Sensory-motor Anticipation and Local Information Fusion for Reliable Humanoid Approach

Hendry Ferreira Chame, Christine Chevallereau

Sensory-motor Anticipation and Local Information Fusion for Reliable Humanoid Approach

H. F. Chame[1] and C. Chevallereau[2]

IRCCyN, Ecole Centrale de Nantes, CNRS, Nantes, FRANCE,
[1]*e-mail: Hendry.Ferreira-Chame@irccyn.ec-nantes.fr*
[2]*e-mail: Christine.Chevallereau@irccyn.ec-nantes.fr*

Abstract. The possibility of developing increasingly sophisticated robots, and the availability of cloud-connected resources, have boosted the interest in the study of real world applications of service robotics. However, in order to operate under natural or less structured conditions, and given the information processing bottleneck and the reactivity required for a secure execution of the task, it is desirable that the agent can exploit more efficiently the local information available, so that being more autonomous, and relying less on remote computation. This study explores a strategy for obtaining reliable approach tasks. It considers the anticipation of perception, by taking into account the statistical regularities and the information redundancies induced in the sensory-motor coupling. From an initial perception of the object assisted by remote computation, contextual features are defined for capturing bodily sensations emerging in the task. The observations based on proprioceptive and visual data are fused in a Bayesian Network, which is in charge of assessing the saliency during the object approach, thus constituing a local discriminative processing of the object. The strategy proposed reduces dependency on context-free models of behavior, while providing an estimate on the degree of confidence in the progress of the task.

Key words: Cognitive robotics, Embodied cognition, Humanoid robotics, Ego-localization, Top-down visual attention, Robot Vision.

1 Introduction

Ubiquitous computing is now a reality, given the progresses in the fields of information technology and artificial intelligence. In everyday life we have access to various applications offered on mobile devices. Just to mention a few, we can obtain information about a product from a captured image, or identify a song from an audio sample. In such applications, given the computational limitations of the local device, the processing is performed remotely on dedicated servers, equipped with extensive knowledge data-bases and a vast computing power. The local device is simply in charge of running the client application, thus, ensuring the aspects related to the usability in the human-machine interaction.

From the success of ubiquitous computing, several efforts are aiming at developing solutions to more challenging scenarios in heath-care, assistance, or service

robotics. Hence, robot applications can share knowledge or be assisted by cloud-connected resources. In this sense, there are currently initiatives that focus on the definition of robot architectures (e.g. Vasiliu et al[20]), that integrate distributed resources to the task, so mitigating specific constraints of the robot platform. In parallel, several research communities are engaged in the definition of ontologies for knowledge sharing, distributed learning, and the collection and reuse of information for practical applications (e.g. Waibel et al[21]).

However, since robots take action on the environment, the safety aspect is particularly affected by the information bottleneck during remote computing. When moving, the agent would ideally not depend at all on remote resources, since the communication with the server can be interrupted. It would also have the least possible of top-down deliberation, since it must be reactive to unexpected disturbances. Thus, it has to be endorsed with autonomy in order to ensure the maximum local progress on the task.

In this study, inspired by the research on embodied cognition, we explore the emergency of information during the task. From a first-person perspective definition of the approach, and taking into account the redundancies and the statistical regularities induced in the sensory-motor coupling, we examine the possibility of exploiting the anticipation of multi-sensory, contextual, and more diversified evidence about the object. To this end, we study a Bayesian network structure for information fusion, to discriminate the saliency related to the object.

This paper is organized as follows. In Sec. 2 some related contributions from the cognitivist approach to artificial intelligence are discussed, and contrasted to the point of view of embodied cognition research, that considers the emergent aspects of behavior. In Sec. 3 the definition of the approach task is presented. In Sec. 4 the aspect of autonomy is tackled, where the design of the features and the structure of the Bayesian network are detailed. A case-study has been developed and conducted with the robot Nao, which is going to be discussed in Sec. 5. Finally, the conclusions are given in Sec. 6.

2 Related work

Vision-based locomotion control is a challenging task for walking robots. Unlike natural beings, which are in possession of extremely sophisticated sensory organs, the vast majority of the research in robotic vision has been carried out with quite inferior equipment, usually employing general purpose cameras. Moreover, the body structure and the actuation system utilized is much less stable, fine, and accurate, when compared to the natural musculo-skeletal system. In view of such limitations, some studies (e.g., Lewis & Simo[12], and Michel et al.[14]) have resorted to capture information from extra-corporal sensory, for achieving higher quality observations. Unfortunately, robot motion may occlude the cameras thus compromising the solution. Furthermore, the generality of the solution is affected once the scene is adapted to the task.

