Gaze Stabilization for Humanoid Robots: a Comprehensive Framework

Alessandro Roncone[1], Ugo Pattacini[1], Giorgio Metta[1] and Lorenzo Natale[1]

Abstract—Gaze stabilization is an important requisite for humanoid robots. Previous work on this topic has focused on the integration of inertial and visual information. Little attention has been given to a third component, which is the knowledge that the robot has about its own movement. In this work we propose a comprehensive framework for gaze stabilization in a humanoid robot. We focus on the problem of compensating for disturbances induced in the cameras due to self-generated movements of the robot. In this work we employ two separate signals for stabilization: (1) an anticipatory term obtained from the velocity commands sent to the joints while the robot moves autonomously; (2) a feedback term from the on board gyroscope, which compensates unpredicted external disturbances. We first provide the mathematical formulation to derive the forward and the differential kinematics of the fixation point of the stereo system. We finally test our method on the iCub robot. We show that the stabilization consistently reduces the residual optical flow during the movement of the robot and in presence of external disturbances. We also demonstrate that proper integration of the neck DoF is crucial to achieve correct stabilization.

I. INTRODUCTION

Efficient gaze stabilization in mammals is fundamental because it reduces image blur elicited by the movement of the body during locomotion. The brain senses external motion through the vestibular system and the generated optical flow and performs compensatory movements with the eyes and the head to maintain stable fixation. The effect of the absence of stabilization can be easily measured by taking a picture or shooting a video while walking or running.

Gaze stabilization is therefore a fundamental capability for a humanoid robot. Conventionally, algorithms and behaviors for visual stabilization have been designed drawing inspiration from biological systems. Due to its relative simplicity the brain circuitries involved are relatively well understood [1]. Broadly speaking compensatory movements are obtained with two main contributions. The vestibulo-ocular reflex (VOR) exploits the information about the head movement coming from the vestibular system. The whole control loop in this case involves a few synapses and it is therefore very fast. The opto-kinetic reflex (OKR) uses on the other hand retinal slip from the eyes to generate compensatory movement and maintain stable fixation. The computation in this case involves more complex computations, it has larger latency and is less efficient. However these contributions perform best at different frequencies and are therefore integrated for efficient stabilization.

Early work on oculomotor control in robotics has focused on replicating various type of eye movements like vergence, smooth-pursuit, saccades [2], [3], [4] and gaze stabilization reflexes obtained using inertial and visual input [5], [6], [7].

Computation of the eye velocity command for proper stabilization depends on several parameters: eye-head geometry, relative distance between the fixation point and the head but also non-linearities due to lens distortions and delays in the plant. If the eyes and the head do not rotate around the same axes, the compensation signal must take into account the translational velocity due to parallax. This can be done analytically [5] or with Feedback Error Learning [6], [7]. The advantage of the latter methods is that it can also optimally integrate visual and inertial information and compensate for delays in the plant.

Only in a few cases the attention has been devoted to the problem of gaze stabilization during legged locomotion [8], [9]. In [8] the authors implement a controller based on an oscillator which is adapted to match the frequency and phase of the optical flow generated by the robot gait, in the assumption that the latter is periodic. In [9] the authors use genetic algorithms to evolve a central pattern generator that optimally reduces head shaking during locomotion of a quadruped. Previous work on gaze stabilization has focused on the control of the eyes and has ignored a third source of information useful for gaze stabilization, i.e. the motor signals issued to the robot during walking and generic whole-body movements. This information, however, provides important cues for stabilizing motion due to the robot own movement. With respect to inertial and visual signals this information is predictive in that it allows anticipating and planning compensatory movements in advance.

In this paper we solve the problem of gaze stabilization by integrating a feedback component coming form the sensory system with a feedforward component derived from the commands issued to the motors. We build upon the gaze controller implemented on the iCub [10] and extend it to stabilize gaze during active movements of the iCub [11]. The system uses all 6 DoF of the head and it relies on two sources of information: i) the inertial information read from the IMU placed on the robot's head (*feedback*) and ii) an equivalent signal computed from the commands issued to the motors of the torso (*feedforward*). For both cues we compute the resulting perturbation of the fixation point and use the Jacobian of the iCub stereo system to compute the motor command that compensates the perturbation. Retinal slip (i.e. optical flow) is used to measure the performance of
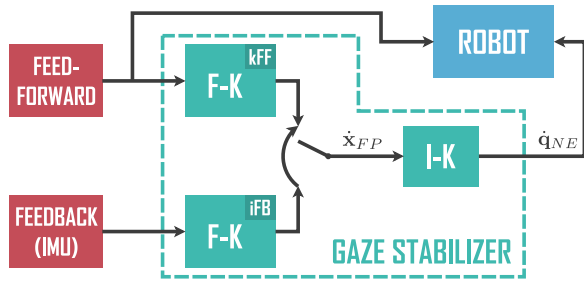
Fig. 1. Block diagram of the framework presented. The Gaze Stabilizer module (in green) is designed to operate both in presence of a kinematic feedforward (*kFF*) and an inertial feedback (*iFB*). In both cases, it estimates the motion of the fixation point and controls the head joints in order to compensate for that motion.

