Proceedings Paper:

Lepora, N.F., Martinez-Hernandez, U. and Prescott, T.J. (2013) A SOLID case for active bayesian perception in robot touch. In: Biomimetic and Biohybrid Systems. Second International Conference, Living Machines 2013, July 29 – August 2, 2013, London, UK. Lecture Notes in Computer Science, 8064 . Springer Berlin Heidelberg , pp. 154-166. ISBN 9783642398018

https://doi.org/10.1007/978-3-642-39802-5-14

A SOLID case for active Bayesian perception in robot touch

Nathan F. Lepora[1,2], Uriel Martinez-Hernandez[1,3], and Tony J. Prescott[1,2]

[1] Sheffield Center for Robotics (SCentRo), University of Sheffield, UK.
[2] Department of Psychology, University of Sheffield, UK.
[3] Automatic Control and Systems Engineering (ACSE), University of Sheffield, UK.
{n.lepora,uriel.martinez,t.j.prescott}@sheffield.ac.uk

Abstract. In a series of papers, we have formalized a *Bayesian perception* approach for robotics based on recent progress in understanding animal perception. A main principle is to accumulate evidence for multiple perceptual alternatives until reaching a preset belief threshold, formally related to Bayesian sequential analysis methods for optimal decision making. Here we describe how this approach extends naturally to active perception, by moving the sensor with an active control strategy according to the accumulated beliefs during the decision making process. This approach can be seen as a method for solving problems involving Simultaneous Object Localization and IDentification (SOLID), or 'where' and 'what'. Considering an example in robot touch, we find that active perception gives an efficient and accurate solution to the SOLID problem, whereas passive perception was inaccurate and non-robust when the object location was uncertain. Thus, this general approach enables robust and accurate robot perception in unstructured environments.

1 Introduction

Twenty five years after Bajcsy's landmark paper on active perception [1], it remains the case that most machine perception involves static analysis of passively sampled data. Certainly, there has been progress on passive approaches to pattern recognition in relation to machine learning and uncertainty, and there is a diverse body of work on active vision; nevertheless, a search through recent progress in robot vision, audition or touch reveals the majority of papers still rely on wholly forward perceptual processes without any sensorimotor feedback.

Why this slow uptake, when early arguments for active control of perception were compelling [1, 2] and, as Bajcsy said, it should be axiomatic that perception is active? One factor is the required complexity of the robot hardware, which must involve actuated sensors and sensorimotor control loops. However, this should not be a barrier, since the technology is readily available and many standard robot platforms have these capabilities. A more likely explanation is that researchers have focussed on sensing problems, *e.g.* identification, that can be solved adequately in many scenarios without introducing active methods for sensorimotor control. That being said, conventional robotics is reaching an impasse

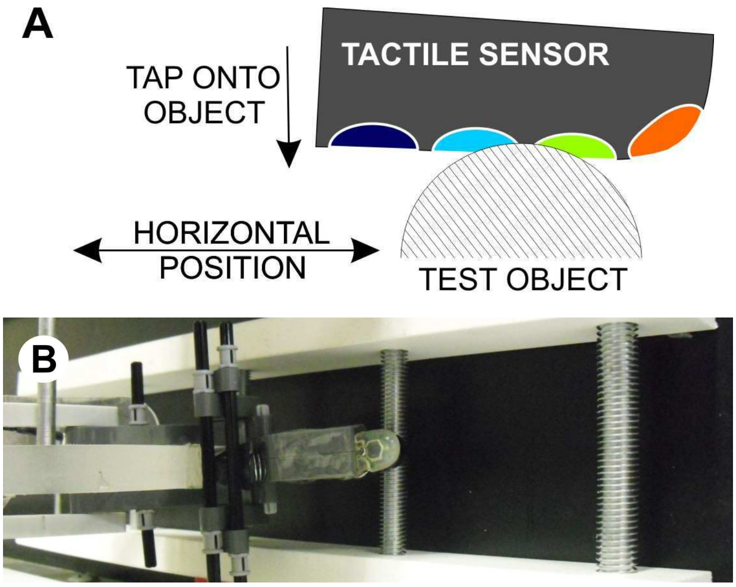

Fig. 1. Experimental setup. (A) Schematic of tactile sensor contacting a cylindrical test object. The fingertip is moved horizontally to sample object contacts from different positions. (B) Top-down view of experiment, with the fingertip mounting on the arm of the Cartesian robot visible to the left.

with present methods, such as poor performance in unstructured environments, which is preventing wider robot utilization beyond traditional factory settings.

