Reconstrucción 3D basada en técnicas de Disparidad U-V en sistemas de visión estéreo. Análisis de robustez y aplicaciones prácticas.

Jonatán Felipe García

Dirigida por: Dr. Leopoldo Acosta Sánchez
Dra. Marta Sigut Saavedra

Departamento de Ingeniería Informática y de Sistemas
Universidad de La Laguna
Programa de Doctorado en Ingeniería Industrial, Informática y Medioambiental

Memoria para la obtención del grado de
Doctor por la Universidad de La Laguna

abril 2024

A Esther, Mario y Diego.

Agradecimientos

Agradezco profundamente el apoyo incondicional de Esther, Mario y Diego, cuyo amor, comprensión y paciencia fueron fundamentales durante todo este proceso.

Quiero expresar mi más sincero agradecimiento a mis padres por su constante aliento, inspiración y sacrificio, que me han permitido alcanzar este logro académico.

Agradezco a mis estimados compañeros del Departamento de Ingeniería Informática y de Sistemas y de los diferentes proyectos de investigación bajo cuyo paraguas se fraguó esta investigación por su colaboración, amistad y el ambiente de trabajo enriquecedor que han creado. En especial al Dr. Pedro A. Toledo Delgado por aportar un rayo de luz en los momentos más oscuros.

Mi gratitud se extiende a mis directores de tesis, Marta Sigut Saavedra y Leopoldo Acosta Sánchez, por su orientación experta, dedicación y apoyo inquebrantable a lo largo de esta investigación.

También quiero agradecer a mis compañeros de trabajo del Instituto Tecnológico y de Energías Renovables por su colaboración, apoyo y comprensión, que han sido un pilar fundamental en este camino hacia la culminación de esta tesis doctoral. Y al director del Área de Tecnología, Jesús F. Rodríguez Álamo por tener confianza en que alcanzara este objetivo.

A mis amigos y colegas, por su amistad y colaboración a lo largo de este viaje.

¡Trata de arrancarlo!
Luis Moya. 24 de noviembre de 1998.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

En este capítulo se explica, en primer lugar, cómo y cuándo surge la idea de llevar a cabo el trabajo que se recoge en la presente memoria. A continuación se presenta un resumen de los principales aspectos de esta tesis doctoral, tanto en español como en inglés, así como los objetivos y principales contribuciones de la misma. Finalmente, se expone el rendimiento científico de la tesis y se explica cómo se ha estructurado el presente documento.

1.1. Motivación

El germen de esta tesis doctoral nace en torno al desarrollo del prototipo Verdino, un vehículo autoguiado diseñado para circular por una urbanización cerrada. El proyecto de urbanización de 25 viviendas bioclimáticas en el ITER (Instituto Tecnológico y de Energías Renovables) lleva aparejado una serie de acciones paralelas para dotar de servicios a los residentes y visitantes de esta urbanización. Uno de estos servicios consiste en proporcionar un sistema de transporte interno que permita un acceso desde el centro de visitantes hacia las viviendas, teniendo en cuenta que pueden darse distancias mayores a un kilómetro entre este centro y las viviendas más distantes al mismo. Fieles al compromiso con el medio ambiente y el uso de energías renovables, se pretende que el sistema de transporte permita garantizar un nivel de emisión cero de CO_2 a partir de la utilización de vehículos eléctricos.

Partiendo de la premisa del uso exclusivo de este tipo de vehículos para el transporte interno en la urbanización, se plantea la posibilidad de permitir que puedan transportar a pasajeros aun cuando éstos no sepan o no puedan conducirlos. Una solución podría consistir en que el ITER tuviera una plantilla de conductores, siempre a disposición de los visitantes y residentes, para realizar el transporte en caso necesario. Otra solución, más acorde con las características del Instituto, consistiría en dotar a estos vehículos de un sistema de conducción automática que permitiese que los mismos pudiesen realizar los trayectos necesarios sin intervención humana, es decir, planteando el uso de vehículos autoguiados para el transporte de personas dentro de la urbanización.

El entorno del ITER en el que se planteaba que se desenvolvieran los vehículos autoguiados se puede clasificar en lo que ha sido denominado como "carreteras no estructuradas" por autores como Crisman and Thorpe (1991). Lo que caracteriza a este tipo de carreteras es la ausencia de líneas que delimitan los bordes de la misma, bordes degradados, posibles defectos en la superficie del asfalto, problemas causados por sombras y la no disponibilidad de un mapa del camino.

En este marco se planteó la colaboración del ITER con un grupo de investigadores del Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna para el desarrollo del sistema de transporte autoguiado. En el seno de esta colaboración se han concedido múltiples proyectos de investigación dentro del programa nacional de I+D. El primero de ellos, desarrollado entre los años 2004 y 2007, y que lleva por título GUIado de un Sistema de Transporte en una Urbanización Bioclimática cerrada (GUISTUB), se centró en la adaptación del vehículo para poder conducirlo de manera automática y la incorporación de los primeros sensores. Este proyecto tuvo su continuación en el proyecto Sistema Inteligente de Bajo coste para el TRAnsporte y la vigilancia en entornos ecológicos no estructurados (SIBTRA) que finalizó en el año 2010. Como resultado del segundo proyecto el prototipo Verdino disponía de un sistema sensorial más completo y diferentes mejoras en las adaptaciones electromecánicas del vehículo. Estos proyectos fueron sucedidos por Sistema Autónomo con Gestión de la Energía para la Navegación en Entornos Dinámicos con Inteligencia Ambiental (SAGENIA) y Hacia un Sistema de Transporte Inteligente en Urbanizaciones y Recintos Peatonales (STIRPE), finalizando este último en 2017. Estos dos últimos proyectos se focalizaron en hacer uso de la plataforma desarrollada y explotar el sistema sensorial para que el vehículo fuera capaz de circular en entornos con peatones y llevar a cabo tareas distintas a la del transporte de personas como, por ejemplo, la vigilancia de un recinto.

Tomando como referencia las fechas en las que se comenzó a desarrollar el prototipo, es importante tener en cuenta los proyectos punteros en el campo de vehículos autoguiados de aquel momento. Una referencia importante a nivel internacional fue el Grand Challenge, organizado por el Defense Advanced Research Projects Agency (DARPA) del Departamento de Defensa de Estados Unidos. El objetivo de esta competición era acelerar la investigación y desarrollo de vehículos terrestres, con el fin de utilizar esta tecnología en el campo de batalla. La prueba en entornos no estructurados, que se celebró por última vez en el año 2005, consistía en realizar un recorrido por el desierto definido por una serie de puntos de control con la restricción de no poder dañar el entorno o las infraestructuras a lo largo del camino, lo que obligaba a los participantes a incorporar sistemas de detección de obstáculos. La siguiente edición, conocida como 2007 Urban Challenge, cambió el entorno en el que debían desenvolverse los vehículos. De este modo se introdujeron entornos estructurados al incluir un circuito urbano en el que los participantes debían obedecer las normas de tráfico y comportarse responsablemente en presencia de tráfico.

A nivel nacional, son numerosos los grupos de investigación que en aquel momento trabajaban en el campo de la navegación autoguiada. Dos de las contribuciones más destacadas fueron el vehículo autoguiado ROMEO de la Universidad de Sevilla y el programa AUTOPÍA, desarrollado por investigadores del Centro Superior de Investigaciones Científicas. El proyecto AUTOPÍA [Rosa et al. (2003)] contaba con una pequeña flota de vehículos adaptados (dos furgonetas eléctricas Berlingo y dos C3 Pluriel de combustión) en un circuito cerrado que imitaba las calles de una urbanización y centraba sus objetivos en el guiado de los vehículos en entornos estructurados. Por su parte, los prototipos ROMEO-3R y ROMEO-4R [Ollero et al. (1999)] se construyeron sobre la base de vehículos eléctricos como los utilizados en los campos de golf. A estos prototipos se les dotó de un sistema de actuación y sensado para poder navegar de manera autónoma en entornos exteriores.

El vehículo escogido para el desarrollo del prototipo Verdino también es un coche eléctrico de los que habitualmente se usan para transitar por campos de golf al que se le realizaron las adaptaciones necesarias para poder controlarlo desde un ordenador. Del mismo modo, el vehículo se dotó de una serie de sensores tanto para el posicionamiento como para el reconocimiento del entorno, ya que ambos aspectos son necesarios de cara a la navegación autoguiada.

Figura 1.1: Fotografía del prototipo Verdino frente a la Escuela Superior de Ingeniería y Tecnología de la ULL.

Como parte del equipamiento sensorial se incluyó un sistema de visión estereoscópica con el objetivo de extraer información tridimensional del entorno. Tomando este problema como punto de partida, se definieron los primeros objetivos de trabajo que, tras varias evoluciones, desembocaron en los resultados que se presentan en esta tesis doctoral.

1.2. Resumen de la tesis

La recuperación de información tridimensional del entorno es un problema que ha sido ampliamente abordado desde diferentes perspectivas y haciendo uso de múltiples técnicas. La reconstrucción de escenas 3D basada en estereovisión es una de ellas y en la bibliografía existen múltiples aproximaciones tanto al tratamiento de la información de las cámaras como al tipo de información que se obtiene tras su procesamiento. En esta tesis se ha optado por un enfoque que utiliza técnicas basadas en Disparidad U-V para recuperar la información tridimensional de un entorno, diferenciar los diferentes elementos que componen la escena y representarlos de manera simplificada mediante un plano con la orientación del elemento. Es decir, a partir de una pareja de imágenes de una escena se obtiene un conjunto de planos de los cuales cada uno de ellos representa un elemento presente en la escena.

Una vez reconstruida la escena, se lleva a cabo un análisis de sensibilidad para determinar cómo afectan posibles errores en el ángulo de cabeceo de las cámaras a los planos resultantes de la reconstrucción. El objetivo es determinar la robustez de la técnica empleada, es decir, la capacidad para mantener su eficacia en condiciones no ideales, y evaluar su aplicabilidad práctica. El estudio realizado ha permitido concluir que la desviación experimentada por los planos resultantes de la reconstrucción de la escena como consecuencia de un error en el ángulo de cabeceo del sistema de visión no es nunca superior a dicho error.

La aplicación práctica de la reconstrucción de escenas 3D basada en estereovisión va más allá de la mera generación de modelos tridimensionales. Por ello se han explorado aplicaciones en dos ámbitos bastante diferenciados. Por una parte se ha desarrollado una metodología de calibración del ángulo de cabeceo de un sistema estereoscópico que permite corregir las desviaciones existentes entre el ángulo real y el que reporta el sistema de posicionamiento de las cámaras. Dicho sistema es independiente de las técnicas que se quieran aplicar a las imágenes obtenidas del par estéreo, por lo que se trata de una herramienta de gran versatilidad. Por otro lado, se han realizado una serie de simplificaciones sobre el sistema de reconstrucción para adaptarlo a la detección de obstáculos en tiempo real, particularizado para el uso en un vehículo autoguiado.

En resumen, en esta tesis doctoral se presentan los resultados de la investigación llevada a cabo en el campo de la reconstrucción de escenas 3D basada en estereovisión y empleando Disparidad U-V. Dicha investigación se complementa con un exhaustivo estudio de robustez y un análisis detallado de dos aplicaciones (calibración del sistema de visión y detección de obstáculos), revelando así su potencial para contribuir al campo de la visión computacional y la percepción tridimensional.

1.3. Abstract

The recovery of three-dimensional information from the environment is a problem that has been widely approached from different perspectives and using multiple techniques. The reconstruction of 3D scenes based on stereovision is one of them and in the literature there are multiple approaches to both the treatment of the information from the cameras and the type of information obtained after processing. In this thesis we have opted for an approach that uses techniques based on U-V Disparity to recover the three-dimensional information of an environment, differentiate the different elements that make up the scene and represent them in a simplified way by means of a plane with the orientation of the element. In other words, from a pair of images of a scene, a set of planes is obtained, each of which represents an element present in the scene.

Once the scene has been reconstructed, a sensitivity analysis is carried out to determine how possible errors in the pitch angle of the cameras affect the resulting reconstruction planes. The aim is to determine the robustness of the technique used, i.e. the ability to maintain its effectiveness under non-ideal conditions, and to assess its practical applicability. The study carried out has led to the conclusion that the deviation experienced by the planes resulting from the reconstruction of the scene as a consequence of an error in the pitch angle of the vision system is never greater than this error.

The practical application of stereovision-based 3D scene reconstruction goes beyond the mere generation of three-dimensional models. For this reason, applications have been explored in two quite different areas. On the one hand, a methodology has been developed for calibrating the pitch angle of a stereoscopic system to correct the existing deviations between the real angle and the angle reported by the camera positioning system. This system is independent of the techniques to be applied to the images obtained from the stereo pair, making it a highly versatile tool. On the other hand, a series of simplifications have been made to the reconstruction system to adapt it to the detection of obstacles in real time, particularly for use in a self-guided vehicle.

In summary, this doctoral thesis presents the results of the research carried out in the field of 3D scene reconstruction based on stereovision and using U-V Disparity. This research is complemented with an exhaustive robustness study and a detailed analysis of two applications (vision system calibration and obstacle detection), thus revealing its potential to contribute to the field of computer vision and 3D perception.

1.4. Objetivos y principales contribuciones

En esta tesis se plantea la recuperación de información del entorno haciendo uso de sensores pasivos que no interfieran en el mismo. Por ello, se plantea como objetivo principal el desarrollo de un método de reconstrucción de escenas tridimensionales a partir del uso

de las imágenes obtenidas de las cámaras de un par estéreo. Para alcanzar este objetivo se han abordado diferentes aspectos del problema, que van desde el diseño del sistema de visión hasta la formulación del método de reconstrucción mediante el uso de técnicas basadas en la Disparidad U-V.

Además del objetivo principal, durante el desarrollo de la tesis se han establecido una serie de objetivos secundarios:

- Estudio de la robustez del método de reconstrucción propuesto frente a la posible aparición de errores en el ángulo de cabeceo de las cámaras.
- Desarrollo de aplicaciones prácticas del sistema de reconstrucción, más allá del propósito con el que fue concebido:
  - Un sistema de calibración del ángulo de cabeceo de un sistema de visión estéreo.
  - Un sistema de detección de obstáculos en tiempo real mediante el empleo de una versión simplificada del método de reconstrucción de escenas.

Teniendo en cuenta los objetivos presentados, las principales aportaciones de la tesis se pueden resumir en los siguientes puntos, que se desarrollarán a lo largo de la presente memoria:

- Contribución a la teoría de reconstrucción de escenas 3D y, en particular, a la relacionada con estereovisión introduciendo una forma diferente de obtener y representar el contenido de una escena haciendo uso de planos.
- Contribución a la teoría de reconstrucción de escenas 3D mediante el análisis exhaustivo de la sensibilidad del sistema frente a la presencia de errores en el ángulo de cabeceo de las cámaras.
- Contribución al campo de la estereovisión, presentando un sistema de calibración genérico del ángulo de cabeceo de un sistema estereoscópico, lo que lo hace independiente del uso y las técnicas de procesamiento que se quieran realizar a posteriori con las imágenes.
- Contribución al ámbito de detección de obstáculos al introducir un sistema capaz de funcionar en tiempo real que determina la distancia y ubicación estimadas del obstáculo.

1.5. Rendimiento científico de la tesis

El trabajo llevado a cabo en esta tesis doctoral ha dado lugar a dos publicaciones en revistas de impacto que se enumeran a continuación.

- Felipe, J.; Sigut, M.; Acosta, L. Influence of Pitch Angle Errors in 3D Scene Reconstruction Based on U-V Disparity: A Sensitivity Study. Sensors 2022, 22, 79. https://doi.org/10.3390/s22010079
- Felipe, J.; Sigut, M.; Acosta, L. Calibration of a Stereoscopic Vision System in the Presence of Errors in Pitch Angle. Sensors 2023, 23, 212. https://doi.org/10.3390/s23010212

1.6. Estructura de la memoria

La tesis se compone de ocho capítulos. En este primer capítulo, se ha introducido el ámbito en el que contextualiza la investigación, presentando los objetivos y las contribuciones fundamentales, así como las publicaciones derivadas de su progreso.

El segundo capítulo realiza un repaso al estado de la técnica del concepto fundamental abordado en este trabajo, que es la reconstrucción de escenas 3D. Para el resto de tópicos se ha optado por incluir una sección al comienzo de cada capítulo con las referencias bibliográficas más relevantes.

En el tercer capítulo se describe el sistema de visión considerado y las decisiones de diseño adoptadas para llegar hasta el mismo. Además se presentan las bases geométricas y relaciones matemáticas entre los diferentes elementos a considerar para el desarrollo del método propuesto.

El cuarto capítulo introduce el concepto del espacio de disparidad. Al tratarse de un concepto común en el dominio del tratamiento de imágenes estéreo, el capítulo comienza con un análisis del estado de la técnica para, posteriormente presentar los sistemas de ecuaciones que posteriormente permitirán llevar a cabo el modelado de la escena.

En base al espacio de disparidad presentado en el capítulo anterior, el quinto capítulo describe la metodología propuesta para reconstruir el contenido de una escena tridimensional y obtener una representación de la misma basada en los planos que la componen. Además, se lleva a cabo el análisis de los efectos de un posible error en el ángulo de cabeceo de las cámaras sobre los resultados de reconstrucción que produce el método.

Los dos capítulos siguientes describen dos aplicaciones diferentes que derivan del método de reconstrucción. En el capítulo seis se presenta un sistema novedoso de calibración del sistema estereoscópico para la corrección de errores en el ángulo de cabeceo haciendo uso de regresores. Por su parte, el séptimo capítulo describe el uso de una versión simplificada de la metodología de reconstrucción de escenas completas para la detección de obstáculos en un vehículo autoguiado.

Finalmente, el último capítulo presenta las principales conclusiones de esta tesis.

Capítulo 2

Estado de la Técnica

En el capítulo 1 se ha presentado la estructura de la presente tesis doctoral. Debido a las diferentes temáticas abordadas, y para facilitar la lectura del documento, en cada uno de los capítulos que siguen se incluye una sección denominada "Trabajo relacionado", que presenta al lector el estado de la técnica del ámbito del conocimiento tratado en cada capítulo.

No obstante, dado que el elemento vertebrador de este trabajo es la reconstrucción de escenas 3D a partir de imágenes estereoscópicas, se presenta a continuación el estado de la técnica en este campo.

2.1. Reconstrucción de escenas

La reconstrucción tridimensional de escenas a partir de imágenes estereoscópicas ha experimentado avances significativos en las últimas dos décadas gracias al desarrollo continuo de tecnologías de visión por computador y algoritmos de procesamiento de imágenes. La combinación de algoritmos avanzados, sensores mejorados y enfoques específicos para diferentes aplicaciones ha llevado a avances notables en la capacidad de capturar y modelar el mundo tridimensional a nuestro alrededor.

Uno de los primeros desafíos en la visión estereoscópica fue el desarrollo de algoritmos de correspondencia estéreo eficientes y precisos. Investigaciones tempranas, como el trabajo de Scharstein et al. (2001), proporcionaron un enfoque sólido para la determinación de correspondencias, utilizando técnicas basadas en la minimización de energía. Este método formula el problema como la optimización de una función de energía, considerando restricciones de suavidad y discontinuidades. A pesar de su efectividad, estos enfoques inicialmente carecían de la capacidad para manejar texturas repetitivas y áreas de oclusión.

Para abordar las limitaciones de la correspondencia estéreo basada en energía, se introdujeron algoritmos que trabajan con volúmenes de costo, como el propuesto por Hirschmuller (2008). Este enfoque calcula los costos de correspondencia para diferentes disparidades y utiliza la información acumulada para determinar la disparidad final. La inclusión de infor-

mación contextual mejoró la robustez, especialmente en regiones con texturas repetitivas y oclusiones.

El trabajo de Labayrade et al. (2002) ha contribuido significativamente al campo de la representación estéreo mediante el enfoque de Disparidad V. La Disparidad V es una técnica que representa las disparidades a lo largo del eje vertical de la imagen. Esta representación proporciona una visión compacta y efectiva de la información de profundidad en la escena.

Asimismo, el trabajo de Hu et al. (2005) ha explorado la representación estéreo mediante la Disparidad U-V, considerando tanto las disparidades horizontales (U) como verticales (V). Esta técnica proporciona una visión más completa de la escena, abordando desafíos específicos relacionados con la variación vertical en la disparidad.

Con el objetivo de adaptarse a la complejidad de las escenas, los algoritmos de correspondencia estéreo adaptativa han ganado prominencia. Investigaciones más recientes, como la de Diaz-Ramirez et al. (2022), proponen un método de emparejamiento estéreo basado en correlación morfológica adaptativa, capaz de determinar correspondencias de puntos con alta precisión en regiones homogéneas y en los bordes de objetos de la escena, y recuperar con éxito correspondencias desconocidas de puntos ocultos o no coincidentes en la escena mediante un post-procesamiento.

En la última década, el auge de las redes neuronales ha influido considerablemente en la correspondencia estéreo. La aplicación de redes convolucionales, como la arquitectura PSMNet propuesta por Chang and Chen (2018), ha llevado a mejoras notables en la velocidad y precisión de la correspondencia estéreo. Estas redes aprenden representaciones profundas de las imágenes, superando desafíos tradicionales y ofreciendo soluciones más adaptables a diversas condiciones de iluminación y textura.

El progreso en la calidad y la resolución de los sensores estéreo también ha sido fundamental. Investigaciones como la de Hirschmuller (2008), que se centraron en la calibración precisa de sistemas estéreo, contribuyeron al desarrollo de sensores más precisos y algoritmos más eficientes para la obtención de datos tridimensionales.

Asimismo, la fusión de información proveniente de múltiples fuentes se ha vuelto esencial para mejorar la precisión de las reconstrucciones 3D. Trabajos como el de Seitz et al. (2006), que propone métodos de fusión de imágenes múltiples para escenas complejas, han demostrado la importancia de la integración de datos para lograr representaciones más completas y precisas de la escena tridimensional.

Con el tiempo, la investigación se ha dirigido hacia aplicaciones específicas, como la reconstrucción de escenas en tiempo real para sistemas de realidad aumentada. Investigaciones como la de Ding et al. (2011) han abordado los desafíos asociados con la reconstrucción rápida y precisa de escenas con un enfoque más próximo al hardware. La técnica propuesta se basa en el concepto de 'adaptive support weight' (peso de soporte adaptativo). En lugar de asignar pesos uniformes a todas las disparidades, los autores introducen una estrategia adaptativa que ajusta los pesos de acuerdo con la variabilidad local de la textura y la intensidad en la

escena. Esto permite adaptarse a cambios dinámicos en la escena y mejorar la precisión de la generación del mapa de profundidad. Dada la complejidad computacional de la solución propuesta, es implementada mediante el uso de una FPGA (Field Programmable Gate Array).

En los últimos años, el empleo de técnicas de Inteligencia Artificial ha transformado significativamente la reconstrucción de escenas 3D y la detección de obstáculos, impulsando la precisión y la robustez de los sistemas de percepción.

La aplicación de redes neuronales convolucionales profundas (CNN) ha revolucionado la reconstrucción 3D a partir de datos de imágenes. El trabajo de Flynn et al. (2015) introduce DeepStereo, un enfoque que aprende a predecir nuevas vistas a partir de imágenes existentes. Varios trabajos han destacado la eficacia de estas técnicas en la generación precisa de modelos tridimensionales a partir de datos de imágenes.

Un ejemplo significativo es el método propuesto por Yariv et al. (2020). Esta investigación emplea una CNN para la reconstrucción de superficies 3D a partir de múltiples vistas de una escena. Al aprender representaciones latentes complejas, el modelo logra una reconstrucción más detallada y precisa de la geometría tridimensional.

Otro enfoque innovador es el presentado por Dai et al. (2017). Aquí se utiliza una CNN para realizar reconstrucciones 3D detalladas de escenas interiores a partir de datos de escaneo. La red aprende a generalizar y reconstruir con precisión objetos y estructuras en entornos complejos, superando las limitaciones de métodos tradicionales.

Investigaciones como la de Žbontar and Lecun (2015) han demostrado que entrenar una CNN para comparar parches de imagen puede llevar a resultados sobresalientes en términos de precisión y velocidad de procesamiento.

Si bien las CNN han avanzado significativamente en diversas tareas de visión por computador, aún tienen limitaciones, como la susceptibilidad al cambio de dominio. Para abordar esto, Wang et al. (2023) proponen un marco de trabajo que utiliza la traducción de imagen a imagen para superar la brecha de dominio en el emparejamiento estéreo, incorporando un módulo de generación atenta horizontal que considera las restricciones epipolares para mejorar la consistencia entre las vistas. Esto hace que el método sea más robusto para diferentes conjuntos de datos.

Las redes convolucionales completas (FCN) son una técnica derivada de las CNN con la capacidad de generar salidas de la misma dimensión que la entrada, lo que las hace especialmente útiles en tareas de segmentación semántica y procesamiento de imágenes a nivel de píxel. Zhang et al. (2024) describen un método para detectar y medir grietas en componentes de puentes de hormigón utilizando visión estéreo binocular y una FCN. El enfoque propuesto mejora la flexibilidad, eficiencia y precisión del proceso de inspección sin contacto, permitiendo una reconstrucción tridimensional confiable de la morfología de las grietas.

Además, el uso de técnicas de Aprendizaje Profundo (Deep Learning) se extiende a la reconstrucción 3D a partir de imágenes médicas. En el trabajo de Singh et al. (2020) se describe

el creciente uso de modelos de Aprendizaje Profundo en el ámbito médico, especialmente las redes neuronales convolucionales tridimensionales (3D CNN), que han sido adoptadas para mejorar la eficiencia de los médicos en el análisis de imágenes.

Además, se ha avanzado en la integración de técnicas de Aprendizaje Profundo en la reconstrucción 3D estéreo. Trabajos como el de Ye et al. (2022) han propuesto un enfoque de red estéreo que utiliza la consistencia local de disparidad para mejorar la estimación de disparidad en tiempo real, mediante módulos de refinamiento de consistencia espacial y temporal, logrando una alta precisión y velocidad de procesamiento de más de 40 FPS.

Se observa también un aumento en la aplicación de métodos de fusión de datos basados en inteligencia artificial para mejorar la precisión y robustez de las reconstrucciones 3D estéreo. Investigaciones como la de Zhao et al. (2020) han explorado la combinación de datos de múltiples sensores, como cámaras estéreo y LiDAR, utilizando técnicas de inteligencia artificial para obtener modelos tridimensionales más completos y detallados.

Otro ejemplo de desarrollo de técnicas de reconstrucción 3D estéreo en tiempo real aplicando técnicas de Inteligencia Artificial es el trabajo de Dovesi et al. (2020). Los autores presentan método que pueden realizar la reconstrucción semántica de escenas en tiempo real, lo que tiene aplicaciones potenciales en áreas como la navegación autónoma y la realidad aumentada.

Por otro lado, en las técnicas tradicionales de reconstrucción 3D y en particular en aquellas basadas en disparidad, se ha observado un enfoque continuo en el refinamiento de algoritmos clásicos de correspondencia estéreo. Aunque los métodos basados en inteligencia artificial han ganado terreno, aún se realizan avances en técnicas como la optimización de algoritmos de matching block y la mejora de la precisión mediante el ajuste de parámetros y la incorporación de estrategias de posprocesamiento. Un ejemplo es el trabajo de Guo et al. (2023) que presenta un algoritmo mejorado de Coincidencia Semiglobal (I-SGM) para cámaras binoculares en rovers lunares, diseñado para detectar obstáculos de manera precisa y eficiente en condiciones difíciles en la superficie lunar. Aunque la investigación reciente ha avanzado en el desarrollo de técnicas más sólidas y efectivas, sigue existiendo interés en el desarrollo de técnicas de reconstrucción estéreo basadas en aproximaciones más clásicas, como las que se abordan en esta tesis doctoral.

Capítulo 3

Sistema de visión

En este capítulo se aborda detalladamente el sistema de visión que se ha considerado para el desarrollo de esta tesis doctoral. De este modo se presentan los diseños que se han considerado y se realiza una comparativa que justifica la configuración seleccionada. Además, se describen los modelos geométricos utilizados para la estimación de la profundidad o la posición de los objetos en el entorno visual, así como la formulación de ecuaciones matemáticas que describen las relaciones espaciales entre los diferentes elementos del sistema. Este análisis teórico es fundamental para comprender el funcionamiento del sistema de visión y proporciona una base sólida para la implementación práctica del método propuesto.

3.1. Fundamentos

En la figura 3.1 se presenta el modelo general de un sistema de un sistema de visión binocular. Se observan una serie de parámetros, que se definen a continuación:

- d: distancia entre los centros ópticos $(O_i, i = r, l)$ de las cámaras.
- a: altura del centro óptico de la cámara sobre el nivel del suelo.
- θ: ángulo entre el centro óptico y el plano horizonatal (ángulo de cabeceo).