On-board solutions have been attempted under the visual servoing (VS) framework (see Chaumette & Hutchinson[4] and Corke[6]), which relates variations on image features to the spatial instantaneous velocity of a flying (dismembered) camera. A study by Dune et al.[7] has considered a monocular vision task with the robot HRP2. Given the walk style of the robot, the solution involved the cancellation of the oscillatory contribution to the control signal (also called the *sway motion*). A work by Moughlbay et al.[15] has considered approaching tasks with the robot Nao. In both studies, in order to handle the image noise, the tracking technique by Comport et al.[5] was employed. The algorithm is based on visual odometry and required of a realistic 3D model of part of the room.

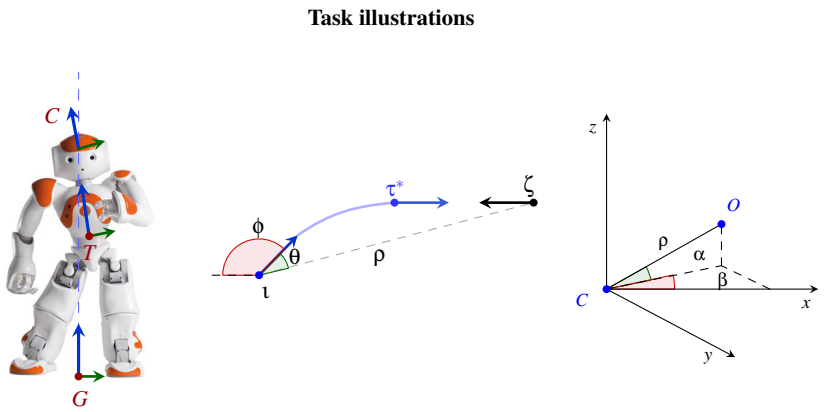
Model-based navigation have been explored in the simultaneous localization and mapping (SLAM) research (see Thrun et al.[19]). Examples of contributions in the field are numerous. Just to mention a few, in the work by Hornung et al.[10], starting from a volumetric map of the environment, precise indoor localization is obtained by adapting a range sensor to the robot's head. A work by Oriolo et al.[16] considered building the map on-line by fusing proprioceptive, inertial, and visual information, within an extended Kalman filter. In general, map-based navigation has produced impressive results, but it has also received some criticism. According to Shapiro[18], EC researchers disagree with the premise that organisms must firstly represent the environment for then navigating its topology. Indeed, this would be inconvenient to unstructured or reactive situations. Moreover, from the practical point of view, map-based solutions present as a drawback requiring maintenance, where environmental changes must be systematically acknowledged.

The works discussed so far fall within the so-called cognitivist research paradigm of artificial intelligence, that addresses the problem of automation under a representational focus, in which, the solution of the task is modeled explicitly. In the last decades a different perspective has been adopted through the multi-disciplinary research in embodied cognition (EC). Under the EC methodology, behavior is viewed as a complex system, emerging from the interactions with the environment (Anderson[1], Hoffmann & Pfeifer[9]). Thus, knowledge representation is thought to be grounded in the physical coupling (Brooks[2]), so emergent behavior would be neither explicitly described nor planned in advance.

The analysis of the sensory-motor coupling in natural tasks, from a dynamic system perspective, appears as a promising research direction, that can provide more efficient, easier to implement, and robust solutions. In this sense, a work by Lungarella & Sporns[13] has explored the relation between sensory-motor coordination, body morphology, and information processing. The study reported that higher levels of information correlation occurred when actions and perceptions were coordinated. Moreover, sensory-motor coordination reduced the dimensionality of the information content, given the perceptual regularities induced in the task. From these results, the information emerging in the task, including the motor activity, the body configuration, and the visual saliency, can be exploited to anticipate the evolution of the object in the field of vision. Thus, reducing the dependency on intensive context-free computations for perceiving the object in the scene.

3 Task definition

The on-board localization is obtained from the definition of a sensory ego-cylinder, that heuristically takes the z-axis perpendicular to the ground, under the assumption of motion over a plane surface. Figure 1a illustrates the frames defined in the study. The frame G was placed at the ground level between the feet of the robot, it corresponds to the origin of the ego-cylinder. The frame T was set at the center of mass of the robot. The frame C is placed at the camera location on the forehead. The transformation between the reference frames G, T, and C, depend on the current joint configuration q of the robot. They are obtained by the classical modeling approach (see Khalil & Dombre[11]), that considers the body as a set of interconnected serial structures departing from a common reference, which is taken as frame T.



(a) (b) (c)

Fig. 1: (a) Frontal view of the task reference frames (camera C, trunk T, ground G). The z-axis is represented in blue and the y-axis in green. The x-axis goes towards the reader and is plotted as a red dot. Notice that the frames are not necessarily aligned. (b) Upper view of the task. The localization of the object is denoted by ζ. The black dot represents the object's center, the black arrow illustrates the direction of the projection on the motion plane of the mean normal direction to the tracked face of the object. The agent with the heading direction is represented in blue. A desired configuration in relation to the object is represented by τ^*. The trajectory followed to approach the object is illustrated in light blue. (c) Illustration of the look-at task. The position of the object center corresponds to O. After the head correction, the x-axis of the camera frame C will be aligned with the direction $\overline{CO}$.

