



Modelos y Herramientas para Sistemas de Rehabilitación Visual

Christian A. Morillas Gutiérrez

TESIS DOCTORAL

Directores: Francisco José Pelayo Valle
Alberto Prieto Espinosa

Granada 2006



Dpto. Arquitectura y Tecnología de
Computadores

D. Francisco José Pelayo Valle, Catedrático de la
Universidad de Granada, y D. Alberto Prieto Espinosa,
Catedrático de la Universidad de Granada,

CERTIFICAN

Que la memoria titulada

**Modelos y Herramientas para Sistemas
de rehabilitación Visual**

ha sido realizada por D. Christian A. Morillas Gutiérrez bajo
nuestra dirección en el Departamento de Arquitectura y
Tecnología de Computadores de la Universidad de Granada,
para optar al grado de Doctor.

Granada, a 25 de Julio de 2006

Fdo. Francisco José Pelayo Valle

Fdo. Alberto Prieto Espinosa



ugr

Universidad
de **Granada**



Modelos y Herramientas para Sistemas de Rehabilitación Visual

Memoria presentada por

Christian A. Morillas Gutiérrez

Para optar al grado de

**Doctor en Informática por
la Universidad de Granada**

Fdo: Christian A. Morillas Gutiérrez

*A toda mi familia
y a Loli*

*“Science may set limits to knowledge,
but should not set limits to imagination”*

*“La ciencia puede poner límites al conocimiento,
pero no debería ponerlos a la imaginación”*

Bertrand Russell (1872-1970)

Agradecimientos

Mi agradecimiento se extiende a todas aquéllas personas que con su ayuda, apoyo y muchas veces comprensión, paciencia y esfuerzo, han hecho posible mi dedicación a este trabajo, y la realización del mismo.

En el plano personal, en primer lugar quiero dar muy especialmente las gracias a mis padres por su ilusión vertida en este proyecto. Sólo gracias a ellos ha sido posible todo el proceso hasta llegar aquí. Tampoco puedo olvidarme de mis hermanos, gracias a su ánimo nunca decayó mi empeño. Y por supuesto, a Loli, por todo el tiempo que hemos tenido que sacrificar para que este trabajo llegara a buen puerto, por estar a mi lado apoyándome en todo momento. Muchísimas gracias a todos, extendiéndolas a los demás familiares y amigos que siempre encontraron una palabra de aliento.

Muchas personas son las que profesionalmente han dejado su 'marca' en este trabajo, y no quiero comenzar sin disculparme ante quienes no distingua nominalmente. Quiero agradecer a Francisco Pelayo por su empeño, su dedicación e ilusión que ha depositado en este trabajo, así como en todos en los que me ha guiado. Gracias por encontrar siempre ese punto de vista, ese aspecto, ese razonamiento que me ha permitido ver las cosas de una forma más clara. Tampoco puedo olvidar a Alberto Prieto, que con su inestimable desempeño ha guiado a un variado y extenso grupo hacia objetivos científicos comunes. Desearía extender mi gratitud hacia el resto del equipo del proyecto CORTIVIS en Granada, Samuel y Antonio, por todos los momentos vividos y los trabajos *sufridos*. No puedo olvidarme del tesón que siempre ha mostrado Eduardo Ros por su trabajo, así como las infinitas *discusiones* mantenidas con el equipo de los demás proyectos europeos: Javi, Sonia, Guille, Eva, Richard, Rodrigo y Rafa, gracias a todos.

Finalmente quiero dar las gracias a todo el departamento de Arquitectura y Tecnología de Computadores de la Universidad de Granada por acogerme durante todo este tiempo y contar conmigo en innumerables ocasiones. Muchísimas gracias a todos.

Este trabajo ha contado con la financiación del proyecto europeo CORTIVIS (QLK-CT-2001-00279) y el nacional DEPROVI (DPI2004-07032).

Resumen

Esta tesis doctoral versa sobre el procesamiento de información visual destinado a la obtención de sistemas de ayuda para invidentes, bien basados en neuroestimulación mediante prótesis visuales, bien basados en la transducción sensorial de información visual relevante.

Aporta soluciones en el campo del modelado y simulación de funciones de procesamiento bioinspirado de información visual, así como implementaciones hardware orientadas a la obtención de sistemas empotrados basados en FPGA, portátiles, configurables para cada usuario, y procesando la información visual procedente de una o dos cámaras en tiempo real.

Los resultados concretos obtenidos se pueden categorizar en los tres grupos siguientes:

a) Diseño de módulos software para el modelado de los sistemas de procesamiento, que han conducido al desarrollo de un entorno completo para definición, simulación, análisis y comparación de resultados de este entorno con experimentos biológicos.

b) Diseño mediante un lenguaje de descripción de hardware de alto nivel de los diferentes módulos de procesamiento que intervienen en el sistema global.

c) Diseño e implementación en chips reconfigurables (FPGA) de diversas soluciones completas para ayuda a invidentes, incluyendo un sistema de generación de eventos relativos a estimulación nerviosa, y un nuevo dispositivo electrónico-informático de transducción sensorial que genera patrones sonoros que codifican información espacial.

Abstract

This dissertation deals with visual information processing intended for the implementation of visual aid systems for the visually impaired, whether they are based on neurostimulation by means of visual prostheses or on sensorial transduction of significant visual information.

This thesis contributes with solutions in the field of modeling and simulation of bioinspired visual information processing functions, as well as hardware implementations oriented to obtain FPGA-based embedded systems, that are portable, customizable for every user, and being able to process visual information coming from one or two cameras, in real time.

Particular results obtained can be classified into the three following categories:

a) Design of software modules to model processing systems, which have led to the development of a complete environment for the definition, simulation, analysis and comparison of the results of this platform to those obtained from biological experiments.

b) Design of the different processing modules arranging the global system, using a high level hardware description language.

c) Reconfigurable circuitry (FPGA) based design and implementation of several complete visual aids for the blind, including an event generation system for neural stimulation, and a novel electronic-computing device for sensorial transduction that delivers sound patterns encoding spatial information.

Índice General

AGRADECIMIENTOS	I
RESUMEN.....	III
ABSTRACT	V
ÍNDICE GENERAL.....	VII
ÍNDICE DE FIGURAS	XI
ÍNDICE DE TABLAS	XVII
CAPÍTULO 1: INTRODUCCIÓN.....	1
1.1 PLANTEAMIENTO DEL PROBLEMA	1
1.2 ANTECEDENTES	3
1.3 ORGANIZACIÓN DE LA MEMORIA	6
CAPÍTULO 2: VISIÓN Y REHABILITACIÓN VISUAL.....	9
2.1 DESCRIPCIÓN DEL SISTEMA VISUAL HUMANO.....	10
2.1.1 <i>El ojo y la retina</i>	10
2.1.1.1 La retina y su función	12
2.1.1.2 Las células bipolares	15
2.1.1.3 Las células horizontales.....	16
2.1.1.4 Las células amacrinas.....	17
2.1.1.5 Las células interplexiformes	17
2.1.1.6 Las células ganglionares.....	18
2.1.2 <i>La visión en color</i>	21
2.1.3 <i>El núcleo geniculado lateral (LGN)</i>	25
2.1.4 <i>La corteza visual</i>	26
2.1.4.1 Arquitectura de la corteza visual.....	27
2.1.4.2 Las células de V1.....	29
2.1.4.3 Organización funcional de la corteza.....	32
2.1.5 <i>Visión estéreo</i>	34

2.1.5.1	Neuronas sensibles a la disparidad.....	37
2.2	PATOLOGÍAS QUE AFECTAN AL SISTEMA VISUAL.....	39
2.2.1	<i>Trastornos de la retina</i>.....	40
2.2.1.1	Degeneración macular.....	41
2.2.1.2	Desprendimiento de retina.....	42
2.2.1.3	Retinosis pigmentaria.....	43
2.2.2	<i>Glaucoma</i>.....	43
2.2.2.1	Glaucoma de ángulo abierto.....	45
2.2.2.2	Glaucoma de ángulo cerrado.....	45
2.2.3	<i>Incidencia en la población</i>.....	46
2.3	SISTEMAS PARA LA REHABILITACIÓN VISUAL.....	46
2.3.1	<i>Neuroprótesis visuales</i>.....	47
2.3.1.1	Prótesis retinianas.....	48
2.3.1.2	Prótesis para el nervio óptico.....	50
2.3.1.3	Prótesis corticales.....	51
2.3.2	<i>Sistemas electrónico – informáticos</i>.....	57
2.4	CONCLUSIONES.....	63
CAPÍTULO 3: PROCESAMIENTO DE INFORMACIÓN VISUAL: RETINER.....		65
3.1	INTRODUCCIÓN: EL PROYECTO CORTIVIS.....	66
3.2	ESTRUCTURA DE LA PRÓTESIS VISUAL.....	68
3.3	ENTORNO DE MODELADO Y SIMULACIÓN SOFTWARE: RETINER.....	72
3.3.1	<i>Respuesta temporal central y periférica</i>.....	74
3.3.2	<i>Determinación de los campos receptivos</i>.....	75
3.3.3	<i>Codificación neuromórfica</i>.....	78
3.3.4	<i>Salida visual</i>.....	80
3.3.5	<i>Simulación temporal</i>.....	82
3.3.6	<i>Visualización de spikes. Registros biológicos</i>.....	85
3.3.7	<i>Extensiones de Retiner</i>.....	85
3.4	CONCLUSIONES.....	89
CAPÍTULO 4: MODELO DE PROCESAMIENTO DE VISIÓN ESTÉREO.....		93
4.1	VISIÓN BINOCULAR Y ESTEREOVISIÓN.....	94
4.1.1	<i>Modelo de cámara. Calibración</i>.....	94
4.1.1.1	Planteamiento general.....	96
4.1.1.2	Método de la matriz de perspectiva.....	97
4.1.1.3	Comparación de métodos de calibración.....	102

4.1.1.4	Calibración débil.....	103
4.1.2	<i>Geometría de un sistema estéreo</i>	103
4.1.2.1	Geometría epipolar.....	106
4.2	MÉTODOS DE CÁLCULO DE CORRESPONDENCIAS O MATCHING	108
4.2.1	<i>Métodos basados en características</i>	109
4.2.2	<i>Métodos basados en área</i>	111
4.2.3	<i>Algoritmos clásicos</i>	111
4.2.3.1	Marr – Poggio.....	111
4.2.3.2	Marr – Poggio – Grimson.....	112
4.2.3.3	PMF (Pollard – Mayhew – Frisby).....	113
4.2.3.4	Nishihara.....	114
4.2.4	<i>Métodos de correlación</i>	114
4.2.4.1	Funciones de similitud.....	116
4.2.4.2	Selección del tamaño de la ventana de correlación.....	117
4.2.4.3	Aproximaciones jerárquicas. Matching multietapa.....	119
4.2.5	<i>Método stretch-correlation para estereovisión</i>	121
4.2.5.1	Resultados.....	122
4.2.6	<i>Método del INRIA para la obtención de mapas densos de profundidad</i>	123
4.2.6.1	Preprocesamiento. Calibración y rectificación.....	123
4.2.6.2	Criterios de correlación.....	124
4.2.6.3	Algoritmo de correlación. Validación de las correspondencias.....	124
4.2.6.4	Aproximación jerárquica.....	125
4.2.6.5	Implementación del algoritmo.....	125
4.2.7	<i>Algoritmo SSSD para múltiples líneas base</i>	126
4.2.7.1	Descripción del método.....	126
4.2.7.2	Una aplicación práctica del algoritmo SSSD.....	128
4.2.8	<i>Otros métodos de matching</i>	130
4.2.8.1	Estéreo usando razones fotométricas.....	130
4.3	UTILIDAD Y USO DE LA INFORMACIÓN DE PROFUNDIDAD	132
4.4	CONCLUSIONES	137
CAPÍTULO 5: SISTEMA DE GENERACIÓN DE SONIDOS 3D		139
5.1	LOCALIZACIÓN ESPACIAL DE SONIDOS	140
5.2	CÁLCULO DE LOS INDICIOS DE LOCALIZACIÓN	143
5.2.1	<i>Procedimiento de medida</i>	145
5.3	GENERACIÓN DE SONIDOS 3D DESDE RETINER	148
5.4	CONCLUSIONES	152

CAPÍTULO 6: SISTEMAS PARA REHABILITACIÓN VISUAL INTEGRADOS EN FPGA.....	153
6.1 IMPLEMENTACIÓN HARDWARE DE UN SISTEMA DE PROCESAMIENTO VISUAL BIOINSPIRADO	154
6.1.1 Plataformas de diseño hardware	154
6.1.1.1 RC-100	155
6.1.1.2 RC-200	156
6.1.1.3 RC-300	157
6.1.2 Dispositivo de captura de imágenes	158
6.1.3 Diagrama de bloques	160
6.1.3.1 Captura de video o <i>frame-buffer</i>	160
6.1.3.2 Lectura de la memoria RAM	161
6.1.3.3 Filtrado espacial	162
6.1.3.4 Combinación de los filtros	164
6.1.3.5 Cómputo de la matriz de actividad y visualización.....	166
6.1.3.6 Suavizado gaussiano.....	167
6.1.4 Resultados de síntesis de los circuitos.....	168
6.2 RETINA ARTIFICIAL	169
6.3 VIS2SOUND: SISTEMA COMPLETO DE TRANSDUCCIÓN SENSORIAL.....	171
6.3.1 Subsistema de generación de sonidos 3D.....	172
6.3.1.1 Implementación en la RC-100.....	172
6.3.1.2 Implementación en la RC-200.....	178
6.3.2 Integración del sistema final.....	190
6.4 CONCLUSIONES	195
CAPÍTULO 7: CONCLUSIONES Y PRINCIPALES APORTACIONES	197
REFERENCIAS BIBLIOGRÁFICAS	205

Índice de figuras

Figura 2.1: Sección horizontal del globo ocular	11
Figura 2.2: Proyección del campo visual.....	12
Figura 2.3: Estructura en capas de la retina	14
Figura 2.4: Funcionamiento de conos y bastones	16
Figura 2.5: Respuesta de las células ganglionares	19
Figura 2.6: Respuesta de dos células X e Y	19
Figura 2.7: Tipos de conos según su respuesta	22
Figura 2.8: Distintos tipos de células: de banda ancha, con antagonismo al color y de entradas mixtas	25
Figura 2.9: Localización de la corteza visual	28
Figura 2.10: Tinción que muestra la línea de Gennari.....	28
Figura 2.11: Estructura en capas de la corteza visual.....	29
Figura 2.12: Experimento y respuesta de una célula de V1.....	31
Figura 2.13: Respuestas de una célula simple y una compleja.....	31
Figura 2.14: Respuestas de las células corticales según la velocidad y dirección del estímulo.....	32
Figura 2.15: Representación de las columnas en la corteza cerebral.....	34
Figura 2.16: Proyección en la retina de dos puntos del espacio visual	35
Figura 2.17: Círculo de Vieth-Müller, horóptero y área de Panum.....	36
Figura 2.18: Visión de pacientes que presentan degeneración macular	41
Figura 2.19: Visión con desprendimiento parcial de retina.....	42
Figura 2.20: Ojo afectado por retinosis pigmentaria	43
Figura 2.21: Campo visual afectado por glaucoma	44
Figura 2.22: Micro-fotodiodo y un array de varios.....	48

Índice de figuras	XII
Figura 2.23: Localización de un implante epi-retiniano.....	49
Figura 2.24: Prótesis para el nervio óptico.....	50
Figura 2.25: Matriz de electrodos planos de Dobelle	52
Figura 2.26: Esquema del sistema completo de Dobelle	53
Figura 2.27: Array de microelectrodos de Troyk	54
Figura 2.28: Esquema de los electrodos de Schmidt.....	56
Figura 2.29: Matriz de microelectrodos de Utah.....	56
Figura 2.30: Sistema de ayuda a la visión LVES	58
Figura 2.31: Dispositivo JORDY de Enhanced Vision.....	58
Figura 2.32: Dispositivo de visualización see-through	60
Figura 2.33: Esquema de generación de sonido de The vOICe.....	61
Figura 2.34: Soporte para las cámaras de EAV	62
Figura 2.35: División del campo visual realizado por EAV	63
Figura 3.1: Esquema de la prótesis visual de CORTIVIS.....	67
Figura 3.2: Estructura en bloques del sistema completo.....	68
Figura 3.3: Diferencias de una cámara lineal y una logarítmica	69
Figura 3.4: Distintas etapas en el procesamiento con Retiner	70
Figura 3.5: Proceso de envío de la información de estimulación.....	71
Figura 3.6: Interfaz de usuario de Retiner	73
Figura 3.7: Filtrado temporal	75
Figura 3.8: Herramienta para la definición de campos receptivos.....	77
Figura 3.9: Efecto de los campos receptivos de tamaño variable	78
Figura 3.10: Módulo de producción de impulsos	80
Figura 3.11: Montaje para experimentos psicofísicos.....	81
Figura 3.12: Aproximaciones a la percepción evocada	82
Figura 3.13: Configuración de los parámetros temporales.....	83
Figura 3.14: Resultados con distinta resolución temporal.....	84
Figura 3.15: Representación de estímulos con Retiner.....	86
Figura 3.16: Procesamiento estéreo con Retiner.....	87

Figura 3.17: Arquitectura de referencia para el modelado con Retiner	88
Figura 4.1: Modelo de cámara pin-hole.....	95
Figura 4.2: Sistema de referencia de la imagen	96
Figura 4.3: Geometría de la proyección.....	99
Figura 4.4: Sistema de visión estéreo simple	104
Figura 4.5: Geometría del sistema estéreo	105
Figura 4.6: Geometría epipolar de un sistema estéreo	106
Figura 4.7: Técnica de correlación.....	115
Figura 4.8: Tamaño de ventana y curva de correlación	118
Figura 4.9: Distorsión de bloques mediante stretch	122
Figura 4.10: Imágenes del proceso de calibración	133
Figura 4.11: Variantes para el cálculo de la profundidad.....	136
Figura 5.1: Indicios para la localización espacial de sonidos	141
Figura 5.2: Modelo de pinna y tragus.....	142
Figura 5.3: Maniquí KEMAR utilizado en el cálculo de las HRTF	144
Figura 5.4: Características antropométricas medidas.....	147
Figura 5.5: Ilustración de los filtros y su procesamiento	149
Figura 5.6: Módulo de generación de sonidos	150
Figura 6.1: Plataforma RC-100 de Celoxica.....	155
Figura 6.2: Detalle de todos los elementos incluidos en la tarjeta RC-200 .	157
Figura 6.3: Tarjeta RC-300	159
Figura 6.4: Cámara a color modelo C-Cam8 de Comedia	159
Figura 6.5: Diagrama de bloques de la implementación hardware del sistema de procesamiento visual bioinspirado	160
Figura 6.6: Máscara del operador diferencia de gaussianas.....	162
Figura 6.7: Arquitectura de cómputo de la operación de convolución.....	164
Figura 6.8: Esquema de cómputo de la matriz de actividad y de la visualización de los resultados obtenidos.....	167

Figura 6.9: Instantánea del procesamiento visual que realiza la plataforma hardware portable.....	167
Figura 6.10: Incorporación del suavizado gaussiano al procesamiento visual	168
Figura 6.11: Esquema de procesamiento para la generación de eventos de estimulación para un electrodo individual.....	170
Figura 6.12: Arquitectura para la generación de impulsos para diferentes electrodos, utilizando un módulo individual, memorias de almacenamiento y esquema de cómputo secuencial.....	171
Figura 6.13: Izquierda: Detalle del sistema inicial con la RC-100 de Celoxica. Derecha: esquema eléctrico que implementa la interfaz con las señales de audio, para un solo canal	172
Figura 6.14: Arriba: A la izquierda tenemos parte de la señal original a reproducir, y a la derecha los coeficientes de un filtro (muestreados con 8 bits). Abajo: Resultado de filtrar la señal original con el filtro mostrado ..	173
Figura 6.15: Representación del experimento implementado en la tercera versión de sistema.....	176
Figura 6.16: Representación de la señal seno (arriba) y el resultado de filtrarla con ambos canales de uno de los filtros (abajo), resultando una señal con el mismo periodo que la original	177
Figura 6.17: Estructura diseñada para almacenar las señales filtradas. Presenta dos niveles de indexación, el primero determina la elevación, y el segundo el azimut	179
Figura 6.18: Interfaz del entorno software para test del sistema.....	180
Figura 6.19: Estructura lógica en memoria para el almacenamiento de las señales.....	182
Figura 6.20: Representación de las distintas duraciones configurables.....	184
Figura 6.21: Detalle de la interfaz para configurar el volumen de los canales de audio	187

Figura 6.22: Entorno software para la conexión y transmisión de señales	188
Figura 6.23: Detalle del interfaz con una malla sin configurar	188
Figura 6.24: Dispositivo de captura de imágenes utilizado en VISION2SOUND. En la zona central podemos distinguir el sensor de ultrasonidos incorporado.....	191
Figura 6.25: Arquitectura de procesamiento de la información implementada en VISION2SOUND	192
Figura 6.26: Prototipo del sistema de transducción sensorial implementado en una plataforma portable.....	194

Índice de tablas

Tabla 5.1: Medidas realizadas con el maniquí KEMAR	145
Tabla 5.2: Características físicas medidas.....	148
Tabla 6.1: Resumen de recursos consumidos por las dos implementaciones	168
Tabla 6.2: Resultados de síntesis para distintas versiones del circuito en la tarjeta RC-100.....	178
Tabla 6.3: Resultados de síntesis de las distintas versiones implementadas en la RC-200	189

Capítulo 1

Introducción

El objeto de este primer capítulo introductorio es el de presentar el planteamiento para la elaboración del trabajo de tesis y resumir la estructura de esta memoria. En la sección 1.1 se introduce la problemática abordada, así como los objetivos planteados para este trabajo. En la sección 1.2 se expone el contexto de aplicación del trabajo desarrollado, y finalmente, la sección 1.3 muestra la estructura en capítulos de esta memoria de tesis doctoral.

1.1 Planteamiento del problema

La visión es uno de los principales órganos sensoriales del ser humano, como lo demuestra el hecho de que aproximadamente el 40% de toda la información que llega al cerebro está relacionada con el sistema visual. Aceptamos con toda naturalidad el ver, y pensamos que ello no comporta ningún esfuerzo especial, aunque existen aproximadamente 50 millones de personas ciegas en el mundo. Sin embargo las cifras que más preocupan no son estas, sino la previsión de que en 15-20 años esta cifra se puede duplicar, debido fundamentalmente a la falta de recursos en algunas sociedades, el

crecimiento demográfico, y el envejecimiento progresivo de las sociedades industrializadas. La incidencia en nuestro país es especialmente elevada no sólo debido a las enfermedades neurodegenerativas asociadas a la edad de nuestra población autóctona, sino también por el creciente porcentaje de personas de edad avanzada procedentes de otros países europeos que fijan aquí su residencia.

Conscientes de la necesidad de abordar soluciones para la rehabilitación visual, nos encontramos en un momento de las nuevas tecnologías idóneo para experimentar y desarrollar nuevos dispositivos capaces de mejorar la calidad de vida de las personas con distintas carencias visuales. Este es el contexto en que se centra el trabajo propuesto.

Como resultado de investigaciones realizadas en el seno del proyecto europeo CORTIVIS [*Cortivis*] se han obtenido elementos útiles para el desarrollo de modelos y prototipos de retinas artificiales, orientados fundamentalmente al diseño de neuroprótesis visuales y de sistemas artificiales (electrónico-informáticos) de apoyo a la visión mediante realce selectivo de la información visual.

Entre las herramientas y dispositivos desarrollados en el marco de CORTIVIS, destaca la creación de un entorno, llamado RETINER [*Pelayo 2004*], para la definición y ensayo de modelos de visión bioinspirados, basados en el procesamiento espacio-temporal que la retina humana realiza sobre las imágenes incidentes para su traducción a impulsos nerviosos, que serán posteriormente enviados a las áreas del cerebro especializadas en visión.

Este resultado sirve como punto de partida que inspire nuevos desarrollos, encaminados a investigar aplicaciones en distintas patologías del sistema visual humano, con objeto de obtener cierto grado de rehabilitación. A modo de ejemplo mencionamos la posible aplicación de transformaciones automatizadas de entradas visuales para corregir defectos del campo visual; por ejemplo los producidos por enfermedades como la degeneración macular o la retinosis pigmentaria.

En concreto, se ha desarrollado un sistema de transducción sensorial automática, de forma que las entradas visuales que el sistema natural es incapaz de procesar adecuadamente por su patología, son redirigidas hacia otros sentidos que conserven total o parcialmente su funcionalidad, como por ejemplo, el oído.

Este tratamiento requiere una transformación y selección de la información visual 'óptima' para su envío por un canal de comunicación de capacidad restringida. De esta forma, se puede efectuar un realce y

selección de la información que se desea extraer de la entrada de video de una cámara; por ejemplo contrastes de color o información sobre distancias u obstáculos.

Se trata de un trabajo de investigación realizado en el seno de un grupo multidisciplinar en el que se conjugan tareas de procesamiento de imágenes, modelos de percepción sensorial, desarrollo de herramientas software, e implementación digital de sistemas en hardware reconfigurable.

Dentro de este marco, los **objetivos** perseguidos con este desarrollo son:

- a) Proveer de un entorno de modelado, refinamiento y simulación de transformaciones visuales bioinspiradas que nos permitan extraer la información más relevante del entorno.
- b) Ofrecer la capacidad de transformar la información visual seleccionada para ser transmitida, en información auditiva, en unas condiciones altamente parametrizables y ajustables por el usuario final.

Los resultados obtenidos en relación con los objetivos a) y b), son de aplicación tanto para el desarrollo de neuroprótesis visuales (para invidentes, tal como se planteaba en el proyecto Cortivis), así como para el desarrollo de otros dispositivos no invasivos de ayuda a personas con discapacidades visuales.

1.2 Antecedentes

El desarrollo de sistemas de realidad aumentada y de redirección sensorial, para personas ciegas con o sin resto visual funcional, es un campo de investigación especialmente activo. El creciente número de personas que padecen enfermedades que afectan a la visión, especialmente debido a accidentes y al envejecimiento progresivo de la población, hace de este tema una prioridad entre los grupos de investigación.

Existen diversos sistemas previamente desarrollados cuyo objetivo encaja en este campo de aplicación. Entre ellos citamos a los más destacados:

SeeHear: un circuito integrado específico desarrollado por Carver Mead y Misha Mahowald [Nielsen 1989] para transformar la detección de movimiento visual en señales audibles.

LVES: sistema basado en tres cámaras montadas sobre un 'casco' de realidad virtual, que proporciona una transformación de la imagen de entrada para proyectar sobre la porción funcional del campo visual [LVES].

The vOICe: proporciona una salida audible que responde a niveles de intensidad de la imagen de entrada, sobre la que se realiza un barrido secuencial, codificando el barrido de columnas en el retardo del sonido y el de filas en la frecuencia del mismo [vOICe].

EAV: sistema que realiza un análisis en estéreo de la imagen de entrada y codifica la presencia de objetos asignándoles un sonido concreto y su distancia representada con la amplitud del sonido [EAV].

Sistemas como el LVES, desarrollado por la Universidad John Hopkins de Baltimore en colaboración con la NASA, o más recientemente el JORDY, el FlipperPort, o el MaxPort, comercializados por [Enhanced Vision], consiguen que la persona con baja visión emplee de forma más eficiente la poca visión que le queda a través del aumento de las imágenes y la mejora del contraste entre luz/oscuridad así como de los colores, pero no consiguen desarrollar un control de ganancia local eficiente que proporcione imágenes claras y nítidas en todas las situaciones de iluminación. Además suelen ser relativamente pesados (0.5-1 Kg.), caros y difíciles de manejar en movimiento.

Los sistemas de realidad aumentada basados en mini cámaras y visores Head-Mounted Displays (HMD) del tipo 'goggles' o de dimensiones más reducidas, con o sin transparencia parcial (*see through*), constituyen una alternativa muy válida investigada por varios grupos de I+D dentro y fuera de nuestro país. Estos sistemas electrónicos y opto-electrónicos de ayuda a la baja visión, cuando se basan en chips FPGA o DSP, tienen la posibilidad de reconfigurarse o reprogramarse para distintos tratamientos de imagen. Sin embargo la forma de definir el tratamiento óptimo y su traslación al hardware, programando o configurando el circuito, es una tarea muy laboriosa y relativamente compleja. Una necesidad para su aplicación práctica es disponer de herramientas que faciliten dicha reprogramación, o que la realicen *automáticamente*, pudiéndose

particularizar ágilmente para distintos problemas de la visión, o para distintos estadios de un mismo déficit visual degenerativo.

También existen múltiples grupos, incluido el nuestro, investigando en prótesis visuales para personas con ceguera total, que actúan a nivel de retina, de nervio óptico y de corteza cerebral. Aunque se trata de una tecnología que exige aún años de investigación para llegar a soluciones útiles, los resultados obtenidos en los últimos años hacen prever en un futuro próximo avances muy importantes como los ocurridos con los implantes cocleares, que se han convertido en una técnica clínica habitual con la que han podido rehabilitarse muchos miles de personas sordas.

Por ahora, como se recoge en el estudio publicado por la O.N.C.E. [Graw 2004], los sistemas de ayuda existentes para ciegos están orientados mayormente a intentar presentarles la información visual a través de otros sentidos como el auditivo o el táctil. Y para personas con baja visión o resto visual funcional, se ha recurrido principalmente a dispositivos ópticos que pretenden presentar la información visual en condiciones de mayor aprovechamiento.

Como hipótesis de trabajo, el desarrollo de cualquier dispositivo de ayuda a la visión debe tener en cuenta desde el principio a los potenciales usuarios al objeto de tratar de obtener un sistema que sea realmente adecuado a las necesidades de los mismos. Tener en cuenta y conocer al usuario final es esencial en nuestro abordaje y será una de nuestras prioridades.

Esta es una recomendación recogida en [Graw 2004], que debe tenerse en cuenta para el desarrollo de cualquier ayuda a la visión o tecnología específica para ciegos. El éxito de un producto de este tipo va a depender de (a) que el usuario se sienta cómodo con la ayuda (sencillo de manejar, que no falle, etc.), y (b) que la ayuda proporcione los resultados esperados. Para conseguirlo es necesario:

1. Un diseño *en base a y para* los usuarios, haciéndoles partícipes del mismo.
 2. Primar la evaluación funcional y no dar por bueno un desarrollo si no se cumplen sus especificaciones.
 3. 'Feedback' por parte del usuario. En la fase de pruebas, y antes de sacar una nueva ayuda, hay que recoger las impresiones, pegas, mejoras, defectos, etc., que los usuarios ven en la ayuda.
-

El actual desarrollo de la tecnología, en especial del hardware reconfigurable y las arquitecturas paralelas, nos permiten modelar el funcionamiento del sistema visual de una forma bioinspirada y abstraer la información extraída del entorno para ser posteriormente transformada y transmitida a las personas con carencias visuales, de la forma adecuada, siempre con la intención de complementar al resto de facilidades, no de sustituirlas ni anularlas. Gracias a la posibilidad de realizar sistemas completos de tratamiento de imagen en un solo chip, funcionando en tiempo real, planteamos soluciones portátiles capaces de realzar aquellas componentes o características visuales que sean necesarias para mejorar el grado de percepción visual.

Dada la naturaleza del procesamiento a llevar a cabo sobre la información visual, los requisitos de portabilidad y re-programabilidad o re-configuración de estos dispositivos, consideramos oportuno el uso de soluciones hardware con circuitos integrados de lógica programable (FPGA). La complejidad actual de estos circuitos y su velocidad de funcionamiento los convierte en una opción de implementación muy válida para sistemas portables, alimentados con baterías, fácilmente reconfigurables para distintos pacientes, o para distintos estadios de la deficiencia visual.

1.3 Organización de la memoria

El presente trabajo se encuentra organizado en capítulos con la siguiente estructura:

- Capítulo 1 – *Introducción*: Presente capítulo que nos acerca al problema tratado en el trabajo de tesis y los antecedentes sobre el tema.
 - Capítulo 2 – *Visión y Rehabilitación Visual*: Abarca una introducción al sistema visual humano, distintas patologías por las que puede verse afectado, y una revisión acerca del estado del arte sobre los sistemas de rehabilitación visual.
 - Capítulo 3 – *Procesamiento de Información Visual: Retiner*: Este capítulo está dedicado a la plataforma de definición,
-

simulación y ajuste de modelos de visión, en cuyo desarrollo ha participado el doctorando, constituyendo una de las aportaciones de la tesis.

- Capítulo 4 – *Modelo de Procesamiento de Visión Estéreo*: En este capítulo se plantea una extensión del modelo de procesamiento visual definido en Retiner, al objeto de incorporar visión estereoscópica. Presenta una descripción del sistema de visión estereo y los algoritmos utilizados, y justifica su utilidad y mejoras que aporta al sistema global.
 - Capítulo 5 – *Sistema de generación de sonidos 3D*: En este capítulo se incluye una descripción del sistema de generación de sonidos localizados espacialmente, abarcando desde su fundamento teórico, herramientas y material utilizados hasta la implementación software propuesta y las posibilidades de adaptación a cada usuario.
 - Capítulo 6 – *Sistemas para Rehabilitación Visual integrados en FPGA*: Este capítulo se centra en la implementación en hardware reconfigurable de los diferentes modelos de procesamiento descritos en capítulos anteriores, abarcando todo el procesamiento de información visual bioinspirado y dos posibilidades de aplicación: su integración con un módulo de generación de impulsos para la estimulación nerviosa, o con un sistema de transducción sensorial que produzca patrones sonoros localizados espacialmente.
 - Capítulo 7 – *Conclusiones y principales aportaciones*: Capítulo dedicado a recopilar las aportaciones del trabajo doctoral realizado y exponer las conclusiones, en base a los distintos resultados expuestos en distintos capítulos de toda la memoria, así como algunas líneas de trabajo futuro.
 - *Referencias bibliográficas*: Recopilación de las publicaciones científicas, sitios web e informes técnicos utilizados para la elaboración del trabajo de tesis doctoral.
-

Capítulo 2

Visión y Rehabilitación

Visual

Una buena parte del trabajo que presentamos se centra en el modelado de distintas facetas del sistema visual humano. Dentro de este capítulo nos acercamos al sistema visual humano. La sección 2.1 nos presenta su estructura y funcionamiento, para comprender las posibles patologías que pueden afectarle, las cuales las podemos encontrar en la sección 2.2. La última parte del capítulo está plenamente dedicada a una revisión sobre los distintos sistemas de rehabilitación visual que nos podemos encontrar en la literatura (sección 2.3), atendiendo a sus características principales y punto del tracto visual en que se hace la interfaz.

2.1 Descripción del Sistema Visual Humano

Los procesos neuronales que nos permiten entender el mundo exterior por medio de la visión engloban el funcionamiento de numerosas estructuras. Imaginemos un observador ante una escena cotidiana. Su *campo visual*, la zona del espacio que puede ver, se encuentra poblado de objetos (formas) estáticos o en movimiento y repletos de color. ¿Cómo recibe esta información nuestro sistema visual, y cómo la interpreta posteriormente? Cualquier objeto que se sitúe delante de nuestros ojos se proyectará en la *retina*. En ella, las células se disponen formando un mosaico, de manera que cada una de ellas recibe información de una zona muy pequeña del campo visual, su *campo receptivo*, y al ser estimulada produce un aumento de su actividad eléctrica. Esta información se envía hacia la corteza cerebral, pasando por el núcleo geniculado lateral.

Pasamos a ver más detalladamente cada uno de los elementos que hemos comentado.

2.1.1 El ojo y la retina

El primer elemento que forma parte del sistema visual es el ojo. Su estructura está diseñada para que los rayos de luz alcancen la retina con la mínima distorsión óptica posible. Los sistemas ópticos actúan como un grupo de lentes cuya función principal es enfocar la imagen visual, de forma invertida, en la retina.

Una sección horizontal (ver Figura 2.1) del ojo humano nos permite observar cómo los rayos de luz atraviesan los elementos ópticos antes de alcanzar la retina. Allí podemos encontrar una zona central llamada *mácula* que se caracteriza por la abundancia de un tipo de fotorreceptores, los conos. En el ser humano, la extensión de esta área es aproximadamente de 5.5 mm de diámetro [Gil 1995].

En la parte central de la mácula podemos distinguir otra zona, la *fóvea*, en la que los cuerpos de las células más cercanas están desplazados, permitiendo que llegue la luz de forma más directa a los fotorreceptores. Esta parte de la retina tiene forma circular con un diámetro que puede variar entre 0.5 y 1 mm [Dowling 1987], representando aproximadamente 3.5° del campo visual [Hubel 1988]. En la fóvea se presenta la máxima agudeza visual, debido a la mejor recepción de luz y también por tener la

mayor densidad de fotorreceptores de toda la retina. Una particularidad, es que todos los fotorreceptores que podemos encontrar en la fovea son conos, responsables de la percepción de los detalles de formas con una resolución espacial elevada y el color.

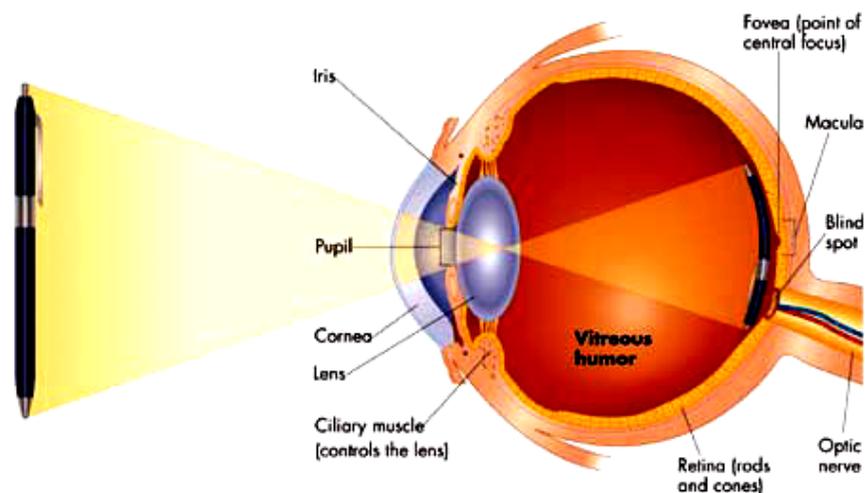


Figura 2.1: Sección horizontal del globo ocular [Cudeiro 2005].

Las terminaciones nerviosas de las células de la última capa de la retina se reúnen formando el nervio óptico, que abandonan el globo ocular por una zona en la que, lógicamente, no existen fotorreceptores. Esta ausencia nos priva de la información visual correspondiente a dicha zona, conocida como *punto ciego*.

El campo visual es el cono espacial que puede percibir un ojo manteniendo la mirada fija en un punto, y se expresa en grados angulares. Como vemos en la figura 2.2, el campo visual puede dividirse en dos zonas: nasal, y temporal.

La parte nasal de la retina recibe la luz del campo temporal de visión, y viceversa. Como el ojo es esférico, la retina nasal del ojo izquierdo y la retina temporal del ojo derecho reciben información de la mitad izquierda del campo visual (*hemicampo* izquierdo), mientras que la retina nasal del ojo derecho y la temporal del ojo izquierdo miran hacia el *hemicampo* derecho.

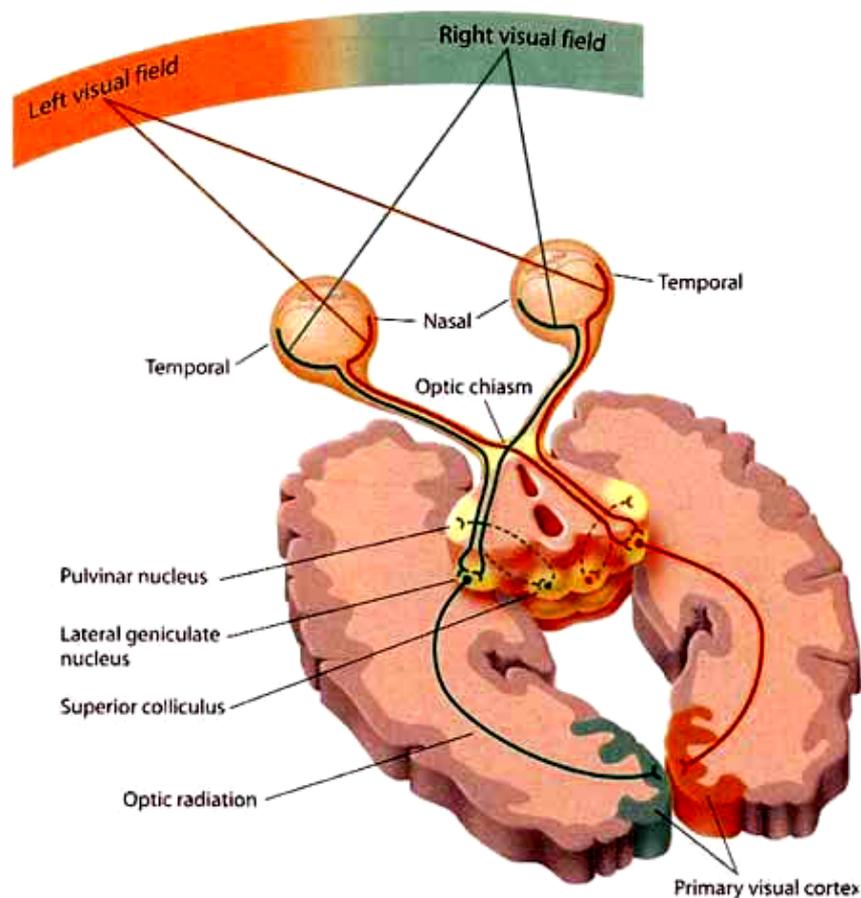


Figura 2.2: Proyección del campo visual. Se ha resaltado el camino visual desde las retinas hasta la corteza primaria (en la región occipital del cerebro) [Goldstein 1999].

