

COGNITIVE AID SYSTEM FOR BLIND PEOPLE (CASBLiP)

Víctor M. Santiago Praderas, Nuria Ortigosa, Larisa Dunai, Guillermo Peris-Fajarnés

Centro de Investigación en Tecnologías Gráficas / Universidad Politécnica de Valencia
Valencia / España
vicsanpr@upv.es

CASBLiP nace con el objetivo de diseñar una herramienta tecnológica que permita a las personas invidentes la integración en la vida social mejorando su calidad de vida y aportándoles herramientas adicionales a las que ya poseen, bastón, perros, etc. para mejorar su movilidad. El 1,9 % de la población europea tiene discapacidad visual, total o parcial, motivo por el cual la Universidad Politécnica de Valencia y sus socios decidieron abordar la investigación, mezclando para ello metodologías basadas en visión artificial, mecanismos de análisis acústicos, sistemas de posicionamiento GPS/HPS y sensores para la adquisición de información 2D. Antes de iniciar el desarrollo se realizó un estudio de mercado, llegando a revisar una decena de proyectos con antecedentes similares a los del proyecto CASBLiP. Se vio que nuestra propuesta podía mejorar las prestaciones e incluso incrementar la funcionalidad de los proyectos originales. La conclusión a la que se ha llegado tras 3 años de investigación, implementación e integración, ha sido que, gracias al gran desarrollo cognitivo de estas personas, junto con el interés y motivación que tienen por mejorar su estilo de vida, es posible ayudarles mediante herramientas tecnológicas hasta llegar a conseguir una integración, casi total, en la sociedad moderna. Todas las pruebas se han realizado contando con el apoyo de personas invidentes de diferentes perfiles y edad, mezclando distintos escenarios y usando técnicas de representación 2D y 3D. El proyecto ha sido desarrollado gracias a fondos europeos y ha estado coordinado por el Centro de Investigación en Tecnologías Gráficas (Universidad Politécnica de Valencia), contando con un equipo de 6 socios europeos: Universidad Politécnica de Le Marche (Italia), Universidad de La Laguna (España), Universidad de Bristol (Inglaterra), SIEMENS (Alemania), Instituto Francesco Cavazza (Italia) y DBSV (Alemania). Como resultado de esta investigación, merece la pena destacar la creación de dos prototipos que, conjuntamente con sistemas GPS, permiten a personas invidentes moverse tanto en espacios abiertos como en espacios cerrados, detectando la presencia de objetos móviles y permitiéndoles diferenciar entre espacios abiertos (pasillos por donde puedan circular), y objetos estáticos a distancias comprendidas entre 0,5 y 15 metros.

Invidente, ciego, acústica, visión, 2D, 3D, tecnología, integración social, herramientas, estéreo visión, hrtf, mapas de profundidad, GPS, ayudas integración, investigación, gráficas, visión artificial, tratamiento de imágenes, sonificación.

CASBLiP born with the goal of designing a technological tool to let the blind people the integration into social life by improving their quality of life and providing additional tools besides already they have (cane, dogs, etc.) for better mobility. 1.9% of the European population is visually impaired, in whole or in part, is because the Polytechnic University of Valencia and its partners decided to research in this field, mixing methodologies based on vision, mechanisms for analysis acoustic, system GPS / HPS and sensors for the acquisition of 2D information. Before starting the development of the project, a state of the art study was carried out a, reviewing at least 10 projects with a history similar to the draft CASBLiP. We notice that our proposal could improve benefits and even increase the functionality of the original projects. After 3 years of implementation, integration and research, the conclusion, thanks largely to the cognitive development of

these people, along with the interest and motivation they have to improve their lifestyle, we can help them by reaching technological tools to achieve integration in modern society. All tests have been conducted with the support of blind people with different social characteristics and ages, mixing different scenarios and using representation techniques of 2D and 3D. The project has been developed thanks to European budget funds and has been coordinated by the Research Center in Graphic Technology (Polytechnic University of Valencia), with a team of 6 European partners: University of Le Marche (Italy), University of La Laguna (Spain), University of Bristol (England), Siemens (Germany), Francesco Cavazza Institute (Italy) and DBSV (Germany). As a result of this investigation, it should be emphasized the creation of two prototypes, which in conjunction with GPS systems, allow blind people to move in open spaces or in indoor, detecting the presence of moving objects and allowing them to differentiate between open spaces (free paths where they can move), and static objects between distances from 0.5 to 15 meters.

Blind, sound, vision, 2D, 3D, technology, social integration, tools, stereo vision, hrtf, depth maps, GPS, integration aid, research, graphics, artificial vision, image processing, sonification.

