XLIV Jornadas de Automática 2023

https://jautomatica.es

Simulador inmersivo de visión protésica modelando estímulos espacio-temporales

Santos-Villafranca, M., Tomas-Barba, J., Perez-Yus, A., Bermudez-Cameo, J., Guerrero, J.J.

Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, España.

To cite this article: Santos-Villafranca, M. Tomas, J., Perez-Yus, A., Bermudez-Cameo, J., Guerrero, J.J. 2023. Im-mersive prosthetic vision simulator modelling spatiotemporal stimuli.
XLIV Jornadas de Automática, 879-884. https://doi.org/10.17979/spudc.9788497498609.879

Resumen

Recientes avances han demostrado que, en ciertos casos de deficiencia visual, la visión puede ser parcialmente restituida mediante prótesis visuales. Debido a sus limitaciones, surge el interés por desarrollar métodos de visión por computador para extraer información relevante del entorno y adaptarla a las prótesis. Para poder evaluar la eficacia de estos métodos, debido a la escasez de personas operadas, se utilizan simuladores de prótesis visuales, que permiten experimentar con personas de visión sana. En este trabajo, presentamos un nuevo simulador realista, integrado en un *framework* de robótica, que permite probar distintos modos de representación mediante gafas de realidad virtual. Una de las principales novedades es la inclusión de un modelo temporal, inspirado en experimentos con pacientes reales, para transmitir la dimensión del tiempo en la generación de estímulos visuales. Además, permite la inmersión total del usuario en un entorno virtual en el que se ha integrado una red neuronal de segmentación semántica para ayudar a detectar objetos y personas.

Palabras clave: Navegación, programación y visión de robots, Redes neuronales, Algoritmos en tiempo real, Trabajar en entornos reales y virtuales, Interfaces inteligentes, Tecnología asistiva e ingeniera de rehabilitación

Immersive prosthetic vision simulator modelling spatiotemporal stimuli

Abstract

Recent advances have shown that, in certain cases of visual impairment, vision can be partially restored by means of visual prostheses. Due to its limitations, there is interest in developing computer vision methods to extract relevant information from the environment and adapt it to the prosthesis. In order to evaluate the effectiveness of these methods, due to the scarcity of operated people, visual prosthesis simulators are used, which allow experimenting with healthy sighted people. In this work, we present a new realistic simulator, integrated in a robotics framework, which allows testing different modes of representation through virtual reality goggles. One of the main novelties is the inclusion of a temporal model, inspired by experiments with real patients, to convey the dimension of time in the generation of visual stimuli. In addition, it allows total immersion of the user in a virtual environment in which a semantic segmentation neural network has been integrated to help detect objects and people.

Keywords: Robot Navigation, Programming and Vision, Neural networks, Real-time algorithms, Work in real and virtual environments, Intelligent interfaces, Assistive technology and rehabilitation engineering

1. Introducción

El sentido humano que más información aporta es la visión, pero algunas personas lo han perdido total o parcialmente debido a accidentes o enfermedades degenerativas. Según la OMS, en el mundo hay al menos 2200 millones de personas con deterioro de la visión cercana o distante (WHO, 2022). 88,4 millones por ceguera o deterioro moderado o grave de la visión distante, 94 millones por cataratas, 8 millones por degeneración macular relacionada con la edad, 7,7 millones por glaucoma y 3,9 millones por retinopatía diabética (Steinmetz et al., 2021).

*Autor para correspondencia: m.santos@unizar.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Numerosos trabajos han demostrado que al estimular diferentes partes de la vía visual mediante electrodos, se puede conseguir que los pacientes perciban formas luminosas llamadas *fosfenos*. Este fenómeno ha permitido el desarrollo de prótesis visuales que pueden implantarse en el córtex visual (Brindley and Lewin, 1968), el nervio óptico (Veraart et al., 1998) o la retina (Humayun et al., 1996). Hay que tener en cuenta que estas últimas provocan una remodelación retiniana (Marc et al., 2008; Jones et al., 2016) no dejándola intacta.

Los experimentos demuestran que los pacientes son capaces de detectar fosfenos en electrodos individuales. Desafortunadamente, la resolución de la rejilla de fosfenos producidos se ve limitada por la biología, la tecnología y la seguridad del paciente (Lui et al., 2012) y tienen serias limitaciones (Meffin, 2013; Beyeler et al., 2017b) como la baja resolución, un campo de visión reducido, bajo rango de luminosidad, o ruido (Dagnelie, 2006) (Figura 1). Además, estos no tienen forma circular perfecta como se había supuesto hasta la fecha, si no que tienen formas distorsionadas y alargadas que se desvanecen con el tiempo y dependen de la prótesis y del paciente (Beyeler et al., 2019)

