Towards Assistive Robotics Through Embodiment and Computational Models: On-board Visually Approaching Objects in the Scene

Hendry Ferreira Chame[1]   Christine Chevallereau[1]

[1]Ecole Centrale de Nantes, IRCCyN, CNRS
Robotics Team
{Hendry.Ferreira-Chame, Christine.Chevallereau}@irccyn.ec-nantes.fr

June 24, 2014

Outline

Our research interest

- Exploring tasks at human environments.
- Assistive robotics.
- Certain degree of complexity and abstraction.
- Feasible solutions.
- Exploration of the concept of embodiment.

Outline

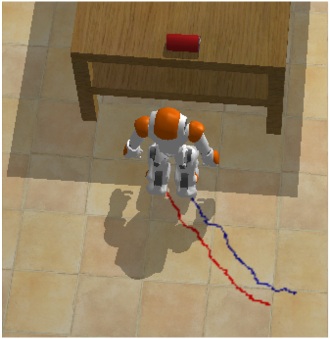
Embodiment

- Real-world thinking occurs in particular situations with specific practical ends.
- Sensory and motor functions are relevant aspects of intelligent behavior.
- Cognition is a distributed process.

What is the task to be solved?

To approach objects in the scene using vision:

- Converging to a desired 2D pose by following planar motion.



- Three degrees of freedom task.

What are the available resources?

Table: Resources available to solve the task.

Type	Resource
Brain (models and algorithms)	Object representation
	Top-down feature attention
	Ego-centered stimuli reference
Body	Proprioception
	Vision
	Motion primitives
Environment	Planar surface for motion
	Static objects
	Human supervisor

How are does the agent solve the task?

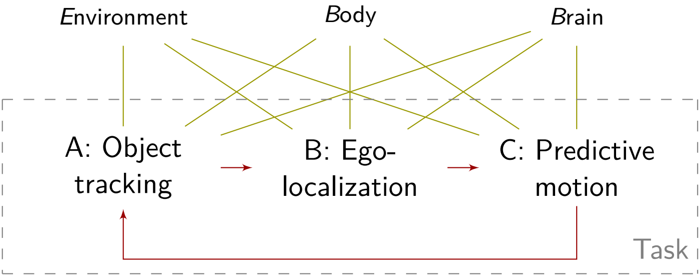


Figure: Relationship between available resources and the task.

Outline

Challenging

Certain difficulties are reported when using on-board visual systems:

- Limited control over the head's direction [1].
- Sway motion for continuous visual servoing [2].
- Delays in visual feedback [3].
- Physiological evidence suggests considerable delays in the human visuo-motor (e.g. 130 ms for ocular-motor)[4].

Computational visual tracking

The tracking behavior assumes:

- Robust color-based object segmentation.
- No relation between successive frames.
- Intermittent control.

Markof Random Fields

The label of interest $\hat{\varphi}$ is the one that maximizes the a posteriori probability $P(\varphi \mid F)$:

$$\arg\max_{\varphi \in \Phi} \prod_{s \in I} P(f_s \mid \varphi_s) P(\varphi), \tag{1}$$

where $F = \{f_s \mid s \in I\}$, and Φ is the set of possible labellings.

Markof Random Fields

Singleton and doubleton cost

$$C(\varphi, F) = S(\varphi, F) + D(\varphi) \tag{2}$$

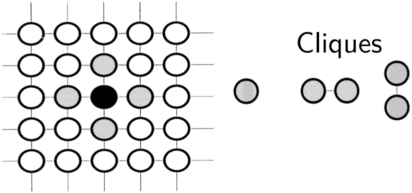


Figure: First-order neighborhood system. Single pixel cliques are called singletons, horizontal and vertical cliques are called doubletons [5].

