XLIV Jornadas de Automática 2023

https://jautomatica.es

Predicción de deslizamiento mediante la segmentación de imágenes táctiles

Julio Castaño-Amorós.[a,*], Pablo Gil[a,b]

[a]*Grupo de AUtomática, RObótica y Visión Artificial (AUROVA), Inst. Univ. de Investigación Informática, Universidad de Alicante, Alicante, España.*
[b]*Depto. Física, Ingeniería de Sistemas y Teoría de la Señal, Universidad de Alicante, Alicante, España.*

To cite this article: Castaño-Amorós, J., Gil, P. 2023. Slip prediction using tactile image segmentation. XLIV Jornadas de Automática, 551-556. https://doi.org/10.17979/spudc.9788497498609.551

Resumen

El uso de sensores táctiles está comenzando a ser una práctica común en tareas complejas de manipulación robótica. Este tipo de sensores proporcionan información extra sobre las propiedades físicas de objetos que están siendo agarrados y/o manipulados. En este trabajo, se ha implementado un sistema capaz de medir el deslizamiento rotacional que pueden sufrir objetos durante su manipulación. Nuestra propuesta emplea sensores táctiles ópticos DIGIT, a partir de los cuáles se capturan imágenes de contacto que luego se procesan e interpretan. En concreto, nuestro método hace uso de un modelo neuronal para la detección de la región de contacto. Y posteriormente, mediante extracción de características visuales de la región detectada, se estima el ángulo causado por movimientos de deslizamiento. Nuestro método estima correctamente la región de contacto obteniendo un 95 % y 91 % usando las métricas *Dice* e *IoU*. Y es capaz de obtener un error medio máximo de 3^o en agarres de objetos nunca vistos anteriormente.

Palabras clave: Manipulación robótica, Percepción y sensorizado, Robótica inteligente, Tecnología robótica, Sistemas robóticos autónomos

Slippage Prediction from Segmentation of Tactile Images

Abstract

Using tactile sensors is becoming a common practice to achieve complex manipulation in robotic tasks. These kinds of sensors provide extra information about the physical properties of the grasping and/or manipulation task. In this work, we have implemented a system that is able to measure the rotational slippage of objects in hand. Our proposal uses the vision-based tactile sensors known as DIGITs which allow us to capture contact images, which are then processed. In particular, our method is based on a neural network model applied to the detection of touch/contact regions. Afterwards, we extract visual features from detected contact regions and we then estimate the angle generated due to an unwanted slippage. Our method obtains results of 95 % and 91 % in Dice and IoU metrics for contact estimation. In addition, it is able to obtain a mean rotational error of 3 degrees in the worst case with previously unseen objects.

Keywords: Robotic manipulation, Perception and sensing, Intelligent robotics, Robotics technology, Autonomous robotic systems

1. Introducción

Los métodos tradicionales para realizar tareas de manipulación robótica de objetos suelen utilizar sensores de visión 2D o 3D (Du et al., 2021). Éstos solo suelen considerar las propiedades geométricas de los objetos para llevar a cabo el agarre. En cambio, el uso de sensores táctiles posibilita medir y reaccionar ante estas propiedades físicas (masa, centro de gravedad o fricción) para conseguir un agarre estable (Luo et al., 2017).

En la última decada, se han desarrollado tecnologías táctiles muy variadas, tales como sensores resistitivos, capacitivos, barométricos, etc. (Chi et al., 2018). En los últimos años, la tendencia se ha inclinado más hacia la implementación de sensores táctiles ópticos o basados en cámara (Zhang et al., 2022). En este trabajo, se presenta un algoritmo que, haciendo uso de sensores visuo-táctiles sin marcadores, es capaz de estimar el

*julio.ca@ua.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

ángulo de rotación causado por el deslizamientos de objetos en mano en tareas de manipulación. El método propuesto se basa en redes neuronales de segmentación para obtener la región táctil de contacto entre las superficies del objeto y del sensor, y en técnicas de visión artificial para calcular el ángulo de rotación a partir de la región de deslizamiento segmentada. Cabe hacer notar que en este trabajo se han utilizado dos sensores visuo-táctiles, DIGIT (Lambeta et al., 2020), cada uno de los cuales proporciona como salida una imagen RGB sin marcadores.