As illustrated on Fig. 1b, the localization ζ of the object is represented by the four parameters

$$\zeta = \left[\, \rho \;\; \theta \;\; \iota \;\; \phi \,\right]^{\mathrm{t}}, \tag{1}$$

where ρ, θ, and ι are position components. Respectively, the distance, the azimuth, and the height of the center of the object. The parameter ϕ corresponds to the orientation of the object around the z-axis. It is estimated by the difference between the projection on the motion plane of the mean normal direction to the tracked face of the object, and the heading direction of the agent. The observation of the object's pose with respect to the camera frame C is obtained by defining a virtual oriented trapezoid, which delimits the image region containing the tracked face (see Fig. 2).

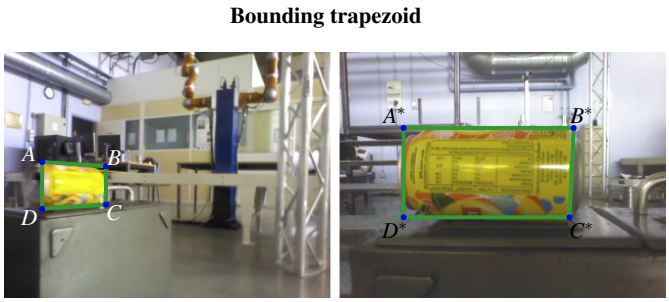


Fig. 2: The localization of the can is estimated by fitting the salient trapezoid to a rough geometric model of the object, which in this case is a cylinder. Notice that the lateral face of the can at the left is not salient. The observation of the approach error $\hat{e}$ (see Eq. (2)) is obtained from the spatial relation between the current view at the left and the desired view at the right.

The task considered is the approach to a specific face of a static object, by walking on a plane, in a scene without obstacles. More specifically, starting from the knowledge of a desired ego-centric perception of the object, the agent has to autonomously return as close as possible to such state once disturbed. The behavior can be viewed as a regulation task where the control parameters include a 2D pose, which is defined with respect to a movable reference frame (on-board). Hence, the localization component ι is assumed to be constant. Formally, the approach error e expresses the desired configuration τ^* of the body (see Fig. 1b) in the actual ego-centric perspective, such that

$$e = \left[\tau_{\rho^*} \;\; \tau_{\theta^*} \;\; \tau_{\phi^*} \right]^{\mathrm{t}}. \tag{2}$$

The solution of the task is based on the parallel execution of two motor behaviors: the walk and the look-at tasks. The former is in charge of steering the robot to ensure convergence toward the object. In order to take into account the aesthetics of motion, human walk is mimicked. That is, non-holonomic motion is used when human is far from the object, but holonomic motion is preferred when human is close enough to the goal. Thus, the Cartesian evolution of the walk task is described by

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \omega \end{bmatrix} = \begin{bmatrix} \lambda(k_1\hat{e}_\rho) + \gamma(k_2\cos(\hat{e}_\theta)\hat{e}_\rho) \\ \gamma(k_3\sin(\hat{e}_\theta)\hat{e}_\rho) \\ \lambda(k_4\sin(\hat{e}_\theta - \hat{e}_\phi) + \hat{e}_\phi) + \gamma k_5\hat{e}_\phi \end{bmatrix}. \tag{3}$$

where $\dot{x}$ and $\dot{y}$ are the linear velocities, and ω is the angular velocity. The holonomic walk includes independent corrections along the 3 degrees of freedom, with different gains k_2, k_3, k_5. The non-holonomic walk does not include correction along the y-axis, since no lateral displacement is permitted. A correction along the walking direction is applied proportionally to the error distance $\hat{e_\rho}$ with gain k_1. The steering correction is chosen to direct the agent toward the object with gain k_4. The transition to the holonomic motion depends on the distance $\hat{e}_\rho$, such that $\lambda = 1/(1+\exp(s_1 * (\hat{e}_\rho - s_2)))$, where s_1 is a proportional gain, and s_2 is the sensitive distance for the transition.

The look-at task is in charge of directing the view towards the object, thus maintaining it centered on the image. As illustrated in Fig. 1c, α and β are respectively the pitch and yaw angles of the Nao robot neck, that affect the pose of the camera frame C. The head motion is described by

$$\begin{bmatrix} \dot{\alpha} \\ \dot{\beta} \end{bmatrix} = \begin{bmatrix} \text{atan2}(l_\iota, \cos(l_\theta)\hat{\zeta}_\rho) \\ l_\theta \end{bmatrix}, \tag{4}$$

where l_θ and l_ι are the azimuth and the height correction desired, and $\hat{\zeta}_\rho$ is the observed distance to the object, expressed with respect to frame T.