Markof Random Fields

A multivariate normal distribution can be considered as the singleton contribution, such as

$$f_{\mathrm{x}}(x_1, ..., x_k) = \frac{1}{\sqrt{2\pi^k|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{t}}\Sigma^{-1}(x-\mu)\right) \quad (3)$$

Resulting in the Kato et al. proposal[5]

$$S(\varphi, F) = \sum_{s \in S} \ln(\sqrt{(2\pi)^3|\Sigma_{\varphi_s}|}) + \frac{1}{2}((f_s - \mu_{\varphi_s})\Sigma_{\varphi_s}{}^{-1}(f_s - \mu_{\varphi_s})^{\mathrm{t}}) \quad (4)$$

Markof Random Fields

Homogeneity of labeling determines the doubleton cost [6]

$$D(\varphi) = \beta \sum_{\{s,r\} \in N} (1 - \delta(\varphi_s, \varphi_r)), \tag{5}$$

where the function $\delta(a, b)$ is known as the "Kronecker delta function" such as

$$\delta(a, b) = \begin{cases} 1 & \text{if} \quad a = b \\ 0 & \text{if} \quad a \neq b \end{cases} \tag{6}$$

Markof Random Fields

Supervised segmentation:

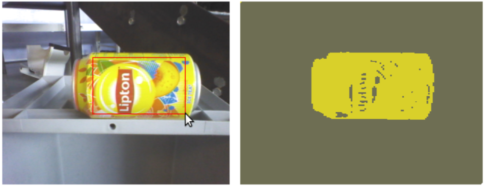
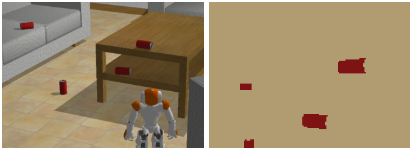


Figure: On the left, the user selects the region from where the color model will be built, in other words $P(f_s \mid \varphi_s)$. On the right a segmentation φ obtained.

Computational paradigm limitations

Color model segmentation is not enough:

- Robot motions can cause the object to leave the field of vision.
- High chance of confounding similar objects in the scene.
- Enriching the model may not solve the problem.



Embodied design principle

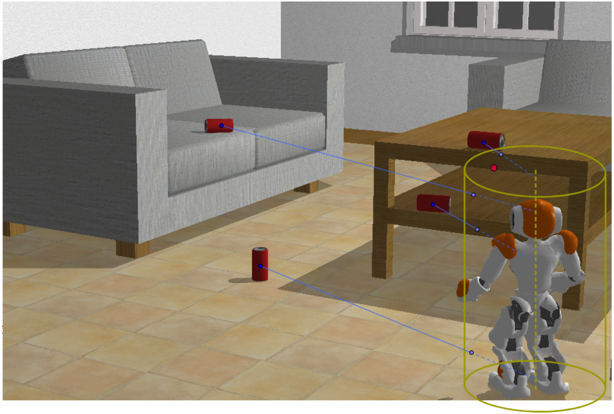
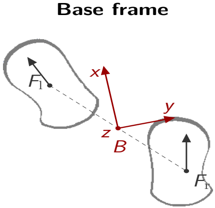


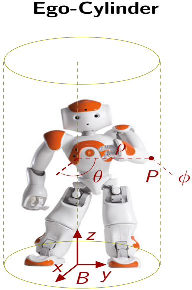
Figure: Projection of blob's center on the ego-space. The red dot corresponds to the prediction.

Outline

Sensory ego-cylinder



(a)



(b)

Figure: Representation of the ego-localization. a) Movable base frame B. b) Representation of the ego-cylinder. $P = \begin{bmatrix} \rho & \theta & z & \phi \end{bmatrix}^{\mathrm{t}}$.

Task Frames

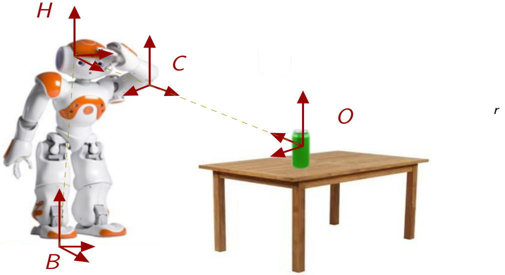


Figure: Definition of the reference frames to solve the localization task. In the image, B corresponds to the base frame, H to the head frame, C to the camera frame, and O to the object frame.