2.1.1.1 La retina y su función

Los primeros estudios de la retina comienzan con Ramón y Cajal [Ramón y Cajal 1892] y se ha seguido tratando con posterioridad debido a tres razones fundamentales:

- Ha resultado muy útil para la comprensión de los mecanismos generales relacionados con el fenómeno de la *transducción sensorial*, dado que los fotorreceptores son, probablemente, las células sensoriales mejor estudiadas.

- La retina, a diferencia de otras estructuras sensoriales, como la cóclea, no es un órgano periférico, sino que forma parte del *sistema nervioso central* (SNC) [Duke-Elder 1963; Mann 1964].
- La retina está constituida de una manera relativamente simple en comparación con otras estructuras del SNC. Contiene seis clases principales de neuronas unidas por un patrón de conexiones que es complicado, pero presentando una ordenación estructural en capas.

La estructura en capas que presenta la retina se puede observar en la figura 2.3. En ella podemos distinguir dos capas plexiformes, la externa y la interna. En estas zonas se producen las sinapsis entre distintas células de la retina. Se encuentran localizadas entre tres capas de cuerpos neuronales.

Los seis tipos celulares se distribuyen desde la zona más posterior del ojo hacia el humor vítreo en el siguiente orden: fotorreceptores, células bipolares, células horizontales, células interplexiformes, células amacrinas y células ganglionares. Éstas constituyen la capa de salida de la retina y sus axones son los que forman el nervio óptico. Detrás de la retina se encuentra el epitelio pigmentado, que se encarga, entre otras funciones, de evitar la reflexión de la luz evitando distorsiones de la imagen percibida.

El número de fotorreceptores existente es mucho mayor que el de células bipolares, y el número de éstas mayor que el de células ganglionares. Esto significa que ha de existir una convergencia importante, de modo que muchos fotorreceptores contactan sobre una misma célula bipolar, y muchas bipolares sobre la misma célula ganglionar. En el ser humano hay aproximadamente 125 millones de fotorreceptores, y sólo algo más de 1 millón de células ganglionares en cada ojo. No obstante la convergencia no es igual en toda la retina sino que varía mucho con la excentricidad [Sterling 1990] [Beaudot 1994], creciendo conforme nos vamos retirando desde la fovea hacia la periferia de la retina.

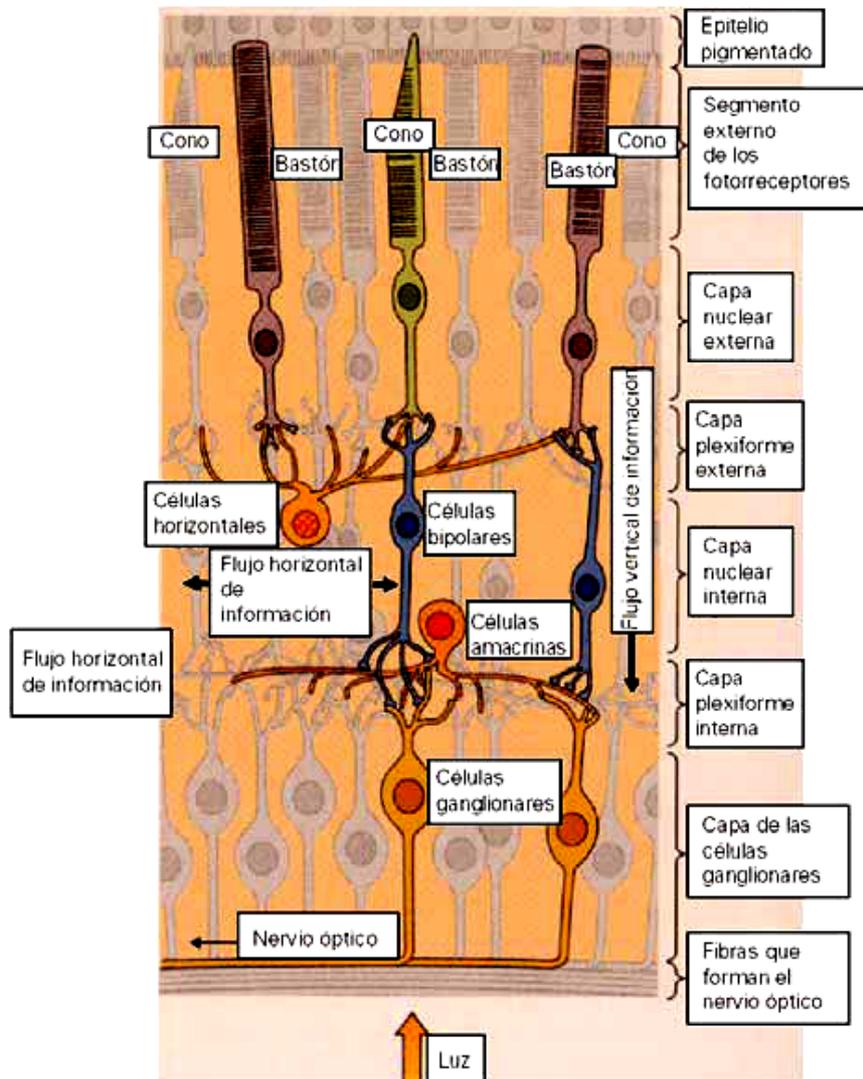


Figura 2.3: Estructura en capas de la retina.

Por tanto, podemos considerar que la entrada a la retina son los conos y los bastones, y que su salida está representada por los axones de las células ganglionares. Los conos y los bastones están conectados entre sí mediante uniones sinápticas eléctricas (o *gap junction*). Entre los fotorreceptores y las células ganglionares, existen varias capas neuronales que no se limitan a actuar como interconexiones entre la entrada y la salida, sino que combinan las señales procedentes de varios receptores, de forma que la salida de las células ganglionares va a depender críticamente

de las características temporales y espaciales de los estímulos luminosos que inciden sobre la retina.

2.1.1.2 Las células bipolares

Además de las sinapsis eléctricas, los fotorreceptores interconectan con las células horizontales y las bipolares. Los bastones terminan en unas estructuras esféricas llamadas *cuerpos sinápticos* y los conos lo hacen en forma de un pie ancho que se conoce con el nombre de *pedículo*. En estos puntos es dónde se realizan las sinapsis con las células bipolares y horizontales. El comportamiento de las células bipolares va a depender de si la sinapsis la realiza con un cono o con un bastón.

Ahora veremos en más detalle cada uno de los dos caminos paralelos de información que van de los fotorreceptores hasta las células bipolares.

Hablamos de la **vía de los conos** para indicar que la información de estos fotorreceptores llega a dos tipos de células bipolares (ver figura 2.4). Un grupo se comporta como los fotorreceptores y se despolariza en la oscuridad, se trata de las bipolares *off*. El otro grupo se comporta de forma opuesta, se hiperpolariza en la oscuridad, lo que significa que estas células se excitarán en presencia de luz, por lo que se les llamó bipolares *on*.

Así podemos decir que la información luminosa actúa por dos vías, la presencia de luz nos llegará a través de las bipolares *on*, mientras que la ausencia de luz se propagará a través de las bipolares *off*.

Todas estas células, independientemente del tipo que sean, presentan una estructura circular concéntrica, antagónica en cuanto a la localización espacial del estímulo que reciben. La respuesta producida cuando el estímulo se encuentra en el centro de su campo receptivo será la opuesta a la que obtenemos al desplazar el estímulo a la periferia. Este comportamiento es conocido como antagonismo centro/periferia, o antagonismo *on/off*. Las células bipolares que se conectan con conos conectan directamente con las células ganglionares.

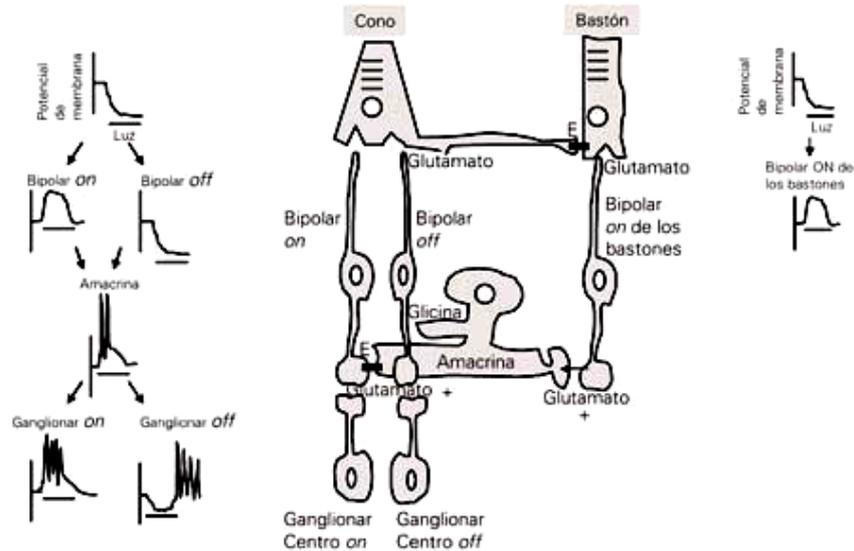


Figura 2.4: Circuitos relacionados con el funcionamiento de los conos y los bastones en la retina de los mamíferos [Cudeiro 2005].

La **vía de los bastones** se caracteriza porque éstos conectan con bipolares sólo de tipo on (figura 2.4), aunque esto no significa que los bastones sean ajenos al procesamiento antagónico on/off; tal procesamiento se realiza de forma indirecta a través de las conexiones con las células amacrinas que se interponen entre estas bipolares y las ganglionares correspondientes.

2.1.1.3 Las células horizontales

Dependiendo de su axón, en los vertebrado nos encontramos con dos clases de células horizontales: las de de axón corto y las que no tienen axón (que no aparecen en los primates). Las dendritas de las primeras están conectadas con conos, mientras que sus terminaciones axónicas establecen contacto con los bastones.

Las células horizontales sin axón, en los mamíferos sólo contactan con los conos. Su comportamiento ante la luz es similar al de los fotorreceptores, por lo que podemos decir que son responsables (junto con las amacrinas) de la inhibición lateral en la retina. Se pueden dividir en dos grupos dependiendo de sus diferentes respuestas ante diferentes longitudes de onda. Las del tipo L (luminosas) responden ante cualquier

longitud de onda, mientras que las del tipo C (cromáticas) tienen respuesta distinta ante estímulos con diferentes longitudes de onda.

La mayoría de células horizontales se caracterizan por tener un campo receptivo mucho más grande del que le correspondería, ya que están acopladas eléctricamente con células vecinas, recibiendo información de zonas alejadas en la retina.

2.1.1.4 Las células amacrinas

Existen multitud de tipos de células amacrinas, pero la mayoría presentan un comportamiento transitorio ante cambios producidos en el estímulo en cualquier parte de su campo receptivo. Dicho campo es bastante grande, respondiendo a estímulos en amplias zonas de la retina y no poseen una periferia antagónica. Este tipo de células, que tampoco son sensibles al color, reciben el nombre de amacrinas transitorias. Además podemos encontrarnos con una minoría de células amacrinas que responden de forma sostenida y dependiendo de la longitud de onda del estímulo.

Ya que la mayoría se activan ante apariciones o desapariciones del estímulo, se piensa que estas células están destinadas a señalar cambios en las condiciones de iluminación, más que a responder a un estímulo mantenido. De esta forma, actuarán sobre un tipo de células ganglionares cuya respuesta también es transitoria.

2.1.1.5 Las células interplexiformes

Se trata de un tipo de células difíciles de observar, aunque se ha visto que sus entradas provienen de las amacrinas, y su salida viaja hacia las horizontales, las bipolares, o ambas. Se conoce muy poco de su funcionalidad, aunque se ha sugerido que disminuyen la inhibición periférica (del componente centro/periferia del campo receptivo) medida por las células horizontales en función de la iluminación. De esta forma juegan un papel importante en la adaptación visual y en el paso de una visión *fotópica* a una visión *escotópica* [Beaudot 1994].

2.1.1.6 Las células ganglionares

Constituyen la última capa celular de la retina, proporcionando el resultado del procesamiento visual que se produce en la misma. Las células ganglionares también presentan campos receptivos más o menos redondeados con un centro y una periferia antagónicos, al igual que ocurre en niveles más cercanos a los fotorreceptores.

De igual manera podemos encontrar células ganglionares de centro *on* y periferia *off*, y viceversa. La figura 2.5 nos muestra el comportamiento de ambos tipos de células. Las células con centro *on* se excitan cuando la luz incide directamente en el medio de su campo receptivo, mientras que cuando la luz incide sobre la periferia la célula disminuye su frecuencia de disparo. Las células de centro *off* se comportan al revés. Un punto de luz localizado en el centro del campo receptivo inhibirá la respuesta de la célula, que se incrementará al apagar la luz o al situarla sobre la periferia. Una iluminación que abarque el centro y la periferia producirá en ambos tipos una respuesta débil. Por medio de estos dos tipos celulares, *on* y *off*, la retina manda una información de salida por dos vías paralelas.

En el dominio temporal, y de forma independiente a la clasificación vista con anterioridad, podemos encontrarnos con dos tipos de células, con respuestas distintas en el tiempo. Unas células ganglionares (ya sean *on* u *off*) responden a los estímulos de forma transitoria, mientras que otras lo hacen de forma más sostenida.

Para estudiar este tipo de respuestas y analizar las interacciones espaciales que ocurren entre el centro y la periferia, se utiliza un estímulo visual denominado *grating*, y que consiste en la repetición periódica de un patrón de bandas alternantes claras y oscuras. Las diferentes respuestas de estas células cuando cambiamos la posición del *grating* dentro de su campo receptivo las podemos observar en la figura 2.6.

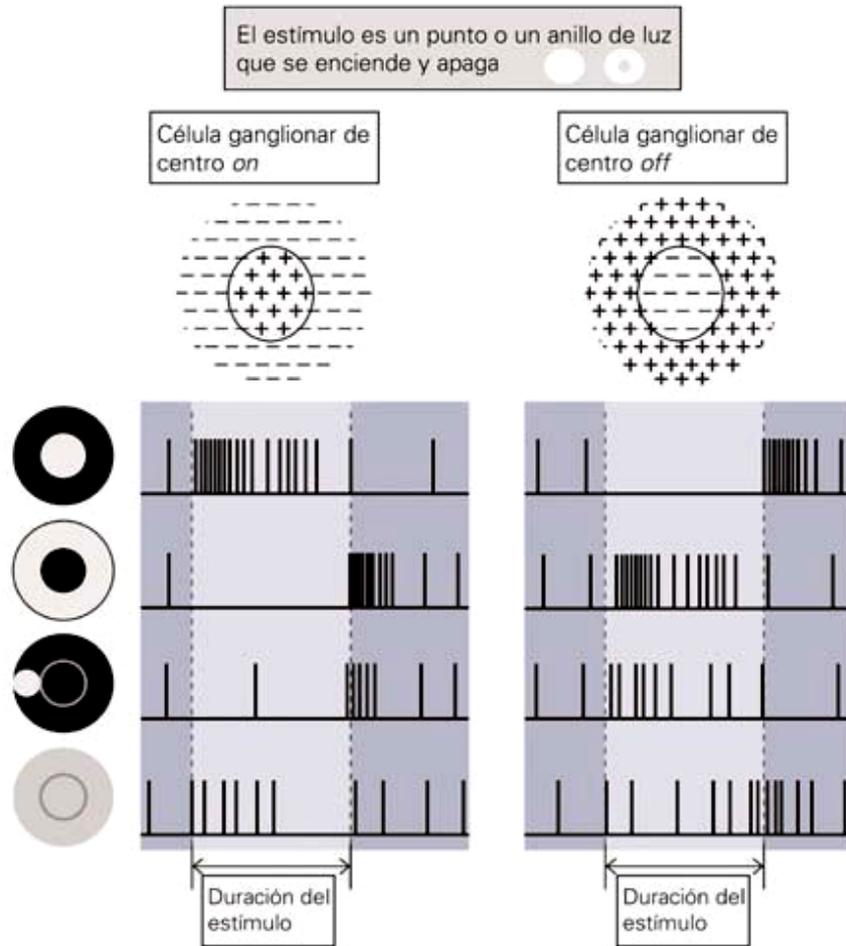


Figura 2.5: Respuesta de las células ganglionares al estimular el campo receptivo: antagonismo centro/periferia.

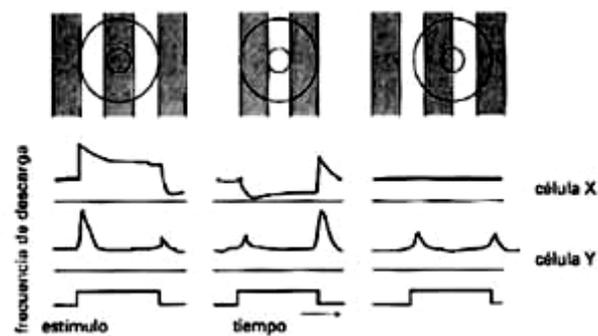


Figura 2.6: Respuesta de dos células ganglionares pertenecientes a dos tipos distintos: X e Y.

En la parte superior podemos distinguir una célula tipo off, ya que responde de forma sostenida cuando la banda oscura del grating se localiza sobre el centro de su campo receptivo; pero, cuando se centra sobre la periferia, la respuesta se anula (llegando a invertir su polarización), y aumenta transitoriamente cuando el estímulo se apaga (rebote). Podemos encontrar una posición del grating en la cual las respuestas centro/periferia se cancelan y la célula no produce ninguna respuesta. Esto se conoce como *punto nulo*. A estas células se les llama, en el gato, en donde fueron descritas, células X, y desde la retina mandan sus axones exclusivamente al núcleo geniculado lateral.

Sin embargo, otro tipo de células ganglionares, tal y como podemos ver en la parte inferior de la figura 2.6, se comporta de forma muy distinta. Su respuesta es transitoria y no existe punto nulo. Estas células se conocen como Y, y proyectan hacia el núcleo geniculado lateral y hacia el colículo superior.

Las diferencias entre este tipo de células no se limitan a las descritas, sino que hay muchas más, por ejemplo, el tamaño del campo receptivo, que es menor en las X; la velocidad de conducción, mayor en las Y; la resolución espacial, que es mejor en las X, o la resolución temporal, que es mejor en las Y. En cuanto al color, ambos tipos se consideran acromáticos, aunque sabemos que algunas células X del gato tienen una cierta preferencia por determinadas longitudes de onda.

También se ha detectado una pequeña población que no encuadra en ninguno de los tipos celulares descritos anteriormente, difíciles de estimular y con una respuesta tardía, sin una clara organización centro/periferia, y con una velocidad de conducción menor que la de las células X. Se llamaron células W, y teniendo en cuenta sus lugares de proyección fuera de la retina – el colículo superior, el pretectum y el núcleo óptico accesorio – se ha sugerido que están relacionadas con la actividad visuomotora.

Los tres tipos de células ganglionares clasificados por sus parámetros funcionales, se pueden clasificar también desde un punto de vista morfológico. Se distinguen tres tipos principales de células ganglionares: α , con soma grande y expansiones dendríticas amplias, β , con soma pequeño y dendritas densamente agrupadas en un espacio reducido, y γ , con un soma pequeño y amplias ramificaciones dendríticas (en realidad se han descrito, además, un tipo δ y otro ϵ , muy similares a las γ). Existe, por supuesto, una correspondencia morfofuncional, de manera que las células Y se corresponden con las α y representan un 4% de la población

total, las X con las β , con un 50-65%, y finalmente las W con las γ , que representan un 45% del total [Kolb 1984] [Sterling 1986].

Diversos estudios ([Perry 1981] [Leventhal 1981] [Kaplan 1986]) en los primates llegaron a clasificaciones prácticamente coincidentes entre sí (aunque cada autor le dio una nomenclatura distinta) que se pueden resumir en los siguientes tipos celulares:

- M (*magnocelulares*), P α ó A: células de gran tamaño, dendritas largas y axón delgado que constituyen aproximadamente el 8-12% del total de células ganglionares. Se corresponden con las células α y β descritas en el gato.
- P (*parvocelulares*), P β ó B: células de cuerpo pequeño, dendritas cortas y ramificadas, y axón más delgado que las anteriores. Son mayoritarias, abarcando el 80% del total.
- P γ ó C: otro tipo de células pequeñas, dendritas largas y axón extremadamente corto, que se corresponden con el 10%.

2.1.2 La visión en color

En la retina existen tres tipos de conos, cada uno de ellos con un pigmento visual distinto y con distinto espectro de absorción de la luz. Cada cono contiene solamente uno de los tres posibles pigmentos (figura 2.7): uno de ellos es sensible, fundamentalmente, a longitudes de onda cortas del espectro visible (pico de absorción en 419 nm), y hace posible la percepción del azul (cono A). Otro es selectivo para longitudes de onda medias (531 nm) y se relaciona con la percepción del verde (cono V). El último pigmento absorbe longitudes de onda largas (550 nm) y permite la percepción del rojo (cono R).

Cuando un cono absorbe un fotón, su respuesta eléctrica es siempre la misma, sea cual sea la longitud de onda del fotón. Esto significa que los conos, individualmente, no pueden transmitir información acerca de cuál es la longitud de onda que los excita, ya que para nuestro sistema nervioso lo único que transmiten es una señal eléctrica. Esto no implica que los conos no tengan una sensibilidad preferente hacia una determinada longitud de onda; por supuesto que la tienen, pero lo que esto quiere decir es que la *probabilidad de absorber un fotón* es mayor o menor dependiendo de su longitud de onda, aunque la respuesta eléctrica sea siempre la misma. Ésta es la razón por la cual las personas con un solo tipo de conos no ven todo de un color, sino que su visión es similar a la de los sujetos normales en condiciones escotópicas (con poca luz), cuando la información visual

depende de los bastones y la percepción del mundo es acromática; pero, incluso con muy poca luz, podemos distinguir un objeto si la luz que refleja es mayor que la del fondo. En otras palabras, lo distinguimos comparando el objeto contra su fondo.

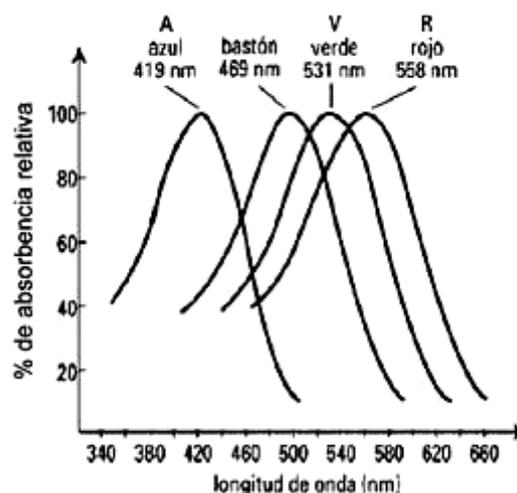


Figura 2.7: Tipos de conos según su respuesta a diferentes longitudes de onda.

La visión en color requiere, al menos, dos grupos de fotorreceptores con sensibilidades espectrales diferentes. De esta manera, un sistema con dos variables proporcionaría dos valores sobre la intensidad de luz reflejada por un objeto. Comparándolos, el cerebro sería capaz de distinguir los colores. Por ejemplo, si un objeto refleja fundamentalmente luz de longitud de onda larga, la respuesta de los conos con pigmento selectivo para dicha longitud será mayor que la respuesta del otro sistema (menor probabilidad de absorción), y las zonas del cerebro encargadas de procesar la información del color entenderán que el objeto es rojo o amarillo. Si el objeto refleja longitudes de onda cortas, será percibido como azul; y si refleja longitudes de onda cortas y largas por igual, lo percibiremos como blanco, gris o negro, dependiendo del brillo del fondo. Hemos conseguido distinguir los colores con un sistema con dos variables o dicromático.

No obstante, muchas combinaciones de color entre el objeto observado y el fondo son indetectables para el sistema de dos variables. Un objeto que refleje luz de ambos extremos del espectro y que aparezca sobre

un fondo que refleje luz de una longitud de onda en el medio del espectro sería invisible, porque tanto el fondo como el objeto producirían la misma respuesta en ambos tipos de fotorreceptores. Todos estos problemas y ambigüedades se pueden obviar con un sistema de tres receptores diferentes.

La teoría que acabamos de describir y que utiliza tres pigmentos distintos, no es suficiente por sí sola para explicar varios aspectos importantes de la percepción del color:

- ***Antagonismo del color.*** Ciertos colores nunca se perciben combinados. Por ejemplo, no percibimos un verde rojizo o un amarillo azulado, pero sí somos capaces de ver un azul rojizo (magenta), un amarillo rojizo (naranja) o un azul verdoso (cian). La luz roja y verde se mezclan de forma que todos los trazos originales de la rojez o del verdor se pierden y obtenemos un amarillo puro; de igual manera, el azul y el amarillo se mezclan para dar blanco sin dejar atisbo de los colores originales. Esta cancelación perceptual de los colores, llevó a proponer la teoría del antagonismo del color, según la cual hay seis cualidades primarias de color, que se procesan en tres pares que se antagonizan mutuamente: rojo-verde, amarillo-azul y blanco-negro. Estos tres pares se analizan en el sistema visual por medio de tres pares de canales de información diferentes antagonistas al color. En un par, las sensaciones de rojo y verde son opuestas: un canal se excita con el rojo y se inhibe con el verde, mientras que el otro se excita con el verde y se inhibe con el rojo. Algo similar ocurre para el amarillo y el azul y el blanco y negro. Las salidas de los tres tipos de conos se combinan de forma antagónica, comenzando en la retina, continuando en el núcleo geniculado lateral y, finalmente, en la corteza.
 - ***Contraste simultáneo del color.*** Aparece cuando dos colores antagónicos emanan de puntos del espacio vecinos, como es el caso de un objeto y su fondo. En un ejemplo de este tipo, un objeto de color verde resaltará más si su fondo es rojo que si es azul. Dicho con otras palabras, los colores antagónicos se resaltan mutuamente cuando se localizan en puntos adyacentes en el espacio. Esta característica se explica, en base a la existencia de células con antagonismo doble en la corteza visual.
-

- **Constancia del color.** Somos capaces de percibir el color de un determinado objeto como relativamente invariable a pesar de los cambios enormes en la composición espectral de la luz ambiente. Un limón será amarillo independientemente de si lo vemos a la luz del día (blanquecina), iluminado por una bombilla incandescente (proporciona una luz rojiza) o bajo una luz fluorescente (azulada). Este fenómeno psicofísico tiene su sustrato neuronal en las células de la corteza visual.

En los primates, la información sobre el color y sobre la intensidad de la luz reflejada por un cuerpo se transmite por vías separadas. Las células ganglionares de la retina y las células del núcleo geniculado se pueden dividir en dos grandes grupos: células de banda ancha y células con antagonismo al color. Las primeras proporcionan información sobre la intensidad de luz, y su campo receptivo tiene una organización concéntrica centro/periferia de carácter antagónico. Un punto de luz blanca localizado en el centro del campo excita (o inhibe) la célula, mientras que un estímulo de luz situado sobre la periferia produce el efecto contrario. Las células de banda ancha reciben señales de los conos V y R (no de los A) en el centro del campo receptivo y en la periferia, y se suman independientemente (figura 2.8). Responden, por tanto, al contraste en intensidad de iluminación, pero no contribuyen a la percepción del color.

En la mayoría de las células con antagonismo al color las entradas procedentes de los conos V y R son de polaridad opuesta y están organizadas en forma de antagonismo concéntrico centro/periferia. El centro recibe una entrada de un cono, por ejemplo R, mientras que la periferia recibe una señal de un cono G. No reciben señales procedentes de los conos A. Estas neuronas se llaman células con antagonismo simple y responden de forma vigorosa cuando estimulamos su campo receptivo con un punto grande (incluso cubriendo la periferia) de luz monocromática de la longitud de onda apropiada. La ganglionar V-centro/R-periferia, por ejemplo, responde óptimamente cuando la luz es verde. Debido a que los conos V y R son capaces de absorber luz blanca en igual medida en el centro y en la periferia, ante un estímulo acromático se comportan como las células de banda ancha.

Finalmente, la información que proviene de los conos A se transmite por medio de un tipo especial de células con antagonismo simple cuyo campo receptivo no presenta zonas separadas de excitación e inhibición, es decir, no tiene la típica estructura centro/periferia. En estas células, las señales de los conos A se oponen a la señal combinada de los conos V y R en todo el campo receptivo.

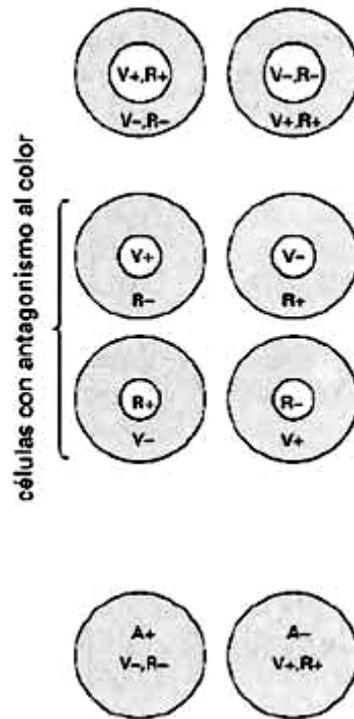


Figura 2.8: Células de banda ancha (arriba), células con antagonismo al color (centro), y células de entradas mixtas (abajo). Adaptación de [Gouras 1991].

2.1.3 El núcleo geniculado lateral (LGN)

En los mamíferos cuyos ojos se disponen frontalmente, como los primates y la mayoría de los carnívoros, los axones que dejan la retina y forman el nervio óptico se organizan de tal forma que la información visual que procede de un mismo lugar del espacio y es recogida por ambos ojos pueda combinarse. La representación del espacio visual tanto en el tálamo como en la corteza es contralateral, ya que las células ganglionares cuyo campo receptivo se localiza en una mitad del campo proyectan hacia el tálamo y la corteza del lado opuesto. Este cruce ocurre en una estructura conocida como *quiasma óptico* (figura 2.2). Por ejemplo, la retina temporal

del ojo izquierdo y la nasal del derecho informan sobre el hemisferio derecho, que estará representado en el tálamo y la corteza izquierdos.

En el LGN, la información visual procedente de la retina llega a las células principales y de aquí proseguirá hacia la corteza. De todas maneras, la inmensa mayoría de las sinapsis que reciben las células principales tienen un origen extrarretiniano. Estas conexiones proceden, fundamentalmente, de la corteza, del tronco del encéfalo y del propio núcleo o su vecindad. Las aferencias retinianas son siempre excitadoras.

El LGN se considera un filtro dinámico capaz de controlar la información que llega a la corteza visual dependiendo del estado del sujeto (despierto o dormido) y de su nivel de atención. Además, debido a las conexiones descendentes, la corteza es capaz de modificar las entradas de la retina al LGN y, por ejemplo, facilitar determinadas respuestas a la vez que desactiva otras, dependiendo de circunstancias como la localización del estímulo en el espacio.

2.1.4 La corteza visual

Desde el LGN, la información visual sigue su camino hasta la corteza cerebral (figura 2.2), donde la mayoría de los axones alcanza la corteza visual primaria, corteza estriada o V1 (algunas fibras llegan directamente a V2, o corteza secundaria en incluso hasta zonas tan alejadas como V5). V1 es la parte de la corteza cerebral mejor estudiada. Esto se debe a los trabajos pioneros de David Hubel y Torsten Wiesel [*Hubel 1988*], quienes en la década de los sesenta sorprendieron a la comunidad científica con una descripción precisa de las características funcionales de las células corticales del gato; descripción en la que enfatizaban las notables diferencias existentes con las estructuras previas en el procesamiento visual. En efecto, las respuestas de las células de la corteza son radicalmente diferentes de aquellas que se pueden registrar tanto en la retina como en el LGN. Si las células del LGN presentan unos campos receptivos concéntricos, de centro on u off según el tipo de neurona y son monoculares (responden a un solo ojo), en la corteza los campos son alargados (de forma mayoritaria, ya que también existen células con campos receptivos concéntricos y antagónicos [*Gouras 1991*]), muchos de ellos son binoculares y, sorprendentemente, la mayoría son capaces de señalar determinadas características de los estímulos visuales como la orientación.

Se han descrito unas 30 áreas de la corteza relacionadas con la visión con una complejidad creciente en cuanto a la respuesta ante los estímulos y en cuanto a la información que nos ofrecen sobre el mundo a nuestro alrededor. Las complejas interconexiones entre estas zonas y las etapas previas en el camino visual fueron descritas por van Essen [*van Essen 1981*]. La complejidad es muy elevada, dado que el sistema visual no solamente está organizado de forma jerárquica con conexiones en un solo sentido, sino que existe una gran cantidad de conexiones hacia atrás (*feedback*). No podemos olvidar además, que el sistema visual se trata de un sistema visuomotor, en el que los ojos se encuentran en constante movimiento y son niveles superiores los que se encargan de la reconstrucción de la escena a partir de la información obtenida en las retinas, permitiendo ver como estáticos los objetos, a pesar de los continuos cambios de su proyección en la retina.

2.1.4.1 Arquitectura de la corteza visual

La figura 2.9 nos muestra la localización de la corteza visual primaria, en el polo occipital. La parte central del campo visual (es decir, la mácula y la fóvea) está representada en la porción de la corteza visual primaria más cercana al polo occipital, y ocupa aproximadamente la mitad de V1. La parte de la corteza correspondiente a V1 está formada por seis capas y se caracteriza por una ancha capa IV. Esta contiene una banda adicional de fibras mielínicas, la estría de Gennari (figura 2.10), que explica su denominación de corteza estriada. Además, la capa más profunda, capa VI, destaca mucho y es el origen de muchas conexiones de retroalimentación hacia el LGN, lo que se conoce como conexiones corticofugales o feedback corticogeniculado.

Las neuronas de la corteza estriada aparecen con sus cuerpos dispuestos en forma de capas o láminas paralelas a la superficie cortical. Estas capas pueden identificarse claramente con técnicas de tinción simples como la de Nissl (figura 2.11). Las capas celulares se denominan con números del 1 al 6 (tradicionalmente se empleaban números romanos), desde la 1, situada más superficialmente, hasta la 6, en la profundidad de la corteza, limitando con la sustancia blanca. En el primate, el espesor completo de la corteza estriada es de unos 2 mm.

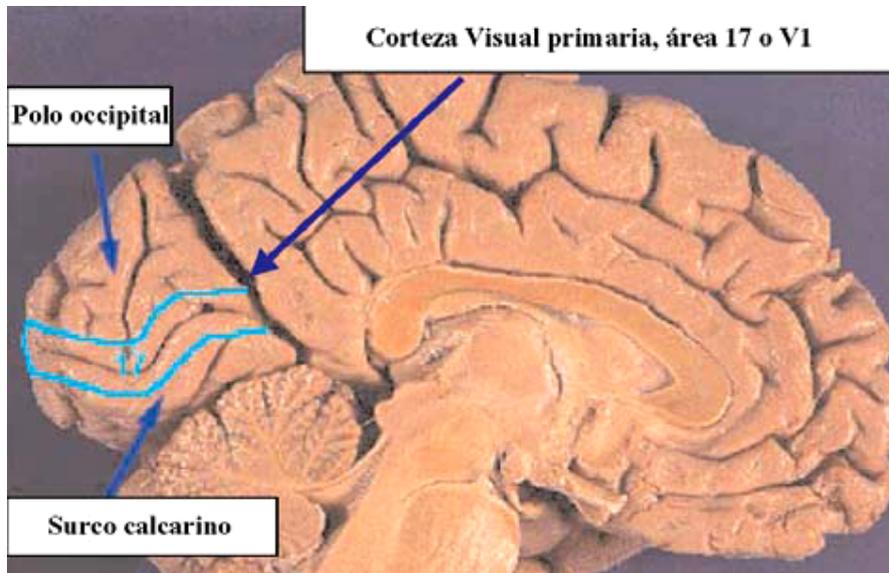


Figura 2.9: Localización de la corteza visual primaria en un corte de cerebro humano.

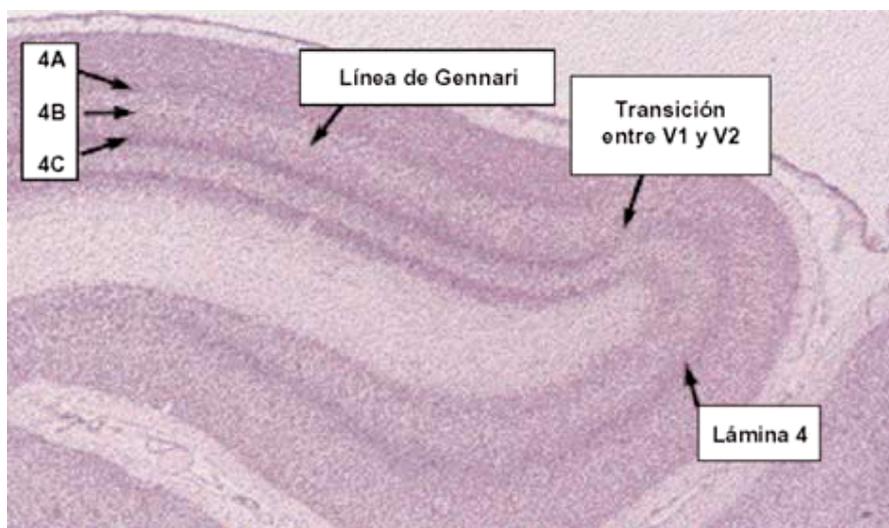


Figura 2.10: Tinción que muestra la línea de Gennari, en cuyo final observamos la transición entre V1 y V2.

Como se muestra en la figura 2.11, en realidad el número de capas de V1 es mayor de seis. Existen como mínimo nueve capas diferentes debido a las subdivisiones de alguna de ellas. De esta forma, se describen

tres subcapas en la capa 4 denominadas 4A, 4B y 4C. La capa 4C se divide adicionalmente en 4C α y 4C β . De una forma similar, también se ha subdividido la capa 6.

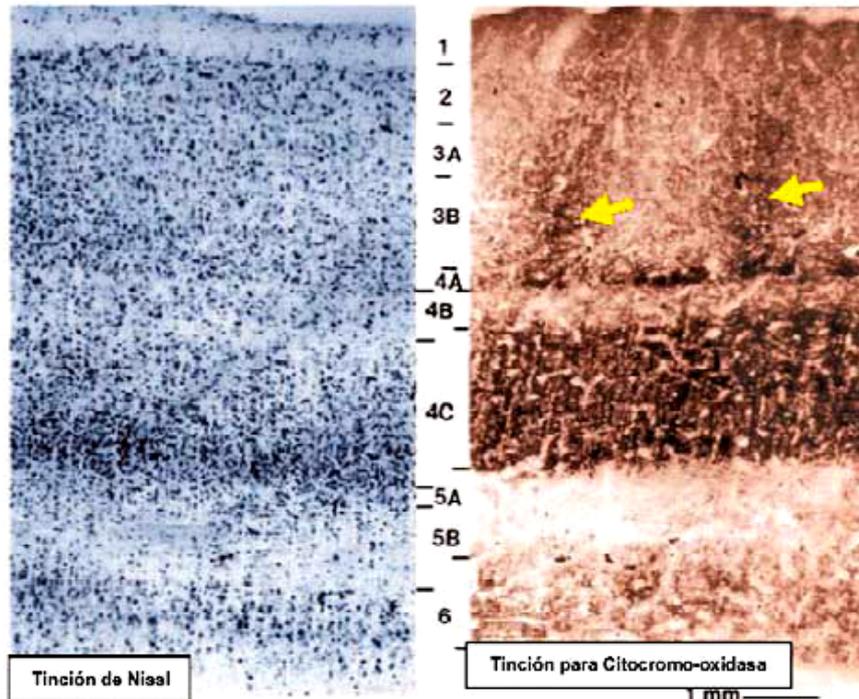


Figura 2.11: Estructura en capas de la corteza visual reveladas por distintas técnicas de tinción.

2.1.4.2 Las células de V1

La extensión del estudio de los campos receptivos, que había comenzado en la retina, a la corteza cerebral reveló algunas diferencias con respecto a los resultados obtenidos en las etapas anteriores. Se han distinguido principalmente dos clases de células con localización, morfología y funcionamiento bien diferenciados [Kandel 1991]:

A) Células sensibles a la orientación

Constituyen la mayoría de células en V1 y sus campos receptivos son alargados, con forma más o menos rectangular, presentando sensibilidad a la orientación del estímulo. Las neuronas estudiadas

responden vigorosamente a la presentación de barras, blancas o negras (no son sensibles al color), en el interior del campo receptivo, pero solamente cuando el estímulo tenía una orientación determinada. Y aún más: cada célula cortical respondía a una gama estrecha de orientaciones, lo que se dio en llamar la curva de orientación (o de sintonía a la orientación) de una determinada neurona. Esto es lo que se representa en la figura 2.12, en la cual, junto al preparado experimental clásico de los experimentos sobre el sistema visual, se muestra la respuesta de una célula de este tipo, ante distintas orientaciones y su curva de sintonía. Este tipo de células, a su vez podemos clasificarlas en dos categorías:

- *Células simples* (figura 2.13.A): tienen campos receptivos pequeños, alargados y bien definidos en subregiones on y off. La actividad espontánea de estas células o está ausente o es muy baja. Cuando una barra de luz se sitúa sobre el campo receptivo, puede producir un brote de actividad (respuesta on o de encendido) o un cese de su actividad (respuesta off o de apagado) dependiendo de la subregión en la que se haya colocado el estímulo.
 - *Células complejas* (figura 2.13.B): son más abundantes que las simples. Sus campos receptivos son rectangulares, mayores que los de las células simples, tienen una actividad espontánea viva, y se activan cuando una barra de luz, con la orientación apropiada, cae sobre el campo receptivo. Sin embargo, los campos receptivos de las células complejas no tienen subregiones on (excitatorias) y off (inhibitorias) bien delimitadas, como las células simples.
-

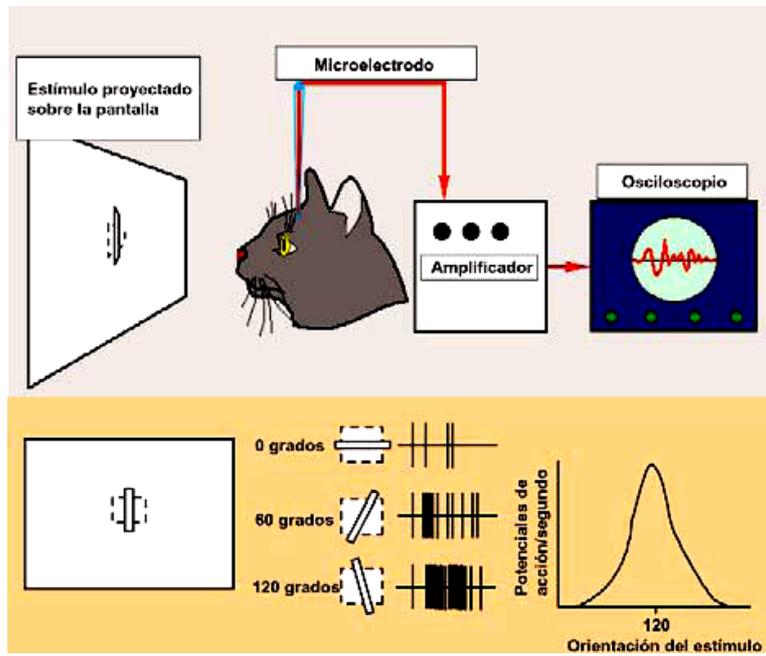


Figura 2.12: Experimento clásico (arriba) y respuesta de una célula de V1 (abajo).