Además, se establecen tres sistemas de referencia diferenciados: R_w, que es el sistema de coordenadas mundial en el que se representan los puntos $P = (XYZ1)^T$ del mundo real, y R_i y R_d, que definen los sistemas coordenados de las cámaras izquierda y derecha, respectivamente. Para establecer completamente el marco de trabajo es necesario tener en cuenta que se puede suponer que se ha realizado la corrección epipolar y, por tanto, ambos planos imagen son coplanarios y están a la misma altura (a) sobre el suelo. En los sistemas R_i y R_d cada punto de la imagen se indica mediante sus coordenadas (u, v), resultantes de la proyección de las coordenadas del mundo real en el plano imagen. Un caso concreto son las proyecciones de los centros ópticos, denotados como (u_0, v_0).

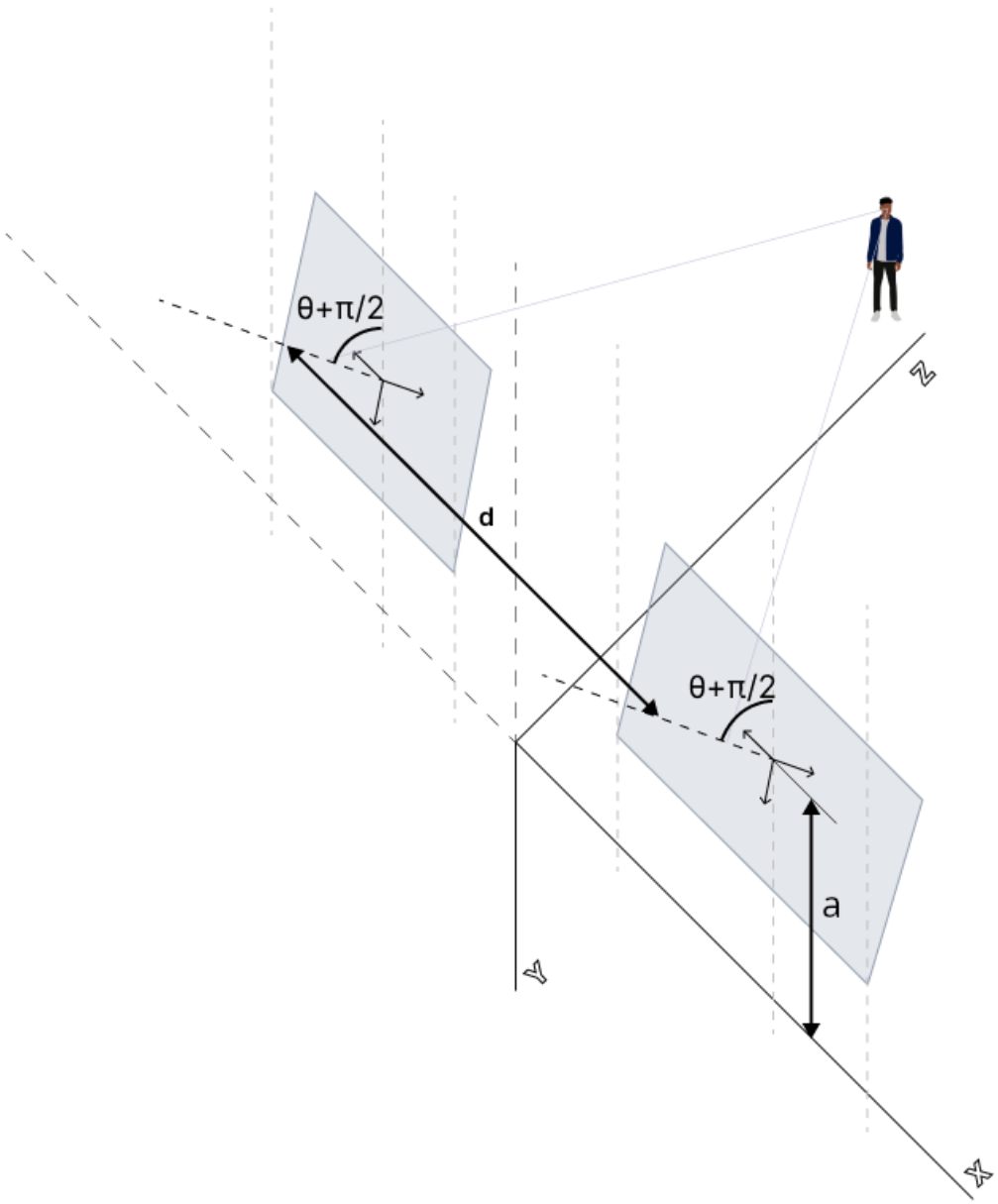


Figura 3.1: Representación esquemática de un sistema de visión estereoscópica.

Los parámetros intrínsecos de la cámara necesarios para la proyección se denotan como:

- f: Distancia focal de las cámaras, que se asume igual en ambas cámaras.
- t_u, t_v: Tamaño de los píxeles en las dimensiones u y v.

Por simplicidad, a la hora de desarrollar las ecuaciones se utiliza $\alpha_u = f/t_u$ y $\alpha_v = f/t_v$, pudiendo asumirse que $\alpha = \alpha_u = \alpha_v$.

Para poder utilizar la información que proporcionan las imágenes de ambas cámaras en la reconstrucción de información tridimensional es necesario establecer la relación entre los sistemas de coordenadas mundial y de las cámaras. Esta transformación se puede descomponer en tres operaciones simples. En primer lugar, una traslación hará que los centros de coordenadas sean coincidentes; en segundo lugar, mediante una rotación sobre el eje X, que tiene una orientación común a todos los sistemas definidos, se consigue que el sistema de coordenadas adopte la orientación del plano imagen. Finalmente, se proyectan los puntos del mundo real sobre los planos imagen para obtener sendas representaciones bidimensionales del mundo real.

Utilizando coordenadas homogéneas, estas transformaciones pueden ser expresadas mediante operaciones matriciales. De este modo, las matrices de traslación T_i y T_d pueden escribirse de la forma genérica T_l, donde q_l tendrá un valor $q_i = -1$ para la cámara izquierda y $q_d = 1$ para la cámara derecha.

$$T_l = \begin{pmatrix} 1 & 0 & 0 & -q_l \frac{d}{2} \\ 0 & 1 & 0 & a \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.1}$$

Por su parte, la matriz de rotación es idéntica en ambos casos, y de la forma:

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\operatorname{sen}(\theta) & 0 \\ 0 & \operatorname{sen}(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.2}$$

La combinación de ambas transformaciones da lugar a la matriz de cambio de sistema de coordenadas (T_{sc_l}) que se muestra a continuación:

$$T_{sc_l} = R \bullet T_l = \begin{pmatrix} 1 & 0 & 0 & -q_l \frac{d}{2} \\ 0 & \cos(\theta) & -\operatorname{sen}(\theta) & a\cos(\theta) \\ 0 & \operatorname{sen}(\theta) & \cos(\theta) & a\operatorname{sen}(\theta) \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.3}$$

Finalmente, se lleva a cabo la proyección del punto del mundo real P en sus coordenadas en cada uno de los sistemas locales R_i y R_d a través de la matriz de proyección M_{proy}, que se encarga de realizar el escalado de las posiciones por efecto de la óptica y de desplazar el origen de coordenadas a la esquina inferior izquierda de la imagen.

$$M_{proy} = \begin{pmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \alpha & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3.4}$$

La combinación de estas tres matrices da lugar a la matriz de transformación denominada T_{trans_l}.

Para proyectar un punto $P = (XYZ1)^T$ en coordenadas homogéneas en el sistema de coordenadas mundial, es necesario realizar una primera operación de traslación para situar el sistema de referencia de la cámara en el mismo punto que el de la imagen correspondiente. Con esto, el punto modificado P' será:

$$P' = \begin{pmatrix} X \\ Y \\ Z + f \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z' \\ 1 \end{pmatrix} \tag{3.5}$$

De esta manera se obtienen sendas proyecciones p_l:

$$p_l = T_{trans_l} \bullet P' = \begin{pmatrix} x_l \\ y \\ z) \end{pmatrix} \tag{3.6}$$

que permiten obtener las coordenadas no homogéneas:

$$\begin{cases} u_l = \dfrac{x_l}{z} & \text{(3.7a)} \\ v = \dfrac{y}{z} & \text{(3.7b)} \end{cases}$$

Del análisis llevado a cabo hasta el momento se desprende que existe una diferencia entre la proyección de un punto en las respectivas imágenes en la dimensión u. A esta diferencia, definida como $\Delta = u_i - u_d$, se la conoce como disparidad.

3.2. Sistema de visión de tres grados de libertad

Inicialmente se planteó un diseño de sistema de visión con tres grados de libertad. Estos permitían un movimiento conjunto de cabeceo de las dos cámaras y la guiñada independiente de cada una de ellas. El modelo de este sistema se muestra en la figura 3.2, siendo θ_1 el ángulo del grado de libertad que representa el cabeceo, y θ_2 y θ_3 los ángulos de guiñada de la cámara izquierda y derecha, respectivamente.

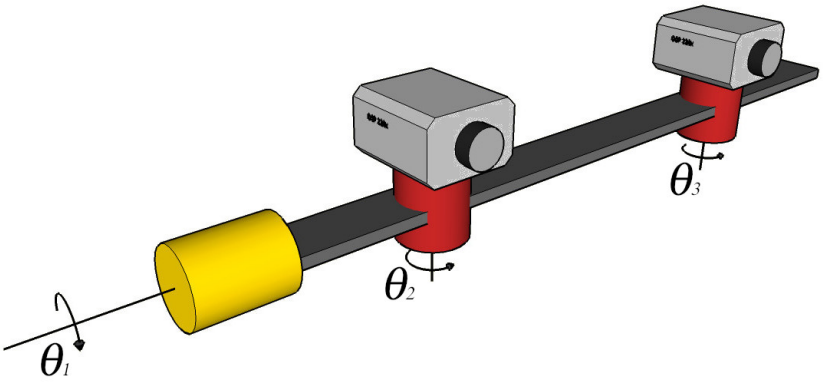


Figura 3.2: Representación simplificada del sistema de visión con tres grados de libertad.

Desde un punto de vista puramente geométrico, el hecho de permitir el giro independiente de cada una de las cámaras en torno al eje vertical hace que las ecuaciones que se obtienen sean más complejas. Por tanto, para simplificar el análisis del sistema se elimina un grado de libertad asumiendo que las dos cámaras realizan el mismo movimiento de guiñada, por lo que $\theta_2 = \theta_3$. Por motivos de claridad en la notación, se usará $\beta = \theta_2 = \theta_3$.

Asumiendo un modelo de cámaras estenopeicas (pinhole) y tomando la vista de planta del sistema de estereovisión con las restricciones impuestas, resulta sencillo determinar la relación entre la distancia hasta el punto P, representado por sus coordenadas (X, Y, Z), y la información que proporcionan las imágenes.

A partir de la representación de la figura 3.3, y empleando relaciones trigonométricas, se puede demostrar que la distancia a un punto con respecto al origen de coordenadas del sistema de referencia de la cámara izquierda es:

$$Z_i = \frac{df}{\Delta}\left(\cos\beta + \frac{u_d - u_0}{f}\,\mathrm{sen}\,\beta\right) - f \tag{3.8}$$

De manera rigurosa es necesario realizar una rotación y una traslación para que la medida de distancia calculada esté referida al origen del sistema de coordenadas mundial R_w. Estas transformaciones quedan descritas por las matrices R_β y $T_{d/2}$ que se muestran a continuación:

$$R_\beta = \begin{pmatrix} cos\beta & 0 & -\,\mathrm{sen}\,\beta & 0 \\ 0 & 1 & 0 & 0 \\ sen\beta & 0 & cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} T_{d/2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{d}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.9}$$

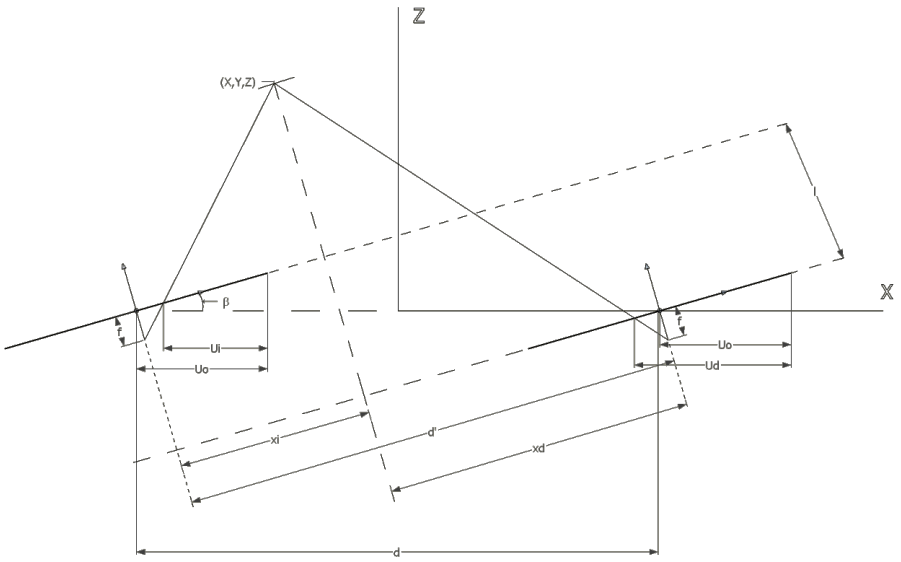


Figura 3.3: Representación estenopeica del sistema de visión con los planos imagen paralelos entre sí y rotados un ángulo β respecto al plano XY.

Aplicando las transformaciones anteriores a un punto cuyas coordenadas están referenciadas al sistema de coordenadas de la cámara izquierda, tal y como se muestra:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = R_\beta \bullet T_{d/2} \bullet \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \tag{3.10}$$

Una vez realizada la transformación, la coordenada Z que representa la distancia hasta el punto se puede calcular como:

$$Z = Z_i \left(\frac{u_i - u_0}{f} \operatorname{sen} \beta + \cos \beta \right) + (u_i - u_0) \operatorname{sen} \beta \tag{3.11}$$

3.3. Sistema de visión de dos grados de libertad

Además de la complejidad indicada anteriormente, el modelo de tres grados de libertad introduce una serie de complicaciones a la hora de garantizar la alineación de las cámaras que, si bien pueden ser compensadas mediante el procesamiento de las imágenes, es algo que debe ser tenido en cuenta a la hora de seleccionar el sistema de visión a utilizar. Junto al desalineamiento, los posibles defectos en la construcción y desviaciones en las rotaciones de los

ángulos θ_2 y θ_3 respecto a la posición comandada al tratarse de elementos electromecánicos diferentes hace que existan otra serie de problemáticas que habría que contemplar en caso de optar por esa configuración.

Por todo esto se propone un modelo estereoscópico en el que se han suprimido los grados de libertad θ_2 y θ_3, dejando fijas las cámaras sobre un mismo eje. De este modo se garantiza que los planos imagen de ambas cámaras sean coplanarios. El movimiento de guiñada es común a ambas cámaras, de modo que toda la estructura girará un ángulo θ_0 manteniendo la restricción de coplanariedad en torno a un eje vertical en el centro del sistema estéreo. El movimiento de cabeceo es el mismo que en el diseño de tres grados de libertad y, por claridad, se mantiene la nomenclatura, denominando θ_1 a este ángulo.

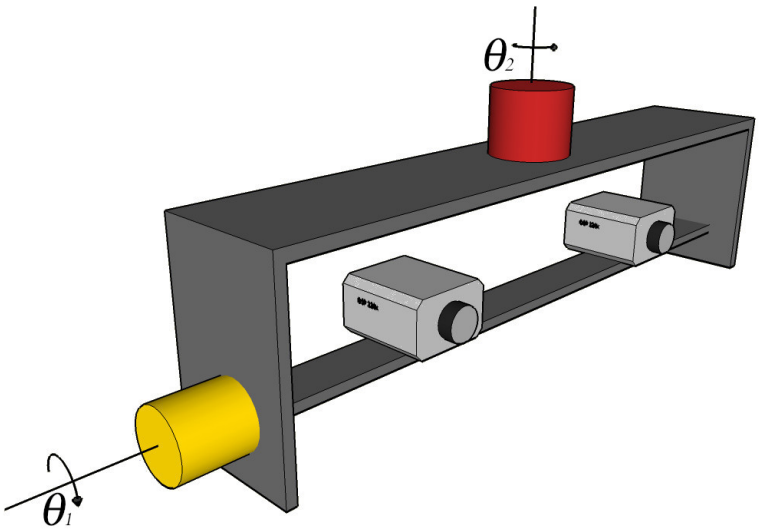


Figura 3.4: Representación simplificada del sistema de visión con dos grados de libertad.

La representación cenital del modelo estenopeico del sistema de visión de dos grados de libertad considerado es la siguiente:

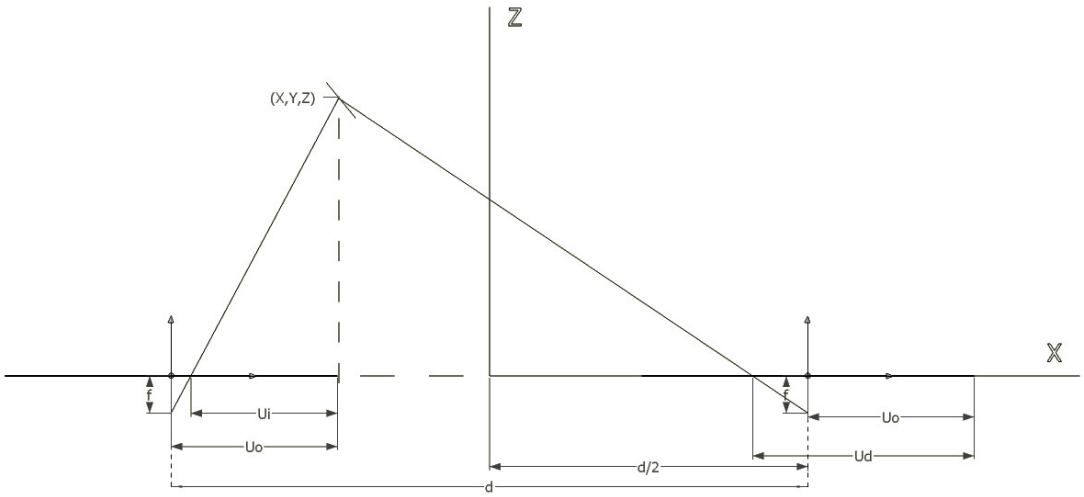


Figura 3.5: Representación estenopeica del sistema de visión con los planos imagen paralelos entre sí y también al plano XY.

Del mismo modo que en el caso del sistema de visión de tres grados de libertad, se puede analizar trigonométricamente la figura 3.5, obteniéndose la distancia Z como:

$$Z = \frac{df}{\Delta} - f \tag{3.12}$$

3.4. Comparación de los sistemas de visión

Además de lo ya reseñado anteriormente, para adoptar la decisión de qué sistema de visión emplear se ha llevado a cabo un estudio comparativo de las capacidades de ambos a la hora de estimar la distancia hasta un punto en el mundo real. A lo largo de esta discusión se denotará como Z_3 a la distancia medida utilizando el sistema de tres grados de libertad para distinguirla de la que es medida con la estructura de dos, que se denotará como Z_2. En el sistema de tres grados de libertad la distancia exacta se puede calcular según la expresión 3.11. Sin embargo, dadas las restricciones mecánicas que impone una estructura de este tipo, el ángulo β puede considerarse lo suficientemente pequeño para que se pueda asumir que $sen(\beta) \to 0$ y $cos(\beta) \to 1$, y, por tanto, Z_3 puede considerarse aproximadamente igual a Z_i, tal como se muestra en la ecuación 3.8. Entonces, la diferencia entre Z_2 y Z_3 se calcula como:

$$Z_3 - Z_2 = \frac{df}{\Delta}\left(\cos\beta + \frac{u_d - u_0}{f}\operatorname{sen}\beta - 1\right) \tag{3.13}$$

Si esta diferencia es negativa, implica que la distancia calculada para el caso del sistema de visión con tres grados de libertad dará un valor inferior al de dos grados de libertad. Por el contrario, si la diferencia es positiva, la distancia calculada para el sistema de tres grados

de libertad será mayor que la que se obtendría con el sistema de dos grados de libertad. Observando el segundo término de la ecuación 3.13, la condición $Z_3 < Z_2$ se dará cuando el factor de la expresión 3.14 sea menor que uno.

$$\left(\cos\beta + \frac{u_d - u_0}{f}\,\text{sen}\,\beta\right) \tag{3.14}$$

El término de la ecuación 3.14 puede interpretarse como un Factor de error sobre el resto de los términos de la ecuación 3.13. En la figura 3.6 se observa cómo evolucionan los valores de dicha expresión en función de los posibles valores de β y u_d:

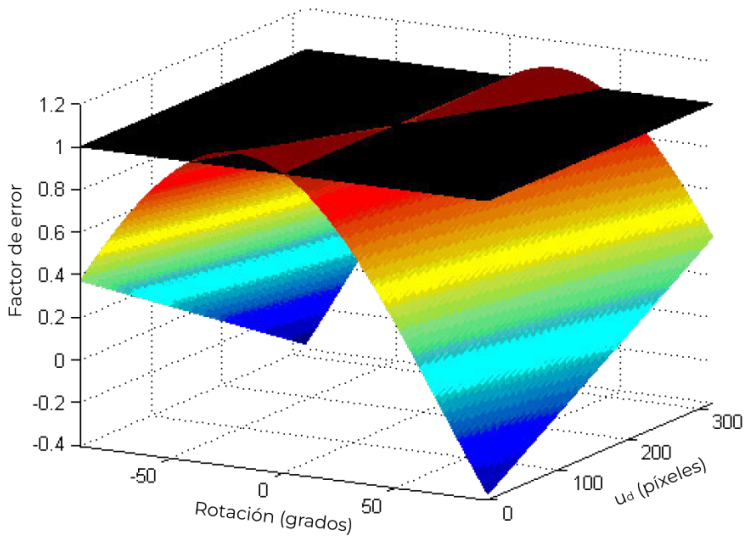


Figura 3.6: Factor de error atendiendo a los diferentes valores que pueden adoptar β u u_d

Al representar 3.14 para diferentes valores del ángulo β y diferentes valores de u_d, se observa que por debajo de la unidad, valor representado por el plano de color negro, el valor real de Z_3 es menor que Z_2. Por tanto, si se utiliza la expresión 3.11 como aproximación se adopta una postura conservadora, ya que en el peor de los casos se indicaría que un obstáculo está más cerca de lo que en realidad está, imposibilitando la colisión. Sin embargo, los puntos por encima de este plano indican que la distancia estimada con el sistema de visión de tres grados de libertad es mayor que la estimada con el de dos, por lo que en ese caso la posición conservadora se adopta tomando la distancia calculada a partir del sistema de dos grados de libertad. En algunas aplicaciones, como la detección de obstáculos, este tipo de errores representa riesgos por una mala percepción de la profundidad.

Para analizar las regiones por encima del plano unidad resulta más sencillo utilizar una vista cenital de la figura anterior.

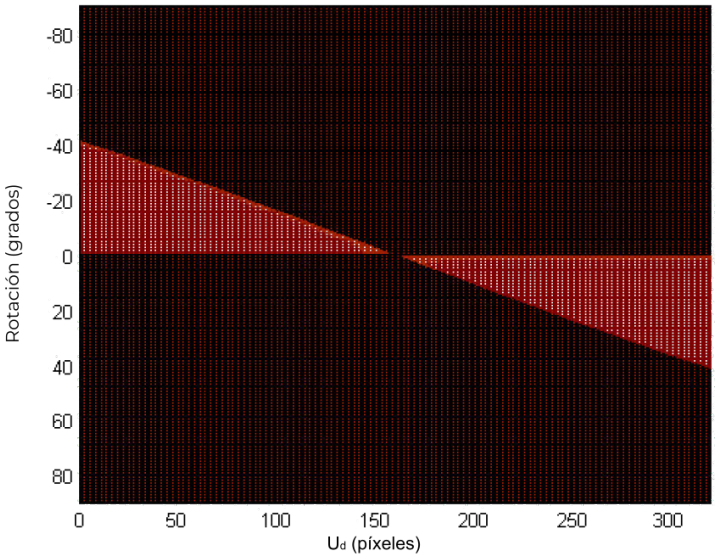


Figura 3.7: Vista cenital de la figura 3.6

En esta figura se observa cómo las regiones cuyo valor está por encima de la unidad aparecen para giros pequeños del ángulo β, que tal como se ha indicado previamente, son los que el sistema de visión realizaría. A la vista de este comportamiento se opta por descartar el sistema de tres grados de libertad.

A la vista de los resultados obtenidos se optó por el diseño de sistema de visión con dos grados de libertad. Esta decisión tiene una serie de implicaciones que serán relevantes a la hora de trabajar con él:

- El movimiento de guiñada es común para ambas cámaras. Esto también ocurre con el sistema de tres grados de libertad.
- Al eliminar un grado de libertad de cabeceo y hacerlo común consigue que los planos imagen de ambas cámaras puedan considerarse paralelos.

Capítulo 4

Espacio de Disparidad

La reconstrucción de escenas tridimensionales mediante técnicas de disparidad ha sido un área de intensa investigación en visión por computador y percepción computacional. La disparidad, que se refiere a las diferencias aparentes entre imágenes capturadas desde distintas perspectivas, constituye la base para la determinación de la estructura tridimensional de una escena.

4.1. Trabajo relacionado

Partiendo de un sistema estereoscópico cuyas características son conocidas, es posible computar el mapa de disparidad completo y, a partir de él, calcular un mapa 3D completo. El problema de estas técnicas es que son muy costosas computacionalmente por lo que no han sido una de las más empleadas en detección de obstáculos en tiempo real. En Hancock (1997) se propone un método que, a pesar de estar basado en disparidad, disminuye el coste computacional al trabajar con un conjunto de datos más pequeño que el que se emplea en las técnicas tradicionales de disparidad.

En términos generales, el proceso de reconstrucción 3D a partir de disparidades comprende varios pasos cruciales. En la etapa de emparejamiento estéreo, algoritmos como SIFT (Scale-Invariant Feature Transform) de Lowe (2004) y SURF (Speeded Up Robust Features) de Bay et al. (2006) son comúnmente utilizados para identificar correspondencias entre puntos clave en imágenes estéreo. Posteriormente, la generación de mapas de disparidad, empleando métodos como Block Matching y enfoques basados en optimización, se convierte en una tarea esencial para representar las diferencias de coordenadas entre las imágenes estéreo.

La calidad de los mapas de disparidad a menudo se mejora mediante técnicas de filtrado y refinamiento. Estrategias como el filtrado bilateral presentado en Tomasi and Manduchi (1998) y métodos basados en la consistencia de disparidad como los expuestos en Scharstein and Szeliski (2002) son implementados para mitigar posibles errores. La reconstrucción 3D final implica el cálculo de las coordenadas tridimensionales de los puntos en la escena

utilizando la información de disparidad. Métodos de triangulación y técnicas de optimización son comúnmente aplicados en esta fase.

Un posible enfoque para abordar el problema de reconstrucción empleando disparidad son los métodos basados en Disparidad V descritos en Labayrade et al. (2002), que simplifican la forma en la que se interpreta el mapa de disparidad. Concretamente, partiendo de la premisa de que los puntos pertenecientes a un mismo objeto deberán presentar un valor de disparidad similar, se propone una nueva forma de reconstruir la información tridimensional. Esta teoría ha sido extendida en la dimensión horizontal además de en la vertical. Es la denominada Disparidad-U-V introducida en Hu et al. (2005).

4.2. Definición del Espacio de Disparidad

Una vez justificada la elección de un sistema de visión estereoscópica con dos grados de libertad, se tiene que tanto los ángulos de guiñada como de cabeceo son comunes para ambas cámaras, tal y como se explicó en la sección 3.3. Como consecuencia de ello, los planos imagen rotan de manera solidaria, viéndose únicamente afectados por el ángulo de cabeceo (pitch).

Por ello, es posible limitar el análisis del sistema de visión en relación al comportamiento de este ángulo, puesto que bastaría con aplicar una matriz de rotación sobre el eje vertical para cambiar entre el sistema de coordenadas rotado en el movimiento de guiñada y el mundial. En base a esto, y para simplificar la notación, en adelante se hará referencia únicamente al ángulo de cabeceo (pitch) como θ.