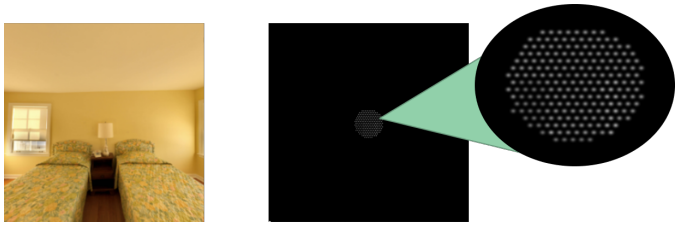

Figura 1: A la izquierda, imagen del entorno como lo vería una persona de visión sana, a la derecha, mapa con modelo *Scoreboard* de fosfenos de como vería una persona operada con prótesis

Dos de las prótesis más usadas y que cuentan con aprobación clínica para salir al mercado, son el Argus II (Humayun et al., 2012) y el Alpha IMS (Rothermel et al., 2008; Zrenner et al., 2011). Argus II consta de una cámara montada en el centro de unas gafas que capta la información visual, la cual es convertida en estimulaciones eléctricas que llegan a la rejilla de electrodos que provocan el mapa de fosfenos. Por otro lado, Alpha IMS usa microfotodiodos dentro del globo ocular para capturar la información luminosa. A pesar de que la resolución de Alpha IMS es mayor que la de Argus II (1500 frente a 60 electrodos), algunos estudios han demostrado que este mayor número de fosfenos no correla con una mejora significativa en la agudeza visual en algunas tareas (Meffin, 2013). La razón puede ser que el control de las variables en el Alpha es limitado mientras que el Argus permite ajustarlas para cada electrodo (Stronks and Dagnelie, 2014). Para nuestro trabajo nos hemos inspirado por el esquema en el que la información se captura con una cámara, y aprovecharemos métodos de visión por computador, deep learning y robótica para extraer información del entorno y ayudar a comunicársela al paciente.

Estas prótesis se encuentran en una fase temprana de investigación, por lo que existen muy pocas personas operadas, lo cual hace que la experimentación con estos pacientes sea prácticamete inviable. Por eso, los simuladores de prótesis visuales (SPV) han surgido para poder realizar pruebas en sujetos de visión sana de tal manera que puedan visualizar lo que vería una persona con prótesis a través de una pantalla (Sanchez-Garcia et al., 2021) o de un dispositivo montado en la cabeza (HMD) (Perez-Yus et al., 2023). La finalidad consiste en evaluar los factores más limitantes a la hora de realizar diversas tareas, para contribuir con la línea de investigación que se centra en mejorar estas prótesis, y en introducir y testear nuevos modos de representación fosfénica para aumentar la información relevante (Barnes, 2013) y evaluar su viabilidad dentro de prótesis ya existentes, con las limitaciones actuales.

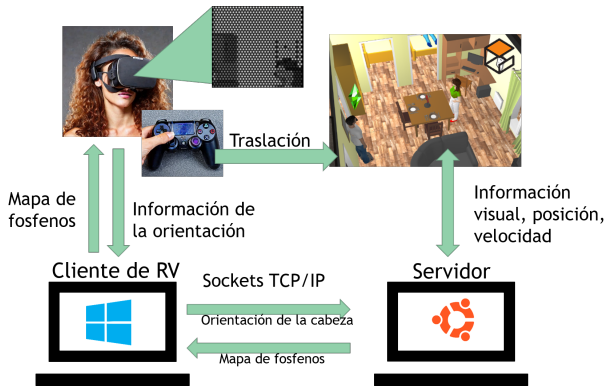

Figura 2: Esquema general del SPV

Algunos ejemplos de experimentos realizados para determinar los factores limitantes se citan a continuación. En (Cha et al., 1992) se concluye que con un campo de visión de 30º e introduciendo una serie de píxeles de 25×25 distribuidos dentro del área visual foveal, se puede conseguir una movilidad útil en un entorno simple en el reconocimiento de patrones. En (Dagnelie et al., 2007) se exploró el número mínimo de fosfenos para la movilidad en entornos reales y virtuales. En (Zhao et al., 2017) se simuló un entorno virtual en el que los sujetos realizan tareas de identificación y navegación con un campo de visión de 50º. En (van Rheede et al., 2010) se utiliza un HMD con *eye tracking* para estudiar su impacto en el escaneo visual con movimientos de cabeza a la hora de explorar el entorno. También se han usado técnicas de *deep learning* (Leo et al., 2018) para intentar representar el máximo de información a modo de realidad aumentada en el mapa de fosfenos. En (Feng et al., 2017) se ha usado para detectar la información de bordes estructurales.