4 Behavior autonomy

As discussed earlier, the tasks in charge of the agent are moving toward the object and controlling the gaze direction. The available sources of information include the image acquired by the camera and the proprioceptive registry of the robot. The difficulty of the task consists in finding the object of interest on the image, by relying mostly on contextual representations. As illustrated in the diagram of Fig. 3, the idea is to provide the agent with an autonomous implementation of the behavior that can be continuously assessed. Thus, the agent would resort to remote services only when the level of confidence about the correctness of the task is low. For this, we will rely on the coupling between action and perception. The perception is based on the analysis of properties related to the image blobs (e.g. the area, the radio-aspect, and the topology), which results from the color saliency. It is also based on the comparison of body postures, and the spatial relation between the agent and the object. The actions concern the motion control and a prediction of the localization, from the motion undertaken.

Table 1 details the definition of contextual features for assisting the object perception. The first two measurements are obtained from a particularly useful class of image features, which is known as moment (Corke[6]). For a binary image $B[x,y]$ the $(p+q)^{\text{th}}$ order moment is defined by

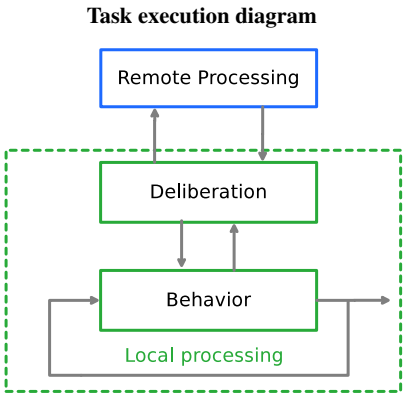


Fig. 3: The deliberative process evaluates the consistency of the autonomous local execution of the behavior. Once inconsistency is detected, remote processes are queried for support.

$$m_{pq} = \sum_{y=0}^{y_{\max}} \sum_{x=0}^{x_{\max}} x^p y^q B(x,y) \quad . \tag{5}$$

The radio-aspect F_3 is defined from the width and height of the minimum bounding-box (MBB) enclosing the blob, where the angle between the MBB's principal axis and the image x-axis is $\gamma = 0.5(\text{atan}(2m_{11}/(m_{20} - m_{02})))$. The feature F_4 includes proprioceptive information from the instantaneous posture of the neck. The feature F_5 represents the topographic relation between the blobs, it is a descriptor of the presence of saliency at a four cardinal neighborhood.

Table 1: Saliency features

Expression	Description
$F_1 = (\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}})$	Centroid.
$F_2 = m_{00}$	Area.
$F_3 = B_{height}/B_{width}$	Radioaspect, where B denotes the oriented bounding box.
$F_4 = (\alpha, \beta)$	Posture, with α and β the pitch and yaw neck angles.
$F_5 = v(S,s)$	Topology, where v attributes a 4-bit vicinity code according to the saliency set S around the blob s.

The anticipation process involves a deterministic motion model, that assumes an ideal noise-free robot, moving at constant velocity $v = [\dot{x}, \dot{y}, \omega]^t$, along the time interval Δt. Thus, a prediction for the localization of the object $\tilde{\zeta}$ is obtained from the last observation available $\hat{\zeta}$, and the expected displacement $m = v\Delta t$. Table 2 presents the definition of a set of features that relate the actual saliency to the antic-

ipated information flow. Figure 4 illustrates the feature $\overline{F_1}$. In a periphery-to-center flow, information from salient regions (e.g, the blobs centroids) are projected to the sensory ego-space. Notice that among the features defined, some are not directly related to the motion of the agent, such is the case of $\overline{F}_3$ and $\overline{F}_5$.

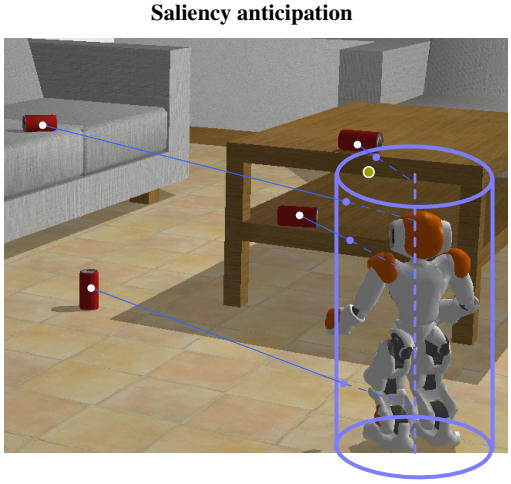


Fig. 4: The agent is approaching the red can on the top of the table. The white dots correspond to the center of the salient objects. The estimate on the distance to the blob center is unavailable during the saliency analysis, thus, the last observation $\hat{\zeta}_\rho$ is employed. The projection of the blobs in the ego-cylinder is represented by the blue dots, whereas the predicted localization is represented by the yellow dot.

Table 2: Anticipation features.

