

Mechanism for pedestrian detection for vehicle automation

Mecanismo para detección de peatones para la automatización de vehículos

Celso Alejandro Maldonado Mendoza^{*1} and Juan Antonio Guerrero Ibáñez¹

¹Universidad de Colima, Maestría en Tecnologías del Internet,

Facultad de Telemática - Universidad de Colima, Av. Universidad 333. Col. Las Víboras, Colima, Col. México, 28040.

{cmaldonado0,antonio_guerrero}@uclm.mx

Abstract

In recent years, vehicle automation has been one of the activities on which the automotive industry has focused on with the objective of creating a more friendly and safer driving environment for drivers, passengers, and pedestrians. This article focuses on one automation task: pedestrian detection. A computer vision-based mechanism for pedestrian detection within a road driving environment is presented. The results obtained by applying the mechanism showed a high accuracy in the detection and identification of pedestrians, which contributes to the vehicle automation process.

Keywords— Computer vision, Deep learning, pedestrian detection.

Resumen

En los últimos años, la automatización de vehículos ha sido una de las actividades en la cual la industria automotriz se ha centrado con el objetivo de crear un entorno de conducción más amigable y seguro para conductores, pasajeros y peatones. Este artículo se centra en una tarea de automatización: la detección de peatones. Se presenta un mecanismo basado en visión computacional para detección de peatones dentro de un entorno de conducción vial. Los resultados obtenidos al aplicar el mecanismo mostraron una alta exactitud en la detección e identificación de los peatones, lo que contribuye al proceso de automatización de vehículos.

Palabras clave— Visión computacional, redes neuronales, detección de peatones.

^{*}Autor de correspondencia

I. Introducción

Uno de los problemas más importantes que enfrenta la sociedad moderna en nuestros días es el tráfico vial. De acuerdo con la Organización Mundial de la Salud, cada año mueren cerca de 1.3 millones de personas como consecuencia de accidentes de tránsito a nivel mundial [1]. Además menciona que estos accidentes generan un serio impacto en la economía ya que para muchos países estos accidentes les cuestan cerca del 3 % del producto interno bruto [2]. Se estima que entre el 71 % y el 91 % de estos accidentes son ocasionados por errores humanos [3].

Esta situación ha planteado la necesidad de mecanismos que ayuden a disminuir el número de accidentes viales y que al mismo tiempo genere un entorno de tráfico más seguro para todos sus usuarios. Es por esto que la industria automotriz está enfocando sus esfuerzos en generar vehículos cada día más automatizados que asistan a las personas en la actividad de conducción. El objetivo es llegar a la automatización completa de los vehículos. La automatización de vehículos está centrada en la reducción de accidentes de tráfico, la optimización de la congestión de tráfico, entre otras funciones y se visualiza como un conjunto de subtarefas incluyendo la localización y el mapeo, la detección de objetos estáticos y dinámicos, así como la detección y predicción de la trayectoria de movimiento.

De esta manera los esfuerzos están dirigidos a generar de forma paulatina vehículos con un cierto nivel de automatización hasta lograr crear un coche 100 % automatizado, que tenga capacidad de moverse y transportar de forma segura, sin la intervención de un humano. Parte del esfuerzo lo ha desarrollado el SAE (Society of Automotive) que ha definido un conjunto de niveles de

autonomía de los vehículos de acuerdo al grado de participación del conductor en el proceso de conducción como se muestra en la Fig. 1 [4].



Figura 1: Niveles de autonomía de un vehículo

A continuación se explican de forma general los niveles de automatización del vehículo:

- El nivel 0 hace referencia a que el conductor tiene control completo del auto, por lo que no tiene ningún tipo de asistencia por parte de los sistemas del vehículo.
- En el nivel 1 se especifica que el conductor aunque el conductor sigue teniendo el control sobre el coche, puede ser asistido por el vehículo en ciertos escenarios específicos como control de crucero.
- El nivel 2 muestra un escenario donde el vehículo tiene capacidades para controlar la dirección y el desplazamiento del coche, por lo que el conductor podría quitar las manos del volante o quitar el pie del pedal en algún momento.
- El nivel 3 hace referencia a que el coche puede tomar algunas decisiones como cambio de carril, control de frenado y tiene la capacidad de detectar objetos.
- El nivel 4 hace referencia que en varias situaciones el conductor ya no es necesario ya que el vehículo comienza a controlar el tráfico y las condiciones del entorno. Sin embargo, el conductor puede retomar el control bajo solicitud.
- Finalmente, el nivel 5 es conocido como completamente automatizado, el vehículo tendrá la capacidad de realizar todas las tareas de conducción sin la necesidad de que un conductor esté presente.