Teniendo en cuenta el cambio señalado en la notación, desarrollando a partir de las ecuaciones 3.7a y 3.7b y haciendo uso de la restricción epipolar, se deduce que la coordenada v resultante de proyectar el punto P es la misma en ambas imágenes y de la forma:

$$v = \frac{y}{z} = \frac{Y\left(\alpha\cos\theta + v_0\,\mathrm{sen}\,\theta\right) + \left(v_0\cos\theta - \alpha\,\mathrm{sen}\,\theta\right)Z' + \alpha a\cos\theta + v_0 a\,\mathrm{sen}\,\theta}{Y\,\mathrm{sen}\,\theta + Z'\cos\theta + a\,\mathrm{sen}\,\theta} \tag{4.1}$$

Asimismo, las coordenadas u_l, para la imagen izquierda y derecha son:

$$u_l = \frac{x_l}{z} = \frac{\alpha X + Y u_0\,\mathrm{sen}\,\theta + u_0 Z'\cos\theta - \alpha q_l\frac{d}{2} + u_0 a\,\mathrm{sen}\,\theta}{Y\,\mathrm{sen}\,\theta + Z'\cos\theta + a\,\mathrm{sen}\,\theta} \tag{4.2}$$

En la figura 4.1 se observa el efecto conseguido al garantizar el alineado de los planos imagen. Se observa como la proyección del punto P en cada uno de los planos imagen posee una coordenada v común, mientras que la coordenada u_l presenta diferentes valores u_i y u_j.

Particularizando para las imágenes izquierda y derecha ($q_l = -1$ en la imagen izquierda y $q_l = 1$ en la derecha) se tiene que:

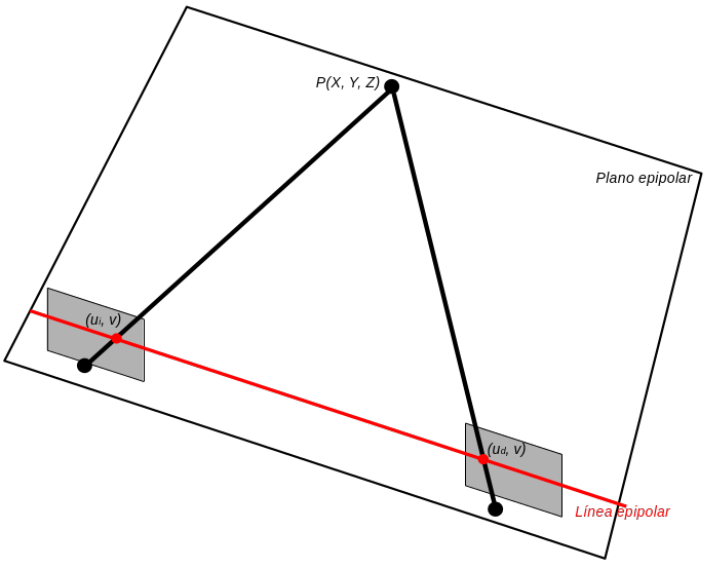


Figura 4.1: Representación simplificada del efecto obtenido al alinear los planos imagen. Se observa como la proyección del punto P en casda uno de los planos imagen tiene su propia coordenada u_l mientras que la coordenada v es coincidente.

$$\begin{cases} u_i = u_0 + \dfrac{\alpha X + \alpha \frac{d}{2}}{(Y+a)\operatorname{sen}\theta + Z'\cos\theta} & \text{(4.3a)} \\ u_d = u_0 + \dfrac{\alpha X - \alpha \frac{d}{2}}{(Y+a)\operatorname{sen}\theta + Z'\cos\theta} & \text{(4.3b)} \end{cases}$$

Y de aquí, tal como se definió anteriormente, la disparidad $\Delta = u_i - u_d$ es:

$$\begin{aligned} \Delta &= \frac{\alpha d}{(Y+a)\operatorname{sen}\theta + Z'\cos\theta} \\ \Delta &= \frac{\alpha d}{(Y+a)\operatorname{sen}\theta + (Z+f)\cos\theta} \end{aligned} \tag{4.4}$$

De esta manera se establece como espacio de disparidad al definido en (u_d, v, Δ).

Si se analiza la expresión de Δ, de manera intuitiva se deduce que aquellos píxeles pertenecientes a objetos o elementos de la imagen más cercanos presentan un valor de disparidad más alto que los pertenecientes a objetos lejanos. Así pues, cuanto más cercano está el punto proyectado menor es el valor de la coordenada Z, y por tanto, mayor es la disparidad medida. Este resultado, que es bastante evidente, resulta de utilidad a la hora de utilizar la información contenida en el espacio de disparidad para la obtención de información tridimensional a partir de un par estereoscópico.

Capítulo 5

Modelado de la escena

Una forma de simplificar la composición de una escena del mundo real es interpretarla como una combinación de planos con diferentes orientaciones distribuidos en el espacio. Cada objeto presente en la escena se representa por un único plano con una orientación proporcional al promediado de la posición de los puntos en el espacio que responden a la expresión general de un plano:

$$rX + sY + tZ + u = 0 \tag{5.1}$$

En el estado de la técnica, las implementaciones y desarrollos basados en Disparidad V o Disparidad U-V se concentran en un conjunto de planos concretos que responden a versiones reducidas de la ecuación 5.1. Por ejemplo, Labayrade and Aubert (2004) presenta la aplicación de la técnica de disparidad en V descrita previamente en Labayrade et al. (2002). En este artículo, los autores reducen los planos reconocidos a aquellos que responden a la ecuación $tZ + u = 0$. De esta forma, son capaces de estimar la inclinación de la vía y la existencia de obstáculos haciendo uso únicamente de la información contenida en el plano de Disparidad V.

Por otro lado, en Hu et al. (2005) no solo se consideran los planos oblicuos, como en el caso anterior, sino que también se hace una distinción en los planos verticales, que responden a $Z = -u/t$, y en las perpendiculares al plano imagen. En este caso hacen uso tanto de las proyecciones de disparidad en V como de disparidad en U para no solo detectar los elementos de la imagen que responden a los casos considerados, sino también para determinar sus dimensiones.

La técnica propuesta en esta tesis no se limita al estudio de un subconjunto de planos que responda a alguna variación de la ecuación 5.1, sino que se propone un tratamiento totalmente generalizado de la información. De esta forma, es posible llevar a cabo la composición de una escena completa y no solo de determinados elementos que la conforman.

5.1. Identificación de elementos de la escena

Para que un conjunto de puntos puedan agruparse e interpretarse como un plano, es necesario identificar cuáles de ellos pertenecen a cada elemento de la imagen en el espacio de Disparidad U-V. Asumiendo esta simplificación del mundo real y teniendo en cuenta el teorema que se presenta a continuación, se puede deducir que independientemente de la orientación de un plano dado en el mundo real, siempre es posible encontrar una línea paralela a ese plano que también sea paralela al plano de la imagen.

Teorema 1. *Dados dos planos que son paralelos a una recta, la intersección de esos planos será otra recta paralela a la primera.*

Introduciendo la restricción de la ecuación 5.1 y ampliando la ecuación 4.1, la disparidad también se puede calcular como:

$$\Delta = \frac{-2d}{2u + rd - 2sa} \cdot \left(r\left(u_d - u_0\right) + \left(v - v_0\right)\left(s\cos\theta - t\,\mathrm{sen}\,\theta\right) + \alpha\left(s\,\mathrm{sen}\,\theta + t\cos\theta\right)\right) \quad (5.2)$$

La ecuación 5.2 se puede particularizar con las restricciones discutidas anteriormente utilizadas por otros autores (como son la consideración únicamente de planos verticales u oblicuos), comprobándose que es una generalización de las de 4.1 y 4.2.

Se observa que la disparidad es una función que depende de u_d y v, que son variables de dominio discretas. Debido a esto, el gradiente de la disparidad se define como la derivada de Δ con respecto a ambas variables:

$$\nabla\Delta = \left(\frac{d\Delta}{du_d}, \frac{d\Delta}{dv}\right) \quad (5.3)$$

donde:

$$\frac{d\Delta}{du_d} = \frac{-2d}{2u + rd - 2sa} r \quad (5.4)$$

$$\frac{d\Delta}{dv} = \frac{-2d}{2u + rd - 2sa}\left(s\cos\theta - t\,\mathrm{sen}\,\theta\right) \quad (5.5)$$

Estas ecuaciones permiten postular un teorema que será fundamental para esta representación abstracta del contenido de una escena:

Teorema 2. *El gradiente de cada elemento (u_d, v) perteneciente a la proyección y transformación de puntos coplanares en el espacio de coordenadas del mundo real es constante y sólo depende del propio plano y su orientación con respecto al sistema de visión.*

En la figura 5.1 se muestra un ejemplo de la aplicación de los teoremas 1 y 2. El parelelepípedo coloreado dentro del plano de la imagen se construye uniendo las proyecciones de la línea paralela al plano P en cada imagen del par estéreo. Dado que la línea es paralela al

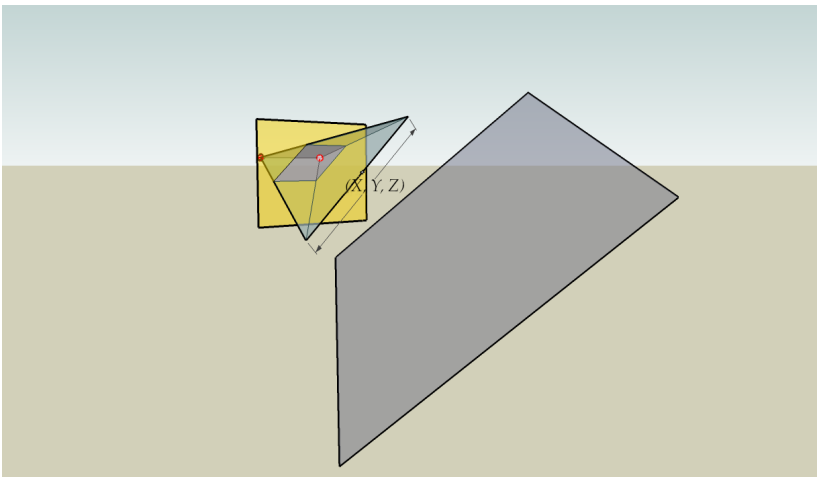


Figura 5.1: El plano P con una orientación arbitraria, representado en gris, se proyecta sobre el plano imagen, en amarillo. Las esferas, destacadas en rojo, por detrás del plano imagen representan los centros ópticos de las cámaras izquierda y derecha, respectivamente

plano de la imagen, estará contenida en un plano descrito por la ecuación $0X + 0Y + tZ = u$ y el ángulo entre los planos será 0 grados. Con esta configuración y según las expresiones anteriores, se observa que $\nabla\Delta = (0, 0)$, lo que significa que la disparidad es constante para cada punto de la línea, siendo las dos proyecciones paralelas en el plano de la imagen. Teniendo esto en cuenta, es posible dividir el espacio de disparidad en función de los diferentes gradientes de disparidad obtenidos. Así se puede obtener la ecuación del plano que genera ese gradiente, lo que permite reconstruir completamente la escena a partir de la información del espacio (u_d, v, Δ).

5.2. Reconstrucción de la escena

Los teoremas anteriores permiten sentar las bases para reconstruir cualquier escena representándola mediante un conjunto de planos. Dado que el objetivo es reconstruir una escena en coordenadas mundiales, se debe poder calcular dichas coordenadas en función de los puntos proyectados en el sistema de visión. Expandiendo las ecuaciones para Δ, u_d y v (ver ecuaciones 4.1, 4.2 y 4.4), se obtienen las expresiones para obtener las coordenadas de estos puntos con respecto al sistema mundial de referencia:

$$X = \frac{d}{\Delta}(u_r - u_0) + \frac{d}{2} \tag{5.6}$$

$$Y = \frac{d}{\Delta}((v - v_0)\cos\theta + \alpha \operatorname{sen}\theta) - h \tag{5.7}$$

$$Z = \frac{d}{\Delta}(\alpha\cos\theta - (v - v_0)\operatorname{sen}\theta) \tag{5.8}$$

Haciendo uso de la propiedad descrita en el teorema 2, se pueden reconocer aquellos puntos proyectados que pertenecen al mismo plano en la escena real. Esto hace posible tener tres pares de puntos $(u_{ij}, u_{dj} | j \in \{1:3\})$ proyectados en ambas imágenes que, una vez reconstruidos, forman parte de un mismo plano. Además, se puede calcular el valor de la disparidad $\Delta_1..\Delta_3$ para cada par de puntos.

Particularizando las ecuaciones 5.6, 5.7 y 5.8 para $(u_{ij}, u_{dj} | j \in \{1:3\})$, es posible calcular los coeficientes de la ecuación del plano. Esto se puede hacer determinando aquellas regiones del espacio (u_d, v, Δ) cuyo gradiente es constante y seleccionando tres puntos cualquiera dentro de esas regiones. A partir de las expresiones que representan los puntos proyectados sobre los planos de la imagen (ecuaciones 4.1, 4.3a y 4.3b) y la ecuación general del plano (ecuación 5.1), es posible calcular analíticamente los coeficientes del plano que contiene los tres puntos, como se describe en las ecuaciones 5.9, 5.10, 5.11 y 5.12.

$$r = \frac{d^2}{\Delta_1 \Delta_2 \Delta_3} \alpha \left((v_2 - v_1)\,\Delta_3 + (v_1 - v_3)\,\Delta_2 + (v_3 - v_2)\,\Delta_1 \right) \tag{5.9}$$

$$s = \frac{d^2}{\Delta_1 \Delta_2 \Delta_3} \left(A \operatorname{sen}\theta + B\alpha \cos\theta \right) \tag{5.10}$$

$$t = \frac{d^2}{\Delta_1 \Delta_2 \Delta_3} \left(B\alpha \operatorname{sen}\theta + A cos\theta \right) \tag{5.11}$$

$$u = \frac{d^2}{\Delta_1 \Delta_2 \Delta_3} u_1 \tag{5.12}$$

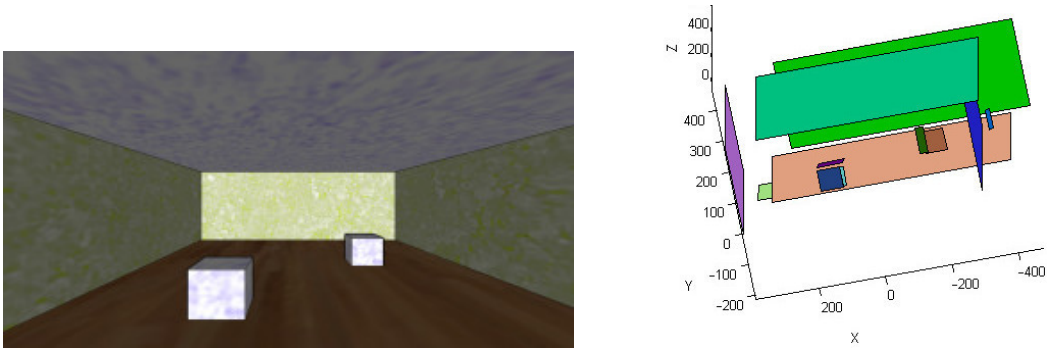
siendo:

$$\begin{aligned} u_1 = \Big(& (v_2 - v_1)\,\frac{\Delta_3}{2} + (v_1 - v_3)\,\frac{\Delta_2}{2} + (v_3 - v_2)\,\frac{\Delta_1}{2} \\ & + (u_{d1} - u_{d2})\,v_3 + (u_{d3} - u_{d1})\,v_2 + (u_{d2} - u_{d3})\,v_1 \Big)\,\alpha d \\ & + (B\alpha \cos\theta + A \operatorname{sen}\theta)\,h \end{aligned} \tag{5.13}$$

$$\begin{aligned} A = & \left((u_{d1} - u_0)\,v_2 + (u_0 - u_{d2})\,v_1 + (u_{d2} - u_{d1})\,v_0 \right) \Delta_3 \\ & + \left((u_0 - u_{d1})\,v_3 + (u_{d3} - u_0)\,v_1 + (u_{d1} - u_{d3})\,v_0 \right) \Delta_2 \\ & + \left((u_{d2} - u_0)\,v_3 + ((u_0 - u_{d3})\,v_2 + (u_{d3} - u_{d2}\mathrm{t})\,v_0 \right) \Delta_1 \end{aligned} \tag{5.14}$$

$$B = (u_{d2} - u_{d1})\,\Delta_3 + (u_{d1} - u_{d3})\,\Delta_2 + (u_{d3} - u_{d2})\,\Delta_1 \tag{5.15}$$

En la figura 5.2 se muestra de forma sencilla el comportamiento del método descrito al aplicarlo sobre una imagen sintética.



(a) Escena sintética de prueba.

(b) Reconstrucción obtenida de 5.2a

Figura 5.2: Demostración del funcionamiento del método de reconstrucción de escenas 3D mediante el uso de planos sobre una escena sintética.

5.3. Análisis de error de ángulo de cabeceo

Tal como se ha presentado, el problema de reconstrucción 3D puede abordarse con éxito haciendo uso de la Disparidad U-V. Por ejemplo, en Li et al. (2020), los autores proponen un sistema para segmentar escenas del mundo real basándose en una novedosa técnica de segmentación basada en el análisis de Disparidad U-V para diferenciar objetos de las superficies que los sostienen en ambientes ruidosos. Tanto en Li et al. (2020) como en otros artículos similares, se asume que se conocen exactamente todos los parámetros que caracterizan el sistema de visión.

Sin embargo, es bien sabido que en la práctica existen o pueden existir discrepancias entre los valores "teóricos" y "reales" de ciertos parámetros. Estas discrepancias pueden tener diferentes orígenes. Pueden deberse a errores en la determinación del valor de estos parámetros con total exactitud, pero también pueden venir dados por a los cambios que dichos parámetros sufren a lo largo del tiempo debido al envejecimiento normal que experimenta cualquier sistema físico. Por tanto, parece razonable concluir que ser capaz de determinar cómo los errores o incertidumbres en el conocimiento de ciertos parámetros del sistema afectan al rendimiento de un algoritmo en condiciones reales es tan importante como disponer de un procedimiento que resuelva satisfactoriamente el problema de reconstrucción 3D en condiciones ideales. En otras palabras, se quiere medir la sensibilidad del método diseñado a los errores considerados.

En concreto, en sistemas de visión estereoscópica el ángulo de cabeceo juega un papel muy importante en los resultados de reconstrucción de la escena. Esto es debido a que una buena estimación de dicho ángulo contribuye a aumentar la precisión en la medición de la distancia hasta los elementos que componen la escena, lo que es esencial para múltiples aplicaciones y, en particular, para aplicaciones relacionadas con la seguridad vial.

Una de las aplicaciones estudiadas en esta tesis doctoral es la detección de obstáculos en un vehículo autoguiado. En este ámbito, la estimación precisa del ángulo de cabeceo es crucial en los sistemas de detección de peatones para evitar interpretaciones erróneas de objetos en el plano horizontal como obstáculos tridimensionales. Del mismo modo, un ángulo de cabeceo incorrectamente estimado puede provocar falsas alarmas al interpretar objetos como obstáculos o, inversamente, puede evitar la detección de los mismos debido a una estimación errónea de la inclinación de la carretera.

5.3.1. Trabajo relacionado

La literatura contiene numerosos trabajos que apoyan la importancia de realizar un análisis de sensibilidad, como en Iooss and Saltelli (2017), Borgonovo and Plischke (2016), Ferretti et al. (2016) o Becker and Saltelli (2015). Específicamente en Becker and Saltelli (2015), los autores afirman que el análisis de sensibilidad se utiliza para medir la solidez de la inferencia basada en modelos, es decir, el grado en que los resultados generados por el modelo dependen de los supuestos utilizados para construirlo.

Este tipo de análisis también se aplica a sistemas en los que se obtiene información de profundidad de una escena, lo que se puede conseguir mediante diferentes técnicas. Por ejemplo, en Nagatomo et al. (2010) los autores proponen lo que llaman "Método de medición 3D estéreo relativo" (The Relative Stereo 3D Measurement Method). Este método muestra una buena tolerancia a los errores de calibración del sistema de visión estéreo y es muy preciso a la hora de calcular la distancia relativa. En otras aplicaciones, como las que se encuentran en Khoshelham and Elberink (2012) o Park et al. (2012) , se utilizan sensores Kinect para obtener información de profundidad. En Khoshelham and Elberink (2012), los autores presentan una discusión sobre la calibración del sensor Kinect y analizan la precisión y resolución de la información de profundidad que proporciona. Los resultados experimentales obtenidos muestran que el error aleatorio en las mediciones de profundidad aumenta con la distancia al sensor. Por su parte, Park et al. (2012) presenta un modelo matemático de incertidumbre para la medición espacial de características visuales utilizando sensores Kinect. Gracias a este modelo se dispone de un análisis cualitativo y cuantitativo del uso de sensores Kinect como sensores de percepción 3D. En Belhaoua et al. (2010), se presenta un enfoque diferente en el que los autores proponen un método para analizar y evaluar las incertidumbres encontradas en escenas reconstruidas basadas en visión. Este estudio considera principalmente los errores en la etapa de segmentación de la imagen, que se propagan a lo largo del procedimiento de reconstrucción.

En Zhao and Nandhakumar (1996) se presenta una aportación que, a pesar de su antigüedad, resulta de gran interés. En este trabajo, los autores realizan un estudio exhaustivo de la influencia de ciertos parámetros de calibración en el rendimiento de un sistema de reconstrucción 3D estereoscópico. En concreto, consideran un sistema de visión binocular

con dos cámaras montadas sobre un soporte rígido y estudian los errores de profundidad debidos al cabeceo, alabeo y guiñada entre las dos cámaras. También estudian la magnitud de los errores provocados por el hecho de que en una de las dos cámaras el conjunto CCD no sea paralelo a la lente, así como los errores provocados por la distorsión de la misma. Su estudio cuantitativo permite a los autores determinar los requisitos que deben cumplir los parámetros antes mencionados para que los errores se mantengan dentro de unos umbrales determinados. Finalmente, Santoro et al. (2012) proporciona un procedimiento para corregir la desalineación con el fin de reducir los errores de profundidad debido a los desplazamientos de la cámara. Las imágenes reales obtenidas con un sistema de cámara estéreo muestran que el método propuesto por los autores reduce significativamente las desalineaciones causadas por el cabeceo, alabeo y guiñada de la cámara.

Finalmente, el trabajo de Llorca et al. (2009) resulta interesante ya que aborda explícitamente la importancia de corregir el ángulo de cabeceo en un sistema de visión instalado en un vehículo autoguiado. Los autores describen dos métodos de compensación del ángulo de cabeceo diseñados para operar sobre una representación no densa del entorno: uno basado en el mapa de proyección YOZ y otro en una imagen de disparidad virtual. El análisis comparativo de los métodos indica que el segundo método supera al primero en términos de rendimiento. Además de este análisis, concluyen que la compensación del ángulo de cabeceo ofrece dos ventajas principales. En primer lugar, aumenta la precisión en la estimación del tiempo de colisión en accidentes de automóviles con peatones. En segundo lugar, mejora la separación entre los puntos de la carretera y los puntos de los objetos, lo que reduce las tasas de detección de falsos positivos y falsos negativos.

5.3.2. Medición de la desviación entre los planos ideal y calculado

Tal y como se indicó en el capítulo 4, en base a la discusión presentada en la sección 3.4 para el desarrollo y análisis de la técnica propuesta el único grado de libertad considerado es el ángulo de cabeceo. Por este motivo es necesario caracterizar el comportamiento de la técnica frente a posibles errores en dicho ángulo.

Para introducir el efecto de este error en las expresiones anteriores (5.9, 5.10, 5.11 y 5.12), el valor de θ es reemplazado por $\theta + \epsilon$, donde ϵ representa la diferencia entre los ángulos de cabeceo calculado e ideal. El plano ideal se obtiene asumiendo que no hay error en el ángulo de cabeceo, es decir, que ϵ vale cero.

Dados tres puntos cualesquiera $P1 - P3$ pertenecientes a un mismo elemento en la escena real identificados por sus coordenadas con respecto al sistema de referencia mundial $[X_j, Y_j, Z_j]$, es posible determinar los coeficientes de la ecuación del plano que los contiene $P_{ideal} = [r, s, t, u]_{ideal}$. Si estos puntos se proyectan utilizando las ecuaciones 4.1, 4.3a y 4.3b, es posible determinar las coordenadas proyectadas en ambas imágenes para todos los puntos.

Introduciendo el factor de error (ϵ) y recurriendo a las ecuaciones modificadas (5.9, 5.10, 5.11 y 5.12), se obtienen los coeficientes del plano reconstruido $P_{calculado} = [r, s, t, u]_{calculado}$.

Dado que los planos se caracterizan por cuatro parámetros (r, s, t, u), una forma posible de comparar los planos ideal y calculado sería obtener la diferencia entre los valores de estos parámetros en los dos casos. En el análisis que se ha llevado a cabo se ha optado por emplear el ángulo que forman los vectores normales de los dos planos involucrados para realizar esta comparativa de forma más sencilla e intuitiva. Este ángulo se calcula de la siguiente manera:

$$\varepsilon_{normal} = \arccos\left(\frac{|\vec{n}_{ideal} \cdot \vec{n}_{calculado}|}{|\vec{n}_{ideal}|\,|\vec{n}_{calculado}|}\right) \tag{5.16}$$

La ecuación anterior permite evaluar cuantitativamente cómo responde el sistema de reconstrucción propuesto a errores en el ángulo de cabeceo del sistema de visión.

5.3.3. Construcción del conjunto de prueba

Para realizar un estudio exhaustivo de la influencia del error en el ángulo de cabeceo en la reconstrucción de la escena, es necesario evaluar cómo responde el sistema en una variedad de situaciones que representan los casos que pueden ocurrir en cualquier escena real. Este problema se aborda mediante la construcción de un conjunto de prueba que consta de planos en un amplio rango de orientaciones con respecto al sistema de visión. Nótese que desde el punto de vista del conjunto de prueba no tiene relevancia que los planos que lo componen provengan de un escenario sintético simple, como el que se muestra en la figura 5.2, o de un escenario real más complejo.

El elemento de partida para construir el conjunto es un plano paralelo al plano de la imagen. Este plano semilla se rota alrededor de los tres ejes cartesianos, variando los ángulos ρ_X, ρ_Y, ρ_Z entre 0° y 90° a intervalos de 5°. El resultado es un conjunto de prueba que consta de 6859 planos con diferentes orientaciones, todos ellos a la misma distancia del plano de la imagen. Por lo tanto, para que el conjunto de prueba pueda representar fielmente cualquier escena real, es necesario introducir un grado de libertad adicional que es la distancia de los planos al plano de la imagen. Esto implica multiplicar el tamaño del conjunto de prueba por un factor N, donde N es el número de distancias consideradas.

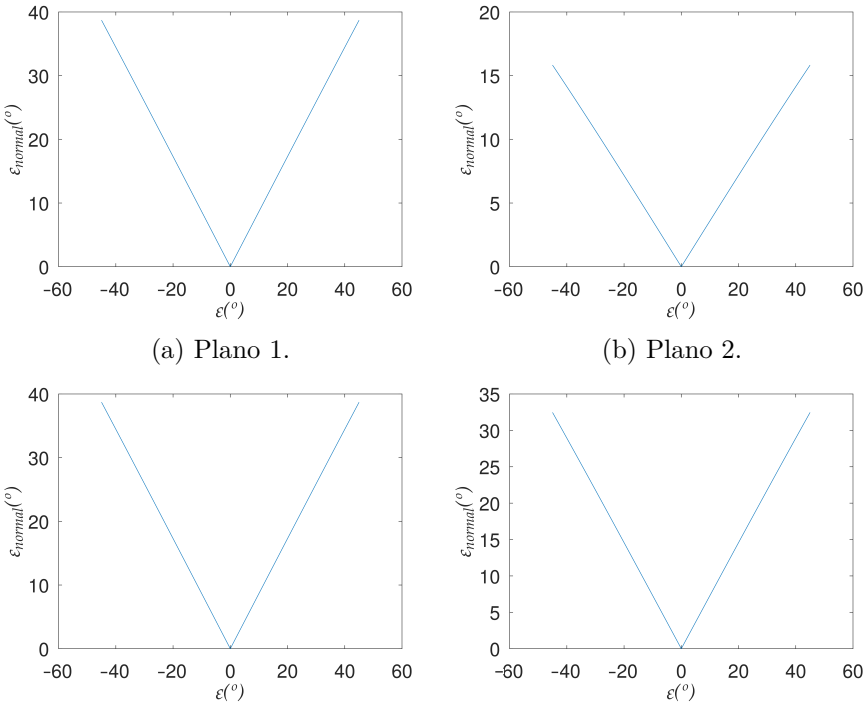
5.3.4. Procedimiento y resultados

Para analizar cómo afecta un error en el ángulo de cabeceo a los planos que se utilizan para reconstruir una escena, se mide la diferencia entre las normales de los planos ideal y calculado con este valor del error para todos los planos del conjunto de prueba. Al hacer esta comparativa, se observa que la distancia entre el plano en coordenadas mundiales considerado y el plano de la imagen no tiene influencia en la desviación entre los planos ideal y calculado. En otras palabras, dicha desviación sólo depende de la orientación del plano, y no de su

distancia al plano imagen. Esto permite reducir el conjunto de prueba a los 6859 planos originales con diferentes orientaciones, como se describe en la sección 5.3.3, y a una distancia de 5 m del plano de la imagen. Dado que, como se acaba de comentar, es un parámetro que no influye en la desviación entre los planos ideal y calculado, esta distancia se ha elegido de forma arbitraria.