Expression	Description
$\overline{F_1} = \lvert\sigma(F_1' - \tilde{\zeta})\rvert$	F_1' denotes the projection of the blob centroid in the ego-space, $\tilde{\zeta}$ is the predicted localization of the object, and σ weights the contribution of each component.
$\overline{F_2} = 1 - \frac{F_2}{\left(\frac{\hat{\zeta}_\rho + m_\rho}{\hat{\zeta}_\rho}\right) F_{2(k-1)}}$	Relation between the actual blob's area F_2 and the simulated area from the expected motion m. Here $F_{2(k-1)}$ denotes the saliency during the last observation $\hat{\zeta}$.
$\overline{F_3} = \lvert F_{3(k)} - F_{3(k-1)}\rvert$	Difference between the current and the last detected radio-aspect.
$\overline{F_4} = \lvert\check{F}_4 - \tilde{F}_4\rvert$	Difference between the simulated posture of the neck $\check{F}_4$, that would center the blob on the visual field, and the predicted attitude of the neck $\tilde{F}_4$.
$\overline{F_5} = \sum_{i \in N} \delta(F_{5(k-1)i}, F_{5i})$	Estimate of the topographic relation through the Kronecker delta function $\delta(a,b)$. The neighborhood set is defined by $N = \{left, right, up, down\}$.

The information fusion for the discriminative process is accomplished in a Bayesian Network (BN), which is a directed acyclic graph, that represents the conditional probabilities of interconnected random variables. The BNs have been employed for diverse automatic diagnosing and recognition tasks (see Ertel[8]). In a BN, a node is assumed to be conditionally independent from non-successors, given its parents. The joint probability $p(X_1,...,X_n)$ of the nodes X_i is expressed by

$$p(X_1, \ldots, X_n) = \prod_{j=1}^{n} p(X_j | parents(X_j)). \tag{6}$$

One important advantage of knowledge representation through BNs is that the information contained is directly understandable by humans, which facilitates doing future modifications (e.g. including new features, or more complex observations). As illustrated in Fig. 5, the structure of the network corresponded to a tree of height 2. The root node is a binomial random variable, which represents the probability that the blob saliency matches up with the object of interest. The intermediate nodes B_i are binomial random variables that represent the a posteriori probability of the features, given the observation of the object. This layer is included in order to simplify adjustments to the contribution of the features to the discriminative process. Probabilistic independence is assumed between the nodes B_i, which is also known as a *naive Bayes classifier*. The leaves O_i are multinomial random variables that represent the a posteriori probability of observing a particular intensity of $\overline{F_i}$, given B_i. The tree can easily accommodate new features by horizontal expansion. The most likely blob s among the saliency set S is obtained by maximizing the expression

$$s = \text{argmax}_{s \in S} p(Object | \, B_i, O_i). \tag{7}$$

Thus, the BN can be used to classify the salience, while providing an estimate of the certainty in such classification. As we shall see in the case study, this information is of crucial importance to the deliberative process, once it has to decide whether or not resorting to remote processing.

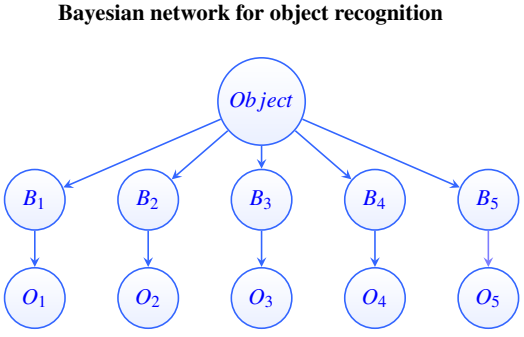


Fig. 5: Bayesian network for contextual information fusion.

5 Case Study

The case study included simulations of the task in Webots version 7.4.0, and a real approach task with the robot Nao. The robot is 58 cm tall, weighs 4.8 kg, has 25 degrees of freedom, and is equipped with a CMOS digital camera with a 58 degree field of view. The algorithms were implemented in the C++ programming language, and run under Ubuntu version 12.04.5 LTS. The vision processing was obtained with the support of the OpenCV library version 2.4.8. The Bayesian network implementation was provided by the dlib C++ Library version 18.13.

In the scenario envisaged, the agent is expected to be assisted by remote resources. Thus, the first detection of the object, including the initial segmentation, and the virtual oriented trapezoid as seen from the desired configuration, are assumed to be provided remotely. This requirement was simulated in the study. The initial segmentation was manually provided through the GrabCut technique (Rother et al. [17]), and the desired pose was shown by demonstration (i.e. by placing the robot in front of the object). A color-based segmentation technique, defined under the Markof random field formalism, has been employed from our prior work (see Chame & Chevallereau[3]), for detecting the color saliency. The resulting algorithm presented a computational complexity $O(kn^2)$, where k is the maximum number of iterations allowed in the optimization process, and n is the number of pixels.