La estimación de la región táctil de contacto entre los dedos del robot y el objeto agarrado es un problema que, desde el estado del arte, se ha tratado resolver con diversas aproximaciones. Por ejemplo, restando imágenes táctiles de contacto y no contacto (Lambeta et al., 2021), detectando el movimiento de marcadores visuales presentes en el sensor (Ito et al., 2014), a través de reconstrucción 3D y algoritmos fotométricos (Wang et al., 2021), o utilizando redes neuronales (Bauza et al., 2019), (Lin et al., 2022). Aunque nuestro trabajo está inspirado en estos trabajos mencionados, se diferencia de ellos, en que se hace uso de imágenes táctiles que no disponen de marcadores visuales (Ito et al., 2014) para medir el movimiento, tampoco se dispone de información de profundidad (Wang et al., 2021). Nuestra propuesta usa modelos neuronales para segmentación, que proporcionan más robustez ante incertidumbres que operaciones de morfología matemática (Lambeta et al., 2021), que redes neuronales convolucionales (CNN) simples como en (Bauza et al., 2019), y además, cuyo entrenamiento tiende a converger antes comparado con el entrenamiento de las redes neuronales generativas (GAN) (Lin et al., 2022).

El deslizamiento es un evento físico, concretamente un movimiento no deseado del objeto en mano, que se produce con más frecuencia de la deseada, por ejemplo cuando un objeto resbala durante una tarea de manipulación como agarrar y levantar. La detección e interpretación de este tipo de eventos está cobrando especial relevancia en la actualidad. Así, recientemente, se ha trabajado en la detección de eventos de deslizamiento binarios mediante técnicas de preprocesamiento de imagen (Castaño-Amorós et al., 2023) y, también, se han implementado redes recurrentes (LSTM) y redes recurrentes con capas convolucionales (Conv-LSTM) para clasificar tipos de deslizamiento tanto traslacionales como rotacionales (Zapata-Impata et al., 2019). Estos métodos, como debilidad, no llegan a medir o caracterizar numéricamente el evento de deslizamiento. No obstante, hay otras aproximaciones como (Kolamuri et al., 2021) y (Toskov et al., 2023) que sí tratan de proporcionar información medible sobre el deslizamiento, como la estimación del ángulo de rotación de deslizamiento. Para hacerlo, (Kolamuri et al., 2021) hizo uso de sensores visuo-táctiles con marcadores para seguir el movimiento, y (Toskov et al., 2023) se apoyó en sensores de fuerza. Este trabajo, también, busca caracterizar y medir el deslizamiento, y en concreto inspirándose en estos últimos trabajos, presenta un método para medir el ángulo rotacional de deslizamiento.

2. Metodología

En este artículo se propone un método de dos etapas para la segmentación de la región de contacto táctil y la estimación del ángulo de deslizamiento rotacional utilizando los sensores táctiles DIGIT. Estos sensores están formados por un gel, una ventana de metacrilato, una cámara, una tira de LEDs y una carcasa de piezas 3D. La cámara captura imágenes en color a 30 fps con una resolución de 320x240.

La primera etapa del método está basada en redes neuronales de segmentación aplicadas a la sensorización táctil basada en visión. Como nuestro objetivo solo es segmentar la región de contacto, hemos decidido construir una red TSNN (Tactile Segmentation Neural Network) que hace uso de una arquitectura DeepLabV3+ (Chen et al., 2018). Esta arquitectura es conocida por emplear un esquema *encoder – decoder* para realizar la segmentación táctil, y por introduccir una capa adicional en la que se combinan capas convolucionales dilatadas y capas convolucionales separables en profundidad. Esta combinación permite reducir la complejidad computacional. Con respecto al *encoder*, se ha utilizado una versión modificada de la conocida arquitectura Xception, llamada *AlignedXception*, en la cual se sustituyen todas la capas de agrupamiento máximo por capas convolucionales separables en profundidad para extraer las características. Por otro lado, el *decoder* es una parte más simple que el encoder debido a que está formado simplemente por una serie de capas de convolución, concatenación y muestreo ascendente para transformar las características intermedias en la salida final.

