Where am I Touching on An Object? Solved by Localizing Tactile Features in a Visual Map

Shan Luo (shan.luo@kcl.ac.uk), Centre for Robotics Research, King's College London

Abstract—The Ph.D project is aimed to localize the contact location of a tactile sensor on an object. By taking advantage of certain correspondences between visual and tactile sensing, it is proposed to localize tactile readings in visual images by sharing same sets of feature descriptors through two sensing modalities. Thus the recursive Bayesian filtering is employed and feature-based measurement model and Gaussian based motion model are built. In our tests, a tactile array sensor is utilized to generate tactile images during interaction with objects and test results have proved its feasibility.

I. INTRODUCTION

Humans tend to utilize both vision and the tactile sensations of hands during grasping and/or manipulating objects. Key features, e.g., corners and edges, are first captured by our eyes and direct hands to an appropriate pose to handle objects. However, these visual features become unobservable during the manipulations in hand as vision is occluded. In this case, touch sensation makes up the information loss caused by the occlusion. The corresponding features are sensed in the tactile modality and poses of objects can thus be inferred. In this process prominent features act as a bridge between vision and tactile sensing, making our hand movements consequent. To apply this mechanism in robotics, we propose to use same feature descriptors in both vision and tactile sensing to localize tactile features in visual images.

Early researchers have attempted to fuse vision and touch since decades ago but tactile sensors were only utilized to verify contacts due to the low resolution [1] [2]. Thanks to the increasing spatial resolution and spatiotemporal response, recently developed off-the-shelf tactile sensors have shown the ability to serve as “imaging” devices. Schneider et al. [3] took tactile images as features directly to recognize objects in a framework of Bag-of-Words (BoW) originated from computer vision. By use of the same streamline, Pezzementi et al. [4] took one step further: multiple kinds of features were extracted from tactile readings and their performances were compared. Liu et al. [5] recognized objects by classifying local features through the covariance analysis of pressure values in tactile images. All of these works show the potential of taking both visual and tactile readings as images and extracting features of the same type through two modalities for object contact estimation. Thus we propose a framework to localize tactile readings in visual images using recursive Bayesian filtering, in which feature descriptors of the same type are extracted from both visual and tactile. It is promising to be used to facilitate robotic hand manipulations.

Figure 1 illustrates our experimental system: a webcam (not shown here) and a Weiss tactile sensor of 14×6 sensing elements are utilized to obtain visual and tactile images of objects respectively; a Phantom Omni haptic device to which the tactile sensor is attached is employed to acquire the positions of the tactile sensor. A 3D printed gecko model is utilized to test our algorithm, which has a 2D shape that protrudes from its base 4mm. The sensor is controlled to press at multiple steps to follow its surface and the sensor plane is kept normal to the surface of the gecko.

II. METHODOLOGY AND EXPERIMENTS

The recursive Bayesian filtering is employed to estimate the location (state) \( x_t \) of the tactile sensor in the visual map \( m \) at each time step \( t \), as illustrated in Fig. 2. The tactile readings \( z_t \) and movements of the tactile sensor \( u_t \) in the 3D space are collected to update the belief distributions over possible locations (states) in the visual map. Based on the input streams, it can be divided into two steps, i.e., control update and measurement update.

In control update, the belief \( \hat{p}(\mathbf{x}_t|u_t) \) at \( x_t=(x_i,x_j) \) is obtained by summing each product of the probability at the deterministic state \( (x_i,a_i,x_j-b) \) and its nearest \( k \) states \( (k=8 \) in our case) at time \( t-1 \), measured in Euclidean distance, and the Gaussian probability density \( f(i,j) \). Here, \((i,j)\) is the shift from the deterministic state.
In the measurement update, the belief \( p(x_t) \) is obtained as the product of \( \hat{p}(x_t|u_t) \) and the probability \( p(z_t|x_t) \) that the measurement \( z_t \) may have been observed for each hypothesical posterior state \( x_t \). \( p(z_t|x_t) \) is inverse to the Euclidean distance \( d(p_{l_{ij}}, z_t) \) between descriptors extracted from segmented sub-image \( p_{l_{ij}} \) with index \((i,j)\) and the tactile image \( z_t \).

These features are adapted Scale Invariant Feature Transform (SIFT) descriptors [6], taking the invariance to translation and rotation into consideration. Compared to the classic SIFT algorithm, scale-space pyramids and key point localization are removed as in [7]. As shown in Fig. 3, each tactile image is segmented into three overlapped sub-images of the same size and one SIFT descriptor is extracted from each sub-image, taking sub-image centers as “key points”. Different from classic SIFT descriptors with a dimension of 128, descriptors of 32 elements are obtained instead in our case by reducing sampling areas from a 4×4 grid to a 2×2 grid to minimize the computation time, as illustrated in Fig. 3. More details can be found in our previous work [8].

![Fig. 3. Each tactile image is segmented into three overlapped sub-images and one 32-dimensional feature (\( f_1, f_2, f_3 \)) is extracted from each sub-image.](image)

The same approach is also applied to the visual images, extracting three SIFT descriptors from each sub-image. As shown in Fig. 4, a sliding window with the size of tactile image is carried out to get matching probability for each pixel.

![Fig. 4. Feature matching between visual and tactile images. (a) A sliding window (marked in red) in the visual image. (b) Tactile image \( z_t \) at time \( t \).](image)

In total, 22 experiments were carried out and in each of the experiment the tactile sensor explored the gecko model in different exploration paths. At the end of the exploration process, the robot can locate the tactile sensor in the gecko map with a large certainty. As shown in Fig. 5, it can be noticed that the localization errors tend to decrease through the localization process. To conclude, the experimental results prove the feasibility of our proposed framework in localizing the tactile sensor in a visual object map.

Fig. 5. Sample localization process and results. First row: obtained tactile images at each step. Second row: corresponding ground truth locations of the tactile sensor (marked as red) in the visual map. Last row: Corresponding probability distribution of locating the tactile sensor at different states.

### III. CONCLUSION

This paper proposes to integrate the vision and tactile sensing applied in localizing the tactile readings in the visual object image, which is viewed as a probabilistic estimation problem and solved by the recursive Bayesian filtering. Gaussian noises are used to model the motion; in measurement update, revised SIFT descriptors are extracted from both tactile and visual images and belief distributions are updated by feature matching. Test results prove its feasibility and it is promising to be used to facilitate robotic grasping and hand manipulations. More details can be found in [9].

### REFERENCES


