Superresolution with an optical tactile sensor

Nathan F. Lepora, Member, IEEE, Benjamin Ward-Cherrier

Abstract—Although superresolution has been studied to huge impact in visual imaging, it is relatively unexplored in tactile robotics. Here we demonstrate a novel optical sensor design (the TacTip) capable of achieving 40-fold localization superresolution to 0.1 mm accuracy compared with a 4 mm resolution between tactile elements. This superresolution is reached for localizing a 40 mm diameter hemicylinder with a tactile finger pad also of 40 mm diameter. Deformations of the sensor surface are measured as displacements of molded internal pins, with pin separation thus defining sensor resolution. Active Bayesian perception for classifying object location was used to ensure robust localization and hence the magnitude of the superresolution. These results are comparable with those for capacitive tactile sensors, which we interpret as originating from a convergence in the taxel-based design of the optical sensor and capacitive tactile sensors. The attained superresolution is comparable to the best perceptual hyperacuity in humans.

I. INTRODUCTION

Superresolution sensing encompasses a range of techniques for transcending the resolution limit of a sensor and earned the 2014 Nobel Prize in Chemistry (for superresolved fluorescence microscopy). Visual superresolution has revolutionised the life sciences by enabling the imaging of nanoscale features within cells below the resolution limit of optical microscopes. Tactile superresolution has the potential to drive a step-change in tactile robotics, by enabling tactile sensing at accuracies far finer than the spacing between tactile elements (taxels), which has commonly been viewed as limiting the accuracy of tactile devices.

Recently, a 35-fold superresolution to 120 \( \mu \)m localization sensitivity has been demonstrated with a capacitive touch sensor with 4 mm resolution and 12 taxels [1], the iCub tactile fingertip [2]. Three principles helped the perceptual acuity surpass sensor resolution [1]: (i) the sensor is constructed with multiple overlapping, broad but sensitive taxels; (ii) the tactile perception method interpolates between taxels to attain superresolution; (iii) active (closed-loop) touch [3]–[5] ensures robustness to initial contact location. These factors follow from applying statistical techniques for robot touch [6]–[8] to biomimetic sensors with an elastomeric covering that spreads the contact over many taxels.

The purpose of this study is to demonstrate superresolution on a completely different design of tactile fingertip: the TacTip, an inexpensive, 3D-printed optical tactile sensor that uses an internally-mounted webcam to image deformations of a flexible contact pad [9]–[12]. Prima facie an optical tactile sensor would seem better suited to methods for visual superresolution; however, we show that tactile superresolution can be attained with principles derived from the study of capacitive touch sensors. Crucially, the TacTip functions by tracking an array of pins on the inner surface of the finger pad, with the displacements of these pins analogous to the sensor readings from a taxel-based device. In consequence, the TacTip functions like a taxel-based fingertip.

II. BACKGROUND AND RELATED WORK

Superresolution (SR) encompasses a range of techniques for transcending the resolution limit of a sensor (geometrical SR) and diffraction limit (optical SR) [14]. An example

![Fig. 1. Experimental setup. The tactile fingertip (TacTip) taps vertically down against a 40 mm hemicylinder to determine its horizontal location relative to the test object.](image-url)
of geometrical SR is sub-pixel localization [15], by interpolating location over many activated pixels (Fig. 2). Our expectation is that geometrical SR will be directly relevant to robot touch, because resolution is a principal limitation of tactile sensors.

Standard geometrical SR techniques from image processing do not apply well to tactile robotics. Imaging techniques are based on hardware assumptions for conventional vision sensors which have planar two-dimensional arrays, minimal cross-talk between pixels and high pixel densities [14]. Conversely, most tactile sensors have complex geometries (e.g. Figs 1,3), significant cross-talk between tactile elements induced by the surface medium [16] and low tactile element densities. Moreover, tactile perception differs from computer vision in being innately active, in that all tactile measurements result from directly contacting an object, so that the nature of the physical contact and how the robot is controlled during perception cannot be ignored.