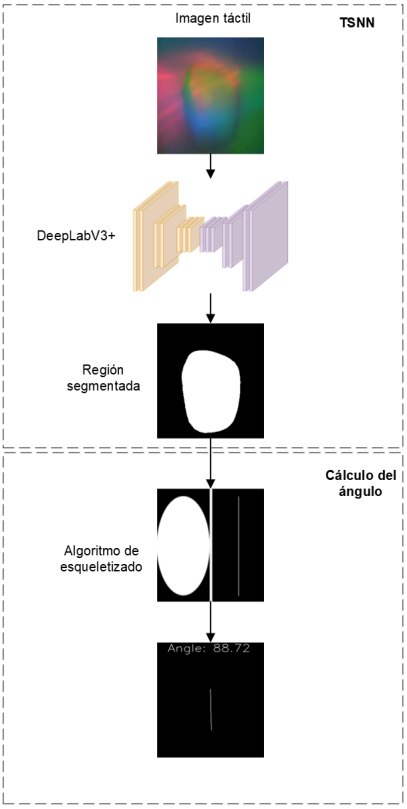


Figura 1: Esquema de nuestro método combinando ambas etapas.

La segunda etapa del método propuesto se encarga de estimar el ángulo de rotación, a partir de la región de contacto segmentada, utilizando un algoritmo de adelgazamiento del esqueleto (Skeleton Thinning) (Guo and Hall, 1989) que ennegrece los píxeles de una región binaria utilizando una 8-vecindad

cuadrada y diferentes condiciones de conectividad. Este algoritmo de esqueletizado convierte la zona de contacto táctil en una línea recta en el mejor de los casos. Por lo tanto, se puede obtener el ángulo de rotación del objeto agarrado como la orientación de esta línea recta con respecto al eje x horizontal utilizando un método de cálculo de mínimos cuadrados.

Otras aproximaciones, basadas en diferentes redes neuronales para segmentación como Unet++ (Zhou et al., 2018) y PSPNet (Zhao et al., 2017), así como otros algoritmos para estimar el ángulo, tales como el análisis de componentes principales o la estimación de una elipse, fueron también evaluados durante el diseño del método propuesto. El pipeline completo viene descrito en la Figura 1.

3. Experimentación y resultados

Para llevar a cabo el ajuste y evaluación del método propuesto, se ha diseñado un conjunto de imágenes que contiene muestras táctiles en las que se ha etiquetado la región de contacto de forma manual. Cada una de estas imágenes han sido obtenidas a partir de secuencias táctiles capturadas con sensores DIGIT reales, en el transcurso de la ejecución de tareas de manipulación que conllevan acciones como agarrar en diferentes poses y levantar el objeto agarrado. Nuestro conjunto de datos, el cual se divide en tres particiones de entrenamiento (70 %), validación (20 %) y testeo (10 %), contiene 3675 imágenes táctiles con sus respectivas regiones de contacto etiquetadas. Para obtener las imágenes táctiles, se han utilizado 16 objetos del conocido dataset YCB (Calli et al. (2017)), a razón de en torno a 200-250 imágenes táctiles por objeto. Estos objetos contienen diferente textura, rigidez, peso, geometría, etc. Se ha optado por crear nuestro propio conjunto de muestras táctiles de evaluación, ante la dificultad de disponer de imágenes táctiles procedentes de sensores DIGIT, ya que en el estado del arte no hemos encontrado muestras que no fueran sintéticas (obtenidas desde simulador) y/o que registraran información de toque similar a la de las acciones robóticas que contemplabamos.

Para entrenar nuestro modelo TSNN, se han escogido las métricas *Dice* e *IoU* muy usadas para este tipo de tareas en el estado del arte. A nivel de hardware se ha hecho uso de una tarjeta gráfica NVIDIA A100 Tensor Core GPU con 40 gb de RAM para el ajuste paramétrico de nuestro modelo neuronal TSNN. Los hiper-parámetros de ajuste escogidos que maximizan los resultados fueron: un batch size de 32, un ratio de aprendizaje de 1e-3, el optimizador Adam, la función de coste conocida como Focal, y 30 épocas de entrenamiento.

