VISUAL TRACKING ON AN AUTONOMOUS SELF-CONTAINED HUMANOID ROBOT

MAURO RODRIGUES, FILIPE SILVA
Department of Electronics, Telecommunications and Informatics, University of Aveiro
Aveiro, 3810-193, Portugal

VÍTOR SANTOS
Department of Mechanical Engineering, University of Aveiro
Aveiro, 3810-193, Portugal

This paper describes the hardware and software setups that allow a humanoid robot developed from scratch to perform visual tracking based exclusively on onboard components. The robot which was started on earlier projects was finally given its full autonomy in what concerns perception and vision capabilities. An embedded PC104-based controller running Linux is now able to interface a IEEE1394 color camera and, using the OpenCV library, can now perform visual tracking of some objects moving on its neighborhood. This embedded controller, besides being responsible for image acquisition and processing, serves as an interface between external monitoring and the distributed control architecture based on a master-multi-slave CAN bus of local controllers for joint actuation and sensor monitoring.

1. Introduction

Most roboticists, at least among the young community, have certainly dreamed about building a humanoid robot able to move, perceive and possibly act like humans do. The dream has already come true for some people and is promising every day to become true for a broader set of serious enthusiasts, mainly scientists and engineers [1-3]. Concerning the engineering perspective, the ultimate challenge is certainly related to full real autonomy, both in power and decisions. Current trends in electronics and embedded systems ensure that is it indeed possible with specially developed hardware, but a real extra to the mentioned challenge is to use off-the-shelf components [4-5].

Within earlier activities of the authors [6-7] a humanoid prototype has been developed. That includes the mechanical structure, the entire hardware for a distributed control architecture, and also some simple force sensors along with a fairly complex control system for RC servomotors that are used for joint actuation. This 22 joint robot, depicted in Figure 1, has proven able to sustain

balance based on simple local control using the feet force sensors, and is reaching a stage where locomotion is the inevitable next step.

This paper addresses the problem of tracking a moving target using a single camera mounted on a pan-tilt unit (PTU) which is actually the robot's neck. The whole perception-decision-action process is discussed, including the image processing and the control loop. The development of effective methods for performing this task represents a challenging testbed when considering full autonomous self-contained small-size platforms. An intense research activity has been done aimed at providing soccer playing humanoid robots with vision capabilities. One common approach is the use of local/global omni-directional vision to detect the objects of interest [8-10]. Instead, the core of this proposal is the distributed control architecture that provides sufficient computing power to allow the implementation of simple stimulus-based behaviors. Here, the monocular active vision system can align with and track a moving target, following a natural procedure closer to the biological counterparts.

This paper describes the first steps of a long term effort towards developing an active observer using vision to interact with the environment, in particular capable of visually guided navigation. The remainder of the paper is organized as follows: Section 2 describes the hardware and software components of the vision system and central unit. Section 3 discusses the visual tracking approaches and the proposed method to use the inclination of the trunk. Section 4 presents experimental results illustrating the system's performance obtained with these onboard capabilities. Section 5 concludes the paper and presents the perspectives of future work.

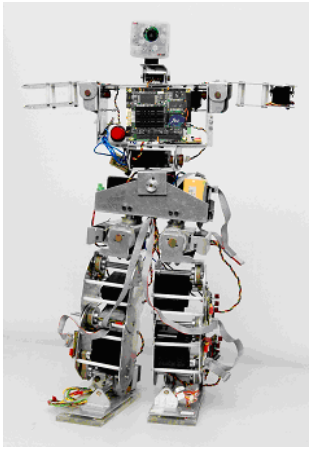


Figure 1. The humanoid robot with 20-DOF. Note the head-mounted CCD camera and the pan-tilt head system (2-DOF).