In a series of papers [3–8], we have formalized an approach for robot perception based on recent progress in understanding animal perception [9, 10]. A main principle is to accumulate evidence for multiple perceptual alternatives until reaching a preset belief threshold that triggers a decision, formally related to Bayesian sequential analysis methods for optimal decision making [11]. Here we describe how this perception approach extends naturally from passive to active perception and some implications of this theory of active Bayesian perception.

Our proposal for active Bayesian perception is tested with a simple but illustrative task of perceiving the identity (diameter) and location (horizontal) position of a test rod using tapping movements of a biomimetic fingertip with unknown contact location (Fig. 1). We demonstrate first that passive perception can solve this task, but the perceptual acuity and reaction time depend strongly on the location of the fingertip relative to the rod. We then show that an active 'fixation point' control strategy can substantially improve the overall quality of the perception, by moving the fingertip to locations with good perception independent of the starting position. Thus, active perception gives far superior robustness, accuracy and speed to the decision making than passive methods.

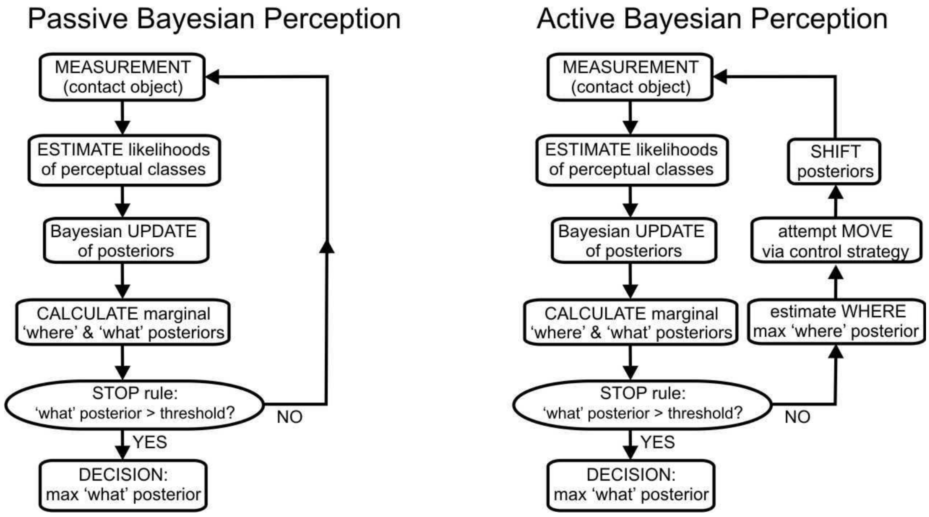

Fig. 2. Active and passive Bayesian perception applied to simultaneous object localization and identification. Passive Bayesian perception (left) has a recursive Bayesian update to give the marginal 'where' and 'what' perceptual beliefs, with decision termination at sufficient 'what' belief. Active Bayesian perception (right) has active control of the sensor location depending on the current beliefs (here just the 'where' beliefs).

2 Methods

2.1 Conceptual foundations

The main goal of this work is to advance our understanding of the role of active perception for situated agents in determining the 'where' and 'what' properties of objects. We refer to the computational task that must then be solved by Simultaneous Object Localization and IDentification (SOLID), to emphasize a similarity with SLAM of having two interdependent task aims, in that knowledge of localization aids the computation of identification (mapping) and similarly that knowledge of object identity (mapping) aids localization.

Passive Bayesian perception accumulates belief for distinct 'where' and 'what' classes by making successive taps against a test object until at least one of the marginal 'what' beliefs crosses a belief threshold, when a 'where' and 'what' decision is made. The passive nature of the perception means that the 'where' position class is constant over this process (Fig. 1A).

Active Bayesian perception also accumulates belief for the 'where' (horizontal position) and 'what' (cylinder diameter) perceptual classes by successively tapping until reaching a predefined 'what' belief threshold. In addition, it utilizes a sensorimotor loop to move the sensor according to the online marginal belief estimates during the perceptual process (Fig. 1B). For example, the sensor could be controlled with a 'fixation point' strategy, with the marginal 'where' beliefs used to infer a best estimate for current location and thus a relative move towards a preset target position on the object.