La Tabla 1 muestra los resultados experimentales obtenidos utilizando la partición de testeo de nuestro conjunto de datos con nuestra TSNN basada en DeepLabV3+, comparándolos con los obtenidos si se hace uso de otros dos modelos neuronales distintos, mencionados anteriormente. Todas las aproximaciones son capaces de segmentar las imágenes táctiles con una alta precisión, ejecutándose en tiempo real. Sin embargo, nuestra TSNN basada en DeepLabV3+ alcanza un mejor balance entre precisión de segmentación y tiempo de predicción en inferencia.

Tabla 1: Resultados de nuestra TSNN y los otros dos modelos neuronales con la partición de testeo de nuestro conjunto de datos y según las métricas Dice, IoU, y el tiempo de inferencia.

	Dice	*IoU*	*Time(s)*
TSNN(DeepLabV3+)	0.956 ± 0.013	0.914 ± 0.023	0.006 ± 0.002
PSPNet	0.951 ± 0.014	0.907 ± 0.025	0.006 ± 0.002
Unet++	0.958 ± 0.011	0.920 ± 0.020	0.009 ± 0.001

La Figura 2 muestra diferentes ejemplos de la segmentación de la región de contacto táctil realizada con nuestro modelo TSNN, mientras que la Figura 3 muestra el setup utilizado para este trabajo.

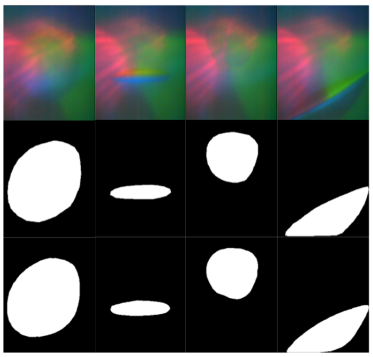


Figura 2: Ejemplos de segmentación táctil: (primero) imágenes DIGIT, (segundo) regiones de contacto etiquetadas- *ground-truth*, (tercero) regiones de contacto obtenidas por nuestra TSNN.

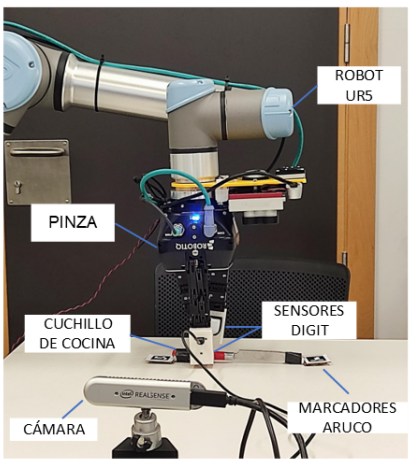


Figura 3: Setup robótico utilizado en este trabajo.

Este setup está compuesto de un brazo robótico UR5e, dos sensores táctiles basados en visión DIGIT, una pinza ROBOTIQ de tres dedos, aunque solo se utilizan dos de ellos para así generar agarres con máxima inestabilidad. También, se ha incoporado en el objeto a agarrar marcadores ArUco para calcular el ángulo de deslizamiento de referencia (que se usará como *ground-truth*) a partir de la detección y localización con una cámara Intel RealSense externa a la escena. El ángulo de referencia se calcula a partir de la posición 2D proporcionada por los marcadores visuales ArUco según la fórmula del ángulo entre dos vectores que viene representada en la ecuación 1.

$$\theta = acos(\frac{\vec{p} \cdot \vec{q}}{|\vec{p}| \cdot |\vec{q}|}) \quad (1)$$

donde $\vec{p}$ y $\vec{q}$ son los vectores actual e inicial obtenidos a partir de la posición de cada marcador ArUco, respectivamente.

La tarea objetivo, que se ha propuesto para poner a prueba nuestro método, para calcular el ángulo de deslizamiento, consiste en agarrar y levantar verticalmente un objeto. Estas acciones son repetidas para objetos de distintas características físicas, por lo tanto, se escogen diferentes puntos de agarre según el objeto.

