector that would be induced by the body motion, computed as $\dot{x}_{pred} = -J_{body}\dot{q}_{body}$, where $\dot{q}_{body}$ is the velocity of the body joints and J_{body} is the Jacobian relating this joints velocity to the end effector velocity. In contrast to [10] and [12], here, we use the measured velocity of the body joints rather than the joint velocity commands. This releases the constraint of controlling the body joints (e.g. legs, torso) in velocity and barely affects the control since the encoder measurements are almost delay free. The redundancy of the inverse kinematics is solved through a combined minimization of the optical flow and head velocity. For more details, refer to [12]. The block diagram of this controller is represented in Fig. 3.

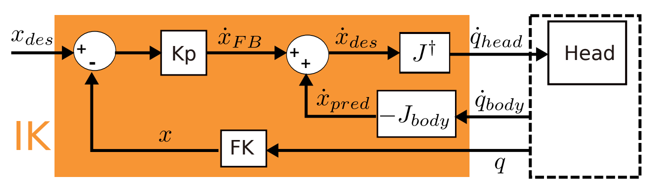


Fig. 3. Inverse kinematics method (IK) for gaze stabilization. A corrective velocity $\dot{x}_{des}$ is computed with a feedback on the fixation point pose x and a predictive term $\dot{x}_{pred}$. FK is a forward kinematics model.

As opposed to both reflex-based methods controlling only the eye degrees of freedom, IK method controls all the head joints. This allows a higher reactivity by exploiting the redundant motors. Another advantage of IK stabilization methods is that all the theoretical framework of redundant serial manipulator control can be adapted to it. For instance, null space projection or joint limit avoidance can be implemented to solve the inherent redundancy [21]. Finally, task space control offers to control the gaze (i.e. changing the view point), on top of stabilizing it [14]. However, a limitation of stabilization methods based on joint kinematic measurement is that they can only measure and thus stabilize self-induced perturbations.

III. Combination of gaze stabilization methods

A straightforward manner to combine reflexes and inverse kinematics control is to sum or to average their respective contributions (Section III-A). However, such naive combination methods suffer from limitations that would eventually degrade the stabilization performances. Taking inspiration from neuroscience, this section introduces a more appropriate combination approach based on the reafference principle [16] in Section III-B.

A. Combination by summation and average: limitations

The OKR consists in a feedback contribution stabilizing the image from a direct measurement of it. Regarding the VOR, it can be seen as a predictive contribution triggered by the head velocity. Hence, VOR and OKR can be combined by summing their respective output [20]. In such a configuration, the VOR compensates for the perturbations due to the head motion and the OKR stabilizes the remaining motion perceived in the image.

However, using the same summation method for the IK contribution degrades the overall performance, because it corresponds to a mechanism of different nature. On the one hand, the IK controller captures voluntary control of the gaze through neck and eye coordination. Its feedback component offers to control (and thus to change) the view point (i.e. the line of sight) while its feed-forward component compensates for self-induced perturbations. On the other hand, VOR and OKR correspond to reactive eye movements aiming at stabilizing the gaze.

Due to these differences, summing the IK contribution and the reflex ones would produce ineffective gaze stabilization. First, summation of the commands would overcompensate self-induced perturbations. Indeed, if the IK predictive model is accurate enough, it should compensate for a large fraction of the voluntary body motion. But at the same time, if the VOR gain is well tuned, it would also generate a command stabilizing the self-induced body motions measured by the induced head velocity. Summing the contributions of these two pathways would thus produce a command twice too large.

On the contrary, averaging the commands would under-compensate the externally induced perturbations. Indeed, the IK method would not generate any command since it cannot detect externally induced perturbations.

Furthermore, the reflexes would by nature counter-act any voluntary change of gaze direction. For example, a voluntary eye rotation to the right would generate an optical flow in the left direction. This optical flow, if directly fed to the OKR, would thus generate an eye rotation to the left, counteracting the initial desired eye motion to the right.

B. Combination by reafference: principle

Facing this paradox of reflexes counter-acting voluntary motions, neuroscientists identified the principles of *reafference* [16], [22] and *forward model* [17].

