EAI Endorsed Transactions
on Creative Technologies
Research Article ICST.ORG

CuriousMind photographer: distract the robot from its initial task

Vincent Courboulay[1,*], Matei Mancas[2]

[1]L3I, University of La Rochelle, Av Michel Crepeau, 17042 La Rochelle, France.
[2]Numediart Institute, University of Mons (UMONS) 20, Place du Parc, 7000, Mons, Belgium

Abstract

Mainly present in industry, robots begin to invade our every-day lives for very precise tasks. In order to reach a level where more general robots get involved in our lives, the robots' abilities to communicate and to react to unexpected situations must be improved. This paper introduces an attentive computational model for robots as attention can help both in reacting to unexpected situations and to help improving human-robot communication. We propose to enhance and implement an existing real time computational model. Intensity, color and orientation are usually used but we have added information related to depth and isolation. We have built a robotic system based on LEGO Mindstorm platform and the Kinect RGB-D sensor. This robot, called CuriousMind, is able to take a picture of the most interesting part of the scene and it can also be distracted from its first goal by novel situations mimicking in that way the human (and more precisely small children) behaviour.

Keywords: Attentional system, robotic implementation, 3D saliency.

Received on 09 May 2014, accepted on 05 September 2014, published on 27 February 2015

doi: 10.4108/ct.2.2e4

1. Introduction

1.1. Context

Robots will help us in the future with all the boring daily tasks: housekeeping, shopping, classification, etc. This implies that we will have many interactions with intelligent robots. To do this they need to fit into our lives with comprehensive abilities: vision, grasping, motion, etc. For us, human beings, all of these capabilities are often conditioned by our ability to pay attention to something (person, object, word, etc). If we cannot pay attention to the world around us we can neither anticipate dangers, nor share with others. Visual attention is, by the way, an important phenomenon to be able to understand our environment. It corresponds to the mechanisms that enable us to select visual information in order to process some clues in particular and also that we use to attract someone's else attention and communicate with him. The ability to pay attention is thus used by humans to understand their environment, to adapt to environment changes and to communicate with others.

While machine vision systems are becoming increasingly powerful, in most regards they are still far inferior to their biological counterparts. Attention is important for robots for 1) its functional objective related to their mechanic possibilities and 2) their limited abilities to process information.

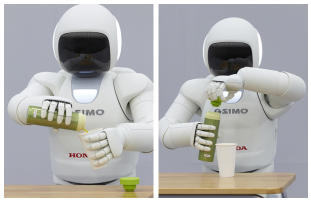
The first point considers that our processing capabilities are unlimited. For proponents of this theory ([2], [19], [28]), attention would not be a filter for our limited brain capacities, but would be a filter for our limited capacities of action. Motor skills are limited by morphology, for example hands can only handle one (or two) objects simultaneously (cf Figure 1 (a)). Thus, action capacities are limited and require the collecting of a selection of information in order to treat it accurately.

The second theory considers irrelevant messages are filtered out before the stimulus information is processed for meaning. In other words, if our brain were bigger and/or more powerful, we would not need attentional mechanisms [3]. In this context, attention selects some information in order not to overload our cognitive system. This is also the basic premise of a large number of computational models of visual attention [1] [12] [15] and [24].

1.2. Hypothesis

The objective of this article is to propose an attentive computational model for robots. This model is an enhancement of [15] and [6]. The main difference between the above mentioned models and what is actually implemented on a robot mainly relies on the presence of spatial information extracted from a depth map acquired

*Corresponding author. Email: vincent.courboulay@univ-lr.fr



(a) Asimo pouring out a glass of water

(b) Isolated snooker blue ball

Figure 1. Future robots will have to focus their attention either on close object such a glass of water or farther like the blue ball.

by a low-cost sensor. We propose to integrate two new conspicuity maps in the existing framework:

- one for the depth of objects itself,
- one for the objects isolation in the 3D scene.

The depth map helps to promote the nearest elements. The depth map acts like top-down information stating that closer objects need to be taken into account first as they are the most likely to collide with the robot.

The isolation map brings out an element, even banal or diffuse, but clearly separated from the rest of its surroundings in terms of depth (cf Figure 1 (b)). This map is a bottom-up approach and tends to highlight objects which are in a different configuration than the others independently from the object distance from the robot.

In the following section we describe a few computational models of attention as well as our contributions concerning a model of attention for a robotic system. In section 3 we describe how we have integrated our model in a robotic system. Section 4 provides first experiments. Finally, section 6 presents conclusions and some outlooks.

2. Attentive robots

The tasks of the robot which involves visual attention might be classified roughly into three categories [7]:

- low-level category: uses attention to detect salient landmarks that can be used for localization and scene recognition,
- mid-level category: considers attention as a front-end for object recognition,
- highest-level category: attention is used in a human-like way to guide the actions of an autonomous system like a robot.