A recent study of tactile SR demonstrated a 40-fold improvement over sensor resolution for a tactile fingertip and a 20-fold improvement for a larger section of tactile skin (both developed for the iCub [2]). These accuracies are much improved over previous attempts at tactile SR. One previous study used the same tactile fingertip [13], but suffered from an incomplete location range with a low sampling density. Another previous study used digital image processing techniques for multi-exposure noise reduction [17]; however, those methods relied on having flat tactile arrays, whereas both the tactile sensor considered here and the iCub fingertip are curved. Moreover, tactile data is commonly from dynamic contacts, such as taps, whereas imaging techniques rely on single images (or many versions of the same image).

III. METHODS

A. Details of the tactile sensor and data collection

1) The Tactile fingertip: The TacTip is an optical tactile sensor developed at Bristol Robotics Laboratory [9]–[12] that has several highly useful properties (Fig. 3): (i) the casing is 3D-printed and hence readily customizable and inexpensive; (ii) it uses a standard CCD web-camera (LifeCam Cinema HD, Microsoft) to collect data, which is also inexpensive and connects to a PC via a USB interface; (iii) it has a molded silicon outer membrane that is robust to wear and easily replaced if damaged; and (iv) between the outer membrane and the electronics is a clear compliant gel (RTV27906, Techsil UK) that both enables tactile sensing through compression and protects the delicate parts of the sensor.

The particular design of TacTip used here has a 40 mm diameter hemispherical sensor pad with 532 pins arranged in a regular array on its underside. Six LEDs are mounted on a ring around the base of the pad to illuminate the pins, whose tips have been coated with white paint to give good contrast with the black silicon outer membrane.

2) Data collection: The TacTip is mounted as an end-effector on a six degree-of-freedom robot arm (IRB 120, ABB Robotics). The arm can precisely and repeatedly position the sensor (absolute repeatability 0.01 mm). As such, it is an ideal platform to probe tactile sensing.

A smooth hemicylinder (diameter 40 mm) was fixed in place and used as a test object (Figs 1,3). Data were collected while the tactile sensor tapped 8 mm down onto the test object followed by a move back up and then a horizontal displacement $\Delta x = 0.01$ mm, before making the next tap. A horizontal $x$-range of 40 mm was used, giving 4000 taps across the cylindrical object. From each tap of a tactile sensor against the test object, a 2 sec time series of pressure readings ($N_{\text{samples}} = 40$) was extracted. The data used later in this paper were collected twice to give distinct training and test sets. This approach for validation ensured that the results are based on sampling from an independent data set to that used to train the classifier.

3) Data preprocessing: The TacTip collects tactile data as images (resolution 640×480 pixels, sampled at ~20 fps), which are filtered to detect and track displacements of pins molded to the underside of the outer membrane. Images were captured and filtered using opencv (http://opencv.org/). For pin detection, a Gaussian spatial filter with adaptive threshold was applied to each image (pins shown as circles in Fig. 4); the adaptive threshold allowed for varying luminosity.
across the image field. For pin tracking, the Lucas-Kanade algorithm [18] was then applied to the stack of images for each tap, to give the optical flow of pins during a contact. The displacements of individual pins could then be inferred by applying the flow field to the initial pin locations (assumed initially the same for all taps).

The tactile elements were sub-sampled from the detected pins, reducing the computational requirements and removing redundancy from the data. Selected pins were constrained to have at least 4 mm separation from each other (achieved by using an iterative selection algorithm starting from the central pin). This procedure resulted in 40 selected pins (shown colored in Fig. 4) that were used for all analysis. Then the two dimensional \( s_k \) and \( s_j \) displacements of these selected pins \( s_k(j) \) are treated as distinct data dimensions, so that

\[
1 \leq k \leq N_{\text{dims}} = 80 \quad \text{and} \quad 1 \leq j \leq N_{\text{samples}} = 40.
\]

### B. Active and passive algorithms for perception

We use a Bayesian approach for tactile perception that is based on sequential analysis methods for optimal decision making. Sequential analysis is a statistical technique for hypothesis selection over data that is sequentially sampled until reaching a stopping condition [19], which commonly takes the form of a threshold on the posterior belief. Related methods are used in operations research and to model decision formation in humans and animals [6], [20].

