

Detección de pose de objetos usando cámaras RGB para aplicaciones industriales

Object pose detection using RGB cameras for industrial applications

Rebeca Yuan

Universidad Tecnológica Nacional Facultad Regional San Francisco - Argentina
ryuan@facultad.sanfrancisco.utn.edu.ar

Micaela Mulassano

Universidad Tecnológica Nacional Facultad Regional San Francisco - Argentina

Bruno Chiabrando

Universidad Tecnológica Nacional Facultad Regional San Francisco - Argentina

Ibrahim Jaime

Universidad Tecnológica Nacional Facultad Regional San Francisco - Argentina

Gonzalo Cervetti

Universidad Tecnológica Nacional Facultad Regional San Francisco - Argentina

Javier Redolfi

Universidad Tecnológica Nacional Facultad Regional San Francisco - Argentina
javierredolfi@sanfrancisco.utn.edu.ar

Resumen

El avance y descubrimiento de distintos materiales y componentes tecnológicos, junto con el acoplamiento de algoritmos inteligentes, impacta en forma directa en la innovación de soluciones a distintos tipos de problemas. Un ejemplo de ello, es el uso de las cámaras RGB en el sector industrial. La presente investigación busca aplicar en el sector industrial, cámaras RGB junto a algoritmos de aprendizaje profundo, para la detección de la ubicación y la pose de los objetos que circulan por cintas transportadoras, propias del proceso productivo. El objetivo es reconocer la pose de los objetos ante diferentes variables como, por ejemplo, la velocidad de la cinta transportadora y la iluminación de la planta. En paralelo al armado del dataset de entrenamiento, se exponen los posibles modelos inteligentes a utilizar para alcanzar los objetivos planteados.

Palabras clave: cámaras RGB, visión artificial, aprendizaje automático.

Abstract

The advancement and discovery of different materials and technological components, together with the coupling of intelligent algorithms, has a direct impact on the innovation of solutions to different types of problems. An

example of this is the use of RGB cameras in the industrial sector. The present research seeks to apply RGB cameras in the industrial sector, together with deep learning algorithms, for the detection of the location and pose of objects that circulate on conveyor belts, typical of the production process. The objective is to recognize the pose of the objects in the presence of different variables such as, for example, the speed of the conveyor belt and the lighting of the plant. In parallel to the assembly of the training dataset, the possible intelligent models to be used to achieve the objectives are presented.

Keywords: RGB cameras, machine vision, machine learning.

Introducción

Actividades como el control de la calidad de piezas, manejo de inventarios, ordenado de piezas resultan tareas tediosas o peligrosas, la implementación de visión por computadora está cada vez más establecida en las industrias, que buscan mejorar sus procesos [1, 2, 3, 4, 5, 6]. Estas aplicaciones se encuentran acompañadas por algoritmos de aprendizaje automático que dan soporte a la detección de objetos en las imágenes que la visión por computadora captura. Si bien existen muchos algoritmos para la detección de objetos, su uso en ambientes industriales está poco explorado. Para resolver el problema de percepción, se han utilizado muchos enfoques como segmentación, clasificación, etiquetado, etc [9]. Pero recientemente, los enfoques basados en redes neuronales se están volviendo populares gracias a que muestran resultados muy prometedores en comparación a métodos anteriores. Sin embargo, una de las desventajas es que requieren de muchos datos de entrenamiento, los cuales son difíciles de obtener [10].

En el trabajo realizado en [14] para el reconocimiento de paquetes de salchichas en la cinta de producción, se utilizó un algoritmo de detección de objetos genérico conocido como YOLO (You Only Look Once, en inglés). El algoritmo YOLO es un sistema de detección de objetos en imágenes que tienen dos etapas. Primeramente, hace uso de una única red neuronal convolucional para detectar objetos y luego plantea la detección como un problema de regresión para predecir las regiones con objetos y las probabilidades para cada clase. Recientemente, los enfoques basados en redes neuronales se están volviendo populares gracias a que muestran resultados muy prometedores en comparación a métodos anteriores. Sin embargo, una de las desventajas es que requieren de muchos datos de entrenamiento, los cuales son difíciles de obtener [10]. Además de esto, la mayoría de los trabajos se basan en el uso de sensores del tipo RGB-D [8, 10, 7, 11, 9, 12], los cuales son sensores relativamente accesibles para aplicaciones hogareñas (Microsoft KINECT, Asus XTION), pero son costosos si necesitamos que los mismos trabajen en ambientes industriales (SICK Visionari, Basler ToF camera).