2.2 Algorithmic foundations

Our algorithm for active perception is based on including a sensorimotor feedback loop in an optimal decision making method for passive perception derived from Bayesian sequential analysis [3].

Measurement model and likelihood estimation: Each tap against a test object gives a multi-dimensional time series of sensor values across the K taxels (Fig. 3). The likelihood of a perceptual class $c_n \in C$ for a test tap z_t (with samples s_j) is evaluated with a measurement model [3, 4]

$$P(z_t|c_n) = \sqrt[JK]{\prod_{j=1}^{J} \prod_{k=1}^{K} P_k(s_j|c_n)}. \quad (1)$$

The sample distribution is determined off-line from the training data using a 'bag-of-samples' histogram method

$$P_k(s|c_n) = \frac{h_k(b(s))}{\sum_b h_k(b)}, \quad (2)$$

with $h_k(b)$ the occupation number of a bin (and $b(s) \ni s$), taking 100 bins across the full data range. We will use $K = 12$ taxels and $J = 50$ time samples per tap.

Bayesian update: Bayes' rule is used to recursively update the beliefs $P(c_n|z_t)$ for the N perceptual classes c_n with likelihoods $P(z_t|c_n)$ of the present tap z_t

$$P(c_n|z_t) = \frac{P(z_t|c_n)P(c_n|z_{t-1})}{P(z_t|z_{t-1})}. \quad (3)$$

The likelihoods $P(z_t|c_n)$ are assumed i.i.d. over time t (so $z_{1:t-1}$ drops out). The marginal probabilities are conditioned on the preceding tap and given by

$$P(z_t|z_{t-1}) = \sum_{n=1}^{N} P(z_t|c_n)P(c_n|z_{t-1}). \quad (4)$$

Iterating the update (3,4), a sequence of taps $z_1, \cdots, z_t$ gives a sequence of posteriors $P(c_n|z_1), \cdots, P(c_n|z_t)$ initialized from uniform priors $P(c_n) = P(c_n|z_0) = 1/N$. We will use $N = 80$ classes over 16 positions and 5 object curvatures.

Marginal 'where' and 'what' posteriors: The perceptual classes have L 'where' (position) and M 'what' (curvature) components, with each class c_n an (x_l, w_m) 'where-what' pair (*i.e.* $C = X \times W$). Then the beliefs over the individual 'where' and 'what' classes are found by marginalizing

$$P(x_l|z_t) = \sum_{m=1}^{M} P(x_l, w_m|z_t), \quad (5)$$

$$P(w_m|z_t) = \sum_{l=1}^{L} P(x_l, w_m|z_t), \quad (6)$$

with the 'where' beliefs summed over all 'what' classes and the 'what' beliefs over all 'where' perceptual classes. Here we use $L = 16$ position classes and $M = 5$ curvature classes.

Stopping condition on the 'what' posteriors: Following methods for passive Bayesian perception using sequential analysis [3], a threshold crossing rule on the marginal 'what' posterior triggers the final 'what' decision, given by the maximal *a posteriori* (MAP) estimate

$$\text{if any } P(w_m|z_t) > \theta_W \text{ then } w_{\text{MAP}} = \arg\max_{w_m \in W} P(W|z_t). \tag{7}$$

This belief threshold θ_W is a free parameter that adjusts the balance between decision speed and accuracy.

Move decision on the 'where' posteriors: Analogously to the stop decision, a sensor move requires a marginal 'where' posterior to cross its decision threshold, with a MAP estimate used for the 'where' decision

$$\text{if any } P(x_l|z_t) > \theta_{\text{X}} \quad \text{then} \quad x_{\text{MAP}} = \arg\max_{x_l \in X} P(X|z_t). \tag{8}$$

Here we consider two particular cases, termed

$$\begin{aligned} \text{passive perception}: \ \theta_{\text{X}} = 1 \qquad & \text{(never moves)} \\ \text{active perception}: \ \theta_{\text{X}} = 0 \qquad & \text{(always tries to move).} \end{aligned}$$

There are many possible strategies to control the movement Δ depending on the task. Whatever the strategy, the 'where' posteriors should be kept aligned with the sensor by shifting the joint 'where-what' posteriors with each move

$$P(x_l, w_m|z_t) = P(x_l - \Delta, w_m|z_t). \tag{9}$$

For simplicity, we recalculate the posteriors lying outside the original range by assuming they are uniformly distributed.

