

Anteproyecto de Trabajo Fin de Máster

Miguel Antunes García

18 de enero de 2023

Título: Estudio de sistemas basados en cámara para la detección tridimensional de objetos en conducción autónoma

Departamento: Departamento de Electrónica

Autor: Miguel Antunes García

Tutor: Luis Miguel Bergasa Pascual

1 Introducción

El uso de sistemas de visión se ha convertido en una pieza clave en los últimos años, especialmente gracias al desarrollo de sistemas robóticos y de la inteligencia artificial. En el caso de la conducción autónoma, contar con un sistema de percepción robusto es un punto clave para el correcto desempeño del resto de las capas del vehículo.

Los tres sensores fundamentales para este tipo de tareas son: la cámara, el LiDAR y el radar. Cada uno de ellos presenta diferentes ventajas e inconvenientes, por ejemplo, la cámara presenta una resolución elevada en las imágenes a costa de perder prácticamente toda la información tridimensional, mientras que el LiDAR y el radar adquieren la información del entorno con mayor precisión representándola en nubes de puntos 3D.

Existen multitud de técnicas que utilizan la información de estos sensores para identificar elementos del entorno que sean de interés, como pueden ser otros vehículos o peatones. Los sistemas basados en Deep Learning logran obtener una precisión bastante elevada si se someten a un proceso de entrenamiento con los datos correctos. Varias de estas técnicas van a ser objeto de estudio en el trabajo que se plantea.

Gracias a que el desarrollo del trabajo se va a realizar en el grupo RobeSafe será posible realizar evaluaciones de rendimiento sobre el simulador de conducción autónoma CARLA y el vehículo autónomo del grupo.

2 Objetivos y desarrollo

El objetivo principal del trabajo es la realización de un estudio, mejora e implementación de diferentes sistemas y métodos de detección 3D basados en información principalmente de cámara.

Las imágenes de entrada a los sistemas pueden provenir de dos tipos de cámara:

1. **Cámara monocular:** conformada por una única lente, otorgando una única imagen.
2. **Cámara estéreo:** conformada por dos lentes separadas una distancia determinada, permitiendo generar dos imágenes. Permite realizar estimación de profundidad para extraer información 3D.

Para cada tipo de cámara se plantean dos líneas de trabajo en el estudio:

1. Sistemas end-to-end: a partir de la información de la cámara se realiza directamente la detección 3D de elementos del entorno. Su funcionamiento depende fuertemente del proceso de entrenamiento inicial y no tienen gran margen de configuración fuera del propio modelo.
2. Sistemas de estimación de profundidad: generan valores de profundidad a partir de las imágenes de entrada. Esta información se puede transformar para generar nubes de puntos similares a las que se utilizan en LiDAR o radar para su posterior procesamiento, aplicando técnicas de filtrado e incluso rectificación con información de otros sensores. Gracias a estas etapas es posible aplicar sistemas de detección originalmente pensados para otros sensores. Se espera que esta línea de trabajo sea la que mas tiempo abarque debido al gran margen de estudio y mejora que presenta un sistema multietapa.

Con todos los resultados de las evaluaciones, se planteará una migración de los mejores sistemas al proyecto de vehículo autónomo de Robesafe, de tal manera que se pueda comprobar su funcionamiento sobre simulación y entornos reales.

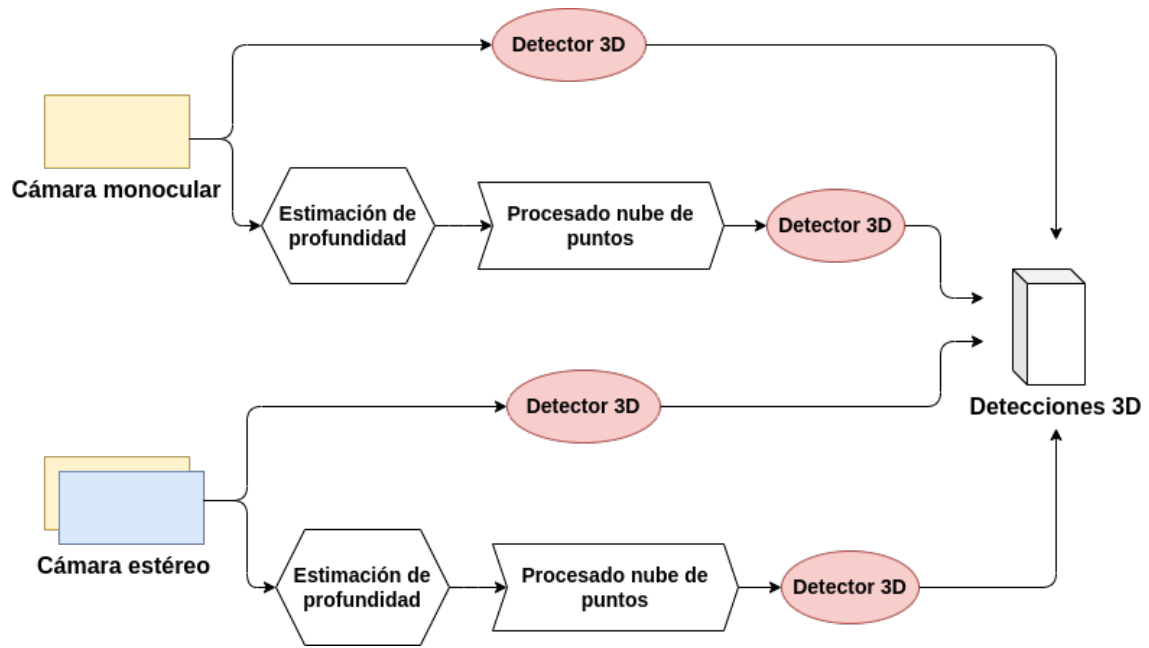


Figura 2: Esquema del funcionamiento de los diferentes tipos de sistemas a estudiar

3 Plan de trabajo

1. Descarga y adaptación de datasets.
 - KITTI[1], NuScenes[2] y AdPerDevkit
 - Creación de herramienta de conversión de datasets.
2. Estudio del estado del arte monocular: selección de modelos end-to-end para detección y de estimación de profundidad.
3. Entrenamiento y evaluación de modelos monoculares (métricas de detección y profundidad).
4. Estudio del estado del arte estéreo: selección de modelos end-to-end y de estimación de profundidad.
5. Entrenamiento y evaluación de modelos estéreo (métricas de detección y profundidad).
6. Generación y procesamiento de nube de puntos a partir de los modelos de profundidad.
7. Proceso de fine-tuning del detector con las nubes generadas.
8. Implementación y evaluación (métricas de detección) de la línea de trabajo completa.
9. Migración al proyecto y evaluación.

4 Material y herramientas

El proyecto se desarrollará utilizando material del laboratorio del grupo RobeSafe, concretamente se utilizará un PC que cumpla con los requerimientos para los procesos de entrenamiento:

- CPU Intel i7-9700K.
- GPU NVIDIA GeForce GTX 1080 Ti.
- Sistema operativo Linux.
- Discos de almacenamiento para albergar los datasets.
- Entorno de desarrollo Visual Studio Code.
- Control de versiones git.
- Lenguaje python con librerías requeridas.

Referencias

- [1] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [2] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “nusenes: A multimodal dataset for autonomous driving,” in *CVPR*, 2020.