Vision Based Motion Control for a Humanoid Head

L. C. Visser, R. Carloni and S. Stramigioli

Abstract— This paper describes the design of a motion
control algorithm for a humanoid robotic head, which consists
of a neck with four degrees of freedom and two eyes (a stereo
pair system) that tilt on a common axis and rotate sideways
freely. The kinematic and dynamic properties of the head are
analyzed and modeled using screw theory. The motion control
algorithm is designed to receive, as an input, the output of a
vision processing algorithm and to exploit the redundancy of
the system for the realization of the movements. This algorithm
is designed to enable the head to focus on and to follow a target,
showing human-like motions. The performance of the control
algorithm has been tested in a simulated environment and, then,
experimentally applied to the real humanoid head.

I. INTRODUCTION

In the last years, the research interest on humanoids
has increased and, within that, the interest in developing
humanoid heads. In the literature, there are two categories of
robotic head systems which basically differ in the complexity
of the mechanical design, i.e. structure and number of
degrees of freedom (DOFs), and in the movement speed they
can perform. For example, ASIMO [14] and Maveric [10]
rely, respectively, on 2 and 3 DOFs and can realize fast
movements while tracking objects. iCub [1] and QRIO [5]
can move slowly but they have, respectively, 3 and 4 DOFs so
to interact with humans by mimicking human-like motions.

The University of Twente, in collaboration with an in-
dustrial partner, has developed a humanoid head (i.e. a
complete system comprising a neck and two cameras). In
the purpose of developing a research platform for human-
machine interaction, the humanoid head should not only be
able to focus on and track targets, but also it should be able to
exhibit human-like motions, e.g. observing the environment
by expressing interest/curiosity and interacting with people.

The general mechanical design of the Twente humanoid
head is presented in [2] and it had to be a trade-off between
having few DOFs enabling fast motions and several DOFs
enabling complex, human-like motions/expressions. The final
choice was to have a four DOFs neck structure and three
DOFs for the vision system, as shown in Fig. 1. The
major contribution in the mechanical structure is due to the
introduction of a differential drive, which combines in a
small area the lower tilt (a rotation around the y-axis) and
the pan (around the z-axis) motions of the neck. The other
two degrees of freedom of the neck are the roll (around
the x-axis) and the upper tilt motions. Finally, the cameras,
mounted on a carrier, share an actuated tilt axis and can
rotate sideways freely, realizing three DOFs more.

{l.c.visser,r.carloni,s.stramigioli}@utwente.nl, IMPACT Institute, Faculty
of Electrical Engineering, Mathematics and Computer Science, University
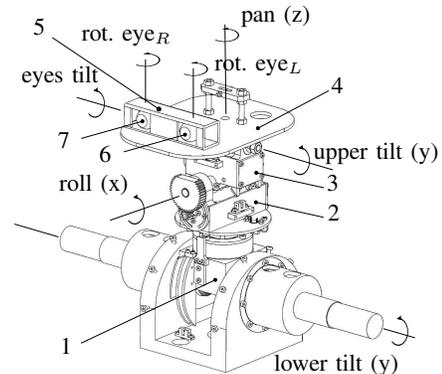of Twente, 7500 AE Enschede, The Netherlands.

Fig. 1. Mechanical design of the humanoid head - The mechanical design
consists of seven rigid bodies: a differential housing (body 1), two neck
elements (bodies 2, 3), the head plate (body 4), the eye carrier (body 5) and
two eyes (bodies 6, 7). The system has seven DOFs: the differential drive
combines the lower tilt and pan movement, the neck realizes the roll and
the upper tilt, the cameras tilt on a common axis, but rotate independently.

In this paper, we build the kinematic and the dynamic
model of this humanoid head based on screw theory and we
propose a motion control algorithm which uses the kinematic
properties of the model by exploiting the redundancy of the
mechanical structure. In particular, the control algorithm pro-
cesses the information from the vision system and, while the
target position is changing in the image plane, the humanoid
head can track it. Moreover, human-like motions/expressions
can be performed while looking at the target. This means
that through the control algorithm we are proposing, we
properly exploit the mechanical structure of the system so
to make the humanoid head move as described in biological
studies. In particular, we are aiming to realize the behavior
proposed in [6], according to which human beings use both
their head and eyes to track targets: the gaze (i.e. the angle
of the eyes with respect to a fixed reference) changes fast
due to the the fast and light-weight eyes moving towards
the target quickly, while the heavy head follows later and
slower. Fig. 2 shows a simulated one dimensional saccade,
i.e. an abrupt gaze change, which we reproduce in both a
simulation environment and the real setup.

II. MODEL OF THE HUMANOID HEAD

As shown in Fig. 1, the humanoid head consists of seven
rigid bodies, interconnected by joints. In order to facilitate al-
gorithm development and simulation testing, a kinematic and
dynamic model of the complete system has been developed
using screw theory, which provides the mathematical tools
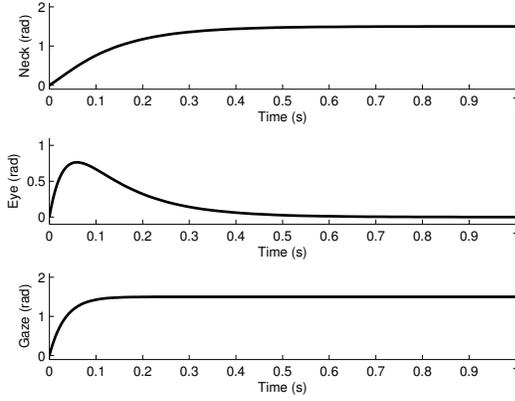to describe the relations between connected rigid bodies.