'Fixation point' active perception strategy: Here we consider a control strategy with fixation point x_{fixed} that the sensor attempts to move to. Then the appropriate move Δ is found from the 'where' decision x_{MAP} of sensor location

$$x \to x + \Delta\left(x_{\text{MAP}}\right), \quad \Delta(x_{\text{MAP}}) = x_{\text{fixed}} - x_{\text{MAP}}. \tag{10}$$

This move Δ defines both the motor command to the sensor and the shift (9) applied to the posteriors.

2.3 Data collection

The tactile sensors used in this study have a rounded shape that resembles a human fingertip [12], of dimensions 14.5 mm long by 13 mm wide. They consist of an inner support wrapped with a flexible printed circuit board (PCB) containing 12 conductive patches for the touch sensor 'taxels'. These are coated with PCB

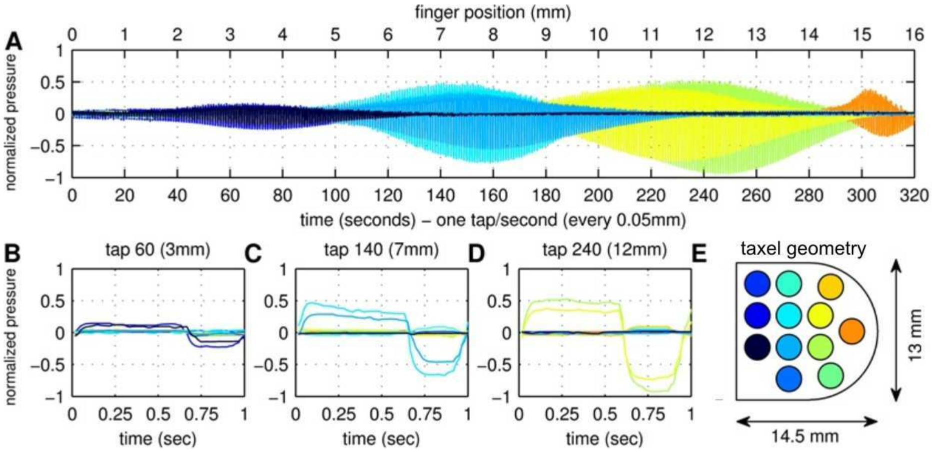

Fig. 3. Fingertip pressure data recorded as the finger taps against a test rod (diameter 4 mm) at a constant rate of 1 tap/sec. The range of finger positions spanned 16 mm over 320 s, giving 320 taps spaced every 0.05 mm. Tickmarks are shown every 1 mm displacement, or 20 taps. Data from the different taxels are represented in distinct colors depending on the taxel position shown on the diagram to the right.

and silicone layers that together comprise a capacitive touch sensor to detect pressure via compression. Data was collected at 50 samples per second with 256 vales, and then normalized and high-pass filtered before analysis [12].

The present experiments test the capabilities of the tactile fingertip mounted on an xy-positioning robot. This robot can move the sensor over a horizontal plane in a highly controlled and repeatable manner onto various test stimuli ($\sim$50 μm accuracy), and has been used for testing various tactile sensors [13]. The fingertip was mounted at an angle appropriate for contacting axially symmetric shapes such as cylinders aligned along the z-axis perpendicular to the plane of movement (Fig. 1A). Five steel rods with diameters 4 mm, 6 mm, 8 mm, 10 mm and 12 mm were used as test objects (Fig. 1B). They were mounted with their axes vertically upwards but their centers offset in the y-direction (by 4 mm, 3 mm, 2 mm, 1 mm and 0 mm) to align their closest point to the fingertip in the direction of tapping.