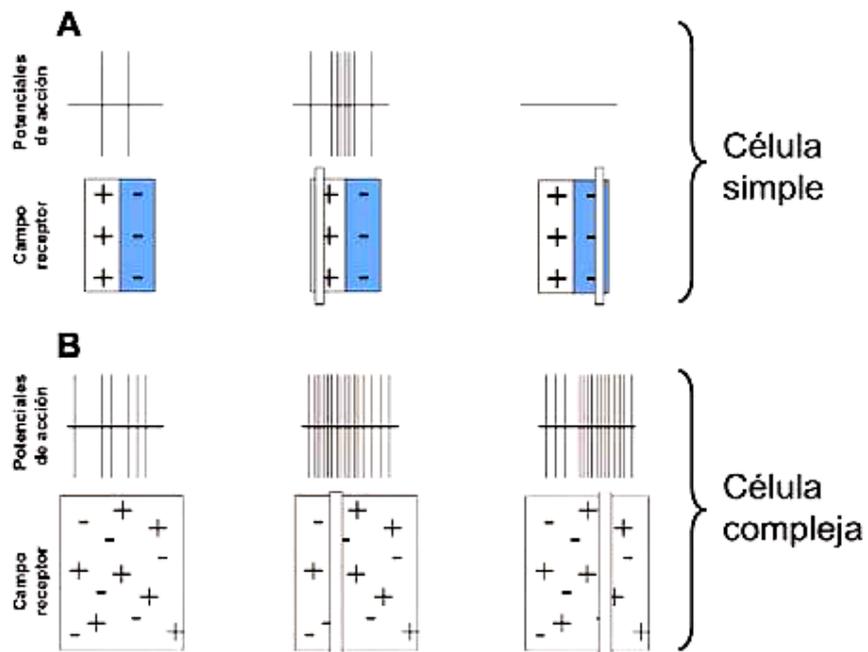


Figura 2.13: Respuestas de una célula simple (A) y una compleja (B).

Además de las propiedades descritas, también podemos encontrar neuronas que son sensibles a la dirección del movimiento, respondiendo, por ejemplo, hacia la derecha pero no hacia la izquierda. De igual manera la mayoría de las neuronas son sensibles a la velocidad con la que los estímulos se mueven sobre su campo receptivo. En la figura 2.14 se muestra un ejemplo de las propiedades mencionadas, detección de la velocidad y de la dirección.

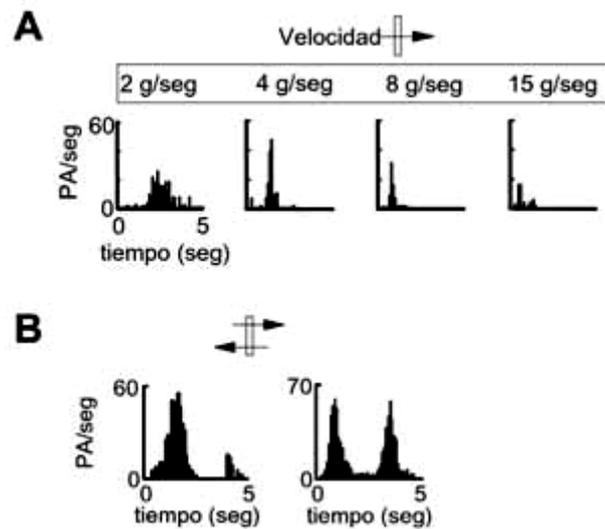


Figura 2.14: Respuestas de las células corticales según la velocidad y dirección del estímulo.

B) Células no orientadas

Este tipo de células presentan un campo receptivo circular concéntrico, como ocurría en etapas anteriores del camino visual, y presentan un comportamiento antagónico frente al color, aunque de una forma más elaborada, que se conoce como *dobles oposición*. Este concepto puede estudiarse con más detalle en [Gouras 1991].

2.1.4.3 Organización funcional de la corteza

Las células de la corteza visual cuyos campos receptivos tienen propiedades similares, tienden a agruparse. Este es el caso de la

sensibilidad a la orientación, el contraste cromático y a la respuesta celular dependiendo del ojo del que le llega la información. Existen otras organizaciones relativas a la dirección de los estímulos o a la sensibilidad a las frecuencias espaciales pero no las consideraremos aquí.

A) Organización en columnas de orientación

Al introducir un electrodo de registro perpendicularmente a la superficie de la corteza visual, se puede observar que todas las células que se encuentran, capa tras capa, tienen campos receptivos que responden a estímulos con la misma orientación (figura 2.15). Esta idea es consistente con la organización en columnas de la corteza cerebral que fue descrita por primera vez por Lorente de Nó, quien la llamó 'unidad elemental' [Lorente 1933]. La describió como una pequeña banda vertical perpendicular a la superficie cortical, donde están representados todos los elementos de la corteza. En 1957, Mountcastle describió esta organización en la corteza somestésica bajo el punto de vista funcional y la denominó 'columna' [Mountcastle 1957]; en 1962, Hubel y Wiesel la describieron en la corteza visual [Hubel 1988], y desde entonces se considera como una unidad organizativa fundamental del procesamiento de la información en la corteza cerebral. Cada banda de orientación corre perpendicularmente a la superficie cortical y paralelamente unas a otras en bandas de 400 micras de ancho. En realidad, y dado que tenemos que considerar la corteza como una estructura tridimensional, no hablaremos de bandas sino de columnas más o menos paralelepípedicas de unas 400 micras de lado. Si se tiene en cuenta lo anterior, se puede predecir que si en lugar de introducir un electrodo de registro perpendicular a la corteza lo hacemos de forma oblicua, a medida que avancemos tendremos que encontrar neuronas con orientaciones distintas, y en efecto eso es lo que ocurre. Si conseguimos que el electrodo avance tangencialmente, lo más horizontal posible, comprobaremos que aparecen representadas todas las orientaciones a medida que avanzamos (es decir, vamos atravesando columnas) y que, a una distancia de aproximadamente 1 mm, se repetirá la primera orientación que hayamos observado.

B) Columnas de dominancia ocular

Si nos fijamos en el ojo al que responden las células cuando se presenta un estímulo visual, también observaremos que dicha propiedad se distribuye en columnas, en este caso, columnas de dominancia ocular (DO). Si bien la mayoría de las neuronas de V1 responden a ambos ojos, la potencia relativa de las aferencias desde cada uno varía de una célula a otra. En los extremos de esta serie continua se encuentran neuronas que responden casi exclusivamente al ojo derecho o al izquierdo; en el centro

están las que responden por igual. La figura 2.15, junto con las columnas de orientación, muestra esquemáticamente la organización de las columnas de DO, para el ojo derecho y para el ojo izquierdo. Lógicamente existe una relación entre las columnas de dominancia ocular y las de orientación de manera que para cada ojo existirá todo un grupo de orientaciones a las que responderá.

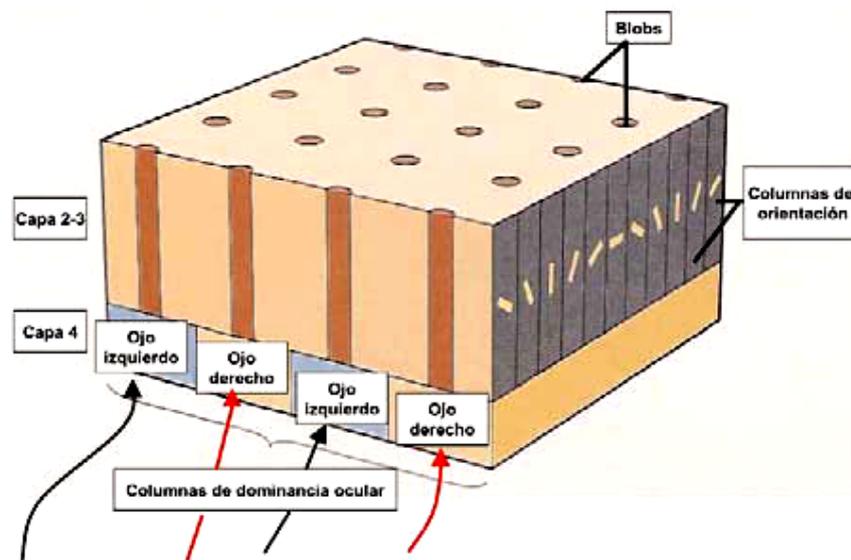


Figura 2.15: Representación de las columnas en la corteza cerebral.

C) Blobs

En la figura 2.15 también podemos distinguir unas estructuras cilíndricas entremezcladas con las columnas de orientación y de dominancia ocular, conocidas como *blobs*. Estas estructuras están formadas por la concentración de las células sin sensibilidad a la orientación, encargadas del procesamiento cromático que vimos en el apartado anterior.

2.1.5 Visión estéreo

La localización de los ojos en el humano dividen nuestro campo visual en tres zonas: una monocular para el ojo izquierdo, otra para el derecho y una porción, binocular, en la que se superpone la visión de ambos ojos y se percibe como una imagen única. Esta distribución sacrifica

la visión panorámica que poseen numerosos animales, pero como compensación poseemos visión estereoscópica, información muy importante para poder construir una visión en profundidad o tridimensional.

Observando la figura 2.16 podremos entender mejor todo el proceso de obtención de la información de profundidad. Supongamos que nuestros ojos convergen para fijarse un punto F (*punto de fijación*). La imagen de ese punto se recoge en zonas equivalentes de la retina de ambos ojos, y por ello hay una imagen única. Mientras fijamos el punto anterior consideremos ahora lo que ocurre con otro, P, que está un poco más lejos y desplazado ligeramente hacia un lado. La imagen de este nuevo punto se recoge en zonas distintas de la retina, esto es lo que se conoce como *disparidad* retiniana. Existen disparidades horizontales y verticales, pero las más relevantes para la visión en profundidad son las primeras.

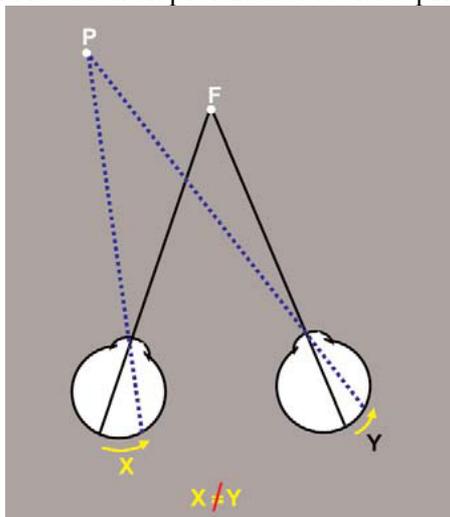


Figura 2.16: Proyección en la retina de dos puntos del espacio visual.

En nuestro cerebro, la disparidad se interpreta como componente en profundidad del objeto observado. La *estereopsis* se produce al integrarse en el cerebro dos imágenes que presentan disparidades, creando la sensación de profundidad. Los diferentes ángulos de proyección de un objeto sobre los dos ojos, nos permiten calcular distancias; pero además es necesario, para que haya *visión binocular* y estereopsis, que las imágenes sean suficientemente nítidas.

Si atendemos a los aspectos geométricos de la visión binocular, podemos hacer unas observaciones muy interesantes. En primer lugar, en el espacio que podemos ver existirán siempre unos lugares en que al situar

un punto, éste caerá siempre en zonas correspondientes de las dos retinas, es decir, sin disparidad o con disparidad cero. Se representa como una circunferencia, aunque en realidad sería una superficie esférica, y se conoce como círculo de Vieth-Müller (figura 2.17.A).

No obstante, como el ojo no es totalmente esférico, se comprobó que el círculo de Vieth-Müller no es exactamente un círculo, y se pudo determinar con exactitud el lugar del espacio que tiene disparidad cero. Esto es lo que se llama horóptero (figura 2.17.B). Así podemos distinguir tres zonas del campo visual, atendiendo a las posibles disparidades que se pueden medir: disparidad cero, disparidades positivas que son las que se forman por detrás del horóptero, y disparidades negativas, las producidas por puntos localizados entre el horóptero y el observador.

Al intentar determinar experimentalmente el horóptero real se pudo delimitar, en profundidad, una cierta zona dentro de la cual tiene lugar la fusión binocular; por fuera de esta zona, por delante o por detrás, aparece diplopía. Esta zona de fusión se llama zona de Panum (figura 2.17.B), y los puntos que caigan en esta zona siempre se verán como únicos (aunque presenten disparidad), los que caigan fuera de ella se verán dobles.

De todas formas esta diplopía (visión doble) fisiológica no produce molestias dado que existen mecanismos como el de la atención selectiva a una zona de la imagen que hacen que no sea percibida.

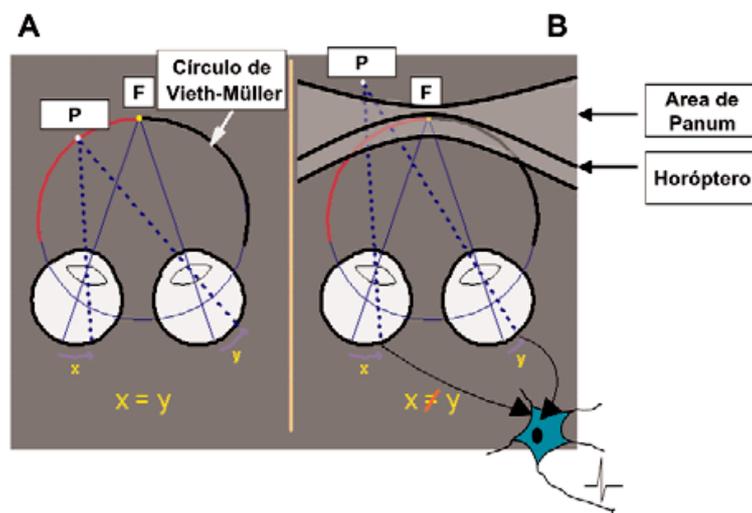


Figura 2.17: Círculo de Vieth-Müller (A), horóptero y área de Panum (B).

2.1.5.1 Neuronas sensibles a la disparidad

En la corteza visual, áreas V1, V2 y V3 (incluso en V4 y V5) existen neuronas sensibles a las disparidades retinianas. Los primeros procesos de información estereoscópica dependen de tres sistemas funcionales principales de neuronas: El primero de ellos está formado por células que detectan las disparidades dentro del área de Panum elevando su frecuencia de descarga (sintonizadas excitatoriamente). Algunas detectarían las disparidades positivas o lejanas, otras las negativas o cercanas, y un tercer subgrupo se encargará de señalar aquellos objetos que se localizan a disparidad cero. El segundo sistema está formado por neuronas que señalan los objetos dentro del campo visual exclusivamente con disparidad cero, pero se comportan al revés, es decir, disminuyen su frecuencia de descarga, por lo que se conocen como células sintonizadas inhibitoriamente. El último sistema lo integran aquellas neuronas que responden de forma excitatoria para cualquier disparidad, ya sea positiva o negativa.

Ahora pasamos a comentar diversos mecanismos que ayudan a obtener información útil sobre la percepción tridimensional de nuestro entorno.

Aun en ausencia de visión binocular, podemos tener información acerca de la relación de los objetos entre sí y de su posición en el espacio. Algunos de los mecanismos monoculares que utilizamos son:

- Interposición. Es el ocultamiento parcial de un objeto por otro. Origina la clara impresión de que el objeto parcialmente oculto está más lejos. Depende de la percepción de contornos continuos y de la detección de sus puntos de intersección.
 - Señales basadas en las sombras. Éstas proyectan los objetos y son una pista evidente del espacio que se prolonga más allá del objeto y que se utiliza en la percepción de profundidad.
 - Perspectiva lineal. La representación gráfica de profundidad por perspectiva lineal, muy conocida por los pintores, es otra pista monocular. Cuando existen líneas paralelas en perspectiva lineal, cuanto más lejanas están, mayor es su convergencia, hasta que a una distancia infinita se encuentran en el llamado 'punto de fuga'.
-

- Perspectiva aérea o atmosférica. Cuando se observa un objeto distante, se percibe menos nítido y más azulado que los cercanos a causa de las partículas suspendidas en la atmósfera entre el objeto y el observador.
- Tamaño relativo. Los objetos que presentan un tamaño más pequeño en relación con otro hacen que parezcan más lejanos; y, viceversa, si el objeto tiene un mayor tamaño relativo, parece más próximo.
- Tamaño familiar. La experiencia que se ha tenido con determinados objetos ejerce influencia en la percepción de la distancia.
- Altura relativa. Además de que los objetos lejanos ocupan una menor porción del campo de visión, también se ubican en las porciones más elevadas del campo visual. Esto es válido para los objetos que se sitúan por debajo de la línea del horizonte, pero no para los situados por encima de la línea. En este último caso es al contrario: parece más cercano al observador cuanto más elevado esté en el campo visual.
- Paralaje de movimiento. Esta pista interviene cuando se observan objetos mientras la persona está en movimiento, por ejemplo cuando se viaja en automóvil. Los más cercanos pasan en el campo de visión más rápidamente que los lejanos. Los movimientos de la cabeza con respecto a objetos estacionarios también son muy útiles para la discriminación en profundidad. Por ejemplo, al movernos con respecto a objetos que están muy cerca, la dirección de éstos cambia mucho, pero cuando nos movemos con respecto a objetos situados a mucha distancia, su dirección cambia muy poco.
- Gradiente de textura. Cuando observamos la textura de una superficie, los elementos más cercanos parecen estar más espaciados, aumentando su densidad con la distancia.

Además existen señales oculomotoras que nos ayudan a la visión estéreo. La convergencia de los globos oculares y la acomodación (modificación de la curvatura del cristalino y enfoque) son dos fuentes potenciales de señales para estimar las relaciones espaciales de los objetos localizados en nuestro entorno, ya que dichas relaciones están condicionadas por la distancia desde el observador al punto de fijación. La

idea es que el esfuerzo de acomodación o el esfuerzo de convergencia pueden generar percepción en profundidad. La primera es una referencia monocular, y la segunda, binocular.

El grado de vergencia ocular es una referencia importante para la visión estereoscópica. Las disparidades retinianas por sí solas no contienen la información completa para el cálculo de esas distancias, ya que las disparidades se modifican con el punto de fijación. Por ejemplo, dos objetos separados 1 cm tendrán una disparidad retiniana menor si son vistos a una distancia de 2 m que si son vistos a una distancia de 30 cm. Por lo tanto, la estimación de la distancia real entre los objetos debería de estar determinada por el grado de vergencia ocular o por la acomodación o por ambas. Existen evidencias que indican que el cambio de la distancia de visión produce cambios en la actividad de neuronas de la corteza visual sensibles a la disparidad retiniana.

2.2 Patologías que afectan al sistema visual

Cuando hablamos de una persona con una discapacidad visual, puede tratarse de una persona ciega (o con *ceguera*) o una persona con disminución visual (o *baja visión*).

Para hablar con propiedad de este tema necesitamos conocer previamente algunos conceptos, como son:

- *Agudeza visual*: concepto referido a la 'calidad' de la visión del sujeto. Se toma como referencia la percepción de una persona con agudeza máxima, de objetos situados a 6m, y se compara con la del sujeto en cuestión. Una agudeza visual 20/20 (ó 1) indica que podemos ver los objetos a 6m con la misma calidad que una persona con visión máxima, mientras que una agudeza de 20/200 (ó 0.1) indica que observamos los objetos a 6m tal como lo haría una persona con visión máxima a 60m.
 - *Campo visual*: hace referencia a la 'cantidad' de visión. El campo visual máximo es de 90° en la parte temporal, 60° en la nasal, 50° en la superior y 70° en la inferior.
-

Legalmente se define la **ceguera** como una agudeza visual inferior a 20/200, incluso después de una corrección con lentes. La **baja visión** es considerada con agudeza visual máxima de 3/10 (0,3) y/o un campo visual menos o igual a 20°.

La ceguera puede producirla cualquiera de las siguientes razones:

- La luz no llega a la retina
- Los rayos de luz no se concentran correctamente sobre la retina
- Falla la transducción a señales eléctricas de la luz incidente en la retina
- Los impulsos nerviosos de la retina no son transmitidos al cerebro normalmente
- El cerebro no puede interpretar la información enviada por el ojo

A continuación vamos a ver algunos de los trastornos que pueden causar estos problemas, centrándonos en aquéllos no tratables quirúrgicamente.

2.2.1 Trastornos de la retina

Como hemos detallado en secciones anteriores, la retina es la membrana sensible a la luz localizada en la superficie interna de la parte posterior del ojo. Los axones de las células ganglionares de la retina se agrupan en el denominado nervio óptico. El área central de la retina, llamada mácula, contiene la mayor densidad de nervios sensibles a la luz y en consecuencia produce la mejor resolución visual. La vena y la arteria retinianas llegan a la retina cerca del nervio donde se ramifican siguiendo el recorrido de los nervios. Al igual que el nervio óptico, la retina también cuenta con una importante cantidad de vasos que llevan sangre y oxígeno.

La córnea y el cristalino, que se hallan cerca de la parte más anterior del ojo, concentran la luz sobre la retina. Luego, las ramificaciones del nervio óptico captan las señales de luz y el nervio óptico las transmite hacia el cerebro, donde se interpretan como imágenes visuales.

2.2.1.1 Degeneración macular

La degeneración macular es una enfermedad en la cual la mácula, la parte central y más vital de la retina, se deteriora. La enfermedad afecta a los ancianos, es igualmente común entre los varones y las mujeres, y es más frecuente en las personas de piel blanca que en las de piel negra. La causa es desconocida, pero la enfermedad tiende a aparecer en el seno de una misma familia.

La degeneración macular produce una pérdida indolora de la visión. Ocasionalmente, el primer síntoma es una distorsión en un solo ojo, por el que las líneas delgadas y rectas parecen sinuosas. A veces el médico puede notar claramente ciertos cambios físicos cerca de la mácula incluso antes de que aparezcan los síntomas. La degeneración macular puede dañar gravemente la visión, pero rara vez conduce a la ceguera total. La visión en los bordes externos del campo visual (visión periférica) y la capacidad para discernir el color no resultan afectadas (véase figura 2.18).



Figura 2.18: Visión de pacientes que presentan degeneración macular.

Prácticamente no existe tratamiento para la degeneración macular. Sin embargo, cuando crecen nuevos vasos sanguíneos dentro o alrededor de la mácula, la fotocoagulación con láser en ocasiones puede destruirlos antes de que causen más daño.

2.2.1.2 Desprendimiento de retina

El desprendimiento de retina es la separación de la retina de la parte posterior que la sostiene. La red de nervios que conforman la parte de la retina sensible a la luz forma una delgada película que se adhiere firmemente al tejido que lo sostiene por debajo. Cuando estas dos capas se separan, la retina no puede funcionar y, a menos que se las vuelva a unir, puede resultar dañada para siempre.

El desprendimiento puede comenzar en un área reducida, pero si no se trata, toda la retina puede desprenderse. En una variedad de desprendimiento, la retina literalmente se rompe. Esta variedad en general se presenta en personas miopes o que han sido sometidas a una operación de cataratas o han sufrido una lesión ocular. En otra variedad, la retina no se rompe pero se separa del tejido que tiene debajo. La retina se separa cuando el movimiento del líquido dentro del ojo tira de la retina o bien cuando el fluido se introduce entre la retina y el tejido subyacente y las separa, arrancando la retina.



Figura 2.19: Visión normal (izquierda), y visión con desprendimiento parcial de retina (derecha).

El desprendimiento de retina es indoloro pero puede crear imágenes de formas irregulares flotando o destellos luminosos y puede nublar la visión. La pérdida de la visión comienza en una parte del campo visual y a medida que avanza el desprendimiento la pérdida de visión se extiende (ver figura 2.19). Si el área macular de la retina se desprende, la visión se deteriora rápidamente y todo se torna borroso.

El desprendimiento puede ser detectado por el oftalmólogo mediante observación directa con el oftalmoscopio, y en cualquier caso, con una prueba de ultrasonidos.

2.2.1.3 Retinosis pigmentaria

La retinosis pigmentaria es una rara alteración hereditaria en la que la retina degenera de forma lenta y progresiva, conduciendo finalmente a la ceguera. Algunas formas de retinosis pigmentaria son dominantes, y requieren de un solo gen de cualquiera de los padres; otras dependen del cromosoma X, y por tanto necesitan un solo gen de la madre. En algunas personas, en su mayoría varones, también aparece asociada una forma hereditaria de pérdida de la audición.

Los bastones, que son responsables de la visión cuando hay poca luz, gradualmente degeneran, por lo que la visión empeora en la oscuridad. Los primeros síntomas suelen comenzar en la primera infancia. Con el paso del tiempo, tiene lugar una pérdida progresiva de la visión periférica. En las últimas etapas de la enfermedad, la persona cuenta con una pequeña área de visión central y un pequeño resto de visión periférica (*visión túnel*).

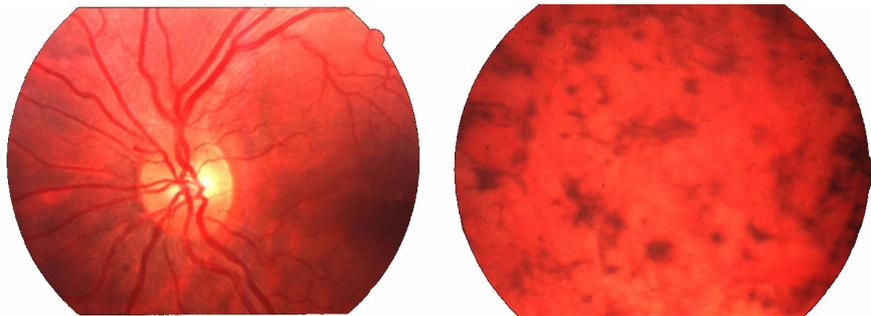


Figura 2.20: Ojo normal (izquierda) y afectado por retinosis pigmentaria (derecha).

Al examinar la retina con un oftalmoscopio, el médico nota cambios específicos que sugieren el diagnóstico (ver figura 2.20). Varias pruebas pueden ayudar a determinar el diagnóstico y el examen de los miembros de la familia puede establecer la modalidad de la herencia. Ningún tratamiento puede detener la progresión del daño a la retina.

2.2.2 Glaucoma

El glaucoma es un trastorno en el que aumenta la presión dentro del globo ocular, dañando el nervio óptico y causando pérdida de visión.

Tanto la cámara anterior como la posterior del ojo están llenas de un fluido llamado humor acuoso. Normalmente, el fluido es producido en la cámara posterior, pasa por la pupila hacia la cámara anterior y luego sale del ojo a través de unos canales específicos (canales de salida). Si la corriente de fluido resulta interrumpida, generalmente debido a una obstrucción que evita que el fluido salga fuera de la cámara anterior, la presión aumenta.

En general, el glaucoma no tiene una causa conocida; sin embargo, en ocasiones afecta a miembros de una misma familia. Si los canales de salida están abiertos, el trastorno recibe el nombre de glaucoma de ángulo abierto. Si los canales están bloqueados por el iris, la enfermedad se denomina glaucoma de ángulo cerrado.



Figura 2.21: Campo visual normal (izquierda) y afectado por glaucoma (derecha). En las figuras de la parte inferior se representan las zonas del campo visual con percepción deficiente. En el ojo normal se reduce al punto ciego.

El glaucoma produce una pérdida de la visión periférica o puntos ciegos en el campo visual (figura 2.21). Para determinar si dichos puntos ciegos existen, el especialista pide a la persona que mire de frente hacia un punto central y le indique cuándo puede ver luz. La prueba puede ser

llevada a cabo usando tanto una pantalla y un puntero como un instrumento automático que usa puntos de luz.

2.2.2.1 Glaucoma de ángulo abierto

En el glaucoma de ángulo abierto, el fluido drena demasiado lentamente desde la cámara anterior. La presión se eleva gradualmente (casi siempre en ambos ojos) lesionando el nervio óptico y causando una lenta pero progresiva pérdida de la visión. La pérdida de visión comienza en los extremos del campo visual y, si no se trata, acaba extendiéndose por todo el resto del campo visual y finalmente produce ceguera.

Al principio, la mayor presión ocular no produce ningún síntoma. Los síntomas posteriores pueden incluir un estrechamiento de la visión periférica, ligeros dolores de cabeza y sutiles trastornos visuales, como ver halos alrededor de la luz eléctrica o tener dificultad para adaptarse a la oscuridad. Finalmente, la persona puede acabar presentando visión túnel.

2.2.2.2 Glaucoma de ángulo cerrado

El glaucoma de ángulo cerrado provoca ataques súbitos de aumento de presión, en general en un ojo. En las personas que padecen esta enfermedad, el espacio entre la córnea y el iris (por donde sale el fluido fuera del ojo) es más estrecho de lo normal. Cualquier factor que provoque la dilatación de la pupila (una escasa iluminación, las gotas oftálmicas indicadas para dilatar la pupila antes de un examen ocular o ciertas medicaciones orales o inyectadas) puede hacer que el iris bloquee el drenaje de fluido. Cuando ello sucede, la presión intraocular aumenta de improviso.

Un episodio de glaucoma de ángulo cerrado agudo produce síntomas repentinos. Puede provocar un ligero empeoramiento de la visión, halos de color alrededor de las luces y dolor en el ojo y la cabeza. Estos síntomas pueden durar sólo unas pocas horas antes de que tenga lugar un ataque más grave. Éste produce una rápida pérdida de la visión y un repentino y agudo dolor pulsátil en el ojo. El párpado se hincha y el ojo se torna lloroso y rojo. La pupila se dilata y no se cierra normalmente en respuesta a la luz intensa.

2.2.3 Incidencia en la población

Según un informe de la O.N.C.E. sobre sus afiliados en 2003 (que eran un total de 64404 afectados), aproximadamente el 76% correspondía a individuos cuya deficiencia legalmente fue catalogada como baja visión, mientras que el 24% restante padecían ceguera total. La incidencia dependiendo del sexo es más igualada, afectando casi por igual (la diferencia está en torno a un 0.2%) a hombres y a mujeres. Por grupos de edad es significativo que el 40% de afectados superan los 65 años, mientras que sólo el 8% son menores de 18. Es de esperar, por tanto, que en las sociedades con un progresivo envejecimiento, como es nuestro caso, en un futuro próximo crezca el porcentaje de afectados.

Si atendemos a patologías concretas, algo más de un 35% de sus afiliados sufren alguna patología que afecta a la retina (por ejemplo, el 3.39% sufre desprendimiento de retina, y el 8.08% sufre de retinopatía diabética), un 14.5% tienen alguna afección en el nervio óptico; en un 3.5% de los casos el origen de la deficiencia se encuentra en el glaucoma.

Datos procedentes de Estados Unidos ([AFB 2003], [NFB 2003]), nos informan que en ese país son más de 1.3 millones los individuos con alguna deficiencia visual, llegando a 130000 los afectados por ceguera total. También nos informan de que aproximadamente uno de cada 3500 americanos sufre retinosis pigmentaria, llegando a un total de 1.5 millones de afectados en todo el mundo. En cuando a la degeneración macular asociada a la edad, los datos son que cada año se diagnostican 700000 nuevos casos, y 70000 evolucionan hasta la ceguera total.

2.3 Sistemas para la rehabilitación visual

Muchos son los modelos, dispositivos y procedimientos que se han desarrollado, y se están desarrollando, para intentar paliar de una u otra forma las distintas deficiencias visuales.

Todos estos sistemas pueden ser clasificados en dos grandes grupos, atendiendo a la forma en que interaccionan con el paciente. Como veremos con más detalle, muchas son las diferencias entre estos dos grupos

de sistemas, tanto por el modo de funcionamiento, los requisitos que se le exigen, etc. Una de las características que deben compartir todos es el bajo consumo y la portabilidad, ya que serán sistemas que deberá usar de forma regular el paciente.

A continuación veremos con más detalle los sistemas de ambos tipos.

2.3.1 Neuroprótesis visuales

Se trata de sistemas conectados directamente a terminaciones nerviosas del sistema visual del paciente, que estimulan las células correspondientes mediante descargas eléctricas. Los experimentos de estimulación en corteza cerebral [Brindley 1968], dieron como resultado la percepción por parte del paciente de destellos luminosos que fueron denominados *fosfenos*. Este fue el punto de partida para la investigación de la neuroestimulación visual.

Dependiendo de la patología que afecte al paciente, la disfunción visual estará localizada en un punto diferente de su tracto visual, siendo necesario colocar la prótesis en un nivel superior al que esté afectado. Como hemos visto, las características de las señales en cada punto del camino visual son distintas, por lo que el procesamiento de la información visual y la traducción en señales eléctricas que haga el sistema de rehabilitación, será distinto en la retina del que se haga en el nervio óptico o la corteza cerebral. De esta forma, podemos hacer una clasificación de las neuroprótesis visuales atendiendo al punto en el que se produzca la interfaz con el tejido nervioso.

Dado que las neuroprótesis visuales estimulan tejido neuronal, las características de las señales eléctricas que generen deben ser lo más parecidas posibles a las señales que viajan a través del sistema nervioso (fundamentalmente trenes de impulsos); por otro lado, el medio de comunicación de estas señales debe ser biocompatible (materiales que no deterioren el tejido en periodos largos de tiempo, durante años).

Todos los sistemas neuroprotésicos visuales están actualmente en fase de desarrollo, habiéndose experimentado en algunos casos en implantes agudos y crónicos con humanos.

2.3.1.1 Prótesis retinianas

Cuando el problema del paciente está localizado en alguna de las estructuras del globo ocular, pero permaneciendo operativa, al menos en parte, la retina, una alternativa sería implantar una prótesis en ésta.

Dependiendo del resto funcional de la retina, podemos distinguir dos tipos de prótesis:

A) Sub-retiniana

Este tipo de prótesis está indicada cuando el problema que afecta al paciente es una degeneración de los fotorreceptores (retinosis pigmentaria, degeneración macular,...), haciendo que el sistema conecte con las primeras capas neuronales de la retina, como las células bipolares.

Su función será la de traducir la luz incidente en la prótesis en señales eléctricas. El método de estimulación es mediante un array de micro-fotodiodos (figura 2.22), con microelectrodos encargados de transmitir la corriente eléctrica.

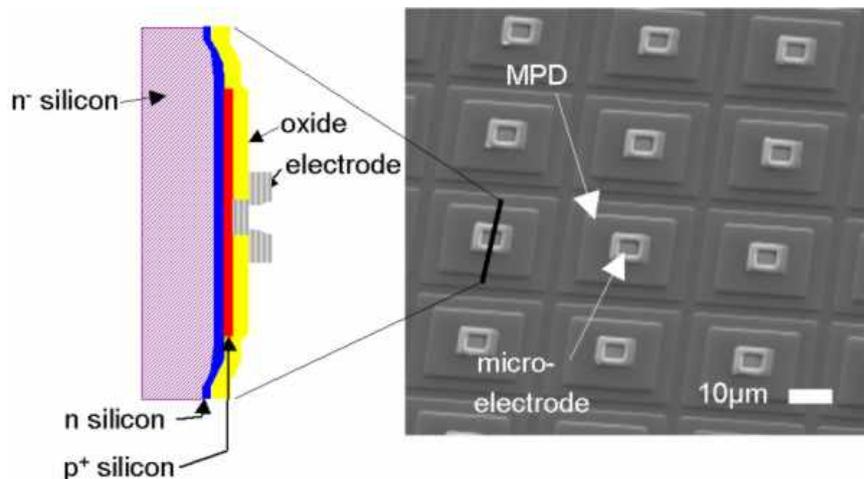


Figura 2.22: Visión esquemática de un micro-fotodiodo (izquierda) y fotografía de un array conteniendo varios (derecha) [Zrenner 2001].

Diversos grupos en todo el mundo se han dedicado a la investigación sobre los implantes sub-retinianos, como es el caso de Zrenner, que ha presentado resultados de experimentos in-vitro [Zrenner

2000]. El equipo de Chow ha ido más allá, llegando a implantar animales con prótesis autónomas [Chow 2001] [Chow 2002] (sin necesidad de alimentación externa).

Las principales ventajas de estas prótesis radican en que aprovechan todo el procesamiento de la retina, y mantienen la organización retinotópica del sistema visual. Por el contrario, el acceso quirúrgico para colocar el implante es muy complicado, y existen problemas para una buena recepción de luz en ese punto.

B) Epi-retiniana

En el caso de que la prótesis esté colocada en el otro extremo del tejido nervioso de la retina, contactando con la capa de células ganglionares, estaremos hablando de una prótesis epi-retiniana.

La figura 2.23 muestra de forma esquemática la localización de un implante epi-retiniano dentro del globo ocular.

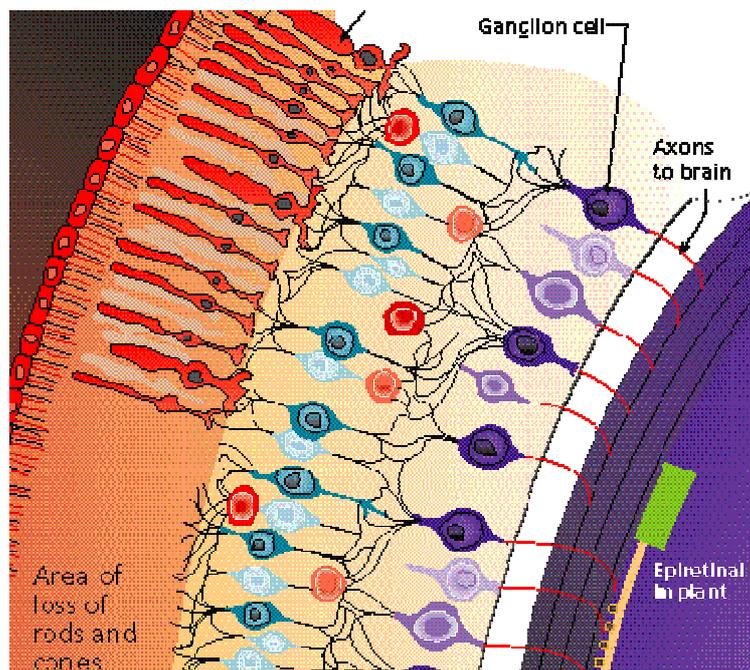


Figura 2.23: Localización de un implante epi-retiniano [Humayun 1994].

El implante está formado por un elemento externo que se encarga de capturar la imagen a transmitir, procesar la imagen de una forma

similar a como lo haría la retina y enviar vía radio la información resultante al implante interno [Eckmiller 1997].

De la misma forma que las sub-retinianas, el acceso para la cirugía es muy complicado, sumándose la dificultad para fijar el implante en la zona deseada. Los principales grupos investigando este tipo de implantes son el de Humayun, el de Eckmiller y el de los profesores Rizzo y Wyatt [Rizzo 1999].

Humayun ha seguido realizando experimentos, como el de 1999 ([Humayun 1999]), en el que implantó a nueve pacientes con retinosis pigmentaria y uno con degeneración macular, en sus etapas finales. Hizo pruebas con arrays de microelectrodos de tamaño 3x3 y 5x5, llegando uno de los pacientes a *percibir* una letra ('H' en este caso), mientras que las sensaciones que se evocaron en los demás pacientes, fueron formas más simples, como puntos (con diferentes intensidades).

2.3.1.2 Prótesis para el nervio óptico

El grupo que más esfuerzos ha dedicado al desarrollo de un implante visual que actuara a nivel del nervio óptico ha sido el del profesor Veraart (Bélgica). Un esquema de cómo quedaría instalada la prótesis puede verse en la figura 2.24.

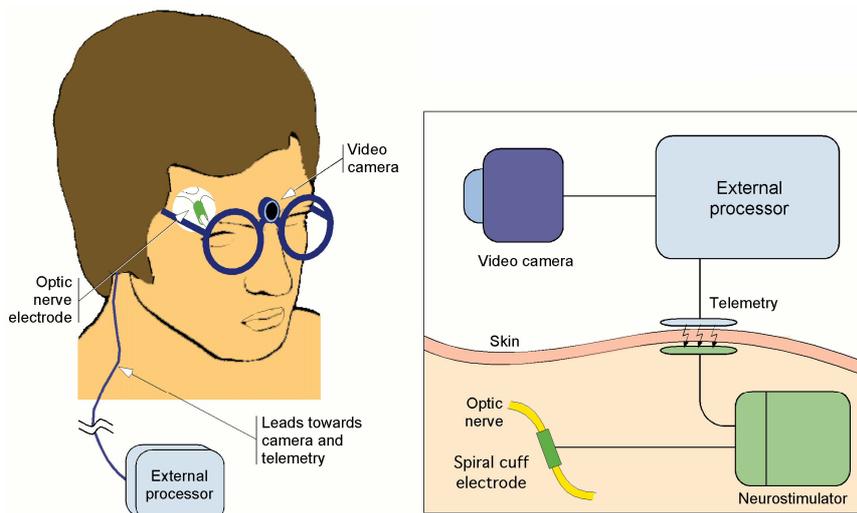


Figura 2.24: Prótesis para el nervio óptico [Veraart 1998].

