Stochastic Extension of the Attention-Selection System for the iCub

Technical Report (in preparation)

H. Martinez, M. Lungarella, and R. Pfeifer
Artificial Intelligence Laboratory
Department of Informatics
University of Zurich, Switzerland

Abstract

To provide humanoid robots with means to act in the environment, they need to be endowed with some kind of mechanism to select and extract meaningful information. In this report, we describe a stochastic extension of the attention-selection system of the iCub platform. A super-diffusive stochastic process restricted by a multimodal saliency map, drives the robot's visual exploration based in the actual salient conditions. Our results show that such a stochastic extension yields visual explorative behavior which (a) is faster than to the standard winner-take-all saliency-driven scheme, and (b) does not require storing the location of the previous focus of attention.

Introduction

The attention system serves as a front gate to select from an abundance of streaming sensory input and thus is an important prerequisite for the ontogenesis of sensory-motor coordinated behaviors such as reaching and grasping. In this report, we describe a biologically-inspired stochastic attention selection mechanism and its integration in the iCub's multimodal attention framework (Ruesch et al., 2008). Based on existing results that demonstrate that human eye saccades can be modeled as a super-diffusive process (Brockmann et al., 1999), we conjectured that the incorporation of a Levy-flight random walk strategy in the attention selection mechanism, would lead to an increase of the performance of the robot's in terms of: (a) speed of the visual exploration process, and (b) the related energy consumption. Using manually generated images, we performed a systematic parameter selection for the Levy-flight-based attention selection algorithm. We conducted an experiment in order to evaluate the Levy Flight-based attention selection algorithm while embedded in the iCub's multi-modal attention and sensory-motor framework. We also compared the performance against the iCub's existing winner-take-all saliency-driven attention selection mechanism. In this experiment, the robot performed a visual exploration task, which integrates the low-level sensory processing, multi-modal sensory integration, bottom-up saliency, and attention selection mechanism. The results demonstrate that the stochastic Levy Flight-based attention selection mechanism enables the robot to explore its environment up to three times faster compared to the standard winner-take-all saliency-driven attention mechanism.

Materials and Methods

The Robot. Our experimental test bed was the 6 degrees of freedom (DOF) iCub robot head with the attention system described in (Ruesch *et al.,* 2008). This attention system builds a saliency map based on visual and auditory information. The visual part consists of a series of filters that extract information such as color, movement, orientation and intensity from the scene. The auditory part computes the saliency in a time-frequency domain, a high saliency is assigned to short and long tones, as well as to the absence of frequencies in a broad band noise, and to temporally modulated tones compared to stationary tones. In both cases (visual and auditory), each point in the space is accumulated to build a visual saliency map and an auditory saliency map respectively. Using these maps, a final multimodal saliency map is generated by taking the

highest value at each point. Finally, given this information, a winner-take-all with inhibition-of-return algorithm is used to select the relevant point in the environment to focus on. The inhibition is important in order to avoid that the robot keeps looking at the same point all the time. This implementation uses two functions to modulate (1) how much time the robot can spend looking at something before it is inhibited from the multimodal saliency map, and (2) how much time it should be inhibited.

The Attention Selection Algorithm. The winner-take-all with inhibition-of-return algorithm can drive the robot to cyclical attention behaviors. In order to avoid this situation, we proposed the implementation of a stochastic process restricted by the saliency map as the mechanism in charge to select the new point of attention. The Levy- flight random walk has been demonstrated to produce a distribution similar to the one produced by human eye saccades and also minimizes the time needed to process the visual environment (Brockmann et al., 1999; Brockmann et al., 2000). Boccignone (2000) proposed an algorithm to constrain the random walk using the saliency map, in this algorithm, the direction of the jump is first selected based on a normal distribution. In contrast, in our Levy flight-based algorithm, all the jump characteristics are restricted by the saliency map.