1. OBJETIVOS DEL PROYECTO

El principal objetivo del proyecto ha sido, desde el inicio del mismo, desarrollar un sistema capaz de gestionar e interpretar el mundo real y la información procedente de diferentes fuentes para ayudar a personas ciegas, o con deficiencias visuales, ofreciéndoles mapas de audio como traducción de escenas reales, con el único fin social de mejorar de manera significativa la independencia y la calidad de vida de los ciegos y deficientes visuales.

Para la consecución de los objetivos generales del proyecto, y dada la importancia de la continuidad de líneas de investigación en este terreno, es importante diferenciar entre objetivos científicos y objetivos tecnológicos.

1.1. Objetivos Científicos

Los objetivos científicos que se han pretendido conseguir durante el desarrollo del proyecto van dirigidos especialmente al incremento del conocimiento en el campo cognitivo de las personas ciegas para, a corto medio plazo, llegar a entender como funciona su percepción del mundo real y desarrollar herramientas de apoyo que les ayude a su integración y desenvolvimiento en la vida social, intentado aprovechar todos los servicios que la sociedad pone a su disposición. Para ello, se ha investigado para desarrollar un sistema cognitivo que pueda interpretar la información obtenida del mundo real y transformarla en representación acústica.

Por otra parte, se han llevado a cabo estudios sobre la capacidad de percepción del espacio circundante de las personas ciegas, y así poder definir la cantidad mínima de información necesaria a transmitir para su movilidad.

Por todo ello, ha sido necesario desarrollar algoritmos de procesamiento de imágenes en tiempo real para la transformación de los entornos habituales al aire libre y calcular la distancia desde el sistema de adquisición de estas áreas.

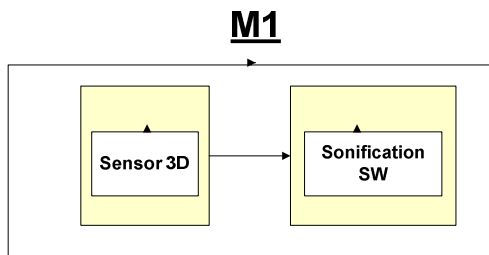
Todo lo anterior no hubiera tenido sentido si no se crea, en tiempo real, un modelo 3D a partir de la combinación de toda la información procesada mediante los diferentes dispositivos de captura de datos, con el fin de diseñar sistemas inteligentes para determinar los posibles riesgos y los obstáculos para una persona ciega en movimiento para, finalmente, crear una interfaz de usuario final con las principales características de ser capaz de representar espacios reales utilizando información acústica [2].

1.2. Objetivos Tecnológicos

Utilizar mecanismos de tratamiento de imagen, desarrollados por la Universidad de Bristol y por la Universidad Politécnica de Valencia, y sistemas de representación del espacio en audio, por la Universidad de La Laguna, con el fin de desarrollar un prototipo y el propósito de ayudar a las personas con discapacidad visual en la movilidad en espacios al aire libre.

2. PROTOTIPOS GENERADOS DURANTE EL DESARROLLO DEL PROYECTO

Para conseguir los objetivos mencionados, así como para generar un producto tangible y real que pueda demostrar el resultado de todo el desarrollo del proyecto, se ha dividido en dos fases la fabricación e implementación del mismo. En primer lugar se diseñó un primer prototipo aislado denominado M1.



Este primer prototipo, está formado por dos procesos; la captura de información procedente del mundo real y la transformación de dicha información en sonidos interpretables por los usuarios finales, personas ciegas en nuestro caso.

La información procedente del mundo real es capturada mediante un sensor de imagen CMOS (Complementary Metal Oxide Semiconductor), capaz de adquirir información de la distancia de los objetos situados en un rango de 0 a 5 metros.

En comparación con los métodos tradicionales de sensores, el sensor CMOS 3D propuesto combina una alta percepción de sensibilidad, con bajos costes de producción y construcción y una integración hardware de pequeño tamaño.

Esta información, básicamente distancias de los objetos situados en el punto de mira del hardware CMOS, ha de ser transformada en mapas acústicos. Para ello es enviada a una FPGA (Field Programmable Gate Array) con el software adecuado para finalmente, transformar y enviar la información a las personas invidentes.

El sistema final está compuesto por unas gafas, las cuales incluyen toda la electrónica láser necesaria para la captura de la información de las distancias, el ensamblado de los circuitos que componen la FPGA encargada de sonificar la información procedente del láser y unos auriculares, además de la batería para alimentar la

circuitería y demás cableado (ver figura 2.)



Fig. 2. Gafas M1.



Fig.3. FPGA M1.