5469

Fig. 2. A simulated saccade of a human - The gaze (bottom) is defined as the angle of the eyes with respect to a fixed reference. The sum of the angle of the eyes with respect to the neck (middle) and the angle of the head with respect to the fixed reference (top) gives the gaze. The gaze quickly reaches the desired angle because of the fast movement of the eyes. The eyes keep the angle of the gaze constant by counter rotating to compensate for the relatively slow movement of the neck.

A. Rigid Bodies

With the aim of modeling a chain of rigid bodies, we recall here some notation of screw theory, see [11] for more details.

As depicted in Fig. 3, each rigid body i is characterized by a reference coordinate frame Ψ_i, centered in the joint connecting body i to a previous body $i-1$ and aligned with the joint rotation axis. Moreover, a principal inertia frame, Ψ_{i_p}, is centered in the center of mass (COM) of the body. This coordinate frame is chosen such that it is aligned with the principal inertia axes of the body so that the inertia tensor of body i, denoted by $\mathcal{I}^i$, is diagonal when expressed in this frame.

B. Kinematic Model

The humanoid head is made by seven rigid bodies interconnected by actuated joints. Each joint is characterized by a twist (generalized velocity) and wrench (generalized force) pair which defines the relative motion of the bodies connected by the joint. The relation between the scalar rotational velocity ω of the output shaft of each joint motor and its twist is given by $\mathbf{T} = \mathbf{J}\omega$, in which $\mathbf{J}$ is a six-dimensional column vector equal to the unit twist $\hat{\mathbf{T}}$.

The kinematic model of the complete system, i.e. the kinematic relation between the joint movements and the movements of each camera, gives the twist $\mathbf{T}^{0,0}_{\{L,R\}}$ of the left (Ψ_L) and right (Ψ_R) camera coordinate frame with respect to the global coordinate frame Ψ_0, expressed in Ψ_0

$$\mathbf{T}^{0,0}_{\{L,R\}} = \mathbf{J}_{\{L,R\}}(\mathbf{q})\,\dot{\mathbf{q}}, \qquad \mathbf{J}_{\{L,R\}} \in \mathbb{R}^{6\times7} \quad (1)$$

where $\mathbf{q} \in \mathcal{Q} \subseteq \mathbb{R}^7$ denotes the generalized joint states defined in the vector space $\mathcal{Q}$, $\dot{\mathbf{q}} \in T_q\mathcal{Q}$, the tangent space to $\mathcal{Q}$, its time derivative defined as the vector of the joints
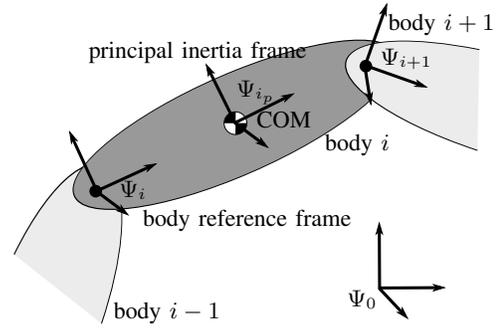

Fig. 3. Representation of a rigid body - The body coordinate frame Ψ_i is chosen to be coincident with the joint connecting it to the previous body. A principal inertia coordinate frame is defined in the center of mass, aligned with the principal axes of the body.

angular velocities

$$\dot{\mathbf{q}} = \begin{bmatrix} \omega_{\text{lower tilt}} \\ \omega_{\text{pan}} \\ \omega_{\text{roll}} \\ \omega_{\text{upper tilt}} \\ \omega_{\text{eyes tilt}} \\ \omega_{\text{roll left eye}} \\ \omega_{\text{roll right eye}} \end{bmatrix} \quad (2)$$

and the Jacobian matrices $\mathbf{J}_{\{L,R\}}$ are, for the left camera,

$$\mathbf{J}_L(\mathbf{q}) = \begin{bmatrix} \mathbf{J}_{\text{dec}} & \mathbf{J}_3 & \mathbf{J}_4 & \mathbf{J}_5 & \mathbf{J}_6 & \mathbf{0} \end{bmatrix} \quad (3)$$

and, for the right camera,

$$\mathbf{J}_R(\mathbf{q}) = \begin{bmatrix} \mathbf{J}_{\text{dec}} & \mathbf{J}_3 & \mathbf{J}_4 & \mathbf{J}_5 & \mathbf{0} & \mathbf{J}_7 \end{bmatrix} \quad (4)$$

where $\mathbf{J}_{\text{dec}} \in \mathbb{R}^{6\times2}$ is the Jacobian corresponding to the differential drive and $\mathbf{J}_i \in \mathbb{R}^6, i = 3,\dots,7$ are the Jacobians corresponding to the other joints. Note that the indices refer to the body numbers as defined in Fig. 1.