Como podemos ver en el esquema, se trata también de un sistema externo de captura y preprocesamiento que comunica a la parte interna la información necesaria para producir la estimulación. La corriente eléctrica es inducida mediante contactos ubicados en un electrodo en forma de espiral que recubre el nervio óptico.

Este sistema, por su diseño, tiene una baja resolución para localizar el estímulo; es decir, es muy difícil de focalizar en un punto determinado la corriente generada por el estimulador. Otra desventaja es la dificultad en la cirugía para colocar el implante.

Los resultados obtenidos han dejado patente que para que se pueda recuperar una forma útil de visión, los implantes de nervio óptico tendrían que basarse en otro tipo de electrodos para realizar la estimulación.

2.3.1.3 Prótesis corticales

La principal ventaja de este tipo de sistemas es que pueden utilizarse para casi cualquiera de las patologías causantes de la ceguera, ya que sólo requiere que conserve su funcionalidad la corteza visual primaria, para poder estimularla mediante una matriz de electrodos. Además el acceso quirúrgico es más fácil que en los dos casos anteriores.

No obstante, al realizar la estimulación a un nivel superior, eliminando el procesamiento de la retina y del LGN, debiéramos incorporar esta tarea al sistema de procesamiento de las imágenes capturadas, antes de transmitir las al estimulador.

En este apartado veremos con más detalle las distintas aproximaciones que existen para la estimulación cortical, ya que parte del trabajo que se presenta se enmarca dentro del ámbito de un proyecto europeo sobre este tipo de implantes. Podemos distinguir dos tipos de implantes corticales, dependiendo de la forma de realizar la estimulación: con electrodos superficiales sobre la corteza, o con electrodos penetrantes en la corteza cerebral (intracortical). Veamos sus características:

A) Electrodos superficiales

El grupo de William Dobbelle desarrolló un array de electrodos superficiales (figura 2.25) en la universidad de Utah. Esta estructura plana

de silicona albergaba 64 electrodos hechos de platino, y de ella sale todo el cableado necesario para hacer llegar las señales a los electrodos.

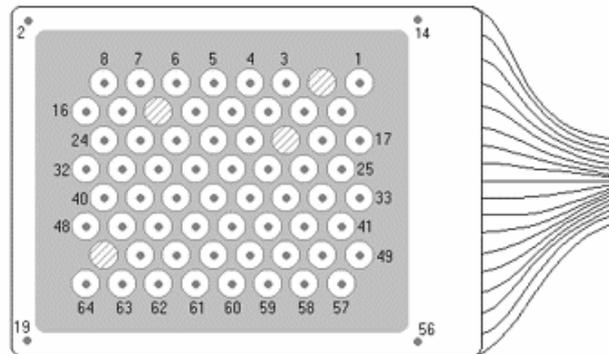


Figura 2.25: Matriz de electrodos planos de Dobbelle [Dobbelle 2000].

Una cámara captura imágenes del entorno, que son preprocesadas (por ejemplo, una extracción de bordes), se detectan los electrodos que deben ser activados, y se envían los trenes de impulsos necesarios a cada uno de ellos.

Para poder realizar convenientemente la estimulación, son necesarias grandes cantidades de corriente eléctrica, ya que los electrodos se encuentran sobre la corteza, sin contactar con la misma, generando la corriente de estimulación por inducción. Estos niveles elevados de corriente pueden ser peligrosos, llegando a producir ataques en el paciente implantado, por lo que la estimulación 'simultánea' de electrodos puede ser prohibitiva.

Otras deficiencias detectadas en este implante son el reducido número de electrodos disponibles, haciendo imposible evocar patrones de una resolución mayor; la distancia mínima a la que se pueden evocar dos fosfenos contiguos es demasiado elevada, lo que dificulta la percepción de formas continuas.

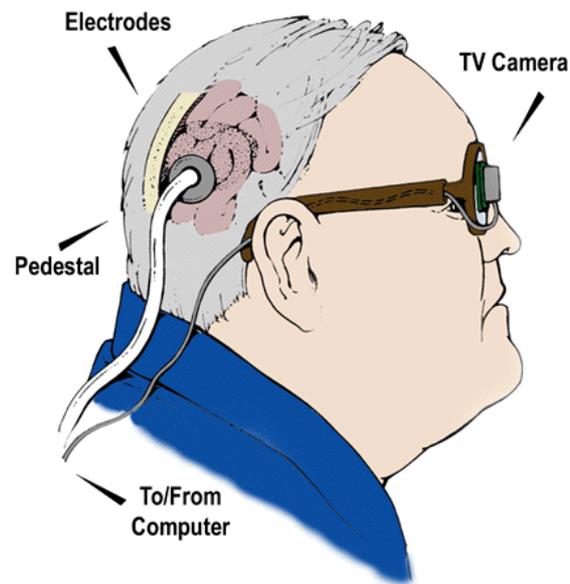


Figura 2.26: Esquema del sistema completo de Dobbins y foto de uno de los invidentes implantados (desde 1974).

Socialmente tampoco ha tenido mucha aceptación, principalmente por la estética de la parte externa del implante (figura 2.26). Como se

puede apreciar el paciente lleva un manajo de cables que enlaza con un conector de tipo pedestal fijado al hueso del cráneo, atravesándolo. Además se ha visto que este sistema de fijación puede presentar problemas de asepsia. No obstante, los trabajos de Dobelle, iniciados en los años 70, han servido para demostrar la viabilidad de este tipo de implantes a largo plazo (décadas), el desarrollo de procedimientos quirúrgicos y de configuración y ajuste de la prótesis, así como su eficiencia en la generación de fosfenos de manera estable en el tiempo.

B) Electrodo penetrantes

Este tipo de electrodos hacen contacto con la corteza, de forma que su extremo alcance la capa 4-A de V1, donde se realiza la estimulación. Al contactar muy próximos a los cuerpos celulares, la corriente que hay que inyectar para evocar fosfenos es mucho menor que en el caso de los implantes superficiales. De esta forma es menor el riesgo de que el paciente implantado sufra ataques.

Otra característica positiva de las matrices de microelectrodos penetrantes es que se consiguen diseñar con una separación menor entre ellos, aumentando la resolución espacial, posibilitando una percepción mejor de los patrones complejos. Asimismo, dado un tamaño determinado del implante, podremos integrar un mayor número de canales de estimulación en el mismo.

Tres son los grupos más importantes que han dedicado sus esfuerzos al diseño y fabricación de implantes de este tipo. A continuación los describiremos con más detalle.

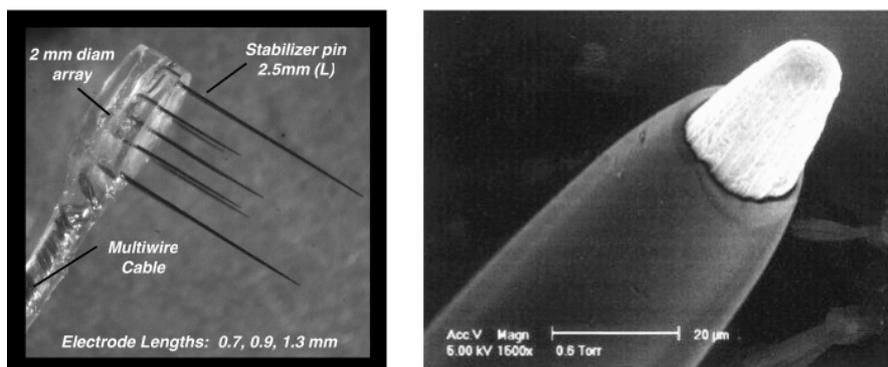


Figura 2.27: Array de microelectrodos de Troyk (izquierda) y detalle del extremo de un electrodo (derecha) [Troyk 2003].

El grupo de **Troyk** ha desarrollado matrices de 8 a 16 electrodos, cuya longitud es variable (ver figura 2.27). Como podemos observar, además de los electrodos de estimulación, el implante consta de unos terminales más largos cuya función es la de estabilizar el array en su posición.

Estos implantes han sido probados en primates, con el fin de comprobar las características de estabilidad, biocompatibilidad y capacidad de estimulación, aunque la validez de los resultados de experimentos realizados con animales siempre ha sido muy cuestionada.

Una desventaja de este implante es que al colocar varias matrices en la corteza, no siguen una estructura regular, lo que dificulta la cobertura uniforme de todo el campo visual.

Otro grupo que trabaja con implantes intracorticales es el de **Schmidt**. Su experiencia con un invidente permitió la caracterización de los diversos parámetros relacionados con la estimulación cortical en humanos. En [*Schmidt 1996*] se describe cómo se mantuvo durante 4 meses implantada a una mujer de 42 años, que llevaba 22 años con ceguera total generada por glaucoma. Se consiguió evocar fosfenos con 34 de los 38 electrodos que había implantado.

La figura 2.28 nos muestra el tipo de electrodos con los que realizó los experimentos. Como podemos ver se trata de electrodos duales, de 2 mm. de longitud.

El grupo de **Normann** (Universidad de Utah) también se dedica al desarrollo de matrices de electrodos intracorticales. Obtuvieron un dispositivo de 4x4 mm², que alberga 100 microelectrodos [*Campbell 1991*]. La separación entre electrodos es de 400 μm, que se corresponde con la distancia entre las columnas corticales; y la longitud de los electrodos es de 1.5 mm, distancia desde la corteza a la que se encuentran, aproximadamente, los cuerpos celulares de las neuronas de V1.

El implante está descrito en [*Normann 1999*], junto con el procedimiento y materiales para su colocación (incluyendo el conector transcraneal que sirve como interfaz con el exterior del cerebro), y resultados obtenidos de experimentos agudos y crónicos.

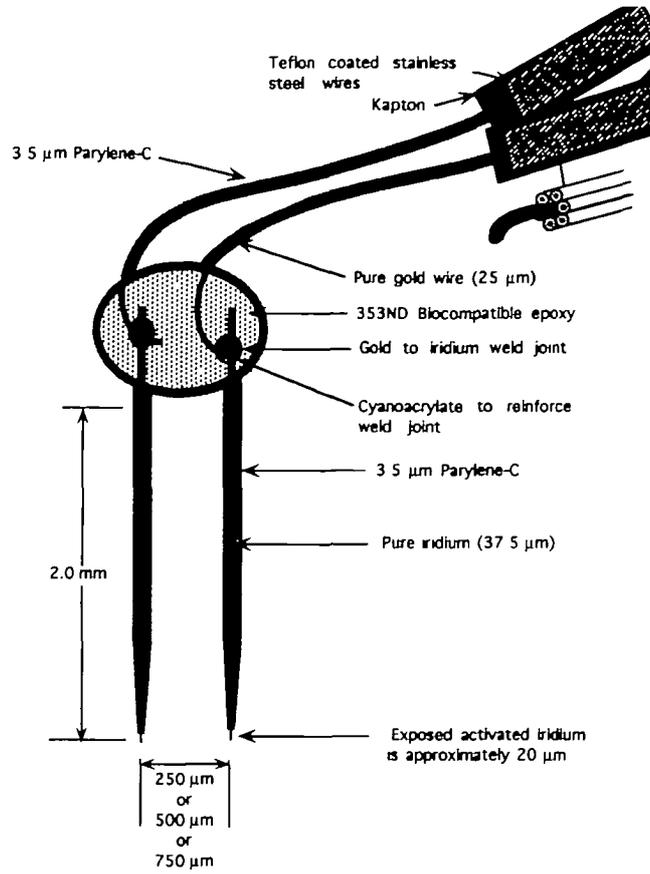


Figura 2.28: Esquema de los electrodos de Schmidt [Schmidt 1996].

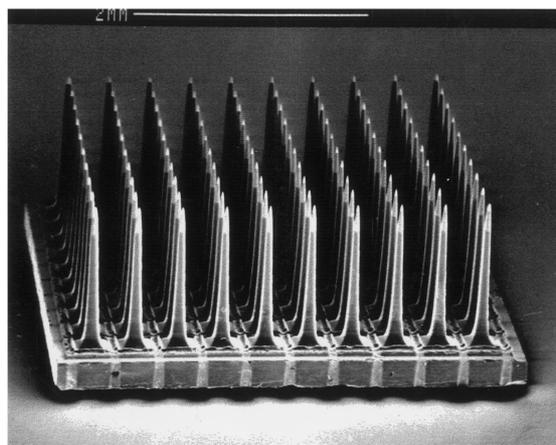


Figura 2.29: Matriz de microelectrodos de Utah.

Su principal ventaja es la mayor resolución espacial, junto con su disposición regular, que facilita la extensión a implantes formados por varias de estas matrices. Ha sido probado en animales tanto para registrar señales nerviosas como para estimular.

Sin embargo, tienen como inconveniente su estructura rígida que dificulta una penetración homogénea en profundidad de todos los electrodos. Este problema hace que aunque es tecnológicamente factible, no sea útil fabricar matrices de mayor tamaño.

2.3.2 Sistemas electrónico – informáticos

Cuando hablamos de sistemas electrónico – informáticos de ayuda a la visión, nos referimos a dispositivos diseñados con el fin de transformar imágenes tomadas del entorno de paciente, e intentar transmitirle la mejor información posible de una forma no invasiva, es decir, sin hacer interfaz directamente con el sistema nervioso.

De entre las posibilidades que se presentan, destacamos dos:

- Aprovechar el *resto visual funcional* del afectado, aplicando distintas transformaciones a la imagen de entrada
- En el caso que no se disponga de un resto funcional visual suficiente, la alternativa es realizar una *transducción sensorial*, transformando la información a transmitir de forma que pueda ser percibida a través de otro sentido que conserve su funcionalidad.

A continuación veremos algunos dispositivos englobados en alguno de los dos tipos que acabamos de describir.

El sistema LVES, desarrollado por la Universidad John Hopkins de Baltimore en colaboración con la NASA [LVES], es un dispositivo del tipo *goggles* (gafas con visores HMD, parecidas a las de realidad virtual), que pueden utilizar pacientes con una agudeza visual comprendida entre 20/80 y 20/100. Se engloba dentro del primer tipo, ya que permite realizar distintas operaciones sobre la imagen de entrada, que finalmente será expuesta al sujeto a través del dispositivo de visualización, para aprovechar su capacidad visual. Este sistema incorpora un total de 3 cámaras (ver figura 2.30), todas en blanco y negro. Las dos que se encuentran localizadas a la altura de los ojos se utilizan para generar una vista tridimensional del entorno del paciente, mientras que la tercera,

situada en la zona central de las otras dos, pero más elevada, proporciona información acerca de los detalles, por ejemplo la lectura. Permite ajustar una magnificación entre 1.5 y 10, además del contraste de la imagen.



Figura 2.30: Sistema de ayuda a la visión LVES.

La empresa **Enhanced Vision** ha desarrollado una amplia gama de dispositivos de ayuda a la baja visión, destacando entre ellos la familia conocida como JORDY [*Enhanced Vision*].



Figura 2.31: Dispositivo JORDY de Enhanced Vision.

Como podemos observar en la figura 2.31, se trata también de un dispositivo tipo goggles. Funciona con baterías e incorpora imágenes en color. Las principales características son:

- Factor de aumento hasta 30x
- Auto-enfoque
- Alto contraste
- Cuatro modos de visualización: en color, en blanco y negro, y alto contraste, con imágenes en positivo y en negativo (sin y con inversión de color).

Opcionalmente se puede adquirir un accesorio que amplía sus posibilidades, consistente en una plataforma de sobremesa que permite un factor de aumento hasta 50x, e incorpora un sistema de iluminación que incrementa el brillo y el contraste de las imágenes.

Si analizamos los dos dispositivos que hemos descrito, vemos que se limitan a realizar transformaciones sencillas de la imagen de entrada, amplificándola en tamaño, intensidad o contraste.

Otro tipo de sistema de ayuda a la visión es el desarrollado por Eli Peli [Peli 2000], que utiliza un visor 'see-through HMD' como el mostrado en la figura 2.32 (*arriba*), que permite superponer imágenes generadas artificialmente sobre las percibidas por el usuario.

En concreto, el procesamiento realizado es una extracción de bordes de la imagen captada (que coincide con la percibida a través del dispositivo HMD), y su transformación de forma que se proyecte sobre la zona del campo visual del usuario que conserve su funcionalidad. Como podemos observar en la parte inferior de la figura 2.32, cuando el paciente sufre visión túnel, sólo percibe de la zona central de su campo de visión. Si en esta zona le proyectamos una versión reducida en tamaño del resultado de la extracción de bordes, aplicado a la imagen de todo el campo visual, puede hacerse una idea de la escena global; de esta forma puede relacionar la información que le suministra su sistema visual con el resto del entorno que le es imperceptible.

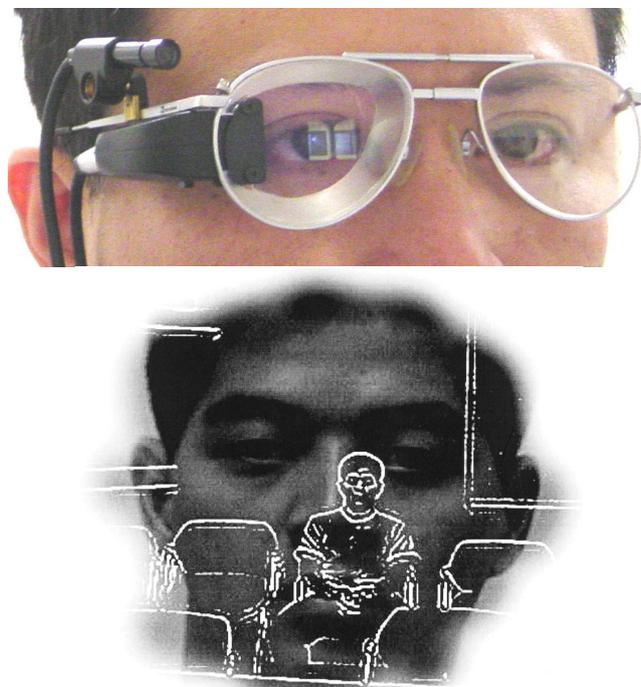


Figura 2.32: Dispositivo de visualización see-through (arriba) y un ejemplo del tipo de imagen que percibiría un paciente que presente visión túnel (abajo).

A continuación pasamos a describir algunos sistemas de transducción sensorial, en concreto transforman la información visual en información acústica.

Carver Mead y Misha Mahowald diseñaron **SeeHear** [Nielsen 1989], un circuito integrado específico (ASIC) que engloba un sensor para capturar la imagen, la etapa de procesamiento y la generación de sonidos. La característica principal de los sonidos generados es que llevan información espacial (binaurales). Modificando la amplitud y la latencia de la señal enviada a cada uno de los oídos se percibe como si el sonido procediera de distintos lugares.

El procesamiento imita al que realiza la retina, y consiste en detectar la posición de la fuente de luz que llega hasta el sensor, realzar los cambios temporales y modular el sonido generado para que parezca proceder de dicha posición.

Otro de los sistemas es **The vOICE** [*vOICE*], constituido por una cámara que captura imágenes de forma continua del entorno, y una aplicación software que traduce la intensidad de la imagen en cada píxel en un nivel de intensidad.

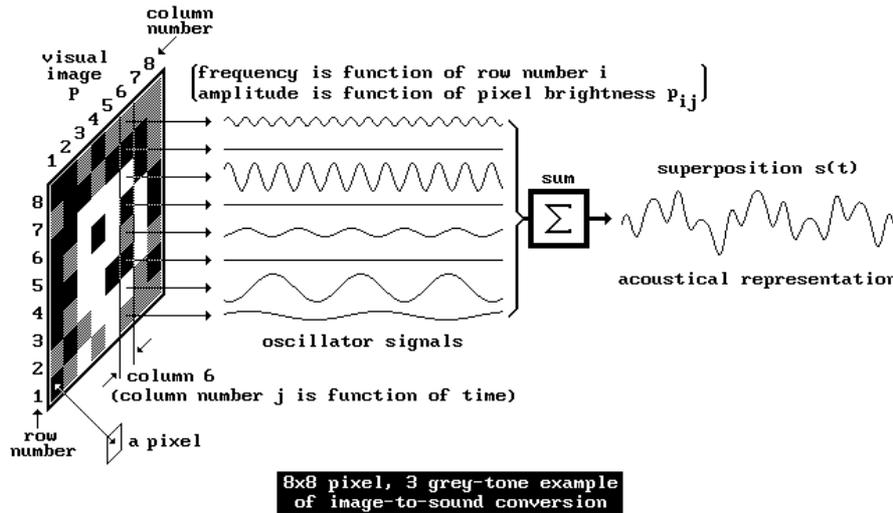


Figura 2.33: Esquema de generación de sonido de The vOICE [Meijer 1992].

El funcionamiento consiste en hacer un barrido de la imagen de izquierda a derecha, de esta forma, el sonido asociado a objetos que aparezcan a la izquierda en la imagen, serán percibidos antes en el tiempo que los asociados a los objetos situados a la derecha de la imagen.

En un instante concreto, estaremos generando los sonidos para los píxeles localizados en una de las columnas de la imagen. La forma de poder generarlo todos simultáneamente es codificando la localización dentro de la columna con el tono del sonido, haciendo corresponder los tonos más graves con la filas inferiores de la imagen y los más agudos con las filas superiores.

La figura 2.33 nos ilustra el procedimiento de generación de los sonidos. Vemos que la columna que transmitimos depende del tiempo, que la frecuencia del sonido es función de la fila en que nos encontremos, y la amplitud depende de la luminosidad del píxel en cuestión.

El principal problema que detectamos en este sistema es la elevada cantidad de información que se transmite al usuario, un sonido por cada

píxel de la imagen, lo cual es excesivo, a la vez que poco específico. Por ejemplo, no se envía información de profundidad.

El último sistema del que vamos a hablar es el denominado por sus autores 'Espacio Acústico Virtual', **EAV**, desarrollado por la Universidad de La Laguna y en Instituto de Astrofísica de Canarias [EAV].



Figura 2.34: Soporte para las cámaras de EAV.

La figura 2.34 nos muestra una estructura construida para poder colocar las dos cámaras que utiliza, de forma que se encuentren orientadas hacia el campo visual frontal del sujeto que las lleva.

Las dos cámaras se utilizan para formar un mapa de profundidades basándose en la visión estéreo. De esta forma, se construye una nueva imagen en la que la información que tenemos es la profundidad a la que se encuentra el objeto que aparecía en ese píxel en la imagen original. Trabaja con una versión de menor resolución que la imagen de entrada, y restringe las posibles profundidades a ocho.

Con esta información se generan sonidos binaurales de todas las zonas de la imagen (figura 2.35), de tal forma que a cada una de las posiciones se le asigna un sonido (que previamente hemos seleccionado, o que resulta del filtrado de un sonido base con el filtro de espacialización correspondiente), y todos los sonidos son presentados al sujeto de forma simultánea.

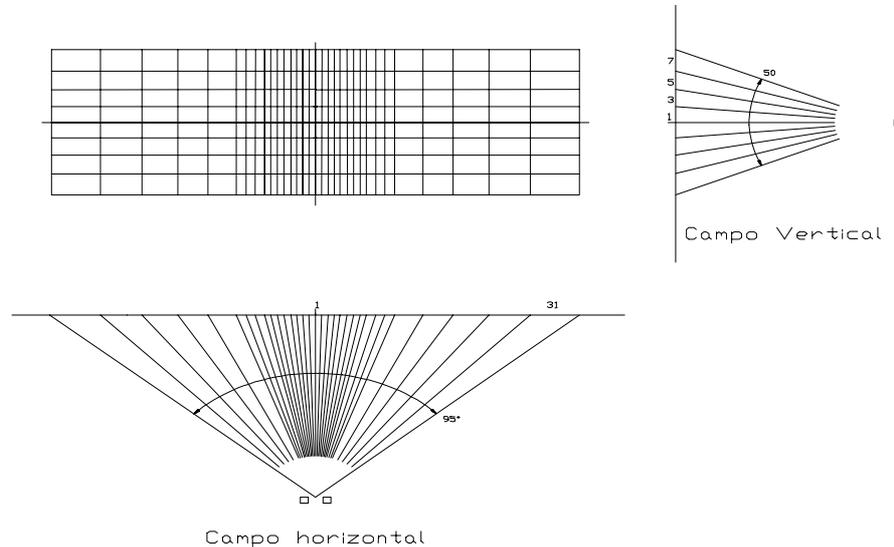


Figura 2.35: División del campo visual realizado por EAV.

En este caso también es demasiada la información que se intenta transmitir al sujeto que lo utilice, ya que una escena completa se representa mediante un patrón sonoro construido como la suma de todas las señales de los filtros, lo que implica una fuerte compresión de la información, y presenta grandes dificultades para el entrenamiento del usuario del sistema.

2.4 Conclusiones

En este capítulo se ha presentado una descripción tanto de la anatomía como de la fisiología del sistema visual humano, tratando de destacar los elementos de mayor importancia. El análisis de las funciones realizadas por los distintos tipos de células que forman los caminos visuales nos ha permitido describir distintos modelos de procesamiento discretos que las imitan.

Posteriormente se ha hecho una revisión de las patologías más comunes que afectan al sistema visual, indicando qué elementos se ven afectados, y detallando los efectos que producen sobre la capacidad de percepción visual de los pacientes que las sufren.

Los contenidos que presenta la última sección del capítulo nos han acercado a distintas alternativas de sistemas para la rehabilitación visual, haciendo una revisión de las diferentes implementaciones que se conocen. En primer lugar nos hemos acercado a los distintos tipos de neuroprótesis, recogiendo sus características y algunas de las necesidades de procesamiento que serían deseables. De la misma forma, se han descrito varios sistemas electrónico-informáticos (no invasivos) de ayuda a la visión en los que actualmente se está investigando.

Tras su estudio podemos definir algunas características que debe presentar el procesamiento visual que realicen los distintos sistemas para la rehabilitación, bien para las neuroprótesis, bien para los sistemas de transformación/transducción sensorial. En todos los dispositivos se imponen las necesidades de portabilidad y de funcionamiento en tiempo real, entendiendo por tiempo real el necesario para evocar percepciones visuales continuas. En cuanto a las neuroprótesis visuales, uno de los grandes retos es llevar a cabo una codificación neuromórfica de los estímulos a aplicar, de manera que permitan la percepción de formas en lugar de fosfenos aislados. Respecto a los patrones generados, es deseable que sean específicos, significativos y fáciles de interpretar por el paciente, sin que el proceso de entrenamiento-aprendizaje presente muchas dificultades. Para que estos sistemas puedan extraer la información más relevante, consideramos que deben tenerse en cuenta las características perceptivas del sistema visual humano. En este sentido, consideramos que deben atender a aspectos como el procesamiento basado en imágenes en **color**, la detección de **cambios** temporales en las secuencias de imágenes capturadas, la extracción de información sobre la **profundidad** en base a procedimientos que combinen medidas absolutas con medidas relativas que permitan modelar el funcionamiento del sistema visual humano. Éste extrae dicha información de la convergencia ocular y las neuronas sintonizadas por la disparidad.

Procesamiento de Información Visual: Retiner

Nuestro grupo ha trabajado en el proyecto europeo CORTIVIS (Cortical Visual Neuroprosthesis for the Blind), cuyo objetivo era diseñar una neuroprótesis cortical para invidentes. Una parte de esta tesis se ha desarrollado en el contexto de CORTIVIS, con aportaciones en el modelado software y hardware para sistemas bioinspirados de visión. El contenido de este capítulo se estructura en tres secciones. La sección 3.1 nos acerca a la organización del proyecto CORTIVIS, para continuar con la estructura global de la prótesis en el apartado 3.2. Finalmente hablaremos del entorno para modelado, simulación y síntesis en hardware de modelos de procesamiento visual que hemos desarrollado en el seno del proyecto, denominado Retiner. La sección 3.3 detalla este entorno, destacando las aportaciones propias de esta tesis doctoral.

3.1 Introducción: El proyecto CORTIVIS

El proyecto CORTIVIS (referencia QLK6-CT2001-00279) nace a finales de 2001, y consigue financiación V Programa Marco de la Unión Europea para el periodo 2002-2004.

Su nombre, formado por el acrónimo de *Cortical Visual Neuroprosthesis for the Blind*, ya nos acerca su intención, conseguir diseñar y demostrar la viabilidad de una prótesis visual a nivel de corteza cerebral para invidentes.

El proyecto, que destaca por la interdisciplinariedad, se lleva a cabo por ocho socios de cinco países europeos. Siete son grupos de investigación asociados a áreas tan dispares como la neurofisiología, medicina, tecnología electrónica, etc., y una empresa que desarrolla productos relacionados con la biomedicina. Los socios son:

- Instituto de Bioingeniería, Universidad Miguel Hernández, Alicante.
- Dpto. Arquitectura y Tecnología de Computadores, Universidad de Granada.
- Dpto. Biología, Universidad de Oldenburg, Alemania.
- Dpto. Fisiología, Universidad de Viena, Austria.
- Centro Nacional de Investigación Científica (CNRS), Francia.
- Facultad de Medicina, Universidad de Montpellier, Francia.
- Instituto de Ingeniería de Sistemas y Computadores (INESC-ID), Portugal.
- Biomedical Technologies, S. L., Alicante.

La alternativa que propone CORTIVIS a las neuroprótesis corticales ya existentes (detalladas en el capítulo anterior), pretende solventar los inconvenientes que presentaban las aproximaciones estudiadas. Más concretamente:

- El preprocesamiento de las imágenes de entrada se hace de forma bioinspirada, intentando imitar a la retina humana.
 - Aunque en principio podría utilizarse cualquier interfaz de neuroestimulación, se decide usar la tecnología de electrodos penetrantes, por ejemplo la matriz de microelectrodos de Utah [Normann 1999].
 - La comunicación de los estímulos (incluyendo la energía necesaria para su generación) se envía al interior del cuerpo vía radio, con el fin de evitar dos grandes problemas: las
-

infecciones, y el rechazo psicológico que pueden producir los sistemas basados en cables con vías transcutáneas abiertas.

- El proyecto abarca el diseño de los protocolos necesarios para seleccionar a potenciales pacientes humanos que puedan ser implantados.

La figura 3.1 nos muestra el esquema de prótesis cortical propuesto por CORTIVIS, detallando en la parte externa el sistema de preprocesamiento bioinspirado que debe codificar la información generada en trenes de impulsos, y un sistema de telemetría encargado de transmitir a la parte interna la energía de alimentación y la información necesaria para la estimulación (y posiblemente de devolver hacia el exterior información sobre el estado del implante). La parte interna del sistema completa la circuitería de telemetría y los circuitos de electro-estimulación que generan las formas de onda a aplicar a cada electrodo. Estos circuitos conectan finalmente con la interfaz neuronal, consistente en microelectrodos intracorticales (penetrantes).

En la siguiente sección se describen con mayor detalle los distintos elementos constituyentes del sistema global de estimulación.

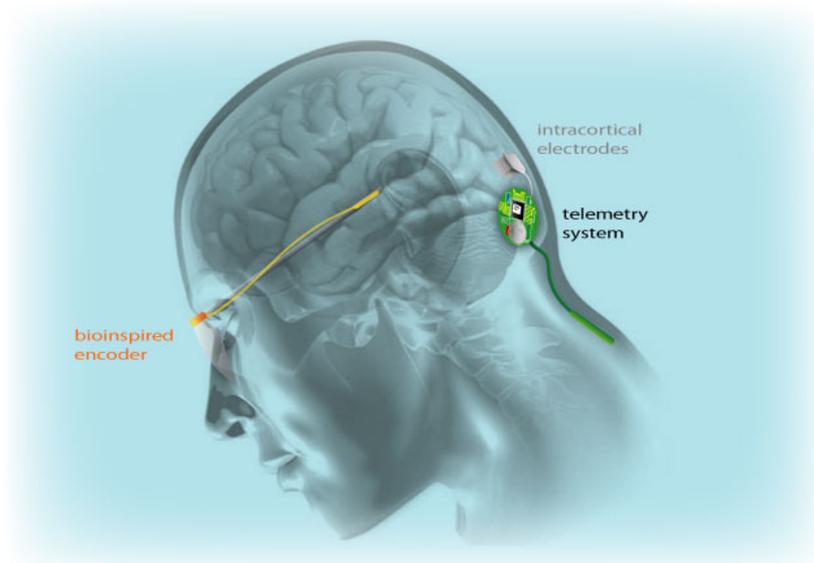


Figura 3.1: Esquema de la prótesis visual de CORTIVIS.

3.2 Estructura de la prótesis visual

La figura 3.2 nos muestra el diagrama de bloques en que podemos estructurar el sistema de neuroestimulación concebido en el proyecto CORTIVIS.

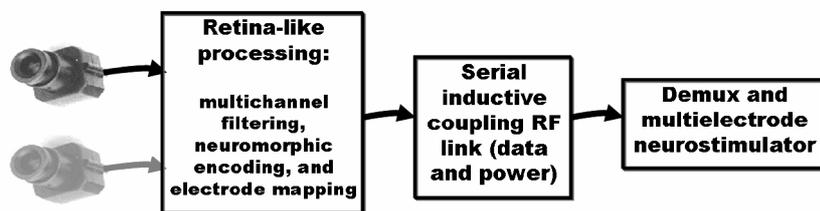


Figura 3.2: Estructura en bloques del sistema completo [Morillas 2006a].

La entrada a todo el sistema se realiza a través de una o dos cámaras que capturan imágenes digitales y en color del entorno, que pasan secuencialmente a las etapas posteriores de procesamiento. La calidad de los sensores utilizados para la captura va a influir notablemente en los resultados que se obtengan, por ejemplo, los sensores CMOS y CCD comerciales que comúnmente utilizamos tienen características lineales que no son muy apropiadas en escenas que presenten fuertes diferencias de luminosidad, llegando a saturar los elementos de captura; se han evaluado sensores cuya respuesta sigue una curva logarítmica [*Ims Chips*], que permite abarcar un rango más amplio de intensidades luminosas en una misma imagen. Podemos observar la diferencia en la figura 3.3.

Las imágenes que se han tomado se pasan a una etapa que realiza un procesamiento bioinspirado, imitando a la retina biológica. Las operaciones que utilizamos para modelar dicho procesamiento son un conjunto de filtros espacio-temporales, que realzan los contrastes de color, de intensidad luminosa y los cambios producidos en el tiempo. Todas las características extraídas las combinamos entre sí para constituir un único mapa de características en el que destaquen las zonas de la imagen con mayor actividad detectada, en base a la combinación de filtros seleccionados (véase figura 3.4(2)).



Figura 3.3: Diferencia en la captura de imágenes por una cámara CCD lineal (izquierda) y una cámara CMOS logarítmica (derecha). Gracias a la compresión logarítmica es posible captar en la misma imagen (sin saturación) zonas que difieren en varios órdenes de magnitud en intensidad luminosa.

Después de filtrar las imágenes de entrada, tendremos que hacer una proyección de las mismas sobre la matriz de microelectrodos (figura 3.4(3)) que se va a utilizar para la estimulación, seleccionando los píxeles que van a contribuir al nivel de actividad en cada canal. De esta forma definimos el *campo receptivo* asociado a cada electrodo.

Una vez que hemos determinado el nivel de actividad que cada fotograma va a aportar a cada electrodo, incluimos un sistema de traducción de estos niveles de actividad en impulsos, mediante un modelo neuronal de actividad del tipo ‘íntegra y dispara’ [Pelayo 2004]. A partir de esta etapa, la información es transmitida como eventos asociados a un electrodo determinado, que se representa mediante su dirección, siguiendo un esquema de codificación similar a AER (Address Event Representation). Este método de comunicación ha sido ampliamente estudiado e implementado tanto en cócleas artificiales [Lazzaro 1993], como en otros sistemas VLSI bioinspirados de visión [Mahowald 1994], [Boahen 2000], [Mortara 1995].

Justo a la salida de esta etapa se ha pensado que podría incluirse un método de redireccionamiento, implementado por ejemplo con una memoria o tabla de consulta, que nos permita redirigir la salida de una zona del mapa de actividad generado hacia electrodos que se encuentren físicamente en un lugar distinto de la matriz. Este proceso se hace necesario para acondicionar el sistema a cada paciente implantado, variando la proyección de su campo visual en la corteza cerebral. En general esta proyección se verá afectada por deformaciones y no linealidades [Normann 1996], no obstante, los estudios realizados [Warren 2001] demuestran que en gran medida se conserva la propiedad de

proyección retinotópica (zonas de la retina que recogen áreas cercanas del campo visual, se corresponden con zonas cercanas de la corteza visual).

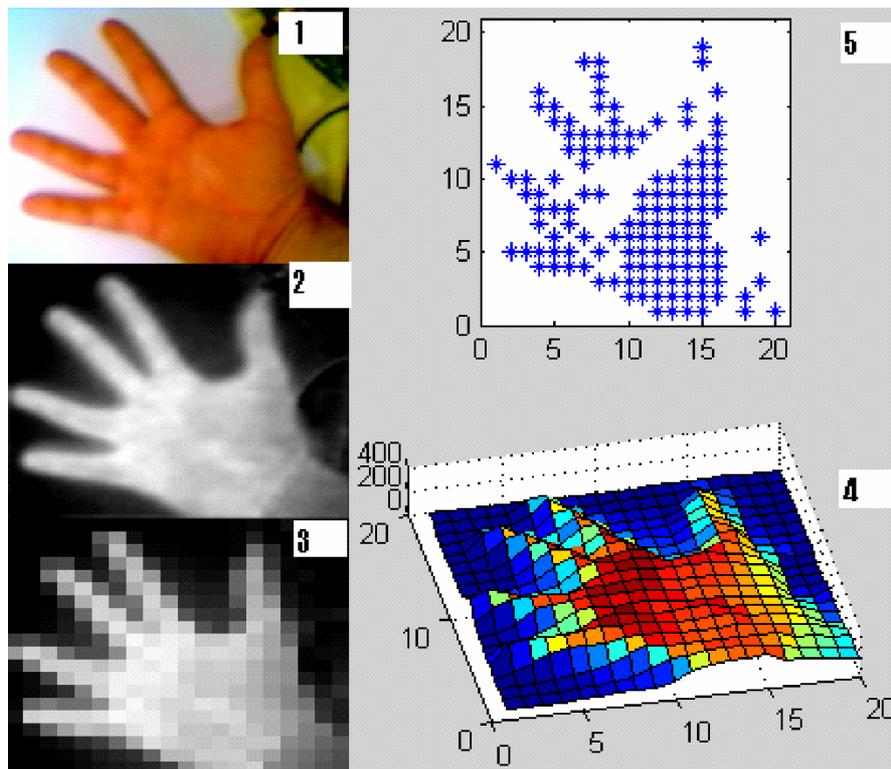


Figura 3.4: Distintas etapas en el procesamiento con Retiner, entorno software que será descrito en la sección 3.3. (1) Imagen original de entrada. (2) Combinación de varios filtros espaciales aplicados a la imagen original. (3) Matriz de actividad que representa la contribución de la salida de los filtros a cada electrodo. (4) Representación tridimensional de la matriz de actividad. (5) Conjunto de eventos producidos en uno de los instantes de tiempo ante la entrada recibida [Morillas 2006b].

El siguiente paso a realizar dentro del sistema global consiste en enviar la identificación del electrodo que queremos activar al módulo de estimulación. Dicho módulo estaría ubicado en el interior del cuerpo del paciente implantado, y la alternativa que propone CORTIVIS es enviar dicha información a través de un enlace serie de radiofrecuencia, que no sólo reciba la información para la estimulación del exterior, sino que

además sea capaz de proveer al implante interno con la energía necesaria para su funcionamiento, y para la generación de las corrientes de estimulación. De esta forma se solucionaría el problema de tener que intervenir quirúrgicamente al paciente cada cierto tiempo para cambiar su fuente de alimentación, como ocurre en el caso de los marcapasos cardiacos.

Finalmente, una vez que el implante ha recibido la información necesaria para la estimulación, debe ser capaz de generar una onda eléctrica con las características que tenga configuradas cada electrodo individualmente, por ejemplo, la corriente de estimulación, la duración en el tiempo, la polaridad inicial, etc. Dependiendo de las características concretas del implante, existe la posibilidad de tener un estimulador por cada electrodo. Si es el caso, podremos estimular cuantos canales deseemos simultáneamente (siempre sin superar unos límites máximos de inyección de carga, para no poner en peligro al paciente). De todas formas, el número de canales simultáneos que podemos estimular estará limitado por la cantidad de módulos estimuladores que incorpore el sistema final.

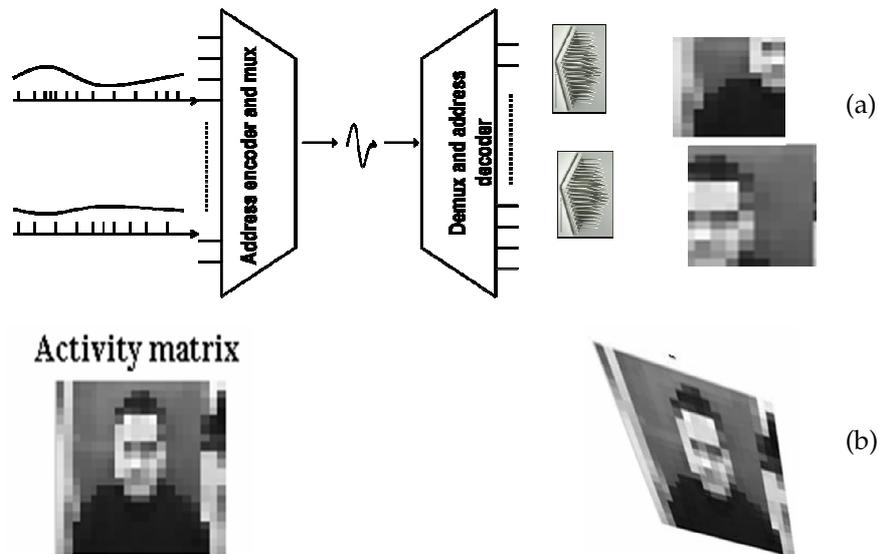


Figura 3.5: Esquema del proceso de envío de la información de estimulación. Podemos distinguir la generación de direcciones, decodificación de la misma y demultiplexación para determinar el array de microelectrodos destino (a) [Romero 2004], así como corregir deformaciones mediante una reasignación de las direcciones (b).