Durante la ejecución de esta tarea, nuestra TSNN segmenta la región de contacto tras la acción de agarre y mientras se lleva a cabo la acción de levantamiento. Después, la segunda etapa de nuestro método estima el ángulo de la región de contacto. El ángulo estimado se calcula como la diferencia entre el ángulo medido en cada iteración y el ángulo referencia (en el instante inicial). Como ya se comentó, el ángulo se calcula en cada iteración aplicando un algoritmo de esqueletizado a la región de contacto obtenida por segmentación mediante la TSNN. Notar que para obtener el ángulo de referencia se han utilizando dos marcadores visuales ArUco situados en el objeto, que son observados por una cámara externa.

El método completo propuesto (Fig. 1) ha sido evaluado con siete objetos nunca vistos antes. Es decir, no fueron utilizados en las fases de ajuste, entrenamiento y validación de la TSNN. Estos objetos, etiquetados como de *1* a *7*, se pueden observar en la Figura 4. Además, otros dos objetos etiquetados como *8* y *9* y que sí son conocidos por la TSNN (es decir, fueron empleados anteriormente en las fases de entrenamiento y validación) fueron añadidos al conjunto de test.

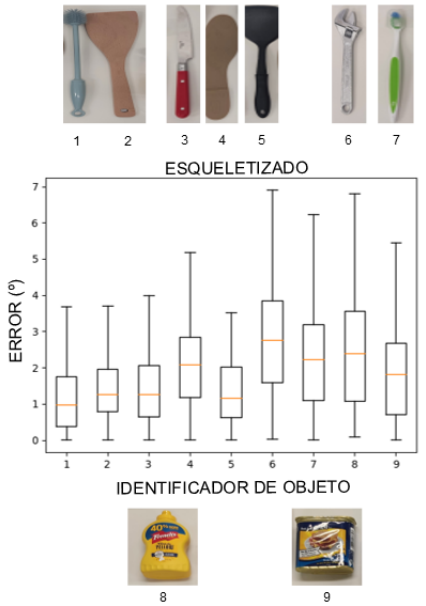


Figura 4: Errores de rotación obtenidos para cada objeto utilizando el método del esqueleto.

Aunque, a priori, se podría pensar que los objetos vistos *8* y *9* pueden proporcionar mejores resultados que los objetos no vistos *1-7*, observando la Figura 4 se desprende que esto no es así, ya que la desviación del error de los ángulos está cercana a los peores casos. Esto se debe, a que los objetos nunca vistos, tienen superficies de contacto más amplias que la superficie de contacto del sensor. Este tema se discute más adelante.

La experimentación ha consistido en realizar 45 agarres y levantamientos, 5 por cada objeto. Sin embargo, no se puede garantizar que los puntos de agarre sean iguales entre diferentes objetos ya que tienen distintas propiedades físicas como la geometría. La Figura 4 muestra el error medio rotacional calculado en grados. El resultado de las pruebas para cada objeto se muestran como un gráfico de caja, para así ver con claridad valores como los rangos intercuartiles de la distribución del error (límites de la caja), sus desviaciones (bigotes) y la mediana del error (naranja).

Cabe destacar que los objetos *6* y *8* generan un mayor error y desviación debido al mayor peso del objeto *6*, y a la mayor curvatura de la superficie del objeto *8*, la cual causa una zona de contacto táctil más grande con mayor ambigüedad y que podría llegar a saturar el sensor. Esto dificulta el proceso de esquelizado y por ende la estimación de la orientación la región.

Nuestro sistema es capaz de estimar el ángulo de deslizamiento rotacional con un error medio de 1.854º ± 0.988º, es decir, con un error medio máximo de 3º en el peor de los casos. La Figura 5 muestra algunos ejemplos de la estimación del ángulo de rotación por deslizamiento en 4 de los objetos comentados que aparecen en la Figura 4.