1: Compute the saliency map s(x, y) of the image

2: Compute φ(x,y) = $\begin{cases} \exp(-\beta(s(x,y)-s(x_{new},y_{new})))/\sum_{r_{new}}\exp(-\beta(s(x,y)-s(x_{new},y_{new}))) \\ s(x,y) \\ \exp(-\beta(s(x,y)))/\sum_{r}\exp(-\beta(s(x,y))) \end{cases}$

3: r ←image center; n ← 0
4: **repeat**
5: Current fixation ←r; accepted ←false
6: **while** not accepted **do**
7: Generate randomly a jump $p(x,y)=\frac{D\phi(x,y)}{x^2+y^2+D^2}$
8: Compute $\Delta\hat{s}=\hat{s}(r_{new})-\hat{s}(r)$

$\hat{s}(x_s,y_s)=\sum_{x,y}s(x,y)\exp(-\frac{(x-x_s)^2-(y-y_s)^2}{\sigma^2})$

9: **if** $0 \le \Delta\hat{s}$ **then**
10: Store rnew;r ←rnew;
accepted ← true;
n ← n + 1
11: **else**
12: Generate a random number ρ
13: **if** $\rho \le \exp(\Delta\hat{s}/T)$ **then**
14: Store rnew; r ←rnew;
accepted ← true; n ← n + 1
15: **until** n<K

Figure 1: Proposed algorithm.

The main structure of the algorithm is as follows: (1) compute a function φ(x,y) which transforms the saliency data, and calculate the probability density function of the jump p(x,y). This function is made by the multiplication of φ(x,y) with a Levy probability density function (point 7 in Fig. 1 p(x,y)); (2) p(x,y) is used to randomly select the next point of attention, (3) this new point of attention has to be compared against an acceptance criteria based in the value of the saliency of the actual neighborhoods. If the saliency weighted addition of the points close to the chosen point, is greater than, the saliency weighted addition of the points close to actual attention point, then the selected point is accepted like the new point of attention. However, if it is not the case, a

random number between 0 and 1 is generated. If this is lower than a function that depends on the weighted saliency and on a "temperature" T, the point is accepted. The T value determines the amount of randomness in the acceptance process.

In Fig. 1 (step 2), the three different types of φ(x,y) that we used in this research are shown. With all these functions, it is necessary to set the "initial" values of the various parameters. The parameter D is part of the Levy probability density function, with this variable it is possible to control the volatility of the random walk, the higher D, the higher the probability for big jumps. β is a parameter to scale the difference between the actual and the next point of attention. If β is high the algorithm is more likely to select a point with high saliency and if β is low the algorithm behaves more like Levy random walk. All those parameters have to be selected appropriately to allow the robot to attend the environment "smoothly" according with the data from the saliency map.

Tuning and Validation of the Algorithm. In order to establish optimal values for these parameters and also optimal function φ(x,y), this algorithm was tested using the static image shown in Fig. 2. To evaluate the performance of the algorithm, we performed 10'000 iterations of the test and produced a statistics figure (Fig. 4). To calculate this figure, we calculated frequency map, based on the number of times that it takes for the algorithm to visit a point in the visual scene.

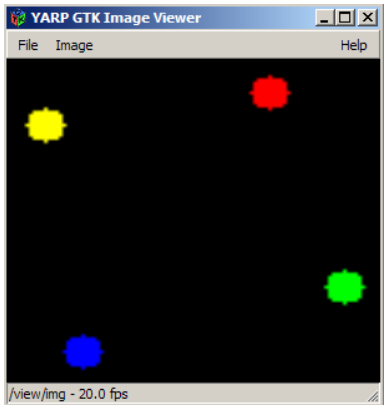


Figure 2: Test Image. Image used to tune and test the proposed algorithm.

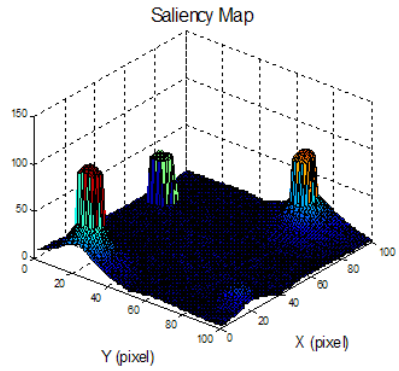


Figure 3: Saliency Map. Saliency map produced by the attention system for the Test Image.