The touch data were collected while having the fingertip repeatedly tap in the y-direction onto and off each test object with rate 1 tap/sec, while moving at constant speed of 0.05 mm/sec in the x-direction across the closest face of the object. The fingertip was angled so the rod axis lay across the fingertip (down the taxels in Fig. 3), and moved so that the rod initially contacted the fingertip at its base and finally contacted only the tip. In each case, an x-range of 16 mm was considered. This gave 320 taps per object at increments of 0.05 mm. Each tap of the fingertip against the object resulted in a 1 sec time series of pressure readings for all 12 taxels covering the fingertip (Figs 2B-E). This data was sampled at

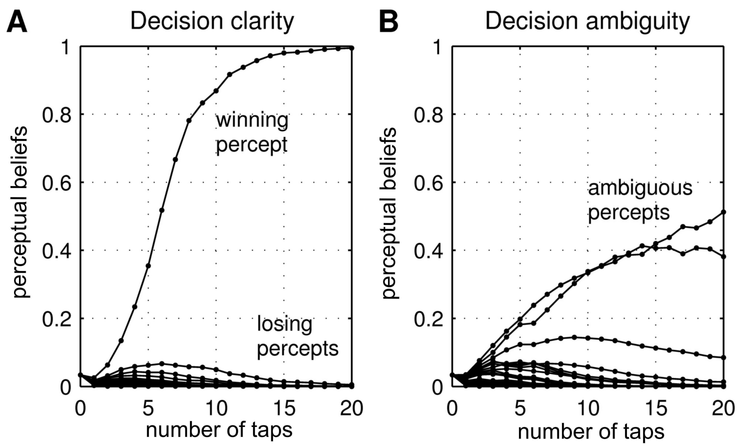

Fig. 4. Examples of the accumulating/depreciating beliefs for the (80) distinct 'where' and 'what' percept classes as data from more taps is included. Panel (A) shows an example with a clear winning percept and panel (B) of ambiguity.

50 Hz, giving 50 samples per taxel per tap. Distinct training and test sets were collected for all 5 cylinder diameters.

3 Results

3.1 Evidence accumulation for robot perception

The 'where' and 'what' perceptual task is to identify the diameter of the rod being sensed and the location of the contact using tactile fingertip data over a sequence of test taps. Examples of the perceptual beliefs derived from tap sequences from two different test cases are shown in Fig. 4, which plots the beliefs for each putative percept against the number of taps. These probabilities begin at equality corresponding to uniform priors and then evolve smoothly with some rising gradually towards unity and others falling towards zero. In the first example (Fig. 4A), the decision given by the largest perceptual belief remains the same after applying 2 taps or more, while the second example (Fig. 4B) flips between the two leading choices.

There are two common methods for making decisions from sequential data of this type: (i) set in advance the number of taps that will be used, or (ii) set in advance a belief threshold that will trigger the decision, so that the reaction time is a dynamic quantity that depends on the data received. Recent progress in perceptual neuroscience strongly supports that animals use a belief threshold to make decisions [9], which is related to sequential analysis methods for optimal decision making [11]. Accordingly, a comparison of these two methods on tactile robot data found that the belief threshold method gave superior performance in perceptual acuity [3]. This can be seen intuitively from Fig. 4: if for example 10