Forward models (also known as internal models) receive copies of the motor commands (efference copies) and predict the expected sensory outcome of self-induced motions (predicted reafference). These reafferences are then subtracted from the actual sensor measurements, thus isolating the sensory consequences of externally induced perturbations (called exafference). Interestingly, feeding the reflexes with these exafferences rather than directly with the sensor measurements does no longer induce a counter-action of voluntary motions.

From the seminal work of von Holst [16], evidence of such a sensory cancellation mechanism feeding the optokinetic reflexes has been widely demonstrated in animals (see [23] for details). More recently, it was also shown for vesibulo-ocular reflexes on monkeys [24]. Similar sensory cancellation is also observed in humans [25].

Inspired by these concepts of reafference and forward models from neuroscience, we implement such a sensory cancellation mechanism to combine voluntary gaze control from the IK with reflexive control from VOR and OKR. Consequently, the limitations mentioned in subsection III-A no longer impact the stabilization. An overview of the proposed control scheme is provided in Fig. 4 and is further detailed in the following subsections.

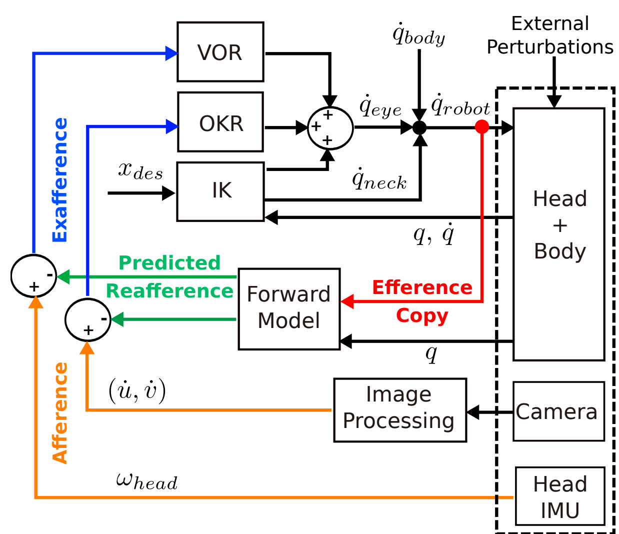


Fig. 4. Combination of the inverse kinematics (IK) with the optokinetic (OKR) and vestibulo-ocular (VOR) reflexes by the reafference method. A forward model predicts the sensory outcome of self-induced motions (reafference). The reflexes are fed with the exafference, i.e. the difference between the sensory measurement (afference) and the reafference prediction. $\dot{q}_{robot}$ is the vector containing the velocity command for the robot joints, i.e. $\dot{q}_{robot} = [\dot{q}_{body}, \dot{q}_{neck}, \dot{q}_{eye}]$. In practice, in the implemented model the efference copy is replaced by a direct measurement of the joint velocity. This canonical scheme is kept for the sake of simplicity.

C. Forward model

The proposed gaze stabilization requires a forward model predicting the sensory consequences of the self-induced movements, known as the reafferences (Fig. 4). In the present case, the forward model must thus predict the self-induced contribution on the head rotational velocity ω_{head_reaf} and on the optical flow of the camera images $(\dot{u}_{reaf}, \dot{v}_{reaf})$.

As a first step toward a full forward model, we considered here a forward model being purely kinematic, and thus embedding no dynamical contributions. As opposed to a kinematic model, a dynamic model would take the joint torques as input and compute the resulting joint trajectories through direct dynamics integration. Instead, our kinematic model uses joint velocities as input and predicts their effect through forward kinematics. This approach is similar to the feed-forward component of [10] taking the velocity reference commands as input. Moreover, rather than using the velocity reference commands as input (efference copies) we propose to use the actual encoder measurements. In practice, measurements obtained from the encoders are accurate and almost delay-free. Interestingly, this does not limit the whole robot to be controlled in velocity but allows any kind of control (e.g. force control for the lower body).