Localization

The object's pose can be known with respect to the base frame B through the definition of the homogeneous transformation matrix

$$^{B}T_{\mathrm{O}} = {}^{B}T_{\mathrm{H}}(q)^{\mathrm{H}}\mathrm{T}_{\mathrm{C}}{}^{\mathrm{C}}T_{\mathrm{O}}, \tag{7}$$

A pose P in the ego-cylinder is calculated from $^{B}T_{\mathrm{O}}$ and is given by

$$P = \begin{bmatrix} \rho & \theta & z & \phi \end{bmatrix}^{\mathrm{t}}. \tag{8}$$

Localization

The transformation ${}^{\mathrm{C}}T_{\mathrm{O}}$ expresses the object frame O in frame C, and is determined from the 3D pose

$$
{}^{\mathrm{C}}O = \begin{bmatrix} \zeta & \omega \end{bmatrix}^{\mathrm{t}} = \begin{bmatrix} \begin{bmatrix} x & y & z \end{bmatrix} & \begin{bmatrix} \gamma & \beta & \theta \end{bmatrix} \end{bmatrix}^{\mathrm{t}}, \tag{9}
$$

Cylindrical container

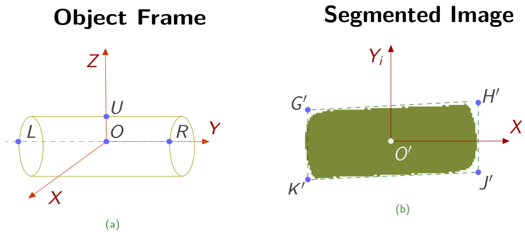
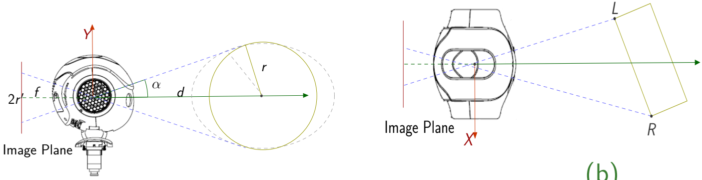


Figure: Definition of the cylindrical container object model. a) 3D representation of the object frame O and the definition of four points of interest. b) Segmented blob and image features defined from the oriented bounding box.

Cylindrical container

Depth estimation



(a) (b)

Figure: Estimation of the object's depth. a) The model assumes ${}^{\mathrm{C}}O_{\theta} = 0$ in (9), b) XZ visualization of the scenario where the circumference corresponds to an ellipse and the distance from the projective ray and the center O is larger than r.

Cylindrical container

Thus, the position component ${}^{\mathrm{C}}O_{\zeta}$ in (9) is given by

$$ {}^{\mathrm{C}}O_{\zeta} = \begin{bmatrix} M_{\mathrm{x}} & M_{\mathrm{y}} & M_{\mathrm{z}} + r \end{bmatrix}^{\mathrm{t}}, \tag{10} $$

where $M = \mathrm{mean}({}^{\mathrm{C}}L, {}^{\mathrm{C}}R)$.

Cylindrical container

The orientation component ${}^{\mathrm{C}}O_\omega$ in (9) is obtained from the relation between ${}^{\mathrm{C}}R$, ${}^{\mathrm{C}}L$, ${}^{\mathrm{C}}U$, and ${}^{\mathrm{C}}O$. It is extracted from the rotation matrix

$$ {}^{\mathrm{C}}R = \begin{bmatrix} s & n & a \end{bmatrix} = \begin{bmatrix} \hat{H} & \hat{V} & (\hat{H} \times \hat{V}) \end{bmatrix}, \qquad (11) $$

with $\hat{H} = ({}^{\mathrm{C}}R - {}^{\mathrm{C}}L)/|{}^{\mathrm{C}}R - {}^{\mathrm{C}}L|$, and $\hat{V} = ({}^{\mathrm{C}}U - {}^{\mathrm{C}}O)/|{}^{\mathrm{C}}U - {}^{\mathrm{C}}O|$.

Outline

Predictive motion

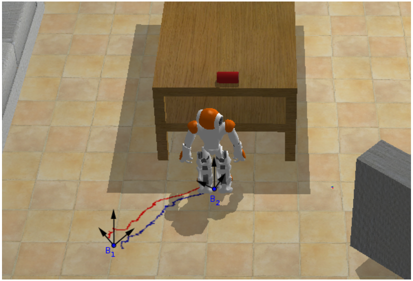


Figure: Initial and desired pose of frame B.

Predictive motion

The transformation ${}^{B}T_{\mathrm{B}*}$ between the mobile frame B and the desired location B^{*}

$$ {}^{B}T_{\mathrm{B}*} = {}^{B}T_{\mathrm{O}}{}^{\mathrm{O}}T_{\mathrm{B}*}, \tag{12} $$

The transformation ${}^{\mathrm{O}}T_{\mathrm{B}*}$ is defined by demonstration.

