Active touch for robust perception under position uncertainty

Nathan F. Lepora, Member, IEEE, Uriel Martinez-Hernandez, Student Member, IEEE, Tony J. Prescott

Abstract—In this paper, we propose that active perception will help attain autonomous robotics in unstructured environments by giving robust perception. We test this claim with a biomimetic fingertip that senses surface texture under a range of contact depths. We compare the performance of passive Bayesian perception with a novel approach for active perception that includes a sensorimotor loop for controlling sensor position. Passive perception at a single depth gave poor results, with just 0.2 mm uncertainty impairing performance. Extending passive perception over a range of depths gave non-robust performance. Only active perception could give robust, accurate performance, with the sensorimotor feedback compensating the position uncertainty. We expect that these results will extend to other stimuli, so that active perception will offer a general approach to robust perception in unstructured environments.

I. INTRODUCTION

The dream of robots that can work alongside or replace people in unstructured environments has long evaded researchers in artificial intelligence [1]. While robots are successful at tasks where they can be rigidly controlled for predictable structured environments like factory assembly lines, they have failed to make significant impact in unpredictable unstructured environments like our homes, hospitals and workplaces. Solving this problem will revolutionize the use of robotics in society, with radical implications for automatization in the home and industry.

The main proposal in this paper is that active perception is a key ingredient for attaining autonomous robotics in unstructured environments. We need only look to humans and animals to see the effectiveness of active perception, with our extensive use of feedback between perception and movement of the sensory organs. Stated succinctly: ‘we do not just see, we look’ and ‘we do not only touch, we feel’ [2].

In active touch [3], we use our fingertips to stroke or tap to perceive texture, trace edges to judge shape, and press to determine compliance. Furthermore, as shown in this study, in unstructured environments there needs to be continual feedback between the perceptual process and the control of movement, as too light/strong a tap, stroke or press will impair recognition of the desired tactile property.

To implement active perception in an unstructured environment, we extend a recent bio-inspired approach for robot perception based on Bayesian sequential analysis [4], [5], [6], [7]. Previous versions of this framework have focussed on a feed-forward process for passive perception in which the robot position is independent of the sensed data. Here we introduce a sensorimotor feedback loop that attempts to place the sensor in a preset ‘good’ target position relative to the perceived object, analogously to heuristics for active touch that regulate contact force [8]. In consequence, we find the active perception compensates the object-sensor positioning uncertainty while improving the perceptual decision making.

The aim of this study is to uncover some general principles that are useful for achieving robust perception in unstructured environments. To do so, we choose a simple but illustrative task of perceiving texture with tapping movements of a biomimetic fingertip [9] when its initial depth relative to the surface is random and unknown. As far as we know, all previous robot studies of texture classification have structured the environment to allow the training and test regimes to have the same contact depth [4], [10], [11], [12], [13], [14], [15], [16], as is also common for other percepts such as shape, location and compliance. However, we find that a depth change of just 0.2 mm between the sensor and object seriously impairs tactile perception of the texture class. Moreover, merely extending the training to classify over these changes in sensor depth is not sufficient to give robust perception, because some depths are worse than others for texture classification.

Thus, we demonstrate that only for active perception, with the sensory data feeding back to move the sensor, can the position uncertainty inherent in an unstructured environment be compensated to attain robust perception.
II. METHODS

A. Data collection

The present experiments use a fingertip sensor designed for the iCub robot [17], mounted instead on a cartesian robot (Fig. 1A) [5]. The cartesian robot (2-axis PXYx, Yamaha Robotics) can move the fingertip in a plane with highly accurate and reproducible positioning (accuracy $\sim50\ \mu m$ under open-loop current control). We mounted the sensor at a fixed angle relative to the planar test surfaces (Fig 1), with tapping movements along the surface normal. The fingertip has an inner support wrapped in a flexible printed circuit board (PCB) having 12 conductive patches for the sensor ‘taxels’ [9], [18], with geometry shown in Fig. 3. This PCB is covered first with a $\sim2\ mm$ layer of non-conductive soft silicone foam and then with a thin layer of conductive silicone rubber. The PCB and silicone layers together comprise a capacitive touch sensor that detects pressure via compressing the non-conductive foam between the two conductive layers.