In order to build the kinematic model of the system, we derive the expression of the Jacobians $\mathbf{J}_{\{L,R\}}$. Except for the differential drive, the Jacobians describing the joint motions take the form of the unit twist with only one non-zero element, because the body coordinate frames are chosen to be aligned with the joint rotation axis. For example, for the roll motion of the neck (body 3), i.e. a rotation about the local x-axis, the Jacobian is

$$\mathbf{J}_3 = \hat{\mathbf{T}}^{0,2}_3 = \text{Ad}_{\mathbf{H}^0_2}\hat{\mathbf{T}}^{2,2}_3 = \text{Ad}_{\mathbf{H}^0_2}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T \quad (5)$$

where the unit twist $\hat{\mathbf{T}}^{2,2}_3$ gives the relative twists of the bodies 2 and 3 connected by the joint and expressed in Ψ_2, $\text{Ad}_{\mathbf{H}^0_2}$ is the adjoint of the matrix $\mathbf{H}^0_2$, the homogeneous matrix which defines the coordinate change from Ψ_2 to Ψ_0.

For the differential drive, the expression of the Jacobian $\mathbf{J}_{\text{dec}}(\mathbf{q})$ is different since the twist of the body attached to the differential drive is a function of two actuators. In order to explain the derivation of the twist of the differential drive, we refer to the schematic representation depicted in Fig. 4.

The generalized velocity of frame Ψ_1 (located in the center of the common gear) with respect to frame Ψ_0 as a function
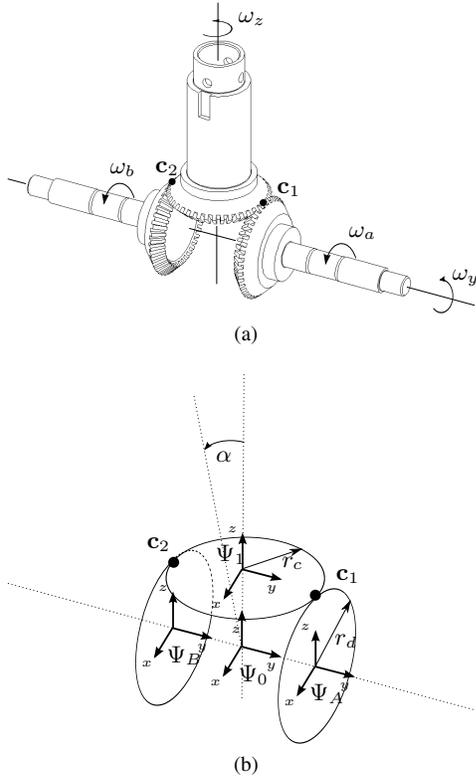
5470

(a)

(b)

Fig. 4. Schematic representation of the differential drive - Fig. 4a presents a schematic drawing of the design in which the motions of the common (upper) gear are constrained by the motion of the two driven gears. Fig. 4b presents a schematic representation of the differential drive with the definition of the coordinate frames.

of the rotational velocities of the driven gears, ω_a and ω_b, can be found by considering the constraints imposed on the contact points c_1 and c_2, which can be expressed in homogeneous coordinates in Ψ_0 as

$$
\begin{aligned}
\mathbf{c}_1^0 &= \begin{bmatrix} r_d \sin \alpha & r_c & r_d \cos \alpha & 1 \end{bmatrix}^T \\
\mathbf{c}_2^0 &= \begin{bmatrix} r_d \sin \alpha & -r_c & r_d \cos \alpha & 1 \end{bmatrix}^T
\end{aligned}
\tag{6}
$$

where the angle α is the angle of the z-axis of frame Ψ_1 with respect to the z-axis of frame Ψ_0, r_c and r_d are the radii of the common and driven gears. Note that, from Fig. 4, α is given by $\alpha = \frac{1}{2}(\theta_a + \theta_b)$, where θ_a and θ_b denote the angle rotated by the driven gears, i.e. the integral of ω_a and ω_b.

Let $\mathbf{p}_1$ be a point fixed in Ψ_1 and $\mathbf{p}_A$ be a point fixed in Ψ_A (on gear A). Furthermore, let both $\mathbf{p}_1$ and $\mathbf{p}_A$ be coincident with the contact point c_1. The linear velocity of $\mathbf{p}_1$ and $\mathbf{p}_A$, expressed in Ψ_0, must be equal when the gears are assumed to be ideal, i.e. no backlash. The linear velocity of $\mathbf{p}_1$ expressed in Ψ_0 is given by

$$
\dot{\mathbf{p}}_1^0 = \frac{d}{dt}\left(\mathbf{H}_1^0 \mathbf{p}_1^1\right) = \dot{\mathbf{H}}_1^0 \mathbf{p}_1^1 = \tilde{\mathbf{T}}_1^{0,0} \mathbf{p}_1^0
\tag{7}
$$

where $\mathbf{H}_1^0$ is a homogeneous matrix that defines the change of coordinates from Ψ_1 to Ψ_0 and $\tilde{\mathbf{T}}_1^{0,0}$ is the skew-symmetric twist corresponding to the motion of Ψ_1 with respect to Ψ_0, expressed in Ψ_0. The same result is obtained for $\mathbf{p}_A$

$$
\dot{\mathbf{p}}_A^0 = \frac{d}{dt}\left(\mathbf{H}_A^0 \mathbf{p}_A^A\right) = \tilde{\mathbf{T}}_A^{0,0} \mathbf{p}_A^0
\tag{8}
$$