Uno de los elementos claves para generar la automatización de los vehículos es la percepción. La percepción hace referencia a la capacidad del vehículo para recolectar información relevante que le facilite el aprendizaje del entorno en el que se está desplazando. Esta tarea le permite al vehículo tener una visión general y entendimiento total del entorno de conducción, de tal forma que proporcione las capacidades al vehículo de detectar y clasificar los objetos, estimar su posición, localización y mapeo simultáneo de objetos [5]. La percepción es considerada como un proceso crucial ya que la toma de decisiones se realizará basada en la información de todos

los datos recolectados por el proceso de percepción y definirán el comportamiento del vehículo ante situaciones detectadas (tráfico, peatones, señales de tráfico, entre otras).

Este artículo forma parte del proyecto DetoAVE, plataforma para Detección de Objetos aplicado a la Automatización de Vehículos. En este trabajo nos centramos en un algoritmo basado en visión computacional para la detección de peatones en un entorno de conducción vial.

El resto del artículo está organizado de la siguiente manera: la sección II hace un breve análisis de los trabajos relacionados con el objeto de este desarrollo. La sección III presenta de forma general lo que es la percepción ambiental de un vehículo. Una explicación general de la visión computacional es presentada en la sección IV. La sección V describe la propuesta de este artículo, mostrando todo el proceso para la detección de peatones. Los resultados preliminares de este trabajo se presentan en la sección VI. Finalmente terminamos el artículo con las conclusiones del trabajo en la sección VII.

II. Trabajo relacionado

La detección y rastreo de peatones en un entorno vehicular ha sido de gran interés para la comunidad científica y la industria automotriz. En [6] y [7] los autores se centran en la taxonomía de funcionalidades de sistemas considerando la estructura de un sistema de captura de movimiento y la diferente información que se debe de procesar. Otro trabajo presenta un estudio sobre los sistemas de detección de peatones enfocándose en las metodologías para la selección de las regiones de interés (ROIs), los métodos de clasificación y rastreo [8].

En [9], los autores presentan los métodos principales para la detección de peatones en imágenes y en [10] se presenta un trabajo que se centra en las arquitecturas deep para la detección de peatones.

Entre toda la variedad de objetos que se pueden identificar, la detección de humanos es de mucho interés debido a que los procesos de detección y segmentación de humanos en imágenes y videos son difíciles por la gran cantidad de condiciones y variables a considerar para esta tarea (ruido, distancia, forma, entre otros) [11]. En [12] se describe un sistema de detección de peatones que integra información de intensidad de la imagen con detalles de movimiento, combinan las características Haar-like [13] con información de movimiento los cuales se procesaron considerando dos frames de video consecutivos. Para la detección de peatones aplicaron detectores de rostros y para la clasificación consideraron una secuencia de AdaBoost.

III. Percepción ambiental

Una de las funciones esenciales de la automatización de los coches es la percepción ambiental (Fig. 2). El enfoque de la percepción radica en recopilar toda la información relevante que esté ocurriendo alrededor del coche cuando se está desplazando por las calles. La información recolectada permite la detección y localización de objetos (tales como señales de tráfico, peatones, ciclistas, motociclistas, otros vehículos, objetos como conos de señalización, entre otros), la movilidad y dirección de desplazamientos de dichos objetos y de esta forma poder realizar predicciones de sus estados futuros.