The touch data were collected while having the fingertip tap briefly (0.1 sec) and periodically against a planar surface, upon which were attached 10 different 25 mm by 30 mm textured patches from a range of grades of silicon carbide abrasive papers (Table I). By positioning the fingertip along the plane, the robot could tap against each texture in turn, with sufficient pause between each to allow transients to decay (examples in Fig. 2). By positioning the fingertip perpendicular to the plane, a range of contact depths (or equivalently contact pressures) could be sampled, for which we used 150 taps for each texture over a 3 mm range of contact depths from no to strong contact (examples in Fig. 3). One training and one test set was collected for each of the 10 textures. Each set was then separated into 15 distinct position classes at 0.2 mm intervals (tick-marks on Fig. 3A).
Active Bayesian Perception

![Diagram](image)

**Fig. 4.** Algorithm for active Bayesian perception. After each tap, the likelihoods of the ‘what’ (texture) and ‘where’ (depth) perceptual classes are used to sequentially update the posteriors. The marginal ‘where’ and ‘what’ beliefs are then used to make two decisions: (i) if a ‘what’ belief crosses its threshold, ‘what is the texture?’ is decided, and the process stops; (ii) if a ‘where’ belief crosses its threshold, ‘where is the sensor?’ is decided, to determine a sensor move and compensatory shift of the posteriors.

### B. Active Bayesian Perception

We use a statistical method based on Bayesian sequential analysis that is related to leading models of perceptual decision making in neuroscience and psychology, and has been applied successfully to robot perception [4], [5], [6]. Previous implementations have concentrated on passive perception with no feedback between the sensory data and sensor position. Here we present an algorithm for active perception based on the sensorimotor feedback loop shown in Fig. 4. The details of this algorithm are explained below.

**Measurement model and likelihood estimation:** Each tap against a test object gives a multi-dimensional time series of sensor values across the $K = 12$ taxels (Figs 2, 3). The likelihood of a perceptual class $c_n \in C$ for a test tap $z_t$ is evaluated with a measurement model estimated off-line from the histogram of sensor values over the training data for that class [19]. For a sample $s$,

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

(1)

with $h_k(b)$ the occupation number of bin $b$ (where $b(s) \supseteq s$) from the $k$th dimension of the training data, with 100 bins. Applied to a test tap, this measurement model gives the likelihoods evaluated over all samples $s_j$ in that tap

$$P(z_t|c_n) = \frac{\prod_{j=1}^{J} \prod_{k=1}^{K} P_k(s_j|c_n)}{\sqrt{\prod_{j=1}^{J} \prod_{k=1}^{K} P_k(s_j|c_n)}},$$

(2)

where $J = 50$ and $K = 12$ are the time samples per tap and the number of taxels respectively. This model assumes a bag of measurements in which all samples are treated as independent and identically distributed for each taxel. The geometric mean prevents the product (2) from producing vanishingly small likelihoods by ensuring that the probabilities remain almost invariant to sample number (and approximates a combinatoric factor for reordering the samples [4]).

**Bayesian update:** Bayes’ rule is used to update the posterior probabilities (beliefs) $P(c_n|z_t)$ for the $N$ perceptual classes $c_n$ with the likelihoods $P(z_t|c_n)$ of the present tap $z_t$. The prior is from the posterior probability at the previous tap, giving a sequential update

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

(3)

The likelihoods $P(z_t|c_n)$ are assumed identically distributed and independent over time $t$ (so $z_{t-1}$ drops from the posteriors). The marginal probabilities are conditioned on the preceding tap and calculated by summing

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

(4)

to give properly normalized posteriors $\sum_{n=1}^{N} P(c_n|z_t) = 1$.

Taking a sequence of measurements $z_1, \cdots, z_t$ gives a sequence of posteriors $P(c_n|z_1), \cdots, P(c_n|z_t)$ for each class, which are calculated by iterating over the relations (3,4) starting from uniform priors $P(c_n) = P(c_n|z_0) = 1/N$.

**Marginal ‘where’ and ‘what’ posteriors:** For active perception, we suppose the perceptual classes have $L$ ‘where’ (i.e. depth) and $M$ ‘what’ (i.e. texture) components, such that each class $c_n$ corresponds to one $(x_l, w_m)$ ‘where-what’ pair. Then the posteriors are joint distributions over these joint classes, such that the beliefs over the individual ‘where’ and ‘what’ perceptual classes are given by the marginal posteriors

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

(5)

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

(6)

with the ‘where’ beliefs summed over all ‘what’ classes and the ‘what’ beliefs over all ‘where’ perceptual classes.