Since $\mathbf{p}_1$ and $\mathbf{p}_A$ are both coincident with $\mathbf{c}_1$, Eq. (7) and Eq. (8) must be equal

$$
\tilde{\mathbf{T}}_1^{0,0} \mathbf{p}_1^0 = \tilde{\mathbf{T}}_A^{0,0} \mathbf{p}_A^0 \quad \rightarrow \quad \tilde{\mathbf{T}}_1^{0,0} \mathbf{c}_1^0 = \tilde{\mathbf{T}}_A^{0,0} \mathbf{c}_1^0
\tag{9}
$$

Analogously for $\mathbf{c}_2$

$$
\tilde{\mathbf{T}}_1^{0,0} \mathbf{c}_2^0 = \tilde{\mathbf{T}}_B^{0,0} \mathbf{c}_2^0
\tag{10}
$$

Note that $\tilde{\mathbf{T}}_A^{0,0}$ and $\tilde{\mathbf{T}}_B^{0,0}$ are the skew-symmetric twists corresponding to the rotations ω_a, ω_b and that the linear velocity of Ψ_1 expressed in Ψ_0 is zero by design

$$
\mathbf{T}_1^{0,0} = \begin{bmatrix} \omega^T & \mathbf{0} \end{bmatrix}^T
\tag{11}
$$

From this equation, Eqs. (6), (9), (10), and the definition of the Jacobian matrix, it follows that

$$
\mathbf{T}_1^{0,0} = \mathbf{J}_{\text{diff}} \begin{bmatrix} \omega_a \\ \omega_b \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\frac{r_d}{r_c}\sin\alpha & \frac{1}{2}\frac{r_d}{r_c}\sin\alpha \\ \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2}\frac{r_d}{r_c}\cos\alpha & \frac{1}{2}\frac{r_d}{r_c}\cos\alpha \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \omega_a \\ \omega_b \end{bmatrix}
\tag{12}
$$

In the mechanical design, there is a differential housing with a non-neglectable mass that rotates only along the y-axis of the differential drive. Therefore, $\mathbf{T}_1^{0,0}$ in Eq. (12) should be decoupled into two separate rotations along the y- and z-axes as

$$
\mathbf{T}_1^{0,0} = \mathbf{J}_{\text{dec}} \begin{bmatrix} \omega_y \\ \omega_z \end{bmatrix} = \begin{bmatrix} 0 & \sin\alpha \\ 1 & 0 \\ 0 & \cos\alpha \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2}(\omega_a + \omega_b) \\ \frac{1}{2}\frac{r_d}{r_c}(\omega_b - \omega_a) \end{bmatrix}
\tag{13}
$$

C. Dynamic Model

In order to simulate the complete humanoid head system, it is necessary to build the dynamic model, which is based on the kinematic model derived in the previous Sec. II-B. Bond graph theory provides the tools to describe the dynamics of the rigid bodies connected by the actuated joints.

The dynamic model of the complete system is based on screw theory. In particular, for each rigid body i, we can define the moment screw $\left(\mathcal{P}^i\right)^T = \mathcal{I}^i \mathbf{T}_i^{i,0}$ where $\mathcal{I}^i$ is the diagonal inertia tensor of the rigid body and $\mathbf{T}_i^{i,0}$ is the twist of the body fixed in its coordinate frame Ψ_i, with respect to the global coordinate frame Ψ_0 and expressed in Ψ_i.

By applying the second law of dynamics, it follows that the momentum of body i expressed in frame Ψ_0, and therefore its dynamics, is $\dot{\mathcal{P}}^{0,i} = \mathbf{W}^{0,i}$ where $\mathbf{W}^{0,i}$ represents the total wrench acting on body i, expressed in the global coordinate frame Ψ_0.

Finally, the system dynamics is given by

$$
\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{G}(\mathbf{q}) = \tau
\tag{14}
$$

where $\mathbf{M}(\mathbf{q})$ is the symmetric, positive definite mass matrix, $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}}$ describes the centrifugal and Coriolis forces, $\mathbf{G}(\mathbf{q})$ the forces due to gravity, τ is the torque applied to the joints.

5471

III. MOTION CONTROL

The vision processing algorithm determines where the humanoid head should look at by always choosing the most salient target point $\mathbf{x}$ in the image space $\mathcal{X}$ (see [13] for more details on the target selection algorithm). The output of this algorithm is supplied as input to the motion control algorithm which calculates the generalized desired joint velocities $\dot{\mathbf{q}}_{\mathbf{d}}$. In particular, the relation between the time derivative of vector $\mathbf{x}$, i.e. $\dot{\mathbf{x}}$, and $\dot{\mathbf{q}}$ is

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{q})\,\dot{\mathbf{q}} \tag{15}$$

where $\dot{\mathbf{x}} \in T_x\mathcal{X}$, the tangent space to $\mathcal{X}$, and the map $\mathbf{F} : T_q\mathcal{Q} \to T_x\mathcal{X}$.