the system. We show that the feedforward component allows for better compensation of the robot's own movements and, if properly integrated with inertial cues, may contribute to improve performance in presence of external perturbations. We also show that the DoF of the neck must be integrated in the control loop to achieve good stabilization performance.

The article is structured as follows. In Section II, the proposed framework is defined. The experimental protocol and the related experiments are presented in Section III, followed by Conclusions and Future Work (Section IV).

II. METHOD

We define the stabilization problem as the stabilization of the 3D position of the fixation point $\mathbf{x}_{FP}$ of the robot. It is achieved by controlling the cameras to keep the velocity $\dot{\mathbf{x}}_{FP}$ equal to zero. The velocity of the fixation point is 6-dimensional, and is composed of a translational component v_{FP} and a rotational part ω_{FP}.

A diagram of the proposed framework is presented in Fig. 1. As highlighted in Section I, the gaze stabilization module has been designed to operate in two (so far mutually exclusive) scenarios:

- a *kinematic feed-forward (kFF)* scenario, in which the robot produces self-generated disturbances due to its own motion; in this case motor commands predict the perturbation of the fixation point and can be used to stabilize the gaze.
- an *inertial feed-back (iFB)* scenario, in which perturbations are (partially) estimated by an Inertial Measurement Unit (IMU).

As result, the *Gaze Stabilizer* is realized by the cascade of two main blocks: the first block is used for estimating the 6D motion of the fixation point $\dot{\mathbf{x}}_{FP}$ by means of the forward kinematics, while the latter exploits the inverse kinematics of the neck-eye plant in order to compute a suitable set of desired joint velocities $\dot{\mathbf{q}}_{NE}$ able to compensate for that motion. The forward kinematics block represents a scenario-dependent component, meaning that its implementation varies according to the type of input signal (i.e. feed-forward or feedback). Conversely, the inverse kinematics module has a unique realization.
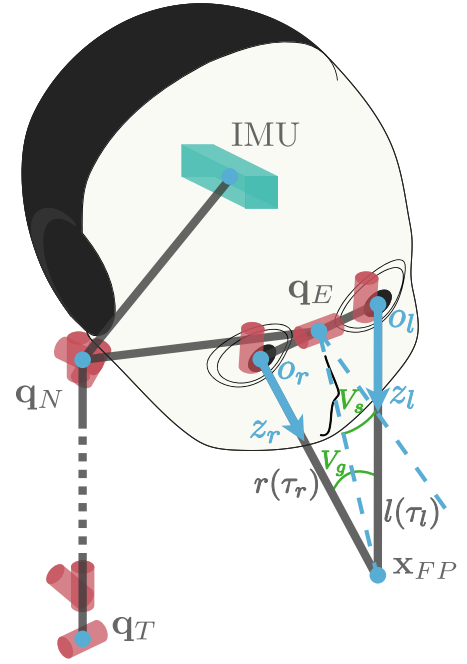

Fig. 2. Kinematics of the iCub's torso and head. The upper body of the iCub is composed of a 3 DoF torso, a 3 DoF neck and a 3 DoF binocular system, for a total of 9 DoF. Each of these joints, depicted in red, are responsible for the motion of the fixation point. The Inertial Measurement Unit (IMU) is the green rectangle placed in the head; its motion is not affected by the eyes.

Crucial to this work is the computation of the position of the fixation point and its Jacobian. Section II-A provides a complete formulation of the kinematic problem occurring at the eyes, whereas Section II-B and II-C analyze the forward and the inverse kinematics modules composing the Gaze Stabilizer.

A. Forward and Differential Kinematics of the iCub stereo system

To derive the Jacobian of the fixation point we start from the forward kinematic law of the eyes as illustrated in Fig. 2. The position of the fixation point $\mathbf{x}_{FP}$ is computed in two steps. The first step computes the position of the frame of reference of the eyes. This uses a representation of the forward kinematics of the iCub head in standard Denavit-Hartenberg notation (the DH parameters of the iCub are reported here: [10]). The second step computes $\mathbf{x}_{FP}$ as the intersection of the two rays joining the cameras optical centers and the projection of the target on the camera planes.