En cuanto a las representaciones de fosfenos, en (Chen et al., 2009b,a) se propone un marco estandarizado según la descripción ofrecida por sujetos implantados. Más recientemente, se han realizado modelos más complejos de la percepción de fosfenos, incluyendo aspectos espaciales y temporales Horsager et al. (2009); Nanduri et al. (2012); Beyeler et al. (2019), algunos de los cuales aparecen implementados en la librería *pulse2percept* (Beyeler et al., 2017a). Dicha librería ha sido usada por trabajos como (Kasowski and Beyeler, 2022). Utilizando estos trabajos como inspiración, hemos realizado una nueva implementación del modelo temporal de (Horsager et al., 2009) en tiempo real, teniendo en cuenta la respuesta dinámica no lineal de los fosfenos.

Un esquema general del funcionamiento del simulador es presentado en la Figura 2. Se simula la visión protésica con un modelo temporal permitiendo la exploración del mismo moviendo la cabeza, gracias a las gafas de realidad virtual, y trasladándose con un mando o teclado.

El objetivo es crear un framework para experimentar y evaluar algoritmos y técnicas que pudiesen llegar a ser implementadas en prótesis reales en teimpo real. En el simulador propuesto, se puede visualizar lo que el usuario percibe y monitorizar el movimiento del mismo por el entorno virtual desde una vista aérea. Por otro lado, el uso de un *framework* de desarrollo de robots (ROS) permite un desarrollo modular y expandible, la emulación o uso directo de sensores y cámaras reales, y la incorporación de métodos ya desarrollados en el campo de la robótica.

La visión por computador puede utilizarse para asistir en la identificación de objetos o personas. En este trabajo se ha integrado una red neuronal de segmentación semántica en el simulador, que permite detectar los elementos relevantes de la escena y resaltarlos en visión fosfénica.

Un resumen de las principales contribuciones de este trabajo son las siguientes:

- Es inmersivo gracias a la introducción de las gafas de realidad virtual Oculus Rift DK2.
- Al conectar dos equipos, permite realizar todo el procesamiento pesado en un ordenador o servidor externo al cual se conecta en remoto.
- El modelo temporal de (Horsager et al., 2009) permite representar dinámicas temporales de fosfenos realistas.
- Se ha introducido una red neuronal para segmentación semántica que ayuda a la percepción del entorno.

El contenido de este documento se estructura en, una explicación detallada del modelo espaciotemporal integrado en el simulador (Horsager et al., 2009), una descripción de la arquitectura del sistema considerando el flujo de información entre equipos, y finalmente se comenta la implementación de la segmentación y resaltado de objetos. Para terminar, se presentan las conclusiones obtenidas en este estudio y se plantean las posibles líneas de investigación futura.

2. Modelado de fosfenos

Los primeros estudios de prótesis visuales describían los fosfenos como pequeños puntos luminosos aislados (*Scoreboard model*). Sin embargo, en estudios más recientes se han introducido modelos temporales, espaciales y espacio-temporales más avanzados (Thompson et al., 2003; Nanduri et al., 2008; Horsager et al., 2009; Nanduri et al., 2012; Beyeler et al., 2019; Granley and Beyeler, 2021) que, a partir de ensayos con pacientes implantados, han demostrado que los fosfenos creados al estimular un único electrodo suelen tardar en aparecer y desvanecerse, y además se perciben con formas distorsionadas y alargadas. Esto se debe a que los electrodos provocan una activación en las fibras axónicas (*Axon Map Model*). Además, si se estimulan varios electrodos a la vez, no se puede conocer el comportamiento de ellos simplemente combinando linealmente las percepciones independientes de cada uno.

El modelo espacio-temporal de (Granley and Beyeler, 2021) simula el encendido y apagado de los fosfenos así como la variación de su elongación dependiendo de la amplitud y frecuencia del estímulo. Sin embargo, simplifica la complejidad de la parte temporal prediciendo el estímulo más intenso, sólo tiene la opción de estimular electrodos individuales y necesita de bastante tiempo de procesamiento. Como además la forma y disposición de los fosfenos dependen de la prótesis y del paciente, lo cual varía demasiado, hemos decidido representarlos con el modelo espacial *Scoreboard*. En este modelo, cada fosfeno representa como un círculo luminoso sin anomalías cromáticas, cuya luminosidad sigue una distribución gaussiana de dos dimensiones (Hayes et al., 2003), de tal forma que el centro está mas iluminado y se difumina conforme se aleja de él (Chen et al., 2009a). Otros trabajos han representado los fosfenos como círculos con intensidad constante en toda su área (Thompson et al., 2003; Dagnelie et al., 2006) o como píxeles en imágenes de muy baja resolución (Sommerhalder et al., 2004), sin embargo, es una sobresimplificación (Dagnelie et al., 2006). Se ha introducido la posibilidad de añadir ruido gaussiano en las posiciones de los fosfenos para indicar que la localización no dispone una malla perfectamente ordenada, y se ha añadido *drop-out* para simular que no todos los fosfenos elicitados se llegan a activar (Thompson et al., 2003; Dagnelie et al., 2006). Por otra parte, el campo de visión en los dispositivos protésicos actuales es de 10º-20º. Cabe mencionar que todos estos parámetros son fácilmente modificables.