In the first category robots use landmarks to compute their position in space. In [20] or [27], authors used static maps in which specific landmarks are located. In [11], the robot has to build a map of its environment and to localize itself inside. Salient regions are tracked over several frames to obtain a 3D position of the landmarks, and match them to database entries of all previously seen landmarks.

In the second category, attention methods are of special interest for all tasks in object detection and localization, or in classification of non pre-segmented images. [13] has integrated attentive object detection on the robot. In the same way, *Curious George*, developed by the laboratory for computational intelligence of the University of British Columbia was ranked first in the robot league of the Semantic Robot Vision Challenge both in 2007 and 2008, and first in the software league for 2009[1] [18].

Finally, the highest-level category is dedicated to robots which have to act in a complex world facing the same problems as a human. One of the first active vision systems which integrated visual attention was presented by [5]. They describe how a robot can fixate and track the most salient regions in artificial scenes composed of geometric shapes. In [29], authors present an attention system which guides the gaze of a humanoid robot. The authors consider only one feature, visual flow, which enables the system to attend to moving objects. In [25], the humanoid robot *iCub* bases its decisions to move eyes and neck on visual and acoustic saliency maps. Other works concerning joint attention were done by [14] and [26].

As mentioned before, many methods exist, but most of them need either strong information concerning the locations of landmarks or concerning objects to recognize. What we propose is to enhance an easily tunable model which works in real time in order to integrate 3D information.

3. Our model and its extension

In this section we present the model, its evaluation and the way we implemented it in CuriousMind. We used the first steps of Laurent Itti's work [15]. The first part of its architecture relies on the extraction of three conspicuity maps based on low level characteristics computation that correspond to the production of information on the retina. These three conspicuity maps

[1] http://google-opensource.blogspot.fr/2010/01/2009-semantic-robot-vision-challenge.html

are representative of the three main human perceptual channels: color, intensity, and orientation. The second part of Itti's architecture proposes a medium level system which allows merging conspicuity maps (C^n), and then simulates a visual attention path on the observed scene. The focus is determined by "*winner-takes-all*" and "inhibition of return" algorithms. This method suffers from numerous lacks. First, it is not dynamic and do not support evolution, it cannot model dynamic path of focus of attention. This method avoids to return to previous sites already observed, and finally it is very difficult to add new information. This is mainly why we have developed our own model.

We have substituted this second part of the initial algorithm by our optimal competitive dynamics evolution equation [22], in which a predator density map I represents the level of interest the image contains and C^n represent respectively color, intensity and orientation prey populations *i.e.* which are the sources of interest (Figure 2).

For each of the conspicuity maps (color, intensity, orientation), the preys population C^n evolution is governed by the following equation:

$$\frac{d_{C^n_{x,y}}}{dt} = C^n_{x,y} + f\,\triangle_{C^{*n}_{x,y}} - m_C C^n_{x,y} - sC^n_{x,y} I_{x,y} \qquad (1)$$

with $C^{*n}_{x,y} = C^n_{x,y} + w{C^n_{x,y}}^2$ and $n \in \{c, i, o, m\}$, which means that this equation is valid for C^c, C^i, C^o and C^m which respectively represent color, intensity and orientation populations. w is a positive controlled feedback. This feedback models the fact that provided that there are unlimited resources the more numerous a population, the better it is able to grow. m^n_C is a mortality rate that allows to decrease the level of interest of regions in the conspicuity map C^n. The population of predators I, which consumes the three kinds of preys, is governed by the following equation:

$$\frac{d_{I_{x,y}}}{dt} = s(P_{x,y} + wI^2_{x,y}) + sf\,\triangle_{P_{x,y}+wI^2_{x,y}} - m_I I_{x,y} \qquad (2)$$

with $P_{x,y} = \sum_{n\in\{c,i,o,m\}}(C^n_{x,y})I_{x,y}$.
This yields to the following set of equations, modeling the evolution of prey and predator populations on a two dimensional map:

$$\begin{cases} \frac{d_{C^i_{x,y}}}{dt} & = bC^i_{x,y} + f\,\triangle_{C^i_{x,y}} - m_{C^i} C^i_{x,y} - sC^i_{x,y} I_{x,y} \\ \frac{d_{I_{x,y}}}{dt} & = sC^i_{x,y} I_{x,y} + sf\,\triangle_{P_{x,y}} - m_I I_{x,y} \end{cases} \qquad (3)$$

As already mentioned, the positive feedback factor w enforces the system dynamics and facilitates the emergence of chaotic behaviors by speeding up saturation in some areas of the maps. Lastly, the maximum of the interest map I at time t is the location of the focus of attention. This system has been implemented in real time, see [6, 22, 23]. A demonstrator of our model can be downloaded and tested on this web page `http://www.perreira.net/matthieu/downloads/vico-visual-attention-model/`.

3.1. Initial model evaluation

In [22], we have presented a very complete evaluation of our model. We used the cross-correlation, Kullback-Leibler divergence and normalized scanpath saliency measures between 6 efficient algorithms and an eye-tracking ground-truth.