2. Vision System Architecture

It is a challenge to operate a humanoid robot in dynamic environments that require continuous decision making and feature updates in real-time. In order to provide the robot with basic navigation abilities and to perceive the structure of the environment, visual information is considered both essential and appealing. In the context of robot-soccer competitions (e.g., RoboCup, FIRA), the vision system plays a key role to sense the environment and to identify target locations, which can be used to update the motion sequence to reach the current goal.

In this section, the role of image acquisition and processing is described in more detail, as well as the motion directives to the minimal joints to perform visual tracking.

2.1. *Hardware Setup*

The distributed control architecture relies on a master controller that manages a CAN bus where several slave controllers interface locally at the low-level, providing tasks like PWM generation, all joint control and gathering of all sensor information. This master unit has no superior autonomy on its own since it only keeps the state and maintains dialog with all the slave controllers. Motion directives must be decided by a higher level unit - the central unit.

Up to now, that role has been ensured by an external ordinary computer that communicated with the master by means of a serial link. From now onwards, the central unit is embedded in the system making it able to reach complete autonomy, including complex perception such as vision and related algorithms. The central unit consists of an embedded PC104-based controller running Linux (2.6.18 kernel) and it is based on an AMD Geode Processor at 500 MHz. This processing unit has the following major roles: capture images and process them, interfaces to remote monitoring and possibly control and finally communicates with the master controller of the distributed architecture of slave controllers seeded all over the robot's body.

The vision system is essentially composed of a low cost FireWire camera yielding images up to VGA resolution in RBG or several compressed (YUV) formats at multiple frame rates (maximum is 30 fps). The camera lies atop a 2-DOF pan-and-tilt system that comprises the "neck" of the robot, being linked to the central unit through a standard PCMCIA interface.

2.2. *Software*

The vision system is implemented as a monolithic architecture, with all the image processing carried out within one process. The purpose of the system is to direct attention to a particular object in space - the ball. Color images sampled by the CCD camera were processed at a frame rate of approximately 25Hz. The raw visual image is first sampled at a resolution of 320x240 pixels, and then down-sampled to 160x120 pixels. The object of interest is detected using simple color-based analysis done in the HSV color space. The next step is the calculation of the ball's centroid coordinates in the image plane. In order to reduce the region of search during visual tracking, a variable region of interest (ROI) is adopted between successive images. Vision abilities were also added with the *OpenCV* framework installed on-board. An example of the results obtained with the image processing is shown in Figure 2.

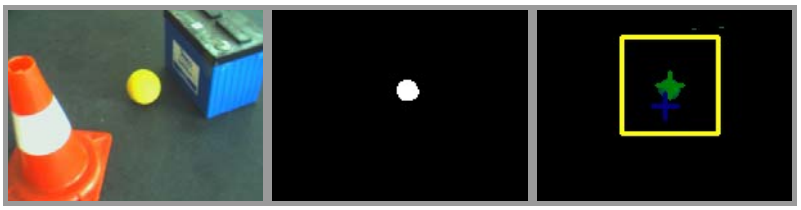


Figure 2. The image processing main steps: (*a*) raw-image acquisition (left); (*b*) color segmentation and noise filtering (middle); and (*c*) ball's centroid location in the image plane and the ROI (right).

3. Visual Detection and Tracking

A major goal of this work is to formalize and implement a strategy in order to achieve an intelligent task-oriented active vision system. This section describes the vision-based approach to address the problems of tracking a moving target through visual feedback. The desired functionalities for such a system can be summarized as follows: the vision system must be able to identify the ball in the scene, align its camera towards the object by properly controlling the pan-tilt unit and maintain its focus on the selected object.

3.1. *Stimuli-Based Actions*

A humanoid robot faces complex challenges when playing soccer games, depending mostly on visual information to guide many of their behaviors and actions. First, the robot has to search the ball that might be in any position on the ground of the field (or even jumping). Second, the moving target has a

generic trajectory which is unknown in advance. The visual detection and the alignment with the ball are necessary tasks whenever the ball disappears from the field of vision. In this case, the pan-tilt system receives orders to perform a scanning procedure searching for the ball in the joints' physical limits.

