Rapid Crack Reconstruction Using Tactile Sensing and Vision

Jiaqi Jiang, Guanqun Cao, Daniel Fernandes Gomes and Shan Luo

I. INTRODUCTION

Cracks in the infrastructures surface are important indicators for assessing the condition of buildings and need to be repaired timely for preventing the expansion of potential risks. However, manual crack detection and maintenance are not only time-consuming and expensive, but also pose health risks to human workers in harsh and complex environments such as dams and underground pipelines. Hence, development of an effective and robust crack detection system will be significant for the substitution of inspection workers

In recent years, computer vision techniques such as edge detection [1], object detection [2] and segmentation [3], have been applied in detecting cracks in concrete structures. However, they suffer from variances in light conditions and shadows, lacking robustness and resulting in many false positives. To address those issues, some researches [4][5][6] detect cracks with tactile sensing which is less susceptible to light and noise. In [4], a Bayesian approach is proposed to actively estimate crack width. Nevertheless, when the initial contact location is far away from cracks, it will take a long time to find the crack, or even fail to find it. In [5][6], tactile sensing was explored for crack detection and characterization, but tactile data was collected passively and the shape of the crack cannot be reconstructed accurately.

In crack detection tasks, skilled inspectors usually first look at the surface to find areas with similar color or shape characteristics to cracks, and then use their fingers or specific tools such as a hammer or ultrasonic device to further inspect those areas instead of traversing all regions. However, there has been no works on crack detection using both vision and touch yet. In this paper, we propose a novel vision-guided tactile perception approach for crack detection and reconstruction, with an overview of the framework illustrated in Fig. 1.

II. METHODOLOGY

A. Visual Guidance for Touch

Visual Crack Segmentation. The visual crack segmentation is treated as a semantic segmentation problem that predicts each pixel of the input image into one of two semantic classes: (a) background (b) cracks. To this end, we use the Deeplabv3+ model [7] to segment the cracks in visual images that is a state-of-art deep learning model for semantic image segmentation. We use a weighted cross-entropy loss with crack pixels weighted 10x more than back ground pixels.

All the authors are with the Department of Computer Science, University of Liverpool, Liverpool L69 3BX, U.K. Emails:{jiaqi.jiang, g.cao, danfergo, shan.luo}@liverpool.ac.uk.

Contact Points Generation. Given the predicted pixel-wise crack mask in the color image, we can extract the skeleton of each crack mask with pattern thinning method. We define two types of keypoints (i.e., end points and branch points) and minimal edges that represent the topology of the crack pattern:

- End points: if they have less than two neighbors.
- Branch points: if they have more than two neighbors.
- Minimal edge E_{ij}: if there is a continuous path between two keypoints p_i and p_j and all points on the path are neither end points nor branch points.

For every minimal edge E_{ij} which consists of a number of ordered points, the keypoint p_i is initially selected as the current contact point $p_{current}$. Then we iteratively choose the next contact point p_k using the following formula:

$$\begin{aligned} &\max_{k} D[p_{current}, p_k] \\ &\text{s.t. } D[p_{current}, p_k] < d \end{aligned} \tag{1}$$

where $D(p_{current}, p_k)$ is the distance between two points in world frame. The hyper-parameter d is the threshold of the distance between two points that is related to the coverage and speed of tactile perception. A smaller d will increase the coverage while reducing the perception speed. In our case, d is empirically set to four fifths of the tactile sensor's view length. As shown in Fig. 1, the end-point pixels and the generated contact points are tagged with red dots and green dots, respectively. For each contact points p_i, the yaw angle of the end-effector is parallel to the vector $< p_i, p_n >$, where p_n is the nearest contact point to p_i, so that the end-effector can contact the surface perpendicularly.

B. Tactile Crack Perception

Tactile Crack Detection. The collected tactile images are fed to another deep convolutional network (Deeplabv3+ [7] with ResNet-101). Since the number of background pixels is similar to that of the background pixels in tactile images, we use the vanilla cross-entropy loss instead of the weighted cross-entropy loss in Section II-A.