In these evaluations, we have shown that our model is highly stable, robust, exploratory (we can easily define scene exploration strategy), dynamic, plausible, fast, and highly configurable.

All measures were done on two image databases. The first one is proposed in [4]. It is made up of 120 color images which represent streets, gardens, vehicles or buildings, more or less salient. The second one, proposed in [17], contains 26 color images. They represent sport scenes, animals, buildings, indoor scenes or landscapes. For both databases, eye movement recordings were performed during a free viewing task.

Regarding the numerous models that exist in literature, we have decided to benchmark our model. However, it is hard to make immediate comparisons between models. To alleviate this problem, T. Judd has proposed a benchmark data set containing 300 natural images with eye tracking data from 39 observers to compare model performances[2]. She writes, *this is the largest data set with so many viewers per image*. She calculates the performance of 10 models in predicting ground-truth fixations using three different metrics: a receiver operating characteristic, a similarity metric, and the Earth Mover's Distance. We have downloaded the database, have runned our model to create saliency maps of each image and have submitted our maps. We present the results in Table.1. References of algorithms can be found in the web page of the benchmark. Evaluation indicates that our system obtains an average ranking among almost 20 algorithms. In addition to such a benchmark, we have demonstrated in [22], that our model is :

- plausible,
- adaptable,
- invariant,
- rapid,
- extensible,

[2] http://people.csail.mit.edu/tjudd/SaliencyBenchmark

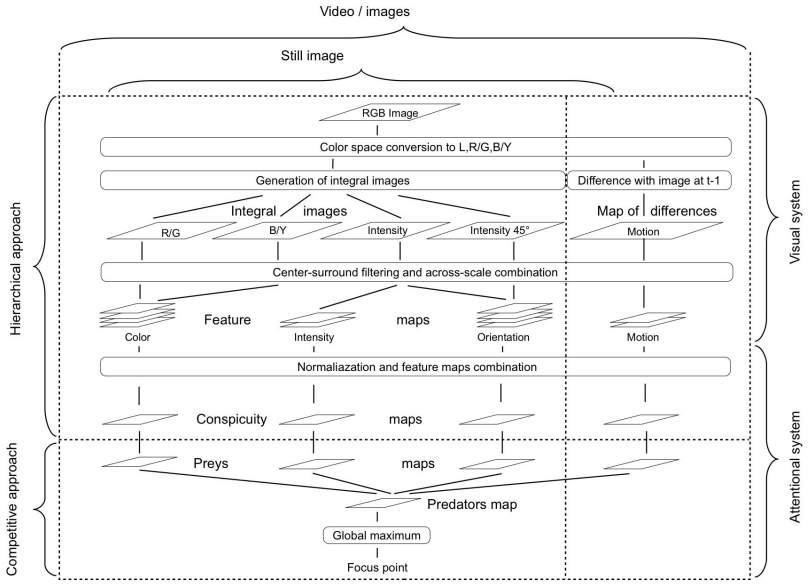


Figure 2. Schematic view of the proposed model of visual attention. After a first step similar to Itti [15] work, a prey-predator approach optimally fuse the information from different conspicuity maps.

Model Name	Area under ROC curve	Similarity
Chance	0.503	0.327
Achanta	0.523	0.297
Itti&Koch	0.562	0.284
SUN saliency	0.672	0.34
Hou & Zhang	0.682	0.319
Torralba	0.684	0.343
Context-Aware saliency	0.742	0.39
Preys/predators model	0.7496	0.4147
Itti&Koch 2	0.75	0.405
Bruce and Tsotsos AIM	0.751	0.39
Narayan model	0.753	0.42
RARE2012	0.7719	0.4363
Saliency for Image Manipulation	0.774	0.439
Center	0.783	0.451
CovSal	0.7999	0.4869
GBVS	0.801	0.472
Judd et al.	0.811	0.506
Humans	0.922	1

Table 1. A comparison of several models realised by T.Judd.

- dynamic.

3.2. Extension to robotic environment

In order to enhance our model, and make it usable for robotic applications, we have integrated with the previous model two new conspicuity maps. One for the depth and one for the isolated objects.

The depth conspicuity map. This map represents the depth of the scene in front of the robot. We have used a Kinect RGB-D sensor [8] from Microsoft which is a low-cost active 3D camera. The SDK provides Kinect capabilities for developers to build applications which include access to low-level streams from the depth sensor, the color camera sensor, and a four-element microphone array. The depth sensor consists of an infra red laser projector combined with a monochrome CMOS sensor, which captures video data in 3D under any ambient light conditions. Let I_d be the depth image. Each pixel represents approximately the distance between the Kinect sensor and each object of the scene. In order to promote close objects rather than distant ones we define the depth consipicuity map as the inverse of I_d between computable distances provided by the Kinect.