De una forma o de otra, el camino a seguir para determinar qué electrodo concreto debe ser activado cuando llega un evento, se determina mediante algún tipo de decodificación/demultiplexación. Este método es también el utilizado cuando el implante consta de más de una matriz de microelectrodos; es la dirección del electrodo la que determina tanto la matriz en la que se encuentra como el estimulador que debe iniciar el proceso.

La figura 3.5 nos ilustra el encadenamiento de las últimas fases, así como un ejemplo gráfico que nos permite comprender mejor su funcionamiento.

3.3 Entorno de modelado y simulación software: Retiner

La primera de las tareas a realizar por nuestro grupo dentro del proyecto CORTIVIS consistía en implementar un modelo de retina que efectuase el procesamiento de las imágenes capturadas por una cámara. Este modelo llevaría a cabo una codificación bioinspirada de la secuencia de imágenes de acuerdo con las directrices que se han descrito en la sección 3.2. En nuestro caso se decidió realizar una primera simulación del mismo por software para validar el correcto funcionamiento del mismo. De esta forma surgió *Retiner*, una aplicación software diseñada en MatlabTM, que simulaba por software un modelo concreto de retina.

Posteriormente y con el avance del proyecto, se determinó la necesidad de convertir esta aplicación en un entorno mucho más general y paramétrico, que permitiese diseñar y simular modelos de procesamiento visual bioinspirados mucho más generales. A lo largo de esta sección vamos a ir conociendo en más detalle esta aplicación en el estado más avanzado que se encuentra, resaltando las posibilidades de modelado que incorpora, y su justificación tanto en el contexto de CORTIVIS, como en otros trabajos de investigación relacionados con la rehabilitación visual.

En la figura 3.6 podemos ver la ventana principal de nuestra aplicación. Se puede observar la organización de una forma intuitiva de los

menús, encontrando uno para cada fase de procesamiento (también ilustradas en la parte superior de la ventana).

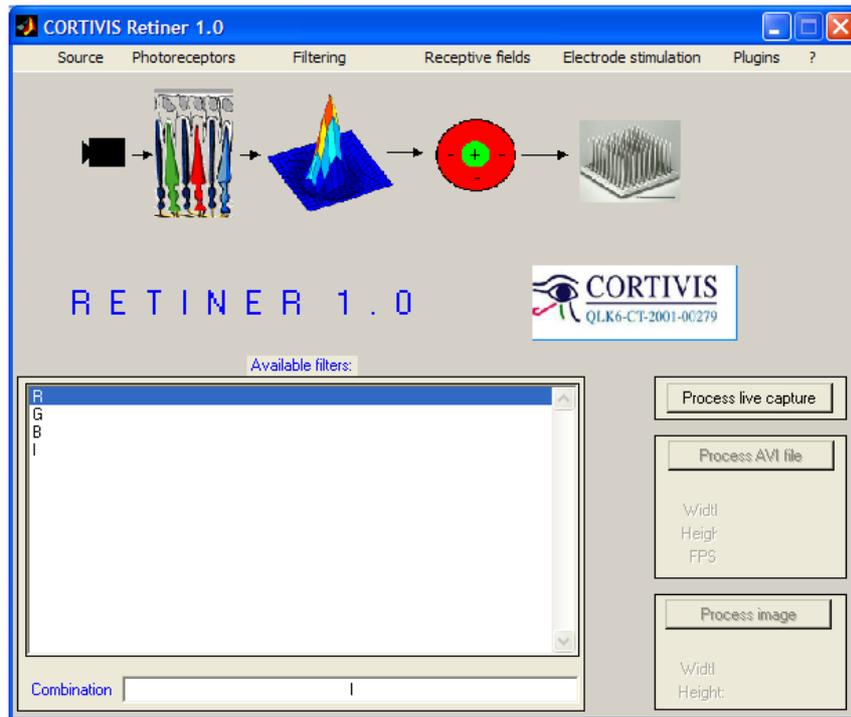


Figura 3.6: Interfaz de usuario del entorno de modelado y simulación de sistemas de visión, Retiner.

Nuestro entorno está diseñado para aceptar como entradas, tanto imágenes estáticas (en cualquiera de los formatos más extendidos), como secuencias, bien almacenadas en archivos con el formato AVI, bien capturadas *en vivo* por cualquier cámara soportada por el sistema operativo. Más adelante veremos algunas de las particularidades del procesamiento de secuencias almacenadas en archivos de vídeo, así como de imágenes estáticas.

Una vez hemos determinado la entrada a nuestro sistema, podemos descomponer las imágenes obtenidas en tres canales de color (rojo, verde y azul), haciéndolos corresponder con los distintos tipos de fotorreceptores existentes en la retina. La ganancia de cada uno de los canales puede ser modificada individualmente, permitiendo desde realzar algunas características cromáticas de la escena, hasta modelar algunas

alteraciones de la visión cromática (*acromatopsia*). A partir de estos canales de color obtenemos uno adicional (acromático) para la intensidad luminosa.

En el siguiente paso podemos definir el conjunto de filtros espacio-temporales que van a formar parte del modelo de procesamiento, así como la combinación de sus salidas, que definirá el comportamiento del mismo. Retiner permite la definición de los siguientes tipos de filtros espaciales:

- Gaussiana: realiza un suavizado espacial.
- Diferencia de Gaussiana: modela el contraste centro-periferia actuando sobre dos entradas distintas, permitiendo simular la oponencia de color.
- Laplaciana de Gaussiana: similar al anterior, pero actuando sobre una única entrada.

Adicionalmente, mediante expresiones Matlab se puede añadir cualquier otro filtro soportado por esta herramienta.

Las entradas a los distintos filtros pueden ser tanto los canales cromáticos y de intensidad que hemos comentado con anterioridad, como la salida producida por algún otro filtro espacial que hayamos definido con anterioridad, posibilitando la estructura en cascada.

3.3.1 Respuesta temporal central y periférica

Podemos incluir también un único filtro temporal aplicado al canal de intensidad luminosa, y que realza los cambios temporales en las secuencias de imágenes (no es aplicable a las imágenes estáticas). Para intentar emular el comportamiento de la retina, en la que los cambios temporales afectan con mayor importancia a las zonas periféricas, se ha incluido la posibilidad de definir un factor de ponderación variable para un área circular de la zona central con respecto a la periferia (ver figura 3.7).

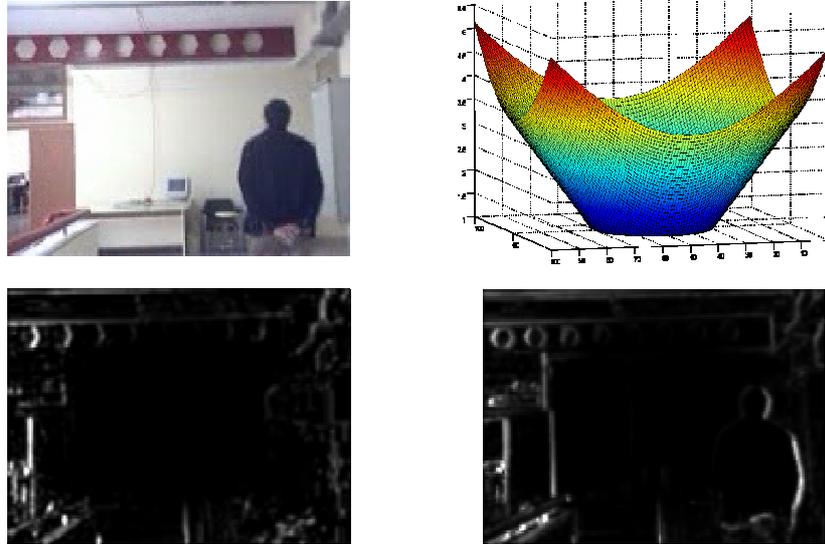


Figura 3.7: Filtrado temporal en el que distinguimos la zona central de la periférica [Morillas 2006a]. Arriba a la izquierda presentamos una de las imágenes de la secuencia original, en la que la cámara avanza por un pasillo; una persona adelanta a la cámara por la derecha. Arriba a la derecha tenemos la máscara aplicada para este procesamiento; el área central es un círculo que ocupa el 30% de la diagonal de la cámara, donde no se afecta a la imagen, mientras que los valores máximos en la periferia llegan a un factor 5. Abajo a la izquierda se observa la salida producida por células OFF transitorias, mientras que a la derecha vemos la salida de células ON-OFF.

3.3.2 Determinación de los campos receptivos

Las salidas obtenidas por los distintos filtros que incluyamos dentro del modelo son combinadas mediante una expresión algebraica definida por el usuario, obteniendo una única figura de información que denominamos 'imagen perceptual' (véase por ejemplo la figura 3.4(2)).

Una vez determinada la combinación de los filtros, debemos decidir qué zonas de la imagen van a contribuir a cada electrodo para

neuroestimulación¹, es decir, tenemos que definir el *campo receptivo* de cada electrodo. Por defecto, se hace una división uniforme y sin solapamientos de la imagen de entrada entre el tamaño de la matriz de microelectrodos, formando un grid (campos receptivos rectangulares). Con el fin de poder estudiar estructuras más complejas, y poder simular el comportamiento del sistema visual de otras especies que presentan campos receptivos más complejos, se ha diseñado un módulo software que puede invocar Retiner, y que amplía las posibilidades para la definición de dichas zonas.

La parte superior de la figura 3.8 nos muestra la interfaz de este módulo, junto con cuatro ejemplos de definición de campos receptivos. Para todos los casos se ha considerado una matriz de 25 x 25 microelectrodos. En la fila central, a la izquierda, tenemos el resultado de la configuración que vemos en la ventana superior: campos receptivos de forma circular cuyo radio varía desde la zona central (5 píxeles) hacia la periferia (hasta 15 píxeles el más lejano). La interpolación para determinar el radio se hace de forma lineal, y vemos que las características de la zona central se han extendido a dos filas y dos columnas más en torno a esta zona.

Si nos fijamos en la fila central, a la derecha, podemos ver el resultado de definir campos receptivos con forma de cuadrado de lado variable desde el centro a la periferia. En este caso, son 10 píxeles el lado del cuadrado de la zona central, y 30 el situado en la zona más exterior. Sin embargo, en esta situación se ha seleccionado un tipo de interpolación mayor que lineal, por lo que el lado de los cuadrados crece muy rápidamente conforme nos vamos alejando de la zona central de la matriz de microelectrodos.

En la parte inferior distinguimos otras dos formas de campos receptivos que se pueden definir, a la derecha hemos definido áreas cuyo contorno está delimitado por elipses con 15 píxeles de altura y 28 de anchura. A la izquierda tenemos formas rectangulares, de 14 píxeles de altura y 18 de anchura, pero sobre estas formas se ha definido una transformación geométrica de rotación. En concreto se han rotado, con respecto a su centro, 30 grados en sentido contrario a las agujas del reloj.

¹ Aunque en esta descripción nos referimos a los electrodos como elementos de destino de salida en el contexto de un sistema para neuroestimulación, sin embargo la definición de campos receptivos es también aplicable para salidas a visualizadores u otros elementos transductores.

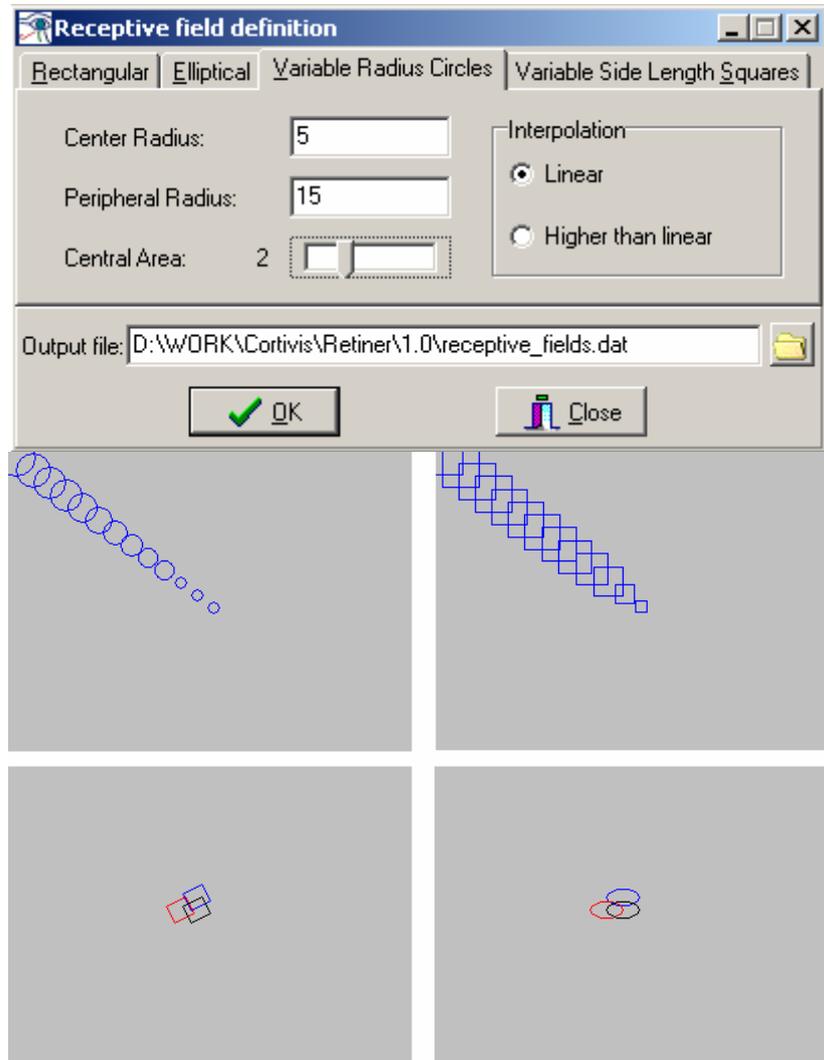


Figura 3.8: Herramienta para la definición de campos receptivos.

Esta herramienta genera un archivo de configuración válido para Retiner que a partir de ese momento calculará la actividad para cada electrodo como la media de todos los valores que presentan los píxeles de su campo receptivo, en la ‘imagen perceptual’. De esta forma llegamos a otro tipo de representación, en el que disponemos de un valor de actividad para cada microelectrodo, asociado al resultado de filtrar la imagen actual con el modelo de visión diseñado. A esta matriz se le conoce como ‘matriz de actividad’ (véase figura 3.4(3)).

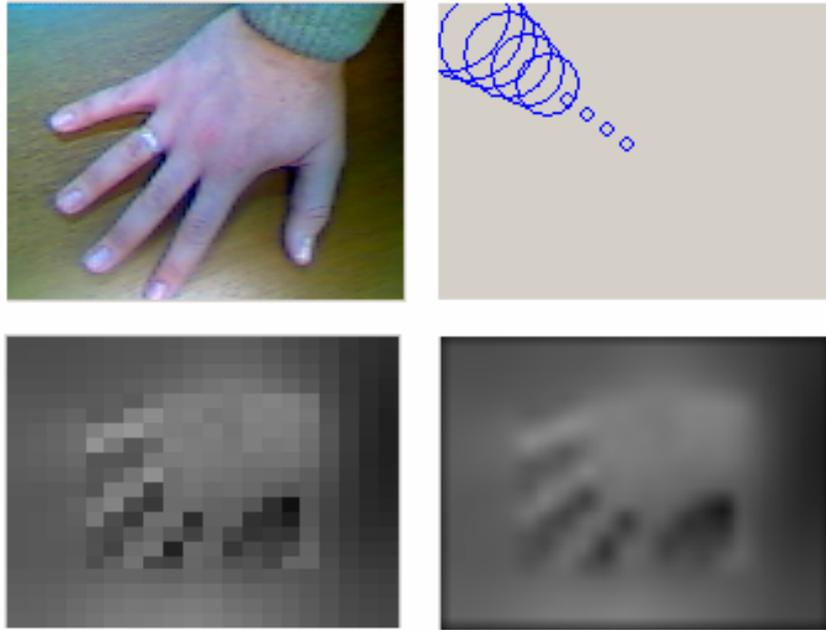


Figura 3.9: Efecto de la aplicación de campos receptivos de tamaño variable centro-periferia (arriba derecha) a la imagen original (arriba izquierda). Abajo a la izquierda tenemos la matriz de actividad resultante y a la derecha, una aproximación a la percepción evocada. Podemos observar la mayor resolución espacial en el centro [Morillas 2006b].

El archivo de configuración se puede generar en su totalidad o modificar uno existente de forma manual, dando la posibilidad de generar cualquier forma de campos receptivos, homogéneos o irregulares. En el archivo se especifica en primer lugar el tamaño de la matriz de electrodos sobre la que se proyecta la imagen, seguido de una lista de píxeles de dicha imagen para cada uno de los componentes de esta matriz.

3.3.3 Codificación neuromórfica

La matriz de actividad es la información de entrada al módulo de conversión a impulsos. Para este módulo se ha realizado una implementación discreta de un modelo básico del tipo integra y dispara, que incluye un término de pérdidas. La elección de este modelo se debe a la simplicidad de su implementación en hardware (FPGA) con los mínimos recursos posibles [Pelayo 2004]. La figura 3.10 ilustra el funcionamiento de

este modelo básico. La contribución de la actividad es lineal, sumándose cada instante de tiempo al valor acumulado; además, el término de pérdidas es constante, lo que conlleva una velocidad lineal en la descarga.

No obstante, bastaría con cambiar el método de cómputo en el módulo Matlab para implementar otros modelos de mayor complejidad, más próximos a la dinámica de las neuronas biológicas. Así, dos posibles modelos que incorporan una dependencia del término de pérdidas con el valor actual del potencial de membrana de la neurona son los que describen matemáticamente las ecuaciones (3.1) y (3.2) [Pelayo 1997]:

$$\frac{\partial V_x}{\partial t} = -\frac{1}{\tau} V_x + I \quad (3.1)$$

$$\frac{\partial V_x}{\partial t} = -\frac{1}{\tau} V_x + I(V_{\max} - V_x) \quad (3.2)$$

donde V_x representa el potencial de membrana (potencial interno (C) en la figura 3.10), τ es la constante de tiempo de descarga, V_{\max} se corresponde con el máximo valor para el potencial de membrana, e I es el nivel de actividad de entrada instantáneo.

La ecuación (3.1) modela una neurona aditiva o lineal con un factor de pérdidas dependiente del potencial de membrana, por lo que, en ausencia de actividad a la entrada, la velocidad a la que decrece el potencial de membrana es exponencial. La característica principal del modelo definido en la ecuación (3.2) es que la actividad de entrada contribuye de forma multiplicativa; así la velocidad de crecimiento del potencial de membrana es menor cuanto más nos acercamos a su nivel máximo, con una evolución exponencial en respuesta a un nivel de entrada excitatoria constante. Ambos modelos se encuentran descritos con más detalle en [Pelayo 1997] y su implementación en circuito digital implica añadir uno o dos multiplicadores al módulo básico incorporado en Retiner.

En la bibliografía se ha descrito un gran número de otros modelos más complejos, por ejemplo incluyendo probabilidades de emisión de impulsos [Gerstner 2002].

A la salida de esta etapa ya disponemos de la identificación de los electrodos en los que se ha producido un evento, que tendremos que comunicar al dispositivo de estimulación.

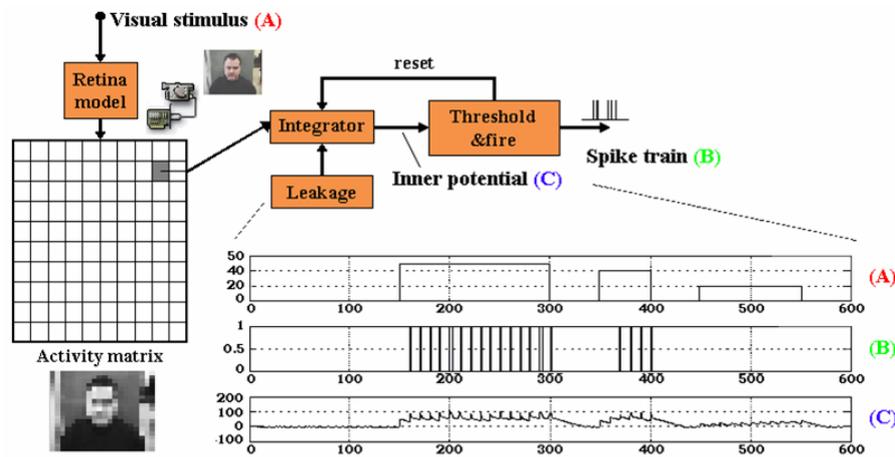


Figura 3.10: Módulo de producción de impulsos implementado en Retiner [Morillas 2005]. La señal (B) nos muestra el tren de pulsos producido por el estímulo de entrada (A). La señal (C) nos muestra el potencial interno. Debido a que se trata de una implementación digital podemos observar que la evolución temporal se produce de forma discreta. Cada vez que el potencial interno supera el umbral de disparo, se produce un pulso. La frecuencia de generación de pulsos es proporcional al nivel de intensidad de la matriz de actividad, y el tiempo desde que se presenta el estímulo hasta que se produce el primer evento, es inversamente proporcional a dicha intensidad.

3.3.4 Salida visual

La herramienta que hemos diseñado, Retiner, ofrece la posibilidad tanto de visualizar como de almacenar (como archivo de secuencia de imágenes, AVI), los distintos puntos intermedios del procesamiento.

En primer lugar podemos visualizar el resultado de la combinación de la aplicación de todos los filtros a la imagen de entrada. De esta forma podemos utilizar esta información en otro tipo de sistemas de rehabilitación visual sin limitarnos a la estimulación cortical. En concreto, podemos realizar este filtrado con el fin de extraer la información más relevante y transmitirla a través de las zonas visuales que conserve un

paciente; o transformarla en otro tipo de información, por ejemplo, acústica.

En resumen, el entorno de simulación que proporciona Retiner ha sido diseñado de manera que sea una herramienta útil para el desarrollo de cualquier sistema de rehabilitación visual.

Otra salida que puede ser muy interesante es visualizar la propia matriz de actividad, o una transformación de ésta, que hemos denominado 'graded output', que no es más que un intento de reconstrucción de la sensación que se evocaría en un paciente estimulado con la información contenida en la matriz de actividad. Tanto una como otra pueden ser utilizadas en el diseño y realización de experimentos psicofísicos.

Retiner permite visualizar a pantalla completa estos resultados, de forma que si conectamos unas gafas de realidad virtual (ver figura 3.11) al PC, podríamos hacer estudios sobre la agudeza visual que se obtiene para un determinado tamaño de array de microelectrodos, o determinar el tamaño mínimo de la matriz para proveer de ciertas capacidades visuales al paciente (como la lectura de caracteres de gran tamaño, navegación por entornos conocidos/desconocidos, etc.).

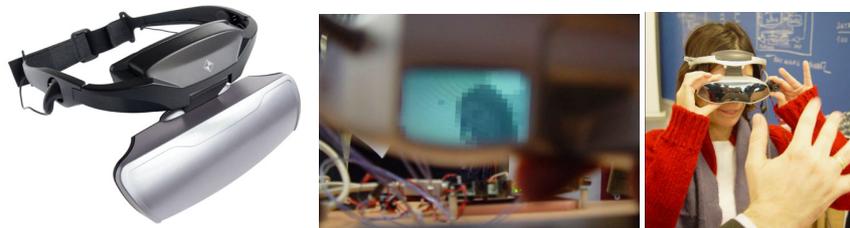


Figura 3.11: Izquierda: Gafas de realidad virtual que se utilizan para realizar los experimentos psicofísicos. Centro: Detalle de la imagen visualizada mediante las gafas. En este caso se visualiza la matriz de actividad con 20x20 componentes. Derecha: Prueba de una implementación portable en un circuito reconfigurable.

La información que hemos denominado 'graded output' se construye a partir de la matriz de actividad mediante diversas técnicas de suavizado que puede seleccionar el usuario, como son bilineal, bicúbico y gaussiano. La figura 3.12 nos muestra las diferencias entre estos tres tipos de procesamiento para la emulación de la percepción evocada en el caso de estimulación eléctrica.

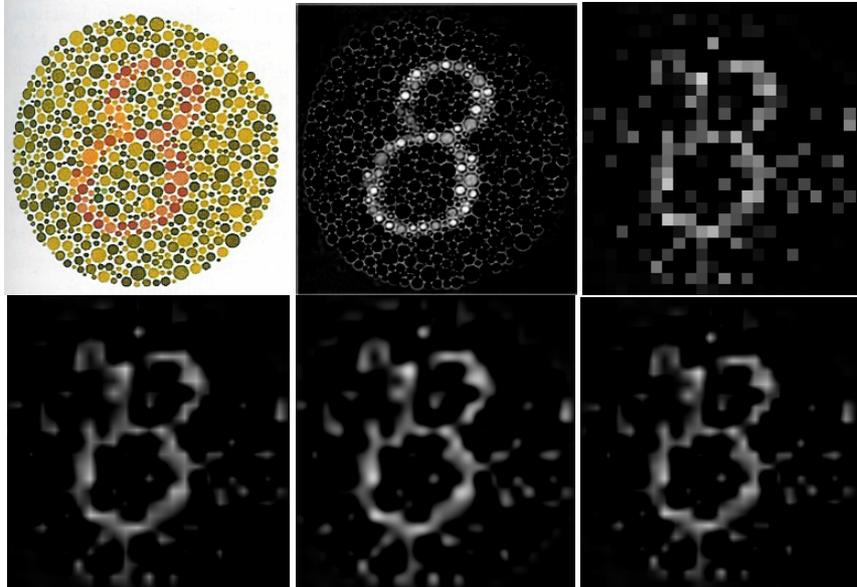


Figura 3.12: En la fila superior, de izquierda a derecha, se muestra la imagen de entrada a Retiner, un tipo de imágenes utilizadas para la detección de anomalías visuales (daltonismo); la salida obtenida con una determinada combinación de filtros espaciales que realzan el contraste de color y una reducción de resolución para una matriz de 25×25 microelectrodos. En la fila inferior tenemos las distintas aproximaciones a la percepción evocada en el siguiente orden: bilineal, bicúbica y gaussiana. Se puede observar que la que produce unos resultados más naturales, con menos rasgos sintéticos, es el suavizado bicúbico.

3.3.5 Simulación temporal

El modelo de procesamiento implementado con Matlab cuando trabajamos con una secuencia de imágenes 'en vivo', directamente capturada por la cámara, presenta un problema: el tiempo necesario (latencia) para procesar completamente una de las imágenes es elevado. La resolución temporal (o paso de simulación) que podemos conseguir viene determinada por dicha latencia, por lo que será reducida, limitando el rendimiento de la respuesta temporal del sistema de generación de eventos. Dada la importancia de las características temporales en los sistemas neuronales, se ha desarrollado un método alternativo de

procesamiento basado en el almacenamiento previo de la secuencia de imágenes en un archivo AVI, y una simulación off-line posterior, que nos permita seleccionar la resolución temporal.

En los archivos de secuencia de imágenes, la resolución temporal viene determinada por la característica denominada como *fps* (imágenes o frames por segundo). Sin embargo, nosotros hemos permitido definir un paso de simulación menor (figura 3.13), de forma que la simulación no está dirigida por las imágenes que forman la secuencia, sino por el tiempo del que llevamos un control estricto. Dado un instante de simulación, t , lo primero que hacemos es determinar el frame de la secuencia que le corresponde temporalmente.

En el caso de que el modelo de visión que definamos con Retiner incorpore el filtrado temporal, la detección de cambios no está fijada a la comparación del frame actual con el que correspondiera en el paso anterior de simulación, sino que se permite seleccionar tanto una resolución temporal distinta, descrita como el número de pasos de simulación hacia atrás para seleccionar el primer frame con el que realizamos la comparación, como un número de frames previos variable, de los que se computará la media, y se compararán con el frame actual.

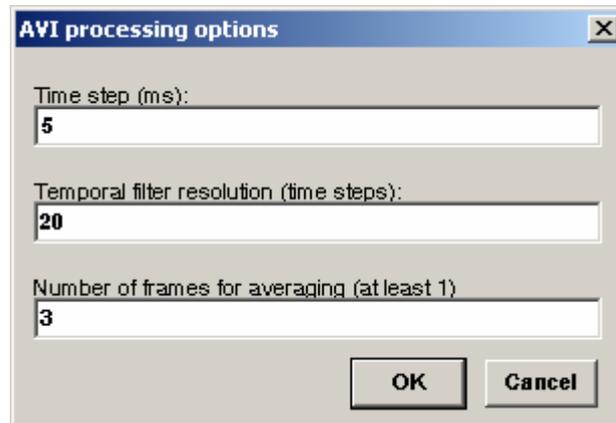


Figura 3.13: Ventana de Retiner para configurar los parámetros de la simulación temporal de secuencias de imágenes almacenadas en archivo.

La mejora en los resultados obtenidos al introducir la posibilidad de realizar la simulación con una resolución temporal superior a la determinada por las características del archivo que almacena la secuencia de imágenes, puede observarse en la figura 3.14.

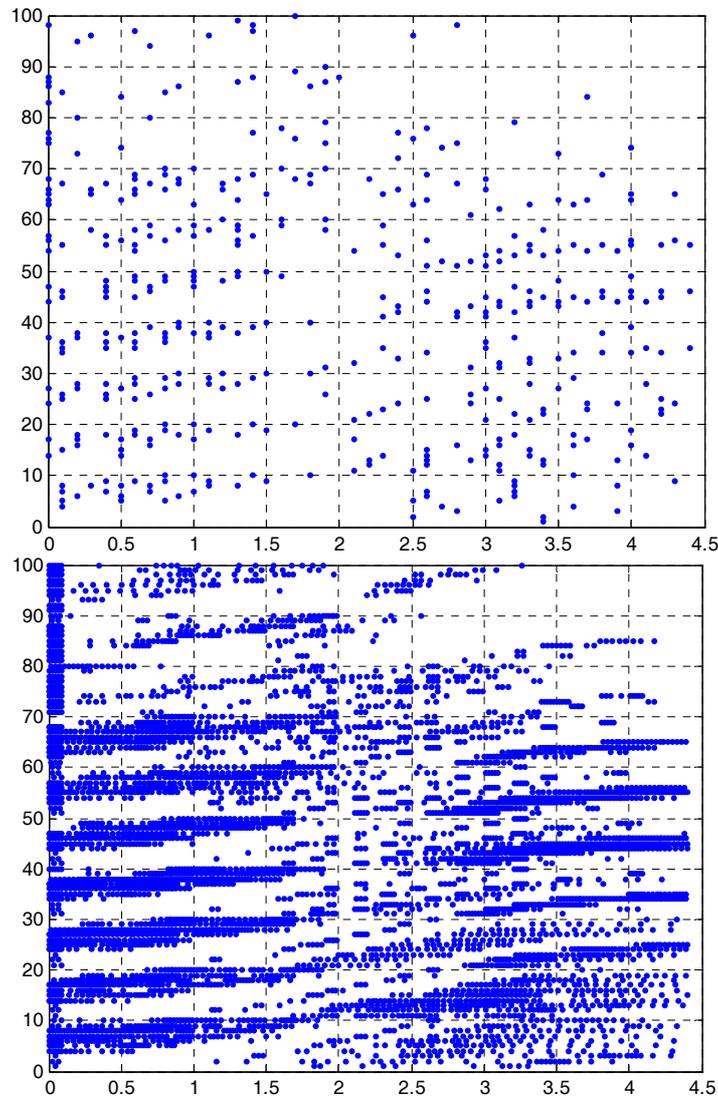


Figura 3.14: En la parte superior tenemos el conjunto de eventos generado para una simulación con una resolución temporal ajustada a las características del archivo que almacena una secuencia. La respuesta temporal, del orden de las décimas de segundo, no se corresponde con los resultados que se obtendrían en un experimento biológico. En particular, se están generando patrones de sincronización ficticios entre neuronas, que coinciden con los periodos de actualización de las salidas. Este problema está resuelto en la figura de abajo, en la que la cadencia de respuesta está en torno a la centésima de segundo.

3.3.6 Visualización de spikes. Registros biológicos

Con la finalidad de poder realizar comparaciones con resultados obtenidos en experimentos biológicos, Retiner dispone de un módulo de representación de eventos (*spikes*) que permite visualizar todos los eventos que se han producido en una sesión de simulación (los gráficos de la figura 3.14 han sido generados con este módulo). Además, permite representar los archivos de registro de spikes obtenidos en experimentos con retinas de animales in-vitro, almacenados en el formato .NEV de los instrumentos fabricados por Cyberkinetics [*Cyberkinetics*]. En la figura 3.15 tenemos un ejemplo. En este caso se aproxima la respuesta de un subconjunto de las células de la retina biológica mediante el modelo Retiner definido por la ecuación (3.3):

$$retina = \frac{1}{4} \cdot I + F_{temp} + 45 \cdot RAND \quad (3.3)$$

En ésta se incluye una componente de intensidades, un filtrado temporal modelando la respuesta transitoria de células ON-OFF, y una componente aleatoria para modelar las frecuencias de disparo en ausencia de actividad.

3.3.7 Extensiones de Retiner

Algunas ampliaciones que está experimentando retiner van dirigidas por tres caminos que se corresponden con aportaciones de los trabajos de tesis doctoral realizados por los miembros del proyecto CORTIVIS en Granada (véase figura 3.17).

Este trabajo de tesis doctoral aporta una extensión del modelo para incorporar información de visión estéreo. Para ello, nuestra aplicación necesita ser capaz de trabajar con dos cámaras simultáneamente, algo que han incorporado las últimas versiones de Matlab™ (7.0, 7.1,...), a través de la herramienta *Image Acquisition Toolbox*. El nuevo módulo obtiene mapas de disparidad con distintas resoluciones, a partir de los pares de imágenes capturadas por las cámaras. Esta información está disponible para utilizarla, por ejemplo, como uno de los filtros más, a la hora de realizar la combinación de los mismos que represente el modelo del sistema visual. La figura 3.16 muestra un ejemplo de los resultados obtenidos.

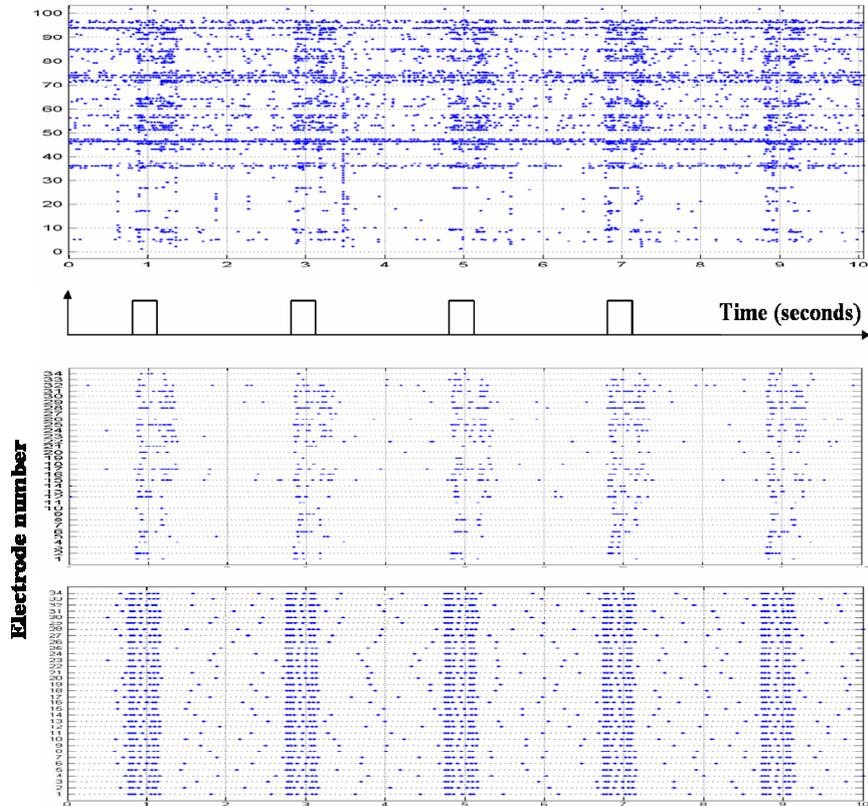


Figura 3.15: Ejemplo de representación de estímulos con Retiner y comparación con experimentos biológicos. En la parte superior tenemos la representación de todos los spikes registrados en un experimento in-vitro con una retina de conejo. En la parte inferior tenemos tres representaciones, arriba la línea de tiempos que indica la activación del estímulo (full-field flash), debajo hemos seleccionado algunas células con un comportamiento ON-OFF, y en la parte inferior tenemos la respuesta simulada con Retiner para el mismo estímulo, con un modelo de retina concreto, y la traducción a impulsos generada por el modelo descrito en la sección 3.3.3 [Morillas 2006a].



Figura 3.16: Procesamiento estéreo con Retiner. Arriba tenemos el par de imágenes de entrada; abajo a la izquierda se visualiza el mapa de disparidades obtenido, y a la derecha, su combinación con otros filtros espaciales, y uno temporal.

En el capítulo 6, en el que se presenta un sistema completo de ayuda para la rehabilitación visual, se hará uso de este nuevo módulo de cálculo de disparidades. Allí podremos ver una implementación hardware del mismo, y resultados de su utilización.

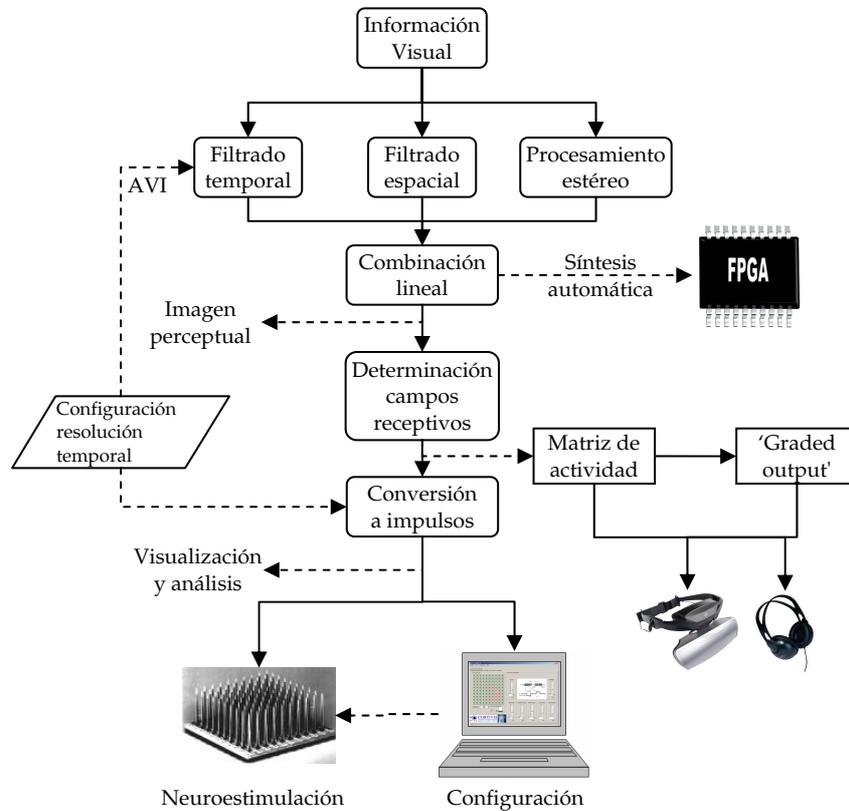


Figura 3.17: Arquitectura de referencia para el modelado con Retiner.

La segunda de las ampliaciones, motivo de la tesis doctoral [Martínez 2006], consiste en un módulo capaz de generar una descripción en un lenguaje de descripción del hardware para el modelo de procesamiento visual que hayamos definido con Retiner, y sintetizarla para su programación en un circuito reconfigurable (véase figura 3.17). No consiste en una mera traducción de la solución software a un circuito con la misma funcionalidad, sino que hay un proceso previo de simulación para optimizar determinados parámetros que van a conseguir una implementación en circuito con el mínimo de requisitos, sin que afecte de forma sustancial a la calidad de los resultados.

Finalmente, se está completando la aplicación con una herramienta adicional concebida para el ajuste de todos los aspectos necesarios para la estimulación en experimentos de implantes agudos en humanos [Romero 2006]. Se determinan los umbrales de corriente para la estimulación, la

cantidad de canales del implante que conservan su funcionalidad tras el proceso quirúrgico y, si es necesario, redirigir los eventos producidos por la herramienta para adecuarse a la posición en la que el paciente perciba los fosfenos. Todo este proceso está ideado para ser efectuado automáticamente a partir de unos bancos de pruebas que debe realizar el paciente, bajo su propio control, eliminando la intervención de terceras personas.

3.4 Conclusiones

En este capítulo se ha comenzado realizando una introducción al proyecto europeo CORTIVIS, en el que ha trabajado nuestro grupo, cuyo principal objetivo era el desarrollo de una neuroprótesis cortical visual para invidentes. Seguidamente hemos presentado el esquema general de la prótesis que se ha concebido en el seno del proyecto.

Posteriormente nos hemos centrado en la descripción de Retiner, un entorno de diseño de sistemas de procesamiento visuales, cuyas capacidades se extienden desde el modelado y la simulación de los mismos, hasta su ajuste y comparación de resultados con datos biológicos obtenidos en experimentos de registro de actividad neuronal.