A. Target Perception

Before entering into the details of the design of the visual servoing control algorithm, it is required to know how the target perception by the cameras changes during the joint movements, i.e Eq. (15). From the pinhole camera model [9], it follows that the target coordinates are defined as the projection of the target on the image plane in the camera coordinate frame, as is shown in Fig. 5. Let

$$\mathbf{p}^{\{L,R\}} = \begin{bmatrix} x & y & z \end{bmatrix}^T \tag{16}$$

be a target point in three dimensional Euclidean space $\mathcal{E}(3)$, expressed in coordinate frame $\Psi_{\{L,R\}}$, for the left and right cameras. The projection of this target point, expressed in the camera coordinate frame, $\mathbf{p}_{\text{proj}}^{\{L,R\}}$, is given by

$$\mathbf{p}_{\text{proj}}^{\{L,R\}} = \begin{bmatrix} y_{\text{proj}} \\ z_{\text{proj}} \end{bmatrix}^{\{L,R\}} = \frac{f}{x^{\{L,R\}}} \begin{bmatrix} y \\ z \end{bmatrix}^{\{L,R\}} \tag{17}$$

where f is the focal depth of the camera.

Assuming that the origin of the camera coordinate frame is located in the center of the image, focus on the target is to be interpreted as $\mathbf{p}_{\text{proj}}$ being in $\mathbf{x_0} = (0,0)$ for both cameras. Therefore, the vector of target coordinates $\mathbf{x}$ is defined as

$$\mathbf{x} = \begin{bmatrix} \mathbf{p}_{\text{proj}}^{\mathbf{L}} \\ \mathbf{p}_{\text{proj}}^{\mathbf{R}} \end{bmatrix} \tag{18}$$

B. Target Perception and Joint Movement

From Eq. (17), it follows that when the camera coordinate frame moves, the projection is affected because $\mathbf{p}^{\{L,R\}}$ changes. An expression for the instantaneous rate of change of $\mathbf{p}^{\{L,R\}}$, denoted by $\dot{\mathbf{p}}^{\{L,R\}}$ and caused by the joint movement, can be found by assuming the situation as depicted in Fig. 5.

Let the homogeneous coordinates of the target, expressed in the left camera coordinate frame Ψ_L, be given by

$$\begin{bmatrix} \mathbf{P}^L \\ 1 \end{bmatrix} = \mathbf{H}_0^L \begin{bmatrix} \mathbf{P}^0 \\ 1 \end{bmatrix} \tag{19}$$

where $\mathbf{p}^0$ denotes the target coordinates in Ψ_0. The linear velocity of $\mathbf{p}^L$ expressed in Ψ_L is found by differentiating Eq. (19) with respect to time, yielding

$$\begin{bmatrix} \dot{\mathbf{P}}^L \\ 0 \end{bmatrix} = \dot{\mathbf{H}}_0^L \begin{bmatrix} \mathbf{P}^0 \\ 1 \end{bmatrix} \tag{20}$$
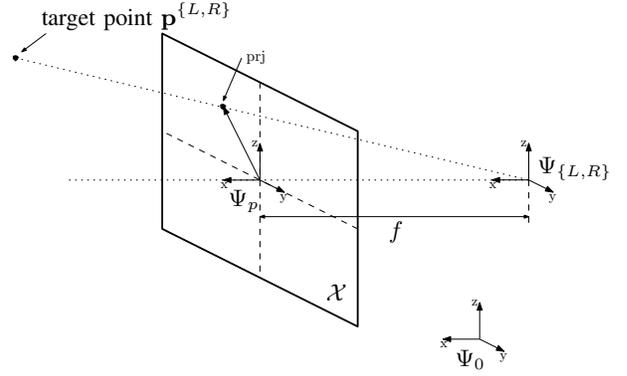

Fig. 5. Target coordinates - The target coordinates as perceived by the cameras can be modeled by a projection on the image plane $\mathcal{X}$ using a pinhole camera model. The image plane is at focal depth f on the x-axis of the camera frame. $\Psi_{\{L,R\}}$ denotes the left and right camera coordinate frames, respectively.

where we are considering the instantaneous case in which $\dot{\mathbf{p}}^0 = 0$. By using the relation $\dot{\mathbf{H}}_0^L = \tilde{\mathbf{T}}_0^{L,L} \mathbf{H}_0^L$ we obtain

$$\begin{bmatrix} \dot{\mathbf{P}}^L \\ 0 \end{bmatrix} = \tilde{\mathbf{T}}_0^{L,L} \mathbf{H}_0^L \begin{bmatrix} \mathbf{P}^0 \\ 1 \end{bmatrix} = \tilde{\mathbf{T}}_0^{L,L} \begin{bmatrix} \mathbf{P}^L \\ 1 \end{bmatrix} \tag{21}$$

that can be also written as

$$\dot{\mathbf{p}}^L = \begin{bmatrix} -\tilde{\mathbf{p}}^L & \mathbf{I}_3 \end{bmatrix} \mathbf{T}_0^{L,L} \tag{22}$$