Formally, the perception algorithms apply to sequences of contact data \( z_{1:t} = \{z_1, \cdots, z_t\} \) that are multi-dimensional time series of sensor values,

\[
z_t = \{s_k(j) : 1 \leq j \leq N_{\text{samples}}, 1 \leq k \leq N_{\text{dims}}\}, \tag{1}
\]

with indices \( j,k \) labeling the time sample and data dimension respectively. This contact data gives evidence for the present location class \( x_t \), \( 1 \leq t \leq N_{\text{loc}} \), considered one of a set of distinct punctual locations (here \( N_{\text{loc}} = 400 \) locations spanning 40 mm are used). Algorithm details are as follows.

1) **Measurement model and likelihood estimation**: The location likelihoods \( P(z_t|x_t) \) use a measurement model of the training data for each location class \( x_t \)

\[
\log P(z_t|x_t) = \sum_{k=1}^{N_{\text{dims}}} \sum_{j=1}^{N_{\text{samples}}} \log P_k(s_k(j)|x_t) \frac{N_{\text{samples}}}{N_{\text{ dims}}} \tag{2}
\]

constructed by assuming all data dimensions \( k \) and samples \( s_k(j) \) within each contact are independent (so individual log likelihoods sum). Here this sum is normalized by the total number of data points \( N_{\text{samples}} N_{\text{dims}} \) to ensure that the likelihoods do not scale with the sample number of a contact.

Following other work on robot tactile perception [6], [21], the probabilities \( P_k(s_k(j)|x_t) \) are found with a histogram method applied to training data for each location class \( x_t \). The sensor values \( s_k \) for data dimension \( k \) are binned into equal intervals \( I_k, 1 \leq b \leq N_{\text{bins}} \) over their range (here with \( N_{\text{bins}} = 100 \)). The sampling distribution is given by the normalized histogram counts \( n_k(b) \) for training class \( x_t \):

\[
P_k(s_k|x_t) = P_k(b|x_t) = \frac{n_k(b) + \varepsilon}{\sum_{b=1}^{N_{\text{bins}}} n_k(b)}, \tag{3}
\]

where \( n_k(b) \) is the sample count in bin \( b \) for dimension \( k \) over all training data in class \( x_t \). Technically, the likelihood is ill-defined if any histogram bin is empty, which is fixed by regularizing the bin counts with a small constant (\( \varepsilon \ll 1 \)).

2) **Bayesian belief update**: Bayes’ rule is used after each successive test contact \( z_t \) to recursively update the posterior location beliefs \( P(x_{1:t}|z_{1:t}) \) for the perceptual classes with the location likelihoods \( P(z_t|x_t) \) of that contact data

\[
P(x_{1:t}|z_{1:t}) = \frac{P(z_t|x_{1:t})P(x_{1:t-1}|z_{1:t-1})}{P(z_t|z_{1:t-1})}, \tag{4}
\]

from background information given by the prior location beliefs \( P(x_{1:t}|z_{1:t-1}) \) (i.e. the posterior beliefs from the preceding contact). The marginal probabilities are given by

\[
P(z_{1:t-1}) = \sum_{l=1}^{N_{\text{loc}}} P(z_{1:t}|x_l)P(x_l|x_{1:t-1}). \tag{5}
\]

Iterating (4, 5), a sequence of contacts \( z_1, \cdots, z_t \) results in a sequence of posterior beliefs \( P(x_t|z_1), \cdots, P(x_t|z_{1:t}) \) initialized from uniform priors \( P(x_t|z_0) := P(x_t) = 1/N_{\text{loc}} \).