The forward model predicting the head rotational velocity is straightforward. Knowing the location of the IMU in the head, it is possible to get the IMU orientation (given as the rotation matrix R_{imu}) and its rotational velocity ω_{imu} as a function of the joint positions q and velocities $\dot{q}$, where R_{imu} and ω_{imu} are both expressed in the world frame. Then, the reafference for the gyroscope velocities (expressed in the IMU frame), is given by:

$$\omega_{head\_reaf} = R_{imu}(q)\ \omega_{imu}(q, \dot{q}) \tag{6}$$

The optical flow can be estimated from the image Jacobian J_{im} (also called interaction matrix), originally developed for visual servoing [26]. This Jacobian linearly maps the cameras linear and rotational velocities (expressed in the camera frame) $\begin{bmatrix} v_{cam}, & \omega_{cam} \end{bmatrix}$ to the optical flow $\begin{bmatrix} \dot{X}, & \dot{Y} \end{bmatrix}$ as:

$$\begin{bmatrix} \dot{X} & \dot{Y} \end{bmatrix}^T = J_{im}(X, Y) \begin{bmatrix} v_{cam} & \omega_{cam} \end{bmatrix}^T \tag{7}$$

where (X, Y) are the coordinates of the point of interest in the image frame.

Assuming that the visual target is centered in the image frame, i.e. that the gaze is properly stabilized, the target image velocity can be estimated as $J_{im}(0, 0) \begin{bmatrix} v_{cam} & \omega_{cam} \end{bmatrix}^T$, thus giving:

$$\begin{bmatrix} \dot{u}_{reaf} \\ \dot{v}_{reaf} \end{bmatrix} = \begin{bmatrix} (f/Z) v_{cam\_x} + f \omega_{cam\_y} \\ (f/Z) v_{cam\_y} - f \omega_{cam\_x} \end{bmatrix} \tag{8}$$

where Z is the distance between the camera and the visual target and f the camera focal length. This captures the contribution of the translations and the rotations along the horizontal and vertical axes of the image, x and y, respectively. The optical flow corresponding to (8) can then be expressed as a function of the robot kinematics using:

$$\begin{bmatrix} v_{cam} & \omega_{cam} \end{bmatrix}^T = R_{cam} J_{cam}(q) \dot{q} \tag{9}$$

where J_{cam} is the Jacobian matrix of the camera-fixed frame and R_{cam} is its rotation matrix given by the forward kinematics.

IV. EXPERIMENTAL VALIDATION

This gaze stabilization method was validated in two experiments. First, the three stabilization modalities (IK, VOR and OKR) were individually assessed. Then, different methods combining these modalities were evaluated, including the one based on reafferences (Section III).

A. *Experimental set-up*

The experiments were performed with the ARMAR-III humanoid robot (Fig. 1A). This robot features a human-like head in terms of both kinematics (range of motion, velocity) and vision (foveal vision) [18]. This makes it a suitable platform for evaluating bio-inspired control. More precisely, the head has 7 degrees of freedom as represented in Fig. 1B. However, the *upper pitch* joint was not used during the experiments due to redundancy already provided by other pitch joints.

Each eye is equipped with a wide and a narrow angle camera. The wide camera video stream is available at $30\,\mathrm{Hz}$. Optical flow computed from feature tracking was used as input for the OKR[3]. An *XSense* IMU was mounted on the head for the VOR.

The VOR gain k_{vor} was set to 1. The OKR gain, k_{okr} was set to 0.8, to avoid instability due to the delay of the visual feedback. The feedback gain of the IK, K_p was set to 0, i.e. no drift compensation was used, since drift was neither compensated with the VOR and OKR reflexes. The full OKR (Eq. 3) was used except for the reafference method that used the reduced version (Eq. 2).

B. *Evaluation scenarios*

Three scenarios were used in order to provide a general assessment of the proposed method.

In the first scenario, the perturbation consisted in a periodic motion of the hip yaw joint, as in [8] and [10]. A sinusoidal motion of $0.48\,\mathrm{rad}$ (amplitude) at $0.125\,\mathrm{Hz}$ was used, which corresponds to peak velocity of about $20\,\mathrm{deg/s}$ (similar order of magnitude as in [10]). This voluntary self-generated perturbation is the only one that the IK method can detect and thus stabilize.

The second scenario captured the unpredictable motions of the robot pose in space (e.g. as would occur with an external push). For the sake of reproducibility, it was generated by controlled rotations of the robot omnidirectional platform. A sinusoidal rotation of the platform around the vertical axis of $0.48\,\mathrm{rad}$ (amplitude) at $0.125\,\mathrm{Hz}$ was used. Importantly, this motion was not sent to the gaze stabilization controllers and can thus be considered as an externally induced unpredictable perturbation.