Tactile Crack Reconstruction. First, we predict the location of cracks on the surface of the GelSight sensor [8], given the detected boundaries of pixel-wise masks in the tactile images. To simplify the problem, we treat the surface of GelSight sensor as a flat plane that is perpendicular to the webcam's z axis. Then we can easily calculate contact point $P = [X_c, Y_c, Z_c]^T$ in the tactile sensor coordinates based on the $P' = [u, v]^T$ in the tactile image coordinates using a pinhole camera model.

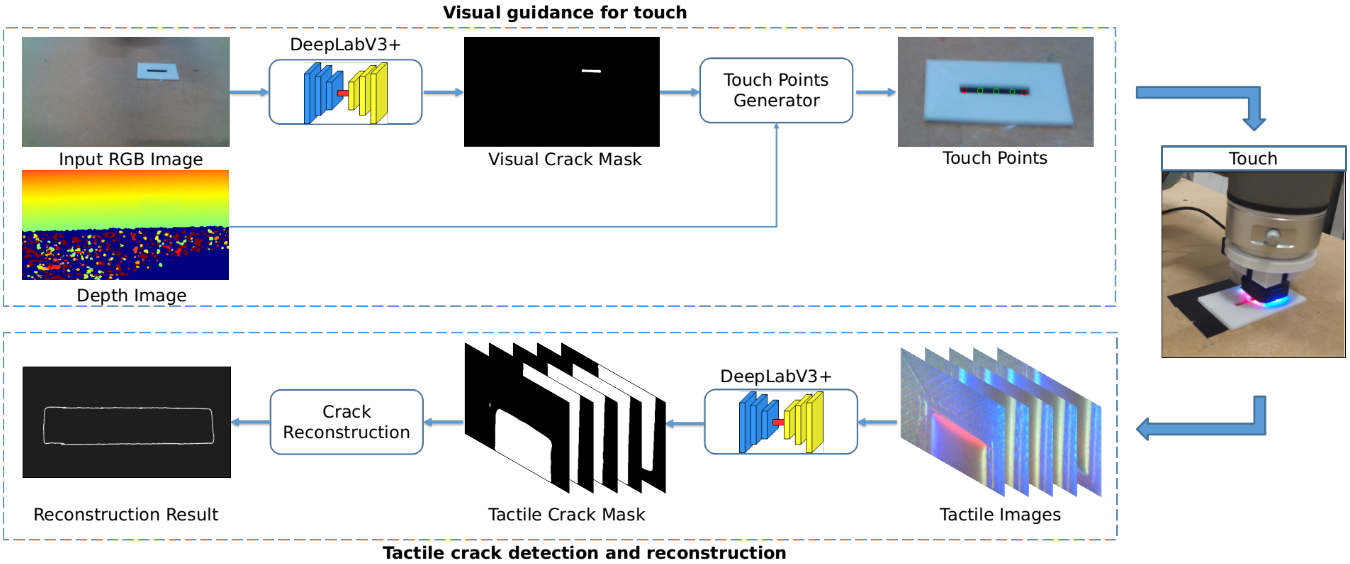

Fig. 1: An overview of our vision-guided tactile crack detection and reconstruction method. **Top row (from left to right):** The Deeplabv3+ model is used to segment the cracks in the visual image. Given the visual crack mask and the depth image, a set of contact points are generated to guide the collection of tactile images. **Bottom row (from right to left):** Another deep convolutional network is used to segment the crack in the collected tactile images. Given the detected tactile crack mask, the crack shape are reconstructed based on the geometrical model of the GelSight sensor and the coordinate transforming relation between the tactile sensor coordinate and the world coordinate.