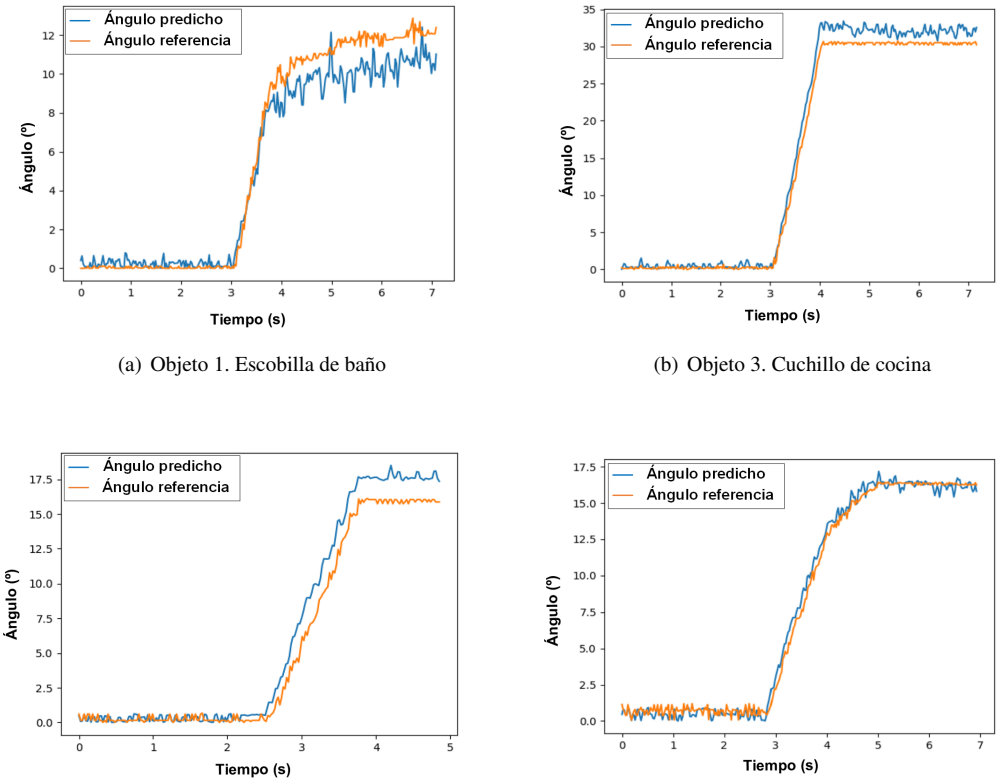
4. Conclusiones

En este trabajo se ha propuesto un método para predecir el ángulo de deslizamiento causado por un objeto que se desplaza o resbala entre los dedos durante la tarea de agarrar y levantar un objeto, cuando el agarre no es estable. El método propuesto consta de dos etapas, un modelo neuronal diseñado para segmentar la región de contacto táctil, TSNN, y una etapa de postprocesamiento para estimar la variación en ángulo de la región de contacto basada en esqueletizado. La medida de esta variación nos permite caracterizar el deslizamiento.

Nuestro método, alcanzó un error medio de **1.854º ± 0.988º**, menor que el error de otros métodos del estado del arte. Por ejemplo, en Kolamuri et al. (2021) se obtuvo **3.96º ± ND**, mientras que en Toskov et al. (2023) se mostró un error de **4.39º ± 0.18º**. Notar en cualquier caso, que esta comparativa es subjetiva, ya que no ha sido posible reproducir la experimentación de esos trabajos del estado del arte. Ya que, aunque nosotros hemos usado algunos objetos iguales desde YCB (Calli et al. (2017)), conocido conjunto de objetos usado para evaluar tareas de manipulación, los sensores empleados en estos trabajos no son los mismos y el código no es de libre acceso o es dificílmente adaptable a nuestros DIGIT.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Ciencia e Innovación a través del proyecto PID2021-122685OB-I00 y por la beca predoctoral UAFPU21-26 de la Universidad de Alicante.



(a) Objeto 1. Escobilla de baño

(b) Objeto 3. Cuchillo de cocina

(c) Objeto 4. Suela de zapato

(d) Objeto 5. Espátula de cocina

Figura 5: Ejemplos de estimación del ángulo de rotación causado por el deslizamiento del objeto durante la acción de levantamiento.