**Stop decision on the ‘what’ posteriors:** Following the methods for sequential analysis [4], a threshold crossing rule

<table>
<thead>
<tr>
<th>Texture: silicon carbide abrasive paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO grit designation</td>
</tr>
<tr>
<td>Description</td>
</tr>
<tr>
<td>Mean particle size</td>
</tr>
</tbody>
</table>

**Table I**

**Textures used in experiment:** 10 grades of silicon carbide abrasive paper (Wetordy, 3M-Trimite).
on the marginal ‘what’ posterior is used to decide when to stop gathering sensory data and make a final decision about the ‘what’ (texture) class. The maximal a posteriori (MAP) estimate is then used to give the ‘what’ decision

\[
\text{if } P(w_m | z_t) > \theta_W \text{ then } w_{MAP} = \arg \max_{w_m \in W} P(W | z_t).
\]  

(7)

As emphasized in previous robot studies [4], [5], [6], this decision rule can optimize the tradeoff between reaction speed and error rates (here texture classes), by minimizing the costs of delaying decisions and making mistakes.

**Move decision on the ‘where’ posteriors:** Analogously to the stop decision, a sensor move requires a marginal ‘where’ posterior to cross its decision threshold. Then the MAP estimate is used for the ‘where’ (depth) decision

\[
\text{if } P(x_l | z_t) > \theta_X \text{ then } x_{MAP} = \arg \max_{x_l \in X} P(X | z_t).
\]  

(8)

We suppose there is a preset target position \(x_{\text{target}}\) that the sensor attempts to move to, with the move \(\Delta\) determined from the ‘where’ decision \(x_{MAP}\) of sensor location

\[
x \rightarrow x + \Delta(x_{MAP}), \quad \Delta(x_{MAP}) = x_{\text{target}} - x_{MAP}.
\]  

(9)

To keep the ‘where’ posteriors aligned with the sensor, it is necessary to shift these probabilities with each move

\[
P(x_l, w_m | z_t) = P(x_l - \Delta(x_{MAP}), w_m | z_t).
\]  

(10)

For simplicity, we recalculate the posteriors lying outside the original range by assuming they are uniformly distributed. Overall, this active perception strategy tries to reposition the sensor to a previously determined ‘good’ location for perception relative to the object being sensed.

### III. RESULTS

#### A. Passive tactile perception in a structured environment

We begin by considering passive Bayesian perception when the fingertip sensor remains in a narrow depth range while tapping the textures. A 0.2 mm depth range of testing and training data is considered at a mean depth of 2.1 mm, comprising data similar to the examples in Fig. 2. Because the depth of the fingertip relative to the test object is finely tuned in advance, we consider robot perception on this data to be within a structured environment.

Performance of passive Bayesian perception in this structured environment is assessed with testing data within the same narrow depth range as the training data. For statistical robustness, a Monte Carlo method is used to draw random sequences of test taps from the test data for each texture class (with 2000 Monte Carlo iterations). For classification, we use the algorithm in Fig. 4, but disregard the move rule because there is only one depth class. This procedure sequentially updates the posteriors for the texture classes until crossing a decision threshold (here 0.5 belief), as in past implementations of passive Bayesian perception [4], [5], [6].

The result of this assessment of Bayesian perception is that individual textures can be discriminated to a mean absolute error of 0.4 over the 10 class range (Fig. 5; central dot). This performance is with a belief threshold of 0.5, giving a mean reaction time of 5 taps. Referring to Fig. 2, one obvious discriminant of texture is the peak pressure of a tap.

Physically, this peak pressure has two origins: the texture thickness increases with mean particle size (Table I), to give greater contact forces; and larger particle sizes also reduce the surface area of contact, increasing pressure.

#### B. An unstructured environment degrades tactile perception

The effect of an unstructured environment on tactile perception is now assessed by modifying the above experiment (Sec. III-A). We suppose that the testing data have a different depth range from the training data, due to uncertainties in the relative sensor-object position in an unstructured environment. Other decision parameters are kept constant.

Under this systematic change of sensor depth, the passive tactile perception of texture degrades as the depth offset between training and testing increases (Fig. 5). As the offset increases from 0 mm to 0.2 mm, the mean absolute error increases from 0.4 to ~1 class. Further increases in offset up to ±0.6 mm result in greater errors, after which the errors are close to a chance level of 3.2 texture classes. Physically, this impairment in performance is not unexpected, because the most obvious discriminant of texture is peak taxel pressure, which depends strongly on contact depth (Fig. 3).

#### C. Using position classes can improve passive perception but not robustly

We now attempt to solve the above performance issue for tactile perception in an unstructured environment by introducing classes for the unknown position between the sensor and object. The robot is trained over several depth classes that together span a broad range of contact depths,
Fig. 6. (A) Dependence of texture classification errors on both texture and depth (rather than the single depth of Fig. 5). (B) Mean absolute errors averaged over texture to show the dependence on depth. Decisions are for 0.5 belief threshold. These results describe classification in an unstructured environment, because the relative sensor-object depth is uncertain.

rather than judging texture from an incorrect assumption of a single contact depth as in Sec. III-B.