The twist $\mathbf{T}_0^{L,L}$ is given by

$$\mathbf{T}_0^{L,L} = -\mathrm{Ad}_{\mathbf{H}_0^L} \mathbf{T}_L^{0,0} \tag{23}$$

by noting that $\mathbf{T}_j^{i,i} = -\mathbf{T}_i^{i,j}$ and $\mathbf{T}_i^{i,j} = \mathrm{Ad}_{\mathbf{H}_j^i} \mathbf{T}_i^{j,j}$. Finally, from Eq. (1), it follows that

$$\dot{\mathbf{p}}^L = \begin{bmatrix} \tilde{\mathbf{p}}^L & -\mathbf{I}_3 \end{bmatrix} \mathrm{Ad}_{\mathbf{H}_0^L} \mathbf{J}_L(\mathbf{q})\,\dot{\mathbf{q}} \tag{24}$$

From Eq. (17), $\mathbf{p}_{\text{proj}}^L$ is found to be a scaled version of $\mathbf{p}^L$, and therefore

$$\dot{\mathbf{p}}_{\text{proj}}^L = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{P}}_{\text{proj}}^L & -\mathbf{I}_3 \end{bmatrix} \mathrm{Ad}_{\mathbf{H}_0^L} \mathbf{J}_L(\mathbf{q})\,\dot{\mathbf{q}} \tag{25}$$

with

$$\mathbf{P}_{\text{proj}}^L = \begin{bmatrix} f \\ \mathbf{p}_{\text{proj}}^L \end{bmatrix} \tag{26}$$

where the projected target $\mathbf{p}_{\text{proj}}^L$ is given by Eq. (16) and where $\dot{\mathbf{p}}_{\text{proj}}^L$ denotes the two dimensional velocity vector, i.e. the instantaneous velocity of the observed target on the image plane.

With the same approach, we find a similar expression for the right camera, and by combining these results we obtain the expression of the matrix $\mathbf{F}$ in Eq. (15)

$$\mathbf{F}(\mathbf{q}) = \begin{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{P}}_{\text{proj}}^L & -\mathbf{I}_3 \end{bmatrix} \mathrm{Ad}_{\mathbf{H}_0^L} \mathbf{J}_L(\mathbf{q}) \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{P}}_{\text{proj}}^R & -\mathbf{I}_3 \end{bmatrix} \mathrm{Ad}_{\mathbf{H}_0^R} \mathbf{J}_R(\mathbf{q}) \end{bmatrix} \tag{27}$$

C. Control Law

The goal of the visual servoing control law is to move the perceived target coordinates $\mathbf{p}_{\text{proj}}^{\{L,R\}}$ to $\mathbf{x_0} = \mathbf{0}$, i.e. in the center of the camera image. This is obtained by a proportional control law in the image space given by

$$\dot{\mathbf{x}}_{\mathbf{d}} = K(\mathbf{x_0} - \mathbf{x}) = -K \begin{bmatrix} \mathbf{p}_{\text{proj}}^{L} \\ \mathbf{p}_{\text{proj}}^{R} \end{bmatrix} \qquad (28)$$

where $\dot{\mathbf{x}}_{\mathbf{d}}$ is the desired target point velocity in the image plane and $K > 0$ is the proportional gain matrix.

In order to apply this control law, it is required to invert the relation (15). Since the system is redundant, the solution is given by [7]

$$\dot{\mathbf{q}}_{\mathbf{d}} = \mathbf{F}^{\sharp} \dot{\mathbf{x}}_{\mathbf{d}} + \left(\mathbf{I}_7 - \mathbf{F}^{\sharp} \mathbf{F} \right) \mathbf{z} \qquad (29)$$

where $\dot{\mathbf{q}}_{\mathbf{d}}$ is the desired joint velocity, $\mathbf{F}^{\sharp} : T_x \mathcal{X} \to T_q \mathcal{Q}$ denotes the weighted generalized pseudo inverse of the map $\mathbf{F}$ and $\mathbf{z} \in T_q \mathcal{Q}$ is an arbitrary vector which is projected onto the null-space of $\mathbf{F}$. Note that $\mathbf{F}^{\sharp}$ is given by

$$\mathbf{F}^{\sharp} := \mathbf{M}_{\mathbf{q}}^{-1} \mathbf{F}^{T} (\mathbf{F} \mathbf{M}_{\mathbf{q}}^{-1} \mathbf{F}^{T})^{-1} \qquad (30)$$

where $\mathbf{M}_{\mathbf{q}}$ is a positive definite diagonal matrix that defines a metric on the tangent space $T_q \mathcal{Q}$. The first right-hand term of Eq. (29) is a minimum norm solution, where the norm is defined by the matrix $\mathbf{M}_{\mathbf{q}}$ [4]

$$\|\dot{\mathbf{q}}\| = \sqrt{\dot{\mathbf{q}}^T \mathbf{M}_{\mathbf{q}} \dot{\mathbf{q}}} \qquad (31)$$

Fig. 6 visualizes the minimum norm solution for a two dimensional system. The plotted surface represents the norm (31) for a given set of values of $\dot{\mathbf{q}}$. The line represents all solutions to Eq. (29) for a given set of values of $\dot{\mathbf{x}}_{\mathbf{d}}$ and $\mathbf{z}$ and its minimum (marked with •) is the minimum norm solution obtained for $\mathbf{z} = 0$.