1) Forward Kinematics: by referring to Figure 2, the 3D Cartesian position of the fixation point $\mathbf{x}_{FP}$ can be intuitively defined as the intersection point of the lines $l(\tau_l)$ and $r(\tau_r)$ that originate from the left and right camera planes passing through the respective optical centers. In a parametric formulation, they are defined as:

$$
\begin{aligned}
l(\tau_l) &: o_l + \tau_l \cdot z_l \\
r(\tau_r) &: o_r + \tau_r \cdot z_r
\end{aligned}
\tag{1}
$$

where o_l and o_r are the centers of the left and right camera planes respectively, and z_l and z_r are the axes perpendicular to these planes, as shown in Figure 2. To address the more general case of skew lines (i.e. $l(\tau_l)$ and $r(\tau_r)$ might not be coplanar due to mechanical misalignments of image planes), the fixation point $\mathbf{x}_{FP}$ can be defined as the mean point of the shortest segment between $l(\tau_l)$ and $r(\tau_r)$. From Eq. 1, it is possible to derive the points $\mathbf{P}_l$ and $\mathbf{P}_r$ that belong to each line and minimize the distance from the other line. They are given by:

$$\tau_l^* = \frac{[z_l - (z_l \cdot z_r) \cdot z_r] \cdot [o_l - o_r]}{(z_l \cdot z_r)^2 - 1}$$
$$\tau_r^* = \frac{[(z_l \cdot z_r) \cdot z_l - z_r] \cdot [o_l - o_r]}{(z_l \cdot z_r)^2 - 1} \quad . \quad (2)$$

Finally, the intersection point $\mathbf{x}_{FP}$ can be found as the mean point between $\mathbf{P}_l$ and $\mathbf{P}_r$:

$$\mathbf{x}_{FP} = \frac{\mathbf{P}_l + \mathbf{P}_r}{2} = \frac{o_l + \tau_l^* \cdot z_l + o_r + \tau_r^* \cdot z_r}{2} \quad . \quad (3)$$

2) Differential Kinematics: the position of the fixation point in the Cartesian space depends on the whole body configuration, namely the legs, the torso, the neck and the eyes: $\mathbf{q} = [\mathbf{q}_L, \ \mathbf{q}_T, \ \mathbf{q}_N, \ \mathbf{q}_E]^T$. It is possible to profitably apply the standard DH notation to the kinematics of all the body parts with the exception of the eyes. On the iCub, indeed, three DoFs (the common tilt t_c, the version v_s and the vergence v_g) account for four coupled joints actuating the eyes (the tilt and pan for the left and right cameras, i.e. $[\mathbf{t}_l, \ \mathbf{p}_l]^T$ and $[\mathbf{t}_r, \ \mathbf{p}_r]^T$ respectively). In particular, $\mathbf{q}_E$ is given by:

$$\mathbf{q}_E = \begin{bmatrix} t_c \\ v_s \\ v_g \end{bmatrix} = \begin{bmatrix} \mathbf{t}_l = \mathbf{t}_r \\ \dfrac{\mathbf{p}_l + \mathbf{p}_r}{2} \\ \mathbf{p}_l - \mathbf{p}_r \end{bmatrix} , \quad (4)$$

and this leads to the inverse relations:

$$\mathbf{t}_l = \mathbf{t}_r = t_c, \quad \mathbf{p}_l = v_s + v_g/2, \quad \mathbf{p}_r = v_s - v_g/2 \ . \quad (5)$$

For what concerns the motion of the fixation point $\dot{\mathbf{x}}_{FP}$, for the purposes of this work we are only interested in finding the relation between the joints velocities $\dot{\mathbf{q}}_E$ and its translational component v_{FP}, as detailed in Section II-C. Under this assumption, the Jacobian matrix J_E that relates the motion of the fixation point $\mathbf{x}_{FP}$ with the eyes joints $\mathbf{q}_E$ will be reduced to a 3×3 matrix. The standard analytical Jacobian matrix is defined as:

$$J_E = \frac{\partial \mathbf{x}_{FP}(\mathbf{q}_E)}{\partial \mathbf{q}_E} = \left[\frac{\partial \mathbf{x}_{FP}}{\partial t_c}, \quad \frac{\partial \mathbf{x}_{FP}}{\partial v_s}, \quad \frac{\partial \mathbf{x}_{FP}}{\partial v_g} \right] \ . \quad (6)$$