Una cosa poco investigada es el uso de cámaras RGB de bajo costo o sistemas de múltiples cámaras en aplicaciones industriales. A diferencia de las cámaras industriales, son menos costosas y existen técnicas de procesamiento de imágenes muy estudiadas las cuales pueden ser usadas para resolver diferentes problemas. Además las cámaras de tipo industrial son sistemas cerrados, los cuales no pueden ser adaptados para resolver problemas particulares. Para la industria, resulta de interés investigar cómo aprovechar los datos generados por sistemas de múltiples cámaras, los cuales permitirán detectar a los objetos desde diferentes ubicaciones, mejorando la exactitud y evitando problemas de oclusiones.

El presente trabajo de investigación busca obtener el dataset que entrene el modelo a través el uso de las cámaras RGB en un ambiente industrial, para establecer la pose de la imagen con distintos algoritmos de clasificación y regresión, buscando determinar si los bosques aleatorios se corresponden al modelo que ofrece la mejor predicción.

Detección de Objetos

Con el objetivo inicial de lograr la detección de pose de objetos, se trabajó en primer lugar con el entrenamiento del modelo YOLOv3 para verificar el reconocimiento de las imágenes. Para ello se armó un conjunto de datos de entrenamiento formado por imágenes las cuales fueron acompañadas por un archivo de texto indicando las coordenadas con la posición de los objetos y la etiqueta de la clase a la que pertenece el objeto. Para nuestro caso, se trabajó con un único objeto. Se tomaron imágenes del proceso de empaquetado de salchichas. Para este etiquetado de las imágenes se utilizó el software Label Me el cual permite marcar en que parte de la imagen se encuentran los objetos de interés (Figura 1).



Figura 1 - Ejemplos de imágenes del conjunto de entrenamiento construido.

Después del entrenamiento del modelo el mismo fue evaluado sobre las 40 imágenes de test y se logró una precisión media promedio usando un valor de intersección sobre la unión de 0.5: $mAP@0.50 = 0.979592(97.96\%)$. En las 40 imágenes de evaluación hay 49 paquetes etiquetados y se obtuvieron los siguientes valores para positivos verdaderos (TP, por sus siglas en inglés), falsos positivos (FP, por sus siglas en inglés) y falsos negativos (FN, por sus siglas en inglés): $TP = 48, FP = 0, FN = 1$, concluyendo que el modelo basado en aprendizaje profundo logró la detección de paquetes en la cinta transportadora.

Planteo Detección de Pose

Actualmente nos encontramos trabajando en el armado de un dataset, con el objetivo de predecir el ángulo de rotación de un paquete con respecto a la imagen. Se busca armar un dataset que contenga la rotación del paquete de salchichas. Label Me, será la herramienta empleada para realizar la anotación poligonal de los objetos. Este dataset, estará formado por 360 imágenes correspondientes a cada ángulo de rotación, con la etiqueta correspondiente al mismo y será generado en forma sintética debido a que resulta muy tedioso realizar de forma manual estas rotaciones.

Con el afán de aplicar otro modelo de detección de pose, se plantea el uso de aprendizaje por regresión y clasificación. Se intentará aplicar el modelo que proponen los bosques aleatorios o técnica de Random Forest. Éstas se han convertido en un método popular en la visión por ordenador debido a su capacidad para manejar grandes conjuntos de datos logrando generalización, rapidez de cálculo y facilidad de implementación. [13].

El aprendizaje automático reúne distintos algoritmos para la clasificación de objetos, como la regresión logística, máquinas de soporte vectorial, clasificadores bayesianos y árboles de decisión. En un escalón más alto se encuentran los bosques aleatorios como herramienta más precisa para la clasificación. El bosque aleatorio, como su nombre lo indica, consiste en un gran número de árboles de decisión individuales que operan como un conjunto. Cada árbol individual en el bosque aleatorio evidencia una predicción de clase y la clase con más votos se convierte en la predicción de nuestro modelo. Un buen resultado se encuentra sujeto a la baja correlación entre los modelos, de tal manera que los modelos no correlacionados pueden producir predicciones de conjunto que son más precisas que cualquiera de las predicciones individuales. Esta característica hace al modelo de bosque aleatorio atractivo para la detección de pose.