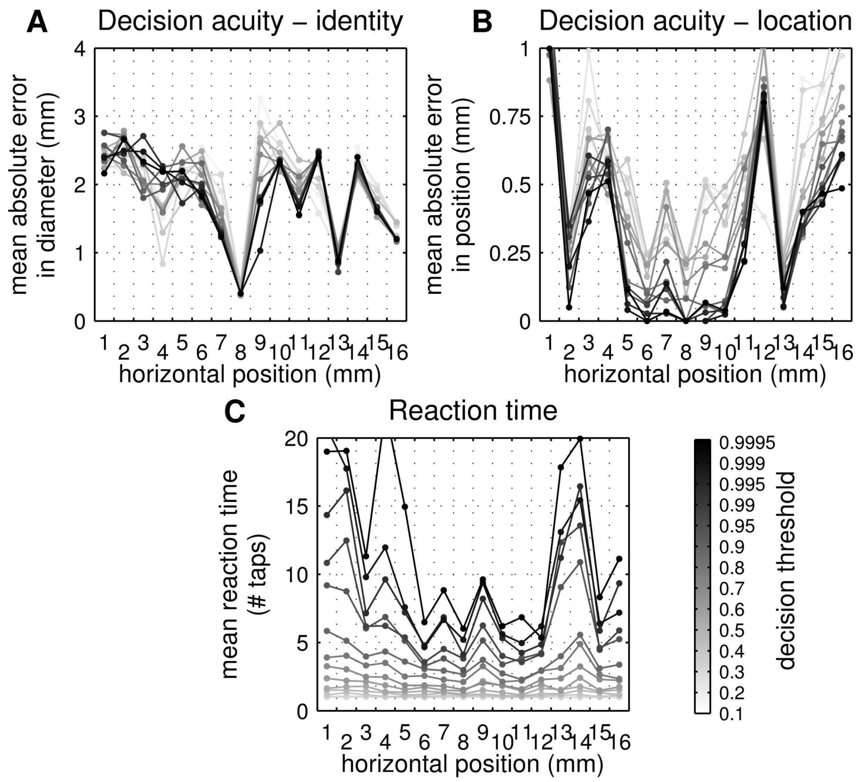

Fig. 5. Active perception results depends on the decision threshold and fixation point. The mean accuracy of identifying the cylinder (A) and the mean reaction time (B) vary with threshold (gray-shade of plot) and fixation point (x-axis). Each data point corresponds to 1000 decision trials.

taps were set in advance, then the decision is unnecessarily slow in situations of clarity (Fig. 4A) and too quick in situations of ambiguity (Fig. 4B). Instead, setting a belief threshold allows the decision time to adjust dynamically to the uncertainty of the situation.

3.2 Passive perception of where and what

This section considers the application of passive Bayesian perception to the 'where' and 'what' perceptual task of identifying the rod diameter and fingertip location relative to the rod. Results are generated using a Monte Carlo method averaged over distinct decision trials (1000 iterations per location and threshold), each with test taps drawn from one 'where' and 'what' class consistent with the sensor not being able to change position during perception.

Average decision errors and reaction times for perceiving shape and horizontal position were examined over belief thresholds ranging from 0.1 to 0.995

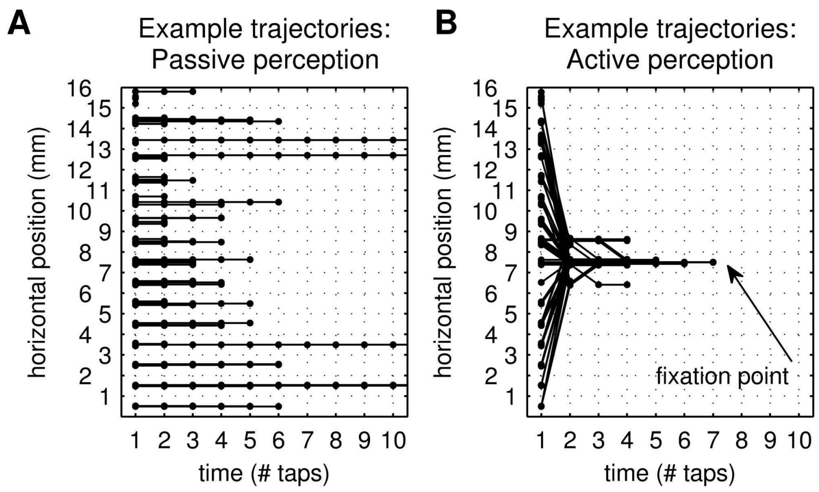

Fig. 6. Example trajectories for passive and active perception. (A) Passive perception, with horizontal position fixed against time. (B) Active perception, with trajectories converging on the fixation point (8 mm) independent of starting position. 100 trajectories were selected randomly for each case.

(Fig. 5). Statistically robust estimates of the mean errors were found by averaging the absolute classification errors over many test instances. These mean errors decreased steadily with belief threshold, reaching a minimum of about 0.4 mm for rod diameter (identity) and 0.1 mm for rod position for the largest thresholds (Figs 5A,B; black curves) The number of taps to reach a decision had a reaction time distribution, such that increasing the belief threshold increased the mean reaction time (Fig. 5C; black curve).

Clearly, the best perceptual acuities and reaction times are in the center of the range of fingertip positions (Fig. 5, position class at 8 mm). However, in passive perception, there is no way to modify the position from where an object is sensed. Hence, average values across all possible sensing positions give a typical perceptual acuity: these are shown in red on Fig. 7, with the overall mean acuities for rod diameter and horizontal position of 2 mm and 0.2 mm respectively.

3.3 Active perception of where and what

This section considers active Bayesian perception in the same scenario as for passive perception above, with a 'where' and 'what' perceptual task of identifying the rod diameter and fingertip location relative to the rod. Results are again generated using a Monte Carlo method (1000 decision trials), now with an active control strategy that aims for a fixation point in the center of the horizontal range (8 mm), as visible in the trajectories for active perception (Fig. 6).