At the same time, the tracking subsystem plays an important role in the future implementation of a scheme for visually guided locomotion control (*e.g.*, interception of a moving target). The image-plane can be used to obtain a coarse estimation of the relative position of the ball with respect to the base coordinate system of the robot. The restriction that all relevant objects are located on the ground of the field provides an one-to-one correspondence between a point on the ground plane and its observation in the image plane.

In what concerns the control of the pan-tilt angles, many different visual servoing approaches have been proposed in literature [11]. Our approach is to design a scheme in which the control values are computed directly on the basis of the image features. Since the camera is equipped with a narrow-angular lens, the angle of inclination of the trunk, in the saggital plane, is used in order to increase the field of vision. This extra degree of freedom allows the robot to see its own feet and enables the execution of a joint limit avoidance task.

3.2. *Image-Based Approach*

The spatial relationship between the target and the camera is directly estimated on the image plane and the error vector is expressed in terms of image features. More concretely, the acquired image is processed in order to obtain the coordinates of the ball's centroid with respect to the centre of the image plane, *i.e.*, the ball offset. The control objective is to keep the target close to the centre of the video image, while reducing the computational delay and errors due to camera calibration and/or sensor modeling. The choice of the relationship between joint angles and ball offset allows finding algorithms with different performances. The simplest method is the straight forward application of a proportional law such as:

$$\dot{\mathrm{q}} = K\mathrm{e}. \tag{1}$$

Here, $\dot{\mathrm{q}} = \left[\dot{q}_p, \dot{q}_t\right]^T$ is the joint velocity vector defined by the pan and tilt angles, K is a diagonal matrix of constant gains and $\mathrm{e} = \left[C_x, C_y\right]^T$ is the error vector defined by the ball offset. This approach has the main advantage of its simplicity: each component of the error vector relates directly and in an independent way to the pan and tilt joints. However, this approach suffers from a major drawback. It turned out to be difficult to use the same controller gains

both when the target is moving slowly/fast and when it is close/far way from the robot. To address this drawback, the control objective is modified aiming to regulate the orientation of the pan-tilt unit (and hence, the camera) so that the projection of the target object in image space coordinates is close to zero. This approach, based on the image Jacobian [12], can be expressed as follows:

$$\dot{\mathrm{q}} = J_R^{-1} J_I^{-1} K e \tag{2}$$

where J_R^{-1} is the inverse of the Jacobian matrix which transforms the differential variation of the camera's orientation into the differential variation in the joint space, J_I^{-1} is the inverse of the image Jacobian which relates the image feature parameters to the camera orientation coordinates and K is a diagonal matrix properly chosen to ensure convergence to a bounded error.

Assuming the camera orientation is restricted so that it is always pointing to the ground, the design of the visual servoing controller reduces to the following expressions:

$$\dot{q}_{\mathrm{p}} = \frac{K_p C_x}{\cos(q_t)}, \; \dot{q}_{\mathrm{t}} = \frac{K_t C_y}{\sin(q_t)}. \tag{3}$$

Once again, for a given position of the ball, the coordinates in the image plane, C_x and C_y, only affect the pan and the tilt angles, respectively. However, the terms in denominator can be seen as adaptive gains, compensating for the lag in the tracking.

4. Experiments and Results

In order to evaluate the vision system's performance, several experiments were carried out with the embedded controller running at 500MHz with 512MB of RAM. Our experimental platform comprises a color CCD video camera and a pan-tilt head, both mounted on the top of the humanoid robot that remains in a static upright posture during the entire execution of each trial.

Table 1 shows the upper bounds on the execution times of the visual processing, control calculations for visual servoing and communication with the master unit. Pan and tilt servos are controlled by a slave unit that receives the commands issued from the central unit to the master unit that manages the CAN bus. The positions of the servos are updated at a frequency of 50Hz.