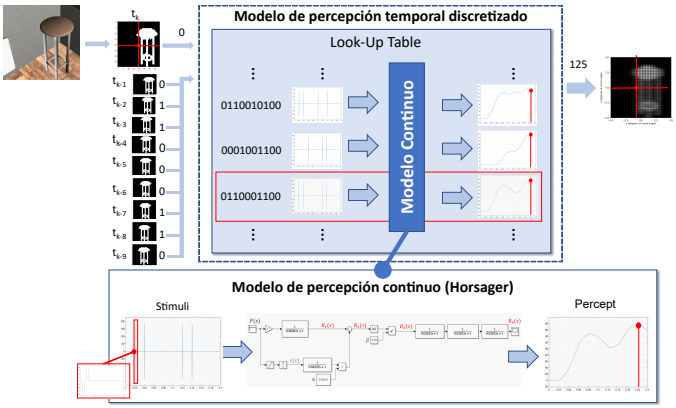

Figura 3: Esquema general del modelo temporal de un fosfeno. En rojo el valor de la Look-Up Table para el fosfeno señalado en cada instante.

En cuanto a la dinámica temporal que sigue la luminosidad de los fosfenos, utilizamos el modelo temporal de (Horsager et al., 2009), en cuyos experimentos revela que el tiempo de respuesta del estímulo visual es de unos 200 ms, mucho mayor que los cambios temporales que la visión es capaz de percibir (20 ms). Partiendo de sus ecuaciones, se ha modelado su comportamiento dependiendo del pulso de entrada (Figura 3).

Hemos simulado el mismo tipo de estímulo eléctrico que con el implante Alpha-AMS de Retina Implant AG, en el cual se utilizan pulsos bifásicos en 1600 fosfenos. La salida de un único pulso representa el nivel de iluminación del fosfeno respecto al tiempo. Se enciende bastante rápido pero el apagado del mismo es mas lento y genera una sensación de estela visual. Sin embargo si se producen varios pulsos seguidos el comportamiento no es lineal y requiere realizar cálculos costosos a un tiempo de muestreo muy pequeño respecto de las frecuencias de trabajo. Para poder implementarlo en tiempo real, se ha construido un modelo de percepción temporal discretizado que almacena el valor de intensidad del píxel según la información obtenida en instantes anteriores. Para ello, se han obtenido los

resultados de iluminación a partir de las posibles combinaciones que pueden haberse dado en el *frame* actual y los 9 anteriores, donde los estímulos se calculan a partir de las imágenes de entrada binarizadas y reescaladas de tal forma que cada píxel se corresponde a un fosfeno. Se ha guardado dicha información en una Look-Up Table cuya entrada es una secuencia de unos y ceros dependiente del frame actual y los 9 anteriores. En orden secuencial se representa con 0 si no ha habido pulso y 1 si sí que lo ha habido.

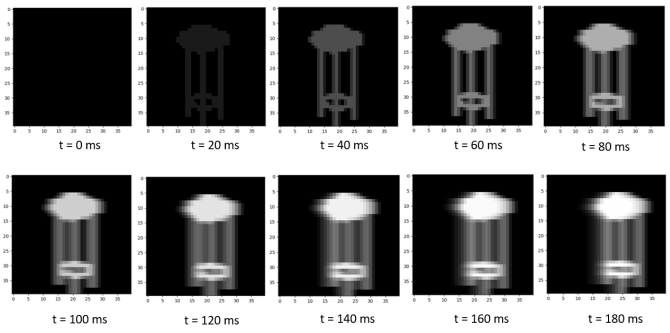

Figura 4: Salida del modelo temporal del taburete desplazándose hacia la derecha

El resultado de aplicar el modelo discreto anterior a una imagen binarizada de un taburete desplazádose hacia la derecha se muestra en la Figura 4.

3. Simulador de realidad virtual

El uso de entornos virtuales es importante a la hora de realizar experimentos de forma segura y barata. La evaluación es mucho más precisa ya que es un entorno fácilmente modificable y en el que siempre se conoce de forma exacta la posición del sujeto, el tiempo y las posibles colisiones con obstáculos. También se pueden realizar experimentos de forma sistemática, en igualdad de condiciones y con más cantidad y variedad de sujetos. Nuestra propuesta consiste en integrar el modelo simulado de prótesis visuales (Sección 2) en un *framework* de robótica (Robot Operating System (ROS, 2022)) de forma que los algoritmos de percepción y representación puedan utilizarse indistintamente en entornos reales como en entornos virtuales. La arquitectura hardware de este simulador se compone de dos sistemas, por un lado un servidor el cual realiza el procesamiento de la imagen para convertirla a fosfenos, y por el otro un cliente que hace de interfaz con las gafas de realidad virtual (Figura 2). ROS es un conjunto de librerías que facilita la comunicación entre módulos de software (nodos) a través de unos canales de información (*topics*). Además hace de interfaz con sensores y actuadores reales y permite integrar librerías de algoritmos del estado del arte de robótica y visión por computador. El entorno virtual utilizado para el simulador es Gazebo [1], un simulador 3D multirrobot, dinámico, *open-source* con visualización de datos, simulación de entornos remotos, con colisiones en entornos tanto internos como externos, en el que se pueden añadir inercias, modificar velocidades e incluso colocar objetos que se muevan según una cierta trayectoria.