Debido al gran tamaño del conjunto de prueba, es imposible representar gráficamente la diferencia entre las normales de los planos ideal y calculado en función del error en el ángulo de inclinación para todos los planos que lo componen. Dado que el comportamiento cualitativo observado es el mismo para todos ellos, la figura 5.3 solo muestra la respuesta para cuatro planos con diferentes orientaciones, lo que basta para ilustrar la diversidad que existe a nivel cuantitativo. Se seleccionaron los siguientes planos:

- *Plano 1*: $\rho_X = 0°; \rho_Y = 45°; \rho_Z = 45°$
- *Plano 2*: $\rho_X = 15°; \rho_Y = 75°; \rho_Z = 0°$
- *Plano 3*: $\rho_X = 30°; \rho_Y = 15°; \rho_Z = 90°$
- *Plano 4*: $\rho_X = 45°; \rho_Y = 75°; \rho_Z = 0°$



(a) Plano 1. (b) Plano 2.

(c) Plano 3. (d) Plano 4.

Figura 5.3: Desviación entre las normales de los planos ideal y calculado como función del error en el ángulo θ para los planos seleccionados.

Como se muestra en la figura 5.3, para los cuatro planos considerados y, por extensión, para todos los planos del conjunto de prueba, existe una perfecta simetría en la desviación entre las normales de los planos ideal y calculado con respecto a ϵ. En otras palabras, el valor de la desviación solo depende del valor absoluto del error en el ángulo θ y no de su signo. Esto permite caracterizar el efecto de este error en la reconstrucción de cada plano utilizando un parámetro escalar, que se denominará *deviationrate*(dr), definido como:

$$deviationrate = \frac{\varepsilon_{normal}}{|\varepsilon|} \tag{5.17}$$

Esta variable es la pendiente de la recta para los valores positivos del error (ϵ) en el ángulo θ (ver figura 5.3). Esto reduce la dimensionalidad del problema y permite una representación compacta en un solo gráfico del efecto de ϵ sobre todos los planos posibles que sirven para reconstruir una escena genérica.

A pesar de esto, todavía hay tres variables independientes que caracterizan cada plano posible en la escena, ρ_X, ρ_Y y ρ_Z, y una variable dependiente, *deviationrate*. Así, para la representación tridimensional es necesario fijar el valor de una de las variables independientes, en este caso ρ_X. La elección de la variable que se fija para la representación es arbitraria y no obedece a ningún criterio en particular.

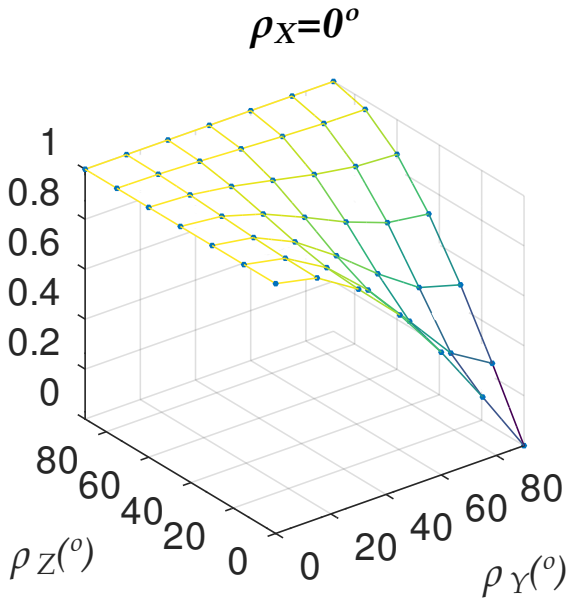


Figura 5.4: *Deviationrate* como función de ρ_Y y ρ_Z, para $\rho_X = 0°$.

La figura 5.4 muestra la gráfica correspondiente a un ángulo $\rho_X = 0°$. El comportamiento observado varía dependiendo del valor de ρ_X considerado, tal como se muestra en la figura 5.5, que ilustra esta misma representación para valores de ρ_X entre 0° y 90° con incrementos de 15° entre un gráfico y el siguiente:

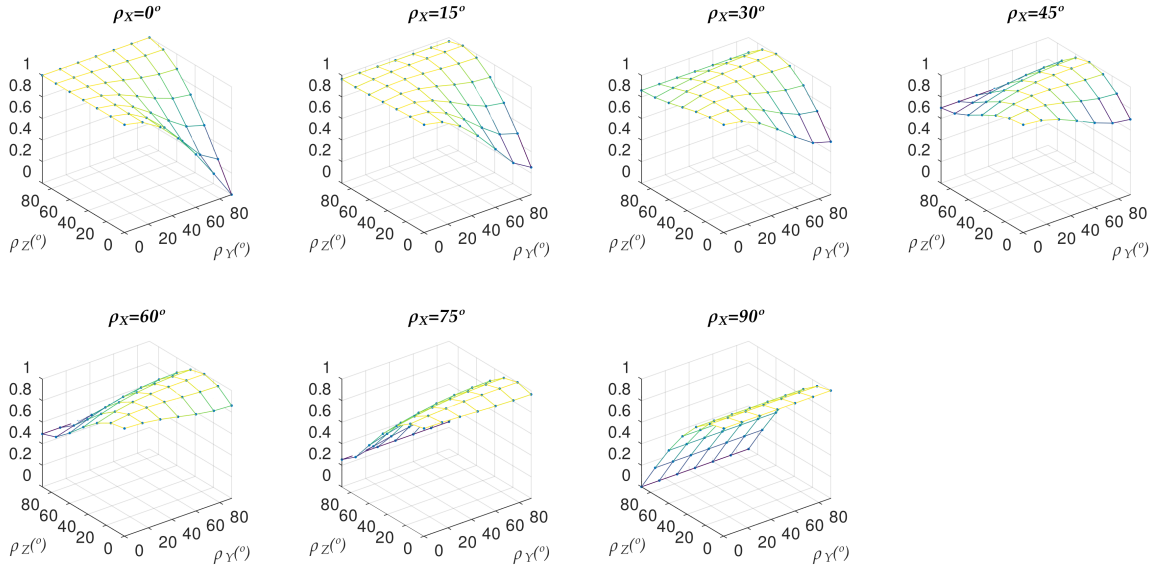
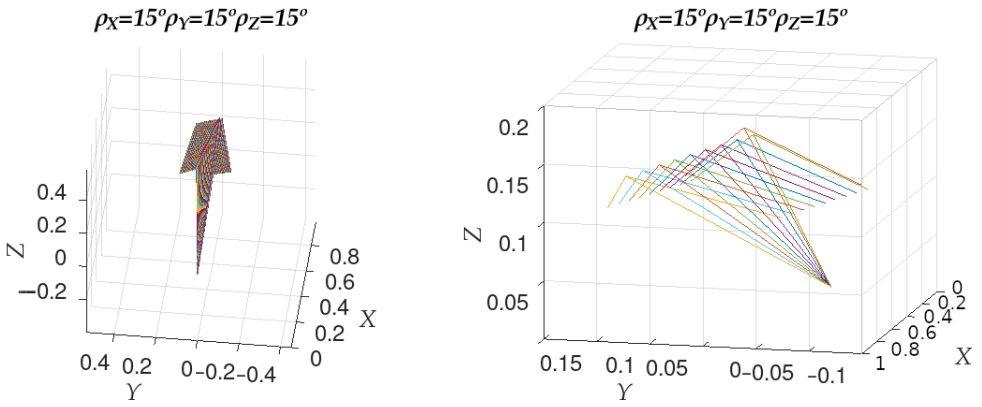


Figura 5.5: *Deviationrate* variando ρ_Y y ρ_Z, para $\rho_X = 0º, 15º, 30º, 45º, 60º, 75º, 90º$.

Como muestra la figura 5.5, el valor de *deviationrate* está acotado entre 0 y 1. La medición de esta variable para los 6859 planos del conjunto de prueba produce el mismo resultado en todos los casos.

Para caracterizar completamente la desviación entre los planos ideal y calculado es necesario determinar también el eje en torno al cual se produce la misma. El eje de rotación se puede obtener como el producto vectorial de la normal de los planos ideal y calculado y se puede representar por sus componentes en cada eje: w_x, w_y y w_z. Los diferentes ejes de giro obtenidos para cada valor de ϵ para un plano ideal específico se muestran en la figura 5.6. La figura 5.6b muestra un primer plano de la figura 5.6a con un rango mucho más pequeño en los tres ejes, lo que proporciona una vista más detallada. Se observa que los ejes de rotación son coplanares y simétricos alrededor del valor de $\epsilon = 0°$ dentro del plano que los contiene. Dado que el eje de giro se obtiene como el producto vectorial de las normales de los planos ideal y calculado, y el plano ideal para toda la observación es el mismo, este será el plano que contendrá todos los ejes de giro.



(a) Vista general. (b) Vista detallada.

Figura 5.6: Orientación de los ejes de rotación entre las normales de los planos ideal y calculados.

La información contenida en la figura 5.6 se puede mostrar superpuesta con el módulo de desviación entre los vectores normales a los planos ideal y calculado. Así, la figura 5.7 muestra cómo se distribuyen los ejes de rotación en relación con la desviación observada y se puede interpretar como una combinación de la representación de información de las figuras 5.3 y 5.6. De este modo, lo que se ha hecho es desplazar la representación del eje de rotación obtenido desde el origen a las coordenadas que (ϵ, ϵ_{normal}) para el que se ha calculado. De este modo tanto el eje X como el eje Y muestran una combinación de dos parámetros. Esto proporciona una representación compacta de la información sobre el ángulo y el eje de rotación que está presente entre las normales a los planos ideal y calculado.

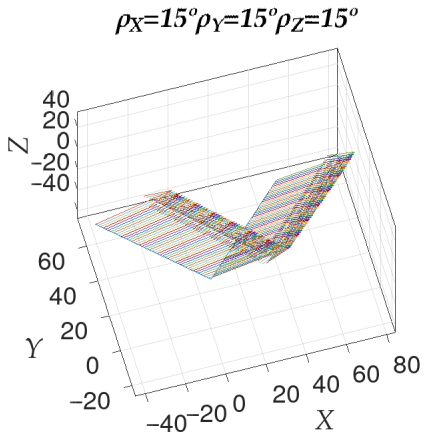


Figura 5.7: Orientación de los ejes de rotación entre las normales de los planos ideal y calculados, superpuesto con el valor de ϵ que produce dicha rotación.

5.3.5. Análisis de los resultados obtenidos

Los gráficos que se muestran en la figura 5.3 demuestran claramente que el efecto de un error en el ángulo q sobre los planos resultantes de la reconstrucción de una escena depende, en general, del plano del que se trate. En realidad, como se explica en la sección anterior, depende de su orientación con respecto al plano de la imagen, pero no de su distancia al mismo.

En la tabla 5.1 se muestra un los valores de *deviationrate* obtenidos y la frecuencia con la que se producen. Dichos resultados deben interpretarse de la siguiente manera: si se observa, por ejemplo, la fila correspondiente al rango 0,4 $<dr <= 0{,}5$, se ve que el 4,23 % de los planos en el conjunto de prueba tienen un tasa de desviación en este rango. Además, el 10,12 % de los planos del mismo conjunto tienen una tasa de desviación menor o igual a 0,5.

Tabla 5.1: Columna izquierda: intervalos de la tasa de desviación (entre 0 y 1); columna central: porcentaje de ocurrencia o porcentaje de planos en el conjunto de prueba que caen dentro de cada intervalo considerado; columna derecha: porcentaje acumulado o porcentaje de planos en el conjunto de prueba cuya tasa de desviación es menor o igual al valor superior del rango considerado.

Deviationrate	**Ocurrencia %**	**Acumulado %**
0,0 <dr <= 0,1	0,90 %	0,90 %
0,1 <dr <= 0,2	1,04 %	1,94 %
0,2 <dr <= 0,3	1,57 %	3,51 %
0,3 <dr <= 0,4	2,38 %	5,89 %
0,4 <dr <= 0,5	4,23 %	10,12 %
0,5 <dr <= 0,6	4,20 %	14,32 %
0,6 <dr <= 0,7	7,38 %	21,69 %
0,7 <dr <= 0,8	8,69 %	30,38 %
0,8 <dr <= 0,9	15,56 %	45,94 %
0,9 <dr <= 1,0	54,06 %	100,00 %

En vista de la tabla 5.1, se verifica que para todos los planos en el conjunto de prueba, el valor de la tasa de desviación siempre está en el intervalo [0,1]. Esto implica que el ángulo formado por las normales de los planos ideal y calculado es, en el peor de los casos, igual al error en el ángulo θ.

Este resultado es importante, ya que garantiza que la técnica presentada no introduce errores en el proceso de reconstrucción y acota el error máximo que se comete durante el proceso al valor de la desviación del sistema de visión real respecto al ideal.

Para profundizar en este análisis, los datos contenidos en la tabla 5.1 se representan gráficamente en la figura 5.8. Poniendo el foco en el porcentaje acumulado de valores posibles para la tasa de desviación variable, se observa que se comporta como la función de densidad

de una variable aleatoria con distribución beta, tal y como se establece en Devore et al. (2009) y Wackerly et al. (2002), cuyos parámetros son $\alpha = 2$ y $\beta = 0,4$.

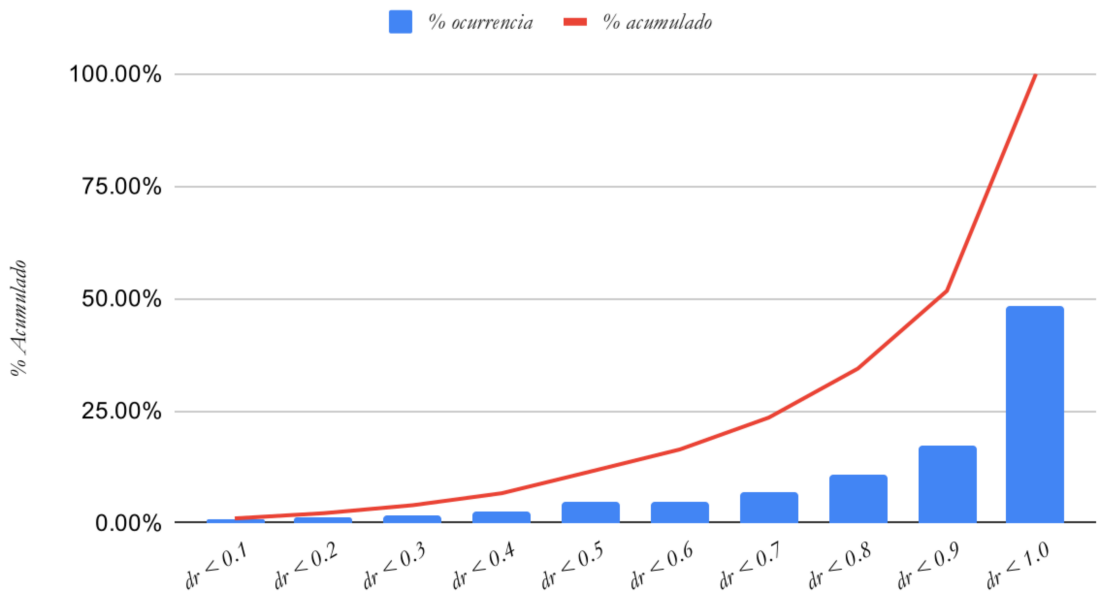


Figura 5.8: Representación gráfica de los datos de la tabla 5.1.

Estos resultados están respaldados por el diagrama de Pareto (ver Wilkinson (2006)) de los resultados anteriores que se muestran en la figura 5.9. Se puede ver claramente que, estadísticamente, más de la mitad de los planos presentarán una tasa de desviación entre 0,9 y 1, lo que significa que el error de reconstrucción será prácticamente idéntico al del sistema de visión utilizado.

Si se analiza el otro extremo del gráfico de Pareto, se detecta un pequeño porcentaje de planos cuya tasa de desviación es nula o prácticamente cero. Si se centra el análisis en estos planos, se comprueba que corresponden a aquellos con $\rho_X = 90°$ y $\rho_Z = 90°$; es decir, son planos verticales perpendiculares al plano de la imagen y, por tanto, insensibles a errores en el ángulo de cabeceo (θ).

Los resultados cuantitativos presentados anteriormente pueden utilizarse para extraer una serie de conclusiones a tener en cuenta a la hora de utilizar la técnica de reconstrucción de escenas descrita. Primero, se ha medido la sensibilidad del sistema propuesto a errores en el ángulo de cabeceo (θ) determinando su límite superior. Esto permite garantizar que el plano obtenido por la reconstrucción nunca se desviará del ideal en un ángulo mayor que el error en el ángulo $\theta(\epsilon)$.

Además del resultado anterior, como ya se ha anticipado, se ha definido un parámetro que permite anticipar el comportamiento de la reconstrucción en función del error del ángulo de inclinación (ϵ). Este parámetro, la tasa de desviación (*deviationrate*), se ha modelado

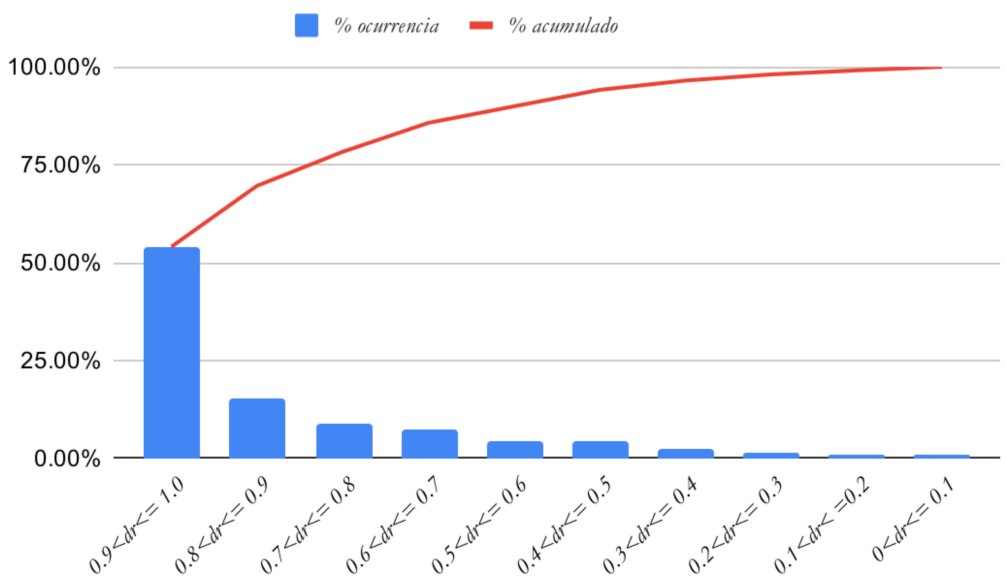


Figura 5.9: Diagrama de Pareto de los datos de la tabla 5.1.

como una variable aleatoria que sigue una distribución beta. Este resultado complementa la conclusión anterior ya que, además de haber encontrado un límite superior para el error, la existencia de un modelo permite estimar la desviación esperada.

Tabla 5.2: Columna izquierda: intervalos de la tasa de desviación (entre 0 y 1); columna derecha: porcentaje de ocurrencia o porcentaje de planos en el conjunto de prueba que caen dentro de cada intervalo considerado.

Deviationrate	**Ocurrencia %**
0,7 <dr <= 0,8	8,69 %
0,8 <dr <= 0,9	15,56 %
0,9 <dr <= 1,0	54,06 %

Analizando los datos de la tabla 5.2 obtenidos a partir de la información mostrada en la tabla 5.1 y el diagrama de Pareto de la figura 5.9, se puede concluir que en el 78,31 % (8,69 % + 15,56 % + 54,06 %) de los casos, el valor de *deviationrate* será mayor que 0,7. Esto significa que para casi el 80 % de los planos del conjunto de prueba, el plano calculado se desviará del ideal en un ángulo de, al menos, el 70 % del valor de ϵ. Además, en más de la mitad de los casos (54,06 %), el valor de dicho ángulo será superior al 90 % del valor de ϵ.

Capítulo 6

Calibración de un sistema de visión estereoscópico

En este capítulo se propone un método novedoso para la calibración de un sistema de cámaras estéreo para reconstruir escenas 3D. Tal como se ha analizado en la sección 5.3, un error en el ángulo de inclinación de las cámaras hace que la escena reconstruida presente cierta distorsión con respecto a la escena real. Para realizar el procedimiento de calibración, cuyo fin es eliminar o al menos minimizar dicha distorsión, se han utilizado técnicas de Aprendizaje Automático (Machine Learning) y, más concretamente, algoritmos de regresión. Estos algoritmos se han entrenado de manera supervisada con una gran cantidad de vectores de características de entrada con sus respectivas salidas, ya que, de cara a la aplicación del procedimiento propuesto, es importante que el conjunto de entrenamiento sea suficientemente representativo de la variedad de planos que pueden existir en una escena real. Para ello se consideran los diferentes valores que puede tomar el ángulo de cabeceo de las cámaras estéreo, el error en dicho ángulo y el efecto que todo ello tiene en el proceso de reconstrucción. Una vez estimado el error, se podría corregir haciendo que la escena reconstruida se ajuste mejor a la real. Aunque el método propuesto se basa en la Disparidad U-V y se emplea esta misma técnica para reconstruir completamente la escena 3D, una de las características más interesantes del mismo es que se puede aplicar independientemente de la técnica utilizada para llevar a cabo dicha reconstrucción.

6.1. Trabajo relacionado

Como es bien sabido, la calibración es un proceso que consiste en comparar los valores obtenidos con un determinado instrumento o técnica de medición con la medición correspondiente de un estándar de referencia. En la literatura existen innumerables aportaciones en las que se proponen diferentes técnicas para la calibración de sistemas de visión tanto monoculares como binoculares. Aunque se pondrá el foco en este último caso, en Itu and

Danescu (2020) se propone la calibración automática de un sistema de visión monocular para extraer y rastrear datos 3D de obstáculos del entorno circundante a un vehículo mientras circula.

En el campo de la visión estereoscópica el problema de la calibración es más amplio, ya que existen diferentes elementos a calibrar. Por ejemplo, las cámaras se pueden calibrar individualmente, en cuyo caso el enfoque adoptado en Itu and Danescu (2020) y artículos similares sería válido; pero también se puede calibrar la posición de una cámara con respecto a la otra, e incluso la posición del par estéreo completo. En cualquier caso, este problema ha sido considerado y resuelto utilizando una amplia gama de enfoques. A continuación se muestra una selección de algunos de ellos en el contexto de la reconstrucción 3D, que simplemente busca ilustrar la gran variedad de trabajos en este campo. En Wang et al. (2013), los autores se centran en estudiar la relación entre la precisión de la reconstrucción y la precisión de la calibración, discutiendo los principales factores involucrados en los errores de reconstrucción. En Bier and Luchowski (2009) se aborda el problema de la propagación de errores de datos de entrada en el proceso de estereovisión y su influencia en la calidad de los puntos 3D reconstruidos. En este caso los autores se centran en la calibración de cámaras y algoritmos de reconstrucción 3D que emplean métodos basados en la descomposición de valores singulares (SVD). En Marita et al. (2007) se presenta un conjunto de métodos originales para evaluar la precisión de los parámetros de la cámara. Algunos de ellos tienen una influencia crítica en la calidad del proceso de reconstrucción 3D y, por tanto, requieren una atención especial durante el proceso de calibración. Si bien los trabajos mencionados se centran más en analizar la influencia de los errores de calibración en la calidad de la reconstrucción 3D, en otros casos se proponen diferentes métodos para realizar esta calibración. Por ejemplo, en Sui and Zhang (2010) los autores emplean una placa plana especial que utiliza puntos circulares de diferentes tamaños para calibrar las cámaras. Basado en un algoritmo existente para calibrar una cámara, en este trabajo los parámetros de distorsión intrínseca y radial de la lente se pueden calcular usando varios pares de imágenes de la placa plana que son capturadas por dos cámaras en diferentes orientaciones. En Prokos et al. (2010) se presenta un escáner de superficie 3D que combina estereovisión y escaneo por hendidura (slit-scanning). El par de cámaras se calibra utilizando pares de imágenes sincronizadas de un tablero de ajedrez codificado.

En otros trabajos, las aplicaciones para sistemas de navegación autónomos recurren con frecuencia a marcas en la carretera para calibrar el sistema de visión. Por ejemplo, en Xu et al. (2014) el sistema propuesto detecta pasos de peatones y extrae sus esquinas en las dos imágenes del par estéreo, que se comparan en un paso posterior. Los autores demuestran que la técnica propuesta es capaz de extraer y hacer coincidir con precisión las esquinas de los pasos de peatones, así como de recalibrar el sistema de estereovisión de forma más precisa y robusta que con otros procedimientos similares.

En esta misma línea también se encuentran varios procedimientos de calibración automática patentados. Por citar dos ejemplos, en Marquet (2007) el autor propone un sistema a bordo para vehículos de motor en el que los errores de cabeceo y guiñada se determinan a partir de las imágenes derecha e izquierda de una misma escena, que debe contener al menos un carril de la vía por la que circula el vehículo. En la invención presentada en Lindgren (2012), también para vehículos de motor, los autores proponen un sistema para estimar con precisión en tiempo real el valor del error en el ángulo de guiñada entre las cámaras izquierda y derecha. Tal y como se describe en la patente, se resuelve un conjunto de ecuaciones utilizando un método de resolución de ecuaciones no lineales para obtener un valor estimado para el error de guiñada intrínseco.Dentro de los problemas abordables mediante técnicas de visión, el problema de detección es más complejo que el de la identificación ya que, además de determinar qué hay en la imagen, hay que indicar dónde se encuentra. Ejemplos de soluciones a este problema con redes convolucionales son el R-CNN descritos en Girshick (2015) y también en Ren et al. (2017) o YOLO: You Only Look Once, presentado en Redmon et al. (2016).

Más allá del uso de técnicas clásicas, las técnicas de Inteligencia Artificial y, más concretamente, las técnicas de Aprendizaje Automático y Aprendizaje Profundo , se vienen aplicando desde hace años para resolver problemas de muy diferente índole, arrojando resultados satisfactorios en muchos casos. Numerosos autores utilizan este tipo de técnicas para calibrar diferentes dispositivos. Por ejemplo, en Kolakowski (2021), se entrenan cuatro modelos diferentes para la calibración de mapas de radio en sistemas de posicionamiento basados en huellas dactilares RSS (RSS-fingerprinting-based positioning system). Estos cuatro modelos son: ajuste de un modelo de pérdida de trayectoria logarítmica de distancia, regresión del proceso gaussiano, red neuronal artificial y regresión de bosque aleatorio. En Amroun et al. (2021), el Aprendizaje Automático se utiliza para calibrar un modelo físico que intenta reproducir el comportamiento vibratorio de un conductor de línea aérea, lo cual es un enfoque novedoso para el problema en cuestión. Para lograr su objetivo, los autores entrenan y prueban diferentes modelos. La literatura también contiene una gran cantidad de aplicaciones en el campo de la visión por computador. En Akinyelu and Blignaut (2020) se presenta un estudio del estado del arte de las técnicas de estimación de la mirada basadas en el Aprendizaje Profundo, centrándose en las redes neuronales convolucionales (CNN). El mismo trabajo presenta una revisión de otras técnicas de estimación de la mirada basadas en Aprendizaje Automático. Este estudio tiene como objetivo brindar a la comunidad investigadora conocimientos valiosos y útiles que puedan mejorar el diseño y el desarrollo de modelos de seguimiento ocular mejorados y eficientes basados en el Aprendizaje Profundo. En Cirillo et al. (2021) los autores proponen una solución robótica basada en visión para la inserción de cables en la que se utiliza el conocimiento a priori del escenario con una cámara RGB calibrada y un brazo robótico. La solución presentada combina diferentes técnicas basadas en gradientes, clasificadores entrenados y estereovisión para obtener imágenes estándar. Otra

aplicación del Aprendizaje Automático en el campo de la visión por computador se presenta en Banús et al. (2021). En este caso se propone una solución basada en Aprendizaje Profundo para controlar automáticamente el cierre y sellado de envoltorios de pizza. Para entrenar las redes los autores proponen una clasificación de los defectos de los envoltorios de pizza centrada en el sellado y el cierre y un método basado en imágenes capaz de detectarlos automáticamente. Por citar tres ejemplos más, en Donné et al. (2016) los autores proponen una red neuronal convolucional (CNN) entrenada con un gran conjunto de imágenes de tablero de ajedrez de ejemplo para la calibración de la cámara. En Wang et al. (2020), se aplican regresión lineal múltiple, regresión de vectores de soporte y bosques de regresión aleatoria para calibrar sensores de monitoreo del aire de bajo costo en el campo. El mismo problema se considera en Vajs et al. (2021). En este caso, los autores exploran métodos para mejorar aún más los algoritmos de calibración para aumentar la precisión de la medición al considerar el impacto de la temperatura y la humedad en las lecturas, mediante el uso del Aprendizaje Automático. Se presenta un análisis comparativo detallado de las prestaciones alcanzadas con regresión lineal, redes neuronales artificiales y algoritmos de bosque aleatorio.