In the walk task the agent had to move to the desired location, and to stop once all the components of the observed localization error $\hat{e}$ (see Eq. (2)) were smaller than a given threshold ε. The tolerance considered was a radial distance $\varepsilon_\rho = 0.05$ meters (m), the azimuth $\varepsilon_\theta = 0.04$ radians (rad), and $\varepsilon_\phi = 0.1$ rad for the orientation component. The walk primitive of the robot receives commands in position, expressed in the Cartesian space. A motion request is sent by the walk task, according to Eq. (3), under the assumption of constant velocity motion. The mean walk velocity was estimated to be around $\bar{v} = [0.022\ \mathrm{m/s}, 0.04\ \mathrm{m/s}, 0.106\ \mathrm{rad/s}]^\mathrm{t}$. Continuous motion was achieved by sending commands at regular time intervals. In order to prevent that unforeseen delays affect the fluidity of the walk, the actual displacement sent considered a larger delay (e.g. 1.5 times the expected value). Thus, a new command would be ideally sent before the routine could finish the previous one. If this would not be the case (e.g. due to losing the object, a program crash, etc.), the robot would stop moving after a while. This strategy ensured a fluid walk while keeping the safety aspects. For speeding up convergence, given the observation noise and the fact that the walk primitive is less precise in continuous motion, once the robot was nearly at the desired location (at $\hat{e}_\rho < 0.1$ m), the walk task switched to a step-by-step policy (i.e. a new correction was sent only after finishing the previous one).

The look-at task was also controlled in position. The correction of the head posture, through the actuation of the neck joints α and β, was obtained from Eq. (4) by assuming constant velocity motion. The tolerance $\varepsilon = 0.03$ rad was admitted for convergence. The head posture is regulated independently from the walk. This means that the motions induced by the walk can affect the convergence of the look-at task, specially at slow turning of the head. Thus, a velocity profile of 4 rad/s was employed and convergence was obtained in few iterations.

The study of the anticipation process was based on the design of a controlled scene that contained a single salient object. The observation of the evidence was accomplished through the definition of data partitions, which were obtained from the statistical analysis of the information flow, as recorded by the agent. The Nearest Neighboor Method (see Ertel[8]) was employed to classify the observation $O_i \in \{0,1,2\}$, with 0 the lowest and 2 the highest intensity. A more complex task was then designed to evaluate the network performance. As illustrated in Fig 6, various stimuli of the same kind were placed on the scene, where the agent had to approach a specific soda can. Despite stimuli were relatively close to each other, with the BN, the agent was able to do the task autonomously (i.e. without remote assistance, as provided in Fig. 3) from different initial configurations, under different feedback delay profiles (between 100 to 2000 ms). Figure 7 shows a comparison of the performance of the network for two delays profiles. As noted, although the task was accomplished in both cases, at higher delay the network is less discriminative, since the anticipation is less consistent.

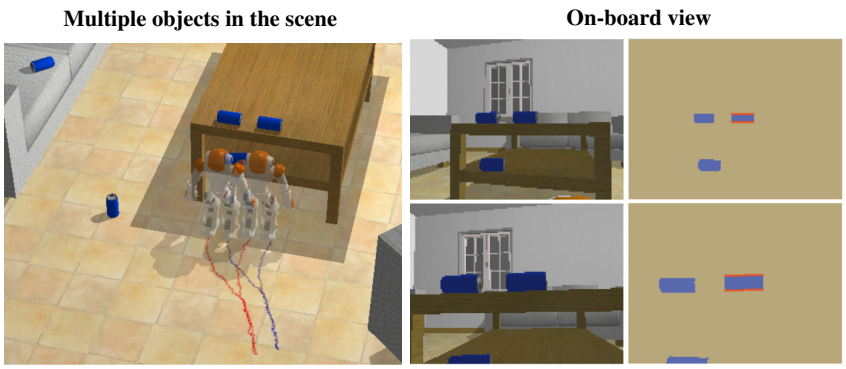


Fig. 6: On the left two trajectories followed are superimposed. Starting from the same position, the agent was required to approach a distinct can over the table. Some on-board views and their corresponding saliency are displayed on the right.

Based on the simulation results, we proceeded to do the experiment in the robotic lab of the IRCCyN, which is an unstructured environment, under natural and artificial illumination. As shown on Fig. 8, two colored tea cans were placed one beside the other at a distance around 4 cm, and the robot was required to approach one of them. Given the characteristics of the robot, the perceptive delay was set to 1700 ms, including at most two iterations for centering the object in the field of vision. The experiment was executed in two motion modalities. In the first one, a step-by-step approach was attempted, where the robot finished the motion before processing additional commands. In the second one, a continuous approach was adopted. The initial trials were conducted off-line, that is, the local deliberative module was disconnected (see Fig. 3), in order to evaluate the autonomous execution of the be-