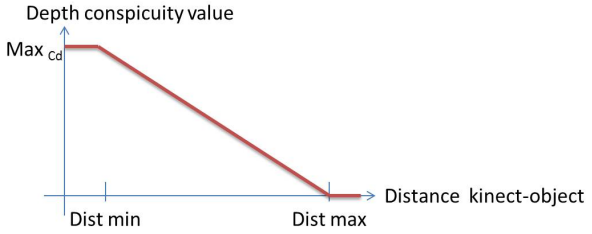
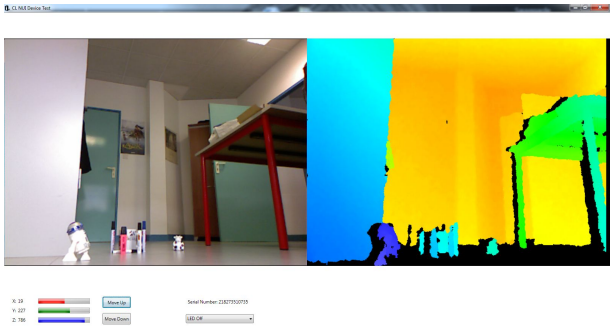


Figure 3. Transformation used to compute depth conspicuity pixel value from original depth one.

$$C_d(i,j) = \frac{-Max_{C_d}}{dyn_{I_d}} * I_d(i,j) + \alpha \quad (4)$$

where dyn_{I_d} represents the dynamic of image I_d and α a coefficient to constraint C_d to be positive. In order to avoid problems due to uncomputed depth in I_d, each null value on I_d remains null on C_d. In that sense, this map is different from the previous one. It is not based on visual perception but distance relevance. It does not include any center surround nor any multi-resolution analysis. Nevertheless, this information can be mixed with the other map thanks to our prey-predator fusion process presented in the previous section.

The isolation conspicuity map. This map has to focus on an isolated element. An isolated element is characterized by a pixel value different from its surroundings (lower or higher). In order to be as coherent as possible we have decided to use the same approach as the one used to detect information in the intensity conspicuity map. The difference is that the input is not the intensity information but the depth map provided by the Kinect. Thus, we compute centre-surround differences to determine contrast, by taking the difference between a fine (center) and a coarse scale (surround) for the depth feature. This operation across spatial scales is done by interpolation to the fine scale and then point-by-point subtraction (Figure 4).



Original image and its depth map

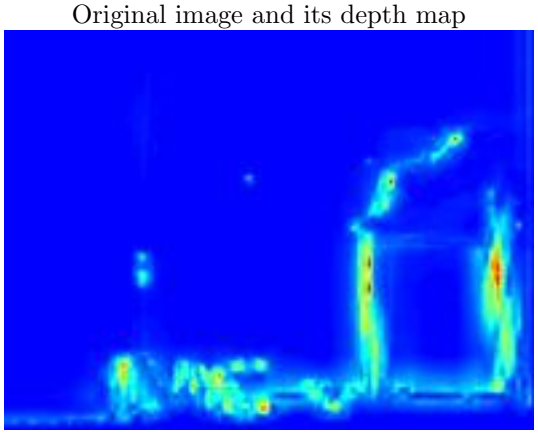


Figure 4. Isolation conspicuity map.

For each of the conspicuity maps (color, intensity orientation, depth and isolation), the prey population C^n evolution is governed by the following equation:

$$\frac{d_{C^n_{x,y}}}{dt} = C^n_{x,y} + f \triangle_{C^{*n}_{x,y}} - m_C C^n_{x,y} - s C^n_{x,y} I_{x,y} \quad (5)$$

4. Experimentation

For our experimentation we have decided to use a mobile system composed by a LEGO Mindstorm platform and a Kinect sensor (Figure 5). The LEGO Mindstorm allows motion, whereas Kinect allows video and depth acquisition. This system will be referred as CuriousMind in the rest of the paper. The LEGO Mindstorms series

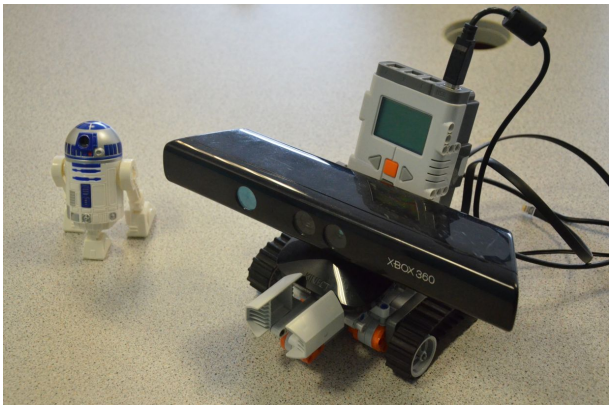


Figure 5. CuriousMind: a LEGO Mindstorm vehicle and a Kinect sensor.

of kits contain software and hardware to create small, customizable and programmable robots. They include a programmable brick computer that controls the system, a set of modular sensors and motors, and LEGO parts from the Technics line to create the mechanical systems. For our test we have decided to link the LEGO Mindstorm and the Kinect to a computer thanks to a USB liaison rather than Bluetooth. Free tools in combination with the Robotics Developer Studio developed by Microsoft [9] enable programming the Mindstorm using the C# language. Concerning Kinect, in 2011 Microsoft announced a non-commercial SDK [10] to build applications with C#, see Figure 6.