En cuanto al uso de la Inteligencia Artificial para calibrar sistemas de visión estereoscópica, existen interesantes aplicaciones en la literatura. Por ejemplo, en Chen et al. (2020) los autores proponen un método basado en una red neuronal de Back Propagation optimizada con un algoritmo de recocido genético simulado mejorado (IGSAA-BP, Back Propagation neural network optimized with an Improved Genetic Simulated Annealing Algorithm) para calibrar una cámara binocular, mejorando la precisión y velocidad de convergencia que se obtienen con redes neuronales de retropropagación (Back Propagation Neural Networks). Los algoritmos genéticos y el recocido simulado son técnicas de optimización que buscan encontrar soluciones de alta calidad en espacios de búsqueda complejos. En Hu (2006), se utiliza una red de función de base radial (RBFN, Radial Basis Function Network) para proporcionar metodologías efectivas para resolver problemas computacionales difíciles tanto en la calibración de la cámara como en el proceso de reconstrucción 3D.

6.2. Aprendizaje Automático

El Aprendizaje Automático (Machine learning) es un subconjunto de la Inteligencia Artificial (IA). Este área de conocimiento se enfoca en enseñar a los computadores para que aprendan de los datos y mejoren con la experiencia, en lugar de ser explícitamente programados para hacerlo. En el Aprendizaje Automático, se usan algoritmos para encontrar patrones y correlaciones en grandes conjuntos de datos, y para tomar las mejores decisiones y previsiones en base a ese análisis. En general, las aplicaciones basadas en Aprendizaje Automático mejoran con el uso y se vuelven más precisas a medida que tienen acceso a más datos.

6.2.1. Tipos de aprendizaje

Como es bien sabido, el Aprendizaje Automático es un campo de estudio muy amplio con un gran número y variedad de aplicaciones. En esta sección se incluye un análisis sucinto de sus principales técnicas y estrategias que permitirá determinar cuáles se adaptan mejor al problema de la estimación del error en el ángulo de cabeceo de las cámaras, necesario para poder llevar a cabo el método de calibración que se presenta.

El Aprendizaje Automático se compone de diferentes tipos de modelos y utiliza varias técnicas algorítmicas. Dependiendo de la naturaleza de los datos y el resultado deseado, se puede utilizar uno de los cuatro modelos de aprendizaje: supervisado, no supervisado, semisupervisado o de refuerzo. Dentro de cada uno de esos modelos se pueden aplicar una o más técnicas algorítmicas, en función de los conjuntos de datos en uso y los resultados que se buscan. Los algoritmos de machine learning básicamente están diseñados para clasificar cosas, encontrar patrones, proyectar resultados, y tomar decisiones fundamentadas. Los algoritmos pueden utilizarse de manera independiente o combinarse para lograr la mayor precisión posible cuando se trata de datos complejos y más impredecibles.

Aprendizaje supervisado

En los algoritmos de aprendizaje supervisado, a la máquina se le enseña mediante ejemplos. Los modelos de aprendizaje supervisados consisten en pares de datos de 'entrada' y 'salida', donde la salida se etiqueta con el valor deseado. Mediante un algoritmo, el sistema compila todos estos datos de entrenamiento a lo largo del tiempo y comienza a determinar similitudes correlativas, diferencias y otros puntos de lógica. Los modelos de aprendizaje supervisado se utilizan en muchas aplicaciones como motores de recomendación para productos y aplicaciones de análisis de tráfico que prevén la ruta más rápida en diferentes horas del día.

Aprendizaje no supervisado

En los modelos de aprendizaje no supervisado, no existe una clasificación de la respuesta. La máquina estudia los datos de entrada, muchos de los cuales no están etiquetados ni estructurados, y comienza a identificar patrones y correlaciones, utilizando todos los datos relevantes y accesibles. En muchos sentidos, el aprendizaje no supervisado sigue el modelo de cómo los humanos observan el mundo, utilizando la intuición y la experiencia para agrupar cosas. A medida que se experimentan más ejemplos de algo, la capacidad de categorizar e identificar se vuelve cada vez más precisa. Para las máquinas, la 'experiencia' se define por la cantidad de datos que se introducen y se ponen a disposición. Ejemplos comunes de aplicaciones de aprendizaje no supervisado incluyen el reconocimiento facial, el análisis de secuencias genéticas, la investigación de mercado y la ciberseguridad.

Aprendizaje semisupervisado

En condiciones ideales, es posible disponer de todos los datos estructurados y etiquetados antes de ser introducidos en un sistema. Pero como en la práctica esto no es siempre factible, el aprendizaje semisupervisado se convierte en una solución viable cuando hay grandes cantidades de datos crudos y no estructurados. Este modelo consiste en introducir pequeñas cantidades de datos etiquetados dentro de los conjuntos de datos sin etiquetar. Esencialmente, los datos etiquetados actúan para dar un inicio de funcionamiento al sistema y pueden mejorar considerablemente la velocidad y precisión del aprendizaje. Un algoritmo de aprendizaje semisupervisado instruye a la máquina para que analice los datos etiquetados según propiedades correlativas que podrían aplicarse a los datos no etiquetados.

Aprendizaje por refuerzo

En el aprendizaje supervisado, la máquina recibe la respuesta de referencia y aprende encontrando correlaciones entre todos los resultados correctos. El modelo de aprendizaje por refuerzo no incluye una respuesta de referencia, sino que introduce un conjunto de acciones permitidas, reglas y estados finales potenciales. Cuando el objetivo deseado del algoritmo es fijo o binario, las máquinas pueden aprender mediante el ejemplo. Pero en los casos en los que el resultado deseado es variable, el sistema debe aprender por experiencia y recompensa. En los modelos de aprendizaje por refuerzo, la 'recompensa' es numérica y se programa dentro del algoritmo como algo que el sistema busca recopilar. Algunos ejemplos de aplicaciones del aprendizaje por refuerzo son la puja de precios automatizada para los compradores de publicidad on-line, el desarrollo de videojuegos, y la negociación bursátil de alto riesgo.

6.2.2. Aplicación de técnicas de Aprendizaje Automático al sistema de calibración propuesto

Para resolver el problema que se aborda en este capítulo se dispone de una gran cantidad de datos etiquetados. Es por ello por lo que de entre los diferentes tipos de algoritmos de Aprendizaje Automático descritos, los que mejor se adaptan a las necesidades del problema de calibración son los correspondientes al aprendizaje supervisado.

Algoritmos de aprendizaje supervisado

El aprendizaje supervisado puede utilizar algoritmos de clasificación o regresión, según el tipo de datos de salida. Los algoritmos de clasificación se utilizan cuando el resultado es una etiqueta discreta, es decir, cuando la respuesta se basa en un conjunto finito de resultados.

Por el contrario, el análisis de regresión es el subcampo cuyo objetivo es obtener un método para establecer la relación entre un cierto número de características y una variable objetivo continua. Esto significa que la respuesta a la pregunta se presenta mediante una

cantidad que puede determinarse de manera flexible en función de las entradas del modelo en lugar de limitarse a un conjunto de etiquetas. En algunos casos, el valor predicho se puede usar para identificar la relación entre los atributos de entrada.

Algoritmos de aprendizaje supervisado

Dado que el sistema de calibración propuesto tiene como objetivo estimar el valor de ϵ (la diferencia entre los ángulos de cabeceo calculado e ideal, tal como se definió en la sección 5.3.2) con la mayor precisión posible, siendo ϵ una variable continua, el problema se abordó utilizando diferentes regresores. En concreto, se realizó un estudio comparativo entre:

- Regresión lineal (LR).
- Árboles de regresión (RT).
- Bosque de regresión (RF).
- Redes Neuronales Multicapa (NN).

Los regresores seleccionados para la comparativa han sido los que, de un conjunto más amplio de candidatos, han obtenido mejores resultados. Otros regresores como por ejemplo las Máquinas de Soporte Vectorial (Support Vector Machine - SVM) o la Regresión de Processos Gaussianos (Gaussian Process Regression) fueron testeados pero debido a las características de los datos utilizados para el entrenamiento no produjeron resultados aceptables.

Las herramientas empleadas para implementar los regresores y analizar su rendimiento han sido MATLAB, MATLAB's Statistics and Machine Learning Toolbox y Regression Learner App (Mathworks (2022)). El uso de estas herramientas en concreto ha permitido trabajar con todos los regresores seleccionados de una forma homogénea e integrada. Es decir, se ha podido garantizar que todos los regresores han utilizado los mismos conjuntos de datos y la misma representación de los mismos ya que no existe diversidad de herramientas para su implementación. Además, esta herramienta integrada ha facilitado el uso de los regresores paramétricos al proporcionar funciones que permiten el ajuste automático progresivo de dichos parámetros.

La aplicación Regression Learner forma parte de la toolbox de Machine Learning de Matlab y facilita el entrenamiento de modelos de regresión, incluidos modelos de regresión lineal, árboles de regresión, modelos de regresión de procesos gaussianos, máquinas de vectores de soporte y conjuntos de árboles de regresión. Además de entrenar modelos, permite explorar sus datos, seleccionar características, especificar esquemas de validación y evaluar resultados. Asimismo, el modelo resultante puede exportarse al espacio de trabajo para utilizarlo con nuevos datos o generar código MATLAB® para su uso de manera independiente.

Para todos los regresores con los que se ha trabajado se consideraron las siguientes características de entrada, las cuales serán descritas en detalle en la sección 6.4:

- Vector normal al plano ideal.
- Vector normal al plano calculado.
- Valor del ángulo θ.
- Valor del ángulo ϵ_{normal}.
- Vector del eje de rotación entre normales.

En todos los casos, la característica de salida del sistema es ϵ.

Debido a las características del problema en cuestión, que no considera cambios de posición del sistema de visión a lo largo del tiempo o imágenes consecutivas con una relación temporal entre ellas, sino que se enfoca en el análisis de posiciones e imágenes instantáneas, se han descartado otros enfoques potenciales. Un ejemplo es el descrito en Kiranyaz et al. (2021) que hace uso de redes neuronales convolucionales. Dado que cada característica de entrada consta de un conjunto discreto de datos, no hay ningún beneficio en hacer convoluciones en las diferentes capas de la red, por lo que no se consideró este enfoque.

Como se explicó en la sección 5.3.4, el eje de rotación de la desviación entre los planos es una variable que se puede calcular a partir de los vectores normales a los planos ideales y calculados. Sin embargo, dada la naturaleza de los regresores seleccionados, se han comparado los resultados de la regresión incluyendo la información de los ejes y prescindiendo de ella, lo que ha permitido determinar si su aportación es o no relevante.

6.3. Técnica de calibración

El método propuesto para la calibración del sistema de visión mediante regresores se resume en la figura 6.1 y consta de los siguientes pasos:

1. Construcción de un conjunto de entrenamiento suficientemente representativo de la variedad de planos que se pueden presentar en una escena real, así como de las distintas orientaciones que el ángulo de cabeceo (θ) y su error (ϵ) pueden tener, y su efecto en el proceso de reconstrucción.
2. Entrenar el regresor seleccionado para que, en función de los datos del conjunto de entrenamiento, estime el valor de ϵ del conjunto de características descritas en la sección 6.2. Este proceso incluye tanto la etapa de entrenamiento como la de validación de dicho entrenamiento.
3. Una vez entrenado el regresor, se presenta al sistema de visión un conjunto de planos de calibración cuyo vector normal se conoce. Dado que también se conoce el ángulo de cabeceo, el proceso descrito en la sección 5.2 devolverá el vector normal calculado para cada uno de los planos.

4. El conjunto de datos de calibración (vectores normales a los planos ideales, vectores normales a los planos calculados y ángulo de cabeceo) se usan como entrada del regresor para que pueda estimar el valor de ϵ.

5. Para cada plano de calibración se obtendrá un valor de ϵ. Este valor será muy similar, aunque no idéntico, para los diferentes planos del conjunto de calibración. Debido a esto, existe una etapa final para determinar la salida del sistema de calibración, que consiste en tomar el valor medio de los ϵ obtenidos para cada plano de calibración.

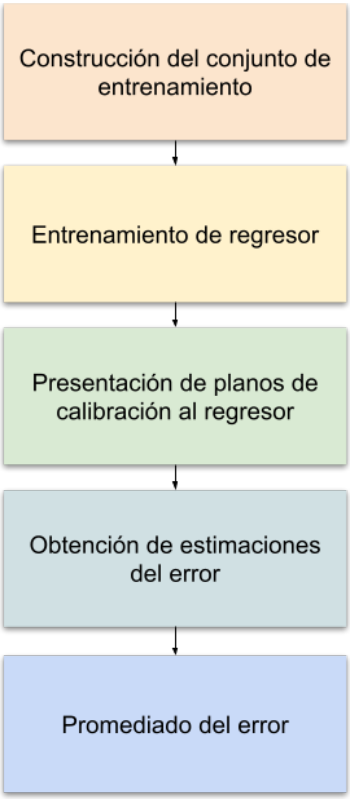


Figura 6.1: Representación esquemática de los pasos que componen el proceso de calibración.

Una vez completado este proceso se dispone de un valor de ϵ, es decir, se conocerá una estimación del error existente en el ángulo de cabeceo. Esta información resulta de interés de cara a corregir las imágenes que proporciona el sistema de visión para cualquier uso posterior. Por ejemplo, se podría incluir esta corrección en un proceso que corrija la proyeción de las imágenes una vez han sido capturadas, de modo que todos los procesos posteriores dispongan de esta información corregida. Otra aproximación al uso de este resultado es en cualquier post procesamiento para extracción de información que se haga de las imágenes en el cual intervenga el ángulo de cabeceo. Puesto que se conoce el error que introduce el sistema de visión, es posible utilizar $\theta_{corregido} = \theta + \epsilon$ en lugar de θ en los cálculos que se realicen, corrigiendo así la influencia del error en el ángulo de cabeceo. Esta versatilidad deriva del

hecho de que el método de calibración propuesto es independiente de las técnicas de análisis de imágenes que se quieran llevar a cabo con la salida del sistema estereoscópico, siendo una característica interesante del mismo.

6.4. Datos de entrada a los regresores

Como se anticipó en la sección 6.2, se ha seleccionado un conjunto de características de entrada para entrenar a los regresores de modo que se pueda caracterizar adecuadamente la relacion entre el contenido de la escena y el estado del sistema de visión.

Dado que el objetivo es determinar el valor del error (ϵ) del ángulo de cabeceo (θ), es necesario conocer la transformación que se produce al reconstruir la información tridimensional una vez que es captada por el sistema de visión sujeto a error.

La forma escogida para representarlo, en base a la descripción dada en la sección 5.2, fue incluir en el vector de características el vector normal al plano ideal que se está considerando, ya que describe completamente dicho plano. De manera similar, el resultado de la transformación se puede representar a través del vector normal al plano calculado. Finalmente, el valor del ángulo de cabeceo (θ) del sistema de visión en el momento en que se evalúa el plano, junto con el error (ϵ), completa la descripción de las características.

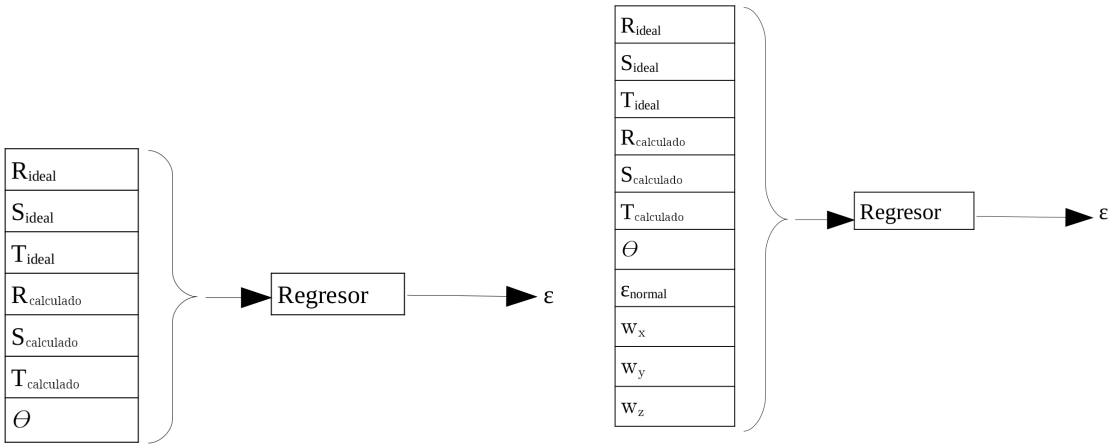
Además de estas características, es posible calcular el ángulo de rotación ϵ_{normal} entre las normales a los planos, y el eje alrededor del cual ocurre la rotación. Como se mencionó en la sección 6.2, se realizará un estudio comparativo para determinar si la inclusión de estas últimas características permite mejorar los resultados de la regresión.

Teniendo en cuenta que tanto el eje de giro entre las normales como el vector normal al plano ideal y el vector normal al plano calculado son variables tridimensionales, el vector de características de entrada al regresor, formado por ambos vectores normales y el valor de θ, tiene 7 componentes. Al incluir el eje de rotación entre las normales y el valor de ϵ_{normal}, se añaden 4 componentes más, tal y como se muestra en la figura 6.2.

Tal y como se muestra en la figura 6.2, independientemente de la dimensión del vector de características, la dimensión del vector de salida es siempre 1×1, ya que la única variable que devuelve el regresor es el error en el ángulo de cabeceo.

6.5. Descripción de los regresores

Una vez presentadas las características de entrada y salida de los regresores, se procede a continuación a describir con más detalle los que han sido considerados en este estudio.



(a) Configuración con 7 entradas. (b) Configuración con 11 entradas.

Figura 6.2: Representación esquemática de la información que el regresor recibe como entrada y devuelve como salida (valor estimado de ϵ). A la izquierda, las entradas son los coeficientes de los planos ideal y calculado y el valor de θ. A la derecha, se han agregado cuatro nuevas características como entradas: los coeficientes del eje de rotación entre normales y el valor de ϵ_{normal}.

6.5.1. Regresión lineal

La regresión lineal (Neter et al. (1996) y Seber and Lee (2012)) es una técnica de modelado estadístico utilizada para describir una variable de respuesta continua en función de una o más variables predictoras. Estas técnicas se utilizan para crear un modelo lineal. El objetivo es determinar una serie de coeficientes que ponderan la contribución de cada característica de entrada o conjunto de características en la estimación del valor de salida.

Dentro de este tipo de regresores, los modelos *Lineal* y *LinealRobusto* aproximan una función lineal que no combina las variables de entrada. La diferencia entre ambos modelos está en el tipo de regresión lineal utilizada. Mientras que el modelo *Lineal* utiliza regresión lineal simple, el modelo *LinealRobusto* utiliza regresión lineal robusta.

Por otra parte, los modelos *Interactionslinear* y *Stepwiselinear* intentan aproximar una expresión que, además de las características de entrada, incluye combinaciones lineales entre ellas, tomadas en pares. En el caso de *Interactionslinear*, el proceso de entrenamiento considera los efectos de interacción entre las entradas para lograr el mejor ajuste. Finalmente, el proceso de entrenamiento *Stepwiselinear* agrega o quita términos al modelo.

6.5.2. Árbol de regresión

Los árboles de regresión (Breiman et al. (1984) y Shalev-Shwartz and Ben-David (2014)) son un método de regresión no paramétrico que crea un árbol binario dividiendo recursivamente

los datos en los valores predictores. Las divisiones se seleccionan de modo que los dos nodos secundarios tengan una variabilidad menor en torno a su valor promedio que el nodo principal.

Se han utilizado tres modelos con características fijas y uno optimizable. En los modelos con características fijas solo varía el tamaño mínimo de la hoja, siendo 4, 12 y 36 los valores predeterminados para los modelos Fino, Medio y Grueso, respectivamente. Por su parte, el modelo Optimizable trata de determinar el tamaño mínimo óptimo de hoja mediante optimización bayesiana.

6.5.3. Bosque de regresión

El bosque de regresión (Breiman (2001)) es un método de aprendizaje por regresión que funciona mediante la construcción de una multitud de árboles de decisión en el momento del entrenamiento. Se ha elegido un modelo optimizable que ajusta automáticamente lo siguiente: el método de ensemble, el tamaño mínimo de la hoja, el número de submodelos (*learners*), la tasa de aprendizaje y el número de predictores que se muestrearán.

6.5.4. Red neuronal multicapa

Como es bien sabido, las redes neuronales (Aggarwal et al. (2018) y Westfall and Arias (2020)) se inspiran en el funcionamiento del cerebro humano. Están formadas por diferentes nodos que funcionan como neuronas y se transmiten señales e información entre sí. Estas redes reciben diferente información de entrada, la procesan como un todo y generan una salida con las predicciones establecidas según hayan sido programadas. En el presente estudio, todos los modelos elegidos tienen una capa de entrada con una neurona por característica de entrada y una capa de salida con una sola neurona, donde la salida es el valor estimado para el error ϵ. En todos los casos se ha utilizado ReLu como función de activación. Lo que diferencia a los modelos es su arquitectura:

1. $NarrowNN$: una sola capa oculta con 10 neuronas completamente conectadas.
2. $MediumNN$: una sola capa oculta con 25 neuronas completamente conectadas.
3. $WideNN$: una sola capa oculta con 100 neuronas completamente conectadas.
4. $BilayeredNN$: dos capas ocultas con 10 neuronas completamente conectadas cada una.
5. $TrilayeredNN$: tres capas ocultas con 10 neuronas completamente conectadas cada una.

6.6. Implementación del sistema de calibración

En este apartado se explica cómo se han utilizado las herramientas software elegidas (MATLAB, MATLAB's Statistics and Machine Learning Toolbox y Regression Learner App) para implementar el sistema de calibración que se propone.

6.6.1. Diseño de experimentos

El propósito de esta parte de la investigación es presentar la técnica de calibración propuesta y demostrar su validez ante el problema planteado. Dado que el punto de partida es la información obtenida del procesamiento del par de cámaras estéreo, se ha optado por una metodología similar a la seguida en la sección 5.3.3, construyendo un conjunto de entrenamiento sintético y parametrizándolo de forma que represente fielmente la variabilidad de elementos que se pueden encontrar en una escena real.

Dicho conjunto de entrenamiento está compuesto por múltiples planos ideales en el sistema de coordenadas global representados por su ecuación paramétrica. El punto de partida es un plano paralelo al plano de la imagen. Este plano es modificado por transformaciones de rotación y traslación, generando nuevos elementos del conjunto de entrenamiento. Realizando este proceso de forma exhaustiva, es posible generar un conjunto de planos que resulte lo suficientemente representativo de los posibles elementos de una escena real. Estos planos se proyectarán en el sistema de visión descrito variando el valor de θ y ϵ para simular los diferentes escenarios a considerar. Estos valores se han limitado para evitar que el conjunto de entrenamiento crezca en exceso. Por ello se han considerado los escenarios más realistas basados en las características estructurales del sistema de visión. Finalmente, el plano proyectado se reconstruye para obtener la ecuación paramétrica del plano calculado correspondiente.

Siguiendo este procedimiento, con los parámetros seleccionados para el paso de los ángulos ρ_X, ρ_Y y ρ_Z así como las variaciones de θ y ϵ a considerar, se ha construido un conjunto de entrenamiento de algo menos de 300.000 (295.323) vectores de características de entrada, con sus respectivas salidas. Como se explica en 5.3.5, se comprobó que el error en el ángulo de cabeceo es invariante a la distancia, por lo que esta variable no se tuvo en cuenta a la hora de construir el conjunto de entrenamiento, ya que haría que este creciera en exceso sin aportar información relevante.

Se utilizó un proceso análogo para el conjunto de test, construyendo un conjunto de datos cuyo tamaño es el 25 % del conjunto de entrenamiento y que contiene datos dentro de los mismos rangos de valores.

Para seleccionar los regresores que mejor se adaptan al problema, se llevó a cabo el proceso de entrenamiento en dos etapas. En primer lugar, todos los regresores se entrenaron mediante un proceso de validación de retención (hold-out), con el 25 % de los datos para el conjunto de

validación. Posteriormente, los regresores más prometedores se volvieron a entrenar con un proceso de validación cruzada de 5 pliegues para obtener la mejor parametrización posible.

Las métricas seleccionadas para determinar qué regresores ofrecen los mejores resultados han sido el Root Mean Square Error (RMSE), que tiene la propiedad de estar en las mismas unidades que la variable de respuesta, y el Relative Absolute Error (RAE), que permite comparar el error promedio frente a los errores que se obtendrían utilizando un modelo trivial.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\epsilon_{calculado_i} - \epsilon_{ideal_i}\right)^2}{n}} \tag{6.1}$$

$$RAE = \frac{\sum_{j=1}^{n} \left|\epsilon_{calculado_j} - \epsilon_{ideal_j}\right|}{\sum_{j=1}^{n} \left|\epsilon_{ideal_j} - \overline{\epsilon_{ideal}}\right|} \tag{6.2}$$

Como es bien sabido, los valores más bajos de RMSE indican un mejor ajuste. De la misma forma, un buen modelo de predicción presentará un valor de RAE cercano a cero. RMSE es una buena medida de la precisión con la que el modelo predice la respuesta, mientras que RAE proporciona información en términos relativos de cuánto se desvía el modelo. Tanto la precisión como la desviación son los criterios más importantes para ajustar, ya que el objetivo principal del modelo es predecir. A la hora de decidir qué regresor es mejor que otro se ha optado por priorizar la precisión en la predicción de la repuesta. Por tanto, el criterio principal será el valor RMSE.

6.6.2. Proceso de calibración

Las métricas presentadas en la sección 6.6.1 pueden utilizarse para seleccionar el regresor que mejor se adapta al problema considerado. Una vez que se entrena esta herramienta, puede continuar el proceso de calibración del sistema de visión. Como se discutió en la sección 6.3, la técnica de calibración propuesta requiere un conjunto de planos de calibración cuyos coeficientes ideales, es decir, en ausencia de error en el ángulo de cabeceo de las cámaras, son conocidos.

El conjunto de planos de calibración se ha elegido seleccionando planos fácilmente reproducibles en una escena real, lo que facilita la aplicación práctica de la técnica descrita. Se debe tener en cuenta que los regresores podrían presentar un sesgo o no generalizar del mismo modo para todos los tipos de planos. Por este motivo, el proceso de calibración se realiza con múltiples planos, en diferentes configuraciones, de modo que se evite o minimice la posibilidad de obtener soluciones parciales.

Una vez seleccionados los planos, se procesan como se describe en la sección 5.2 para obtener los coeficientes calculados correspondientes a la reconstrucción para una configuración conocida del sistema de visión (θ y ϵ conocidos). Todos estos datos se utilizan para construir

el vector de características de entrada que procesará el regresor, devolviendo como salida el valor estimado de ϵ.

Finalmente, se lleva a cabo un proceso de consenso entre los valores de ϵ obtenidos para cada plano de calibración. Como se ha explicado con anterioridad, la herramienta elegida para esta fase es la media aritmética de los valores de ϵ.

Este procedimiento para calibrar un sistema de visión estereoscópica se repitió para múltiples valores de θ y ϵ. Esto permite afirmar que no solo se comporta como se esperaba para un conjunto discreto de casos, sino que puede estimar el valor de ϵ dentro de los límites de error establecidos considerando los escenarios más realistas, que corresponden a valores de ϵ y θ en los intervalos [-5º,5º] en pasos de 0,25 grados y [-10º,10º] en pasos de 1 grado, respectivamente. Estos rangos se han establecido tratando de que sean representativos del movimiento normal del ángulo de cabeceo en la operación de un sistema estereoscópico como el propuesto y unos valores de error no detectables a simple vista. Por todo esto, este es un resultado importante ya que implica que se trata de un método de calibración independiente del uso posterior que se quiera hacer de las imágenes obtenidas y la corrección del ángulo de cabeceo en base al error estimado que se realice.

6.7. Validación del sistema de calibración

En la sección anterior se presentó la metodología y los experimentos realizados para implementar el método de calibración propuesto. El propósito de esta sección es analizar los resultados y mostrar que validan la hipótesis planteada. Es decir, verificar que el sistema permite estimar el valor de ϵ de manera adecuada y, por tanto, es posible utilizar este valor para corregir el defecto detectado en el sistema de visión. Además, como se mencionó anteriormente, se comparan los resultados que se obtienen con dos configuraciones diferentes de características de entrada, usando un modelo cuyo vector de entrada consta de 7 características y otro de 11.

6.7.1. Resultados de los experimentos realizados

La tabla 6.1 muestra los resultados obtenidos para el conjunto de regresores propuesto en la sección 6.5 sin considerar las variables ϵ_{normal} ni el eje de rotación entre normales en el vector de características de entrada y usando una validación de retención, como se explicó anteriormente.