Finally, the last scenario involved motions of the visual target in space, as it typically occurs in dynamic environments. It was generated by a moving chessboard displayed on a TV screen. Once again, no information was provided to the stabilization controllers. The TV and the video were set up to generate perturbation of $0.1\,\mathrm{rad}$ at $0.066\,\mathrm{Hz}$.

[3]From the OpenCV methods *goodFeaturesToTrack* and *calcOpticalFlowPyrLK*.

These three scenarios account for all possible perturbations that can induce image motion: self-induced voluntary robot motions, externally induced robot motions and visual target motions respectively. They will be denoted hereafter as *Self Robot*, *External Robot* and *External Target*.

C. Gaze stabilization assessment

To assess the quality of the image stabilization, the dense optical flow was used like in [10]. It is computed with the Farnebacks algorithm of OpenCV [19] using the actual video stream of the wide camera as input. The dense optical flow ϕ is a 2D vector field capturing the apparent velocity of each pixel in the image frame. This field was then averaged, over a centered window having half of the image width w and height h, using the root mean square error as:

$$\phi_{rmse} = \sqrt{\frac{1}{(w/2)(h/2)} \sum_{-h/4}^{h/4} \sum_{-w/4}^{w/4} \|\phi\|^2} \tag{10}$$

Finally, to get a global stabilization index for the whole experiment, the mean of $\phi_{rmse}(t)$ over the whole video duration was computed (in deg/s). Thus, the better the stabilization, the lower this *stabilization index* should be[4]. Note that the dense optical flow used here is different from the optical flow based on feature tracking used as input of the OKR.

V. Results

A. Individual modalities

In the first experiment, each isolated stabilization method (IK, VOR and OKR) was tested in the three scenarios. As a reference, tests with the stabilization disabled, hereafter referred as *No Stabilization*, were conducted, i.e. all neck and eye joints were kept fixed. The results of this experiment are reported in Fig. 5. For each type of perturbation, a specific method provides better results than the two others.

As expected, the IK is only stabilizing self-induced robot motions. The VOR can also stabilize externally induced robot motions and the OKR can stabilize any type of perturbations.

Interestingly, the less versatile methods are also the most efficient ones. In particular, the IK stabilizes better than the other methods in the *Self Robot* perturbation and the VOR stabilizes better than the OKR for the perturbation induced by robot motions.

This lower performance of the OKR is due to the optical flow computation being both slow and noisy. On the other hand, the good stabilization featured by the IK method — in the *Self Robot* experiment — can be explained by two reasons. First, it is the only method that takes advantage of the whole head's degrees of freedom (neck and eyes). Secondly, it has relatively low delay since encoder measurements are available at high frequency. Regarding VOR, it benefits from the low input delay of the IMU but is limited to the eye joints.

This experiment clearly showed that all stabilization modality are complementary, depending on the type of perturbation. It also strongly suggests that the ideal gaze stabilization method should combine the three sources of information in order to be both versatile and efficient.

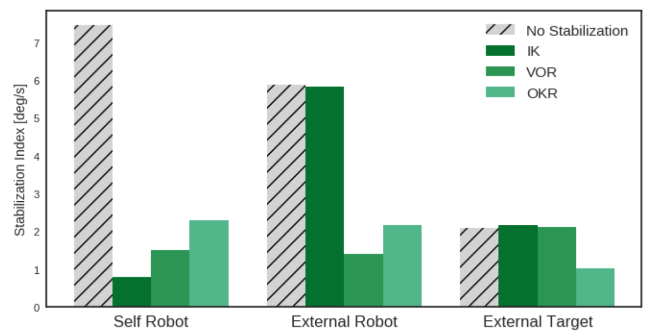


Fig. 5. Stabilization index obtained for each isolated stabilization mechanism (IK, VOR and OKR) in the three scenarios (*Self Robot*, *External Robot* and *External Target*).