Thus, we have decided to use C# to manage our application. It runs in real time on a computer DELL precision M4700 core i5 CPU 2.8 GHz, 8Gb of RAM.

4.1. CuriousMind control

In this section, we present the algorithm we have implemented on CuriousMind.

In the algorithm 1, the first loop is an infinite thread dedicated to saliency analysis that provides three different variables. The first one Sal provides the average value of pixel inside the region of interest. This ROI is defined by a bounding box around the focus of attention. Its size is 10% of the dimension of the acquired image.

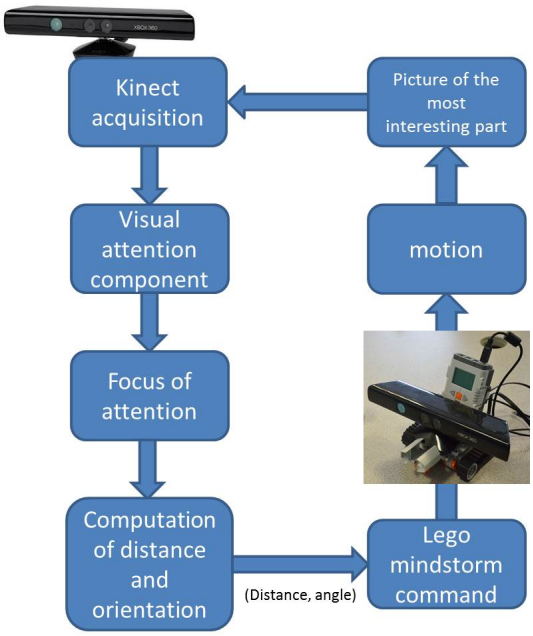


Figure 6. Block diagram of CuriousMind.

Algorithm 1 Take a picture of the nearest salient object

Require: Kinect connected
  loop
    Image $I \leftarrow$ Camera acquisition
    Image $D \leftarrow$ Depth acquisition
    Computation of the most salient region from I and D thanks to our model of attention
    $Sal \leftarrow$<Region of interest saliency value>
    $Dist \leftarrow$<Region of interest depth value>
    $\theta \leftarrow$<Orientation of the ROI>
  end loop
Require: LEGO Mindstrom connected
  loop
    $S \leftarrow Sal$
    Rotate CuriousMind from θ
    while $(S > \alpha * Sal)$ AND $(Distance < \beta * Dist)$ **do**
      CuriousMind moves straight forward
    end while
    if $(Distance >= \beta * Dist)$ **then**
      Take a picture
      loop
        CuriousMind waits for new orders
      end loop
    end if
  end loop

The second one $Dist$ is the minimum distance given by the Kinect inside the same region of interest. Finally, θ is the angle between the normal direction given by the Kinect and the focus of attention (see figure.7).

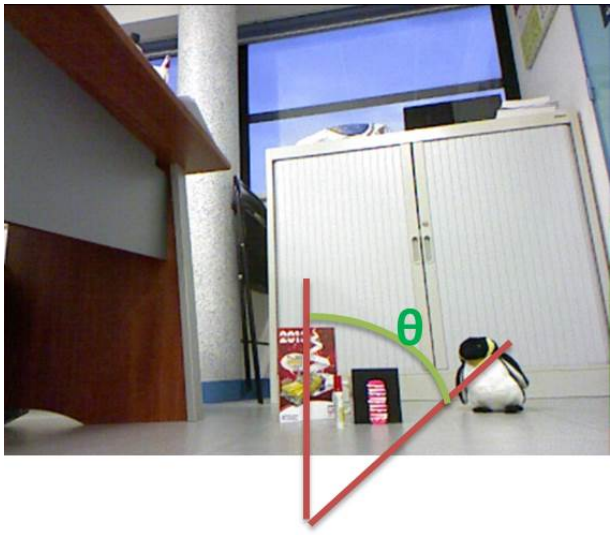


Figure 7. Definition of θ, the angle between the normal direction given by the Kinect and the focus of attention.

The second loop controls the LEGO Mindstorm. Firstly, the level of saliency is saved into a specific variable S. Then, CuriousMind rotates to be facing to the nearest salient region and starts to move towards this region until the distance is smaller than a portion of $Dist$ (for our test, we have chosen $\beta = 90\%$) or until a new salient region appears. If the robot arrives close to the region, it takes a picture and waits for new orders. If a new salient region appears, it starts again the second loop.

In order to avoid too many modification of CuriousMind objectives, we have decide to authorize the robot to be distracted from its initial task only when the new value of Sal is higher than α times the initial one. We have chosen α equal to 1.5. This parameter controls the ability of CuriousMind to be distracted from its initial goal. In humans this parameters varies with age and environment knowledge. It is very easy to distract a small baby from her/his initial goal just by showing her/him a new object as a baby has little knowledge about this environment ans she/he wants to learn a maximum of things. It is not the same with an adult who has a precise task to do and who has a priority hierarchy. In this case, the environment change might need to be drastic to disturb the person.