Tabla 6.1: Resultados del entrenamiento de los regresores sin considerar ϵ_{normal} ni el eje de rotación entre las normales con validación holdout

Regresor	**Variante**	**RMSE (°) Validación**	**RMSE (°) Test**	**RAE (%) Validación**	**RAE (%) Test**
LR	Linear	2,2470	2,2491	70,66	70,72
LR	Interactions	1,2526	1,2600	34,59	34,65
LR	Robust	2,5528	2,5501	68,70	68,85
LR	Stepwise	1,2525	1,2600	34,59	34,65
RT	Fine	0,4801	0,4870	4,11	4,26
RT	Medium	0,4797	0,4870	4,11	4,26
RT	Coarse	0,4851	0,4887	4,33	4,49
RT	Optimizable [1]	0,4797	0,4870	4,11	4,26
RF	Optimizable ensemble [2]	0,4707	0,4804	3,57	3,72
NN	Narrow	0,5684	0,6177	13,43	13,53
NN	Medium	0,4845	0,4930	6,98	7,09
NN	Wide	0,4569	0,4689	3,63	3,75
NN	Bilayered	0,4719	0,4841	6,01	6,15
NN	Trilayered	0,4729	0,4757	5,07	5,19

[1] Hiperparámetros optimizados: Tamaño mínimo de hoja = 14. [2] Hiperparámetros optimizados: Método de ensemble = Bag, Número de submodelos = 10, Tamaño mínimo de hoja = 1 y Número de predictores a muestrear = 3.

La tabla 6.2 muestra los resultados obtenidos para el conjunto de regresores propuesto en la sección 6.5 considerando ϵ_{normal} y el eje de rotación entre normales en el vector de características de entrada y usando una validación de retención, como se explicó anteriormente.

Tabla 6.2: Resultados del entrenamiento de los regresores considerando ϵ_{normal} y el eje de rotación de las normales con validación de retención.

Regresor	**Variante**	**RMSE (°) Validación**	**RMSE (°) Test**	**RAE (%) Validación**	**RAE (%) Test**
LR	Linear	1,4455	1,4479	45,36	45,31
LR	Interactions	0,6252	0,6368	12,59	12,68
LR	Robust	1,4479	1,4510	45,07	45,03
LR	Stepwise	0,6251	0,6368	12,59	12,68
RT	Fine	0,4707	0,4850	3,33	3,49
RT	Medium	0,4705	0,4850	3,33	3,49
RT	Coarse	0,4716	0,4858	3,44	3,60
RT	Optimizable [1]	0,4706	0,4858	3,44	3,60
RF	Optimizable ensemble [2]	0,4665	0,4838	4,08	4,23
NN	Narrow	0,4787	0,5035	7,45	7,57
NN	Medium	0,4587	0,4785	5,16	5,28
NN	Wide	0,4502	0,4692	3,69	3,81
NN	Bilayered	0,4571	0,4799	5,35	5,46
NN	Trilayered	0,4755	0,4795	5,35	5,47

[1] Hiperparámetros optimizados: Tamaño mínimo de hoja = 13. [2] Hiperparámetros optimizados: Método de ensemble = LBoost, Número de submodelos = 494, Tamaño mínimo de hoja = 12421 y Número de predictores a muestrear = 11.

Comparando los resultados que se muestran en las tablas 6.1 y 6.2 , se observa que los algoritmos más simples (regresores lineales) mejoran significativamente sus resultados al

utilizar el vector de características que incluye el ϵ_{normal} y el eje de giro entre las normales. Sin embargo, esta mejora no es tan apreciable en aquellos regresores que ya producían buenas aproximaciones de ϵ. Para tratar de obtener los mejores resultados posibles, todo el conjunto de regresores ha sido reentrenado considerando la ϵ_{normal} y el eje de rotación entre las normales, utilizando un proceso de validación de 5 pliegues. Los resultados se muestran en la tabla 6.3.

Tabla 6.3: Resultados del entrenamiento de los regresores considerando ϵ_{normal} y el eje de rotación entre las normales con validación de 5 pliegues.

Regresor	**Variante**	**RMSE (°) Validación**	**RMSE (°) Test**	**RAE (%) Validación**	**RAE (%) Test**
LR	Linear	1,4458	1,4479	45,36	45,31
LR	Interactions	0,6278	0,6368	12,59	12,68
LR	Robust	1,4484	1,4510	45,07	45,03
LR	Stepwise	0,6280	0,6368	12,59	12,68
RT	Fine	0,4711	0,4850	3,33	3,49
RT	Medium	0,4705	0,4850	3,33	3,49
RT	Coarse	0,4711	0,4858	3,44	3,60
RT	Optimizable [1]	0,4705	0,4850	3,33	3,49
RF	Optimizable ensemble [2]	0,4571	0,4715	3,39	3,52
NN	Narrow	0,4899	0,5173	8,39	8,50
NN	Medium	0,4635	0,4790	5,24	5,36
NN	Wide	0,4544	0,4690	3,64	3,75
NN	Bilayered	0,4650	0,4809	5,59	5,71
NN	Trilayered	0,4613	0,4778	5,11	5,24

[1] Hiperparámetros optimizados: Tamaño mínimo de hoja = 16. [2] Hiperparámetros optimizados: Método de ensemble = Bag, Número de submodelos = 10, Tamaño mínimo de hoja = 1 y Número de predictores a muestrear = 2.

A la vista de los buenos resultados obtenidos con los regresores basados en árboles, y para obtener el mejor resultado posible, se utilizaron dos implementaciones optimizables. Como se discutió en la sección 6.5, este tipo de implementación permite que ciertos parámetros de regresión se ajusten automáticamente para optimizar su operación. Específicamente, el árbol optimizable que consta de un árbol de regresión, y el conjunto optimizable que se basa en una combinación de diferentes árboles (ensemble), se utilizaron para formar un bosque de regresión (RF).

En el caso del árbol Optimizable, el principal parámetro a optimizar es el tamaño mínimo de hoja. El ajuste de este parámetro resulta en un valor muy similar al del árbol fino, confirmando los resultados de los árboles fino, mediano y grueso.

Para el Ensemble Optimizable, además del parámetro de tamaño mínimo de hoja, existen otros importantes como el tipo de ensemble utilizado, el número de subproblemas (*learners*) o el número de predictores a muestrear que, una vez optimizados, arrojan resultados incluso mejores que con un solo árbol de regresión.

Los hiperparámetros de cada modelo optimizable resultantes del proceso de entrenamiento y optimización se detallan al pie de las tablas 6.1, 6.2 y 6.3.

Los datos de las tablas anteriores permiten analizar de manera global los resultados de los diferentes regresores. Para poder interpretar estos resultados, en el gráfico de la figura 6.3 se muestra la discrepancia entre los valores estimados y reales de ϵ. En concreto, se muestra el valor estimado de ϵ frente al valor real para cuatro de los regresores de la tabla 6.3. Concretamente se han seleccionado el regresor lineal, árbol optimizable, ensemble optimizable y red neuronal ancha para ilustrar la variedad de comportamiento observados tanto en aquellos casos que el regresor presenta un buen comportamiento como en los peores casos.

La figura 6.3 muestra que para los regresores de tipo lineal el valor estimado presenta una gran desviación con respecto al valor real de ϵ para muchos de los planos evaluados, por lo que tiene un valor de RMSE y/o RAE mucho mayor que el resto, tal como se observa para todos los regresores de este tipo en la tabla 6.3. Esos casos se identifican como puntos que se separan de manera notoria de la diagonal principal. Sin embargo, otros algoritmos, como el ensemble optimizable o la red neuronal ancha, presentan un número reducido de puntos que se desvían de la respuesta ideal. Esta figura permite tener una visión cualitativa de lo bien se comportan los regresores, reafirmando los resultados presentados en la tabla 6.3.

Estos resultados se han obtenido considerando todo el conjunto de planos de calibración, sin hacer distinción entre las configuraciones de los mismos. En una segunda fase, se vuelve a analizar la información de forma desagregada para discernir si es posible definir algún criterio para seleccionar los planos de calibración que arrojen los mejores resultados. En otras palabras, se lleva a cabo un filtrado del conjunto de entrenamiento.

Para tratar de seleccionar el mejor conjunto de planos de calibración, entendiendo como tales aquellos que, una vez procesados para diferentes configuraciones de θ y ϵ, proporcionan un RMSE y RAE más bajo, se analizó de manera desagregada la información del RMSE para cada configuración de planos considerada. De este modo, en la figura 6.4 los valores de RMSE se representan en función de las rotaciones ρ_X, ρ_Y y ρ_Z que dieron origen al plano considerado. Dado que en un gráfico tridimensional no es posible representar estos cuatro parámetros (ρ_X, ρ_Y y ρ_Z, y RMSE), la información se muestra en diferentes gráficos, correspondientes a diferentes valores de ρ_Z.

La figura 6.4 muestra que, para la mayoría de las configuraciones del plano de calibración, el valor RMSE es inferior al que se indica en la tabla 6.3. El análisis del comportamiento también muestra que cuando $\rho_Z = 0°$ y $\rho_Z = 90°$, los resultados son peores, lo que significa que el regresor no puede estimar el valor de ϵ con la misma precisión que en los demás casos. Es por ello por lo que se toma la decisión de descartar estos planos para la calibración. Este filtrado permite reducir significativamente los valores RMSE y RAE. La tabla 6.4 muestra los valores de RMSE y RAE resultantes de aplicar el filtrado descrito a los regresores de la tabla 6.3. Por lo tanto, para calibrar el sistema de visión se selecciona un cierto número de planos y se estima el valor de ϵ como la media de los valores calculados para cada plano de calibración utilizado.

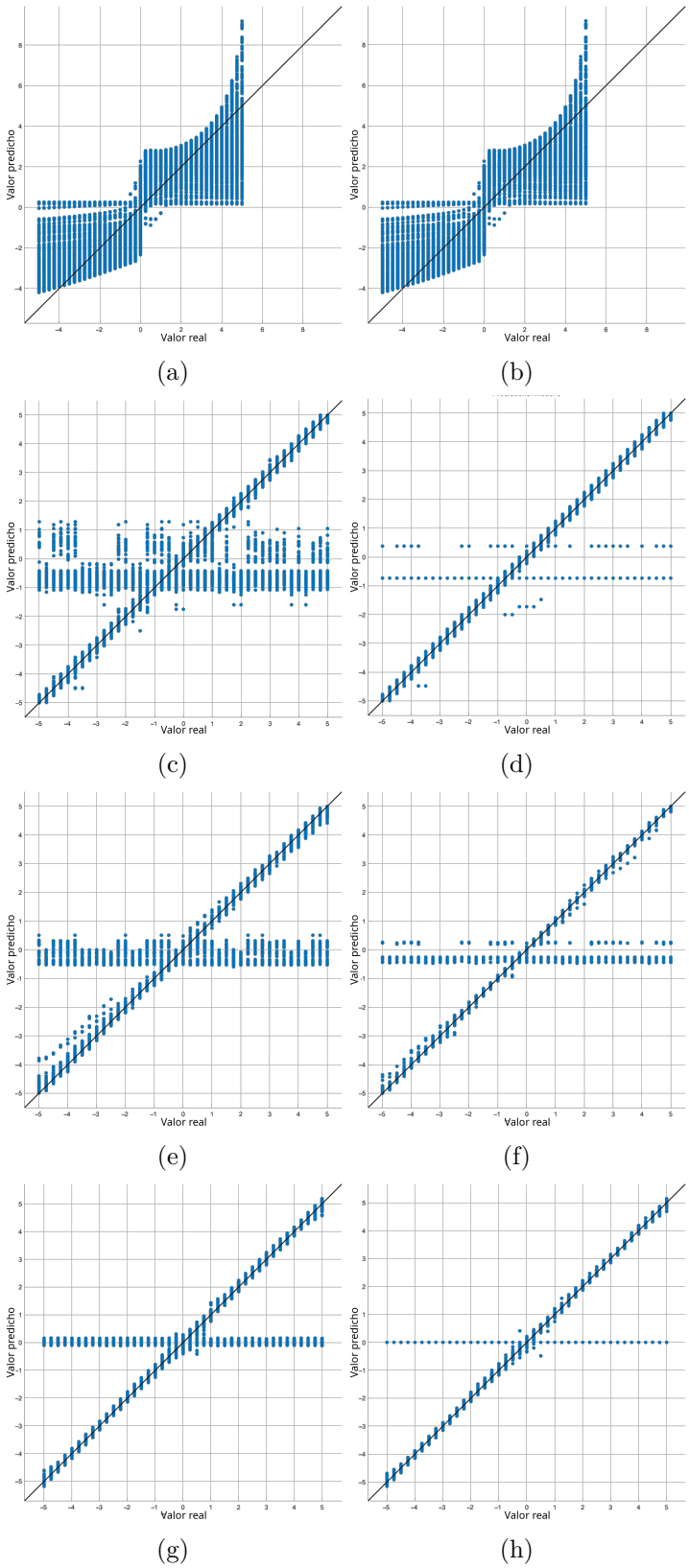
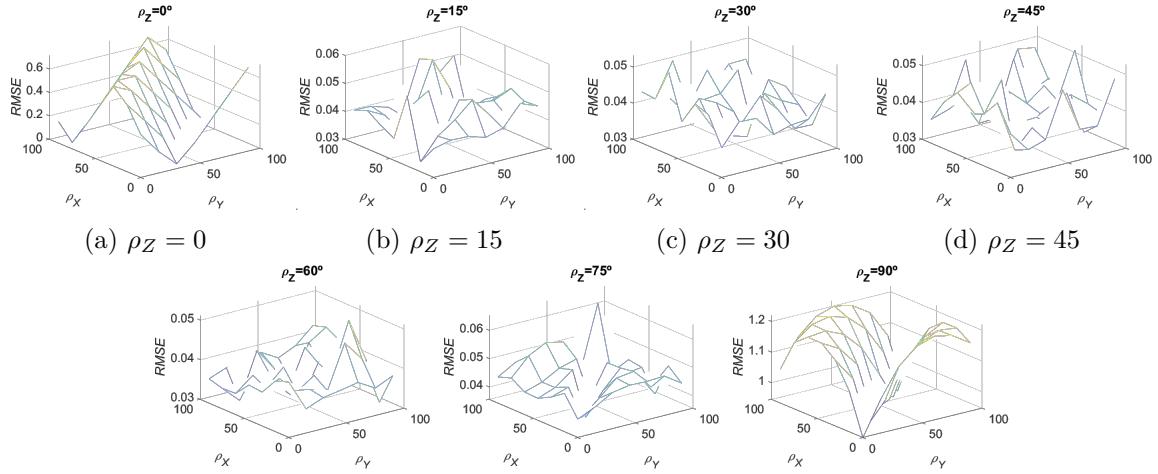


Figura 6.3: Comparación entre la respuesta predicha (eje x) y el valor real (eje y) para cuatro de los regresores de la tabla 6.3. (**a**)(**b**) Validación y test del regresor lineal, (**c**)(**d**) validación y test del Árbol Optimizable, (**e**)(**f**) validación y test del Ensemble Optimizable y (**g**)(**h**) validación y test de la Red Neuronal Ancha.



(a) $\rho_Z = 0$ (b) $\rho_Z = 15$ (c) $\rho_Z = 30$ (d) $\rho_Z = 45$

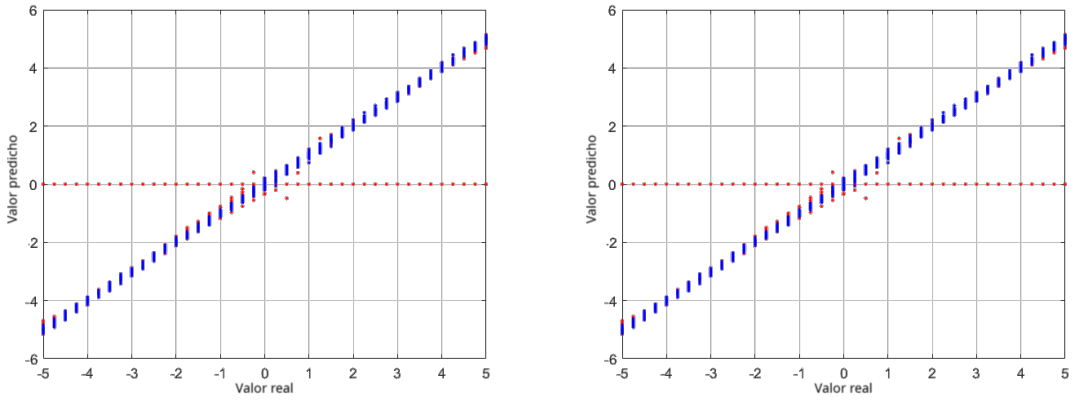
(e) $\rho_Z = 60$ (f) $\rho_Z = 75$ (g) $\rho_Z = 90$

Figura 6.4: RMSE calculado para cada configuración del plano de calibración del regresor de Red Neuronal Amplia. Dada la cantidad de variables a representar, fue necesario separar la representación en múltiples gráficas para los diferentes valores de ρ_Z.

Tabla 6.4: Resultados obtenidos con los regresores de la tabla 6.3 haciendo uso de los planos obtenidos del filtrado de los planos.

Regresor	**Variante**	**RMSE (°) Validación**	**RMSE (°) Test**	**RAE (%) Validación**	**RAE (%) Test**
LR	Linear	1,3479	1,3429	42,96	42,78
LR	Interactions	0,4172	0,4149	9,88	9,85
LR	Robust	1,3491	1,3448	42,61	42,46
LR	Stepwise	0,4172	0,4149	9,88	9,85
RT	Fine	0,0523	0,0521	0,88	0,88
RT	Medium	0,0523	0,0521	0,88	0,88
RT	Coarse	0,0579	0,0578	0,98	0,98
RT	Optimizable	0,0523	0,0521	0,88	0,88
RF	Optimizable ensemble	0,0367	0,0366	0,99	0,99
NN	Narrow	0,2101	0,2093	5,94	5,93
NN	Medium	0,0959	0,0958	2,75	2,75
NN	Wide	0,0432	0,0432	1,28	1,27
NN	Bilayered	0,1092	0,1091	3,20	3,20
NN	Trilayered	0,0942	0,0943	2,67	2,68

La figura 6.5 muestra el efecto de la etapa de filtrado, donde los puntos rojos representan los planos descartados. Gracias a esta técnica, se eliminan la mayoría de los valores espurios. Esto provoca una mejora significativa en los resultados, como se puede observar comparando los valores que se muestran en las tablas 6.3 y 6.4 .



(a) Validación. (b) Test.

Figura 6.5: Comparación entre la respuesta predicha (eje x) y el valor real (eje y) de la red neuronal amplia después de filtrar los planos de entrada. Los puntos azules representan los planos que pasan la etapa de filtrado mientras que los puntos rojos representan los descartados.

6.7.2. Discusión de los resultados

A lo largo de este capítulo, en las secciones anteriores, se han presentado varios resultados que permiten validar el sistema de calibración propuesto.

En primer lugar, se determinó que el uso de regresores, y en particular los derivados de árboles de decisión y ciertas configuraciones de redes neuronales, brindan los mejores resultados. En estos casos, los valores de RAE y RMSE obtenidos son lo bastante pequeños para afirmar que el valor de ϵ se estima con suficiente precisión para ser utilizado para corregir el ángulo de cabeceo. Aunque la información disponible no entra dentro de la clasificación de variables categóricas, donde se sabe que los árboles dan buenos resultados, el problema tiene otras características que lo hacen adecuado para ser abordado utilizando este tipo de algoritmos. Por un lado, la información no se caracteriza por tener un marcado componente de aleatoriedad, como podría ser el caso del preprocesado de las imágenes. Por otro, la dimensionalidad del problema y las características utilizadas para resolverlo hacen que otros regresores no puedan obtener resultados similares. Este es el caso del pobre ajuste obtenido con los regresores lineales. Similarmente, la ausencia de una componente temporal en el procesado, puesto que las imágenes se han considerado independientemente y no como parte de una secuencia, implica que no se puedan aprovechar las ventajas de otros algoritmos como es el caso de las redes convolucionales.

En segundo lugar, las posibles ventajas de incluir el ángulo de rotación entre planos normales en el conjunto de características de entrada a los regresores para estimar el valor de ϵ han sido analizados. Para ello se ha realizado una comparación de los resultados obtenidos con los diferentes regresores tras entrenarlos con y sin estas nuevas características (ϵ_{normal}

y el eje de rotación entre las normales de los planos). Así, se encuentra que los resultados obtenidos con los peores regresores (regresores lineales) mejoran significativamente con la inclusión de las características mencionadas. Por su parte, los resultados de los regresores con mejor desempeño apenas cambian.

Un tercer resultado surge del análisis del comportamiento de los regresores seleccionados y, en concreto, de los que mejores resultados arrojan. Se ha observado que los planos del conjunto de calibración cuyas configuraciones contienen un ángulo $\rho_Z = 0°$ o $\rho_Z = 90°$ producen peores resultados de calibración. Esto se explica analizando cómo se construyen los planos de los diferentes conjuntos y, en concreto, del conjunto de calibración. La primera rotación se realiza sobre el eje Z. Como consecuencia, una rotación de 0 o 90 grados producirá cambios en el plano que no son fácilmente distinguibles en la etapa de regresión. Por este motivo, los valores de RMSE obtenidos son peores en estos casos. Es por esto que la figura 6.5 muestra múltiples errores alrededor del valor 0 ya que el regresor no es capaz de identificar correctamente el valor de ϵ.

Capítulo 7

Aplicación a la detección de obstáculos en un vehículo

La detección de obstáculos es una problemática crítica en campos como la robótica, la conducción autónoma y la navegación de vehículos no tripulados. A lo largo de los años se han desarrollado diversas estrategias para abordar este desafío, utilizando tanto técnicas basadas en sistemas de visión como en otros sensores, así como combinaciones de ambos enfoques.

En el capítulo anterio49 se ha presentado una aplicación directa de la solución de modelización de entornos 3D basada en planos para la calibración del propio sistema estereoscópico. En este capítulo se muestra cómo la información de reconstrucción de escenas puede emplearse para facilitar la navegación de un vehículo. Puesto que este tipo de tarea requiere ser capaz de realizar el procesado en tiempo real, se trata de reducir la cantidad de información a tratar de manera que se logre cumplir con este requerimiento. Esto se consigue proyectando el espacio de disparidad (u_d, v, Δ), descrito en el capítulo 4, sobre el plano Disparidad-V para determinar la existencia de posibles obstáculos y la distancia a los mismos, y sobre el plano Disparidad-U para caracterizar el tamaño del obstáculo.

7.1. Trabajo relacionado

La detección de obstáculos mediante técnicas basadas en visión ha experimentado en los últimos años avances significativos, con una gama de herramientas que abordan desafíos específicos en entornos dinámicos. Los algoritmos basados en técnicas de visión por ordenador han sido ampliamente utilizados para la detección de obstáculos. Métodos tradicionales, como el filtro de Kalman y el filtro de partículas tratados en Thrun et al. (2005), se han aplicado con éxito para rastrear y prever la posición de objetos en movimiento, incluyendo obstáculos.

Las técnicas basadas en visión estéreo, consistente en el uso de dos cámaras para capturar una escena desde perspectivas ligeramente diferentes, han demostrado ser eficaces para la

detección de obstáculos. Algoritmos como Semi-Global Matching (SGM) (ver Hirschmuller (2008)) y Block Matching han sido empleados para estimar la disparidad entre las imágenes estéreo, permitiendo la reconstrucción 3D de la escena y la identificación de obstáculos en el entorno cercano.

A partir del 2010 se produce un resurgimiento del Aprendizaje Profundo con el uso de redes neuronales de algún tipo para la detección de objetos. Esta tendencia se ve impulsada por la aparición de nuevos dispositivos de cómputo cuya capacidad hizo viable la implementación de técnicas que habían sido descritas tiempo atrás. Un hito destacable en este proceso es el de Cireşan et al. (2010), que introdujo el cómputo con GPU usando redes neuronales profundas. En el año 2012 el ganador de ImageNet lo hizo utilizando redes neuronales convolucionales (CNN). Se trata de una competición anual de clasificación visual de imágenes a gran escala, donde los participantes desarrollan modelos de Aprendizaje Profundo para clasificar objetos en una amplia variedad de categorías. El Aprendizaje Profundo es interesante para la detección de objetos porque construyen representaciones de los datos de complejidad creciente de forma automática. En contraposición a las redes neuronales tradicionales, las redes profundas tienen múltiples capas ocultas densamente conectadas.

Dentro de los problemas abordables mediante técnicas de visión, el problema de detección es más complejo que el de la identificación ya que, además de determinar qué hay en la imagen, hay que indicar dónde se encuentra. Ejemplos de soluciones a este problema con redes convolucionales son el R-CNN descrito en Girshick (2015) y también en Ren et al. (2017) o YOLO: You Only Look Once, presentado en Redmon et al. (2016).

Las técnicas basadas en Aprendizaje Automático también pueden ser empleadas en navegación autónoma para la localización de obstáculos en una imagen. En Michels et al. (2005) los autores proponen un sistema de detección basado en el aprendizaje supervisado. Este aprendizaje se ve reforzado a través del uso de imágenes sintéticas de iguales características que las captadas por las cámaras del vehículo. La aplicación es capaz de percibir la profundidad y los obstáculos a partir de imágenes tomadas durante los recorridos automáticos del vehículo, pudiendo generar los comandos de volante adecuados para el guiado.

En Dima and Hebert (2005) los autores proponen la adaptación de algoritmos de aprendizaje activo estándar al dominio de la detección de obstáculos y validan sus resultados mediante una serie de pruebas, concluyendo que estas técnicas pueden jugar un papel importante en los sistemas robóticos.

En Hadsell et al. (2009) se presenta una estrategia de aprendizaje auto-supervisado para construir un sistema de detección de obstáculos a larga distancia y para la interpretación del terreno, pudiendo implementar una estrategia de planificación en base a la información recogida. Se entrena una red jerárquica profunda para extraer las características que se utilizarán para entrenar un clasificador en tiempo real. Este clasificador es capaz de ver obstáculos y caminos en un rango de 5 a 100 metros.

Una aproximación diferente para resolver el problema de detección de obstáculos en vehículos autónomos es la fusión de diferentes tipos de sensores. Tradicionalmente, también se ha hecho uso de sensores de RADAR (RAdio Detection And Ranging) para esta tarea, debido a su capacidad para operar en diversas condiciones meteorológicas y de iluminación. El trabajo de Heuer et al. (2012) explora la viabilidad de sistemas de radar para la detección de peatones en entornos urbanos.

La fusión de datos de múltiples sensores, como cámaras y LIDAR (Laser Imaging Detection and Ranging), ha mejorado la precisión en la detección de obstáculos. Algunas estrategias que combinan información visual con datos de distancia provenientes de sensores LIDAR permiten una representación más completa del entorno y una identificación más precisa de obstáculos. Algunos ejemplos de esta tendencia se presentan en Liu et al. (2023) o Xu et al. (2023).

Otra de las claves en los avances en el problema de la detección de obstáculos es el desarrollo de sistemas de detección de obstáculos en tiempo real. Esto es esencial para aplicaciones como vehículos autónomos y sistemas de asistencia al conductor. Métodos eficientes, como el enfoque propuesto por Wang et al. (2011), utilizan técnicas optimizadas para garantizar la detección en tiempo real.

Finalmente, y aunque no se trate de una herramienta específica para la detección de obstáculos, los sistemas SLAM (Simultaneous Location And Mapping - Mapeo y Localizacion Simultáneos), como Google Cartographer (ver Hess et al. (2016)), no solo contribuyen a la creación de mapas, sino que también desempeñan un papel en la detección de obstáculos al integrar información del entorno en tiempo real.

Si se circunscribe el análisis al uso de técnicas basadas en disparidad hay trabajos de autores como Bertozzi and Broggi (1998) o Labayrade and Aubert (2004) que en su aproximación se limitan a estimar la distancia hasta los obstáculos haciendo uso de la información del plano Disparidad V. De este modo son capaces de proporcionar esta información a un sistema de detección de obstáculos compuesto por múltiples elementos, complementando así la información del entorno.

Una aproximación más parecida a la propuesta en este capítulo es la de Hu et al. (2005), en la que utiliza tanto la información de la Disparidad V como la Disparidad U para la detección de obstáculos.

Más recientemente, en Dinh Nguyen et al. (2016) los autores proponen un sistema que detecta, reconoce y rastrea vehículos y peatones. Lo hacen combinando diferentes fuentes de patrones locales e información profunda utilizando técnicas de Aprendizaje Profundo. En particular, emplean un algoritmo adaptativo de Disparidad U-V, cuyos resultados se aplican a un novedoso sistema de reconocimiento de vehículos y peatones.

7.2. Modelado de la escena

Tal como se describió en el capítulo 5, es posible modelar el mundo real como una combinación de planos. En el caso particular de las escenas viales, se puede asumir que uno de los vectores que definen el eje X es paralelo al sistema de referencia global. Por lo tanto, la ecuación general que describe los planos que satisfacen esta condición (5.1) se puede reducir a:

$$sY + tZ + u = 0 \tag{7.1}$$

En la figura 7.1 se muestra un ejemplo idealizado de un mapa de disparidad que podría obtenerse de una escena en la que hay dos planos que cumplen la condición anterior. Este mapa de disparidad se obtiene a partir de la proyección ideal de la escena en el par de imágenes estéreo y el cálculo posterior de la disparidad entre puntos correspondientes en cualquiera de las imágenes, suponiendo que sea posible determinar la correspondencia exacta entre imágenes.