Figura 5: Izquierda: Ejemplo de un entorno de Gazebo importado desde Sweethome3D. Derecha: Modelo Blindbot

Se ha desarrollado un avatar humano, modelado en Gazebo como un robot (*Blindbot*, ver Figura 5, derecha), para situar al sujeto en el mundo virtual. Se trata de una readaptación del Turtlebot 3, el cual se puede mover con un teclado o un mando joystick y tiene sensores de odometría y colisiones integrados. Para la visualización de imágenes, se ha colocado la cámara RGB-D Asus Xtion Pro Live a la altura de los ojos del sujeto en la que se han añadido movimientos rotatorios entorno a su centro para simular una rótula esférica y que se asemeje todo lo posible al movimiento real de la cabeza. Estos cambios se han conseguido modificando los diferentes *links* (describen cuerpos rígidos) y *joints* (describen la cinemática y dinámica de la articulación) para que los movimientos sean posibles y para que los diferentes elementos se acoplen como uno solo.

La información visual de la cámara se publica en diferentes topics, uno de color, otro de profundidad, y otro de nubes de puntos, información que, posteriormente puede ser usada y procesada para generar diferentes representaciones fosfénicas en función de la tarea que pretendamos abordar (e.g. navegación, detección de objetos, evitación de obstáculos). A partir de esta información visual, se realiza el paso de imagen a mapa de fosfenos. Se simulan cámaras y robots reales, por lo que, es sencillo trasladar todo el trabajo a la realidad y que los dispositivos reales publiquen la información visual en los topics.

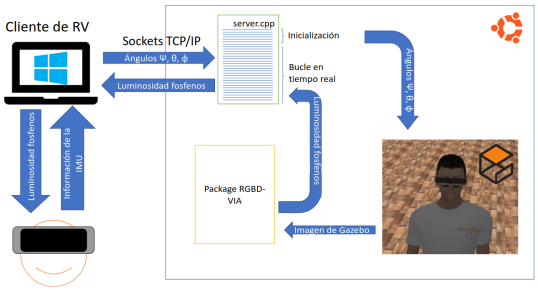

Figura 6: Esquema de envío y recibo de información entre equipos

Para añadir más realismo al entorno virtual, se han importado entornos de Sweethome3D (Sweethome3D, 2022) (Figura 5, izquierda) para simular entornos de interiores más realistas y complejos.

Para poder controlar el desplazamiento y la rotación de la cámara, se necesita la información de orientación de la cabeza del usuario estimada a partir de la unidad de medida inercial (IMU) integrada en las gafas de realidad virtual. El cliente de realidad virtual (Figura 6) es un equipo que hace de interfaz

[1] `https://gazebosim.org`

con las gafas de realidad virtual Oculus Rift DK2. Por un lado, estima la orientación de las gafas a partir de la información inercial de la IMU y la envía al simulador principal (servidor). Por otro lado, es el encargado de recibir la luminosidad deseada de cada uno de los fosfenos y mostrar en la pantalla de las gafas siguiendo el modelo espacial y temporal. El cliente de realidad virtual procesa el modelo espacial de la malla de fosfenos (posición, tamaño, apariencia) actualizando la luminosidad de los fosfenos a partir de la trama de intensidades recibida.

Este equipo se conecta al servidor a través de internet (sockets TCP-IP), con lo que, no es necesario que ambos equipos se encuentren físicamente próximos ni en la misma red local. El servidor envía la información de orientación de las gafas que mueve la cámara de Gazebo. Por otro lado, dicho nodo se suscribe a un topic en el que lee la información de la luminosidad de cada fosfeno y la envía a través del socket al cliente (Figura 6). El proceso completo se puede llegar a ejecutar a unos $10FPS$ considerando la ejecución del entorno virtual, el procesamiento total de fosfenos (que incluye modos de representación aumentada y el modelo espacio-temporal), la visualización en las Oculus y el envío y recibo de información.