5. First evaluation

It is very difficult to objectively evaluate our system. In fact we should evaluate the relevance of our results by using a head-mounted eye tracking solution. Moreover, the view of the robot is very close to the ground which adds practical issues to an eye-tracking based evaluation. As a first attempt of performance evaluation, we prefer

assigning a specific objective to our system, and then subjectively evaluate the result.

The objective assigned to our robot is to go to the *nearest salient* object and take a picture of it when the object is large enough in its field of view. In this way we have a photographer robot which will provide a set of pictures of the most interesting locations in the scene (from its own point of view).

An example is given Figure 9. We have done some experimentation in our lab, our office and hall. Figure 10 represents a small part of experiments. The objective we have assigned to the robot makes us slightly overweight the influence of the depth conspicuity map (twice as color intensity, orientation and isolation), nevertheless, when different objects are located to equal distance from CuriousMind, this one prefers to discover the one which is alone in its region, or the most salient from a color or texture point of view if two objects are alone at equal distance (cf. Figure.10). We present in Figure8, the saliency map and the picture obtained if the weight dedicated to isolation and depth is turned to zero.

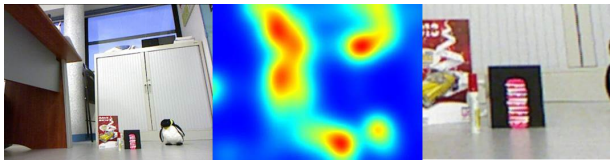


Figure 8. Most interesting part without taking into account isolation and depth.

Indeed, CuriousMind goes towards the areas containing objects of interest. If a new salient object is put in its field of view closer than the existing one it will reach the new object. This kind of behavior is a first step towards adaptation to general purpose events and also to communication. The action list of the robot can easily be updated in case of very salient bottom-up location which will distract the robot from its initial task allowing it to analyze this new event which might be a threat or an attempt to communicate with a human who would intentionally want to attract CuriousMind attention.

The ability to be distracted from the original task by unexpected events is a step towards emerging behavior.

6. Discussion and conclusion

This article proposes a low-cost robotic system based on a LEGO Mindstorm platform and a Kinect sensor which implements an attentive behavior capable of being distracted from its initial task like humans do. A parameter can be set to allow more or less distraction for a robot function of the task priority or the robot environment knowledge. The attentional aspect in robotics is complex and has been addressed by only a few previous works, but represent an important

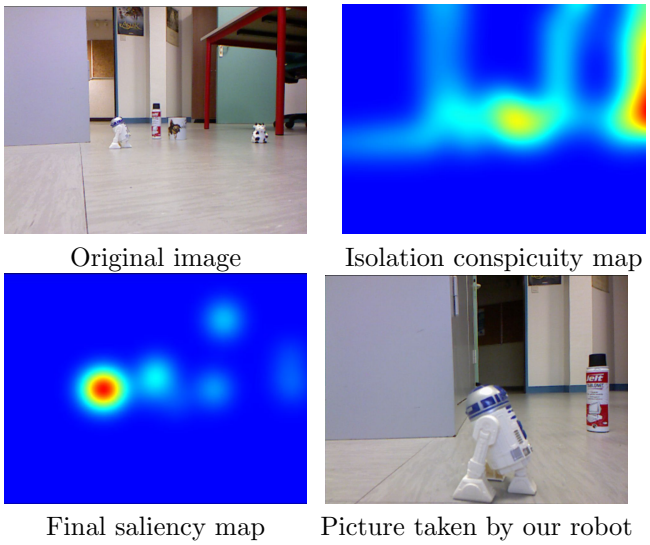


Figure 9. Presentation of different elements of CuriousMind.

milestone for the future. Attention is guided by a real-time computational system inspired by [22] and modified in order to take into account depth and isolation which are crucial in a real-life robotic environment.

We have conducted promising experiments which show that our robot, CuriousMind, is able to reach the most interesting locations for a simple application which is to obtain a panorama of the outstanding views in the lab from a robotic point of view. Moreover, CuriousMind is able to cancel previous tasks/actions to respond to an unexpected distractor if this one is salient enough.

The saliency algorithm provides a generic module capable of reacting in case of novel and unexpected situations for which the robot is not previously trained. This component is complementary to the specific task-oriented classical modules in robotics. It is able to respond to danger and provide reactivity in human-robot communication by a better adaptation to human actions despite an original task to be solved. This first step provides CuriousMind the ability to be distracted in the same way as humans by surprising scenes which can lead to a robot-human empathy improvement. We also look for novel behavior as artificial humor emergence from unexpected situations and unexpected robot behavior.