El subsistema acústico es el encargado de reproducir sonidos espacializados, es decir, aquellos que serán escuchados por el usuario a través de los auriculares pero percibidos como si estuviesen situados en los objetos del entorno. Los sonidos han sido procesados previamente fuera de tiempo real de manera que se perciban situados en el entorno exterior al sujeto. Posteriormente se almacenan en la memoria del subsistema de sonidos. El subsistema selecciona en cada momento los sonidos que correspondan, los suma y reproduce el resultado por los auriculares [1],[2],[3], [4].

La segunda parte del proyecto ha consistido en desarrollar un segundo prototipo denominado M2.

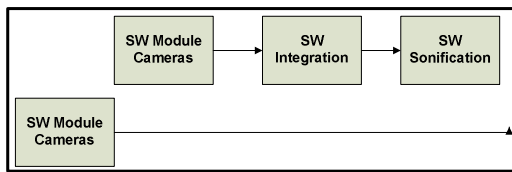


Fig. 4. Prototipo M2.

A diferencia de M1, y aunque el objetivo final es el mismo, es decir, capturar información del mundo real y sonificarla, ahora la información es capturada mediante cámaras de estereovisión y con formato 3D. El sistema segmenta la imagen en diferentes áreas de interés, basándose concretamente en la detección de un objeto (en movimiento con prioridad sobre los estáticos, o el más cercano en caso de varios objetos estáticos) y de pasillos o áreas libre de paso, por las cuales el invidente puede circular sin peligro alguno. El sistema funciona en tiempo real (10 fr / seg) por lo que se puede utilizar mientras se camina por la calle a ritmo normal.

Como se puede observar en la figura 4, el primer paso ha consistido en desarrollar mecanismos de tratamiento de imagen para, mediante la obtención de diferentes mapas de profundidad reconocer objetos estáticos y en movimiento, además de pasillos libres. Tras la obtención de los mapas de profundidad de la escena, los diferentes niveles de gris son indicativos de las diferentes distancias a los objetos de la misma [1].

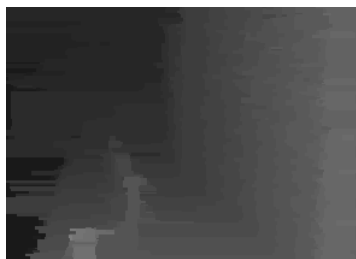


Fig. 5. Mapa de Profundidad.

Tras este análisis de distancias, se localiza el objeto móvil más próximo a la posición de las cámaras, o bien la de aquel que se esté moviendo a mayor velocidad hacia éstas, al igual que se aplican diferentes algoritmos

para la detección de vías libres de obstáculos mediante el análisis de los mapas de profundidad.

Toda esta información ha sido integrada, mediante lenguajes de alto nivel, en un interfaz para mediante protocolos bien definidos traspasar la información al módulo encargado de sonificar las imágenes. Este módulo trabaja de manera similar al descrito en el prototipo M1, pero con diferentes bancos de sonido y para distancias comprendidas entre 5 y 15 metros.

Paralelamente existe un sistema basado en localización GPS para, de forma independiente al prototipo M2, ofrecer al usuario información de su situación y ruta a seguir para llegar a un punto concreto.

3. PROTOTIPO DEFINITIVO

Por último, y como prototipo final resultado de toda la investigación llevada a cabo durante estos últimos 3 años (2005 – 2008), se ha llegado a integrar y fabricar un prototipo formado por la unión de los dos prototipos, M1 y M2, comentados en el apartado anterior.

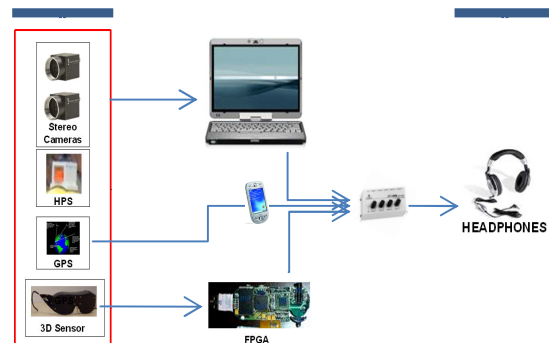


Fig. 6. Elementos del prototipo final.

Finalmente se ha centralizado todo el proceso en un PC portátil con características estándares que actúa como central de proceso. El sistema está formado por: 1 ordenador portátil, 2 cámaras de visión, 1 PDA, 1 sensor 3D CMOS, 1 FPGA, 1 mezclador analógico y auriculares.

