

Sistema de realidad virtual para exploración 3D con visión protésica simulada

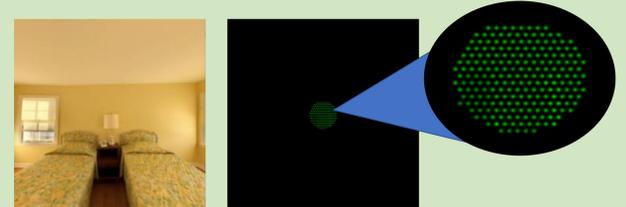
María Santos-Villafranca*, Alejandro Pérez-Yus, Jesús Bermúdez-Cameo, José J. Guerrero

Instituto Universitario de Investigación en Ingeniería de Aragón (I3A), Grupo de Robótica, Percepción y Tiempo Real, Universidad de Zaragoza, C/María de Luna 3, 50018, Zaragoza, España

INTRODUCCIÓN

Los recientes avances en las prótesis visuales han demostrado que es posible restaurar parte de la visión en determinados casos de deficiencia visual, aunque con limitaciones como la baja resolución o un campo de visión reducido. Para experimentar más fácilmente se utilizan los simuladores de prótesis visuales (SPV). En este trabajo, se presenta un SPV con las siguientes características:

- **SPV inmersivo, flexible y realista**
- **Exploración entornos creados por un ordenador**
- **Integración del movimiento real de la cabeza gracias a las gafas de realidad virtual Oculus Rift DK2**
- **Permitir un desplazamiento mediante un mando o joystick**

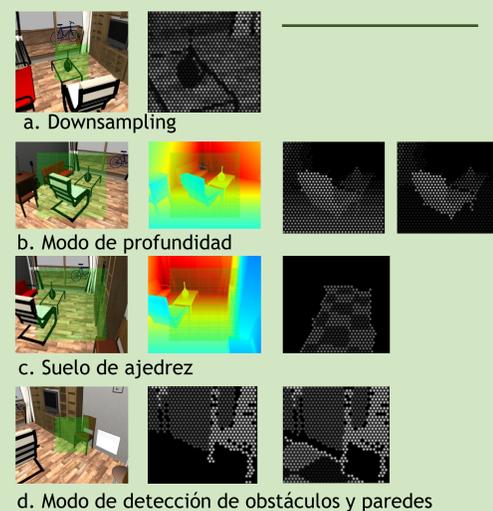


SPV Y ENTORNO VIRTUAL

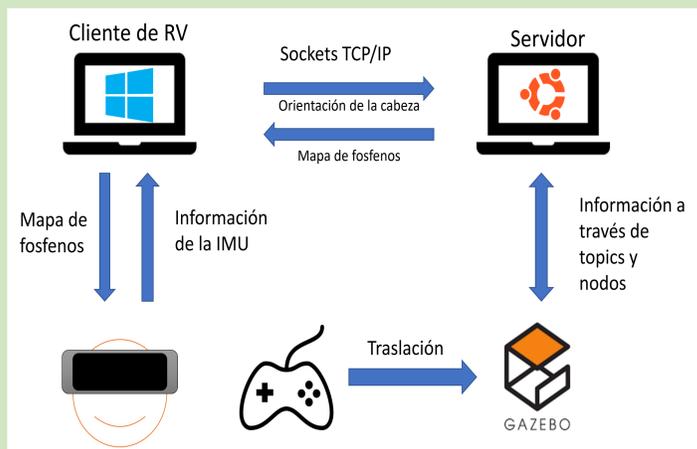
Fosfenos: puntos luminosos que ven las personas operadas. El nivel de intensidad de cada uno depende de la representación de la imagen que se visualice.

Robot: usado para modelar a la persona.

- Estructura montada sobre una base de Turtlebot 3 que integra sensores de odometría
- Cámara de color y profundidad (Asus Xtion Pro Live)
- Modelo de persona con la cámara a la altura de los ojos y articulaciones para emular el movimiento de la cabeza



SISTEMA DE REALIDAD VIRTUAL

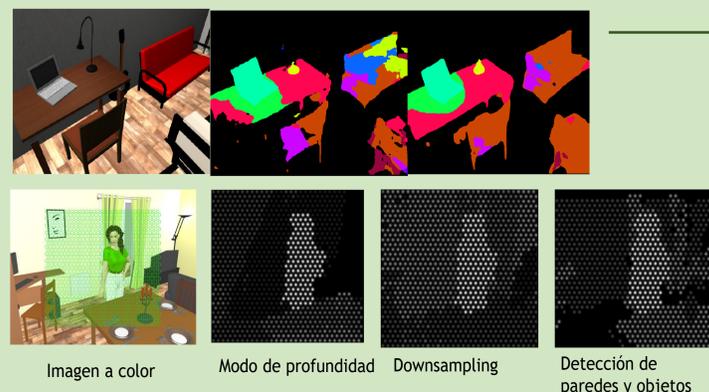


Dos ordenadores servidor/cliente conectados a través de internet (socket TCP/IP).

- **Cliente:** ordenador que sirve de interfaz con el sistema de realidad virtual. Envía la información de la orientación de las gafas y recibe la trama de intensidades del mapa de fosfenos del servidor.
- **Servidor:** dónde se ejecuta el entorno virtual y todo el procesamiento pesado. Los ángulos de giro de las gafas se comunican al entorno virtual, se genera en este equipo todo el procesamiento de la imagen para convertirla a fosfenos [1], [2]

SEGMENTACIÓN SEMÁNTICA

Se ha introducido en un nodo ROS la implementación en pytorch de modelos de segmentación semántica con el dataset de escenas MIT ADE20K [5],[6]. Se usa la red neuronal pre-entrenada DilatedNet [3], [4], capaz de detectar de manera muy eficaz 150 objetos en total. Esta red resulta de gran utilidad para resaltar los objetos mas importantes y facilitar la detección de estos en la visión fosfénica, además se puede ejecutar a tiempo real.



CONCLUSIÓN

En este trabajo se ha presentado un SPV flexible y realista, que permite una experiencia inmersiva gracias a la comunicación con un ordenador que dispone de gafas de realidad virtual. A diferencia de otros simuladores, éste es capaz de representar diferentes entornos virtuales de interiores, con diferentes estancias, con la posibilidad de explorarlos completamente sin necesidad de desplazarse en la realidad. Además, este SPV también tiene la posibilidad de utilizarse con una cámara real. Este SPV habilita para una exploración completa e inmersiva del entorno para evaluar el reconocimiento de objetos, estancias y navegación por el mismo. El objetivo es realizar experimentos que permitan comparar distintos modos de representación. Para ello también se ha implementado un nodo de segmentación semántica que permite resaltar objetos relevantes.

Agradecimientos

1 Este trabajo ha sido financiado por el proyecto RTI2018-096903-B-I00 (MINECO/FEDER, UE)

* Más info: María Santos,
727328@unizar.es

1 L. Montano Oliván, Visual impaired guidance with simulated prosthetic vision in virtual environments, Zaragoza. 2015.
2 A. Perez-Yus, J. Bermudez-Cameo, G. Lopez-Nicolas, and J. J. Guerrero, Depth and motion cues with phosphenes patterns for prosthetic vision, 2017.
3 L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, 2018.
4 Fisher Yu and Vladlen Koltun, Multi-Scale Context Aggregation by Dilated Convolutions, 2016.
5 B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso and A. Torralba, Scene Parsing through ADE20K Dataset, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
6 B. Zhou et al., Semantic Understanding of Scenes through the ADE20K Dataset, 2021.