Using the chain rule, and Equations 3 and 5, leads to:

$$\frac{\partial \mathbf{x}_{FP}}{\partial t_c} = \frac{\partial \mathbf{x}_{FP}}{\partial t_c} = \frac{1}{2} \left(\frac{\partial \mathbf{P}_l}{\partial t_c} + \frac{\partial \mathbf{P}_r}{\partial t_c} \right) \quad (7a)$$

$$\frac{\partial \mathbf{x}_{FP}}{\partial v_s} = \frac{\partial \mathbf{x}_{FP}}{\partial \mathbf{p}_l} + \frac{\partial \mathbf{x}_{FP}}{\partial \mathbf{p}_r} =$$
$$= \frac{1}{2} \left(\frac{\partial \mathbf{P}_l}{\partial \mathbf{p}_l} + \frac{\partial \mathbf{P}_l}{\partial \mathbf{p}_r} + \frac{\partial \mathbf{P}_r}{\partial \mathbf{p}_l} + \frac{\partial \mathbf{P}_r}{\partial \mathbf{p}_r} \right) \quad (7b)$$

$$\frac{\partial \mathbf{x}_{FP}}{\partial v_g} = \frac{1}{2} \cdot \left(\frac{\partial \mathbf{x}_{FP}}{\partial \mathbf{p}_l} - \frac{\partial \mathbf{x}_{FP}}{\partial \mathbf{p}_r} \right) =$$
$$= \frac{1}{4} \cdot \left(\frac{\partial \mathbf{P}_l}{\partial \mathbf{p}_l} - \frac{\partial \mathbf{P}_l}{\partial \mathbf{p}_r} + \frac{\partial \mathbf{P}_r}{\partial \mathbf{p}_l} - \frac{\partial \mathbf{P}_r}{\partial \mathbf{p}_r} \right) \ . \quad (7c)$$

The computation of the quantities presented in Equations 7a, 7b and 7c depends from Equations 3 and 2. For simplicity we derive only the first factor of Eq. 7a; the derivation of the other components has been omitted for brevity but can be derived similarly. $\partial \mathbf{P}_l/\partial t_c$ is given by:

$$\frac{\partial \mathbf{P}_l}{\partial t_c} = \frac{\partial (o_l + \tau_l^* \cdot z_l)}{\partial t_c} = \frac{\partial o_l}{\partial t_c} + \frac{\partial \tau_l^*}{\partial t_c} \cdot z_l + \tau_l^* \cdot \frac{\partial z_l}{\partial t_c} \ . \quad (8)$$

$\partial o_l/\partial t_c$ and $\partial z_l/\partial t_c$ represent, respectively, the geometric Jacobian of the left eye and the analytical Jacobian of the z-axis of the left eye with respect to the tilt; they are described in Equation 11. The second derivative is instead more complex. Let us define:

$$\begin{aligned} \xi_0 &= z_l \cdot z_r \\ \xi_1 &= [z_l - (z_l \cdot z_r) \cdot z_r] = [z_l - \xi_0 \cdot z_r] \\ \xi_2 &= [o_l - o_r] \\ \xi_3 &= (z_l \cdot z_r)^2 - 1 = (\xi_0)^2 - 1 \quad ; \end{aligned} \quad (9)$$

thus, $\partial \tau_l^*/\partial t_c$ becomes:

$$\frac{\partial \tau_l^*}{\partial t_c} = \frac{\partial}{\partial t_c} \left(\frac{\xi_1 \cdot \xi_2}{\xi_3} \right) =$$
$$= \frac{\xi_2 \xi_3 \partial \xi_1/\partial t_c + \xi_1 \xi_3 \partial \xi_2/\partial t_c - \xi_1 \xi_2 \partial \xi_3/\partial t_c}{\xi_3^2} \ . \quad (10)$$

Finally, $\partial \xi_1/\partial t_c$, $\partial \xi_2/\partial t_c$, and $\partial \xi_3/\partial t_c$ can be derived from Equation 9 and are compositions of:

$$\begin{aligned} \frac{\partial o_l}{\partial t_c} &= J_l^G(t_c) & \frac{\partial z_l}{\partial t_c} &= J_l^A(t_c) \\ \frac{\partial o_r}{\partial t_c} &= J_r^G(t_c) & \frac{\partial z_r}{\partial t_c} &= J_r^A(t_c) \end{aligned} , \quad (11)$$

where $J_l^G(t_c)$ and $J_r^G(t_c)$ are the geometric Jacobians of the left and right camera optical centers with respect to the common tilt, whereas $J_l^A(t_c)$ and $J_r^A(t_c)$ are the analytical Jacobians of the left and right z-axis with respect to the tilt. Both J^A and J^G can be retrieved with resort to the standard kinematics libraries as in [10].