As a future work, we will implement our system inside a Nao (Figure 11), an autonomous, programmable humanoid robot developed by Aldebaran Robotics. Moreover, we will integrate a motion conspicuity map, to be reactive when a new element moves inside the robot field of view. A more in-depth validation will be conducted using the NUS3D-Saliency Dataset provided by Tam V. Nguyen [16] which includes depth maps and eye-tracking results on both 2D and 3D scenes.

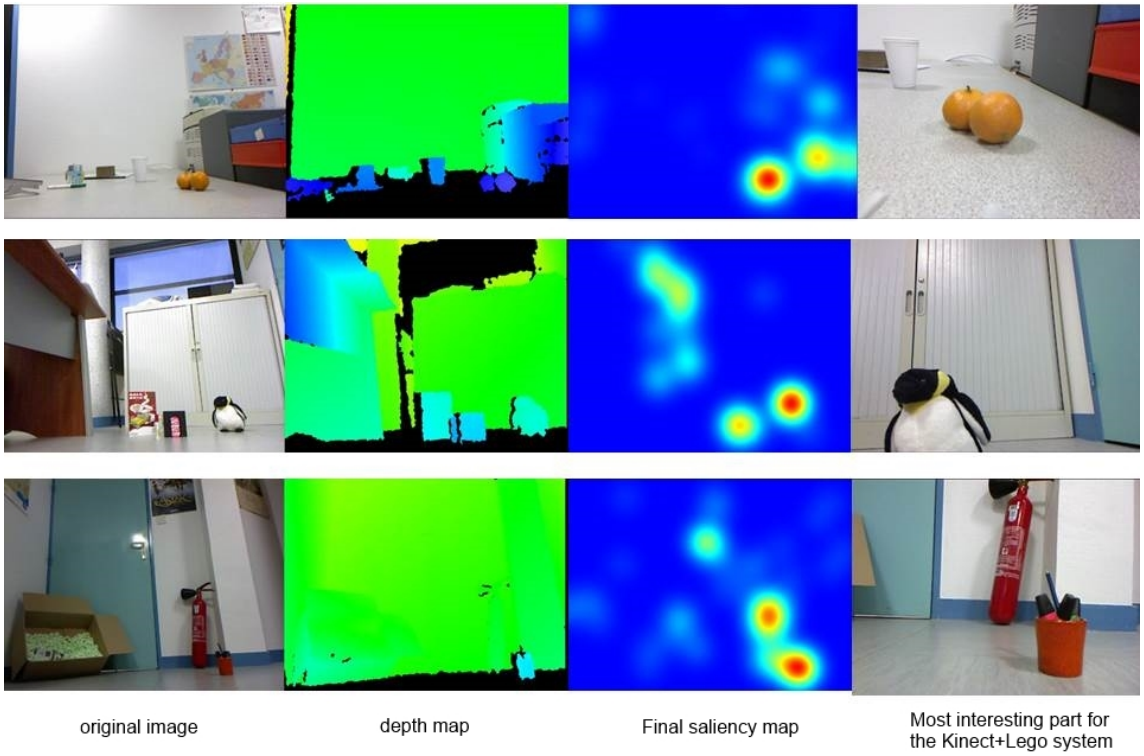


Figure 10. Sample of experiments realised. Left column: initial point of view of CuriousMind. Second column: raw depth map, third column: saliency maps, fourth column: picture taken by CuriousMind once it reached the interesting objects.

Finally, we want to go deeper into the idea of using distractors to add a capacity to develop humor both by detecting situations which might induce humor and by provoking those situation and trying to make a human laugh.

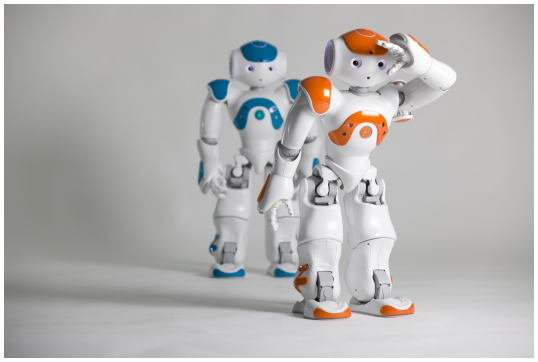


Figure 11. Nao looks at its futures capabilities.

References

[1] Ali Borji and Laurent Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185–207, Jan. 2013.

[2] Allport, D. A. (1987). *Selection for action: Some behavioral and neurophysiological considerations of attention and action*, pages 395–419. Lawrence Erlbaum Associates, Hillsdale, NJ.

[3] Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press, Elmsford, NY, US.

[4] B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, pp. 5, 2009.

[5] Clark, J. and Ferrier, N. J. (1988). Modal control of an attentive vision system. In , *Second International Conference on Computer Vision*, pages 514–523.

[6] Courboulay, V. and Perreira Da Silva, M. (2012). Real-time computational attention model for dynamic scenes analysis: from implementation to evaluation. In SPIE, editor, *SPIE Optics, Photonics and Digital Technologies for Multimedia Applications*, Vol. 8436, Brussels, Belgique.