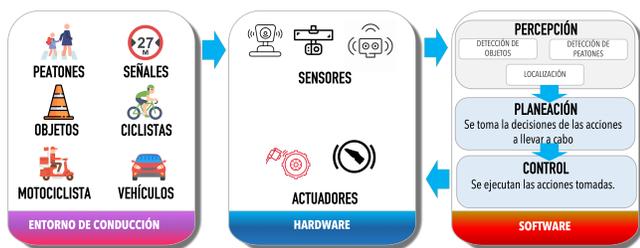


Figura 2: Proceso de percepción ambiental

Para realizar esta función los coches hacen uso de una serie de sensores instalados por todo el vehículo, como se muestra en la Fig. 3. Como se puede observar en la figura, existen diferentes tipos de sensores instalados en el vehículo. Generalmente, los sensores más utilizados para desempeñar la tarea de percepción son el LiDAR, las cámaras y los sensores radar de corto y amplio rango.

Toda la información recolectada es procesada por medio de software y permite la identificación de diferentes tipos de objetos. La información se envía a un proceso de planeación que tiene como objetivo tomar las decisiones más adecuadas de acuerdo a la situación detectada. Finalmente, la función de control hace uso de una serie de comandos que se envían a diferentes actuadores para ejecutar las acciones definidas acordes a la situación detectada.

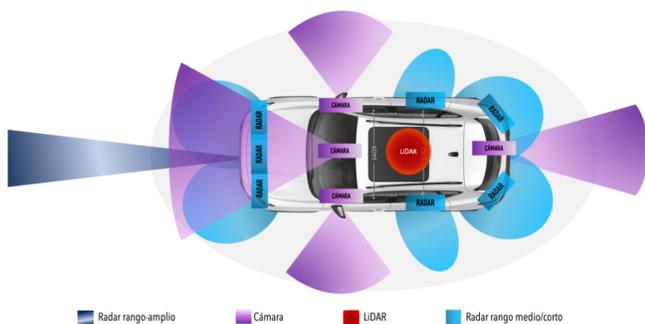


Figura 3: Sensores utilizados para la percepción del vehículo

IV. Visión computacional

De acuerdo a [14], en los últimos años el término de visión computacional ha tomado mucho impulso. La visión computacional o visión por computadora es el proceso mediante el cual se extrae de forma automática información útil contenida en imágenes. Su objetivo es sustraer y transmitir esa información a la computadora de forma comprensible. La visión computacional tiene sus bases en la visión humana y hace uso de software y hardware robusto para procesar toda la información. Su aplicación principal es la detección de objetos. Para esto hace uso de varias tareas como reconocimiento, identificación, detección, movimiento, entre otras.

En forma general, el proceso de detección de objetos basado en visión computacional se conforma de tres pasos fundamentales, los cuales se presentan en la Fig. 4.



Figura 4: Pasos para identificación de peatones

Mediante la cámara se va recolectando información del entorno que se desea monitorear. Las imágenes se envían a la computadora donde se utilizan algoritmos sofisticados para extraer las características principales de los objetos que se detectan en el entorno. Finalmente, se realiza una fase de clasificación para separar los diferentes tipos de objetos y poder tomar las decisiones adecuadas de acuerdo con la actividad a realizar.

V. Descripción de la propuesta

La presente propuesta se centra en la detección de peatones en tiempo real mediante el uso de visión computacional. En específico, los peatones detectados son los que se encuentren en el ángulo de frontal de visión del vehículo. Si bien es cierto que este artículo solamente se centra en el proceso de detección e identificación de peatones, la idea es detectar peatones que permitan posteriormente ser analizados y poder predecir su comportamiento con el objetivo de evitar un accidente.

Para poder detectar a los peatones se definió un mecanismo que consta de varias etapas las cuales se muestran en la Fig. 5. A continuación se describen de forma general cada una de las etapas que conforman el mecanismo propuesto.