4. Segmentación semántica de objetos

Para intentar aportar mayor información en la representación fosfénica, se ha añadido una red neuronal de segmentación de objetos y se han seleccionado algunos para resaltarlos al nivel máximo de luminosidad (Figura 7) indicando así al paciente la ubicación del objeto y su forma. De esta manera, la navegación, el reconocimiento de objetos, y por ende, el de estancias debido al tipo de objetos dentro de ellas, resulta más sencilla.

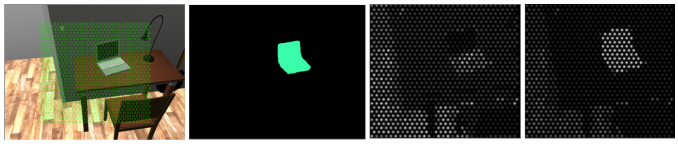

Figura 7: De izquiera a derecha: imagen a color donde la rejilla verde muestra el área que se transforma a fosfenos, resultado de la predicción de la red, mapa de fosfenos base, mapa de fosfenos con el objeto segmentado resaltado.

En el servidor que realiza todo el procesamiento pesado se ha desarrollado un nuevo nodo ROS que ejecuta una implementación en Pytorch de la red pre-entrenada DilatedResNet-18 + PPM. De todas las que prueban en el github y papers oficiales del dataset ADE20K ((Zhou et al., 2017, 2019)), es de los que mejores métricas obtiene manteniendo un FPS razonable para su introducción al simulador (mIoU: 42.19, acc: 80.59, overall score: 61.39, FPS: 6.8). En particular, la arquitectura que utilizamos se compone de Resnet50Dilated, dos redes en cascada propuestas en (Zhou et al., 2017) donde el backnbone son SegNet (Badrinarayanan et al., 2017) y DilatedVGG (Chen et al., 2017; Yu and Koltun, 2015), como encoder y un decoder PPM deepsup (Pyramid Pooling Module) que agrega información contextual multiescala en la escena con el dataset ADE20K (Zhou et al., 2017, 2019). Este dataset contiene 150 clases diferentes de objetos, entre los cuales hacemos una selección, ya que si se resaltasen todos no se distinguirían unos objetos de otros.

Para que la inferencia de segmentación semántica se pueda integrar en un esquema de tiempo real se ha ajustado el número de iteraciones del algoritmo a 3 a una velocidad de $4FPS$ (unos $0{,}25s$). La ejecución completa del nodo incluye la lectura de la información visual del topic de cámara, la inferencia del *score* de cada clase así como la generación y publicación de las máscaras binarias de segmentación semántica.

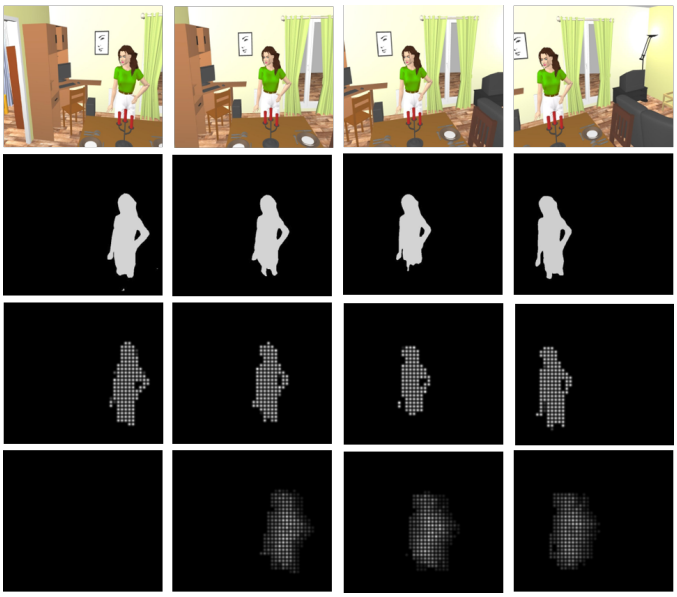

Figura 8: Frames de imagen a color según se mueve la cámara hacia la derecha en la primera línea. Resultado de la segmentación semántica en la segunda línea. Modelo *Scoreboard* sin dinámicas temporales en la tercera línea. Resultado del modelo temporal en la última línea

Un ejemplo de la combinación del resultado de aplicar la segmentación semántica y el modelo temporal de Horsager en nuestro simulador se puede apreciar en la (Figura 8).