B. Combined modalities

In the second experiment, the proposed gaze stabilization combination method based on reafferences was tested in the same scenarios as in the first experiment. In each case, it was compared to the best individual modalities from the first experiment. Furthermore, it was compared to two naive combination methods not relying on reafference prediction, i.e. where the sensory output (afference) is directly fed to the reflexes. The first method, *Sum*, simply sums the output of each modality. The second method, *Mean*, takes the average of the contributions of the three modalities (Section III-A).

The resulting stabilization performances are displayed in Fig. 6. One can observe that the proposed reafference method performs similarly as the best individual modality for each perturbation. In contrast, more naive methods not using reafferences do not perform as well.

The poor quality of the *Sum* method is due to an over compensation of the perturbation, as described in Section III-A. More specifically, for the *Self Robot* and *External Robot* scenarios, more than one stabilization method is active. As a consequence, the sum of the output produces too much compensation.

In contrast, the lower quality of the *Mean* strategy is due to under compensations. Indeed, the IK modality is inactive for the external disturbances. Thus, the mean of the output tends to decrease the velocity command.

More interestingly, the *Reafference* method can automatically detect when it is appropriate to activate or inhibit a reflex, in order to avoid over or under compensations. For example, in the *Self Robot* perturbation, the forward model accurately predicted the inertial and visual feedbacks (see Figs. 7 and 8). Therefore, the input of the reflexes, i.e. the exafferences of Fig. 4 is close to zero, leading to an inhibition of the reflexes. In other words, the reafference

[4]For a full perception of the quality of the stabilization, please refer to the video submitted as supplementary material available at https://youtu.be/WFzg5PzyFDU

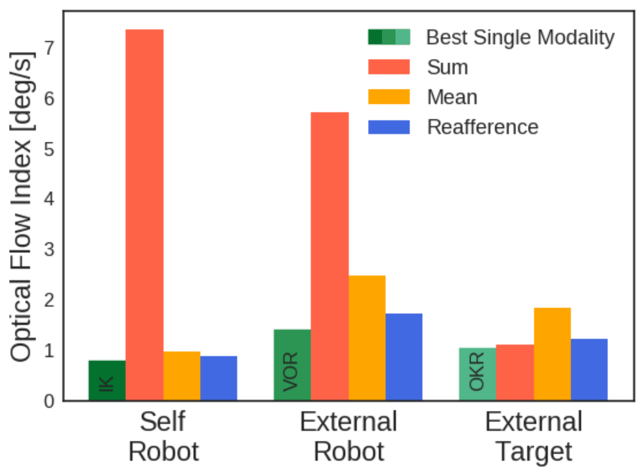


Fig. 6. Stabilization index obtained for the reafference method and two other naive combination methods (*Sum*, *Mean*) in the three scenarios (*Self Robot*, *External Robot* and *External Target*). The stabilization is also compared to the best individual stabilization mechanism from Fig. 5.

method naturally selects the most effective stabilization, i.e. the IK in this case.

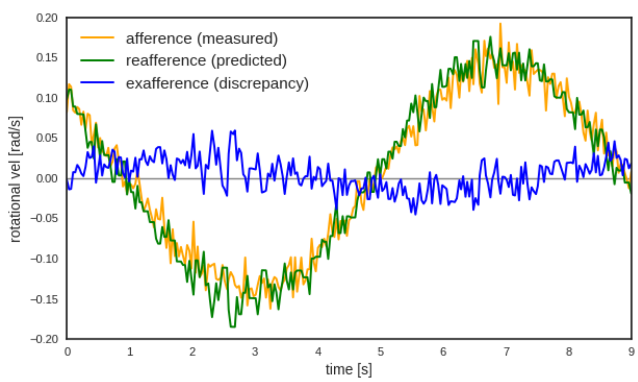


Fig. 7. IMU yaw rotational velocity signals used by the reafference method in the *Self Robot* scenario. The afference is the measurement from the IMU gyroscopes, the reafference is its prediction from the forward model and the exafference is the difference between both used as input for the VOR.

VI. Conclusion

In this contribution, three gaze stabilization controllers were implemented: A classic inverse kinematics (IK) controller along with two bio-inspired reflexes, the vestibulo-ocular reflex (VOR) and the optokinetic reflex (OKR). More importantly, a method combining these three stabilization mechanisms based on the neuro-scientific principles of forward model and reafference was introduced. The stabilization performances obtained was assessed in practical experiments with the ARMAR-III humanoid robot.