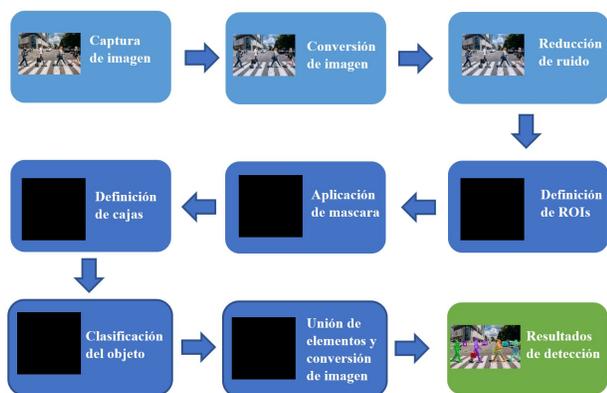


Figura 5: Proceso para detección de peatones

V.1. Captura de imagen

Como primer paso se realiza la adquisición de imágenes. El proceso básicamente consiste en recopilar información del entorno mediante la captura de imágenes. En este caso, los datos que se utilizan de entrada son una secuencia de imágenes a color las cuales son capturadas desde la cámara del vehículo. En este caso, la cámara se colocaría en el espejo retrovisor en la línea central. Toda la información que se captura se procesa por la computadora en tiempo real y es almacenada en el disco duro. Como parte del proceso, el video se divide en fotogramas (Fig. 6).

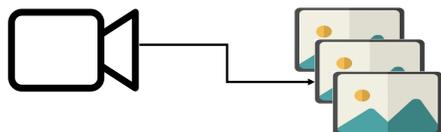


Figura 6: Obtención de fotogramas a partir de videocámara

V.2. Conversión de imagen

Uno de los grandes problemas a los que se enfrenta es el procesamiento de imágenes a color ya que implica muchas dificultades para la detección de bordes y el tiempo de procesamiento. Debido a esta situación, las imágenes se convierten a escala de Red Green Blue (RGB) para un mejor procesamiento además de que el tamaño de la imagen en píxeles debe ser por lo menos divisible 6 veces entre sí.

V.3. Reducción de ruido

Uno de los problemas más comunes que sufren los sistemas es el ruido. La presencia de ruido dentro de la

imagen dificulta la detección correcta de los bordes, por lo que es necesario que se apliquen mecanismos para eliminar el ruido. El objetivo de este proceso es suavizar la imagen para facilitar la detección del mayor número de bordes. Para lograr el objetivo se aplica un filtro Gaussiano que elimina el ruido de la imagen que pudiera generar un borde falso. Cada imagen se almacena como una colección de píxeles discretos a escala de BGR los cuales se representan mediante un número que representa la brillantez del píxel. Para suavizarlo se modifica el valor del píxel con el valor promedio de la intensidad de los píxeles a su alrededor. El promedio se realiza mediante el kernel. El kernel se configuró a un tamaño de 5×5 y se aplica a todos los píxeles de la imagen.

V.4. Detección de objetos

Para el proceso de detección de objetos se hace uso de las técnicas Mask RCNN (Region based Convolutional Neural Network) y Backbone. Por un lado, la técnica RCNN es una variante de una red neuronal profunda que detecta regiones delimitándolas mediante recuadros y máscaras de segmentación para cada instancia de un objeto en la imagen. Mask RCNN define dos etapas para realizar el proceso (Fig. 7). En la primera etapa, la técnica genera una serie de regiones donde podría haber objetos con base en la imagen de entrada. Como segunda etapa, el mecanismo predice la clase de objeto y mejora el recuadro delimitador y finalmente crea la máscara correspondiente para el objeto con base en la primera etapa. Éstas etapas se conectan al backbone.

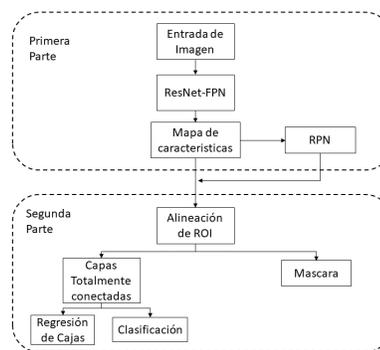


Figura 7: Arquitectura Mask R-CNN

Por otro lado, la técnica Backbone es una red neuronal profunda estilo FPN (Red neuronal convolucional, construida en forma de pirámide). Consiste en un camino de abajo hacia arriba, un camino de arriba a abajo y conexiones laterales.

Para identificar el perímetro de las formas se realiza un análisis de contraste de forma entre la superficie de la carretera y la superficie no pavimentada. Con esto se

reduce la cantidad de datos de aprendizaje requeridos y se simplifica la imagen considerablemente. Para llevar a cabo esta actividad se utilizó el método de Canny el cual es un algoritmo de múltiples etapas que permite detectar una amplia gama de bordes en imágenes.