B. Estimating the motion of the fixation point

As discussed in Sections I and II, in this work we exploited the gaze stabilization in two different scenarios, described in the following Subsections.

1) Kinematic Feedforward: in the first scenario the robot moves autonomously its body and we estimate the motion of the fixation point with resort to the kinematic model of the robot [10]. Under these assumptions, the task is completely defined: given the joints velocities that the robot is actuating at the motors, the fixation point is moving according to the Jacobian of the kinematic chain under consideration. As an example, let us assume that the robot has fixed hips (i.e. no movement at the lower limbs) and is exerting a given set of velocities at the torso ($\dot{\mathbf{q}}_T$), neck ($\dot{\mathbf{q}}_N$) and eyes ($\dot{\mathbf{q}}_E$). At any given instant of time, the motion of the fixation point is given by:

$$\dot{\mathbf{x}}_{FP} = \begin{bmatrix} v_{FP} \\ \omega_{FP} \end{bmatrix} = J_{TNE} \cdot \begin{bmatrix} \dot{\mathbf{q}}_T \\ \dot{\mathbf{q}}_N \\ \dot{\mathbf{q}}_E \end{bmatrix} \quad , \qquad (12)$$

where J_{TNE} is the 6×9 Jacobian of the forward kinematics map relative to the torso, the neck and the eyes.

2) IMU Feedback: in the second application, we exploited the measurements provided by the IMU device to estimate the motion occurring at the head. The iCub head is currently equipped with the MTx sensor from Xsens [12], whose location with respect to the robot kinematic is known [10]. Among the various sensing elements available from such device, the one of interest here is the gyroscope, able to estimate the 3D rotational velocity ω_{IMU} of the sensor at any given instant of time. From this measurement, it is possible to derive the 6D velocity of the fixation point $\dot{\mathbf{x}}_{FP}$:

$$v_{FP} = \omega_{IMU} \times \mathbf{r}, \qquad \mathbf{r} = \mathbf{x}_{FP} - \mathbf{x}_{IMU} \qquad (13a)$$
$$\omega_{FP} = \omega_{IMU} \ , \qquad\qquad\qquad\qquad (13b)$$

where v_{FP} is the 3D translational velocity of the fixation point, ω_{FP} is its 3D rotational velocity, and $\mathbf{r}$ is the lever arm between the position of the fixation point $\mathbf{x}_{FP}$ and the position of the inertial sensor $\mathbf{x}_{IMU}$. It is worth noticing that this is a sub-optimal case: since the inertial sensor measures only a 3D rotational velocity (i.e. ω_{IMU}), we do not have access to the 3D translational component v_{IMU}. In this scenario we can only compensate for the the rotational velocity as it is measured by the sensor (Eq. 13b) and its effect on the translational component (Eq. 13a).

C. Gaze stabilization from the estimation of the fixation point motion

In the previous sections we illustrated how the feedforward and feedback terms produce an estimation of the velocity of the fixation point $\dot{\mathbf{x}}_{FP} = [v_{FP}, \ \omega_{FP}]^T$. Using the inverse kinematics we derive the compensatory motor commands for the head (see Figure 1):

$$\begin{bmatrix} \dot{\mathbf{q}}_N \\ \dot{\mathbf{q}}_E \end{bmatrix} = -J_{NE}^{\#} \cdot \begin{bmatrix} v_{FP} \\ \omega_{FP} \end{bmatrix} \ , \qquad (14)$$

where $J_{NE}^{\#}$ is the 6×6 pseudo-inverse of the Jacobian of the forward kinematics map relative to the neck and the eyes, and $\dot{\mathbf{q}}_N$, $\dot{\mathbf{q}}_E$ are the desired joint velocities at the neck and eyes respectively.

In this work, we chose to decouple the inverse kinematics problem into two sub-problems: instead of using the full 6-DoF chain of the neck and the eyes to stabilize the 6-DoF motion of the fixation point, we designed the controller such that the neck compensates the rotational component ω_{FP}, whilst the eyes have to counterbalance the translational part v_{FP}. The reason is twofold: 1) the neck and the eyes exhibit two different dynamics, the eyes being faster than the neck joints; 2) it is not physically possible for the neck joints alone to stabilize the translational motion v_{FP} and, similarly, the eyes chain can not compensate for the roll of the fixation point by mechanical design. Hence, Equation 14 has been split into:

$$\begin{cases} \dot{\mathbf{q}}_N = -J_N^{\#} \cdot \omega_{FP} \\ \dot{\mathbf{q}}_E = -J_E^{\#} \cdot v_{FP} \end{cases} \ , \qquad (15)$$

with $J_N^{\#}$ and $J_E^{\#}$ being the two independent 3×3 pseudo-inverse matrices of the neck and the eyes respectively. The computed joint velocities $\dot{\mathbf{q}}_N$, $\dot{\mathbf{q}}_E$ are then used as reference signals by the joint-level PID controllers.