Como entradas al sistema tenemos las 2 cámaras de estereo visión, encargadas de capturar las imágenes responsables del prototipo M2. Estas imágenes, en tiempo real son transmitidas al PC, el cual, por cada trama capturada de la imagen real extrae los mapas de profundidad, los procesa para

extraer los objetos finales y aplica los algoritmos necesarios para extraer los pasillos libres. Todos estos datos son almacenados y, mediante protocolos de comunicación son enviados al módulo de sonido para, finalmente ser sonificada por el estrategia de sonido y enviada a los auriculares a través de la tarjeta de sonido del PC.

Al mismo tiempo, y en paralelo con la captura de imágenes, existe un sistema GPS integrado en una PDA la cual funciona de manera independiente, y conectada al mezclador emite la información sonora necesaria para guiar al usuario a su destino final.

Finalmente, el sensor 3D CMOS captura información sobre los objetos próximos y la envía a la FPGA, la cual conectada al banco de sonidos del PC genera información sonora para ser enviada finalmente a los usuarios a través de los auriculares.

Se puede observar que la información que el usuario final va a disponer finalmente va a ser la mezcla de M1, M2 y GPS, o podrá discriminar cualquiera de las tres mediante el mezclador analógico integrado en el sistema.

Resumiendo, el prototipo que actualmente existe operativo, y que esperamos sea mejorado y optimizado en futuros proyectos, dispone de las siguientes características:

- De 0 a 15 metros el sistema es capaz de capturar información real y transformarla a mapas acústicos de un objeto simultáneamente.
- De 5 a 15 metros el sistema puede interpretar información de pasillos libres.
- Toda esta información es sonificada y enviada a los auriculares que finalmente se coloca el usuario.
- Para instalar y ejecutar todo el proceso sólo es necesario un PC portátil y una PDA.



Fig. 7. Prototipo M1 y M2.

4. VALIDACIÓN DEL SISTEMA Y CONCLUSIONES

Con el fin de verificar los resultados de la investigación e integración del prototipo final, se reunió a un grupo de usuarios finales y se realizaron diferentes ejercicios.

Las pruebas básicas de aprendizaje efectuadas basadas en BLP (Basic Learning Protocol) con usuarios finales respecto al prototipo M1, estuvieron preparadas para trabajar con siete ejercicios diferentes y con el fin de localizar distintos tipos de objetos para medir sus anchuras y alturas correctamente mientras caminaban a través de ellos sin golpearlos.

Todos los sujetos fueron capaces de identificar correctamente las dimensiones de los objetos en un corto espacio de tiempo y no tuvieron excesiva dificultad para discernir entre la altura y la anchura de los objetos, incluso en dar la vuelta y seguir el camino andado en orden inverso.

Si que hubo diferencias sin embargo en lo que se refiere al tiempo. El mínimo estuvo en 3:53 minutos y el máximo en 10:35 minutos. Esto puede estar vinculado al nivel de experiencia y de entrenamiento que se reciba, ya que el que mejor resultado obtenido correspondía con el mas experimentado.

Respecto a las pruebas del prototipo M2 y siguiendo las normas de la percepción angular, 60 grados de ancho y 20 m de largo, se realizaron diferentes ejercicios basados en 18 posiciones diferentes. Los sujetos fueron situados en el vértice del ángulo de percepción y se les pidió identificar los cambios en las posiciones que una persona realizaba en el campo de visión. El orden de los cambios que se iban produciendo no estaba pactado y se iba realizando al azar.

Antes de realizar las pruebas referentes a M2, se planifico y se explico con detenimiento la manera en que se iba a trabajar. Las pruebas estaban basadas en posicionar al usuario en una posición fija y sin mover la cabeza. Los resultados fueron bastante prometedores. No se cometieron errores y todos los sujetos fueron capaces de identificar rápidamente y con facilidad la evolución de las posiciones de las personas involucradas en el test, además de indicar los cambios en los movimientos que se estaban realizando.

Sin embargo, fue bastante evidente que los resultados no hubieran sido los mismos en una situación dinámica con el usuario en movimiento o con varias personas participando en el plano de visión.

5. REFERENCIAS

[1] Carlos Merino and Majid Mirmehdi *A framework towards real-time detection and tracking of text. Second International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2007), pages 10--17. September 2007.*

[2] IGonzalez-Mora, J.L., IRodriguez-Hernandez, A.F., Burunat E., Martin F., Castellano M.A. *Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people. Department of Physiology, University of La Laguna.*

[3] Antonio F. Rodríguez Hernández. *Figure perception from real and virtual sounding surfaces* Departamento de fisiología, Universidad de La Laguna, España.

[4] Larisa Dunai, Guillermo Peris Fajarnes, Beatriz Defez Garcia, Nuria Ortigosa Araque, Fernando Brusola Simon, *Perception of the sound source position, Acoustical Physics journal in volume 55, 3rd issue, 2009 ISSN PRINT: 1063-7710.*