Conclusiones

Aplicar distintos modelos de aprendizaje automático, permitirán completar las distintas poses que se busca obtener en la detección de objeto, no solo la identificación del mismo, sino también su posición y rotación con respecto a su ubicación en las cintas industriales.

Utilizando algoritmos de aprendizaje profundo es posible detectar objetos sobre cintas transportadoras, pero uno de los problemas es que estos algoritmos no son capaces de detectar la orientación de los mismos. Pero una de las desventajas de estos algoritmos es que necesitan muchos datos de entrenamiento para dar buenos resultados por lo tanto estos datos se están generando en forma sintética. Creemos que en base a esos datos sintéticos es posible detectar la orientación de los objetos que se trasladan por la cinta transportadora.

Por último, la utilización de cámaras RGB combinadas con modelos de aprendizaje automático, invita a la incorporación de nueva tecnología para la industria a bajos costos y con posibilidades accesibles.

Referencias

- [1] R. Yuan, I. Jaime, B. Chiabrando, and J. Redolfi, "Detección de Paquetes en Movimiento sobre una Cinta Transportadora Usando Visión por Computadora," in V Congreso Bional de la Sección Argentina del IEEE, IEEE ARGENCON, IEEE, 2020.
- [2] D. G. Dondo, J. A. Redolfi, and R. G. Araguás, "Application of deep-learning methods to realtime face mask detection," IEEE Latin America Transactions, vol. 19, no. 6, 2021.
- [3] M. Olmedo, J. A. Redolfi, D. G. Dondo, and R. G. Araguás, "Evaluación empírica de la robustez de diferentes redes neuronales usadas para la detección de objetos," Mecánica Computacional, vol. 37, no. 51, pp. 2045–2054, 2019.
- [4] Y. Hirano, C. Garcia, R. Sukthankar, and A. Hoogs, "Industry and object recognition: Applications, applied research and challenges," in Toward Category-Level Object Recognition, pp. 49–64, Springer, 2006.
- [5] M. H. Ali, K. Aizat, K. Yerkhan, T. Zhandos, and O. Anuar, "Vision-based robot manipulator for industrial applications," Procedia computer science, vol. 133, pp. 205–212, 2018.
- [6] S. Luan, Y. Li, X. Wang, and B. Zhang, "Object detection and tracking benchmark in industry based on improved correlation filter," Multimedia Tools and Applications, vol. 77, no. 22, pp. 29919–29932, 2018.
- [7] S. Kumra and C. Kanan, "Robotic grasp detection using deep convolutional neural networks," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 769–776, IEEE, 2017.
- [8] D. Guo, T. Kong, F. Sun, and H. Liu, "Object discovery and grasp detection with a shared convolutional neural network," in 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 2038–2043, IEEE, 2016.
- [9] M. Schwarz, A. Milan, A. S. Periyasamy, and S. Behnke, "Rgb-d object detection and semantic segmentation for autonomous manipulation in clutter," The International Journal of Robotics Research, vol. 37, no. 4-5, pp. 437–451, 2018.
- [10] H. Karaoguz and P. Jensfelt, "Object detection approach for robot grasp detection," in 2019 International Conference on Robotics and Automation (ICRA), pp. 4953–4959, IEEE, 2019.
- [11] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," in 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 1316–1322, IEEE, 2015.
- [12] B. Yang, H. Wen, S. Wang, R. Clark, A. Markham, and N. Trigoni, "3d object reconstruction from a single depth view with adversarial learning," in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 679–688, 2017.
- [13] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

- [14] Yuan, R., Jaime, I., Chiabrandó, B. J., & Redolfi, J. A. (2020, December). Detección de Paquetes en Movimiento sobre una Cinta Transportadora Usando Visión por Computadora. In *2020 IEEE Congreso Bienal de Argentina (ARGENCON)* (pp. 1-6). IEEE.