Referencias

Bauza, M., Canal, O., Rodriguez, A., 2019. Tactile mapping and localization from high-resolution tactile imprints. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE, pp. 3811–3817.
DOI: doi: 10.1109/ICRA.2019.8794298

Calli, B., Singh, A., Bruce, J., Walsman, A., Konolige, K., Srinivasa, S., Abbeel, P., Dollar, A. M., 2017. Yale-cmu-berkeley dataset for robotic manipulation research. The International Journal of Robotics Research 36 (3), 261–268.
DOI: 10.1177/0278364917700714

Castaño-Amorós, J., Páez-Ubieta, I. d. L., Gil, P., Puente, S. T., 2023. Manipulación visual-táctil para la recogida de residuos domésticos en exteriores. Revista Iberoamericana de Automática e Informática industrial 20 (2), 163–174.
DOI: 10.4995/riai.2022.18534

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818.
DOI: 10.1007/978-3-030-01234-2_49

Chi, C., Sun, X., Xue, N., Li, T., Liu, C., 2018. Recent progress in technologies for tactile sensors. Sensors 18 (4), 948.
DOI: 10.3390/s18040948

Du, G., Wang, K., Lian, S., Zhao, K., 2021. Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review. Artificial Intelligence Review 54 (3), 1677–1734.
DOI: 10.1007/s10462-020-09888-5

Guo, Z., Hall, R. W., 1989. Parallel thinning with two-subiteration algorithms. Communications of the ACM 32 (3), 359–373.
DOI: 10.1145/62065.62074

Ito, Y., Kim, Y., Obinata, G., 2014. Contact region estimation based on a vision-based tactile sensor using a deformable touchpad. Sensors 14 (4), 5805–5822.
DOI: 10.3390/s140405805

Kolamuri, R., Si, Z., Zhang, Y., Agarwal, A., Yuan, W., 2021. Improving grasp stability with rotation measurement from tactile sensing. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 6809–6816.
DOI: 10.1109/IROS51168.2021.9636488

Lambeta, M., Chou, P.-W., Tian, S., Yang, B., Maloon, B., Most, V. R., Stroud, D., Santos, R., Byagowi, A., Kammerer, G., et al., 2020. Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. IEEE Robotics and Automation Letters 5 (3), 3838–3845.
DOI: 10.1109/LRA.2020.2977257

Lambeta, M., Xu, H., Xu, J., Chou, P.-W., Wang, S., Darrell, T., Calandra, R., 2021. Pytouch: A machine learning library for touch processing. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 13208–13214.
DOI: 10.1109/ICRA48506.2021.9561084

Lin, Y., Lloyd, J., Church, A., Lepora, N. F., 2022. Tactile gym 2.0: Sim-to-real deep reinforcement learning for comparing low-cost high-resolution robot touch. IEEE Robotics and Automation Letters 7 (4), 10754–10761.
DOI: 10.1109/LRA.2022.3195195

Luo, S., Bimbo, J., Dahiya, R., Liu, H., 2017. Robotic tactile perception of object properties: A review. Mechatronics 48, 54–67.
DOI: 10.1016/j.mechatronics.2017.11.002

Toskov, J., Newbury, R., Mukadam, M., Kulic, D., Cosgun, A., 2023. In-hand gravitational pivoting using tactile sensing. In: Conference on Robot Learning. PMLR, pp. 2284–2293.

Wang, S., She, Y., Romero, B., Adelson, E., 2021. Gelsight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 6468–6475.
DOI: 10.1109/ICRA48506.2021.9560783

Zapata-Impata, B. S., Gil, P., Torres, F., 2019. Learning spatio temporal tactile features with a convlstm for the direction of slip detection. Sensors 19 (3), 523.
DOI: 10.3390/s19030523

Zhang, S., Chen, Z., Gao, Y., Wan, W., Shan, J., Xue, H., Sun, F., Yang, Y.,

Fang, B., 2022. Hardware technology of vision-based tactile sensor: A review. IEEE Sensors Journal.
DOI: 10.1109/JSEN.2022.3210210

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2881–2890.
DOI: 10.1109/CVPR.2017.660

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer, pp. 3–11.
DOI: 10.1007/978-3-030-00889-5_1