Note that $\mathbf{M}_{\mathbf{q}} = \text{diag}(m_{q_i}) > 0, i = 1, \dots, 7$ in which the first four elements refer to the neck and the remaining three refer to the eyes. This implies that by choosing the matrix $\mathbf{M}_{\mathbf{q}}$ appropriately, we can select the ratio between the velocities of the eyes and of the neck joints. In particular, we select the velocity of the eyes greater than the one of the head since we want the eyes to be faster than the neck.

In target tracking, we select the vector $\mathbf{z}$ in (29) as

$$\mathbf{z} = \mathbf{z}(\mathbf{q}) = \mathbf{W} (\mathbf{q_0} - \mathbf{q}) \qquad (32)$$

where $\mathbf{W} = \text{diag}(w_i) \geq 0, i = 1, \dots, 7$ in which, again, the first four elements refer to the neck and the remaining three refer to the eyes. This means that vector $\mathbf{z}$ is a proportional control that, through motions in the null-space, steers the joint configuration $\mathbf{q}$ to a desired neutral configuration $\mathbf{q_0}$, in which the head is in a upright position and the eyes straight to the target, i.e.

$$\begin{aligned} \mathbf{q_0} &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T \\ \mathbf{W} &= \text{diag}(w_1, 0, w_3, 0, w_5, w_6, w_7) \end{aligned} \qquad (33)$$

Note that w_2, w_4 are equal to zero to require the humanoid head to look at the target by using the pan and the upper tilt.
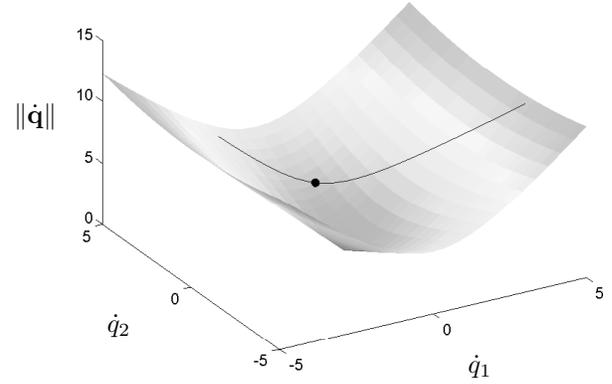

Fig. 6. Minimum norm solution for a two dimensional system - The plotted surface represents the norm (31) for a given set of values of $\dot{\mathbf{q}}$. The line represents all solutions to Eq. (29) for a given set of values of $\dot{\mathbf{x}}$ and $\mathbf{z}$ and its minimum (marked with •) is the minimum norm solution, for $\mathbf{z} = 0$.
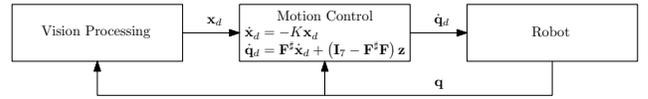

Fig. 7. Controller overview - The vision algorithm provides the motion control algorithm with target coordinates. From this, the controller calculates the joint velocities $\dot{\mathbf{q}}_{\mathbf{d}}$.

The other values w_1, w_3, w_5, w_6, w_7 are positive to steer all the other joints to the neutral position.

The vector $\mathbf{z}$ in Eq. (29) is also used to achieve expression motions while the head is looking at a target, like nodding in agreement, shaking on disagreement, moving the head backwards in surprise or moving the head towards the target in curiosity. These motions can be generated by applying an appropriate time varying function to one or more of the joints. For example, nodding can be achieved by taking $\mathbf{z}$

$$\mathbf{z} = \mathbf{z}(t) = \begin{bmatrix} \alpha \sin t & 0 & 0 & \beta \sin t & 0 & 0 & 0 \end{bmatrix}^T \qquad (34)$$

with the parameters α and β that define the speed of the motion. This will result in a nodding motion of the head while it keeps aimed at the target at the same time.

An overview of the controller structure is shown in Fig. 7.

D. Stability Analysis

In Eq. 14, the torque τ applied to the joints is determined by a PD controller with gravity compensation, i.e.

$$\tau(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{q_d}, \dot{\mathbf{q}}_{\mathbf{d}}) = K_P(\mathbf{q_d} - \mathbf{q}) + K_D(\dot{\mathbf{q}}_{\mathbf{d}} - \dot{\mathbf{q}}) + \mathbf{G}(\mathbf{q})$$

where $\mathbf{q_d}$ is the desired joint position derived from Eq. (29), K_P and K_D are the positive definite gain matrices.

By following the arguments in [8], it is possible to show that, by properly choosing the gain matrices K_P and K_D, the dynamic control law guarantees the asymptotic tracking of a desired trajectory in the image plane.

5473

IV. SIMULATION RESULTS

The dynamic model and the motion control algorithm have been implemented in a simulation environment using 20-sim simulation software [3]. The cameras have been modeled using the pinhole camera model given by Eq. (17) together with a delay due to the time that the vision processing algorithm needs to process the images.