Average decision errors and reaction times for perceiving shape and horizontal position were examined over the same range of 'what' belief thresholds

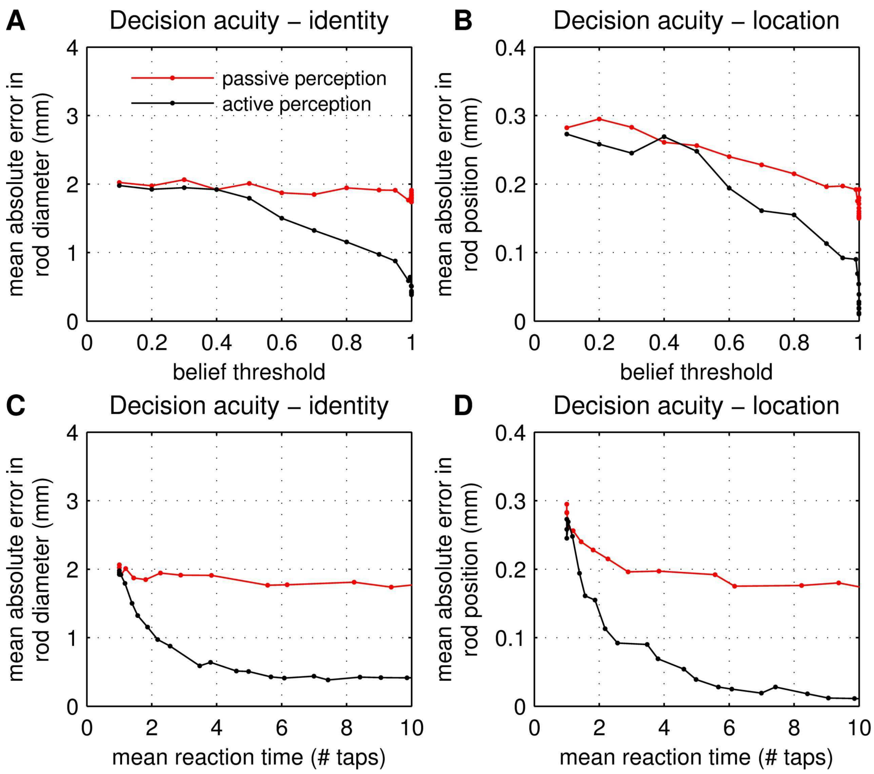

Fig. 7. Decision acuity for active and passive perception. (A,B) Dependence of the mean absolute errors of rod diameter and position plotted against the 'what' belief threshold. Passive perception is shown in red and active perception in black. (C,D) Dependence of these perceptual acuities instead plotted against mean reaction time (with belief threshold an implicit parameter determining all quantities).

as passive perception (0.1 to 0.995), permitting direct comparison of the active and passive approaches (Fig. 7). Once again, the mean absolute errors decreased steadily with belief threshold (black plots), reaching a minimum of about 0.4 mm for rod diameter (identity) and 0.01 mm for rod position for the largest thresholds. The number of taps to reach a decision had a reaction time distribution, with mean reaction time increasing with 'what' decision threshold; treating this threshold as an implicit parameter gave a direct plot of decision error against mean reaction time (Fig. 7C,D).

Comparing active perception with passive perception, the best mean absolute errors improve from 2 mm to 0.4 mm for rod diameter (*cf.* 4-12 mm range) and from 0.2 mm to 0.01 mm for horizontal position (*cf.* 0-16 mm range). Evidently, active perception gives the finest perceptual acuity compared with the passive method when compared at both similar reaction times and belief thresholds.

4 Discussion

The aim of this study is to demonstrate that active perception gives substantial benefits for robot perception over passive methods. We proposed an algorithm for active Bayesian perception that accumulates evidence until reaching a decision threshold while a sensorimotor feedback loop moves the sensor to a 'good' fixation point relative to the perceived object. This algorithm contrasts with standard 'passive' methods for robot perception that lack this feedback loop. We then compared active and passive perception on a simple but illustrative task of simultaneous object localization and identification ('where' and 'what') with a biomimetic fingertip, in which both the position and diameter of a rod is perceived with discrete tapping movements.