Algunas aportaciones del trabajo de esta tesis doctoral abarcan el desarrollo, bien individualmente bien en equipo, de diversos módulos componentes que incorpora Retiner. En concreto, se ha trabajado en los módulos de entrada del sistema, tanto en la extensión de la captura desde una cámara a dos cámaras haciendo uso del toolbox de adquisición de imágenes, como en el que permite procesar secuencias previamente almacenadas en archivos de vídeo en formato AVI. Esta característica se hizo necesaria cuando el tiempo de respuesta en el procesamiento on-line, con captura de cámara, creció por el aumento en la complejidad de los sistemas modelados.

Como hemos comentado con anterioridad, el módulo de cálculo de disparidades y estimación de la profundidad es una de las principales contribuciones de esta tesis al sistema completo. Los resultados obtenidos en esta etapa nos permiten mejorar el proceso de selección de información

visual, incorporando otra fuente de información que está presente en el sistema visual humano.

Otro módulo que ha requerido nuestro esfuerzo ha sido el que permite procesar archivos de vídeo con una configuración temporal personalizada por el usuario, afectando también al filtrado temporal, para el que se permite seleccionar un número variable de imágenes previas con las que computar las diferencias. La resolución temporal configurable nos permite por un lado, afinar el modelo incluso a los tiempos de respuesta de las neuronas observados en experimentos biológicos, y por otro, eliminar las discontinuidades en el procesamiento de imágenes en vivo capturadas directamente de la cámara.

También recogemos entre las aportaciones propias el módulo software encargado de la definición de los distintos tipos de campos receptivos (junto con el formato del archivo de salida de esta información, que permite su modificación manual a gusto del usuario). Gracias a este módulo, Retiner nos permite modelar la diferencia de resolución dependiente de la excentricidad que se observa en la retina humana, además de las diferentes formas de los campos receptivos que presentan otras especies animales. Al mismo tiempo provee a los sistemas de ayuda a la baja visión de la posibilidad de definir transformaciones de la imagen de entrada al objeto de aprovechar el resto visual de los pacientes, permitiendo la derivación de información visual hacia las zonas del campo visual que conservan su funcionalidad.

Finalmente, se ha trabajado en la generación de la matriz de actividad, teniendo en cuenta los campos receptivos previamente determinados, y de la salida graduada, que pueden ser utilizadas mediante visualizadores para la realización de experimentos psicofísicos. Estos experimentos, realizados con personas que presentan visión normal, nos van a permitir determinar la capacidad visual máxima que se va a poder recuperar, para una determinada configuración del sistema.

En resumen, Retiner provee de características específicas para el modelado de sistemas de procesamiento visual bioinspirado, que lo convierten en una potente herramienta para investigadores que trabajen tanto en aproximaciones neuroprotésicas, incluyendo retinianas, para el nervio óptico y corticales, como en otro tipo de sistemas electrónico-informáticos que tengan como finalidad la rehabilitación del sistema visual. El mismo entorno permite la simulación y ajuste de los modelos diseñados, generando resultados que posteriormente pueden ser contrastados con los datos recogidos en experimentos de estimulación visual en animales. En la figura 3.17 hemos esquematizado la arquitectura

de procesamiento de Retiner junto con las distintas posibilidades de aplicación que ofrece en el contexto de los sistemas artificiales de rehabilitación visual.

Modelo de Procesamiento de Visión Estéreo

En capítulos anteriores se ha puesto de manifiesto la importancia que tiene la información de profundidad para el sistema visual humano, además de los mecanismos neuronales y motores que permiten su obtención. El contenido del presente capítulo está orientado, en primer lugar, a los fundamentos de los sistemas de visión estereo por computador, lo que nos ocupará la sección 4.1. En la sección 4.2 se abunda en los detalles relativos a la implementación del sistema estereo, describiendo los distintos algoritmos que intervienen. Finalmente, en la sección 4.3 se pone de relieve la utilidad de la información de profundidad en los sistemas de rehabilitación visual, y se presenta nuestra propuesta de incorporación al modelo de procesamiento que implementan dichos sistemas.

4.1 Visión binocular y estereovisión

El sistema visual humano se caracteriza por ser *binocular*, ya los dos ojos actúan como puntos de entrada de información al mismo.

Esta propiedad, junto a los mecanismos para producir la fusión de las imágenes procedentes de las retinas, los medios para modificar la convergencia ocular, y la evolución de estructuras neuronales que traducen las diferencias de las imágenes en indicios de profundidad relativa, nos provee de una información muy valiosa sobre nuestro entorno, la distancia que nos separa de los distintos elementos presentes en una escena visual.

La búsqueda de métodos y algoritmos para conseguir la información de profundidad de forma artificial a partir de pares de imágenes capturadas de una escena ha sido un área de mucho interés para la investigación durante décadas. La extracción de esta información no es una simple comparación entre las imágenes disponibles, sino que requiere conocimiento acerca de las características de la(s) cámara(s) utilizada(s) para la obtención del par de imágenes, así como de la geometría del sistema formado por las cámaras. Además tendremos que tomar algunas decisiones más durante el proceso que determinarán la solución final a la que lleguemos.

A continuación vamos a detallar los aspectos de la cámara y de la geometría del sistema que intervienen en los cálculos de la estructura tridimensional de la escena.

4.1.1 Modelo de cámara. Calibración

La calibración de la cámara es el problema de determinar los elementos que gobiernan la relación o transformación entre la imagen 2D percibida por una cámara y la información 3D del objeto real. Existen multitud de métodos para llevar a cabo esta calibración. Una recopilación clásica muy completa puede encontrarse en [Tsai 1989].

El modelo de cámara más comúnmente utilizado en visión por ordenador es uno muy sencillo, el denominado 'ojo de aguja' o *pin-hole*, que podemos ver en la figura 4.1.

La imagen se forma sobre el *plano imagen o retiniano* \mathcal{R} a través de una operación llamada *proyección de perspectiva*. La imagen \mathbf{m} de un punto real \mathbf{w} es la intersección del plano imagen con la recta que une \mathbf{w} con el *centro óptico* \mathbf{c} , situado a una distancia f (*distancia focal*) del plano imagen.

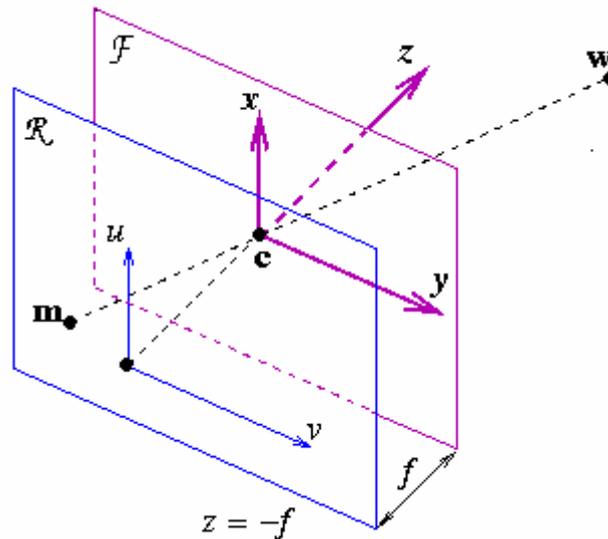


Figura 4.1: Modelo de cámara pin-hole.

El *eje óptico* es la línea recta que pasa por \mathbf{c} y es perpendicular al plano retiniano, al que corta en un punto conocido como *centro de la imagen*. El plano paralelo al retiniano y que contiene a \mathbf{c} es el *plano focal* \mathcal{F} . Los puntos situados en el plano focal tienen su imagen en el infinito.

El modelo se llama pin-hole porque refleja la situación que se produciría si el plano focal fuera un plano físico real con un agujero en \mathbf{c} .

Se ha comprobado que un modelo tan sencillo de cámara puede representar con gran exactitud la geometría y óptica de la mayor parte de las cámaras modernas [Faugeras 1993a]. En algunos casos, sin embargo, es necesario considerar distorsiones radial y tangencial respecto del modelo, para lo que se usan métodos distintos con cada técnica de calibración de cámara.

4.1.1.1 Planteamiento general

Supongamos que tenemos un punto $\mathbf{w}(x, y, z)$ en el espacio y su imagen $\mathbf{m}(u, v)$, como podemos ver en la figura 4.1. Las coordenadas de \mathbf{w} se dan respecto de un sistema de referencia del mundo, con origen en algún punto del espacio, y cuyas coordenadas se expresarán en unidades métricas de distancia.

Las coordenadas de \mathbf{m} se expresan respecto del sistema de referencia de la imagen (o de la cámara), que habitualmente es el de la figura 4.2 y cuyas coordenadas son en píxeles.

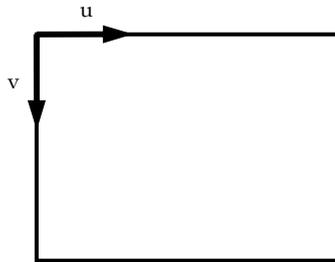


Figura 4.2: Sistema de referencia de la imagen.

El proceso de calibración se encarga de encontrar la relación entre \mathbf{w} y \mathbf{m} . Esta relación será función de una serie de parámetros, cuya determinación es el objetivo de la calibración. Estos parámetros se dividen en *intrínsecos*, que no dependen de la posición y orientación de la cámara, sino que son propios de la misma como dispositivo óptico, y *extrínsecos*, que vienen dados por la posición y orientación de la cámara en el espacio.

Los parámetros intrínsecos son los siguientes:

- Posición del centro de la imagen, respecto del sistema de referencia de la cámara.
- Distancia focal, f .
- Factores de escala para los ejes X e Y (para convertir píxeles en unidades métricas).

Los parámetros extrínsecos quedan perfectamente definidos si conocemos:

- Coordenadas del centro óptico $\mathbf{c}(x, y, z)$ respecto del sistema de referencia del mundo.

- Orientación de la cámara en el espacio, definida por tres ángulos respecto a los ejes X, Y, Z.

El objetivo de la calibración es hallar la relación entre un punto en la escena y su imagen. Los métodos de calibración pueden dividirse en explícitos, donde se hallan los valores de los diferentes parámetros, e implícitos, donde éstos no se calculan explícitamente, sino que se obtienen valores de combinaciones de los parámetros. Otra división que puede hacerse es entre calibración métrica, que utiliza un conjunto de puntos de coordenadas conocidas respecto a un sistema de referencia, y calibración débil, que se lleva a cabo durante el funcionamiento del sistema estéreo usando puntos para los que se conoce su correspondiente. Esto tiene la ventaja de que puede hacerse on-line, lo que permite corregir cambios que puedan producirse en la posición de las cámaras por cualquier causa, y el inconveniente de que los resultados no se refieren a unidades métricas. La calibración débil es habitual en robótica móvil complementada con sensores de distancia que proporcionan la referencia métrica.

Se verá a continuación con cierto detalle el método de la matriz de perspectiva, una de las técnicas de calibración más sencilla, ya que no se considera la distorsión y es totalmente lineal. Esto servirá para tener una idea del problema, independientemente de que existen otras técnicas que también son válidas. Luego se hará una comparación de las ventajas e inconvenientes de cada método.

4.1.1.2 Método de la matriz de perspectiva

Con el método de la matriz de perspectiva se halla la relación entre un punto $\mathbf{w}(x, y, z)$ y su imagen $\mathbf{m}(u, v)$, demostrándose que $\mathbf{m}^* = T \mathbf{w}^*$, siendo \mathbf{m}^* y \mathbf{w}^* las representaciones en coordenadas proyectivas u homogéneas de \mathbf{m} y \mathbf{w} respectivamente, y la matriz T el producto de cuatro matrices: de traslación, rotación, escala y perspectiva. La determinación de esta matriz T será el objetivo de la calibración de la cámara.

El método de la matriz de perspectiva es implícito. La matriz T que relaciona \mathbf{w} con \mathbf{m} es una matriz 3x4, 12 coeficientes que se reducen luego a 11. Estos once coeficientes son función de los nueve parámetros de calibración, cuyos valores no se obtienen en principio de forma explícita. Pueden calcularse posteriormente, pero esto no es necesario en el caso de la estereovisión.

Ahora veremos el proceso de obtención de la matriz T . Debemos considerar dos aspectos: el cambio de sistema de referencia (imagen – mundo) y la proyección de un punto a través del centro óptico (geometría proyectiva).

A) Cambio de sistema de referencia:

Se trata de relacionar el sistema de referencia de la imagen en píxeles con el sistema de referencia del mundo en unidades métricas. La transformación de un sistema al otro viene dada por el producto de tres matrices: matriz de cambio de escala, que tiene en cuenta la relación píxeles – unidades métricas y la posición del centro de la imagen en el sistema de coordenadas de la imagen; matriz de traslación, que define la posición del centro óptico respecto al sistema de referencia del mundo; y matriz de rotación, que define la orientación del eje óptico de la cámara respecto a este mismo sistema. El proceso de obtención de estas matrices se puede ver detalladamente en [Faugeras 1993a].

Con estas tres matrices podemos, dado un punto $\mathbf{m}(u, v)$, conocer sus coordenadas (x, y, z) en el sistema de referencia del mundo. Pero lo que se quiere es conocer las coordenadas del punto \mathbf{w} cuya proyección es \mathbf{m} (en realidad es una recta, existe un número infinito de puntos cuya proyección es \mathbf{m}). Pasamos a estudiar el problema de la proyección.

B) Geometría proyectiva y matriz de perspectiva:

En primer lugar consideraremos un caso muy sencillo. El sistema de referencia de la cámara tiene su origen en el centro de la imagen, y el sistema de referencia del mundo tiene origen en el centro óptico \mathbf{c} , y el eje Z coincidente con el eje óptico (véase figura 4.1).

¿Qué relación existe entre las coordenadas de $\mathbf{w}(x, y, z)$ y su proyección en el plano imagen $\mathbf{m}(u, v)$? Lo vemos en una dimensión en la figura 4.3.

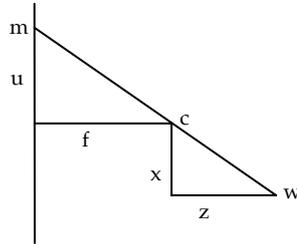


Figura 4.3: Geometría de la proyección.

Por semejanza de triángulos tenemos que:

$$-\frac{u}{f} = \frac{x}{z} \Rightarrow u = -\frac{f \cdot x}{z} \quad (4.1)$$

$$\text{Análogamente, } v = -\frac{f \cdot y}{z} \quad (4.2)$$

Por tanto, para un punto $\mathbf{w}(x, y, z)$, conocemos su proyección $\mathbf{m}(u, v)$. Pero hay un problema con las ecuaciones (4.1) y (4.2). No son lineales, hay una z en el denominador. Es muy recomendable trabajar con ecuaciones lineales para aprovechar las potentes herramientas del álgebra lineal. Definiremos las coordenadas proyectivas y homogéneas para un punto $\mathbf{m}(u, v)$ como $\mathbf{m}^*(U, V, S)$, siendo S un factor de escala tal que $u = \frac{U}{S}$ y $v = \frac{V}{S}$. De este modo, podemos plantear la siguiente ecuación lineal matricial: $\mathbf{m}^* = M \mathbf{w}^*$, siendo \mathbf{m}^* y \mathbf{w}^* los puntos imagen y espacial en coordenadas proyectivas, y M la llamada matriz de perspectiva:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (4.3)$$

Se demuestra que esta relación, que hemos visto para dos sistemas de coordenadas sencillos, se mantiene para cualesquiera otros sistemas de referencia. La matriz de perspectiva M se complicaría.

C) Caso general: cambio de sistema de referencia y aplicación de la matriz de perspectiva:

En un caso general, la relación de un punto $\mathbf{w}(x, y, z)$ con su proyección $\mathbf{m}(u, v)$ deberá tener en cuenta la transformación de perspectiva y las diferentes etapas del cambio de sistema de coordenadas que se vieron anteriormente. Todas estas transformaciones podrán expresarse de forma matricial en coordenadas proyectivas, de modo que tendremos la siguiente ecuación:

$$\mathbf{m}^* = S M R G \mathbf{w}^* = T \mathbf{w}^* \quad (4.4)$$

siendo \mathbf{m}^* : coordenadas proyectivas del punto imagen: $[U, V, S]^T$.

\mathbf{w}^* : coordenadas proyectivas del punto espacial: $[x, y, z, 1]^T$.

M: matriz de perspectiva.

S: matriz de escala, que incluye los factores de escala y el cambio de origen del sistema de coordenadas de la cámara respecto al centro de la imagen.

G: matriz de traslación, que refleja el desplazamiento del centro óptico respecto al origen del sistema referencia del mundo.

R: matriz de rotación, que caracteriza la orientación del eje óptico en el espacio.

T: matriz resultante, que engloba todo.

Dado un punto $\mathbf{w}(x, y, z)$, si conocemos la matriz T podemos obtener con (4.4) su proyección $\mathbf{m}(u, v)$, y a la inversa, conocido $\mathbf{m}(u, v)$, puede obtenerse la recta formada por todos los puntos $\mathbf{w}(x, y, z)$ cuya proyección es $\mathbf{m}(u, v)$. Recordemos que el objetivo de la calibración es la determinación de la matriz T.

Para ello tenemos la ecuación $\mathbf{m}^* = T \mathbf{w}^*$,

$$\begin{pmatrix} U \\ V \\ S \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (4.5)$$

y queremos calcular los doce elementos de T. La matriz T está definida en función de un factor de escala, ya que relaciona coordenadas proyectivas. Puede demostrarse (ver [Newman 1979]) que t_{34} representa ese factor de escala entre los sistemas de coordenadas del mundo y de la imagen, por lo

que no puede ser cero. Podemos dividir los dos lados de la ecuación por t_{34} , con lo que nos queda:

$$\begin{pmatrix} U \\ V \\ S \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (4.6)$$

Debemos calcular los once elementos desconocidos de T. Desarrollando tenemos:

$$\begin{cases} U = t_{11}x + t_{12}y + t_{13}z + t_{14} \\ V = t_{21}x + t_{22}y + t_{23}z + t_{24} \\ S = t_{31}x + t_{32}y + t_{33}z + 1 \end{cases} \quad (4.7)$$

$$u = \frac{U}{S} \Rightarrow U = uS \Rightarrow u(t_{31}x + t_{32}y + t_{33}z + 1) = t_{11}x + t_{12}y + t_{13}z + t_{14} \quad (4.8)$$

$$v = \frac{V}{S} \Rightarrow V = vS \Rightarrow v(t_{31}x + t_{32}y + t_{33}z + 1) = t_{21}x + t_{22}y + t_{23}z + t_{24} \quad (4.9)$$

Para cada punto $\mathbf{w}(x, y, z)$ del que se conozca su proyección $\mathbf{m}(u, v)$, se tienen dos ecuaciones con once incógnitas, los elementos de T. Por tanto, necesitamos al menos seis puntos para resolver el sistema de ecuaciones. En la práctica si se eligen sólo seis puntos los resultados son muy malos; se toman más puntos (24 al menos), con lo que se tiene un sistema de ecuaciones sobredeterminado, que puede resolverse por distintos métodos. Lo más sencillo es estimar los elementos de T por mínimos cuadrados. Hay otros métodos no lineales, habitualmente iterativos, que pueden dar resultados algo mejores, pero con un coste computacional muy alto.

Hay que tener en cuenta que los resultados de la calibración nunca son exactos, en primer lugar porque se trata de ajustar un modelo a un dispositivo real, lo que ya supone un error, y también por los errores en la medición de la posición de los puntos, tanto del mundo como de la imagen. Con el método de mínimos cuadrados se trata de minimizar este error.

4.1.1.3 Comparación de métodos de calibración

En la recopilación de [Tsai 1989] se hacen algunas recomendaciones para la elección del método de calibración. Lógicamente el método de Tsai aparece más favorecido, aunque esto no está tan claro.

Si no se considera la distorsión, el método más sencillo es el de la matriz de perspectiva visto anteriormente, ya que es totalmente lineal. Tiene sin embargo un par de inconvenientes. En primer lugar, al ser un método implícito, el número de incógnitas es superior al número real de grados de libertad, es decir, las incógnitas no son linealmente independientes. Según Tsai, esta parametrización redundante puede ocasionar que una combinación errónea de los parámetros permita un buen ajuste entre la predicción del modelo y los datos experimentales, especialmente en situaciones ruidosas. Un inconveniente práctico es que, como se demuestra en [Faugeras 1993a], los puntos del conjunto de calibración no pueden ser coplanares. Esto implica la necesidad bien de un plano móvil o de un elevador de la cámara en el eje Z.

Por su parte, el método introducido por Tsai ([Tsai 1986], [Tsai 1987]) es un método explícito que incorpora uno o dos términos de distorsión radial. A causa de esto no es totalmente lineal, y para la estimación de cuatro parámetros se debe usar algún método no lineal de métodos cuadrados (como el de Marquardt [Press 1992]). Sin embargo, siempre según Tsai, se necesitan sólo una o dos iteraciones. Este método permite que los puntos de calibración sean coplanares, lo que simplifica el montaje. Hay algún detalle que no está claro en el método, como la elección del centro del buffer de imágenes (frame-buffer) como centro de la imagen, pero éste y otros detalles se corrigen en un estudio posterior [Lenz 1987]. Por otra parte, algún autor ([Shi 1993]) asegura que en caso de que la distorsión sea relativamente pequeña, los resultados del método de Tsai son incluso peores que los del método lineal de la matriz de perspectiva, que no la tiene en cuenta.

Otro método interesante, posterior a la recopilación de Tsai, es el de Sid-Ahmed y Boraie [Sid-Ahmed 1990]. Estos autores consideran, al contrario que Tsai, que la distorsión tangencial no puede despreciarse, e introducen cinco términos de distorsión, tres radiales y dos tangenciales. Se llega a un sistema no lineal con 16 incógnitas, que se resuelve por algún método iterativo, como el de Marquardt, al igual que en el caso de Tsai. Los resultados del método de la matriz de perspectiva se usan como estimación inicial (por tanto los puntos no pueden ser coplanares).

Existen también otros métodos, algunos de ellos incluidos en la recopilación de Tsai y otros posteriores. En general los autores comparan sus resultados con los de Tsai (el método de Tsai se ha convertido en un estándar) y aseguran que los mejoran, pero suelen ser métodos bastante más complejos.

En todo caso, no es fácil hacer comparaciones ya que los resultados pueden ser dependientes en gran medida de la precisión conseguida en el montaje de calibración, lo que incluye la precisión en la medición de las coordenadas 3D de los puntos de calibración y la precisión subpíxel en la determinación de las correspondientes coordenadas en la imagen, entre otros aspectos.

4.1.1.4 Calibración débil

En algunas aplicaciones el uso de un conjunto de puntos de coordenadas conocidas es complicado. Por ejemplo, en Robótica móvil, si el robot debe desplazarse por un terreno desigual, es posible que la posición de las cámaras pueda variar durante la operación, con lo que la calibración previa ya no valdría. Sería necesaria una nueva calibración on-line. Esto es posible mediante las llamadas técnicas de calibración débil, que solamente requieren el establecimiento de un pequeño número de correspondencias entre píxeles de las dos imágenes, a partir de las cuales se reconstruye la geometría epipolar. Evidentemente la reconstrucción 3-D no es posible en un espacio métrico, por lo que para conseguir una información cuantitativa de distancia se precisa la ayuda de algún sensor adicional (por ejemplo un range-finder láser). Debido a la gran importancia actual de la investigación en Robótica móvil (robots para Marte, etc.), hay un gran número de publicaciones sobre calibración débil: [Kimura 2001], [Zhang 1998].

4.1.2 Geometría de un sistema estéreo

Además del problema de calibración de las cámaras que acabamos de tratar, existen otra serie de parámetros relacionados con sus posiciones relativas, que definen un sistema estéreo. En primer lugar vamos a estudiar

un caso simple de esta geometría, que más tarde generalizaremos para abarcar cualquier sistema estéreo que podamos imaginar.

La figura 4.4 nos muestra una visión desde arriba de este sistema simple, un sistema estéreo compuesto por dos cámaras de tipo pin-hole. Los planos imagen izquierdo y derecho son coplanares, y están representados por los segmentos I_l y I_r respectivamente. O_l y O_r son los centros de proyección y los ejes ópticos son paralelos. Por esta razón, el punto de corte de ambos ejes, denominado *punto de fijación* del sistema, se encuentra en el infinito.

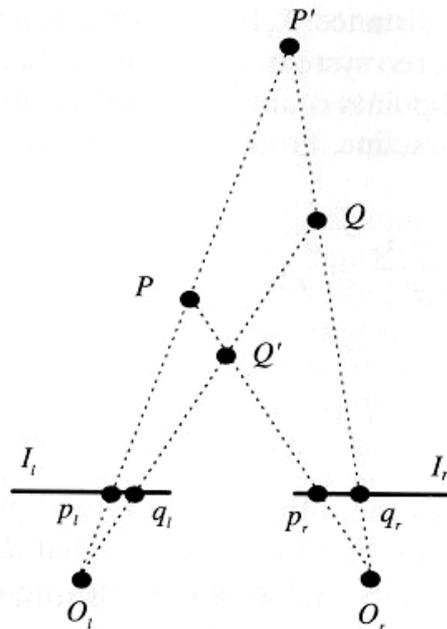


Figura 4.4: Sistema de visión estereo simple.

La forma en la cual el estereo determina la posición en el espacio de P y Q es por triangulación, es decir, por intersección de los rayos definidos por los centros de proyección de cada una de las cámaras y los correspondientes puntos de proyección en las retinas de P y Q , p_l , p_r , q_l , q_r respectivamente. La triangulación depende de forma muy importante del problema de las *correspondencias*: si (p_l, p_r) y (q_l, q_r) son escogidos como parejas de puntos correspondientes en las imágenes, la intersección de los rayos $O_l p_l - O_r p_r$ y $O_l q_l - O_r q_r$ conduce a interpretar los puntos imágenes como proyecciones de P y Q ; pero si (p_l, q_r) y (q_l, p_r) son los puntos que se establecen como correspondientes, la triangulación nos devuelve los puntos P' y Q' . Hemos de señalar que aunque ambas interpretaciones son

muy diferentes, ambas pueden considerarse en pie de igualdad cara a su posible aceptación.

Consideremos ahora la ecuación que nos da el punto P en función de sus proyecciones. Notemos por T a la distancia entre los centros de proyección (ver figura 4.5). Este segmento se denomina *línea base* del sistema estéreo, y a su longitud la denominaremos distancia base.

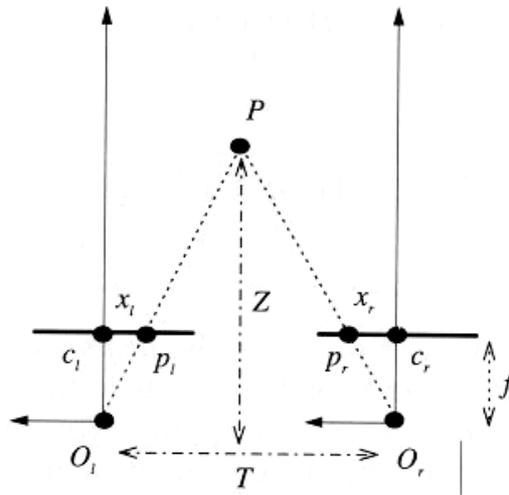


Figura 4.5: Geometría del sistema estéreo.

Sean x_l y x_r las coordenadas de p_l y p_r respecto a los puntos principales c_l y c_r , f la distancia focal común y Z la distancia de P a la línea de base. De las relaciones entre triángulos similares (p_l, P, p_r) y (O_i, P, O_r) tenemos:

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z} \Rightarrow Z = f \frac{T}{d} \quad (4.10)$$

donde $d = x_r - x_l$ la *disparidad*, mide la diferencia en posición de la retina entre los puntos correspondientes de ambas imágenes. De la ecuación (4.10) vemos que la profundidad es inversamente proporcional a la disparidad.

Los parámetros de un sistema estéreo son f , T , c_l y c_r . La calibración de un sistema estéreo consiste en calcular dichos parámetros. Los

parámetros intrínsecos de un sistema estéreo son los ya estudiados para el sistema de una única cámara. Por tanto, en el caso de imágenes estéreo, el problema de reconstrucción puede en muchos casos reducirse a un problema de calibración del sistema.

A pesar de lo dicho, es posible calcular mucha información de un sistema estéreo sin haberlo calibrado previamente. Es decir, se pueden encontrar relaciones geométricas que ligan las posiciones de los rasgos presentes en un par estéreo independientemente de la calibración del mismo. Esta geometría, denominada geometría epipolar, se estudia a continuación.

4.1.2.1 Geometría epipolar

En este apartado vamos a estudiar la geometría de un sistema estéreo de la forma más general, lo que nos permitirá clarificar qué información es necesaria para llevar a cabo la búsqueda de puntos correspondientes de una forma óptima, como veremos, reduciendo la búsqueda de estos puntos a líneas.

La figura 4.6 nos muestra la geometría epipolar:

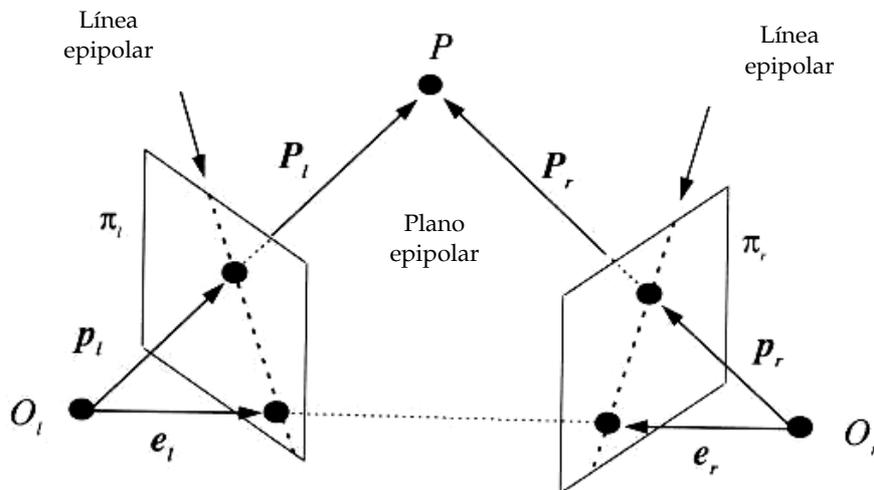


Figura 4.6: Geometría epipolar de un sistema estéreo.

Esta imagen nos muestra dos cámaras de tipo pin-hole, sus centros de proyección, O_l y O_r , y sus planos imagen o retinas en coordenadas normalizadas. Las longitudes focales se notan como f_l y f_r . Como es normal, cada cámara identifica un sistema de referencia 3D, el origen del cual se sitúa en el centro de proyección de la cámara y el eje Z con el eje óptico. Los vectores $P_l = [X_l, Y_l, Z_l]^T$ y $P_r = [X_r, Y_r, Z_r]^T$ se refieren al mismo punto 3D, P, como vectores en los sistemas de referencia de la cámara izquierda y derecha respectivamente. Los vectores $p_l = [x_l, y_l, z_l]^T$ y $p_r = [x_r, y_r, z_r]^T$ definen las proyecciones del punto P en la imagen izquierda y derecha respectivamente y están expresados en su correspondiente sistema de referencia. Evidentemente, para todos los puntos de las imágenes tenemos $z_l = f_l$ y $z_r = f_r$ de acuerdo a la imagen que sea.

Los sistemas de referencia de las cámaras izquierda y derecha están relacionados a través de los parámetros extrínsecos. Éstos definen una transformación rígida en el espacio 3D definida por un vector de traslación $T=O_r-O_l$ y una matriz de rotación R. Dado un punto P en el espacio, la relación entre P_l y P_r es por tanto: $P_r = R (P_l - T)$.

El nombre de la geometría epipolar es debido a que los puntos en los cuales la recta que une los centros de proyección de las cámaras corta a los planos de proyección se llaman *epipolos*.

Notaremos por e_l y e_r el epipolo izquierdo y derecho respectivamente. Por construcción ambos epipolos representan la proyección en su correspondiente plano la imagen del centro de proyección de la otra cámara. En el caso de que uno de los planos imagen sea paralelo a la recta que une los centros de proyección, el epipolo de ese plano estará situado en el infinito.

La relación entre un punto del espacio 3D y su proyección se describe por las ecuaciones usuales de proyección de perspectiva, que en su forma vectorial se escriben $p_l = \frac{f_l}{Z_l} P_l$, $p_r = \frac{f_r}{Z_r} P_r$.

La importancia práctica de la geometría epipolar arranca del hecho que el plano definido por P, O_l y O_r , llamado plano epipolar, interseca cada imagen en una línea llamada línea epipolar. Consideremos la terna P, p_l y p_r . Dado p_l , P puede caer en cualquier punto del rayo definido por O_l y p_l . Pero dado que la imagen de este rayo en la imagen derecha es la recta

epipolar a través del punto correspondiente p_r , dicho punto debe estar sobre la línea epipolar. Esta correspondencia establece una aplicación entre puntos de la imagen izquierda y rectas de la imagen derecha, y viceversa. Una consecuencia de esa correspondencia es, que dado que todos los rayos pasan por construcción por el centro de proyección, todas las rectas epipolares deben pasar por el epipolo.

De esta forma, la búsqueda de correspondencias se reduce a un problema 1D. Alternativamente, este conocimiento también se puede usar para verificar si una potencial pareja de puntos correspondientes, lo son de verdad o no. Esta técnica es normalmente una de las más efectivas para detectar las posibles falsas correspondencias debidas al problema de la *oclusión* (puntos del espacio que tienen proyección sólo en uno de los planos de imagen, quedando fuera del otro, y por tanto, no tiene su correspondiente).

4.2 Métodos de cálculo de correspondencias o *matching*

Ahora que hemos restringido el área de búsqueda de los puntos correspondientes, vamos a estudiar distintos métodos para establecer / detectar estas correspondencias.

Los distintos algoritmos para el cálculo de las correspondencias pueden ser clasificados en las siguientes categorías [Barnard 1982]:

- Basados en características: Extraen características de interés de la imagen (bordes, contornos), con los que realizan el matching. En principio son más rápidos, ya que sólo usan un pequeño subconjunto de los píxeles de la imagen, pero solamente proporcionan mapas de profundidad dispersos y pueden fallar si no se encuentran primitivas de interés en la imagen.
 - Basados en área: Se lleva a cabo la correlación de los niveles de gris en ventanas de las distintas imágenes, considerando que en puntos correspondientes existirá una similitud de niveles de intensidad luminosa. Estos
-

métodos son difíciles de aplicar en zonas de poca textura y en bordes de oclusión. Permiten obtener mapas de profundidades densos.

- Otros: Métodos no basados ni en correlación ni en extracción de características, sino que utilizan alguna otra metodología.

Una comparación de los méritos de cada método no es fácil, ya que los distintos investigadores trabajan con conjuntos de datos muy diversos.

A continuación vamos a comentar distintos problemas que dificultan el matching [Cochran 1992]:

- Variación fotométrica: La luz que recoge la cámara depende de la posición de la cámara respecto a la escena y de ruidos y no linealidades en la cámara. Por tanto si dos cámaras ven una escena desde distinto punto de vista la intensidad en puntos correspondientes puede ser distinta.
- Oclusiones: Debidas a discontinuidades en profundidad que causan que una parte de la escena sea vista por una sola de las cámaras.
- Textura repetitiva: La presencia de texturas repetitivas hace que aparezcan como válidas múltiples correspondencias.
- Falta de textura: En una zona de la escena sin ningún tipo de textura no será posible aplicar ningún tipo de correspondencia

4.2.1 Métodos basados en características

Llevan a cabo la correspondencia sobre características más abstractas y menos sensibles al ruido que la intensidad luminosa de los píxeles (normalmente bordes o contornos, aunque también pueden ser de

más alto nivel). Una vez obtenidas las características se utilizan distintas técnicas para determinar la correspondencia:

- Técnicas de relajación: se hace una hipótesis inicial de las correspondencias, y se deja que estas se reorganicen propagando alguna de las restricciones que se han mencionado (continuidad, gradiente de disparidad, etc.).
- Programación dinámica: El problema de encontrar correspondencias entre primitivas de la imagen puede enfocarse como la minimización de una función de coste con gran número de variables (las posibles correspondencias). La programación dinámica es una técnica que permite abordar este tipo de problemas, sustituyendo la minimización de una función de muchas variables por la de muchas funciones de una variable. Algunos trabajos en este campo son los de Baker y Binford [*Baker 1981*], Ohta y Kanade [*Ohta 1985*] y otros como [*Belhumeur 1996*], [*Birchfield 1996*], [*Birchfield 1998*], [*Cox 1996*], [*Intille 1994*].
- Predicción y verificación. Los elementos que se ponen en correspondencia son de más alto nivel que píxeles. Algunos trabajos son los de Medioni y Nevatia [*Medioni 1985*], Ayache y Faverjon [*Ayache 1987a*], Ayache y Lustman [*Ayache 1987b*], Yachida [*Kitamura 1990*] y Robert [*Robert 1991*].

Los métodos basados en características en general son más exactos, ya que estas pueden extraerse con precisión subpíxel, y por tanto la reconstrucción 3D también es más precisa. Por otra parte el número de píxeles tratado es menor, lo que permite una mayor rapidez. Estas ventajas hicieron que en un principio se dedicara mucha más atención a estos algoritmos. Algunos de ellos son Marr-Poggio [*Marr 1979*], PMF [*Pollard 1985*], Grimson [*Grimson 1981a*] (una implementación mejorada de Marr-Poggio). Estos algoritmos se verán con detalle más adelante.

Los métodos basados en características también presentan inconvenientes:

- Proporcionan mapas de profundidades dispersos.
 - Una de las restricciones que emplean habitualmente es la de continuidad de las superficies (limitación del gradiente de disparidad). Sin
-

embargo, los puntos tratados son normalmente los bordes, precisamente donde la suposición de continuidad es menos evidente.

- Normalmente requieren un flujo de control algorítmico complejo [Lane 1994]. Su implementación rápida es costosa en términos de hardware.

4.2.2 Métodos basados en área

Llevan a cabo la correlación cruzada de niveles de gris en ventanas de las imágenes. Por tanto la característica que usan es la intensidad luminosa de los píxeles. Históricamente se han aplicado en análisis de fotografías aéreas, donde la superficie varía de forma suave y continua. La principal ventaja de estos métodos es que proporcionan mapas de profundidades densos, pero en cambio son más sensibles al ruido y tienen problemas si no hay textura o hay cambios bruscos de profundidad.

La razón de que estos métodos estén ganando popularidad actualmente es que existen a precios razonables dispositivos de procesamiento de imagen capaces de llevar a cabo rápidamente operaciones de convolución y correlación (DSPs y FPGAs), con rendimientos muy superiores a los procesadores de propósito general.

Toda la problemática asociada a los métodos de correlación se tratará con mayor detalle más adelante, en la sección 4.2.4.

4.2.3 Algoritmos clásicos

En este apartado se tratarán los algoritmos de correspondencia que pueden ser calificados como clásicos, debido a su antigüedad. En su gran mayoría están basados en características, usando alguna de las técnicas vistas antes.

4.2.3.1 Marr – Poggio

El algoritmo original de Marr-Poggio [Marr 1979] se basa en técnicas de relajación, usando para la desambiguación las restricciones de unicidad y continuidad.

En primer lugar se extraen las características (bordes), usando el método de zero-crossing de la laplaciana de la gaussiana. Con esto se determinan los píxeles que serán considerados para el matching. Para cada píxel considerado en la primera imagen se establece una medida de confianza para su correspondencia con cada píxel considerado en la segunda imagen. La manera más sencilla de establecer esa medida es por la similitud del nivel de gris. A continuación estas medidas se actualizan usando la restricción de continuidad, de modo que una correspondencia se refuerza si en su vecindad hay otras correspondencias establecidas. Por contra, la restricción de unicidad hace que la medida de confianza disminuya si hay otros píxeles compitiendo para la misma correspondencia o matching.

El funcionamiento de este algoritmo fue bueno con imágenes sintéticas (estereogramas de puntos aleatorios), pero no tanto con imágenes reales.

4.2.3.2 Marr – Poggio – Grimson

Grimson llevó a cabo diferentes implementaciones mejoradas del algoritmo de Marr-Poggio [Grimson 1981a], [Grimson 1985]. Se utiliza la técnica de matching multietapa, con imágenes de distinta resolución. En primer lugar se hace el matching con imágenes de resolución baja, y sus resultados sirven para guiar la búsqueda para resoluciones más altas. En este caso las distintas resoluciones se consiguen variando el parámetro σ en la detección de bordes con la laplaciana de la gaussiana. En el libro de Shirai [Shirai 1987] se encuentra una descripción completa del algoritmo, y se menciona la velocidad como uno de sus problemas.

Existe una nueva implementación del algoritmo Marr-Poggio-Grimson, modificado de modo que se efectúa mediante convoluciones [Stella 1992]. El número de puntos para el que se encuentra correspondencia es menor que para el algoritmo original de Grimson, con lo que los mapas de profundidades son aún más dispersos, pero el tiempo de cálculo se reduce al poder usarse hardware específico de convolución. Se menciona un tiempo de 100 segundos para obtener el mapa de disparidades para un par de imágenes de 256x256, usando una placa MATROX para PC.

4.2.3.3 PMF (Pollard – Mayhew – Frisby)

El algoritmo PMF [Pollard 1985] se basa en la restricción del gradiente de disparidad. La descripción del mismo en el libro de Sonka [Sonka 1993] es la siguiente:

- 1- Extracción de características (bordes) en ambas imágenes.
- 2- Para cada píxel considerado en una imagen, encontrar sus posibles correspondencias en la línea epipolar correspondiente.
- 3- Para cada posible correspondencia, incrementar su puntuación de acuerdo con el número de otras posibles correspondencias que no violen el límite fijado para el gradiente de disparidad.
- 4- El matching con la puntuación más alta para ambos píxeles es elegido como correcto. Por la restricción de unicidad ambos píxeles son eliminados de las demás combinaciones.
- 5- Volver a 2) y recalcular las puntuaciones teniendo en cuenta las correspondencias definitivas ya obtenidas.
- 6- Terminar cuando se hayan efectuado todas las correspondencias posibles.