5. Conclusiones

En este trabajo presentamos un simulador de visión protésica (SPV) que incluye modelos espacio-temporales de estímulos visuales y permite la interacción en un entorno artificial utilizando un sistema de realidad virtual inmersivo. La inclusión de un modelo espacio-temporal aporta mayor realismo al simulador, lo cual permite extraer conclusiones más valiosas de los experimentos al hacer que la experiencia sea más semejante a la de los pacientes con implantes. El uso de Gazebo y ROS permite el funcionamiento tanto en un entorno virtual como en uno real. Por otro lado, su estructura dividida en dos equipos diferentes, hace posible que, aunque ejecute una red neuronal y otros procesos costosos, pueda ser usado en tiempo real y sólo se necesite utilizar un equipo portable de pequeño tamaño. Además, se propone un método para ayudar a encontrar objetos y personas por medio de una red neuronal de segmentación semántica. Se han seleccionado diferentes objetos a resaltar con el fin de facilitar tareas como el reconocimiento de objetos, estancias o navegación. Como trabajo futuro, queda pendiente la realización de experimentos con personas en los que se compruebe la eficacia de usar la segmentación de objetos, así como un estudio de qué objetos es más importante resaltar dependiendo de la tarea a realizar.

Agradecimientos

Este trabajo ha sido realizado gracias a los proyectos de investigación JIUZ2022-IAR-05 y PID2021-125209OB-I00 (MCIN/AEI/10.13039/501100011033 and FEDER/UE) y financiado por el Gobierno de Aragón. Convocatoria Predoctorales 2022-2026.

Referencias

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39 (12), 2481–2495.

Barnes, N., 2013. An overview of vision processing in implantable prosthetic vision. In: 2013 IEEE International Conference on Image Processing. IEEE, pp. 1532–1535.

Beyeler, M., Boynton, G. M., Fine, I., Rokem, A., 2017a. pulse2percept: A python-based simulation framework for bionic vision. BioRxiv, 148015.

Beyeler, M., Nanduri, D., Weiland, J. D., Rokem, A., Boynton, G. M., Fine, I., 2019. A model of ganglion axon pathways accounts for percepts elicited by retinal implants. Scientific Reports 9 (1), 1–16.

Beyeler, M., Rokem, A., Boynton, G. M., Fine, I., 2017b. Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies. Journal of neural engineering 14 (5), 051003.

Brindley, G. S., Lewin, W. S., 1968. The sensations produced by electrical stimulation of the visual cortex. The Journal of physiology 196 (2), 479–493.

Cha, K., Horch, K. W., Normann, R. A., 1992. Mobility performance with a pixelized vision system. Vision research 32 (7), 1367–1372.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40 (4), 834–848.

Chen, S. C., Suaning, G. J., Morley, J. W., Lovell, N. H., 2009a. Simulating prosthetic vision: I. visual models of phosphenes. Vision research 49 (12), 1493–1506.

Chen, S. C., Suaning, G. J., Morley, J. W., Lovell, N. H., 2009b. Simulating prosthetic vision: Ii. measuring functional capacity. Vision research 49 (19), 2329–2343.

Dagnelie, G., 2006. Visual prosthetics 2006: assessment and expectations. Expert review of medical devices 3 (3), 315–325.

Dagnelie, G., Barnett, D., Humayun, M. S., Thompson, R. W., 2006. Paragraph text reading using a pixelized prosthetic vision simulator: parameter dependence and task learning in free-viewing conditions. Investigative ophthalmology & visual science 47 (3), 1241–1250.

Dagnelie, G., Keane, P., Narla, V., Yang, L., Weiland, J., Humayun, M., 2007. Real and virtual mobility performance in simulated prosthetic vision. Journal of neural engineering 4 (1), S92.

Feng, D., You, S., Barnes, N., 2017. Dsd: depth structural descriptor for edge-based assistive navigation. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 1536–1544.

Granley, J., Beyeler, M., 2021. A computational model of phosphene appearance for epiretinal prostheses. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, pp. 4477–4481.

Hayes, J. S., Yin, V. T., Piyathaisere, D., Weiland, J. D., Humayun, M. S., Dagnelie, G., 2003. Visually guided performance of simple tasks using simulated prosthetic vision. Artificial Organs 27 (11), 1016–1028.

Horsager, A., Greenwald, S. H., Weiland, J. D., Humayun, M. S., Greenberg, R. J., McMahon, M. J., Boynton, G. M., Fine, I., 2009. Predicting visual sensitivity in retinal prosthesis patients. Investigative ophthalmology & visual science 50 (4), 1483–1491.

Humayun, M. S., De Juan, E., Dagnelie, G., Greenberg, R. J., Propst, R. H., Phillips, D. H., 1996. Visual perception elicited by electrical stimulation of retina in blind humans. Archives of ophthalmology 114 (1), 40–46.

Humayun, M. S., Dorn, J. D., Da Cruz, L., Dagnelie, G., Sahel, J.-A., Stanga, P. E., Cideciyan, A. V., Duncan, J. L., Eliott, D., Filley, E., et al., 2012. Interim results from the international trial of second sight's visual prosthesis. Ophthalmology 119 (4), 779–788.

Jones, B., Pfeiffer, R., Ferrell, W., Watt, C., Marmor, M., Marc, R., 2016. Retinal remodeling in human retinitis pigmentosa. Experimental eye research 150, 149–165.