Predictive motion

A difference in location ${}^{\mathrm{B}}d$ is obtained from ${}^{B}T_{\mathrm{B}^*}$. A direction of motion $\bar{M}$ can be defined as follows

$$ {}^{B}\bar{M} = sat({}^{B}d, \lambda), \tag{13} $$

where *sat* is a saturation function to λ thresholds.

Predictive motion

A prediction ${}^{B}\hat{P}_{k+1}$ can be calculated such as

$$ {}^{B}\hat{P}_{k+1}(P_k, \bar{M}) = \begin{bmatrix} \sqrt{\rho^2 - 2c\rho\bar{\rho} + \bar{\rho}^2} \\ \operatorname{atan2}(-\rho s, \rho c - \bar{\rho}) \\ z \\ \phi - \bar{\phi} \end{bmatrix}, \qquad (14) $$

with $c = \cos(\theta - \bar{\phi})$, $s = \sin(\theta - \bar{\phi})$, P_k is the current location, $\bar{M}$ is a direction of motion, and ($\bar{}$) denoting the elements of $\bar{M}$.

Outline

The tracking algorithm

The segmentation routine consisted in a customization of the MRF *supervised* technique

- Computational complexity $\varsigma = O(n^2)$, for n pixels, and $|\Phi| = 2$ labels.
- Images were processed in the YUV color-space, so $|\eta| = 3$.
- No background model is used.
- Reasonable performance for naturally illuminated scenes.

The tracking algorithm

Algorithm 1 Segmentation

procedure DOSEGMENTATION
    $\hat{\varphi}(i,j) \leftarrow Initialize()$ ▷ Singleton initialization
    $e_{Old} \leftarrow 0$
    repeat
        $e \leftarrow 0$
        for $i = 0 \rightarrow i < height$ **do**
            $min_e \leftarrow localEnergy(i,j,\hat{\varphi}(i,j))$
            for $j = 0 \rightarrow j < width$ **do**
                for $\lambda = 0 \rightarrow \lambda < |\Phi|$ **do**
                    $c_e \leftarrow localEnergy(i,j,\lambda)$ ▷ current energy
                    if $c_e < min_e$ **then**
                        $\hat{\varphi}(i,j) \leftarrow \lambda$
                        $min_e \leftarrow c_e$
            $e \leftarrow e + min_e$
        $\Delta e \leftarrow abs(e_{Old} - e)$
        $e_{Old} \leftarrow e$ ▷ stop when the change is too small
    until $\Delta e > t$

The simulation environment

The simulated scene

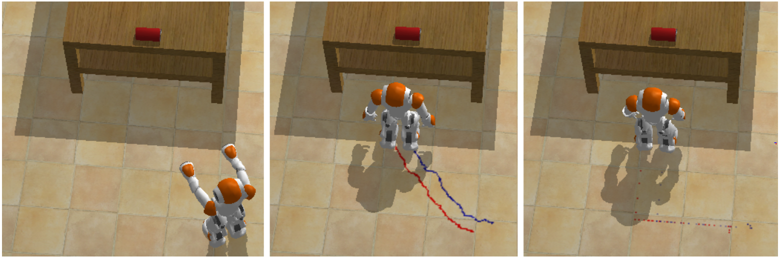


Figure: The approach task modeled in Webots. On the left, the robot's original pose. In the center, the followed path. On the right, the desired pose with respect to the red can on the table

The simulation environment

Multiple objects in the scene

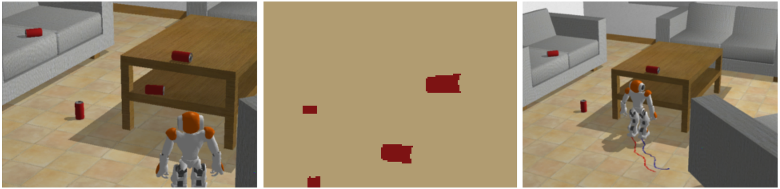


Figure: Top-down feature attention testing. On the left, the robot's original pose. In the center, the on-board view of the frame on the left. On the right, the trajectory followed. Despite the number of red cans, the agent was able to track the one on the top of the table.