This decoupling is beneficial for the stability of the system and it does not affect the final performance. The neck and the eyes are controlled to compensate two different components of the motion of the fixation point but cooperate to achieve the task. The rotational motion that is not compensated by the neck in fact produces translational velocities of the fixation point that are compensated by the eyes.

III. EXPERIMENTAL RESULTS

To validate our work we set up two experiments:

- *Exp. A: compensation of self-generated motion*: we issue a predefined sequence at the yaw, pitch, and roll of the torso and test both the *iKK* and the *iFB* conditions to proved a repeatable comparison between the two.
- *Exp. B: compensation in presence of an external perturbation*: the motion of the fixation point is caused by the experimenter who physically moves the torso of the robot. In this case there is no feedforward signal available, and the robot uses only the *iFB* signal.

For each experiment, two different sessions have been conducted: in the first session the robot stabilizes the gaze only with the eyes, while in the second session it uses both the neck and the eyes. In both the scenarios, a session without compensation has been performed and used as a baseline for comparison. It is worth noticing that Experiment A is obviously a more controlled scenario, and for this reason we have used it to obtain a quantitative analysis. In Experiment B instead the disturbances are generated manually, and, as such, it provides only a qualitative assessment of the performance of the *iFB* modality.

For validation we use the dense optical flow measured from the cameras. This can be used as an external, unbiased measure because as explained in Section I it is not used in the stabilization loop. We used the OpenCV [13] implementation of the dense optical flow algorithm proposed by Farneback [14]. Given an input image at time t, the method finds the 2D optical flow vector $\mathsf{of}(u, v)$ for each pixel in
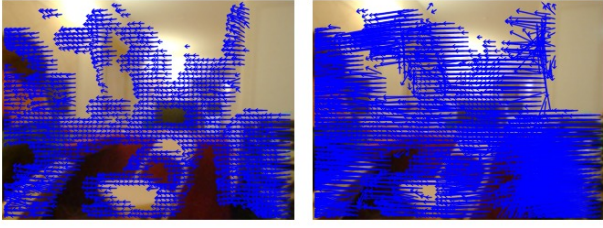
Fig. 3. Optical flow computed from two subsequent image frames from the left camera, baseline experiment (no compensation). Blue 2D arrows represent the optical flow vector $\mathtt{of}^t(u,v)$ at each pixel. For clarity optical flow vectors are reported only for a subset of the pixels (one pixel every five).

Fig. 4. Optical flow computed fro two subsequent image frames from the left camera, *iFB* experiment (compensation using inertial feedback). Blue 2D arrows represent the optical flow vector $\mathtt{of}^t(u,v)$ at each pixel. For clarity optical flow vectors are reported only for a subset of the pixels (one pixel every five).

the image. We derive a measure of performance by averaging the norm of the motion vectors $\mathtt{of}(u,v)$ in the whole image, i.e.:

$$\mathtt{optFl}(t) = \frac{1}{W-40 \times\ H-40} \sum_{u=20}^{W-20} \sum_{v=20}^{H-20} ||\mathtt{of}^t(u,v)||\ , \tag{16}$$

in which we remove from the computation the optical flow vectors of the peripheral region of the image. The reason for this is to compute a performance index that is more appropriate for the task, given that the gaze stabilization is computed for the fixation point (in this work $W = 320$, $H = 240$).

The optical flow computed during an experimental session is shown in Figure 3 and 4 for two consecutive frames in the baseline experiment (no compensation) and the iFb experiment (stabilization with inertial feedback) respectively. This qualitative evaluation shows that the stabilization effectively reduces the motion in the images. In the following Sections we provide a quantitative evaluation of our framework.