Kasowski, J., Beyeler, M., 2022. Immersive virtual reality simulations of bionic vision. In: Augmented Humans 2022. pp. 82–93.

Leo, M., Furnari, A., Medioni, G. G., Trivedi, M., Farinella, G. M., 2018. Deep learning for assistive computer vision. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0.

Lui, W. L. D., Browne, D., Kleeman, L., Drummond, T., Li, W. H., 2012. Transformative reality: improving bionic vision with robotic sensing. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 304–307.

Marc, R. E., Jones, B., Watt, C., Vazquez-Chona, F., Vaughan, D., Organisciak, D., 2008. Extreme retinal remodeling triggered by light damage: implications for age related macular degeneration. Molecular vision 14, 782.

Meffin, H., 2013. What limits spatial perception with retinal implants? In: 2013 IEEE International Conference on Image Processing. IEEE, pp. 1545–1549.

Nanduri, D., Fine, I., Horsager, A., Boynton, G. M., Humayun, M. S., Greenberg, R. J., Weiland, J. D., 2012. Frequency and amplitude modulation have different effects on the percepts elicited by retinal stimulation. Investigative ophthalmology & visual science 53 (1), 205–214.

Nanduri, D., Humayun, M., Greenberg, R., McMahon, M., Weiland, J., 2008. Retinal prosthesis phosphene shape analysis. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 1785–1788.

Perez-Yus, A., Santos-Villafranca, M., Bermudez-Cameo, J., Montano-Olivan, L., Lopez-Nicolas, G., Guerrero, J. J., 2023. Raspv: A robotics framework for augmented simulated prosthetic vision. Tech report.

ROS, 2022. Ros. `https://www.ros.org/`, accessed: 2023-05-23.

Rothermel, A., Liu, L., Aryan, N. P., Fischer, M., Wuenschmann, J., Kibbel, S., Harscher, A., 2008. A cmos chip with active pixel array and specific test features for subretinal implantation. IEEE Journal of Solid-State Circuits 44 (1), 290–300.

Sanchez-Garcia, M., Perez-Yus, A., Martinez-Cantin, R., Guerrero, J. J., 2021. Augmented reality navigation system for visual prosthesis. arXiv preprint arXiv:2109.14957.

Sommerhalder, J., Rappaz, B., de Haller, R., Fornos, A. P., Safran, A. B., Pelizzone, M., 2004. Simulation of artificial vision: Ii. eccentric reading of full-page text and the learning of this task. Vision research 44 (14), 1693–1706.

Steinmetz, J. D., Bourne, R. R., Briant, P. S., Flaxman, S. R., Taylor, H. R., Jonas, J. B., Abdoli, A. A., Abrha, W. A., Abualhasan, A., Abu-Gharbieh, E. G., et al., 2021. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study. The Lancet Global Health 9 (2), e144–e160.

Stronks, H. C., Dagnelie, G., 2014. The functional performance of the argus ii retinal prosthesis. Expert review of medical devices 11 (1), 23–30.

Sweethome3D, 2022. Sweethome3d. `http://www.sweethome3d.com/es/`, accessed: 2023-05-23.

Thompson, R. W., Barnett, G. D., Humayun, M. S., Dagnelie, G., 2003. Facial recognition using simulated prosthetic pixelized vision. Investigative ophthalmology & visual science 44 (11), 5035–5042.

van Rheede, J. J., Kennard, C., Hicks, S. L., 2010. Simulating prosthetic vision: Optimizing the information content of a limited visual display. Journal of vision 10 (14), 32–32.

Veraart, C., Raftopoulos, C., Mortimer, J. T., Delbeke, J., Pins, D., Michaux, G., Vanlierde, A., Parrini, S., Wanet-Defalque, M.-C., 1998. Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode. Brain research 813 (1), 181–186.

WHO, 2022. Ceguera y discapacidad visual. `https://www.who.int/es/news-room/fact-sheets/detail/blindness-and-visual-impairment`, accessed: 2023-05-22.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

Zhao, Y., Geng, X., Li, Q., Jiang, G., Gu, Y., Lv, X., 2017. Recognition of a virtual scene via simulated prosthetic vision. Frontiers in bioengineering and biotechnology 5, 58.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A., 2017. Scene parsing through ade20k dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 633–641.

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A., 2019. Semantic understanding of scenes through the ade20k dataset. International Journal of Computer Vision 127, 302–321.

Zrenner, E., Bartz-Schmidt, K. U., Benav, H., Besch, D., Bruckmann, A., Gabel, V.-P., Gekeler, F., Greppmaier, U., Harscher, A., Kibbel, S., et al., 2011. Subretinal electronic chips allow blind patients to read letters and combine them to words. Proceedings of the Royal Society B: Biological Sciences 278 (1711), 1489–1497.