The main findings are:
(i) The 'where' and 'what' perceptual acuities for both passive and active perception improved with increasing belief threshold. Accordingly, the mean reaction time also increased with belief threshold, corresponding to more data being collected to reach sufficient belief to make a decision.
(ii) Passive perception has an issue that that not all sensing locations are equal: those locations for which the center of the fingertip contacts the rod gave better perception than at the extremities. Thus, passive perception is non-robust.
(iii) Active perception was able to give robust perception whatever the initial contact location of the fingertip against the rod by utilizing a sensorimotor loop to compensate uncertainty in initial sensor placement. Thus, active perception can be robust in unstructured environments [7].
(iv) In consequence, active perception gave an order-of-magnitude improvement in perceptual acuity over the passive method, with best mean errors improving from 2mm to 0.4mm for rod diameter (cf. 4-12mm range) and from 0.2mm to 0.01mm for horizontal position (cf. 0-16mm range).

5 Conclusion

In this paper, we demonstrated that active Bayesian perception with a sensorimotor control loop between the perceptual beliefs and the motion of the sensor can robustly and accurately solve problems of Simultaneous Object Localization and IDentification (SOLID), or 'where' and 'what' objects are in the world. Whereas active perception gave an efficient and accurate solution to the SOLID problem in unstructured environments, passive perception could be inaccurate and non-robust under uncertainty about object location.

In seminal work on active perception, Bajcsy said it is axiomatic that perception (in animals) is active [1]. Robotics is currently in a state of transition from rigidly controlled tasks in predictable structured environments like factory assembly lines, to applications in unpredictable unstructured environments like our homes, hospitals and workplaces. In our opinion, future robots will need active perception to accomplish these tasks in unstructured environments, and thus it may also become axiomatic that robot perception is active too.

Acknowledgments. This work was supported by the FP7 project 'Experimental Functional Android Assistant (EFAA)' (ICT-270490); UMH was supported by the Mexican National Council of Science and Technology (CONACyT).

References

1. R. Bajcsy. Active perception. *Procs of the IEEE*, 76:966–1005, 1988.
2. D. Ballard. Animate vision. *Artificial intelligence*, 48(1):57–86, 1991.
3. N.F. Lepora, C.W. Fox, M.H. Evans, M.E. Diamond, K. Gurney, and T.J. Prescott. Optimal decision-making in mammals: insights from a robot study of rodent texture discrimination. *Journal of The Royal Society Interface*, 9(72):1517–1528, 2012.
4. N.F. Lepora, M. Evans, C.W. Fox, M.E. Diamond, K. Gurney, and T.J. Prescott. Naive bayes texture classification applied to whisker data from a moving robot. *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–8, 2010.
5. N.F. Lepora, J.C. Sullivan, B. Mitchinson, M. Pearson, K. Gurney, and T.J. Prescott. Brain-inspired bayesian perception for biomimetic robot touch. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 5111 –5116, 2012.
6. N.F. Lepora, U. Martinez-Hernandez, H. Barron-Gonzalez, M. Evans, G. Metta, and T.J. Prescott. Embodied hyperacuity from bayesian perception: Shape and position discrimination with an icub fingertip sensor. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 4638–4643, 2012.
7. N.F. Lepora, U. Martinez-Hernandez, and T.J. Prescott. Active touch for robust perception under position uncertainty. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, 2013.
8. N.F. Lepora, U. Martinez-Hernandez, and T.J. Prescott. Active bayesian perception for simultaneous object localization and identification. (under review).
9. J.I. Gold and M.N. Shadlen. The neural basis of decision making. *Annual Reviews Neuroscience*, 30:535–574, 2007.
10. N.F. Lepora and K. Gurney. The basal ganglia optimize decision making over general perceptual hypotheses. *Neural Computation*, (24):2924–2945, 2012.
11. A. Wald. *Sequential analysis*. John Wiley and Sons (NY), 1947.
12. A. Schmitz, P. Maiolino, M. Maggiali, L. Natale, G. Cannata, and G. Metta. Methods and technologies for the implementation of large-scale robot tactile sensors. *Robotics, IEEE Transactions on*, 27(3):389 –400, 2011.
13. M. Evans, C.W. Fox, M. Pearson, N.F. Lepora, and T.J. Prescott. Whisker-object contact speed affects radial distance estimation. In *Robotics and Biomimetics (ROBIO), 2010 IEEE International Conference on*, pages 720–725, 2010.