A. Compensation in presence of predefined torso movements

In experiment A we generate a set of predefined movements with the torso. We then compare the *kFF* and the *iFB* conditions with respect to the baseline. In all three cases we use the same sequence of velocity commands to the three torso joints (yaw, pitch and roll). Joints have been controlled with a velocity commands of 20 $\mathtt{deg/s}$) first independently and then simultaneously. As discussed in Section III, the controller has been tested in two cases: using only the 3 DoF of the eyes, and using all 6 DoF composed by the neck and
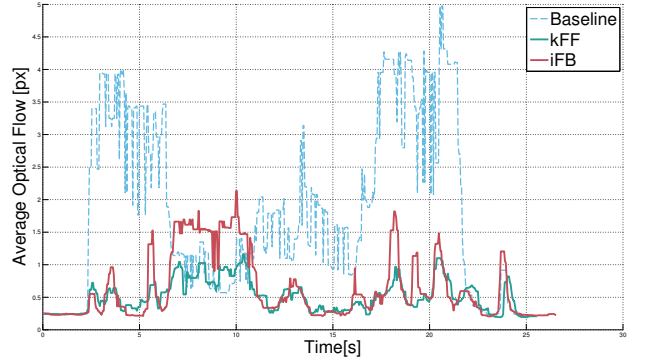

Fig. 5. Average Optical Flow during Experiment A. In this case only the eyes are controlled. The baseline session is the dashed blue line, while the *kFF* and *iFB* conditions are green and the red lines respectively.
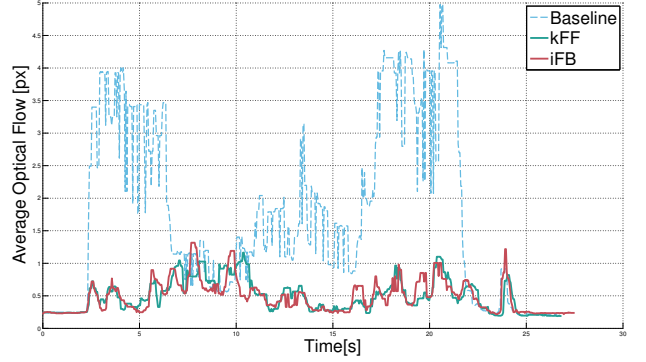

Fig. 6. Average optical flow during Experiment A. In this case stabilization uses all 6 Dof of the head. The baseline behavior is the dashed blue line, while the *kFF* and *iFB* conditions are green and red lines respectively.

the eyes. Figures 5 and 6 report the average optical flow $\mathtt{optFl}(t)$ in the two conditions respectively.

The two plots show the improvement of the stabilization with respect to the baseline (68.1% on average). As expected, the system performed better in the *kFF* condition than in the the *iFB* case (23.1% on average): this is because in the former case the system uses a feedforward command that anticipates and better compensates for the disturbances at the fixation point $\mathbf{x}_{FP}$. Furthermore, a comparison between Figure 5 and Figure 6 confirms that by exploiting all 6 DoFs in the head, the performance of the system improves by 24.4% on average. This occurs in particular when, during the sequence, the robot performs a large movement along the roll with the torso (roughly between $t = 6s$ and $t = 10s$, see also Figure 7). In this situation the optical flow in both the *kFF* and the *iFB* conditions has a peak because the disturbance cannot be compensated with the eyes. Indeed in this case the stabilization fails completely and actually produces unwanted motion (optical flow is higher than the baseline). Notice by comparison with Figure 6 that stabilization is more effective when the robot can exploit the additional DoFs of the neck.

B. Compensation of unknown disturbances

In experiment B the motors of the joints have been deactivated to allow a human operator to produce disturbances
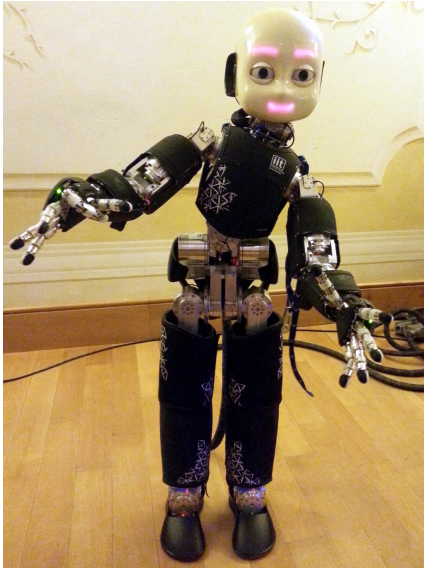
Fig. 7. The iCub compensating for the roll movement at the torso (Exp A, *kFF* scenario). In this particular occurrence, the stabilization is possible only with respect to the rotational component ω_{FP}^{roll}, since it is not physically feasible for the eyes to compensate such a movement.
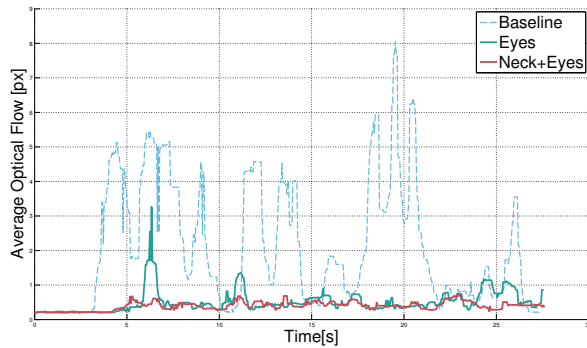