1: Compute the saliency map s(x, y) of the image
2: Compute the frequency map f(x, y)
3: Initialize measure=0;
4: Compute for each column the mean and the standard deviation. (fstd, sstd, fmean, smean)
5: **for** i=1: Total of column
6: $$measure += (fstd(i) - sstd(i))^2 + (fmean(i) - smean(i))^2 + (\sum_{columni} f(x,y) - s(x,y))^2$$
7: **End**
8: $$measure = measure^{0.5}$$

Figure 4: "Statistics" Figure. Measure of the difference between the saliency and frequency maps.

This "statistics" figure shown in Fig. 4 is the tool used to compare the similarity between the saliency map and the frequency map. It provides statistical measurements over the space to

establish when the algorithm is capable of performing attention selection that is close to the saliency map.
Based on this result, we selected the appropriate φ(x,y) function as well as the other parameters of the algorithm. Fig. 5 to Fig. 7 show the results for different combinations of parameters and φ(x,y) functions.

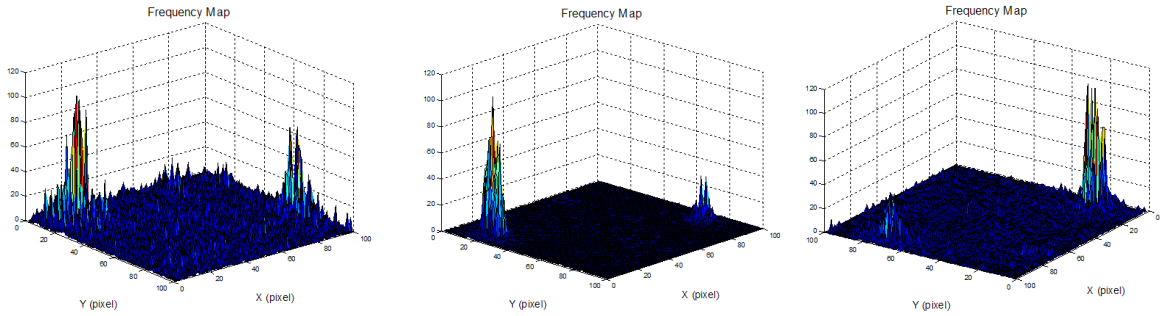


Figure 5. Frequency Maps for Φ1. Different frequency maps developed, using the function Φ1 for different values of parameters using the test image.

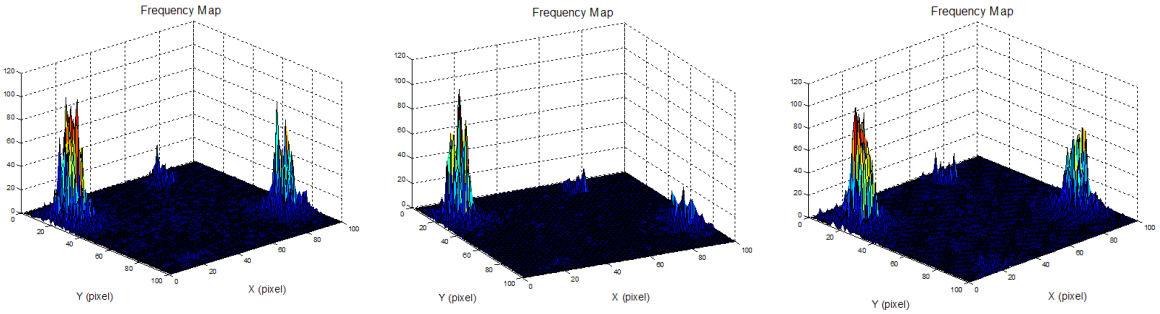


Figure 6. Frequency Maps for Φ2. Different frequency maps developed, using the function Φ2 for different values of parameters using the test image.