3) **Final location decision**: Using sequential analysis methods for optimal decision making, the update (4, 5) stops when the posterior belief \( P(x_{1:t}|z_{1:t}) \) passes a threshold \( \theta_{\text{dec}} \), with location decision \( x_{\text{dec}} \) at decision time \( t_{\text{dec}} \) given by

\[
\text{if any } P(x_t|z_{1:t}) > \theta_{\text{dec}} \text{ then } x_{\text{dec}} = \arg\max_{x_t} P(x_t|z_{1:t}). \tag{6}
\]

This belief threshold \( \theta_{\text{dec}} \) is a free parameter that adjusts the balance between decision time \( t_{\text{dec}} \) and decision accuracy.

4) **Active perception**: Active perception uses a posterior-dependent control policy \( \pi \) to move the sensor \( x \leftarrow x + \pi \) during the perceptual process. For simplicity, we consider this to depend only on an intermediate estimate of location

\[
x_{\text{est}}(t) = \arg\max_{x_t} P(x_t|z_{1:t}). \tag{7}
\]
Two control policies were considered, one passive stationary (no feedback) and the other active:

1. **Passive stationary perception** never moves the sensor from outside the initial location class it contacts the object \( \pi = 0 \).
2. **Active ‘fixation point’ control** [1], [7], [8] attempts to move the sensor to a predefined fixation point \( x_{\text{fix}} \) relative to the object assuming it is at \( x_{\text{est}} \) on the object,

   \[
   x \leftarrow x + \pi (x_{\text{est}}), \quad \pi(x_{\text{est}}) = x_{\text{fix}} - x_{\text{est}}. \tag{8}
   \]

Provided the fixation point is set to be a good location for perception, the perception will progressively improve during the decision making process from an initially unknown location where the perception may be poor. Here we take a fixation point at the center of the location range.

After a move of \( \Delta l \) location classes, the location beliefs \( P(x|z_{1:t}) \) should be kept aligned with the sensor by shifting the class probabilities by the number of classes moved

\[
P(x|z_{1:t}) \leftarrow P(x_{t-\Delta t}|z_{1:t}) \quad \text{if} \quad 1 \leq x_{t-\Delta t} \leq N_{\text{loc}}, \tag{9}
\]

\[
else P(x|z_{1:t}) \leftarrow P(x_{1}|z_{1:t}) \quad \text{or} \quad P(x_{N_{\text{loc}}}|z_{1:t}).
\]

For simplicity, the (undetermined) evidence shifted from outside the location range is assumed uniform and given by the existing probability at that extremity of the range (probabilities can then be renormalized to have unit sum).

5) **Virtual environment validation:** The aim of the data collection is to make a ‘virtual environment’ in which the localization acuity of the TacTip can be evaluated off-line. For analysis, the data are separated into \( N_{\text{loc}} = 400 \) distinct location classes, by collecting groups of 10 taps together, with each group spanning 0.1 mm of the overall 40 mm range. The localization error is quantified with the mean absolute error \( e_{\text{loc}}(x) = |x - x_{\text{dec}}| \) over all classified \( x_{\text{dec}} \) location values for all test runs with true location class \( x \).

The mean localization error \( \bar{e}_{\text{loc}} = \frac{1}{N_{\text{loc}}} \sum_{t=1}^{N_{\text{loc}}} e_{\text{loc}}(x_t) \) is averaged over all location classes \( x_t \). The decision time \( t_{\text{dec}}(x) \) and mean decision time \( \bar{t}_{\text{dec}} \) are defined similarly.

Monte Carlo validation is used to ensure good statistics by averaging errors over test data drawn randomly from all location classes. Typically we averaged over 10000 distinct Monte Carlo iterations for each value of the decision threshold \( \theta_{\text{dec}} \), randomly sampling over all 400 location classes.