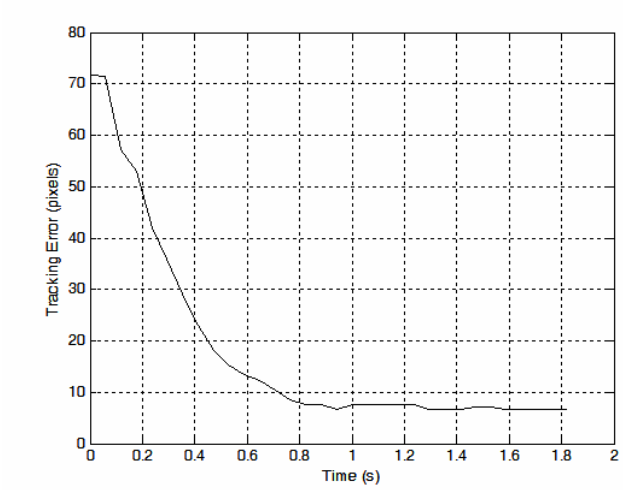
Regarding the vision processing algorithms described above, it could be expected a considerable variation in the process execution times since, in some cases, the ball may be found almost immediately or, in other cases, it may be necessary to analyze the entire image (*e.g.*, when the ball disappears from the field of vision). However, the average execution time to complete the process

rounds about 40 ms. This means that color images sampled by the CCD camera are processed at a frame rate of approximately 25Hz, which is sufficient to allow fast-moving stimuli or other kinds of rapidly changing visual input.

Table 1. Upper bounds on the process execution times of the vision processing in the onboard central unit.

Algorithm	Max. (ms)	Min. (ms)	Avg. (ms)	St. Dev. (m
Acquisition	32.4	11.9	**13.7**	2.0
Pyr Down	25.9	9.5	**9.8**	1. 6
Segmentation/filtering	41.6	9.3	**9.9**	2.4
Centroid location	3.3	0.6	**1. 3**	0.5
Actuation	37.4	2.2	**4.5**	2.7

Several experiments were carried out in order to verify the implementation and to evaluate the performance of the algorithms proposed in the previous section. Figure 3 illustrates the actual image space tracking errors calculated as the norm of the ball offset, both in alignment and tracking. As can be observed from the tracking error plots, the controller reaches the alignment in approximately 1s, while the tracking errors reach approximately 20 pixels when the ball is moving. Decreasing the tracking errors can be achieved by using higher control gains which makes the robustness results, over the whole workspace, poor from a practical point of view. In order to increase the practical applicability of the approach, on going research is aimed at avoiding the high gains paradigm as the only mean of reducing the tracking errors.



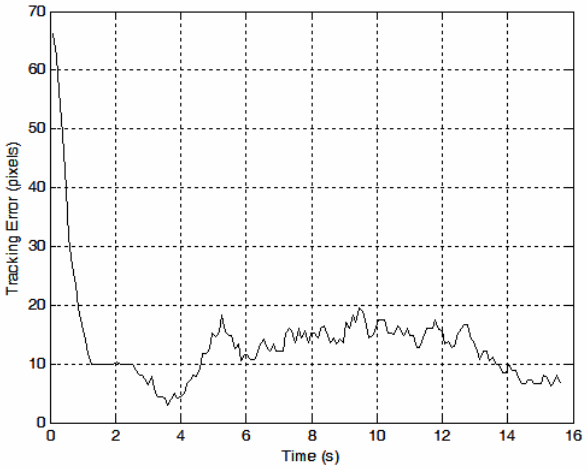


Figure 3. Experimental results of the visual tracking in terms of the norm of the ball offset using the image Jacobian approach: (*a*) the camera aligns with the ball (up); and (*b*) the ball is moving along a straight line with an average velocity of 0.2 m/s (bottom).

5. Conclusions and Future Work

Many robotics applications depend on the use of a camera-based vision system that can provide sense of perception of the environment. In this paper, the emphasis has been given to the design of a vision-based control scheme that would enable autonomous operation of a self-contained humanoid robot. The inclusion of a PC104 Linux based controller proved to be a good compromise among computational power, expandability, power consumption and size.