After obtaining the position P in the tactile sensor coordinates, we can calculate its position P_W in the world frame:

$$P_W = T_E^W T_C^E P \tag{2}$$

where T_C^E and T_E^W are the transformation matrix from the tactile sensor coordinates to the end-effector coordinate system, and from the end-effector coordinate system to the world coordinate system, respectively.

III. Experiment Results

We use the mean, max and standard deviation (SD) of the shortest distance between the actual crack shape and the reconstructed crack location to evaluate the accuracy of our proposed method. There are four methods used for comparison. The vision method uses point clouds recovered through visual detection and depth information to represent cracks. In order to reduce the impact of depth information accuracy on reconstruction, the aligned vision method projects the point cloud to the table surface. The passive tactile method collects the tactile images through traversing the whole 3D printed structure surface.

TABLE I: Reconstruction Accuracy

Method	MeanD(mm)	SD(mm)	MaxD(mm)	time(s)
vision	0.82	0.92	4.87	**1**
aligned-vision	0.55	0.53	3.78	**1**
passive-tactile	**0.20**	0.17	0.99	400
guided-tactile[ours]	0.24	**0.16**	**0.82**	35

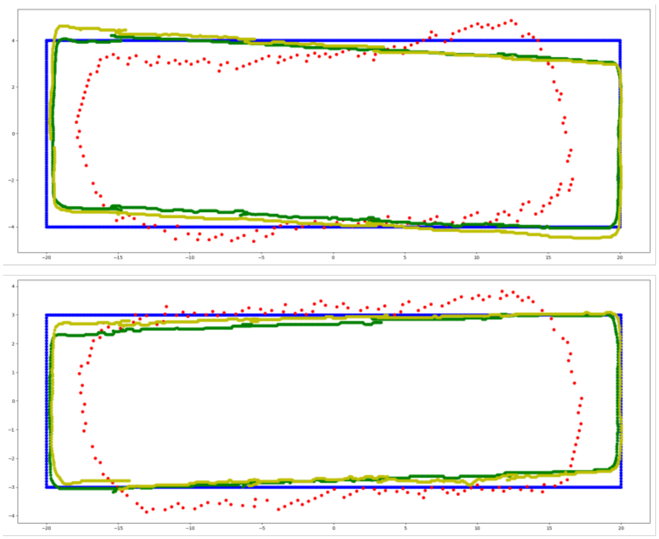

Fig. 2: Visual comparison of different reconstruction methods. blue, red, yellow, green curves represent the ground truth of the crack profile, aligned-vision, passive-tactile and our method for crack reconstruction, respectively.

IV. Conclusion and Future Work

In this paper we introduce a novel vision-guided tactile perception for crack detection and reconstruction. The experiments show that our proposed method can improve the effectiveness and robustness of crack reconstruction significantly, compared to when only vision is used. Future works to improve our method can also be considered, such as the use of weakly supervised learning methods, and fusing the tactile results back into the visual crack detection to improve the next estimate of where to contact.

References

[1] P. Prasanna, K. J. Dana, N. Gucunski, B. B. Basily, H. M. La, R. S. Lim, and H. Parvardeh, "Automated crack detection on concrete bridges," *IEEE Transactions on automation science and engineering*, vol. 13, no. 2, pp. 591–599, 2014.

[2] F.-C. Chen and M. R. Jahanshahi, "Nb-cnn: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4392–4400, 2017.

[3] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "Deepcrack: Learning hierarchical convolutional features for crack detection," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2018.

[4] N. F. Lepora and B. Ward-Cherrier, "Tactile quality control with biomimetic active touch," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 646–652, 2016.

[5] F. Palermo, J. Konstantinova, K. Althoefer, S. Poslad, and I. Farkhatdinov, "Implementing tactile and proximity sensing for crack detection," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 632–637.

[6] ——, "Automatic fracture characterization using tactile and proximity optical sensing," *Frontiers in Robotics and AI*, vol. 7, 2020.

[7] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2018.

[8] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.