We first demonstrated that each of the three stabilization mechanisms (IK, VOR and OKR) presents its own comparative benefit. Indeed, as a function of the perturbation, one sensory information proves to be more appropriate than the two others. While the IK performs best for voluntary self-induced perturbations, inertial sensing makes the VOR most efficient for external pushes on the robot. Finally, visual feedback of the OKR is the only information that can compensate for a moving visual target.

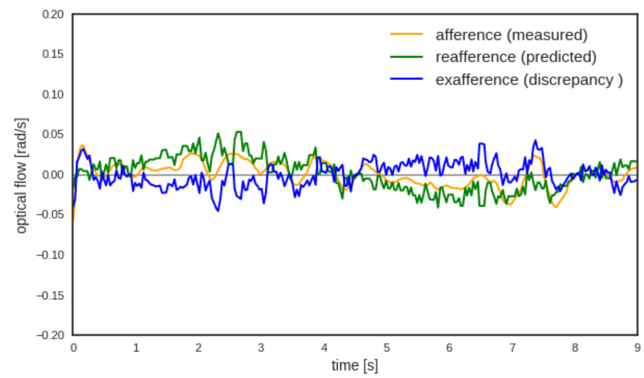


Fig. 8. Optical flow signals (along the horizontal axis) used by the reafference method in the *Self Robot* scenario. The afference is the flow computed from the video stream, the reafference is its prediction from the forward model and the exafference is the difference between both used as input for the OKR. The same scale as for the IMU signals (Fig. 7) is set to allow comparison between both reflexive inputs.

Then, it is shown that combining these individual controllers with the reafference method provides a versatile stabilization. Actually, for each type of perturbation, the reafference method provides stabilization performances of comparable quality as the best individual method. Interestingly, no parameter tuning is necessary for the combination by reafference. The method automatically inhibits the reflexes when appropriate, provided that the forward model is good enough.

In this study, the effectiveness of the reafference method was only evaluated with slow perturbations in one dimension. Our future work will focus on the validation of the proposed approach in more challenging scenarios. Additionally, the integration of this multimodal gaze stabilization on other robots will be considered. Another perspective is to explore the potential of the reafference principle in other tasks than gaze stabilization.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement no 611832 (WALK-MAN) and the support of Wallonie-Bruxelles International.

References

[1] M. Grotz, T. Habra, R. Ronsse, and T. Asfour, "Autonomous view selection and gaze stabilization for humanoid robots," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[2] S. Bazeille, J. Ortiz, F. Rovida, M. Camurri, A. Meguenani, D. G. Caldwell, and C. Semini, "Active camera stabilization to enhance the vision of agile legged robots," *Robotica*, pp. 1–19, 2015.

[3] G. Schweigart, T. Mergner, I. Evdokimidis, S. Morand, and W. Becker, "Gaze stabilization by optokinetic reflex (okr) and vestibulo-ocular reflex (vor) during active head rotation in man," *Vision research*, vol. 37, no. 12, pp. 1643–1652, 1997.

[4] T. Shibata and S. Schaal, "Biomimetic gaze stabilization based on feedback-error-learning with nonparametric regression networks," *Neural Networks*, vol. 14, no. 2, pp. 201–216, 2001.

[5] A. Lenz, T. Balakrishnan, A. G. Pipe, and C. Melhuish, “An adaptive gaze stabilization controller inspired by the vestibulo-ocular reflex,” *Bioinspiration & biomimetics*, vol. 3, no. 3, p. 035001, 2008.

[6] A. Lenz, S. R. Anderson, A. G. Pipe, C. Melhuish, P. Dean, and J. Porrill, “Cerebellar-inspired adaptive control of a robot eye actuated by pneumatic artificial muscles,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 6, pp. 1420–1433, 2009.

[7] E. Franchi, E. Falotico, D. Zambrano, G. G. Muscolo, L. Marazzato, P. Dario, and C. Laschi, “A comparison between two bio-inspired adaptive models of vestibulo-ocular reflex (vor) implemented on the icub robot,” in *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on*, pp. 251–256, IEEE, 2010.