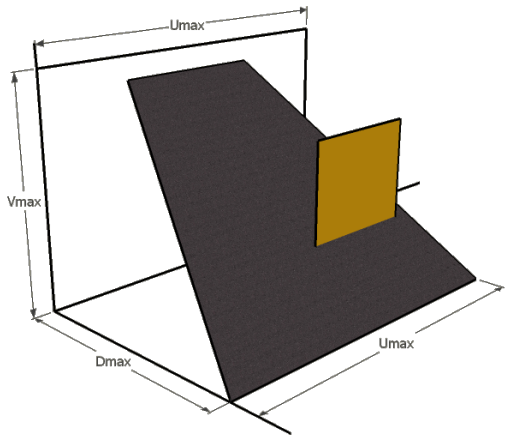


Figura 7.1: Mapa de disparidad ideal de una escena simple que presenta la carretera con un obstáculo. El plano gris oblicuo ilustra una representación simplificada de una carretera, mientras que el plano vertical ocre corresponde a la representación simplificada de un obstáculo.

Teniendo en cuenta la simplificación que introduce la ecuación 7.1 y definiendo dos nuevas constantes m y n para simplificar la notación, es posible reescribir la ecuación 5.2 de la manera siguiente:

$$m = \frac{-s}{t} \tag{7.2}$$

$$n = \frac{-u}{t} \tag{7.3}$$

$$\Delta = \frac{d}{ma - n - f}(v - v_0)(m\cos\theta + \operatorname{sen}\theta)\, v \tag{7.4}$$
$$+ \frac{d}{ma - n - f} + (\alpha(m\operatorname{sen}\theta - \cos\theta) - (v_0(m\cos\theta + \operatorname{sen}\theta)))$$

En esta última expresión se observa que la disparidad Δ tiene una dependencia lineal con v y, a su vez, es independiente de u_d. Por lo tanto se demuestra que todos los planos que responden a la expresión 7.1 se reflejan como una línea recta en el espacio (v, Δ). Si se analiza la ecuación es fácil comprobar que aquellos píxeles pertenecientes a objetos o elementos de la imagen que se encuentren más cercanos presentarán un valor de disparidad mayor que los que se encuentren más alejados. Entonces, cuanto más cerca esté el punto proyectado, menor será el valor de la coordenada Z y mayor será la disparidad calculada.

Además, en base al teorema 2 presentado en la sección 5.1 se sabe que todos los pixeles que son parte del mismo objeto presentarán un valor de disparidad idéntico, o muy parecido, independientemente de la fila o columna de la imagen en la que aparezcan.

En base a estas dos proposiciones es posible definir un método para detectar obstáculos ubicando regiones en las imágenes donde la disparidad es constante. En cambio, si se detecta un área donde la disparidad está cambiando se puede inferir que es algo que se encuentra a una distancia variable del sistema de visión, por lo que probablemente sea la carretera.

La reducción del coste computacional que proporciona el método respecto a otros basados en el cálculo de sistemas completos de mapas tridimensionales radica en la forma en la que se analiza la información de disparidad para evitar tener que reconstruir toda la información de profundidad. Se utilizan algunas estrategias con el fin de mostrar la información de disparidad de una manera fácilmente interpretable, permitiendo extraer información a través de los dos fundamentos extraídos de la ecuación 7.4.

Si se tienen en cuenta estas observaciones, es posible particularizar la ecuación 7.4 para el caso de un obstáculo. En primer lugar, para este caso se verifica que $Z = n$, por tanto 7.4 puede ser reescrita como:

$$\Delta = \frac{d}{-n - f} v \operatorname{sen}\theta + \frac{d}{-n - f}(-\alpha\cos\theta + v_0 \operatorname{sen}\theta) \tag{7.5}$$

Además, en un plano vertical se cumple que $\theta = 0$, por lo que la ecuación anterior se simplifica de la manera siguiente:

$$\Delta = \frac{d}{-n - f} v \operatorname{sen} 0 + \frac{d}{-n - f}(-\alpha\cos 0 + v_0 \operatorname{sen} 0)$$
$$\Delta = \frac{d\alpha}{n + f} \tag{7.6}$$

La ecuación 7.6 obtenida confirma que se cumple que todos los píxeles que forman parte de un mismo objeto presentarán un valor de disparidad idéntico o muy similar, independientemente de la fila o columna de la imagen en la que aparezcan.

Estos resultados se pueden observar de manera cualitativa si se proyecta el espacio (u, v, Δ) sobre el plano (v, Δ).

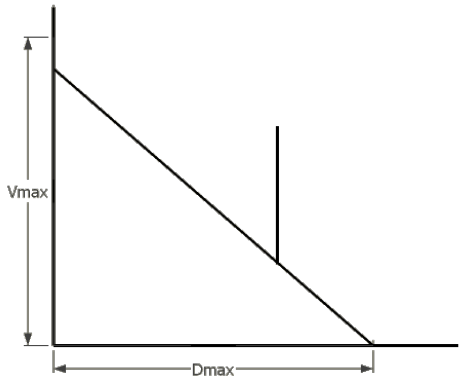


Figura 7.2: Imagen de disparidad correspondiente al mapa de disparidad mostrado en la figura 7.1.

En la figura 7.2 se presenta el resultado de la proyección del mapa de disparidad mostrado en la figura 7.1 sobre el plano (v, Δ) para construir la Imagen de Disparidad correspondiente. Esta figura muestra de manera intuitiva los resultados de las expresiones 7.4 y 7.6. Se observa como el plano oblícuo se proyecta como una recta diagonal en la que el valor de Δ sigue la ecuación 7.4, mientras que para el caso del plano vertical, se verifican las condiciones descritas anteriormente y el valor de Δ se convierte en una constante según la ecuación 7.6. El mismo efecto se observa si se proyecta sobre el plano (u, Δ).

7.3. Implementación del sistema de detección

El proceso de caracterización de los obstáculos en un par de imágenes estéreo se ha abordado en dos fases. El primer paso consiste en extraer información sobre los obstáculos existentes a partir de una proyección Disparidad V del espacio (u_d, v, Δ). En el segundo paso, se utilizará una proyección de Disparidad U combinada con la información recopilada en la primera fase para obtener una mejor caracterización de los obstáculos detectados.

7.3.1. Equipamiento hardware

El objetivo de la implementación que se ha llevado a cabo persigue tanto la validación del método propuesto para la detección de obstáculos como la demostración de que es posible aplicarlo sin necesidad de utilizar equipamiento especializado. Por ello se han empleado los

siguientes dispositivos para construir un par estéreo siguiendo el esquema mostrado en la figura 3.4:

- Cámaras: Dado que se dispone de dispositivos con una buena óptica, se han utilizado dos cámaras analógicas, permitiendo una mejor calidad de imagen.
- Capturadoras de vídeo: Se han empleado un par de capturadoras de vídeo analógico AverMedia, logrando así la captura de imágenes digitales.

Actualmente sería posible reemplazar ambos dispositivos por una pareja de cámaras digitales con un coste razonable.

7.3.2. Fase Disparidad-V

El modo en el que se ha implementado el procesamiento de las imágenes del par estéreo para obtener información en el plano Disparidad V se puede resumir en el diagrama de la figura 7.3.

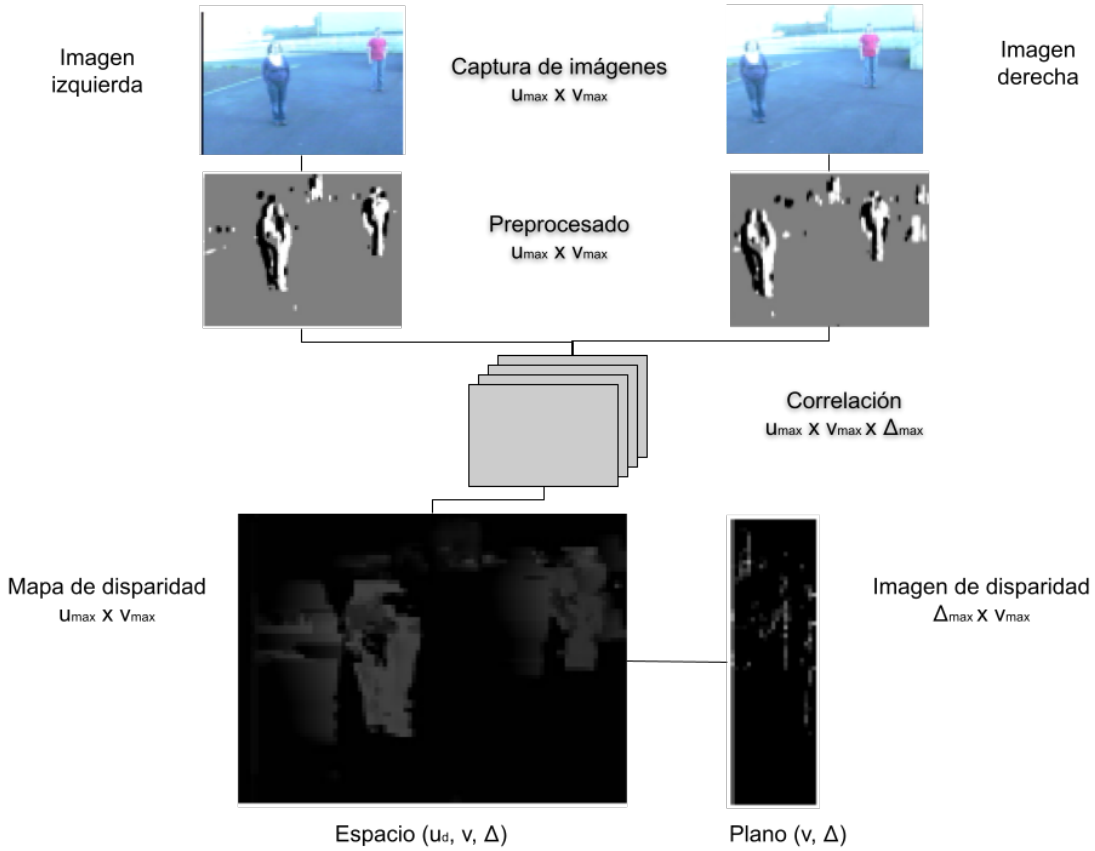


Figura 7.3: Evolución del procesado de las imágenes a lo largo del proceso de cálculo de la Disparidad V.

En los siguientes subapartados se describirá el procesamiento que se realiza en cada una de las etapas mostradas en el diagrama de la figura 7.3, haciendo hincapié en aquellos aspectos que se han tenido en cuenta para la optimización del rendimiento del algoritmo.

Captura de imágenes

El par de imágenes a procesar se captura de forma secuencial, debido a que las cámaras de vídeo analógicas estándar no cuentan con un sistema de sincronización. Para minimizar el tiempo de captura e intentar reducir al mínimo esa desincronización, se hace uso de rutinas de bajo nivel que proporciona el sistema operativo para el manejo de este hardware. Estas imágenes se capturan en color con una resolución de 320x240 píxeles.

Preprocesado

A pesar de que las imágenes se capturan en color (RGB), esta información es irrelevante para el método a aplicar. Para determinar la correlación entre los píxeles de ambas imágenes es suficiente con la información de intensidad en escala de grises, reduciendo así la cantidad de información a procesar. Después de eliminar la información de color, el contraste en ambas imágenes se iguala para que la distribución de los niveles de grises sea lo más similar posible.

A continuación se realiza un proceso de ternarización como el descrito en Broggi et al. (2006). El objetivo que se persigue es simplificar las imágenes para reducir la información al mínimo necesario para localizar planos verticales de acuerdo con la estrategia propuesta. Este procedimiento consta de varias fases:

1. Se realiza un filtrado para eliminar bordes superfluos y débiles que puedan interferir con el procesamiento posterior. Se utiliza un filtro de la media implementado mediante un kernel cuadrado centrado en cada píxel y pesos iguales para todos los elementos del kernel.

2. La imagen resultante se somete a una detección de bordes verticales tipo Sobel. El objetivo es reducir la información de la imagen conservando aquellos elementos que teóricamente deberían simplificar la detección de planos verticales en el mundo real utilizando un umbral automático.

3. El umbral utilizado en la detección de bordes se combina con una clasificación, de modo que los bordes detectados formados por una transición de claro a oscuro se etiquetan de manera diferente a las transiciones de oscuro a claro. Por tanto se obtiene una imagen con tres valores posibles.

4. Para reforzar los bordes detectados se realiza una dilatación de la imagen ternaria.

Construcción del mapa de disparidad

El mapa de disparidad es el nombre que adopta la representación del espacio (u_d, v, Δ) en el proceso de implementación. Dadas las dimensiones de dicho espacio, puede almacenarse como una imagen de dimensiones (u_{max} x v_{max}) en la que cada elemento contenga la diferencia entre las coordenadas v del píxel en una de las imágenes del par estéreo y su contraparte en el otro. Gracias al diseño final elegido para el sistema de visión, es posible hacer uso de la restricción epipolar para reducir el espacio de búsqueda, por lo que se asume que para cada píxel de la imagen izquierda $I_i(u_i, v)$ la correspondencia en la imagen derecha solo se puede encontrar en la fila v de dicha imagen(I_d).

El método elegido para determinar la correspondencia es la maximización de la correlación de niveles de grises o, lo que es lo mismo, minimizar la diferencia entre estos. De este modo, es necesario calcular las diferencias entre el nivel de cada píxel con todos los elementos de la fila correspondiente en la otra imagen que se encuentran dentro del rango máximo de disparidad contemplado (D_{max}). A continuación se ha de seleccionar entre estas diferencias aquella que representa el valor más pequeño. Para mejorar el resultado, en lugar de minimizar la correlación píxel a píxel, se utiliza el método de Suma de Diferencias Absolutas (SAD, Sum of Absolute Diferencies) en una vecindad del píxel. Habiendo asumido la restricción epipolar, esta vecindad se limitará a una ventana de l píxeles de ancho por un píxel de alto, centrada en el píxel que se está analizando. De este modo, el valor de cada píxel de la imagen que representa el espacio de disparidad I_{dm} será:

$$\begin{aligned} &I_{dm}(u,v) = \Delta | \Delta \in [1, \Delta_{max}], \\ &\Delta = argmin_{\Delta} \left(\sum_{i=-l}^{l} |I_d(u,v) - I_i(u+\Delta+i, v)| \right) \end{aligned} \tag{7.7}$$

Para minimizar el coste computacional de estas operaciones, la implementación se ha realizado siguiendo pautas para reducir los cálculos. Para cada valor de disparidad Δ_i a evaluar, una imagen se construye desplazando la imagen original a la izquierda tantos píxeles como se especifica el índice i de la iteración, por lo que se puede calcular la diferencia en el valor absoluto de todos los píxeles a la vez. Además, al convolucionar la imagen resultante con un tamaño de núcleo $2l+1$ cuyos elementos son todos iguales a uno, es posible implementar la suma de la expresión 7.7 para cada uno de los píxeles $I_d(u, v)$ de manera óptima. El tamaño del núcleo de convolución viene determinado por el valor de l, siendo este parámetro ajustable. En la implementación realizada los mejores resultados se han obtenido para $l = 2$

Puesto que en cada iteración se obtienen diferentes valores de correlación para un mismo píxel, habrá que obtener el mínimo, que será el que se corresponda con el valor de disparidad del píxel. Para evitar realizar este procesamiento a posteriori, resulta más

económico computacionalmente hablando mantener estos resultados parciales en una imagen auxiliar en la que se actualiza el valor mínimo en cada iteración, así como una segunda imagen que almacena el valor de disparidad correspondiente. Una vez realizados los cálculos para todos los valores de Δ se obtiene como resultado una imagen que contiene el mínimo de correlación y el mapa de disparidad.

Proyección sobre el plano Disparidad-V

En el capítulo 5 se demostró cómo la información obtenida al proyectar el espacio definido en (u_d, v, Δ) sobre el plano (v, Δ) permite determinar la existencia de planos paralelos al plano imagen, es decir, obstáculos, mediante la búsqueda de líneas verticales en dicha proyección. De cara a la implementación, el plano Disparidad V se representa en la Imagen de Disparidad en la que cada píxel tomará su valor acorde a la siguiente expresión:

$$I_{vdisp}(\Delta, v) = \begin{cases} 1, \exists u \in [1, u_{umax}] \,|\Delta = I_{dm}(u, v) \\ 0, in\ other\ case \end{cases} \tag{7.8}$$

A partir de esta interpretación, una manera simple y fácilmente implementable de detectar los valores de Δ que aparecen en cada una de las filas del mapa de disparidad es calcular el histograma de cada una de ellas. De esta manera se pasa del dominio (u_d, v, Δ) a (v, Δ). Con este método, cada elemento de la imagen de disparidad 7.8 contiene el número de apariciones de un valor de disparidad determinado en una fila. Posteriormente se umbraliza este resultado para eliminar aquellos niveles de disparidad con menos apariciones en la fila, obteniendo un binario I_{vdisp} que satisface las condiciones de la expresión 7.8.

Detección de obstáculos

La información proporcionada por la imagen de disparidad permite establecer una relación entre los posibles obstáculos existentes y las líneas verticales que aparecen en la imagen de disparidad. Para detectar las líneas verticales se ha utilizado la transformada de Hough en la imagen de disparidad. En la parametrización de esta transformada se ha permitido un margen que permita encontrar líneas con una pendiente muy grande, aunque no infinita, como válida para interpretarlas como un obstáculo. Cada línea vertical (L_i) se caracterizará por su coordenada horizontal $(\delta_i\ )$, que es el valor de disparidad representada, por la coordenada vertical donde comienza $(v_i\ )$ y por su longitud (l_i), que es proporcional a la altura del obstáculo.

7.3.3. Fase Disparidad-U

Una vez se han identificado los posibles obstáculos dentro de la imagen haciendo uso de la proyección sobre el plano (v, Δ), en esta etapa se pretende relacionar cada obstáculo

detectado en el último paso de la fase anterior con el correspondiente en la proyección de Disparidad U. Tomando como entrada cada uno de los obstáculos previamente localizados, los pasos que han de repetirse son los siguientes:

Definición de regiones de interés

Para reducir el coste computacional, se define una Región de Interés (Region of Interest - RoI) sobre el Mapa de Disparidad para cada obstáculo. Dado que la coordenada horizontal está directamente relacionada con la coordenada u en la imagen, se encuentra que para cada posición donde la coordenada horizontal u_d es menor que el δ_i de la línea que se está procesando, es imposible que I_{dm} pueda tener un valor de disparidad igual a δ_i. El origen de la RoI y su altura se calcularán directamente a partir de v_i y l_i.

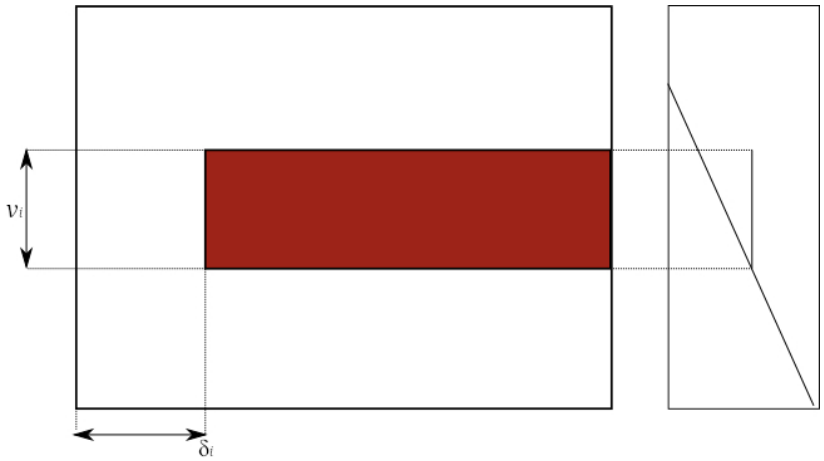


Figura 7.4: Definición de la Región de Interés para la construcción del Mapa de Disparidad Reducido.

Una vez definida la RoI es posible obtener un Mapa de Disparidad Reducido (Reduced Disparity Map - RDM) que contenga la información relativa al obstáculo que se está analizando sin necesidad de procesar toda la imagen. De este modo, el píxel de la imagen adoptará valor 1 en caso de que la intensidad de la misma posición coincida con el valor de la disparidad del obstáculo que se está analizando y cero en otro caso.

$$RDM_i\left(r,s\right) = \begin{cases} 1, I_{dm}\left(\delta_i + r, v_i + s\right) = \delta_i \\ 0, en\ otro\ caso \end{cases} \tag{7.9}$$

Con esta estrategia se consigue filtrar la información del mapa de disparidad de manera que solo se mantenga aquella relativa al obstáculo que se está procesando en la presente iteración.

Proyección sobre el plano Disparidad-U

Una vez obtenido el RDM del obstáculo que se está procesando (RMDi), y siguiendo los mismos principios explicados en la sección 7.3.2, se puede calcular una imagen que represente la proyección de Disparidad U del Mapa de Disparidad Reducida.

Estimación del ancho del obstáculo

El último paso consiste en estimar el ancho de los obstáculos detectando líneas horizontales en la imagen de proyección Disparidad U para lo cual, al igual que en la sección 7.3.2, se hace uso de la transformada de Hough.

7.3.4. Medida de distancia

En esta aplicación, teniendo en cuenta que se basa en el sistema de visión de dos grados de libertad descrito en la sección 3.3, la distancia puede calcularse de acuerdo a la expresión $Z = \frac{df}{\Delta} - f$ (3.12). Sin embargo, el término $-f$ se ha despreciado en esta expresión, ya que la longitud de la focal no resulta significativa en comparación con el resto de la distancia.

7.4. Resultados

La técnica descrita ha sido implementada y probada en un entorno real. La figura 7.5 muestra una imagen donde se han detectado tres obstáculos a tres distancias diferentes. Se trata de dos peatones en movimiento y una farola (al fondo de la imagen). Como se puede observar en la figura, además de detectarse se caracterizan con rectángulos cuyo tamaño se ajusta al de cada uno de los obstáculos.

Figura 7.5: Ejemplo del resultado de la detección y medición de distancia hasta los obstáculos en un fotograma real.

Tabla 7.1: Mediciones obtenidas para el fotograma de ejemplo de la figura 7.5

Obstáculo	**Z Real**	**Z Calculado**	**Error Relativo Z**	**X Real**	**X Calculado**	**Error Relativo X**
1	3,5 m	3,46 m	1,23 %	-0,63m	-0,6m	1,90 %
2	4,9 m	4,83 m	1,52 %	0,98m	1m	1,18 %
3	15 m	14,48 m	3,49 %	0,6m	0,64m	1,31 %

Para evaluar la calidad del sistema de detección de obstáculos se tienen en cuenta tres magnitudes. En primer lugar se comprueba que se detecta correctamente un determinado obstáculo. Una vez verificado el primer punto, la distancia y la posición del obstáculo se comparan con su posición en el mundo real.

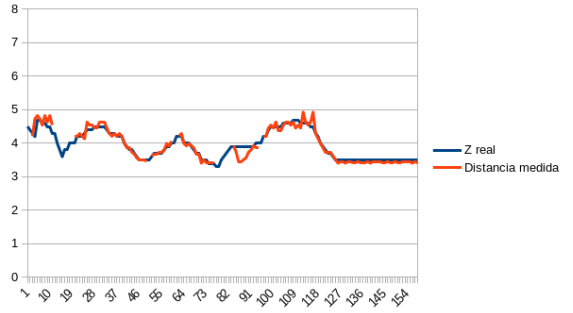
La tabla 7.1 muestra las características medidas para la imagen de la figura 7.5. Los tres obstáculos se detectan correctamente, siendo el obstáculo 1 el peatón más cercano, el obstáculo 2 el peatón más lejano y el obstáculo 3 la farola del fondo de la imagen.

Para validar el método, estas medidas se han comprobado en condiciones de tiempo real con obstáculos móviles y estáticos en escenas de larga duración. La figura 7.6 muestra el comportamiento del sistema en el transcurso de la secuencia de la que forma parte la imagen mostrada mostrada en la figura 7.5.

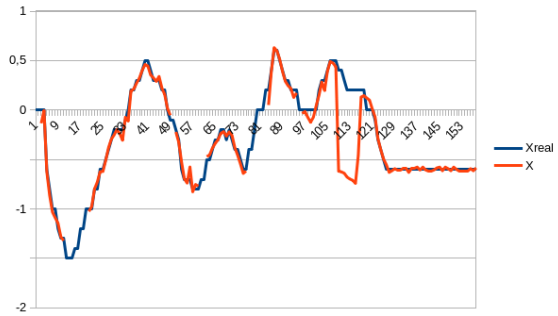
En el eje horizontal de las gráficas de la figura 7.6 se representa el instante de tiempo en el que se tomó el par de imágenes procesadas, mientras que en el eje vertical la representación varía para cada gráfica. En las figuras 7.6b, 7.6d y 7.6f se representa la detección y medición de la distancia a cualquiera de los objetos considerados, mientras que en 7.6a, 7.6c y 7.6e se representa la posición horizontal del obstáculo correspondiente.

La línea azul indica la posición real del objeto en cualquiera de las coordenadas y la línea roja la posición calculada. La ausencia de una línea roja significa que el objeto no ha sido detectado en ese fotograma. Sin embargo, es importante tener en cuenta que este comportamiento no implica un error de detección en todas las situaciones. En muchos fotogramas, los objetos en movimiento pueden salir del foco de una o incluso de las dos cámaras. También uno de los obstáculos puede ocultar al resto en algunos casos. Cuando esto ocurre el resultado esperado es la no detección del objeto. Por esas razones, es importante hacer un análisis más exhaustivo de los resultados. Analizando la secuencia considerada imagen a imagen se observa que el primer obstáculo es correctamente detectado en el 96,84 % de las imágenes, el segundo el 91,77 % y el tercero y más alejado el 79,11 % de las ocasiones. Esta magnitud es la que se ha denominado Tasa de Detección (Detection Rate - DR) en la tabla 7.1.

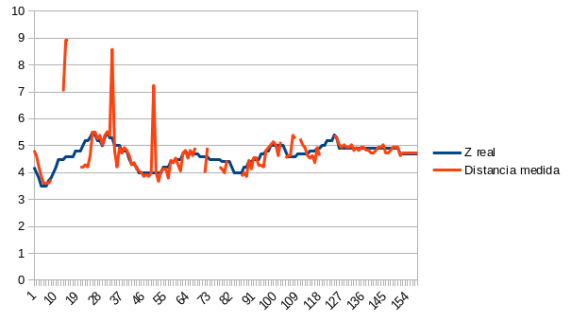
Tal como se ha indicado, la columna Tasa de Detección (Detection Rate - D.R.) de la tabla 7.2 indica el porcentaje de acierto en la detección del obstáculo teniendo en cuenta su aparición o no en cada uno de los fotogramas, mientras que las dos últimas columnas muestran el promedio del error relativo en la distancia y la posición horizontal, respectivamente, para todos los fotogramas de la secuencia del ejemplo.



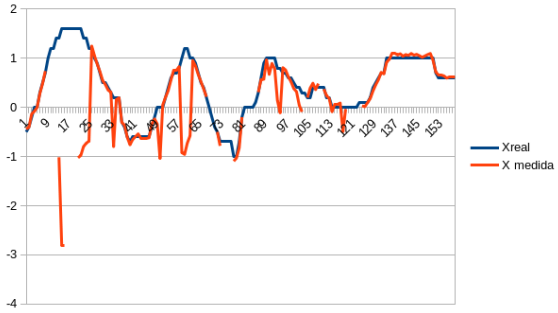
(a) Z Obstáculo 1



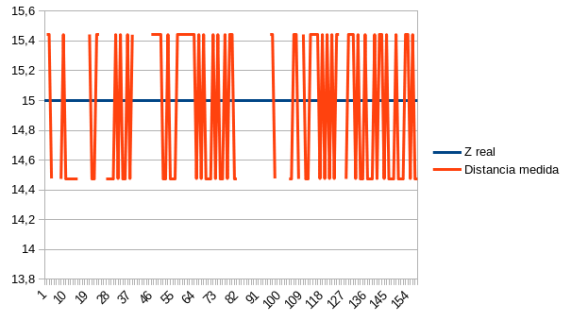
(b) X Obstáculo 1



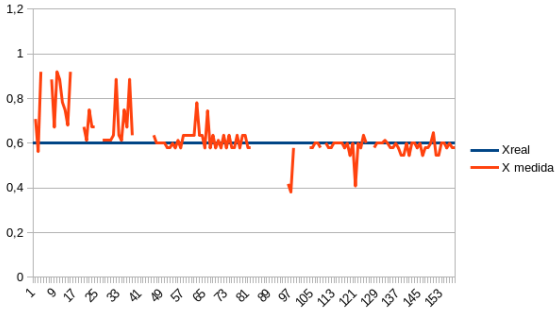
(c) Z Obstáculo 2



(d) X Obstáculo 2



(e) Z Obstáculo 3



(f) X Obstáculo 3

Figura 7.6: Comparativa entre las magnitudes reales y calculadas para una secuencia de fotogramas.