Fig. 8. Average optical flow during Experiment B. The blue dashed line represents the baseline. Green line is the optical flow when the stabilization uses only the eyes while green line is the optical flow when the stabilization uses all 6 DoF of the head.

by manually shaking the torso. This is by design a non-repeatable experiment, but it can act as a confirmation of the performances of the *iFB*. As for Experiment A the improvement of the stabilization with respect to the baseline are remarkable (78% on average), with an improvement of 21.4% when the robot uses all 6 DoF of the head.

IV. CONCLUSIONS AND FUTURE WORK

In this paper we described a framework for gaze stabilization of a humanoid robot. With respect to previous work we focus on the use of feedforward commands derived from the knowledge of the motor commands issued to the robot to improve stabilization when perturbations are generated by the robot own movements (e.g. locomotion or generic whole-body motion). To compensate for external perturbations we also include a feedback component provided by the inertial unit mounted on the head of the robot. Our experiments demonstrate that the feedforward component is effective for

stabilization when perturbations are due to the robot's own movement. We also demonstrate that proper integration of the DoFs of the neck in the control loop is crucial to achieve good stabilization.

In the experiments reported in this paper the robot compensated disturbances induced only by the motion of the upper body and we did not integrate the feedback and feedforward components. In addition optical flow was not used for the stabilization but only as a performance measure. This is therefore only a first step in the implementation of a full gaze stabilization system for a humanoid robot. As part of our future work we will investigate how to optimally integrate feedforward information with feedback coming from the inertial system and optical flow from the cameras. Furthermore, a natural extension of this framework is to integrate the information from the whole body of the iCub, including feedforward commands for all motors, feedback from the inertial units, torque sensors at the arms and legs as well as the tactile feedback from the skin.

REFERENCES

[1] Roger H.S. Carpenter, *Movements of the eyes*, ser. Medical. London, UK: Pion, 1988. [Online]. Available: http://www.ccsds.org/documents/pdf/CCSDS-101.0-B-4.pdf

[2] D. Coombs and C. Brown, "Real-time smooth pursuit tracking for a moving binocular robot," in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, Jun 1992, pp. 23–28.

[3] L. Berthouze, S. Rougeaux, F. Chavand, and Y. Kuniyoshi, "Calibration of a foveated wide-angle lens on an active vision head," in *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, Jun 1996, pp. 183–188.

[4] C. Capurro, F. Panerai, and G. Sandini, "Dynamic vergence using log-polar images," *International Journal of Computer Vision*, vol. 24.

[5] F. Panerai and G. Sandini, "Oculo-motor stabilization reflexes: integration of inertial and visual information," *Neural Networks*, vol. 11, no. 7-8, pp. 1191–1204, 1998.

[6] T. Shibata and S. Schaal, *Biomimetic gaze stabilization*. World Scientific, 2000, pp. 31–52.

[7] F. Panerai, G. Metta, and G. Sandini, "Learning visual stabilization reflexes in robots with moving eyes," *Neurocomputing*, vol. 48, no. 14, pp. 323 – 337, 2002.

[8] S. Gay, A. Ijspeert, and J. Santos-Victor, "Predictive gaze stabilization during periodic locomotion based on adaptive frequency oscillators," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 271–278.

[9] C. P. Santos, M. Oliveira, A. M. A. Rocha, and L. Costa, "Head motion stabilization during quadruped robot locomotion: Combining dynamical systems and a genetic algorithm," in *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, May 2009, pp. 2294–2299.

[10] U. Pattacini, "Modular cartesian controllers for humanoid robots: Design and implementation on the iCub," Ph.D. dissertation, Istituto Italiano di Tecnologia, Genova, Italy, 2011.

[11] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, A. Bernardino, and L. Montesano, "The iCub humanoid robot: An open-systems platform for research in cognitive development," *Neural Networks*, vol. 23, no. 8-9, pp. 1125–1134, 2010.

[12] Xsens website. [Online]. Available: http://www.xsens.com/

[13] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[14] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*, ser. Lecture Notes in Computer Science, J. Bigun and T. Gustavsson, Eds. Springer Berlin Heidelberg, 2003, vol. 2749, pp. 363–370.