IV. RESULTS

A. Inspection of data

Data for the TacTip (Fig. 5) were collected while the sensor tapped repeatedly against the test object (hemicylinder, 40 mm diameter) with a small (0.01 mm) horizontal displacement after each tap to sample across a 40 mm location range with 4000 taps. Individual taps (visible in Figs 5C,D) typically take about 250 ms to reach peak response amplitude, plateau for 250 ms, then return to baseline also taking 250 ms. Contact features from the stimulus are encoded in the time-series response of each taxel (pins colored on Figs 4,5), including its temporal dynamics and peak value reached.

The most obvious effect of varying horizontal contact location \( x \) was a change in the activation of tactile elements and their peak response amplitudes. For each contact, the pattern of taxel pressures depends on the location of the TacTip relative to the object, permitting classification of where the rod is located relative to the TacTip. For example, in the pattern of \( s_x \) and \( s_y \) displacements (Figs 5A,B), the left-most locations activate the taxels plotted in blue on the left of the TacTip (c.f. Fig. 4), changing as the TacTip moves rightwards to the taxels plotted in red on the right.
B. Localization acuity over a test object

The localization acuity of the TacTip is first assessed by applying passive stationary (open-loop) Bayesian perception detailed in the methods (Section III-B) to the contact data (Fig. 5) over the 40 mm location range. Results are generated with a Monte Carlo procedure repeatedly drawing trials from the test data (Sec. III-B.5), such that each contact tap passively remains within its initial location class.

The location error $\epsilon_{\text{loc}}(x)$ varies strongly with test location class $x$ across the 40 mm range (Fig. 6A). The largest location errors are $\epsilon_{\text{loc}}(x) \gtrsim 0.5$ mm for glancing off-center contacts near the extremities ($x < 5$ mm and $x > 35$ mm). The smallest location errors are $\epsilon_{\text{loc}}(x) \sim 0.1$ mm in the central region (roughly $15 \leq x \leq 25$ mm) on top of the test cylinder. Decision times $t_{\text{dec}}(x)$ are also a (noisy) U-shaped function over location $x$ (Fig. 6B), indicating a reliable decision is reached with only a few taps in the central low location error region where the evidence is most reliable.

Therefore the TacTip is capable of attaining 40-fold superresolution sensing, with the best localization acuity $\sim 0.1$ mm about one-fortieth of the $\sim 4$ mm spacing between taxels. However, for passive perception, this accuracy is only reached for a narrow range of central locations directly over the test hemicylinder.

C. Robust superresolution with active touch

The problem with assessing the TacTip from the lowest localization error $\epsilon_{\text{loc}}(x)$ over the entire location range is that it presupposes that the sensor can be placed in that high acuity region, which requires knowledge of location that is the quantity being classified. In other words, because location is unknown a priori, the TacTip could sense from a poor acuity region near the extremities of the location range. In consequence, passive superresolution is non-robust.

As has been described elsewhere [1], [7], this robustness problem can be solved using (closed-loop) active touch. Between contact taps onto the test object, the TacTip could be repositioned via an intermediate estimate of location, using a control policy that attempts to attain a good region for perception (here assumed the center $x_{\text{fix}} = 20$ mm of the location range). Thus we implement active touch (details in Sec. III-B) with the same Monte Carlo procedure (Sec. III-B.5) used to validate superresolution with passive perception.

For active perception, the mean location error reaches $\tilde{\epsilon}_{\text{loc}} \lesssim 0.1$ mm averaged over all location classes $x$ in the 40 mm range (Fig. 7, green curve). The best acuities are reached as the error asymptotes for decision thresholds $\theta_{\text{dec}} \gtrsim 0.2$, with on average $\bar{t}_{\text{dec}} \gtrsim 3$ contacts. As the threshold is lowered, fewer contacts are needed for a decision, and the error increases to $\epsilon_{\text{loc}} \sim 0.6$ mm at $\theta_{\text{dec}} = 0$ for one contact $\bar{t}_{\text{dec}} = 1$. This coincides with passive perception because there is no opportunity to move.