From biological studies [6], it follows that for humans the peak velocity ratio between the head and the eyes is about $1 : 5$. This ratio is used to define the generalized pseudo inverse in Eq. (29)

$$\mathbf{M_q} = \mathrm{diag}(5.0, 5.0, 50.0, 5.0, 1.0, 1.0, 1.0) \qquad (35)$$

in which the contribution of the neck roll motion is penalized more in the implementation since this motion is minimally used by humans.

Fig. 8 presents the simulation results in continuous line. The figure shows the time response of the joint angles for the pan motion of the neck and the pan motion of the left eye during a saccade. It can be seen that the behavior of the simulated head matches the results of the simulated human behavior plotted in Fig. 2. The main difference between the plots of the simulated head and the simulated human is in the time scale. This is principally due to the trade-off between getting as close as possible to the human capabilities and what was really feasible with the real setup, considering the restrictions imposed by the mechanical design/realization.

V. EXPERIMENTAL RESULTS

The tests performed in the simulated environment have been implemented in the real setup and the results are shown in the accompanying video, see also [12] for a complete overview of the system.

The experimental realization of a horizontal saccade is illustrated in Fig. 8 in dashed line, on the top of the simulation results. The time plots of the pan angles for the neck and the left eye of the real setup correspond to the plots of the simulated model.

VI. CONCLUSIONS

A motion control algorithm for a humanoid head has been designed. The algorithm acts on the inputs of a vision processing algorithm and can control the humanoid head according the results of biological studies. This has been achieved by appropriately actuating the redundant joints using a null-space projection method. A kinematic and a dynamic models based on screw theory and bond graphs have been developed and have been used to test the motion control algorithm in a simulated environment. Simulations have shown that both saccades and target tracking tasks can be performed. Finally, experiments on the real setup have validated the model and the control algorithm.
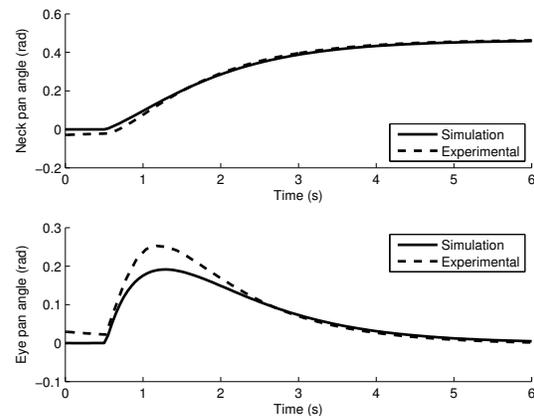

Fig. 8. Time plots of the joint angles in a saccade in simulation and experiment - The upper plot shows the neck pan angle over time while the bottom plot shows the pan angle of the eyes.

REFERENCES

[1] R. Beira, M. Lopes, M. Praca, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi and R. Saltaren, "Design of the robot-cub (iCub) head", *IEEE Int. Conf. Robotics and Automation*, 2006.

[2] D. M. Brouwer, J. Bennik, J. Leideman, H.M.J.R. Soemers and S. Stramigioli, "Mechatronic design of a fast and long range 4 degrees of freedom humanoid neck", *IEEE Int. Conf. Robotics and Automation*, 2009.

[3] Controllab Products B.V., *20-sim*, http://www.20sim.com, 2009.

[4] K. Doty, C. Melchiorri and C. Bonivento, "A theory of generalized inverses applied to robotics", *Int. Jour. Robotic Research*, vol. 12, n. 1, pp 1-19, 1993.

[5] L. Geppert, "QRIO, the robot that could", *IEEE Spectrum*, vol. 41, n. 5, pp 34-37, 2004.

[6] H.H.L.M. Goossens and A.J. van Opstal, "Human eye-head coordination in two dimensions under different sensorimotor conditions", *Jour. Experimental Brain Research*, vol. 114, n. 3, pp 542-560, 1997.

[7] A. Liégeois, "Automatic supervisory control of the configuration and behavior of multibody mechanisms", *IEEE Trans. Systems, Man and Cybernetics*, vol. 7, n. 12, pp 868-871, 1977.

[8] P. Hsu, J. Hauser and S. Sastry "Dynamic control of redundant manipulators", *IEEE Int. Conf. Robotics and Automation*, 1988.

[9] Y. Ma, S. Soatto, J. Kosecka and S. Sastry, *An invitation to 3-D vision*, Springer, 2006.

[10] Maveric, http://www-humanoids.usc.edu/HH_summary.html.

[11] R. Murray, Z. Li and S. Sastry, *A mathematical introduction to robotic manipulation*, CRC Press, 1994.

[12] R. Reilink, L.C. Visser, J. Bennik, R. Carloni, D. M. Brouwer and S. Stramigioli, "The Twente humanoid head", *IEEE Int. Conf. Robotics and Automation*, 2009.

[13] R. Reilink, S. Stramigioli, F. van der Heijden and G. van Oort, "Realtime stereo vision processing for a humanoid", *Internal Report 019CE2008, University of Twente*, 2008.

[14] Y. Sagakami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki and K. Fujimura, "The intelligent ASIMO: system overview and integration", *IEEE Int. Conf. Intelligent Robots and Systems*, 2002.

[15] S. Vijayakumar, J. Conradt, T. Shibata and S. Schaal, "Overt visual attention for a humanoid robot", *IEEE Int. Conf. Intelligent Robots and Systems*, 2001.