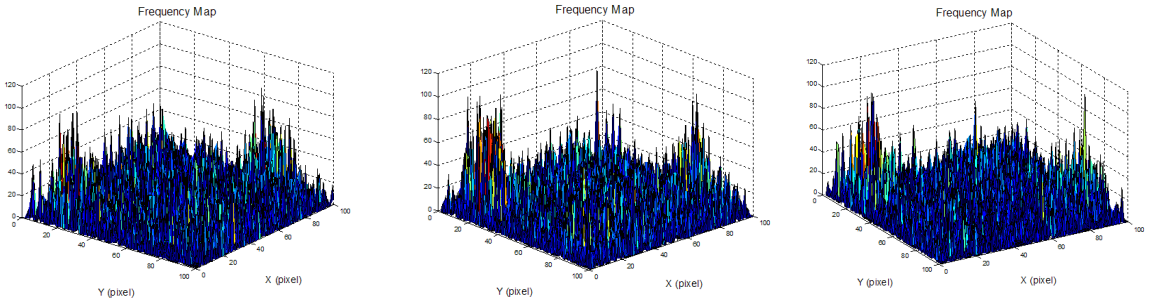


Figure 7: Frequency Maps for Φ3. Different frequency maps developed, using the function Φ3 for different values of parameters using the test image.

The best performance was found for φ(x,y) equal to saliency and the parameters D equal to 2 and a temperature of 1000.

Experiment and Results

Using the parameter selection results described above, we conducted an experiment in order to evaluate the Levy Flight-based attention selection algorithm while embedded in the iCub's multi-modal attention and sensory-motor framework, and to compare the performance against the iCub's existing winner-take-all saliency-driven attention selection mechanism. For the robot's visual environment, we have arranged a black background with up to 8 colored elements, as shown in Fig. 8. We start with 5 elements and incrementally expand the region to the left, by adding up to 3 elements For each group of elements, the robot had 5 different trials to explore the scene. During each trial, we compute three different measures of performance: minimum time to visit all objects, mean power spent by the algorithm, and the difference between the saliency map and the frequency map using the statistical measurement presented in Fig. 4.

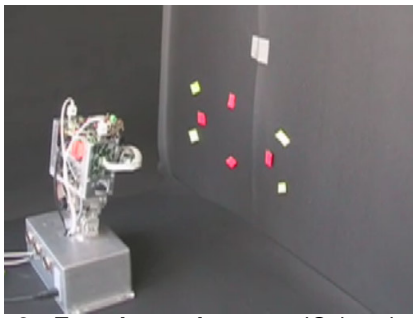
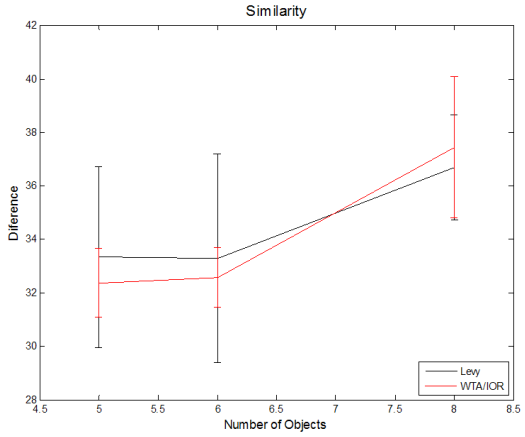


Figure 8: Experimental setup. iCub robot head with six degrees of freedom in the experimental environment.

In order to compare the relation between the saliency map and the frequency of attention, we computed an aggregated saliency map during each trial, which is basically the aggregation of the saliency map projection in the egosphere, multiplied by a fixed scale factor. Using this aggregated saliency map and statistical measurement described in Fig. 4, we compute three measures of performance, as described above. Fig. 9 illustrates the results of the similarity measure between the frequency map and the aggregated saliency map using the Levy flight-based and winner-take-all selection algorithm. Fig. 10 shows the amount of Mean Power employed by both algorithms. Lastly, Fig. 11 shows the minimum time needed for both algorithms to look at all objects in the environment.