Se observa que además de la limitación del gradiente de disparidad se tienen en cuenta las restricciones epipolar y de unicidad. En cuanto al mecanismo de puntuación, debe tenerse en cuenta que cuanto más lejanos estén los píxeles para los que se evalúa el gradiente, es más probable que se cumpla el criterio de limitación. Esto puede hacerse incluyendo en la puntuación un factor inverso a la distancia, o considerando para el criterio sólo los píxeles dentro de un cierto radio.

El algoritmo PMF ha funcionado con éxito en escenas artificiales y naturales. Aparentemente se ha convertido en un estándar, citado para comparación de prestaciones. Por ejemplo, en el artículo de Thacker sobre stretch-correlation [Thacker 1992] es mencionado para comparación de velocidades. Se dice que en una SUN Sparc II una implementación eficiente del PMF tarda dos minutos para una imagen de 512x512. Por otra parte, por su propia naturaleza, el algoritmo PMF podría ser paralelizable, con lo que este tiempo podría reducirse en gran manera.

En [Faugeras 1993a] aparece una descripción ligeramente diferente del algoritmo, en la que para la puntuación de las posibles correspondencias también se usan características propias de los puntos de borde, como la fuerza y la orientación, que se usan para desempate en caso de que haya varios candidatos que cumplan la restricción del gradiente de disparidad.

4.2.3.4 Nishihara

Nishihara [Nishihara 1984] propuso una técnica de correlación que no se aplica directamente sobre la imagen de intensidades luminosas, sino sobre la función signo de la convolución de las mismas con la laplaciana de una gaussiana. La laplaciana de la gaussiana es el operador utilizado por Marr-Poggio para la detección de bordes (zero-crossing de la laplaciana de la gaussiana). Nishihara demuestra que el signo de la laplaciana de la gaussiana es muy estable frente al ruido y que los picos de correlación de esta función son muy agudos, lo que permite el cálculo de la disparidad con gran precisión.

Para probar el rendimiento de su algoritmo, Nishihara se ayuda de la proyección de luz no estructurada sobre la escena, para crear imagen de textura, y utiliza una estrategia de matching multietapa, con gaussianas de distintos tamaños.

El método parece prometedor, pero no se ha encontrado información de trabajos más recientes que lo utilicen.

4.2.4 Métodos de correlación

En este apartado se harán ciertas consideraciones generales sobre los métodos de correlación, que actualmente están ganando popularidad, debido al importante incremento de los recursos de cálculo computacional disponibles.

El principio general de las técnicas de correlación es bien conocido y se ilustra en la figura 4.7.

Se toma una ventana fija en la imagen izquierda en torno al píxel para el que se busca correspondencia y en la imagen derecha se va

desplazando una ventana del mismo tamaño, llevándose a cabo una medida de similitud entre ambas. De este modo se obtiene una curva de correlación, la posición de cuyo máximo nos da el píxel de la imagen derecha que se corresponde con el píxel izquierdo considerado. La diferencia de coordenadas será la disparidad τ_0 buscada. Para obtener una precisión subpíxel puede ajustarse una curva de segundo grado en la vecindad del máximo de correlación.

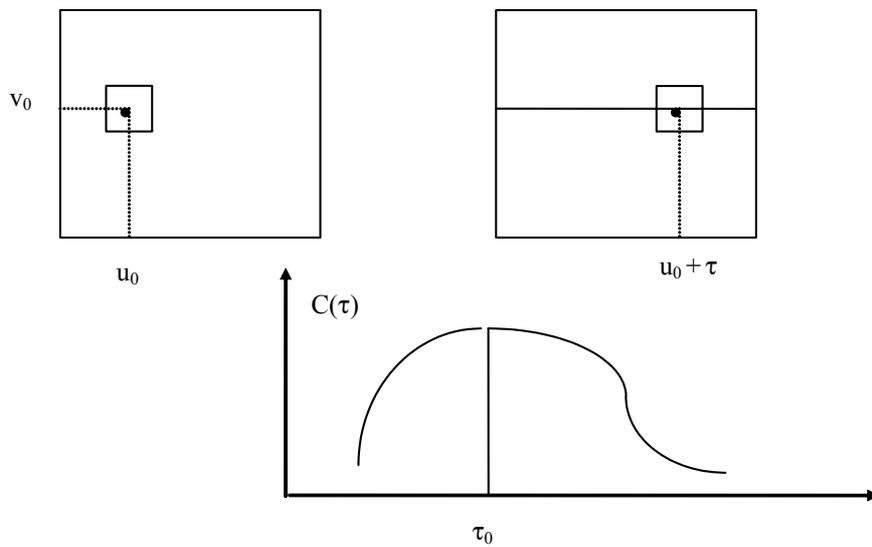


Figura 4.7: Técnica de correlación.

Algunas observaciones: Para que el desplazamiento de la ventana en la imagen derecha pueda realizarse sobre una línea, previamente se ha tenido que efectuar la *rectificación epipolar* (transformación de las imágenes para conseguir que las rectas epipolares sean paralelas entre sí, y horizontales). Por otra parte, si el máximo de la función de correlación no está bien definido la disparidad obtenida no será muy exacta.

También es importante entender que cualquier algoritmo estéreo que implique algún tipo de suavizado o promediado sobre un área en realidad está asumiendo, lo declare explícitamente o no, que las superficies de la escena pueden ser aproximadas localmente por planos fronto-paralelos [Faugeras 1993a], [Kanade 1994]. Esto equivale a decir que la disparidad es constante dentro de la ventana de correlación, lo que puede dar lugar a problemas en caso de gradientes de disparidad elevados.

A continuación se tratarán algunos aspectos críticos de los métodos de correlación: las distintas funciones de similitud, la selección del tamaño de ventana y los pros y contras de las aproximaciones jerárquicas. Posteriormente se abordarán algunos algoritmos concretos que aportan detalles de interés.

4.2.4.1 Funciones de similitud

La función de similitud es la encargada de evaluar la semejanza de las ventanas comparadas. Se han propuesto diferentes funciones para aplicaciones específicas. La más sencilla es la suma de diferencias de intensidad de los píxeles de la ventana:

$$D_s = \sum_1^n \sum_1^m |I_L(i,j) - I_R(i,j)| \quad (4.11)$$

En este caso, a mayor similitud menor es el valor de D_s . Mientras D_s es calculado secuencialmente su valor va incrementándose de forma monótona. Esto puede aprovecharse para reducir el número de cálculos, fijando un umbral por encima del cual la correspondencia es rechazada y no se sigue calculando D_s .

El problema que tiene esta función [Fua 1993] es que sus valores se incrementan notablemente si el contraste de intensidades dentro de las ventanas de correlación es alto. Para compensar este efecto, es posible normalizar con las varianzas:

$$D_w = \sum_1^n \sum_1^m (I_L(i,j) - I_R(i,j))^2 / \sigma_L \sigma_R \quad (4.12)$$

O bien solamente con la varianza de la ventana de la izquierda, que al ser fija permite una computación mucho más rápida:

$$D_v = \sum_1^n \sum_1^m (I_L(i,j) - I_R(i,j))^2 / \sigma_L^2 \quad (4.13)$$

siendo la varianza $\sigma_k^2 = \sum_1^n \sum_1^m (I_k(i,j) - \mu_k)^2 / (mn)$, $k = L, R$.

En todo caso, al incluir las varianzas, ya no es posible la simplificación del cálculo vista anteriormente, ya que estas se calculan sobre toda la ventana.

Si las características de las dos cámaras no son idénticas, las intensidades luminosas para un mismo punto serán distintas en cada ventana. Para solucionar esta dificultad es posible normalizar con la media de la ventana, en la forma:

$$D_n = \sum_1^n \sum_1^m [(I_L(i,j) - \mu_L) / \sigma_L - (I_R(i,j) - \mu_R) / \sigma_R]^2 \quad (4.14)$$

La alternativa a la suma de diferencias es la correlación, que se define así:

$$C = \sum_1^n \sum_1^m I_L(i,j) \cdot I_R(i,j) / (\sigma_L^2 \sigma_R^2)^{1/2} \quad (4.15)$$

También se puede considerar la correlación normalizada:

$$C_n = \sum_1^n \sum_1^m (I_L(i,j) - \mu_L) \cdot (I_R(i,j) - \mu_R) / (\sigma_L^2 \sigma_R^2)^{1/2} \quad (4.16)$$

Una comparación de los resultados de estas funciones puede encontrarse en los trabajos del INRIA sobre el desarrollo de un algoritmo estéreo para mapas de profundidad densos [Fua 1993], [Faugeras 1993b], que se verá más adelante con detalle.

La conclusión es que los criterios normalizados son igualmente insensibles tanto a variaciones en la intensidad media de las imágenes como a variaciones en los parámetros de una de las cámaras (lógico; en realidad son el mismo criterio, ya que puede demostrarse que $D_n = 2 - 2C_n$.) La correlación no normalizada se degrada algo en este último caso, siendo en cambio poco sensible a los cambios en la intensidad media. Por último, el rendimiento de la función de diferencia no normalizada es malo para el caso de cambio de la intensidad media, y se degrada dramáticamente al variar los parámetros de una cámara.

4.2.4.2 Selección del tamaño de la ventana de correlación

El tamaño de la ventana es uno de los aspectos más críticos de los métodos de correlación. La ventana debe ser lo bastante grande para incluir suficiente variación de intensidad para la correspondencia, pero lo

bastante pequeña para evitar los efectos de la distorsión proyectiva [Barnard 1982]. Si la ventana es demasiado pequeña, y no cubre suficiente variación de intensidad, la estimación de la disparidad será pobre, ya que la relación señal (variación de intensidad) a ruido será baja. Si por el contrario la ventana es demasiado grande y cubre una región en que la profundidad de los puntos de la escena (es decir, la disparidad) varía, la posición del máximo de correlación puede no ser muy exacta

Shirai [Shirai 1987] hace una comparación de la forma de la función de correlación, usando la función de diferencia D_v vista en el apartado anterior (la correspondencia viene dada por el mínimo, en este caso).

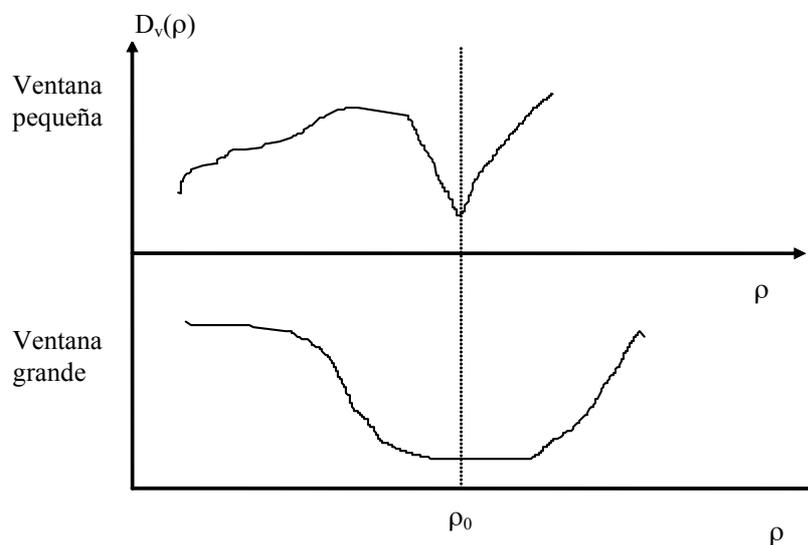


Figura 4.8: Tamaño de ventana y curva de correlación.

Si el tamaño de la ventana es pequeño, la función tiene un valle estrecho, como se muestra en la figura 4.8. Sin embargo, es sensible al ruido, y la diferencia entre el valor en el mínimo y en otros puntos distantes puede ser pequeña, es decir, puede haber múltiples valles (mínimos locales). Si la ventana es grande la función de diferencia cambia suavemente, de modo que el mínimo estará en un valle ancho, y la determinación de su posición ρ_0 no será muy precisa. Otra desventaja de las ventanas grandes es que requieren mayor computación.

Como discute Kanade [Kanade 1994] puede hablarse de dos tipos de ruido en la reconstrucción 3-D, introducidos por ventanas demasiado grandes o demasiado pequeñas. Una ventana demasiado pequeña

introduce un error aleatorio, repartido por toda la imagen. Una ventana demasiado grande hace que la posición de los bordes quede poco definida, con lo que introduce un error sistemático en las zonas de cambio de disparidad, suavizando por tanto los bordes.

El tamaño de ventana ha sido elegido hasta ahora empíricamente para cada aplicación. Esto se debe a que no es fácil su determinación teórica. El tamaño debería ser función de la variación de intensidad y disparidad en la ventana. La variación de intensidad sí es conocida, pero la disparidad es precisamente lo que se busca conocer. Kanade propone un algoritmo para seleccionar adaptativamente un tamaño de ventana en función de las variaciones locales de intensidad y disparidad. Se trata de un proceso iterativo, en que la disparidad se modela de acuerdo a una distribución gaussiana. Partiendo de un tamaño inicial de ventana, éste va modificándose tratando de reducir la incertidumbre en la estimación de la disparidad.

Los resultados del método son muy buenos, tanto en imágenes sintéticas como reales. Un inconveniente es su complejidad, que hace difícil su implementación en tiempo real. De hecho se necesita una estimación inicial de la disparidad, y sólo para este valor inicial ya es necesario llevar a cabo una correlación con ventana fija. Otra posibilidad es estimar como cero la disparidad inicial, con lo que según Kanade para problemas estereo no ambiguos el método sigue funcionando.

Una posibilidad es llevar a cabo diferentes correlaciones con distintos tamaños de ventana y utilizar algún tipo de aproximación jerárquica, de modo que la correlación de menor resolución (ventana grande) sirva de guía para la de más resolución (ventana pequeña), reduciendo su campo de búsqueda de modo que se eviten los errores propios de las ventanas pequeñas. El problema de las aproximaciones jerárquicas se discutirá a continuación.

4.2.4.3 Aproximaciones jerárquicas. Matching multietapa

Las aproximaciones jerárquicas, o matching multietapa, consisten en utilizar imágenes de baja resolución para un matching de baja resolución, y usar los resultados de éste para restringir el área de búsqueda en el matching de alta resolución.

El uso de imágenes de distinta resolución es equivalente a la utilización de ventanas de diferentes tamaños, pero es más eficiente computacionalmente. Las imágenes de menor resolución se consiguen submuestreando las imágenes después de un suavizado gaussiano. Está demostrado [Burt 1982] que esto equivale a llevar a cabo la correlación de la imagen usando varias bandas de frecuencia.

Shirai [Shirai 1987] hace una estimación del ahorro computacional que puede suponer una aproximación jerárquica. En una imagen de alta resolución la búsqueda de correspondencia sólo se efectúa para aquellos píxeles para los que se encontró correspondencia en la etapa previa de baja resolución. Supóngase que cada píxel de la etapa previa corresponde a $m \times m$ píxeles en la imagen actual. El área de búsqueda se restringe a intervalos de longitud m . Supóngase también que la imagen es de tamaño $n \times n$ y que en la etapa previa se encontró correspondencia para todos los píxeles (el peor caso). Por el método convencional (sin multietapa) habría que computar la similitud de dos ventanas en toda la imagen. El área de búsqueda para cada punto es la longitud de la línea, n , por lo que el número total de computaciones de similitud es n^3 . En un matching de dos etapas es necesario un promediado inicial, pero su coste es despreciable frente al resto de cálculos. En la primera etapa el número de cálculos se reduce a n^3/m^3 . En la segunda el número de cálculos será n^2m . Por tanto, la relación respecto al caso convencional será:

$$\frac{\left(\frac{n^3}{m^3} + n^2m\right)}{n^3} = \frac{1}{m^3} + \frac{m}{n} \quad (4.17)$$

que es mucho menor que 1, ya que normalmente $n \gg m$ y $m^3 \gg 1$. Por ejemplo, para un caso de $n = 256$, y $m = 4$, el número de cálculos se reduce por un factor de 32.

Sin embargo, no todo son ventajas en las aproximaciones jerárquicas. Estas asumen que los resultados generados a baja resolución son más fiables, aunque menos precisos, que los de alta. En los informes del INRIA [Fua 1993], [Faugeras 1993b] se cuestiona esta suposición, que puede conducir a errores, especialmente en presencia de oclusiones. Por eso prefieren calcular independientemente cada nivel de resolución, aunque suponga un mayor coste computacional.

4.2.5 Método stretch-correlation para estereovisión

Se describirá en este apartado un método de correlación desarrollada por Lane [Lane 1994], especialmente para problemas estéreo difíciles, que define como aquéllos en que las profundidades de los objetos de la escena son del orden de entre 3 y 10 veces la separación de las cámaras. Esto añade el problema de que la rotación de los objetos entre las vistas es ya considerable, lo que empeora los resultados de los algoritmos de correlación.

El objetivo del algoritmo es relajar la restricción de fronto-parallelismo inherente a las técnicas de correlación, incorporar alguna de las restricciones utilizadas en los algoritmos basados en características, y permitir una implementación eficiente en hardware. Se trata de corregir algunas de las deficiencias comunes a las técnicas de correlación (falta de robustez en los casos de rotación entre vistas, oclusiones y artefactos debidos a la iluminación, resultados pobres en las discontinuidades de profundidad) que pueden causar errores apreciables.

La descripción del algoritmo es la siguiente:

- 1- Rectificación epipolar de las imágenes (las cámaras están calibradas), como paso previo al procesamiento.
- 2- Obtener las diferencias horizontales en la imagen: $I_{out}(i, j) = I_{in}(i, j) - I_{in}(i, j-1)$.
- 3- Suavizado gaussiano, con $\sigma = 1-2$ píxeles.
- 4- Deformar bloques de la imagen derecha, como se verá más abajo (stretch).
- 5- Correlacionar los bloques deformados con un cierto número de bloques a lo largo de la línea epipolar en la imagen izquierda.
- 6- Seleccionar el mejor factor de stretch y la disparidad de bloque, dada por el máximo de la superficie de correlación.
- 7- Calcular las disparidades de los píxeles corrigiendo la distorsión introducida por el stretch, sólo para los píxeles de borde.

Las técnicas de correlación implícitamente asumen superficies fronto-parallelas, con lo que desprecian los efectos de rotación entre las vistas. Por eso en [Lane 1994] han tratado de modelar esta rotación, usando dos técnicas: stretch y shear (expandir y podar). Estas técnicas implican modificaciones menores a la correlación estándar y son implementables en un procesador de imagen optimizado para convolución y correlación. Aquí

se verá únicamente el algoritmo de stretch, que en las pruebas realizadas por los autores ha demostrado ser más robusto frente a errores de posición y falsas correspondencias.

El stretch convierte un bloque de $N \times M$ píxeles en uno de $N \times N$, mediante interpolación en la dirección horizontal. Esto se observa en la figura 4.9, donde se muestran tres bloques con diferente factor de stretch. M varía típicamente entre $2N$ y $N/2$, lo que según los autores es coherente con un límite del gradiente de disparidad de 0.67 a lo largo de la epipolar, que es aplicado habitualmente en algoritmos basados en características [Pollard 1985].

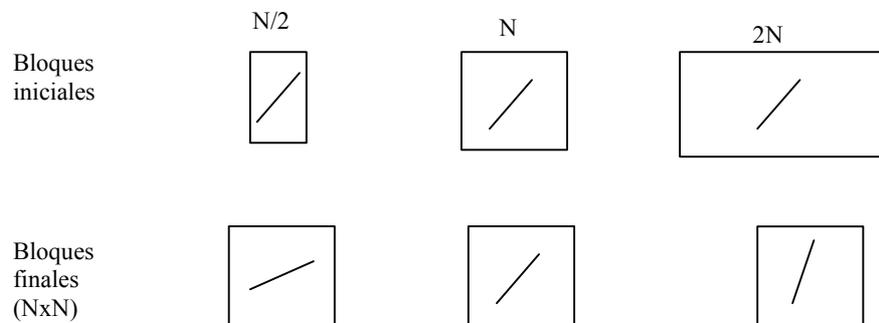


Figura 4.9: Distorsión de bloques mediante stretch.

4.2.5.1 Resultados

En el mismo artículo [Lane 1994] se hace un estudio comparado de los resultados del algoritmo frente a distintas modalidades de correlación, tomando como referencia los resultados de una implementación del PMF (descrito en la sección 4.2.3.3). Se ha utilizado en el análisis una técnica estadística que permite reducir en gran manera el número de imágenes necesario para obtener resultados significativos.

En resumen se concluye que la técnica de deformación considerada devuelve más datos y reduce el error posicional. La reducción del error de posición en comparación con la correlación convencional como resultado de la distorsión puede entenderse intuitivamente. Los beneficios del preprocesamiento con diferencias horizontales y suavizado gaussiano son más sutiles, pero se ven con claridad al comparar con otras formas de preprocesamiento. En cuanto a la capacidad de desambiguación (probabilidad de obtener correspondencias incorrectas) las diferencias no

son importantes, y parece que la superioridad de la distorsión es muy dependiente de la naturaleza de la imagen.

Se mencionan por último tiempos de implementación: en una SUN Sparc II el algoritmo tarda 10 minutos para imágenes de 512x512, frente a los dos minutos de una implementación eficiente del PMF). Se estima que utilizando procesadores específicos para la rectificación y la correlación, capaces respectivamente de 2 y 4 GIPS, el algoritmo generaría resultados a 6 Hz.

4.2.6 Método del INRIA para la obtención de mapas densos de profundidad

En distintos informes [*Fua 1993*], [*Faugeras 1993b*], [*Fua 1991*] de investigadores del INRIA francés (Institut National de Recherche en Informatique et en Automatique) se describe un método de correlación cuyo objetivo es la obtención de mapas de profundidades densos, pero sin embargo fiables (es decir, sin falsas correspondencias), a velocidades que permitan su implementación en tiempo real en aplicaciones como navegación de robots móviles.

En realidad se trata de un algoritmo clásico de correlación, sin grandes sofisticaciones, al que se le da robustez y fiabilidad mediante ciertas mejoras, especialmente en el procedimiento de validación de las correspondencias. A continuación se describirán con cierto detalle los distintos pasos del algoritmo.

4.2.6.1 Preprocesamiento. Calibración y rectificación

La calibración se lleva a cabo de dos maneras: bien por el método de la matriz de perspectiva, o bien con una calibración débil. Parece que los resultados son similares, si bien en el último caso no se tiene información métrica, obviamente.

A continuación se lleva a cabo la rectificación epipolar, re proyectando las imágenes sobre un plano paralelo a la línea entre centros

ópticos, utilizando lógicamente el método de Ayache [Ayache 1988], también del INRIA.

4.2.6.2 Criterios de correlación

Se comparan cuatro criterios de correlación, la correlación propiamente dicha y las diferencias cuadradas, normalizadas y sin normalizar. Como se vio en el apartado 4.2.4.1 se concluye que el mejor comportamiento es el de los criterios normalizados. La correlación sin normalizar consigue resultados similares salvo cuando la diferencia en la distribución de niveles de gris en las dos imágenes es importante.

4.2.6.3 Algoritmo de correlación. Validación de las correspondencias

La filosofía del algoritmo, que lo diferencia de otros similares, es que la correlación no se lleva a cabo sólo en puntos de interés, sino en toda la imagen, intentando lograr un mapa denso de profundidades. Esto ocasiona que la incertidumbre acerca de la validez de las correspondencias encontradas sea mayor, por lo que hay que afinar en la validación de las mismas.

Tras comentar el problema del tamaño de ventana (ver 4.2.4.2) se propone una definición de la validez de una medida de disparidad en la que las dos imágenes juegan un papel simétrico, y que permite el uso de ventanas pequeñas de una manera fiable. La correlación se lleva a cabo en ambos sentidos, y se consideran válidas solamente aquellas correspondencias para las que la correlación inversa confirma los resultados iniciales. Es decir, si buscando el punto correspondiente de un píxel P_1 en la primera imagen se obtiene P_2 en la segunda, al buscar la correspondencia para P_2 mediante la correlación inversa deberá obtenerse P_1 , o la correspondencia no será considerada válida.

De este modo se consigue que el número de correspondencias obtenidas sea también un indicador de la calidad del matching, ya que es muy difícil que se produzca una falsa correspondencia, salvo en presencia de patrones repetitivos. Por tanto puede considerarse que un mapa de disparidades denso es una garantía de que las correspondencias son correctas. Se comprueba [Fua 1993] que en los casos más difíciles (oclusiones, poca textura, gran variación de la disparidad en la ventana de

correlación) la densidad de correspondencias decrece pero la razón de correspondencias correctas a falsas permanece alta hasta que el mapa de disparidades es muy disperso. Pueden conseguirse resultados aún más fiables si se eliminan las correspondencias consideradas válidas pero aisladas (mediante erosiones/dilataciones).

Todavía se utilizan otros dos criterios para contrastar la fiabilidad de las correspondencias. Se define un coeficiente de confianza proporcional a la diferencia de altura entre el máximo de la función de correlación y el siguiente pico. Además se mide la anchura del pico máximo, ya que cuanto más estrecho sea más precisa será la localización del punto correspondiente. Para eso se representa el pico por una gaussiana con desviación estándar σ , y este dato se usa al calcular la precisión subpíxel. Para conseguir una precisión subpíxel se ajusta a la función de correlación en la proximidad del máximo una parábola, y se toma una precisión subpíxel proporcional a σ .

4.2.6.4 Aproximación jerárquica

Para incrementar la densidad de los mapas de disparidad se lleva a cabo el matching con diferentes niveles de resolución, para un tamaño fijo de ventana. Como se vio antes, esto equivale a usar ventanas de distintos tamaños. Las diferentes resoluciones se obtienen submuestreando imágenes a las que previamente se aplica un suavizado gaussiano. Para cada píxel se elige el valor de disparidad correspondiente al máximo nivel de resolución para el que se encontró un resultado válido. La gran fiabilidad de los resultados permite no tener que introducir factores de peso para sobrevalorar los resultados de alta resolución.

La computación se lleva a cabo de modo independiente para cada resolución, sin implementar una aproximación jerárquica en la que los resultados de baja resolución guían a los de alta. Esto se discutió ya en 4.2.4.3.

4.2.6.5 Implementación del algoritmo

El gran inconveniente de este algoritmo es su alta exigencia computacional. En el informe interno del INRIA [Faugeras 1993b] se

detallan algunas simplificaciones para obtener una implementación más eficiente, con las que se consigue hacer el tiempo de procesamiento casi independiente del tamaño de ventana.

Se mencionan distintos tiempos de implementación, en estaciones de trabajo y en hardware específico. Los mejores resultados se obtienen para una placa PeRLe-1, desarrollada en el mismo INRIA a base de chips LCA (Logic Cell Array). Se consigue un tiempo de 0.28 s. para las dos correlaciones.

4.2.7 Algoritmo SSSD para múltiples líneas base

La utilización de más de dos cámaras para eliminar las ambigüedades del matching aparece con cierta frecuencia en la literatura [Yachida 1986], [Milenkovic 1985], [Ayache 1987b]. En general estas técnicas encuentran candidatos a la correspondencia en cada par de imágenes y luego buscan combinaciones correctas de correspondencias entre ellas, utilizando criterios de coherencia geométrica. El inconveniente de este procedimiento es que las decisiones intermedias de correspondencia (para cada par de imágenes) son inherentemente ambiguas, ya que sólo usan dos imágenes, por lo que los criterios de coherencia deben ser sofisticados. Por contra, el algoritmo SSSD [Kanade 1992] no toma decisiones sobre las correspondencias en cada par de imágenes, sino que acumula las medidas de correspondencia de todos los pares en una sola función de evaluación, de la cual se obtienen las correspondencias. Es decir, se integran evidencias para una decisión final, en lugar de filtrar decisiones intermedias.

4.2.7.1 Descripción del método

Una de las funciones más usadas para el matching por correlación, por su sencillez y eficacia, es la suma de diferencias cuadráticas (SSD). La representación normal es en función de la disparidad, de modo que el mínimo de la curva corresponde a la disparidad resultante. Como se vio anteriormente, para el caso de cámaras paralelas la relación entre disparidad y distancia a las cámaras viene dada por $d = \frac{BF}{z}$, siendo B la línea base (distancia entre cámaras) y F la distancia focal (se supone la misma para ambas cámaras).

Suponemos que se dispone de un conjunto de cámaras paralelas, y pueden tomarse pares de imágenes con distintas líneas base. La función SSD puede representarse en función de la distancia inversa, $1/z$, que es una propiedad del punto espacial, y por tanto es constante para cualquier par de imágenes, sea cual sea la distancia entre cámaras. Por tanto es posible obtener la SSD para cada par estéreo, y sumarlas todas, obteniendo una nueva función, la suma de sumas de diferencias cuadráticas (SSSD), función también de la distancia inversa. Se demuestra que la función SSSD tiene un mínimo único y claro en la posición correcta, incluso si las diferentes SSD's no lo tienen.

Las propiedades más importantes de esta nueva función son las siguientes:

- Eliminación de ambigüedades. En el caso de que la función de intensidad luminosa tenga el mismo patrón en torno a dos píxeles, la función SSD no es capaz de discriminar entre esos dos píxeles en el matching, dando dos mínimos iguales. Esto es así sea cual sea la línea base. En cambio, la SSD en función de la distancia inversa sigue presentando dos mínimos, pero mientras el falso mínimo cambia su posición en función de la línea base, el correcto permanece constante. Esta es la propiedad que aprovecha la SSSD, que suma SSD's correspondientes a diferentes líneas base, y presenta el mínimo en la posición correcta. Un caso extremo de ambigüedad se produce cuando la función intensidad es periódica (caso de patrones repetidos en la imagen). Entonces la SSD presenta múltiples mínimos a intervalos dependientes de la línea base. Se demuestra que, si bien la SSSD resultante de sumar SSD's con distintas líneas base también es periódica, su período es mayor que el de cualquiera de las SSD's que la integran. Por tanto, escogiendo las líneas base adecuadamente, puede conseguirse que exista sólo un mínimo dentro del área de búsqueda.
 - Precisión. Se demuestra que el uso de la función SSSD también permite aumentar la precisión en la determinación de la verdadera distancia inversa. En el artículo de Kanade se calcula la varianza de la estimación de la distancia inversa para la SSD, y se demuestra que es menor (estimación más precisa) cuanto mayor es la línea base, y es todavía menor si se usa la SSSD, mejor cuantas más líneas base se consideren. Esto se confirma experimentalmente viendo que las curvas de las SSSD's son más agudas cuantas más líneas base se consideren.
-

4.2.7.2 Una aplicación práctica del algoritmo SSSD

En un artículo de Ross [Ross 1993] se describe la implementación de un sistema estéreo en tiempo real basado en el algoritmo SSSD de Kanade. Se trata de un sistema para un robot móvil, con una doble finalidad: detección de obstáculos entre 3 y 100 m, dando resultados cada 0.1 s, y dar información del terreno, entre 3 y 25 m, generando resultados cada 5 s. Los requisitos de computación son considerables en ambos casos, y exigen una técnica estéreo rápida y simple.

Este artículo es un tanto atípico, ya que teóricamente no aporta novedades, pero se discuten aspectos prácticos de implementación que normalmente no suelen ser publicados, lo que resulta bastante interesante.

Se usa un sistema estéreo con tres cámaras. Se discuten las ventajas sobre un sistema binocular. La calidad de las imágenes de distancia es considerablemente mayor, sobre todo en el caso de objetos cercanos, dado que en este sistema no es posible usar una línea base menor de 1 m. Otras ventajas prácticas son la redundancia, en caso de fallo de una cámara, y el aprovechamiento de los tres canales RGB del frame-grabber para las tres cámaras.

El algoritmo SSSD se usa por su sencillez y buenos resultados, además de ser fácilmente paralelizable tanto para máquinas MIMD como SIMD. El uso de cámaras paralelas simplifica el matching, ya que la búsqueda es sólo en una línea. El tamaño de la ventana de correlación es un parámetro importante. Se usan típicamente ventanas de 10 filas y 20 columnas. Se ha comprobado que es preferible ensanchar la ventana, sacrificando resolución horizontal, que aumentar la altura a costa de resolución vertical.

En esta implementación se prescinde de algunos detalles que mejoraban los resultados a cambio de un mayor tiempo de cálculo. Así no se calculan las disparidades con precisión subpíxel ni se usan ventanas de tamaños variables.

Para incrementar la velocidad se utilizan otras técnicas. Debido a la amplitud de la línea base (1 m) las resoluciones a distancias largas son aceptables, mientras que a distancias cortas son mucho mayores que las necesarias. Es posible saltarse disparidades en las distancias cortas,

mientras que en las largas sí se chequea el rango completo. Esto tiene el efecto de igualar la resolución en todo el rango de distancias.

Otro truco usado para mejorar el rendimiento es invertir el orden de los cálculos. En lugar de hallar las SSD's entre dos conjuntos de ventanas y luego sumar sus valores se computan las diferencias entre las imágenes completas y se suman, produciendo una nueva imagen. Para calcular la SSSD de un píxel se suma entonces la ventana en torno al píxel en cuestión. Esto puede hacerse rápidamente, manteniendo en memoria sumas de columnas.

Por último, para reducir más los cálculos se utilizan imágenes de menor tamaño para el caso de detección de obstáculos, donde los pequeños detalles de la escena no son importantes.

Así pues, el algoritmo es simplemente un SSSD simplificado para reducir el tiempo de cálculo. Tal vez más interesantes son algunas consideraciones prácticas que se mencionan a continuación:

1. No se realiza la rectificación epipolar de las imágenes. Se consigue que las líneas horizontales de las imágenes sean coincidentes mediante un cuidadoso ajuste mecánico previo. No es suficiente con el ajuste externo de las cámaras de modo que sean perfectamente paralelas, ya que se comprueba que las CCD's y lentes típicas muestran diferencias considerables en su alineamiento con el cuerpo de la cámara. Se usa una montura ajustable en dos de las cámaras, que se alinean respecto a la tercera.
 2. La elección de la línea base es crítica. Se busca un compromiso entre corta (resolución en profundidad pobre, pero facilita el matching y aumenta el campo común) y larga (resolución mayor en profundidad, pero disminuye el campo de visión efectivo y complica el matching). La elección de una línea base de 1 m. es un término medio. La resolución a 15 m es sólo aceptable, pero a distancias menores es muy buena.
 3. En el trabajo en exteriores es necesario utilizar autoiris.
 4. Las cámaras usadas son Sony XC-75. Es necesario modificar el hardware de las cámaras y lentes para conseguir una mayor rigidez, de modo que no se permite movimiento entre cámara y lentes. Esto tiene además la ventaja de fijar el foco.
-

5. Se utiliza un frame-grabber para color, con 24 bits. Las tres cámaras monocromas se sincronizan y se conectan a los tres canales RGB. Se utiliza el ajuste de ganancia del frame-grabber para compensar las ganancias de las tres cámaras, ya que se comprueba que si los niveles de gris de las tres imágenes no son lo bastante cercanos los resultados empeoran grandemente. Esto es lógico, ya que se usa una función de diferencias no normalizada (ver 4.2.4.1).

Este sistema se ha implementado en diferentes máquinas, tanto estaciones de trabajo convencionales como máquinas paralelas. Se ha comprobado que el algoritmo se adapta bien a estas últimas, y se consiguen mejoras importantes en los tiempos. Algunos tiempos mencionados son 2.46 s para imágenes de 256x240 y rango de disparidades de 16, en una Sun Sparc II, 8.03 s en la misma máquina con imágenes de 512x240 y rango de disparidades de 27, 0.35 s para imágenes de 256x240 y rango de disparidades 16 en una iWarp de 16 células, y 0.15 s para el mismo caso con una iWarp de 64 células.

En cuanto al funcionamiento del sistema en distintos entornos, los resultados han sido buenos en casi todos, salvo en regiones con pocas características y poca textura. Basta con una pequeña presencia de textura (como pistas en la nieve) para que el funcionamiento sea bueno. Se asegura que este sistema demuestra que la estereovisión es una alternativa competitiva con otros sistemas de percepción en Robótica Móvil.

4.2.8 Otros métodos de matching

Ocasionalmente han aparecido en la literatura métodos no basados ni en correlación ni en extracción de características, sino que utilizan alguna otra propiedad física.

4.2.8.1 Estéreo usando razones fotométricas

En un artículo de Wolff y Angelopoulou [Wolff 1993] se describe este método, que viene a ser una combinación de la estereovisión con otra técnica de reconstrucción 3-D, el estéreo fotométrico (*photometric stereo*). En el estéreo fotométrico se toman imágenes de una escena con una sola

cámara y diferentes iluminaciones. Si estas iluminaciones son conocidas, pueden caracterizarse las superficies presentes en la escena, suponiendo un comportamiento *lambertiano* (la luz reflejada es la misma en todas direcciones) de las mismas.

Varios autores han considerado la posible combinación de estereovisión y estéreo fotométrico, para solucionar aquellos casos en que los datos obtenidos de la estereovisión son escasos (superficies suaves, sin apenas características). Grimson [Grimson 1984] utilizaba la orientación superficial obtenida con el photometric stereo para interpolar con más exactitud los datos dispersos obtenidos con la estereovisión. Smith [Smith 1986] utilizaba valores fotométricos para el matching, mediante una formulación matemática llamada Ecuación Integral Estéreo, obteniendo mapas de profundidades densos.

Existe un cierto número de aspectos prácticos que hacen difícil la utilización de valores fotométricos para la correspondencia estéreo. El principal es que la estereovisión requiere dos cámaras, que no recogen igual las intensidades luminosas, incluso en el caso ideal de que las superficies fueran perfectamente lambertianas. Esto se debe a una serie de causas, tanto ópticas (pequeñas diferencias en la distancia focal y en el diámetro efectivo de la lente), como electrónicas (variación de la ganancia de una cámara a otra y en función de la temperatura, no linealidad). La metodología descrita por Wolff y Angelopoulou [Wolff 1993] permite superar estas dificultades.

Se trata de utilizar como propiedad para el matching las razones de dos imágenes de la misma escena obtenidas bajo distintas condiciones de iluminación. Esta razón fotométrica se calcula para las dos imágenes, izquierda y derecha. En primer lugar se demuestra que la razón fotométrica es un invariante a los parámetros de la cámara y al ángulo de vista, excepto en el caso de reflexión especular. Esto garantiza que si se obtienen con cada cámara dos imágenes I_1 e I_2 bajo dos iluminaciones distintas, la razón I_1/I_2 será la misma para un mismo punto con respecto a las dos cámaras. Por tanto este valor puede ser utilizado para la correspondencia, a lo largo de la línea epipolar. Un aspecto muy importante es que las condiciones de iluminación, al contrario que en otros métodos, no tienen que ser conocidas. El uso de más de dos iluminaciones diferentes permitiría eliminar ambigüedades, proporcionando información redundante.

Se han obtenido muy buenos resultados con el método, con objetos sin textura ni características, cilindros y esferas. Se demuestra que el mejor funcionamiento del método se produce cuando las direcciones incidentes

de las fuentes de luz intersecan la línea base entre las cámaras, o al menos la separación entre las fuentes de luz es paralela a la misma línea base.

Se obtienen mapas de profundidad densos, reduciendo el error a un 0.45%. La razón fotométrica de un píxel en la imagen izquierda se corresponde con precisión subpíxel en la derecha, interpolando razones fotométricas entre píxel.

La gran ventaja del método es que la correspondencia es casi inmediata, sin correlaciones ni extracción de características. Para una implementación en tiempo real habría que sincronizar la adquisición de imágenes con las diferentes condiciones de iluminación.

4.3 Utilidad y uso de la información de profundidad

Históricamente, el proceso de obtención de información tridimensional mediante estereovisión se ha incorporado a multitud de sistemas, desde brazos robóticos para la manipulación y ensamblaje en cadenas de producción, hasta sistemas autónomos de navegación guiados por la visión.

En nuestro caso se ha incorporado en modelos de procesamiento visual bioinspirados que forman parte de sistemas de rehabilitación visual para invidentes. En este ámbito, caracterizado por el limitado ancho de banda del canal (o los canales) disponible por el que se transmite la información, la utilidad de una estimación de la profundidad es muy relevante para poder hacer una mejor selección de las características más realizadas de la escena.

No obstante, en nuestro campo de actuación existe cierta relajación sobre los requisitos de precisión en las medidas que debe realizar el sistema, ya que no se pretende realizar una fiel reconstrucción 3D de la escena que se presenta ante la cámara, sino obtener una estimación fiable de la distancia entre el sujeto que porte el sistema de rehabilitación y los distintos objetos presentes en la escena, con la finalidad de combinar este

dato con el resto de resultados obtenidos en la etapa de filtrado espacio-temporal de las imágenes.

En primer lugar se ha realizado una integración del módulo de cómputo de disparidades dentro del entorno Retiner detallado en el capítulo anterior. La calibración de las cámaras utilizadas se ha hecho mediante la toolbox desarrollada para Matlab™ por Jean-Yves Bouguet [Bouguet] que permite obtener todos los parámetros de calibración a partir de múltiples tomas de un patrón de calibración consistente en una malla tipo tablero de ajedrez. La herramienta intenta extraer las esquinas de todos los cuadrados que forman el patrón, y mediante feedback del usuario se indican las diferencias entre las posiciones predichas de las esquinas y las reales (véase figura 4.10).

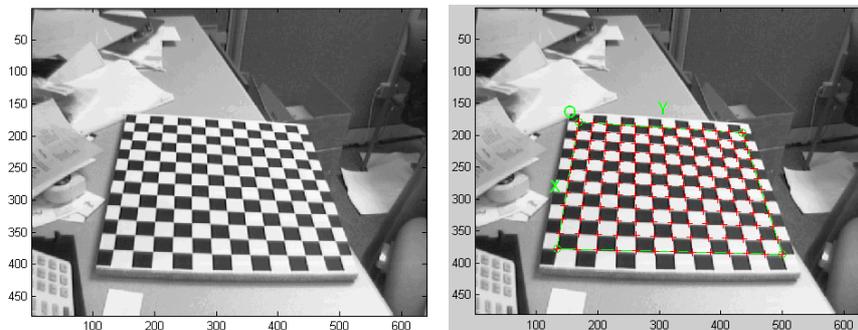


Figura 4.10: Izquierda: patrón utilizado para la calibración. Derecha: detección de características (esquinas) utilizadas para el proceso de calibración.