Tabla 7.2: Datos cuantitativos de la Tasa de Detección y Error Relativo en la detección de los obstáculos considerados.

Obstáculo	**Tasa de Detección**	**Error Relativo Z**	**Error Relativo X**
1	96,84 %	2,77 %	6,95 %
2	91,77 %	6,98 %	16,27 %
3	79,11 %	3,22 %	3,41 %

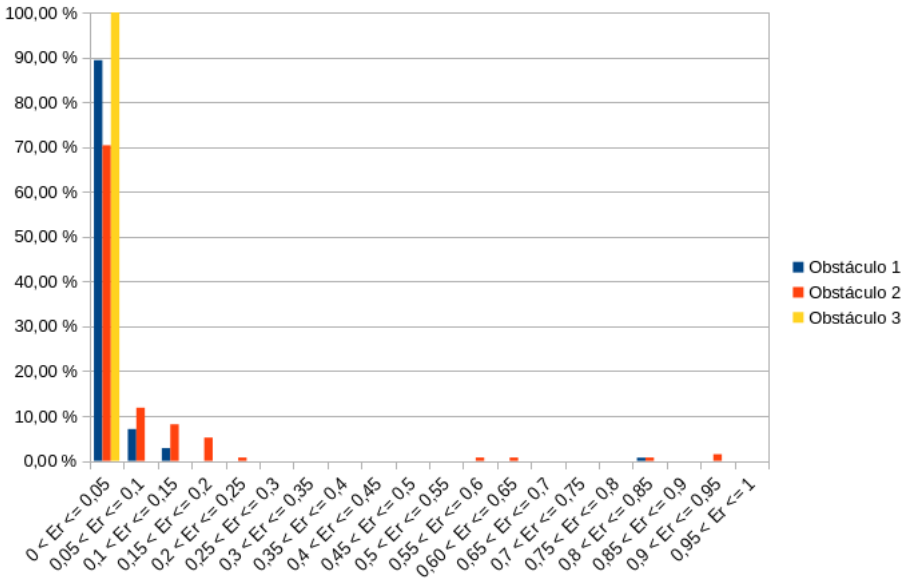


Figura 7.7: Distribución del error relativo en la medida de la distancia para los obstáculos considerados en la secuencia de imágenes correspondiente a la tabla 7.2.

En la figura 7.7 se representa la frecuencia de aparición de los posibles valores de error relativos, agrupados en intervalos de 5 centésimas. Cada una de las barras representa los datos de uno de los obstáculos considerados en la secuencia de imágenes del ejemplo. Esta información complementa la información proporcionada por la tabla 7.2, proporcionando una visión clara del orden de magnitud de los valores de error relativo más frecuentes.

En esta figura se observa que para los tres obstáculos estudiados el error relativo se encuentra por debajo del 5 % en al menos el 70 % de las imágenes analizadas. Además se verifica que son muy pocos los casos que alcanzan o superan el 25 % de error relativo. Al revisar estos casos concretos se trata de falsos positivos en la detección, es decir, se ha detectado algo en la imagen que se ha interpretado como un objeto o posible obstáculo, cuando en realidad no lo es. Este es el caso concreto del obstáculo 2. Debido a que los obstáculos 1 y 2 se mueven lateralmente en la secuencia de imágenes, se producen múltiples oclusiones o salidas de toma que se traducen en este tipo de errores de detección. Este es el motivo por el cual en la figura 7.7 se observan valores espúreos en los rangos de error relativo entre 0,55-0,6; 0,6-0,65; 0,8-0,85 y 0,9-0,95. Puesto que no se está llevando a cabo ninguna estrategia de seguimiento de los elementos que aparecen en las imágenes a lo largo de la secuencia que se está procesando no es posible determinar que se trata de un error de detección.

En la figura 7.8 se muestran dos claros ejemplos en los que el obstáculo 2 queda oculto tras el obstáculo 1 o se sale del plano de una de las imágenes del par estéreo, por lo cual resulta imposible realizar la detección de manera correcta.

(a) Oclusión (b) Fuera de cámara

Figura 7.8: Ejemplos de fotogramas en los que, por la propia disposición de los elementos, se producen errores de detección.

A diferencia de los obstáculos 1 y 2, el obstáculo 3, que es estático y en la secuencia de imágenes capturadas no sufre oclusiones, es el que menor error relativo presenta, tal como se observa en la figura 7.7.

En la tabla 7.2, la distancia medida se compara con la distancia real a los objetos y se muestra el error relativo. En el ejemplo, el error relativo crece con la distancia al objeto. Este comportamiento se debe principalmente a la relación inversa entre distancia y disparidad en la ecuación 3.12. Como consecuencia de esto, el efecto de un error en la disparidad calculada para un objeto será mayor si el objeto está más lejos, ya que la disparidad correspondiente será inferior, y viceversa.

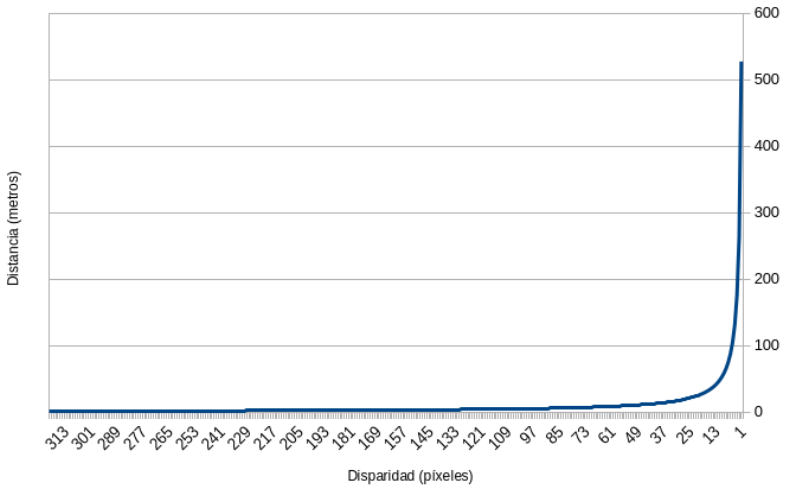


Figura 7.9: Representación de la relación no lineal entre la distancia y la disparidad.

La figura 7.9 muestra este efecto no lineal, representando la distancia calculada para un determinado valor de disparidad.

Se observa que el efecto de esta no linealidad alcanza valores críticos cuando la disparidad es muy baja, es decir, cuando se trata de objetos muy alejados. Por este motivo se puede concluir que el error que se produciría en caso de inexactitudes a la hora de determinar el valor exacto de la disparidad y, por tanto, la distancia a la que se encuentra el posible obstáculo es aceptable para una implementación en tiempo real.

Tabla 7.3: Valores representativos de la relación no lineal entre disparidad y distancia

Disparidad (px)	**Distancia Z (m)**	**Disparidad (px)**	**Distancia Z (m)**
0	∞	80	6,59
1	527,25	90	5,86
5	105,45	100	5,27
10	52,72	110	4,79
15	35,15	120	4,39
20	26,36	130	4,06
25	21,09	140	3,77
30	17,57	150	3,51
35	15,06	160	3,29
40	13,18	170	3,10
54	11,72	180	2,93
50	10,54	190	2,77
55	9,59	200	2,63
60	8,79	240	2,20
65	8,11	280	1,88
70	7,53	320	1,65

En la tabla 7.3 se observa de manera cuantitativa el efecto mostrado en la figura 7.9. Se representa el valor de distancia correspondiente a un conjunto de valores de disparidad, escogidos de manera que se observe el fenómeno de no linealidad descrito. Claramente se observa que para valores altos de la disparidad, que corresponden a objetos cercanos, la variación en la distancia medida es muy pequeña. Por el contrario, para objetos lejanos (disparidades pequeñas) un error en el valor de disparidad determinado representa una desviación mucho mayor en la estimación de la distancia.

En la aplicación propuesta este fenómeno es asumible, ya que prima la detección de obstáculos cercanos para facilitar la navegación del vehículo, aportando información valiosa también de los obstáculos lejanos de cara a una planificación del recorrido. Como efecto colateral, es importante remarcar que dada la poca variación en la distancia medida para valores de disparidad altos, obtener el valor exacto no resulta crítico a la hora de caracterizar el obstáculo.

Finalmente, determinar los posibles valores de distancia a la que se encuentran los obstáculos permite definir márgenes de seguridad de acuerdo a la distancia calculada, de manera que se garantice que la distancia a la que se detecta el obstáculo sea la real o, si no, inferior a ella.

A la luz de los resultados que se presentan, se puede concluir lo siguiente:

- La metodología es capaz de detectar y caracterizar objetos estáticos y en movimiento.
- La precisión en la detección y el error relativo en las dimensiones X y Z depende de la distancia a la que se encuentra el objeto.
- La relación no lineal entre la distancia calculada y la disparidad es bien conocida. Esto permite plantear estrategias de acotación del error introduciendo márgenes de seguridad no uniformes adaptados a la disparidad calculada, de modo que la distancia detectada sea menor o igual que la real. Como consecuencia de ello, el sistema de detección que se propone dirá que el obstáculo está más cerca de lo que realmente está, evitando así accidentes por colisión.

Capítulo 8

Conclusiones

En esta tesis doctoral se han abordado una serie de problemas relacionados con la reconstrucción de escenas 3D basada en estereovisión. En primer lugar, se ha propuesto una metodología para la reconstrucción de la información existente en la escena representando cada elemento de la misma como un plano. Este método se ha realizado haciendo uso de técnicas basadas en Disparidad U-V.

A partir de este punto se ha analizado el comportamiento del método propuesto ante la posible aparición de errores en el ángulo de cabeceo del sistema de visión estéreo. El estudio realizado ha permitido caracterizar el comportamiento esperado para el sistema en presencia de errores, así como extraer una serie de conclusiones sobre el mismo que resultan interesantes de cara a la aplicación práctica de la metodología.

Finalmente se han desarrollado dos aplicaciones prácticas de la técnica de reconstrucción, más allá de la reinterpretación de una escena. Por un lado, se propone una técnica de calibración de un par de cámaras estéreo haciendo uso de técnicas de Aprendizaje Automático. Por otra parte, se presenta una simplificación del método de reconstrucción que permite su utilización para la detección de obstáculos en tiempo real.

En el presente capítulo se recogen las conclusiones extraídas en cada uno de los tópicos citados.

Tal como se ha expuesto, el trabajo presentado se desarrolla sobre la base de una novedosa metodología de reconstrucción de escenas 3D. Dicha metodología se ha descrito completamente demostrando que es posible llevar a cabo la reconstrucción de la escena tridimensional a partir de las imágenes de un par estéreo utilizando técnicas basadas en Disparidad U-V. La técnica de reconstrucción completa de escenas presentada es capaz de proporcionar información detallada sobre todos los elementos presentes en una escena del mundo real, dando como resultado una representación simplificada de los elementos que la componen como una combinación de varios planos.

Para estudiar cómo afectan los posibles errores en el ángulo de cabeceo de las cámaras a los planos con los que se reconstruye una escena, se ha llevado a cabo un análisis de robustez.

Dicho análisis ha consistido en caracterizar el comportamiento de la metodología propuesta frente a la aparición de errores en el ángulo de cabeceo del sistema estereoscópico. Como resultado de este estudio se ha conseguido modelar el comportamiento esperado del sistema de reconstrucción para diferentes configuraciones de planos que se puedan presentar en una escena del mundo real. De este modo se ha medido la sensibilidad del sistema propuesto frente a errores en el ángulo de cabeceo, acotando su límite superior. El resultado obtenido permite garantizar que el plano resultante de la reconstrucción nunca se desviará del ideal, entendiendo como tal el plano al que se tiene en ausencia de error en el ángulo de cabeceo, en un ángulo mayor que dicho error, por lo que se puede concluir que el método propuesto no introduce errores adicionales.

Además del resultado anterior, se ha definido un parámetro que permite anticipar el comportamiento de la reconstrucción en función del error del ángulo de cabeceo. Este parámetro se ha modelado como una variable aleatoria que sigue una distribución beta. Este resultado complementa la conclusión anterior ya que, además de haber determinado que el error de reconstrucción tiene un límite superior, la existencia de un modelo permite anticipar cuál será la desviación esperada.

Por último, el análisis global de los datos obtenidos permite afirmar que el comportamiento dominante, entendiendo como tal el que es más probable que ocurra, es que el ángulo de error de reconstrucción del plano tiene un valor muy próximo al del ángulo de error de cabeceo.

Los resultados de este análisis y la caracterización realizada del comportamiento del sistema han permitido plantear el desarrollo de dos técnicas derivadas de la metodología propuesta. De este modo, se han realizado dos ejercicios de reinterpretación y simplificación de la información proporcionada para el desarrollo de sendas aplicaciones prácticas.

En primer lugar se ha estudiado la posibilidad de utilizar dicha técnica para calibrar un par estéreo frente a errores en el ángulo de cabeceo. Esta propuesta proporciona una herramienta que se puede utilizar para estimar el error en la reconstrucción de un determinado elemento de la escena (un plano) causado por una determinada desviación en el ángulo de cabeceo.

Para calibrar el sistema de visión, se propone utilizar el conocimiento generado para entrenar un regresor para, a continuación, estimar el valor del error utilizando la información del conjunto de planos reconstruido. Las reconstrucciones sucesivas pueden utilizar este valor para compensar el error.

Aunque la técnica propuesta se basa en técnicas de Disparidad U-V, el sistema de calibración es completamente independiente de las técnicas de visualización que se quieran utilizar con el sistema calibrado. Esta es una característica importante ya que el objetivo perseguido es proporcionar una herramienta genérica para corregir un error intrínseco en el sistema de visión estéreo.

Para estimar el valor del error se han entrenado diferentes regresores y se ha comparado su comportamiento ante diferentes situaciones que se pueden producir en una escena del mundo

real. Para ello se han entrenado los regresores utilizando un conjunto de entrenamiento que contiene del orden de 300.000 elementos. Esto permite tener una representación adecuada de las diferentes orientaciones que pueden exhibir los elementos del mundo real. Como resultado de la comparativa realizada se ha determinado que los que mejor se adaptan al problema son determinadas configuraciones de redes neuronales, árboles de regresión y bosques de regresión.

Aunque, a priori, los tipos de características utilizadas para entrenar a los regresores no son las ideales para aquéllos que proporcionaron los mejores resultados, ya que no son variables categóricas, la ausencia de aleatoriedad en la naturaleza de la información y el uso de imágenes independientes explica las buenas prestaciones obtenidas por estos estos algoritmos de regresión.

Ante esta problemática se optó por excluir estos planos del conjunto de calibración. Este proceso de filtrado permitió reducir sustancialmente los valores de RMSE y RAE, mejorando así los resultados del sistema de calibración.

Finalmente se ha presentado una segunda aplicación basada en la simplificación de la interpretación de la información de reconstrucción de la escena. Esta simplificación, consistente en la proyección de la información del espacio de disparidad sobre los planos Disparidad V y Disparidad U, permite identificar los obstáculos de la escena como rectas en las proyecciones.

Tal como se ha comentado previamente en la sección 7.1, esta aproximación es muy similar a la que presentan otros autores que también hacen uso de la información reducida, demostrándose así que el método de reconstrucción basado en planos es una generalización de estas implementaciones.

Se ha validado el funcionamiento del sistema propuesto con imágenes reales obtenidas del sistema de visión de un vehículo autoguiado, verificando así su utilidad para esta aplicación. En este caso, se ha realizado la implementación del algoritmo propuesto y probado el adecuado funcionamiento en tiempo real del sistema de detección de obstáculos, ya que esta es una característica imprescindible para su posible uso en un vehículo en marcha.

Bibliografía

Aggarwal, C. C. et al. (2018). Neural networks and deep learning. *Springer*, 10(978):3.

Akinyelu, A. A. and Blignaut, P. (2020). Convolutional neural network-based methods for eye gaze estimation: A survey. *IEEE Access*, 8:142581–142605.

Amroun, H., Ammi, M., and Hafid, F. (2021). Proof of concept: Calibration of an overhead line conductors' movements simulation model using ensemble-based machine learning model. *IEEE Access*, 9:163391–163411.

Banús, N., Boada, I., Bardera, A., and Toldrà, P. (2021). A deep-learning based solution to automatically control closure and seal of pizza packages. *IEEE Access*, 9:167267–167281.

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer Vision-ECCV 2006*, volume 3951, pages 404–417.

Becker, W. and Saltelli, A. (2015). *Design for Sensitivity Analysis*, pages 627 –673. Chapman and Hall.

Belhaoua, A., Kohler, S., and Hirsch, E. (2010). Error evaluation in a stereovision-based 3d reconstruction system. *EURASIP J. Image and Video Processing*, 2010.

Bertozzi, M. and Broggi, A. (1998). Gold: a parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing*, 7(1):62–81.

Bier, A. and Luchowski, L. (2009). Error analysis of stereo calibration and reconstruction. In *Computer Vision/Computer Graphics CollaborationTechniques: 4th International Conference, MIRAGE 2009, Rocquencourt, France, May 4-6, 2009. Proceedings 4*, pages 230–241. Springer.

Borgonovo, E. and Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3):869–887.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and regression trees. *Biometrics*, 40:874.

Broggi, A., Caraffi, C., Porta, P. P., and Zani, P. (2006). The single frame stereo vision system for reliable obstacle detection used during the 2005 darpa grand challenge on terramax. pages 745 – 752.

Chang, J.-R. and Chen, Y.-S. (2018). Pyramid stereo matching network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5410–5418.

Chen, L., Zhang, F., and Sun, L. (2020). Research on the calibration of binocular camera based on bp neural network optimized by improved genetic simulated annealing algorithm. *Ieee Access*, 8:103815–103832.

Cireşan, D., Meier, U., Gambardella, L. M., and Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. *Neural computation*, 22:3207–20.

Cirillo, P., Laudante, G., and Pirozzi, S. (2021). Vision-based robotic solution for wire insertion with an assigned label orientation. *IEEE Access*, 9:102278–102289.

Crisman, J. D. and Thorpe, C. E. (1991). Unscarf-a color vision system for the detection of unstructured roads. In *ICRA*, pages 2496–2501.

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., and Nießner, M. (2017). Scannet: Richly-annotated 3d reconstructions of indoor scenes.

Devore, J. L. et al. (2009). Probabilidad y estadística para ingeniería y ciencias. *Cengage Learning Editores*.

Diaz-Ramirez, V. H., Gonzalez-Ruiz, M., Kober, V., and Juarez-Salazar, R. (2022). Stereo image matching using adaptive morphological correlation. *Sensors*, 22(23).

Dima, C. and Hebert, M. (2005). Active learning for outdoor obstacle detection. In Elsevier, editor, *Proceedings of Robotics: Science and Systems (RSS '05)*, pages 9 – 16.

Ding, J., Liu, J., Zhou, W., Yu, H., Wang, Y., and Gong, X. (2011). Real-time stereo vision system using adaptive weight cost aggregation approach. *Eurasip Journal on Image and Video Processing - EURASIP J Image Video Process*, 20.

Dinh Nguyen, V., Nguyen, H., Thi Dinh, T., Lee, S.-J., and Jeon, J. (2016). Learning framework for robust obstacle detection, recognition, and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 18.

Donné, S., De Vylder, J., Goossens, B., and Philips, W. (2016). Mate: Machine learning for adaptive calibration template detection. *Sensors*, 16(11):1858.

Dovesi, P. L., Poggi, M., Andraghetti, L., Martí, M., Kjellström, H., Pieropan, A., and Mattoccia, S. (2020). Real-time semantic stereo matching. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10780–10787.

Ferretti, F., Saltelli, A., and Tarantola, S. (2016). Trends in sensitivity analysis practice in the last decade. *Science of The Total Environment*, 568:666–670.

Flynn, J., Neulander, I., Philbin, J., and Snavely, N. (2015). Deepstereo: Learning to predict new views from the world's imagery.

Girshick, R. (2015). Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448.

Guo, Y.-Q., Gu, M., and Xu, Z.-D. (2023). Research on the improvement of semi-global matching algorithm for binocular vision based on lunar surface environment. *Sensors*, 23(15).

Hadsell, R., Sermanet, P., Ben, J., Erkan, A., Scoffier, M., Kavukcuoglu, K., Muller, U., and LeCun, Y. (2009). Learning long-range vision for autonomous off-road driving. *Journal of Field Robotics*, 26(2):120–144.

Hancock, J. A. (1997). *High-Speed Obstacle Detection for Automated Highway Applications*. Carnegie Mellon University.

Hess, W., Kohler, D., Rapp, H., and Andor, D. (2016). Real-time loop closure in 2d lidar slam. pages 1271–1278.

Heuer, M., Al-Hamadi, A., Meinecke, M.-M., and Mende, R. (2012). Requirements on automotive radar systems for enhanced pedestrian protection. In *2012 13th International Radar Symposium*, pages 45–48.

Hirschmuller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341.

Hu, H.-f. (2006). Camera calibration and 3d reconstruction using rbf network in stereovision system. In *International Symposium on Neural Networks*, pages 375–382. Springer.

Hu, Z., Lamosa, F., and Uchimura, K. (2005). A complete u-v-disparity study for stereovision based 3d driving environment analysis. *Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, pages 204–211.

Iooss, B. and Saltelli, A. (2017). *Introduction to Sensitivity Analysis*, pages 1103–1122. Springer International Publishing, Cham.

Itu, R. and Danescu, R. G. (2020). A self-calibrating probabilistic framework for 3d environment perception using monocular vision. *Sensors*, 20(5):1280.

Khoshelham, K. and Elberink, S. O. (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., and Inman, D. J. (2021). 1d convolutional neural networks and applications: A survey. *Mechanical systems and signal processing*, 151:107398.

Kolakowski, M. (2021). Automated calibration of rss fingerprinting based systems using a mobile robot and machine learning. *Sensors*, 21(18):6270.

Labayrade, R. and Aubert, D. (2004). Robust and fast stereovision based obstacles detection for driving safety assistance. *IEICE Transactions*, 87-D:80–88.

Labayrade, R., Aubert, D., and Tarel, J.-P. (2002). *Real time obstacle detection in stereovision on non flat road geometry through "v-disparityrepresentation*, volume 2. Institute of Electrical and Electronics Engineers.

Li, X., Chen, L., Li, S., and Zhou, X. (2020). Depth segmentation in real-world scenes based on u–v disparity analysis. *Journal of Visual Communication and Image Representation*, 73:102920.

Lindgren, L. (2012). Vision system and method for a motor vehicle. US Patent App. 13/265,896.

Liu, H., Wu, C., and Wang, H. (2023). Real time object detection using lidar and camera fusion for autonomous driving. *Scientific Reports*, 13.

Llorca, D. F., Sotelo, M. A., Parra, I., Naranjo, J. E., Gavilan, M., and Alvarez, S. (2009). An experimental study on pitch compensation in pedestrian-protection systems for collision avoidance and mitigation. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):469–474.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–.

Marita, T., Oniga, F., Nedevschi, S., and Graf, T. (2007). Calibration accuracy assessment methods for stereovision sensors used in vehicles. In *2007 IEEE International Conference on Intelligent Computer Communication and Processing*, pages 111–118. IEEE.

Marquet, J. (2007). Method for the automatic calibration of a stereovision system. US Patent App. 11/573,326.

Mathworks (2022). Regression learner app. available online: https://www.mathworks.com/help/stats/regression-learner-app.html (accessed on 12 december 2022).

Michels, J., Saxena, A., and Ng, A. Y. (2005). High speed obstacle avoidance using monocular vision and reinforcement learning. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, page 593–600, New York, NY, USA. Association for Computing Machinery.

Nagatomo, S., Hayashi, J., and Hata, S. (2010). Proposal of calibration error-tolerant 3d measurement method using stereo vision. *Ieej Transactions on Electronics, Information and Systems*, 130:490–495.

Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W., et al. (1996). *Applied linear statistical models*. Irwin Chicago.

Ollero, A., Arrue, B., Ferruz, J., Heredia, G., Cuesta, F., López-Pichaco, F., and Nogales, C. (1999). Control and perception components for autonomous vehicle guidance. application to the romeo vehicles. *Control Engineering Practice*, 7(10):1291–1299.

Park, J.-H., Shin, Y.-D., Bae, J.-H., and Baeg, M.-H. (2012). Spatial uncertainty model for visual features using a kinect™ sensor. *Sensors*, 12(7):8640–8662.

Prokos, A., Karras, G., and Petsa, E. (2010). Automatic 3d surface reconstruction by combining stereovision with the slit-scanner approach. *hand*, 2:2.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.

Rosa, R. G., de Pedro Lucio, T., Fernández-Vallejo, C. G., and Naranjo, J. E. (2003). Autopía control automático de automóviles. resultados conseguidos y actualidad del proyecto cooperativo isaac.

Santoro, M., Alregib, G., and Altunbasak, Y. (2012). Misalignment correction for depth estimation using stereoscopic 3-d cameras. In *2012 IEEE 14th International Workshop on Multimedia Signal Processing, MMSP 2012 - Proceedings*, pages 19–24.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42.

Scharstein, D., Szeliski, R., and Zabih, R. (2001). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*, pages 131–140.

Seber, G. A. and Lee, A. J. (2012). *Linear regression analysis*. John Wiley & Sons.

Seitz, S., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 519–528.

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Singh, S. P., Wang, L., Gupta, S., Goli, H., Padmanabhan, P., and Gulyás, B. (2020). 3d deep learning on medical images: A review. *Sensors*, 20(18).

Sui, L. and Zhang, T. (2010). Camera calibration method based on bundle adjustment. In *Fourth International Conference on Experimental Mechanics*, volume 7522, pages 1497–1501. SPIE.

Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic Robotics*. Intelligent Robotics and Autonomous Agents series. MIT Press.

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 839–846.

Vajs, I., Drajic, D., Gligoric, N., Radovanovic, I., and Popovic, I. (2021). Developing relative humidity and temperature corrections for low-cost sensors using machine learning. *Sensors*, 21(10):3338.

Wackerly, D. D., Mendenhall III, W., Scheaffer, R. L., Yescas Milanés, J., et al. (2002). *Estadística matemática con aplicaciones*. CENCAGE Learning.

Wang, H., Wei, Z., Wang, S., Ow, C., Ho, K., Feng, B., and Lubing, Z. (2011). Real-time obstacle detection for unmanned surface vehicle.

Wang, J. H., Yang, Z., and Wu, Y. P. (2013). Calibration accuracy and reconstruction accuracy of stereovision system. *Applied Mechanics and Materials*, 321:1499–1503.

Wang, S., Seo, J., Jeon, H., Lim, S., Park, S., and Lim, Y. (2023). Horizontal attention based generation module for unsupervised domain adaptive stereo matching. *IEEE Robotics and Automation Letters*, 8(10):6779–6786.

Wang, W.-C. V., Lung, S.-C. C., and Liu, C.-H. (2020). Application of machine learning for the in-field correction of a pm2. 5 low-cost sensor network. *Sensors*, 20(17):5002.

Westfall, P. H. and Arias, A. L. (2020). *Understanding regression analysis: a conditional distribution approach*. CRC Press.

Wilkinson, L. (2006). Revising the pareto chart. *The American Statistician*, 60(4):332–334.

Xu, D., Zeng, Q., Zhao, H., Guo, C., Kidono, K., and Kojima, Y. (2014). Online stereovision calibration using on-road markings. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 245–252. IEEE.

Xu, X., Dong, S., Xu, T., Ding, L., Wang, J., Jiang, P., Song, L., and Li, J. (2023). Fusionrcnn: Lidar-camera fusion for two-stage 3d object detection. *Remote Sensing*, 15(7).

Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Basri, R., and Lipman, Y. (2020). Multiview neural surface reconstruction by disentangling geometry and appearance.

Ye, X., Yan, B., Liu, B., Wang, H., Qi, S., Chen, D., Wang, P., Wang, K., and Sang, X. (2022). Improved real-time three-dimensional stereo matching with local consistency. *Image and Vision Computing*, 124:104509.

Zhang, Z., Shen, Z., Liu, J., Shu, J., and Zhang, H. (2024). A binocular vision-based crack detection and measurement method incorporating semantic segmentation. *Sensors*, 24(1).

Zhao, W. and Nandhakumar, N. (1996). Effects of camera alignment errors on stereoscopic depth estimates. *Pattern Recognition*, 29(12):2115–2126.

Zhao, X., Sun, P., Xu, Z., Min, H., and Yu, H. (2020). Fusion of 3d lidar and camera data for object detection in autonomous vehicle applications. *IEEE Sensors Journal*, PP:1–1.

Žbontar, J. and Lecun, Y. (2015). Stereo matching by training a convolutional neural network to compare image patches. *Journal of Machine Learning Research*, 17.