Overall, the best mean location error with active touch $\epsilon_{\text{loc}} \gtrsim 0.1$ mm gives a better than 40-fold superresolution capability. This active superresolution is averaged over decisions across the entire location range, and is thus robust to initial sensing location.

V. DISCUSSION

In this study, we demonstrated 40-fold superresolution with an optical tactile fingertip (the TacTip) localizing the position of a hemicylindrical surface (40 mm dia.). The best localization acuity was $\sim 0.1$ mm, a forty-fold improvement over the 4 mm sensor resolution used for our analysis.

Active and passive methods for tactile perception were compared, and only active touch found to attain robust superresolution. The best localization acuities occurred when the
TacTip was centered over the test object, deteriorating to poor acuity at the extremities where the TacTip makes glancing contacts. In consequence, passive (stationary) superresolution is non-robust because location is, by definition, unknown. Conversely, active superresolution is able to correct for a poor initial location by using intermediate localization estimates to move to a high acuity region on the object.

These results are consistent with a previous study of superresolution for capacitive tactile sensors [1] that also found ~0.1 mm localization acuity with 4 mm wide taxels. However, a key point is that design and method of data capture for the TacTip optical tactile sensor are completely different from those for capacitive sensors. Whereas capacitive sensors have taxels defined from the internal plates of discrete capacitors [16], for optical sensors the data is captured as a (pixelated) image. Here we showed that a novel design feature of the TacTip, molded pins on the inside of the finger pad [9]–[12], enables the TacTip to also be treated as a taxel-based device with the taxel readings corresponding to the two-dimensional pin displacements tracked over time. Therefore methods for active Bayesian perception developed with capacitive tactile sensors [1], [7], [8] apply also to optical tactile sensors based on the TacTip design. In consequence, we can quantify the superresolution capabilities of the TacTip relative to the pins selected as taxels.

Comparing the magnitude of superresolution for a capacitive tactile sensor [1] with that for the TacTip, the results seem very similar at two orders of magnitude over sensor resolution. However, the iCub tactile sensor at 15 mm long is much smaller than the TacTip (40 mm wide). Moreover, the test stimulus used here (40 mm dia. cylinder) is larger than in the previous superresolution study (6.4 mm dia. cylinder). Also, the iCub fingertip has 12 taxels spaced 4 mm apart, whereas we used 40 pins spaced 4 mm apart in the TacTip.

Why, then, are the results so similar across the two studies? In our view, this was because the resolution (taxel spacing) was similar, and superresolution at two orders of magnitude is about as good as can be reasonably attained. For example, the best superresolution with the human eye is also about two orders of magnitude (about 0.0002 degree acuity over 0.0167 degree resolving power for 20/20 vision) [22], and typically about an order of magnitude with the human fingertip [23]. Accordingly, superresolution at two orders of magnitude is a good goal for artificial tactile sensing.

VI. CONCLUSION

This study demonstrates that a novel design of optical tactile sensor, the TacTip, can perceive horizontal location to 0.1 mm acuity compared with a 4 mm spacing between tactile elements that define the sensor resolution. This result is averaged over a 40 mm horizontal range of sensing locations of the TacTip (with 40 mm diameter finger pad) impinging on a 40 mm diameter test hemicylinder. It was necessary to deploy active (closed-loop) tactile perception methods to attain this forty-fold superresolution, because otherwise the accuracy was non-robust to initial sensing location. The attained superresolution is comparable to the best perceptual hyperacuity in humans.

Acknowledgements: We thank Ben Winstone, Gareth Griffiths, Tareq Assaf and Tony Pipe for help with the TacTip.

REFERENCES