[7] Frintrop, S. (2011). Towards attentive robots. *Paladyn*, 2(2):64–70.

[8] Microsoft Kinect sensor, http://www.xbox.com/kinect

[9] Microsoft Robotics Developer Studio, http://www.microsoft.com/en-us/download/details.aspx?id=29081

[10] Microsoft Kinect SDK, http://www.microsoft.com/en-us/kinectforwindows/develop/

[11] Frintrop, S. and Jensfelt, P. (2008). Attentional landmarks and active gaze control for visual SLAM. *IEEE Transactions on Robotics*, 24(5):1054–1065.

[12] Frintrop, S., Klodt, M., and Rome, E. (2007). A real-time visual attention system using integral images. In *5th International Conference on Computer Vision Systems (ICVS)*, Bielefeld, Germany. Applied Computer Science

Group.

[13] Gould, S., Arfvidsson, J., Kaehler, A., Sapp, B., Messner, M., Bradski, G., Baumstarck, P., Chung, S., and Ng, A. Y. (2007). Peripheral-foveal vision for real-time object recognition and tracking in video. In *Proceedings of the 20th international joint conference on Artifical intelligence*, IJCAI'07, page 21152121, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[14] Heidemann, G., Rae, R., Bekel, H., Bax, I., and Ritter, H. (2003). Integrating context-free and context-dependent attentional mechanisms for gestural object reference. In Crowley, J. L., Piater, J. H., Vincze, M., and Paletta, L., editors, *Computer Vision Systems*, number 2626 in Lecture Notes in Computer Science, pages 22–33. Springer Berlin Heidelberg.

[15] Itti, L., Koch, C., Niebur, E., and Others (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259.

[16] Lang, C., Nguyen, T. V., Katti, H., Yadati, K., Kankanhalli, M., and Yan, S. (2012). Depth matters: Influence of depth cues on visual saliency. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *Computer Vision ECCV 2012*, Lecture Notes in Computer Science, pages 101–115. Springer Berlin Heidelberg.

[17] O. Le Meur, Patrick Le Callet, Dominique Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, 2006.

[18] Meger, D., Forssén, P.-E., Lai, K., Helmer, S., McCann, S., Southey, T., Baumann, M., Little, J. J., Lowe, D. G., and Dow, B. (2007). Curious george: An attentive semantic robot. In *IROS 2007 Workshop: From sensors to human spatial concepts*, San Diego, CA, USA. IEEE.

[19] Neumann, O. (1987). *Beyond capacity: A functional view of attention. Perspectives on perception and action.*, pages 361–394. Lawrence Erlbaum Associates, Hillsdale, NJ, England.

[20] Nickerson, S. B., Jasiobedzki, P., Wilkes, D., Jenkin, M., Milios, E., Tsotsos, J., Jepson, A., and Bains, O. N. (1998). The ARK project: Autonomous mobile robots for known industrial environments. *Robotics and Autonomous Systems*, 25:83104.

[21] Ouerhani, N. (2003). *Visual Attention : From Bio-Inspired Modeling to Real-Time Implementation*. Thèse de doctorat, Université de Neuchâtel.

[22] Perreira Da Silva, M. and Courboulay, V. (2012). Implementation and evaluation of a computational model of attention for computer vision. In *Developing and Applying Biologically-Inspired Vision Systems: Interdisciplinary Concepts*, pages 273–306. Hershey, Pennsylvania: IGI Global.

[23] Perreira Da Silva, M., Courboulay, V., Prigent, A., and Estraillier, P. (2010). Evaluation of preys / predators systems for visual attention simulation. In *VISAPP 2010 - International Conference on Computer Vision Theory and Applications*, pages 275–282, Angers. INSTICC.

[24] Nicolas Riche, Matei Mancas, Matthieu Duvinage, Makiese Mibulumukini, Bernard Gosselin, and Thierry Dutoit, "Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis," *Signal Processing: Image Communication*, vol. 28, no. 6, pp. 642 – 658, 2013.

[25] Ruesch, J., Lopes, M., Bernardino, A., Hornstein, J., Santos-Victor, J., and Pfeifer, R. (2008). Multimodal saliency-based bottom-up attention a framework for the humanoid robot iCub. In *IEEE International Conference on Robotics and Automation, 2008. ICRA 2008*, pages 962–967.

[26] Schauerte, B., Richarz, J., and Fink, G. (2010). Saliency-based identification and recognition of pointed-at objects. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4638–4643.

[27] Siagian, C. and Itti, L. (2009). Biologically inspired mobile robot vision localization. *IEEE Transactions on Robotics*, 25(4):861–873.

[28] van der Heijden, A. H. C. and Bem, S. (1997). Successive approximations to an adequate model of attention. *Consciousness and cognition*, 6(2-3):413–28.

[29] Vijayakumar, S., Conradt, J., Shibata, T., and Schaal, S. (2001). Overt visual attention for a humanoid robot. In *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, volume 4, pages 2332–2337 vol.4.