The same protocol for passive Bayesian perception is used as in Sec. III-A, but with training data over 10 texture and 15 depth classes of 0.2 mm range each. Example training data over depth is shown in Fig. 3A, with the tickmarks denoting the delineation between depth classes. Note that successful classification over these 150 perceptual classes requires discriminants other than the peak pressure in the time series of taxel readings, otherwise ambiguity between variations in texture and depth would give poor perception.

Results for passive perception of texture are shown in Fig. 6A for the mean texture classification error over each of the 150 test classes. The performance seems to improve with contact depth, as is confirmed by plotting the mean absolute error over all textures for each depth class (Fig. 6B). Mean texture classification errors decrease from ~3 classes for depths 0–1 mm to ~0.5 classes for depths 1.5–3 mm. Thus, increasing contact depth improves perception, presumably because of a better signal-to-noise ratio. However, in an unstructured environment, this contact depth is not under the control of the robot, but is determined randomly by the relative sensor-object position. Therefore, consistent performance with passive Bayesian perception would not be achieved in an unstructured environment.

D. Active perception can give robust performance in an unstructured environment

Finally, we use active perception to address the tactile classification of texture in an unstructured environment. The active perception algorithm has a sensorimotor feedback loop (Fig. 4) that attempts to move the tactile sensor to a good position for perception while deciding upon the texture class. The classification performance is assessed with same data and texture/depth classes as in Sec. III-C.

The active perception algorithm requires a target position for attempting a move to, which we set as 2.1 mm depth (the 11th depth class). From Fig. 6, this position seems good for passive texture perception without being too close to the outer range of training data. A belief threshold of 0.5 was used for the stop rule, and three belief thresholds compared for the move rule: >1 (passive perception), 0.999 (weakly active perception) and 0.05 (strongly active perception). Movements become harder to initiate for belief thresholds closer to unity, which we describe as the active perception becoming weaker then passive above unit threshold.

The principal effect of active perception is that texture classification is greatly improved over passive perception when averaged over all starting depths (Fig. 7). In particular, strongly active perception achieves a mean absolute error of 0.4 classes over all textures and all starting positions, which is substantially better than the 1.4 value for passive perception. It is also interesting that as the active perception becomes stronger, with the movement belief threshold closer to zero, the texture perception improves (Figs 7B,C). This result is not obvious in advance, and indicates that active perception favoring quickly responsive but inaccurate repositioning is superior to an accurate but slow movement strategy.
IV. CONCLUSIONS

The goal of this study is to demonstrate that active perception can help attain autonomous robotics in unstructured environments. We propose 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’ position relative to the perceived object. This algorithm contrasts with standard ‘passive’ methods for robot perception that lack this feedback. We apply active perception to a simple but illustrative task of texture discrimination with a biomimetic fingertip when the test depth of the sensor is random and unknown relative to the training depth. The main results are: (i) for a single position, a small difference between testing and training (of just 0.2 mm) seriously degrades the passive perception of texture; (ii) extending the training over a range of contact depths gives non-robust passive perception, because some depths are inherently poorer than others; (iii) active perception can give robust, accurate performance as the sensorimotor feedback loop compensates the uncertainty from the environment.

These results should extend to other types of robot perception; for example, the tactile perception of curvature [6], [20], geometry [21], [22], compliance [23] and flexibility [24]. Robot experiments testing the perception of these properties typically choose an environment in which the relative pose of the sensor and object can be maintained between training and testing. However, the tactile data for these perceptual properties will depend on the relative pose, which outside the laboratory may not be predetermined. And as described above, even if the object pose varies and is perceived, not all poses are good for perception. These issues could underly anecdotal reports that artificial tactile perception can sometimes be unreliable in demos outside the laboratory, since even small disturbances to a robot could spoil its training. Our expectation is that appropriate use of active touch will give robust systems for tactile perception.

Although this study focuses on an active perception strategy that seeks to move the sensor to a ‘good’ location for perception, our proposed algorithm in Fig. 4 is actually more general. Other active control strategies could be implemented with alternative criteria for the move decision (Eq. 9). For example, a recently proposed ‘Bayesian exploration’ method for tactile perception can be considered another special case of this general approach, with movement criterion to disambiguate between possible surface textures [15]. In our view, this emphasizes the importance of considering the task to be solved by active perception: their control strategy gives a way of disambiguating surface texture in structured environments, whereas our control strategy aims to best perceive texture in unstructured environments. Thus we see the main question as not whether to use active control in perception, but what is the right active perception strategy for a particular task?

REFERENCES