On-board full processing and autonomy is now possible with the central unit that interfaces the CAN-bus already operational on the platform. Another key technological concept is the idea of distributed functionality. The proposed architecture separates the high-level vision processing activities from the low level closed-loop control of the actuators. As consequence, it became possible to develop two control laws that ensure global tracking control. Each of the controllers “close-the-loop” in the image space and it is proven to be globally valid. The results are relevant for this class of humanoid platforms (*i.e.*, low-cost and easy-to-design).

A number of open questions remain. One of them concerns the possible role of the force-driven local controller described in [7] to compensate the trunk motion, which was not considered in the present paper. Also, we did not touch upon the interesting problem of on-line calibration under closed-loop control. Other open issue concerns the influence of the robot motion on the visual

information and the tracking system's performance. For instance, a question to address is whether and to what extend the vision system distinguishes between target motion and self motion. Hence, future work will address several research directions to extend the proposed methods.

References

1. Y. Sakagami *et al.*, "The Intelligent ASIMO: System Overview and Integration", in Proceedings of the IEEE International Conference on Intelligent Robots and Systems, pp. 2478-2483 (2002).
2. K. Kaneko *et al.*, "Humanoid Robot HRP-2", in Proc. IEEE International Conference on Robotics and Automation, pp. 1083-1090 (2004).
3. L. Hu and C. Zhou, "Locomotion planning of humanoid robot Robo-Erectus Senior (RESr-1)", in Proceedings of the IEEE-RAS International Conference on Humanoid Robots, Pittsburgh, USA, Nov 29-Dec 01 (2007).
4. T. Furuta *et al.*, "Design and Construction of a Series of Compact Humanoid Robots and Development of Biped Walk Control Strategies", *Robotics and Automation Systems*, **37**, pp. 81-100 (2001).
5. J.-H. Kim *et al.*, "Humanoid Robot HanSaRam: Recent Progress and Developments", *Journal of Computational Intelligence*, **8**(1): 45-55 (2004).
6. F. M. Silva, V. M. Santos, "Multipurpose Small-Cost Humanoid Platform and Modular Control Software Development", book chapter in the publication *Humanoid Robots: Human-like Machines*, [ISBN: 978-3-902613-07-3], edited by Matthias Hackel, pp. 65-88 (2007).
7. M. Ruas, P. Ferreira, F. M. Silva and V. M. Santos, "Local-level Control of a Humanoid Robot Prototype with Force-driven Balance", in Proceedings of the IEEE/RSJ International Conference on Humanoid Robots, Nov. 29 - Dec. 3, Pittsburgh, USA (2007).
8. S. Behnke, J. Stückler, H. Strasdat and M. Schreiber, "Hierarchical Reactive Control for Soccer Playing Humanoid Robots", book chapter in the publication *Humanoid Robots: Human-like Machines*, [ISBN: 978-3-902613-07-3], edited by Matthias Hackel, pp. 625-642 (2007).
9. Y.-T. Su *et al.*, "Omni-Directional Vision-Based Control Strategy for Humanoid Soccer Robot", in Proceedings of the 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON), pp. 2950-2955 (2007).
10. E. Menegatti, A. Pretto, A. Scarpa and E. Pagello, "Omnidirectional Vision Scan Matching for Robot Localization in Dynamic Environments", *IEEE Transactions on Robotics*, [ISSN: 1552-3098], **22**(3):523-535 (2006).
11. S. Hutchinson, G. Hager and P.I. Corke, "A Tutorial on Visual Servo Control", *IEEE Trans. on Robotics and Automation*, **12**(5):651-670 (1996).
12. A. C. Sanderson, L. E. Weiss and C. P. Neuman, "Dynamic Sensor-Based Control of Robots with Visual Feedback", *IEEE Transactions on Robotics and Automation*, **3**:404-417 (1987).