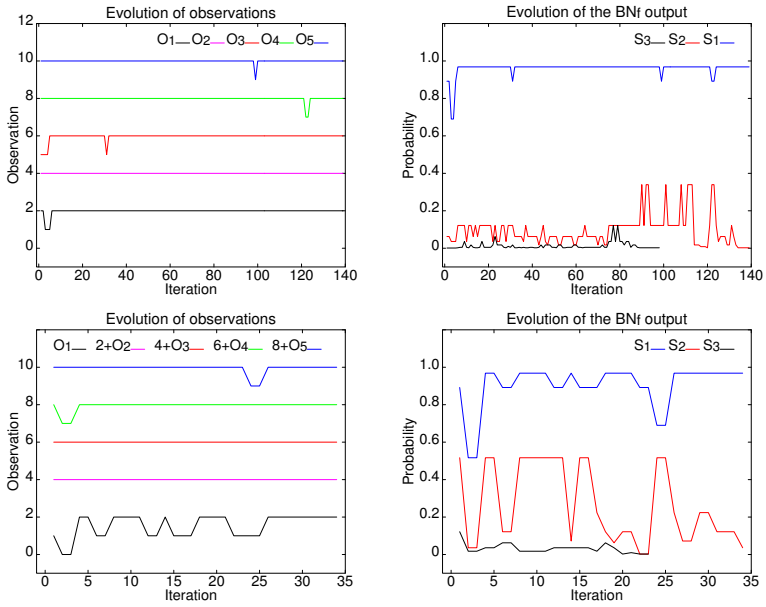


Fig. 7: Comparison of the network response with two delay profiles. The upper row presents the results for a 300 ms delay the bottom the delay at the bottom row corresponds was 1700. The first column shows the evolution of the observations O_i (with 0 the lowest and 2 the highest intensity). The signal are shifted vertically for visualization, such that $O_i = 2(i-1) + O_i$. The second column shows the output of the network. The tracked can over the table is represented by s_1, and s_2 corresponds to the lateral neighbor (see Fig. 6). As it can be noted, the discriminative power of the network is reduced as the delay increases.

havior. Thus, the network was allowed to select the most likely blob conforming to Eq. (7), regardless of the probability obtained. Under these conditions, the task was accomplished 7/10 times with step-by-step motion, and 5/10 times with continuous motion. Among the reasons that affected the off-line execution of the task are: unexpected peaks on the feedback delay (e.g. the expected delay profile was occasionally exceeded in more than 2000 ms), irregular performance of the walk primitive, and momentary degradation on the color saliency detection.

Interestingly, the performance of the network in the off-line trials was very consistent, by yielding a chance $p_o < 0.5$ to blobs selected under degraded conditions. Therefore, it provided a reliable information about the degree of confidence on the task. Though, as shown in Fig. 7, at high delay profiles the discriminative power of the BN drops. Thus, the difference p_d between the two most likely candidates may be employed as an indicator of the actual discriminative power of the network. In a new set of trials on-line execution was attempted by activating the deliberative process. A tolerance $p_d > 0.2$ and $p_o > 0.6$ for the task reliability was set, so the

deliberative process paused the behavior when the threshold was reached and waited for the user, who acting as the remote resource, selected the blob corresponding to the object. In the on-line execution of the experiment the task was accomplished all the times.

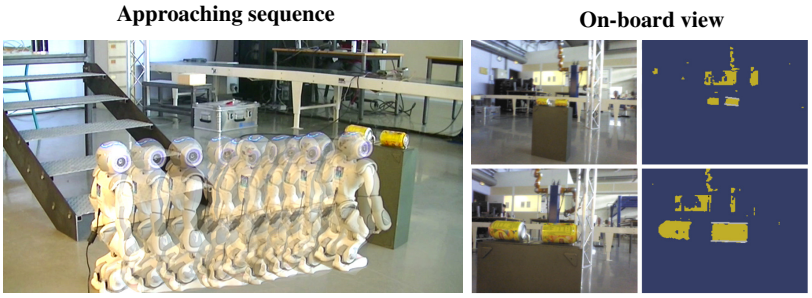


Fig. 8: Experimental task. The trajectory followed is shown on the left, whereas some on-board views and their corresponding color salience are shown on the right.

6 Conclusions

This study has focused on the problem of the autonomy of humanoid approach tasks. It has explored an efficient means of employing distributed resources, towards developing more realistic and robust applications for service robotics. Given the risks in the on-line implementation of the task, we concentrated our efforts in exploiting the emergence of contextual information, notably, the redundancies and the statistical regularities induced in the sensory-motor coupling. We examined the possibility of employing the anticipation of the information flow to assist a local, multi-sensory, perceptive process, thus, reducing the dependency on context-free representations. In view of the stochastic difficulties inherent to the task, in the form of sensory noise, processing delays, and disturbances from the environment, we employed a Bayesian network structure to fuse information. The results obtained suggested that it is a convenient and easy-to-employ technique, which produces reliable information about the degree of confidence on the task. Such information can be evaluated by a local deliberative process, that can resort to remote assistance in case the agent is no longer able to autonomously accomplish the task.