[8] L. Vannucci, S. Tolu, E. Falotico, P. Dario, H. H. Lund, and C. Laschi, “Adaptive gaze stabilization through cerebellar internal models in a humanoid robot,” in *Biomedical Robotics and Biomechatronics (BioRob), 2016 6th IEEE International Conference on*, pp. 25–30, IEEE, 2016.

[9] G. Milighetti, L. Vallone, and A. De Luca, “Adaptive predictive gaze control of a redundant humanoid robot head,” in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pp. 3192–3198, IEEE, 2011.

[10] A. Roncone, U. Pattacini, G. Metta, and L. Natale, “Gaze stabilization for humanoid robots: A comprehensive framework,” in *2014 IEEE-RAS 14th International Conference on Humanoid Robots (Humanoids 2014)*, pp. 259–264, 2014.

[11] D. Omrčen and A. Ude, “Redundant control of a humanoid robot head with foveated vision for object tracking,” in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pp. 4151–4156, IEEE, 2010.

[12] T. Habra and R. Ronsse, “Gaze stabilization of a humanoid robot based on virtual linkage,” in *Biomedical Robotics and Biomechatronics (BioRob), 2016 6th IEEE International Conference on*, pp. 163–169, IEEE, 2016.

[13] N. Marturi, V. Ortenzi, J. Xiao, M. Adjigble, R. Stolkin, and A. Leonardis, “A real-time tracking and optimised gaze control for a redundant humanoid robot head,” in *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*, pp. 467–474, IEEE, 2015.

[14] A. Roncone, U. Pattacini, G. Metta, and L. Natale, “A cartesian 6-dof gaze controller for humanoid robots,” *Proceedings of Robotics: Science and Systems. AnnArbor, Michigan*, 2016.

[15] E. Falotico, N. Cauli, P. Kryczka, K. Hashimoto, A. Berthoz, A. Takanishi, P. Dario, and C. Laschi, “Head stabilization in a humanoid robot: models and implementations,” *Autonomous Robots*, pp. 1–17, 2016.

[16] E. Von Holst, *The behavioural physiology of animals and man: the collected papers of Erich von Holst*, vol. 1. University of Miami Press, 1973.

[17] D. M. Wolpert, Z. Ghahramani, and M. I. Jordan, “An internal model for sensorimotor integration,” *Science*, vol. 269, no. 5232, p. 1880, 1995.

[18] T. Asfour, K. Regenstein, P. Azad, J. Schroder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann, “Armar-iii: An integrated humanoid platform for sensory-motor control,” in *2006 6th IEEE-RAS International Conference on Humanoid Robots*, pp. 169–175, 2006.

[19] G. Farnebäck, “Two-frame motion estimation based on polynomial expansion,” in *Scandinavian conference on Image analysis*, pp. 363–370, Springer, 2003.

[20] D. Robinson, “Linear addition of optokinetic and vestibular signals in the vestibular nucleus,” *Experimental Brain Research*, vol. 30, no. 2, pp. 447–450, 1977.

[21] S. Chiaverini, G. Oriolo, and I. D. Walker, “Kinematically redundant manipulators,” in *Springer handbook of robotics*, pp. 245–268, Springer, 2008.

[22] E. von Holst and H. Mittelstaedt, “Das reafferenzprinzip,” *Naturwissenschaften*, vol. 37, no. 20, pp. 464–476, 1950.

[23] C. R. Gallistel, *The organization of action: A new synthesis*. Psychology Press, 2013.

[24] J. E. Roy and K. E. Cullen, “A neural correlate for vestibulo-ocular reflex suppression during voluntary eye? head gaze shifts.,” *Nature neuroscience*, vol. 1, no. 5, 1998.

[25] S.-J. Blakemore, D. M. Wolpert, and C. D. Frith, “Central cancellation of self-produced tickle sensation,” *Nature neuroscience*, vol. 1, no. 7, pp. 635–640, 1998.

[26] F. Chaumette and S. Hutchinson, “Visual servo control. i. basic approaches,” *IEEE Robotics & Automation Magazine*, vol. 13, no. 4, pp. 82–90, 2006.