V.5. Detección de área de interés

Debido a que dentro del escenario de conducción existen diferentes objetos a parte de los peatones, lo que se hace es aplicar un mecanismo que permita ignorar los objetos no deseados del escenario de conducción. Cada fotograma resultante pasa por un proceso llamado *extract region proposals*, que en términos simples divide cada fotograma en secciones en las que sea posible encontrar un objeto (ROIs), basándose en la zona de la imagen que tiene el mismo color, saturación, entre otros (Fig. 8).

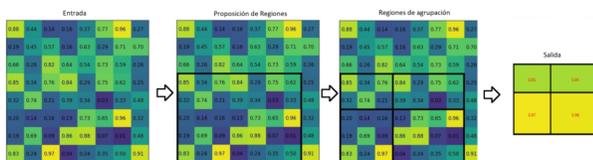


Figura 8: Ejemplo del proceso de *extract region proposals*

Aplicación de máscara, definición de cajas y clasificación del objeto

En cada píxel del objeto encontrado se le agrega una capa de color llamada máscara, la cual es el diferenciador principal de Mask RCNN respecto a otras redes neuronales de detección de objetos (Fig. 9).

El fotograma resultante pasa por un proceso para definir la caja que circula a cada objeto detectado en la figura del proceso (*extract region proposals*).

Dentro de cada caja se clasifica el objeto de acuerdo con su nombre, es decir si Mask RCNN detecta un peatón este debe ser clasificado como un peatón.

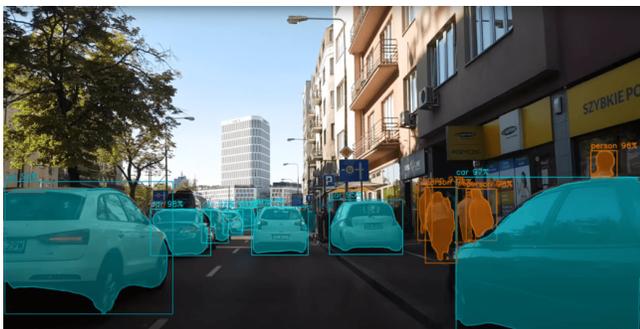


Figura 9: Ejemplo de aplicación de máscara definición de cajas y clasificación del objeto

Todos los procesos vistos van anclados al backbone, al final todas las partes de este proceso son acoplados

en una imagen que engloba la clasificación del objeto, la máscara, la definición de la caja y un porcentaje de probabilidad de que sea el objeto identificado.

VI. Resultados preliminares

Para la implementación del algoritmo se utilizó el lenguaje de programación Python en conjunto con la librería TensorFlow. TensorFlow [14] es una interfaz para expresar algoritmos de aprendizaje automático y una implementación para ejecutar dichos algoritmos. Un cálculo expresado con TensorFlow se puede ejecutar con poco o ningún cambio en una amplia variedad de sistemas heterogéneos, que van desde dispositivos móviles como teléfonos y tabletas hasta sistemas distribuidos a gran escala de cientos de máquinas y miles de dispositivos computacionales como tarjetas GPU. El sistema es flexible y se puede usar para expresar una amplia variedad de algoritmos, incluidos algoritmos de entrenamiento e inferencia para modelos de redes neuronales profundas, y se ha utilizado para realizar investigaciones y para implementar sistemas de aprendizaje automático en producción en más de una docena de áreas de informática y otros campos, incluido el reconocimiento de voz, la visión por computadora, la robótica, la recuperación de información, el procesamiento del lenguaje natural, la extracción de información geográfica y el descubrimiento computacional de fármacos.

Para este proyecto, se utilizó un video de alrededor de 4 minutos donde se mostraba la visión de la cámara durante el proceso de conducción. En tiempo real se aplicó el algoritmo y se visualizó la identificación de las líneas del carril de la carretera. La Fig. 10, muestra una de las imágenes que se utilizaron como entrada para la detección de peatones principalmente. Finalmente, en la Fig. 11 se puede observar el resultado del proceso completo de detección de peatones.



Figura 10: Imagen de peatones antes de procesar