Acknowledgements This research has been funded by the Ecole Centrale de Nantes (ECN) and EQUIPEX ROBOTEX, France; and the CAPES Foundation, Ministry of Education of Brazil, Brasília - DF 700040-020, Brazil.

References

[1] Anderson, M.: Embodied cognition: A field guide. Artificial Intelligence **149**(1), 91–130 (2003). DOI 10.1016/S0004-3702(03)00054-7

[2] Brooks, R.A.: Cambrian Intelligence: The Early History of the New AI, 1 edition edn. A Bradford Book, Cambridge, Mass. (1999)

[3] Chame, H.F., Chevallereau, C.: Embodied localization in visually-guided walk of humanoid robots. In: ICINCO (2), pp. 165–174 (2014)

[4] Chaumette, F., Hutchinson, S.: Visual servo control, part i: Basic approaches. IEEE Robotics and Automation Magazine **13**, 82–90 (2006)

[5] Comport, A., Marchand, E., Pressigout, M., Chaumette, F.: Real-time markerless tracking for augmented reality: the virtual visual servoing framework. Visualization and Computer Graphics, IEEE Transactions on **12**(4), 615–628 (2006). DOI 10.1109/TVCG.2006.78

[6] Corke, P.I.: Robotics, Vision & Control: Fundamental Algorithms in Matlab. Springer (2011)

[7] Dune, C., Herdt, A., March, E., Stasse, O., Wieber, P.b., Yoshida, E.: Vision based control for humanoid robots. In: in "IROS Workshop on Visual Control of Mobile Robots (ViCoMoR (2011)

[8] Ertel, W.: Introduction To Artificial Intelligence. Springer London Ltd (2011)

[9] Hoffmann, M., Pfeifer, R.: The implications of embodiment for behavior and cognition: animal and robotic case studies. CoRR **abs/1202.0440** (2012)

[10] Hornung, A., Wurm, K., Bennewitz, M.: Humanoid robot localization in complex indoor environments. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, pp. 1690–1695 (2010). DOI 10.1109/IROS.2010.5649751

[11] Khalil, W., Dombre, E.: Modeling, Identification and Control of Robots, 3rd edn. Taylor & Francis, Inc., Bristol, PA, USA (2002)

[12] Lewis M.A., Simo L.S.: Elegant stepping: A model of visually triggered gait adaptation. Connection Science **11**(3-4), 331–344 (1999). DOI 10.1080/095400999116287

[13] Lungarella, M., Sporns, O.: Information self-structuring: Key principle for learning and development. In: Development and Learning, 2005. Proceedings., The 4th International Conference on, pp. 25–30 (2005). DOI 10.1109/DEVLRN.2005.1490938

[14] Michel, P., Chestnutt, J., Kuffner, J., Kanade, T.: Vision-guided humanoid footstep planning for dynamic environments. In: Proceedings of the IEEE-RAS Conference on Humanoid Robots (Humanoids'05), pp. 13 – 18 (2005)

[15] Moughlbay, A., Cervera, E., Martinet, P.: Model based visual servoing tasks with an autonomous humanoid robot. In: S. Lee, K.J. Yoon, J. Lee (eds.) Frontiers of Intelligent Autonomous Systems, *Studies in Computational Intelligence*, vol. 466, pp. 149–162. Springer Berlin Heidelberg (2013). DOI 10.1007/978-3-642-35485-4_12

[16] Oriolo, G., Paolillo, A., Rosa, L., Vendittelli, M.: Vision-based odometric localization for humanoids using a kinematic ekf. In: Humanoid Robots (Humanoids), 2012 12th IEEE-RAS International Conference on, pp. 153–158 (2012). DOI 10.1109/HUMANOIDS.2012.6651513

[17] Rother, C., Kolmogorov, V., Blake, A.: GrabCut -Interactive Foreground Extraction using Iterated Graph Cuts. ACM Transactions on Graphics (SIGGRAPH) (2004)

[18] Shapiro, L.: The embodied cognition research programme. Philosophy Compass **2**(2), 338–346 (2007). DOI 10.1111/j.1747-9991.2007.00064.x

[19] Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. The MIT Press, Cambridge, Mass. (2005)

[20] Vasiliu, L., Trochidis, I., Bussler, C., Koumpis, A.: Robobrain: A software architecture mapping the human brain. In: Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on, pp. 160–165 (2014). DOI 10.1109/HUMANOIDS.2014.7041353

[21] Waibel, M., Beetz, M., Civera, J., D'Andrea, R., Elfring, J., Galvez-Lopez, D., Haussermann, K., Janssen, R., Montiel, J., Perzylo, A., Schiessle, B., Tenorth, M., Zweigle, O., van de Molengraft, R.: Roboearth. Robotics Automation Magazine, IEEE **18**(2), 69–82 (2011). DOI 10.1109/MRA.2011.941632