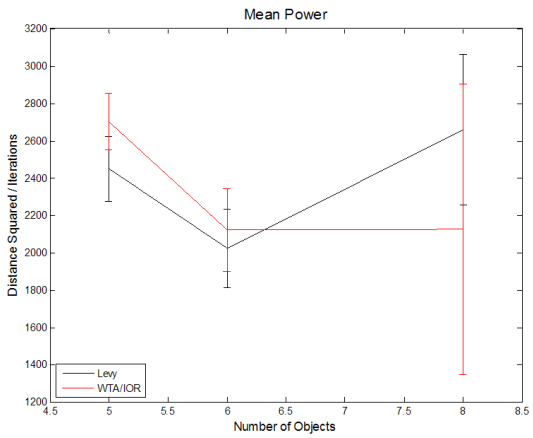


Figure 9. Similarity between the algorithms. Mean and standard deviation of the similarity measure between the frequency map and the aggregated saliency map for both algorithms winner take all with inhibition of return and the restricted Levy fly random walk.

Figure 10. Mean Power of the algorithms. Mean and standard deviation of the Mean Power employed for both algorithms winner take all with inhibition of return and the restricted Levy fly random walk.

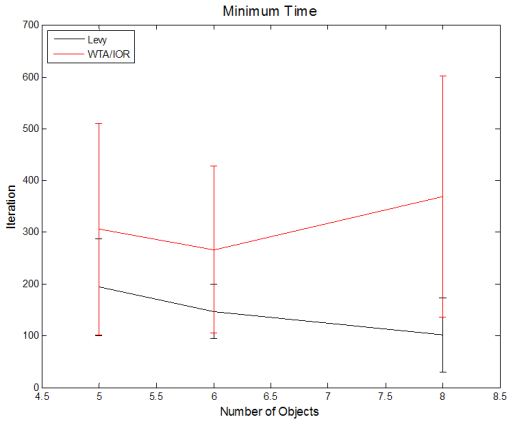


Figure 11: Minimum time of the algorithms. Mean and standard deviation of the Minimum time needed to look at all the objects in the environment for both algorithms winner take all with inhibition of return and the restricted Levy fly random walk.

These results demonstrate that the robot conducts a qualitatively faster exploration when using the stochastic Levy Flight-based attention selection mechanism compared to the standard winner-take-all saliency-driven attention mechanism, as expected from the result in (Brockmann et al., 2000), while keeping similar behavior in terms of the similarity with the saliency map and mean power consumption.

Conclusions and Future Work

This research presents a novel approach to drive the attention of the robot using the Levy fly random walk restricted to the saliency map, implemented on the iCub head that runs in real-time. This enforces the exploration of the environment but also keeps the attention of the robot in the salient points in the environment, without having to keep track of what has been looked at by the robot, This is one major difference between the novel approach and the standard algorithm employed to drive the attention of the attention systems.

For future work, it would interesting to explore how the attention system can be exploited to provide relevant information for an specific task like grasping. Furthermore, because the attention system in itself only delivers retino-centric coordinates, and therefore does not provide any 3D position information, we have begun the development of a real-time depth perception algorithm for the iCub. Depth information is required for efficient reaching and grasping. The implementation relies on a vergence-based real-time stereo-vision algorithm which includes an on-line calibration scheme (Martinez *et al.*, in preparation). We expect that compared to a static stereo-vision setup, our approach will lead to higher precision in depth estimation.

References

Ruesch, J., Lopes, M., Bernardino, A., Hoernstein, J., Santos-Victor, J. and Pfeifer, R. (2008). Multimodal saliency-based bottom-up attention: a framework for the humanoid robot iCub. *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 962-967.

Brockmann, D. and Geisel, T: (1999). Are human scanpaths Levy-flights? *Proc. of 9th Int. Conf. on Artificial Neural Networks*, 1: 263-268.

Brockmann, D. and Geisel, T. (2000). The ecology of gaze shifts. *Neurocomputation*, 32/33: 643.

Boccignone, G. and Ferraro, M. (2004). Modelling gaze shift as a constrained random walk. *Physic A,* 331: pp. 207-218

Martinez, H., Lungarella, M., and Pfeifer, R. (in preparation). Improving real-time depth perception through a vergence-based stereo-vision algorithm (Technical Report).