The simulation environment

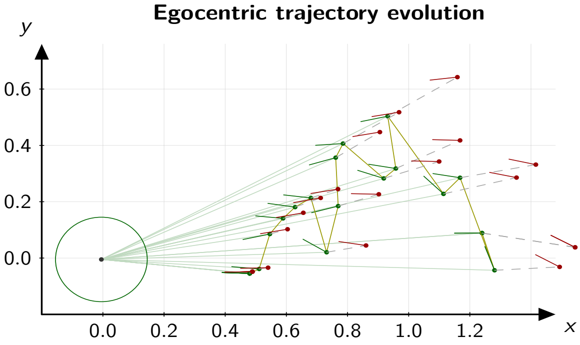


Figure: XY egocentric visualization. The circumference represents the ego-cylinder. In red the real values, in green the estimations. Distances are expressed in m.

The simulation environment

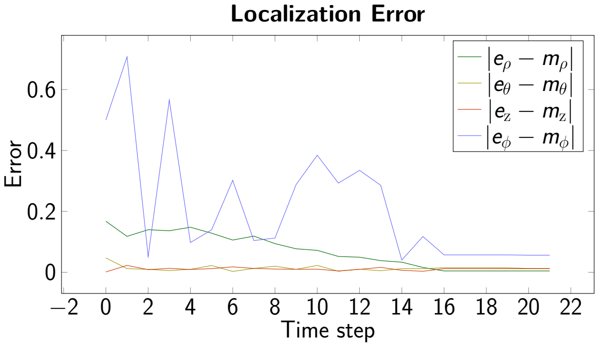
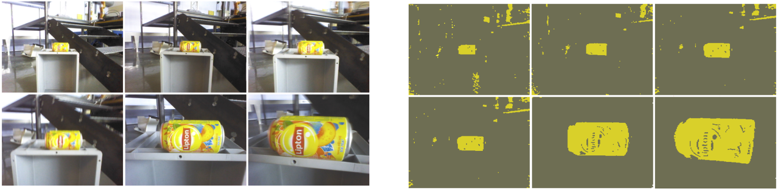


Figure: Evolution of the localization error between estimations e and measurements m.

Card approach

Let's watch the video!

Thea can approach



(a) (b)

Figure: On-board view of the second experimental task where the robot approached a tea can. a) Some of the captured images of the scene. b) Corresponding unfiltered segmentations.

Outline

Conclusions

- First person perspective analysis for problems in assistive robotics.
- Task subdivision intro three sequential steps: object tracking, ego-localization, and predictive motion.

Conclusions

- The MRF formalism ensured reasonably robust color-based segmentation.
- The top-down feature attention mechanism was crucial to discriminate between similar objects.

Conclusions

- The ego-localization representation seemed to be adequate.
- A sequential look-then-move policy appears to be sufficient to perform the task.

Conclusions

- Despite the simplicity of the models and the perturbations involved, the agent was able to accomplish the task.
- This work argued in favor of analyzing the task from the perspective of embodiment, while including simple computational models to ensure more general solutions to assistance robotics problems.

Thank you very much!

References

P. Michel, J. Chestnutt, J. Kuffner, and T. Kanade, "Vision-guided humanoid footstep planning for dynamic environments," in *Proceedings of the IEEE-RAS Conference on Humanoid Robots (Humanoids'05)*, pp. 13 – 18, December 2005.

C. Dune, A. Herdt, O. Stasse, P. B. Wieber, K. Yokoi, and E. Yoshida, "Cancelling the sway motion of dynamic walking in visual servoing," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3175–3180, 2010.

A. Moughlbay, E. Cervera, and P. Martinet, "Model based visual servoing tasks with an autonomous humanoid robot," in *Frontiers of Intelligent Autonomous Systems* (S. Lee, K.-J. Yoon, and J. Lee, eds.), vol. 466 of *Studies in Computational Intelligence*, pp. 149–162, Springer Berlin Heidelberg, 2013.

R. C. Miall, D. J. Weir, D. M. Wolpert, and J. F. Stein, "Is the cerebellum a smith predictor?," *Journal of motor behavior*, vol. 25, pp. 203–216, Sept. 1993.
PMID: 12581990.

Z. Kato, T.-C. Pong, and J. Chung-Mong Lee, "Color image segmentation and parameter estimation in a markovian framework," *Pattern Recognition Letters*, vol. 22, pp. 309–321, Mar. 2001.

J. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 48, no. 3, pp. 259–302, 1986.