El método concreto seguido consta de los siguientes pasos:

1. Cargar todas las imágenes del patrón de calibración tomadas.
2. Para cada imagen, seleccionar de forma interactiva las cuatro esquinas más exteriores del patrón.
3. La herramienta realiza una interpolación lineal de las esquinas intermedias y nos muestra el resultado de este proceso para detectar diferencias notables.
4. Mediante la decisión de la inclusión del cálculo de los términos de distorsión radial, podemos minimizar los errores de cálculo de esquinas detectados en el paso 3.
5. Con todos los puntos seleccionados de las imágenes, se construye un sistema de ecuaciones sobredeterminado, que por métodos de minimización cuadrática se resuelve numéricamente, dando como resultado los parámetros de la cámara.

6. Como paso adicional se puede realizar una rectificación de las imágenes según los coeficientes de distorsión radial calculados.

De esta forma la herramienta puede detectar todos los parámetros, incluyendo la distorsión radial. El método de calibración se corresponde con una variación del método de Tsai (véase sección 4.1.1.3), que incluye el cálculo de cinco términos para la distorsión introducida por las lentes de la cámara y , como caso particular, permite eliminar estas componentes cuando no sea necesario su cálculo.

Una vez se han calibrado las cámaras de forma independiente, y utilizando los resultados de calibración individuales de las dos cámaras se procede a calibrar el sistema de forma automática mediante la herramienta, obteniendo las dos matrices de transformación, R y T , que determinan la geometría epipolar del sistema concreto (ver sección 4.1.2.1). Una vez ha finalizado el proceso tenemos todos los parámetros necesarios para realizar la triangulación de puntos correspondientes en los pares estéreo. Esta toolbox incluye funciones para realizar esta tarea, así como para rectificar las imágenes capturadas (según la geometría epipolar del sistema) y poder buscar los puntos correspondientes en filas con igual índice dentro de las dos imágenes.

De esta forma, el procesamiento que tiene que realizar Retiner es capturar imágenes, hacer el filtrado bioinspirado deseado, detectar puntos correspondientes (matching) en ambas imágenes, triangular la posición de estos puntos para detectar la coordenada en el eje Z (profundidad), y combinar toda la información recopilada, según indique el usuario. A partir de este punto, Retiner seguiría su flujo normal de ejecución que se describió en capítulos anteriores.

En cuanto al método de detección de correspondencias o matching, se ha utilizado uno basado en área (descritos en la sección 4.2.2), con la finalidad de obtener mapas densos de profundidad. Como hemos calculado la geometría epipolar del sistema, se realiza un paso previo de rectificación epipolar para reducir la complejidad del proceso de búsqueda.

Una aportación es que no nos hemos limitado a usar los niveles de grises en el cálculo de la correlación cruzada en las ventanas de píxeles (véase sección 4.2.4), sino que hemos hecho uso de la información de color de las imágenes de entrada, calculando el valor de dicha correlación para cada plano de color y combinando los tres valores obtenidos. La función de similitud que se ha decidido utilizar es la suma absoluta de diferencias o SAD, por su facilidad de implementación tanto en software como en

hardware. La ecuación (4.11) resume el cálculo que realizamos para cada plano de color.

El tamaño de ventana usado para el cálculo de la función de similitud es *fijo*, por su facilidad de implementación y porque permite controlar el tiempo de cálculo evitando sobrecargas introducidas por otros métodos iterativos, y *configurable*, permitiéndonos calcularlo de forma empírica para cada aplicación específica. Las pruebas realizadas con imágenes de entrada de tamaño 160x120 han obtenido los mejores resultados utilizando un tamaño de ventana de 11x11 y un rango de disparidades de ± 15 .

La sobrecarga de incorporar este procesamiento a la carga computacional de Retiner, ha provocado su ralentización de una forma muy notable, disminuyendo su rendimiento a niveles de varios segundos por frame en un Pentium 4 a 2.8 GHz. Por tanto, el funcionamiento 'en vivo' de Retiner, capturando imágenes de las cámaras, procesando y mostrando los resultados, se hace inviable actualmente. Una forma para obtener buenos resultados sobre imágenes reales es el procesamiento off-line de secuencias capturadas previamente y almacenadas en un archivo de vídeo, aprovechándonos también de la particularidad de este tipo de procesamiento que ofrece Retiner para establecer la resolución temporal de los cálculos (ver sección 3.3.5). En este caso, el archivo debe tener almacenado en cada frame un par estéreo.

Otra posibilidad para incorporar esta información es hacerlo en una etapa posterior del procesamiento, después de aplicar los campos receptivos a la imagen, es decir, en la matriz de actividad. La diferencia es que tenemos que aplicar todo el filtrado espacio-temporal a las dos imágenes que forman el par estéreo, aplicarle a ambas la reducción de resolución determinada por los campos receptivos que se hayan definido, y ejecutar un proceso de matching a este nivel. La figura 4.11 ilustra las diferencias entre ambos esquemas de procesamiento.

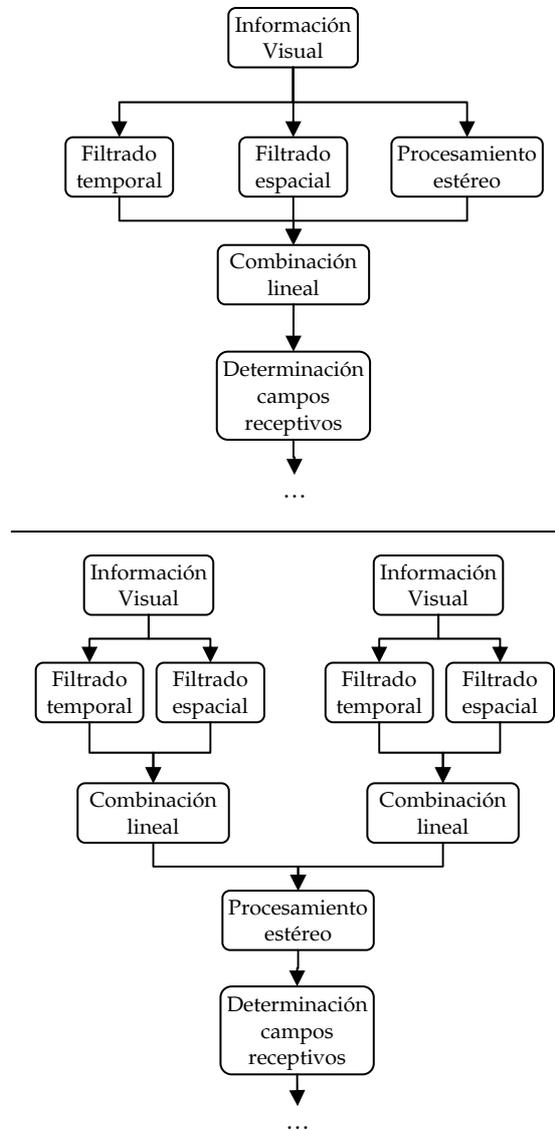


Figura 4.11: Arriba: extracto de la arquitectura de procesamiento normal que implementa Retiner (véase sección 3.3.7). Abajo: arquitectura de procesamiento modificada que permite un cómputo alternativo de la información de profundidad con peor resolución, pero con mejor comportamiento temporal.

Esta alternativa se caracteriza por presentar un mayor paralelismo potencial explotable en implementaciones hardware, de las que

hablaremos más adelante en la sección 6.3, y una complejidad mucho menor en los cálculos introducidos por la estereovisión, debido a la reducción de la resolución de las imágenes sobre las que se aplican los algoritmos. Sin embargo, si se realiza de esta forma, no podremos aplicar los resultados obtenidos de la calibración, ni realizar una triangulación directa de los puntos correspondientes (que se corresponden en este caso con medias ponderadas de amplias zonas de la imagen original).

Si recordamos el comienzo de la sección 4.1.2, se introduce la geometría de un sistema estéreo simple, en el que los planos imagen de las cámaras que lo forman son coplanares. Más adelante se deduce (en la ecuación 4.6) que la profundidad de un punto proyectado en ambas cámaras viene dada por una fórmula que depende de la distancia focal y de la distancia entre los centros de las cámaras (línea base), que son parámetros fijos una vez se ha definido el sistema estéreo; pero además, depende de forma inversamente proporcional de la disparidad, definida como $d = x_r - x_l$, la diferencia en las coordenadas imagen de la proyección de ese punto en ambas cámaras.

Por tanto, basta con realizar un proceso de matching, y calcular la disparidad entre los puntos correspondientes detectados, para tener una estimación de la profundidad del objeto que se proyecta en dichos puntos. Además, la geometría epipolar nos revela que en un sistema estéreo con estas características las líneas epipolares sobre las que se deben buscar las correspondencias son horizontales, facilitando la tarea.

El hecho de trabajar con imágenes de una resolución reducida nos facilita la implementación de este procedimiento de cómputo de las disparidades, ya que la existencia de pequeñas diferencias en el ajuste en la altura de las cámaras no afecta de una forma muy relevante al resultado de los cálculos realizados.

4.4 Conclusiones

En este capítulo se ha realizado una descripción y estudio de distintos aspectos que intervienen en los sistemas de visión estéreo, para poder extraer información sobre la profundidad de los objetos presentes en la escena captada por las cámaras.

En primer lugar nos hemos centrado en la descripción de las características de las cámaras y se han detallado los métodos para su calibración, o extracción de parámetros que nos permiten relacionar la localización de un elemento en el mundo tridimensional, con la posición de su proyección en el plano imagen o retiniano de la cámara.

Una vez que se han caracterizado las cámaras individualmente, se han descrito los aspectos que intervienen en el sistema formado por dos cámaras, y se ha introducido el concepto de geometría epipolar, que nos permite limitar el problema de la búsqueda de puntos correspondientes en los pares de imágenes, a una búsqueda 1D (en una línea).

Seguidamente se ha abordado el problema de la correspondencia, distinguiendo entre métodos basados en características y métodos basados en área, diferenciándose principalmente en el preprocesamiento que exigen y la densidad de los mapas de profundidades que generan. Esta sección se ha completado con un estudio de varios algoritmos clásicos, y algunas mejoras más recientes, incluyendo, en los casos que estaban disponibles, resultados acerca del rendimiento de los mismos.

Finalmente, se completa este capítulo con una discusión acerca de la utilidad de la información aportada por un sistema estéreo, y la forma en que se incorpora a nuestros sistemas de procesamiento orientados a la rehabilitación visual. Se describe el uso de una toolbox para Matlab™ que permite automatizar el proceso de calibrado de las cámaras y el sistema estéreo. Con todo esto se ha construido un módulo estéreo acoplado en Retiner que permite obtener información sobre la profundidad e incluirla de forma ponderada en el modelo de procesamiento visual que hayamos modelado en esta herramienta.

En el modelo de procesamiento implementado en Retiner, el módulo estéreo contribuye realizando las formas más próximas del campo visual. Además, como se verá en los siguientes capítulos 5 y 6, el módulo estéreo es esencial para la construcción del sistema transductor de imágenes a sonidos que se propone.

Sistema de generación de sonidos 3D

En este capítulo se introduce un sistema de generación de sonidos localizados espacialmente, que nos permite evocar en el usuario la percepción de una fuente sonora localizada en cualquier punto del espacio tridimensional. Como veremos con más detalle, este tipo de procesamiento de las señales acústicas se convierte en una potente herramienta para poder comunicar a invidentes cierta información sobre su entorno. En la primera sección, *5.1*, expondremos los fundamentos teóricos y características de este tipo de señales sonoras. A continuación, en el apartado *5.2* hablaremos de la forma en que se realiza el cálculo de los filtros que nos permiten modelar todas las características necesarias para la localización espacial de sonidos, para terminar con la sección *5.3* en la que proponemos la integración de este modelo de procesamiento con el resto del sistema para la rehabilitación visual, describiendo la implementación software del mismo.

5.1 Localización espacial de sonidos

La localización espacial de sonidos consiste en la capacidad de procesar cualquier sonido de forma que el sujeto que lo escuche perciba la sensación de que la fuente de la que procede está situada en una determinada posición espacial, distinguiendo tanto su elevación con respecto al plano horizontal, como el desplazamiento con respecto al plano vertical que pasa por la zona frontal, como la distancia a la que se encuentra el objeto que genera el sonido.

El primer paso consiste en estudiar las bases biológicas en las que se fundamentan los sistemas que realizan este tipo de procesamiento. Afortunadamente, la investigación sobre los aspectos psicofísicos del sistema auditivo ha permitido la comprensión del problema de la localización espacial de los sonidos. Bloom y Kendall, [Bloom 1977a], [Bloom 1977b], [Kendall 1984], consiguieron determinar de una forma cuantitativa los *indicios* o *pistas* (*cues*) que usa el ser humano para localizar los sonidos. Haciendo uso de estos indicios lograron sintetizar sonidos que parecían venir de determinadas posiciones.

Los indicios de los que hablamos son una consecuencia de la interacción de los sonidos con el entorno físico: su propagación a través del aire y alrededor de la cabeza. Una vez llega el sonido a los pabellones auditivos, es reflejado por la *pinna* y el *tragus* del oído externo, para pasar al canal auditivo llegando finalmente al tímpano y el oído interno. Las modificaciones que sufre el sonido en su viaje hasta la parte interna de los oídos izquierdo y derecho son las que nos ofrecen los indicios para localizar los sonidos.

Los principales indicios para determinar la localización en el plano horizontal son dos, y como podemos ver en la figura 5.1, son el resultado de la interacción *binaural* (con los dos oídos). El primer indicio se debe a la interacción del sonido con la cabeza, y hace que las frecuencias más altas se vean atenuadas al viajar alrededor de la cabeza del observador. A esta atenuación de las altas frecuencias se le conoce en inglés como *acoustic headshadow*.

El segundo indicio para la localización horizontal es la diferencia en el tiempo de llegada del sonido al oído izquierdo y al derecho, debido a la diferencia en la distancia recorrida por el sonido en cada uno de los caminos. Si el sonido fuera emitido justo en el frente del observador, la distancia es la misma hasta los dos oídos, y la diferencia temporal sería

nula. En el otro extremo, el retardo máximo se produce cuando la fuente de sonido se encuentra a uno de los lados del oyente. El retardo que se puede observar depende de la distancia entre los oídos, y valores típicos en el ser humano se encuentran entre 350 y 650 μ s.

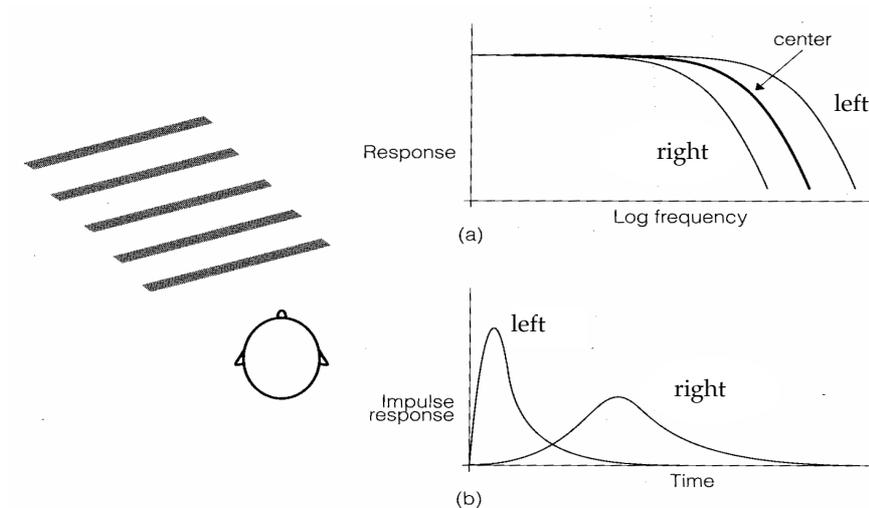


Figura 5.1: Modelo de cabeza a la que llega un sonido con una orientación de 45° a la izquierda. La figura (a) nos muestra la atenuación en frecuencia que sufre la señal percibida en el oído del lado contrario del que proviene el sonido. En (b) tenemos el efecto en el tiempo, como era de esperar, la señal llega retrasada en el tiempo al oído del lado contrario [Nielsen 1989].

Para la localización en la dirección vertical tendremos que fijarnos en los efectos introducidos en la señal sonora por la pinna y el tragus. La figura 5.2(a) nos muestra dos posibles caminos por los que puede entrar el sonido al canal auditivo. Uno de los caminos es lo más directo posible, mientras que el otro es más largo, rebotando el sonido entrante en la pinna, después en el tragus, entrando finalmente por el canal auditivo. Las señales que recorren ambos caminos se combinan en el canal. El efecto producido por esta combinación es una forma especial, un valle, en la representación de la respuesta impulso; la causa de esta depresión se debe a la interferencia destructiva producida entre las diferentes señales.

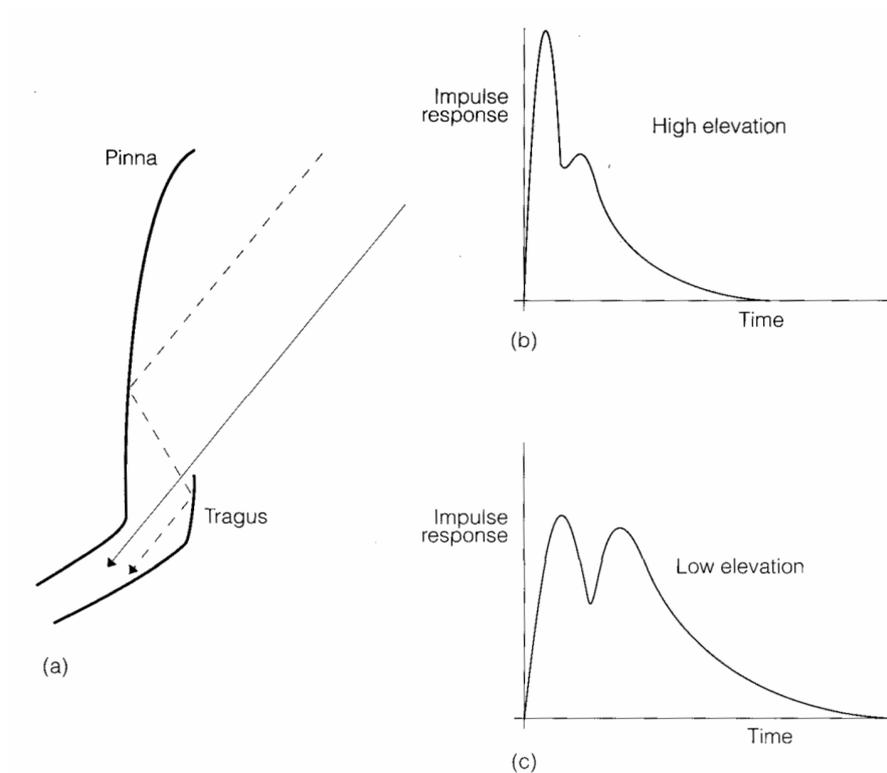


Figura 5.2: (a) Modelo de la pinna y el tragus del oído externo en el que podemos ver la forma en que se producen los rebotes de las señales sonoras entrantes. La respuesta impulsiva para sonidos con elevación alta (b) y baja (c) [Nielsen 1989].

El retardo temporal entre los dos caminos que sigue la señal es una función de la elevación. Debido a que la forma del oído externo es única para cada individuo, los valores para este retardo varían. Sin embargo, para todo el mundo la magnitud del retardo es una función monótona de la elevación, correspondiéndose los pequeños retardos con las elevaciones más altas y retardos más grandes conforme la elevación es inferior.

En la figura 5.2(b) y (c) podemos ver la respuesta impulsiva para distintas elevaciones. Los valores típicos en el hombre para los retardos producidos por los rebotes en pinna-tragus, se encuentran entre 35 y 80 μ s.

En resumen, las pistas o indicios que nos permiten determinar la posición de una fuente sonora son los siguientes:

- Retardo temporal interaural (horizontal).
- Atenuación en frecuencia producida por la forma de la cabeza o acoustic headshadow (horizontal).
- Características pinna-tragus (vertical).

5.2 Cálculo de los indicios de localización

Cuando se ha tratado de llevar a la práctica el procesamiento que modela las características descritas en el apartado anterior, la forma más común en que se ha realizado es mediante la estimación de las HRTF o *Head-Related Transfer Function*.

Cuando hablamos de HRTF nos referimos a un conjunto de funciones de transferencia que representan la respuesta al impulso medida en el canal auditivo de cada oído haciendo llegar sonidos desde múltiples posiciones del espacio previamente conocidas.

En este apartado vamos a describir el método de cálculo de estas funciones. En todo momento tenemos que tener en cuenta que las características físicas que diferencian a todos los individuos hacen imposible la existencia de un conjunto de HRTF universal que sea plenamente válido para cualquier persona. Los mejores resultados se obtienen calculando este conjunto de funciones de forma completa para cada sujeto.

En [Gardner 1994] se nos describe todo el proceso realizado para el cálculo de las funciones de transferencia para un modelo de maniquí conocido como KEMAR (ver figura 5.3). Mucho se ha discutido de la idoneidad de utilizar las medidas obtenidas en este tipo de experimentos. Como ya se ha comentado, la forma física del oído del maniquí no tiene por qué corresponderse con la forma del oído de ningún individuo; asimismo, en este experimento, el maniquí estaba equipado con un modelo de pinna-tragus simétrico, es decir, idéntico en ambos oídos.

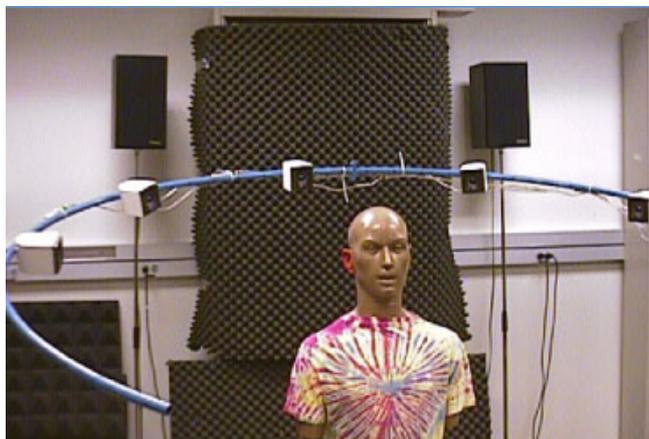


Figura 5.3: Modelo de maniquí KEMAR utilizado en el cálculo de las HRTF.

En el mencionado documento se describe la forma de extraer la respuesta impulsiva a partir de una serie de sonidos emitidos por un altavoz 'Realistic Optimus Pro 7' situado a 1.4 metros del maniquí, y la señal registrada en un par de micrófonos colocados en el canal auditivo del modelo KEMAR. En total realizaron 710 medidas distintas, variando la elevación desde -40° hasta 90° .

Los sonidos emitidos estaban muestreados con una frecuencia de 44.1 kHz y 16 bits para su representación digital. Más concretamente se definen secuencias de longitud $N = 16383$ muestras, se concatenan dos copias de cada secuencia y se reproduce en una posición del altavoz determinada. Simultáneamente se recogen otras $2 \cdot N$ muestras en el micrófono de cada canal auditivo. De estos datos se descartan los N primeros y los restantes se duplican. Para cada canal, mediante una convolución basada en la FFT, se ha realizado la correlación cruzada de la secuencia de muestras recogidas por los micrófonos con la secuencia emitida. Como resultado de este proceso se tiene la respuesta impulsiva que estábamos buscando.

Al igual que en cualquier experimento científico, y con más sentido en el caso de trabajar con señales acústicas, tenemos que contar con la presencia de ruido en todo el proceso. La forma que tenemos en este caso para reducir el efecto producido por el ruido es aumentar el tamaño de las secuencias de reproducción, pero habrá que tener en cuenta el aumento del coste computacional que ello conlleva.

5.2.1 Procedimiento de medida

Todas las medidas de [Gardner 1994] se realizan en una cámara preparada para absorber las ondas sonoras y que no se reflejen, evitando los ecos. El maniquí KEMAR fue montado en un asiento motorizado, capaz de realizar giros a cualquier azimut de una forma automática y muy precisa.

Por otro lado, el altavoz fue colocado en un soporte capaz de posicionar la fuente de sonido en cualquier elevación respecto de la posición del maniquí. La distancia entre ambos elementos era de 1.4 metros.

Una vez se tenían ambas estructuras listas, se realizaron las medidas en el espacio esférico alrededor del KEMAR, desde elevaciones de -40° (40° por debajo del plano horizontal) hasta los 90° (perpendicular al plano horizontal por la parte superior). Para cada elevación, se hizo un muestreo de los 360° alrededor del maniquí, a intervalos regulares. La tabla 5.1 nos muestra el número de muestras y el incremento en el azimut para cada una de las elevaciones, sumando un total de 710 medidas.

Elevación	Nº medidas	Incremento Azimut
-40	56	6.43
-30	60	6.00
-20	72	5.00
-10	72	5.00
0	72	5.00
10	72	5.00
20	72	5.00
30	60	6.00
40	56	6.43
50	45	8.00
60	36	10.00
70	24	15.00
80	12	30.00
90	1	-

Tabla 5.1: Conjunto de medidas realizadas con el maniquí KEMAR [Gardner 1994].

Si el modelo KEMAR fuese totalmente simétrico nos podíamos haber ahorrado la realización de la mitad de las medidas, teniendo suficiente con las de un hemisferio, izquierdo o derecho, e intercambiar las funciones para los canales en caso de querer generar sonidos procedentes del hemisferio opuesto al medido.

A partir de toda la información recopilada en el experimento, se obtuvo para cada modelo de pinna un conjunto de ficheros, uno por cada posición medida, que almacenan los coeficientes de un filtro FIR de 128 muestras para cada canal auditivo.

La forma de localizar espacialmente un sonido, dada una posición deseada, es convolucionar el sonido a localizar con los filtros correspondientes a dicha posición y reproducir cada uno de los resultados de esta operación por el canal de audio correspondiente. La fórmula para realizar la convolución es la siguiente:

$$y(n) = \sum_{i=0}^{N-1} x(n-i) \cdot h(i) \quad (5.1)$$

donde $x(n)$ es la muestra actual del sonido original, $y(n)$ es la muestra actual del sonido a reproducir, N es el número de muestras que tiene el filtro, y h es el vector con los coeficientes del filtro.

En [Algazi 2001] podemos encontrar la descripción de otro conjunto de medidas, conocido como 'CIPIC HRTF database', esta vez realizadas tanto a 43 personas (27 hombres y 16 mujeres), como a un maniquí modelo KEMAR. Del experimento realizado y cuyos resultados se muestran en este artículo, se obtuvieron 1250 medidas para cada individuo. Los filtros obtenidos están representados con 200 muestras para cada canal.

La principal particularidad que podemos encontrar en este conjunto de medidas realizado, es la inclusión de un estudio estadístico antropométrico acerca de 27 parámetros que determinan físicamente a los sujetos que intervinieron en el proceso. La inclusión de esta información nos puede ayudar a seleccionar un conjunto de medidas más acorde a otra persona que no intervino en el experimento, determinando sus características antropométricas y tomando los filtros medidos en el individuo que más se adecue.

La figura 5.4 nos muestra qué características físicas se han medido en este experimento, y la tabla 5.2 resume las medidas realizadas.

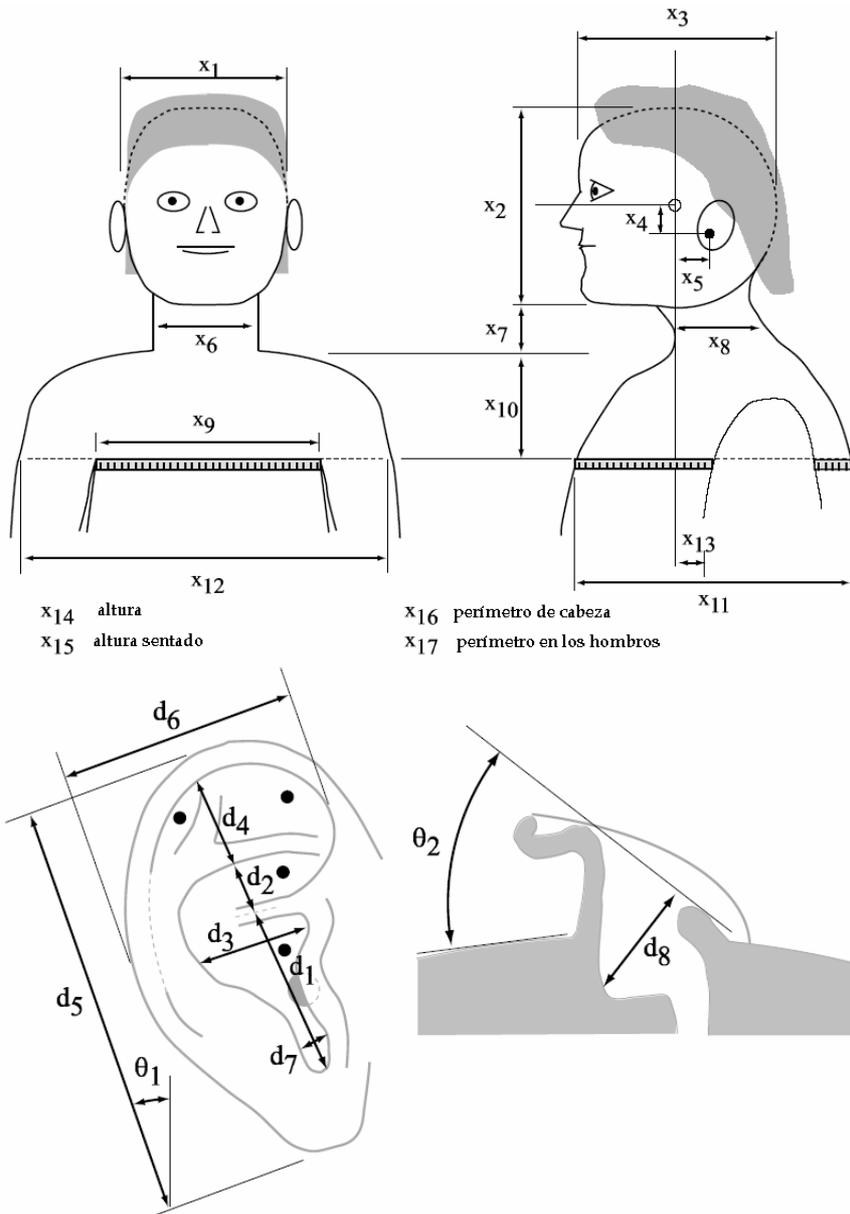


Figura 5.4: Características antropométricas incluidas en el estudio de [Algazi 2001].

Variable	μ	σ	Variable	μ	σ
x_1	14.49	0.95	x_{15}	88.83	5.53
x_2	21.46	1.24	x_{16}	57.33	2.47
x_3	19.96	1.29	x_{17}	109.43	10.30
x_4	3.03	0.66	d_1	1.91	0.18
x_5	0.46	0.59	d_2	0.68	0.12
x_6	11.68	1.11	d_3	1.58	0.28
x_7	6.26	1.69	d_4	1.51	0.33
x_8	10.52	1.22	d_5	6.41	0.51
x_9	31.50	3.19	d_6	2.92	0.27
x_{10}	13.42	1.85	d_7	0.53	0.14
x_{11}	23.84	2.95	d_8	1.02	0.16
x_{12}	45.90	3.78	θ_1	24.01	6.59
x_{13}	3.03	2.29	θ_2	28.53	6.70
x_{14}	172.43	11.61			

Tabla 5.2: Resumen de las características físicas medidas (distancias en cm y ángulos en grados) [Algazi 2001].

5.3 Generación de sonidos 3D desde Retiner

El modelo de procesamiento visual que implementa Retiner está diseñado de forma modular y con un nivel suficiente de abstracción como para poder utilizar los resultados de cada etapa en distintos dispositivos orientados a la ayuda y rehabilitación visual.

En concreto, la información de salida del módulo de cómputo de los campos receptivos, la matriz de actividad, puede aportar información útil para ser transmitida al invidente a través del sistema auditivo, como queda ilustrado en la figura 3.17. De esta forma podemos dividir el espacio visual capturado por la(s) cámara(s) en distintas zonas, asignando a cada una de estas áreas un sonido que le sea al usuario representativo sobre la localización de un objeto en esa zona.

En nuestro caso se ha decidido usar un sonido con forma de onda sinusoidal, que es agradable al oído (véase figura 5.5), que será convolucionada con el filtro correspondiente a la zona en la que se haya

detectado mayor actividad, previamente seleccionado por el usuario. En el momento actual estamos utilizando las funciones HRTF medidas para el maniquí KEMAR, aunque lo ideal sería realizar las mediciones correspondientes para cada individuo que utilice el sistema.

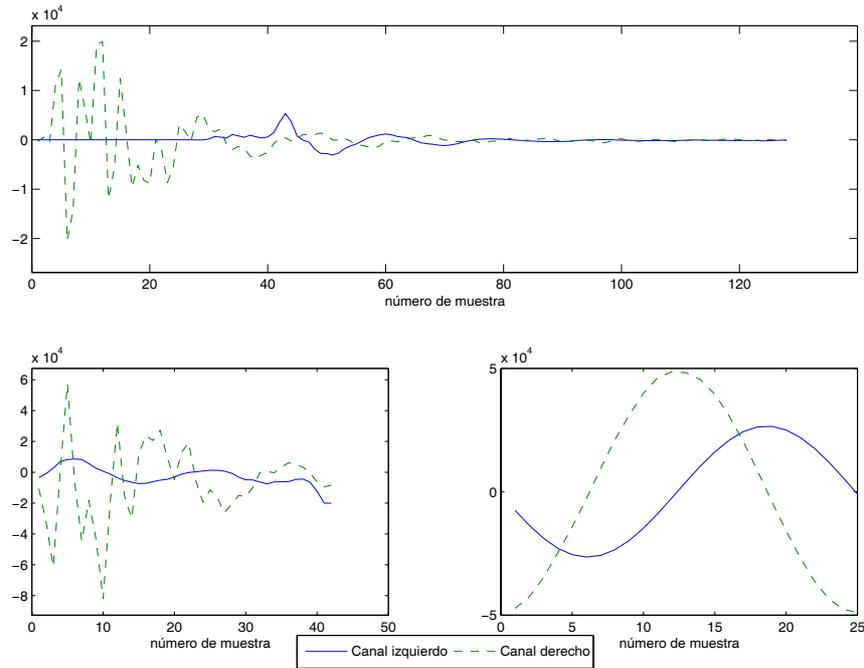


Figura 5.5: Arriba: representación de los coeficientes de ambos canales de un filtro. Abajo izquierda: un periodo del resultado de filtrar una señal cuadrada con el filtro de la figura superior. Abajo derecha: un periodo del resultado de filtrar una señal sinusoidal con el filtro de la figura superior [Morillas 2006c].

Para comprobar la validez del modelo de localización espacial de sonidos se diseñó una aplicación en MatlabTM que permite distintos tipos de pruebas. La decisión de realizar la implementación en este entorno estuvo condicionada por la facilidad en el manejo de los tipos de datos vectoriales, y la presencia de funciones específicas para realizar los cálculos de convolución y para la reproducción de sonidos. Tampoco debemos olvidar que la finalidad era integrarlo en Retiner, que como sabemos, está diseñado con esta herramienta.

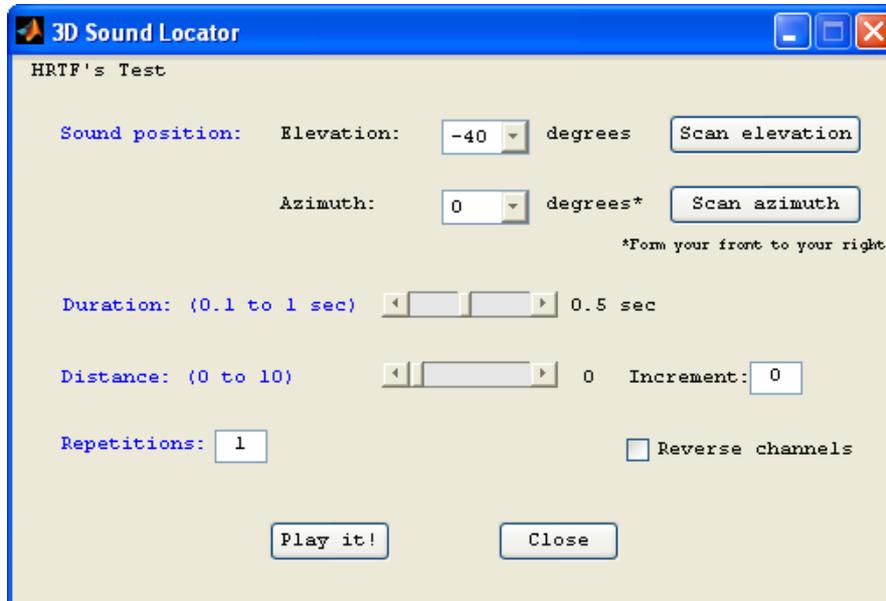


Figura 5.6: Módulo de generación de sonidos que permite realizar distintas pruebas, y comprobar la validez del modelo de procesamiento.

La figura 5.6 nos muestra la interfaz diseñada para el módulo de generación de sonidos. Podemos seleccionar una posición determinada mediante su elevación y su azimut, así como configurar la duración del sonido a reproducir. Además permite modelar una distancia con respecto al sujeto que realiza las pruebas, atenuando en amplitud la señal emitida. Si seleccionamos un incremento distinto de cero realiza un recorrido por las diferentes distancias posibles. Siguiendo la misma filosofía, podemos fijar una elevación determinada y hacer un recorrido por todos los azimut disponibles (*Scan azimuth*), y viceversa, seleccionado un azimut, podemos explorar todas las elevaciones (*Scan elevation*). Los últimos dos parámetros que podemos manipular son el número de repeticiones de cada sonido y si deseamos intercambiar los canales izquierdo y derecho. Como se ha comentado en la sección 5.2.1, en el experimento de Gardner [Gardner 1994] se ha determinado la simetría de las HRTF para ambos lados de la cabeza (por el tipo de maniquí utilizado), lo que reduce las necesidades de almacenamiento a uno de los hemisferios, en este caso el derecho; por tanto, para generar sonidos cuya posición se encuentre en el hemisferio izquierdo, basta con intercambiar los canales acústicos.

Una de las características que en todo momento se ha decidido mantener es la de realizar una selección estricta de la información a

transmitir, pensando en su utilidad para el usuario, y que le sea fácil captar dicha información. Consideramos que el éxito de cualquier sistema de este tipo está totalmente condicionado por la calidad y selectividad del sistema de preprocesamiento visual. Pensando en esto, se ha descartado una traducción a sonidos de todas las posiciones de la matriz de actividad. La primera aproximación que hemos implementado selecciona el máximo de esta figura de información y es el que transmite al sujeto. No obstante, el sistema no está totalmente cerrado y admite la posibilidad de modificar esta característica.

La selección del sonido asociado a cada posición de la matriz de actividad debe representar para el invidente la zona de su campo visual. No obstante, se deja como un grado de libertad al usuario la selección del filtro que desee asociar a cada uno de los elementos de la matriz de actividad.

Algunos grados de libertad más que aún disponemos son la atenuación en amplitud del sonido. Modificándola podemos codificar información sobre la distancia de los objetos al observador; la frecuencia de la función seno, que puede provocar sonidos más agudos o más graves, permitiendo mezclar más de un sonido localizado espacialmente de forma simultánea. Además, podemos elegir cualquier otro sonido base, distinto del seno, que será dotado de las características de localización espacial mediante el filtrado.

Tras múltiples pruebas del sistema se han observado algunas dificultades para detectar de forma precisa la posición que representan los sonidos, cuando se emiten de forma independiente. También se observó que era más fácil de averiguar la procedencia cuando se hacía de forma relativa a otro sonido posicionado previamente, así que se decidió dar la posibilidad al usuario para que seleccionara un sonido invariante y que se emitirá justo antes de cada sonido que incorpore información espacial.

Siguiendo la misma filosofía, todo el sistema permite su adecuación al paciente, la duración de todos los sonidos (referencia y posicionado, además del intervalo entre ambos), el conjunto de HRTF a utilizar (posibilitando la incorporación de los filtros medidos para el invidente de forma particular), el volumen de la señal, la forma de onda del sonido base, su frecuencia, la resolución de la división del campo visual, etc. Como se puede ver, un sistema totalmente orientado al usuario.

5.4 Conclusiones

Se han descrito los aspectos físicos relacionados con la propagación de los sonidos que rigen la percepción de su localización tridimensional.

También hemos revisado algunos estudios orientados a la realización de las medidas experimentales que han dado como resultado un conjunto de filtros FIR para poder sintetizar sonidos que incorporen las características propias de las señales localizadas en una posición arbitraria del espacio, incluyendo azimut y elevación.

Finalmente hemos descrito la forma en que este tipo de procesamiento extiende a nuestro sistema de procesamiento visual bioinspirado, Retiner. Esta extensión está orientada a los sistemas de rehabilitación visual basados en la transducción sensorial, usando como medio de transmisión de la información el sistema auditivo. Cabe destacar el alto grado de adaptabilidad a cada paciente que permite la implementación realizada, así como su adecuación a la implementación en hardware configurable (FPGA), tal como se describe en el capítulo 6. El objetivo principal de esta implementación software del modelo responde a las necesidades de evaluar la generación de sonidos 3D y llevar a cabo el ajuste / configuración de todos los parámetros para cada sujeto.

Como hemos resaltado en la sección 5.3, la efectividad de un sistema de conversión de imágenes a sonidos 3D va a depender en gran medida de la calidad y selectividad del sistema de procesamiento visual. Frente a otras aproximaciones existentes (véase sección 2.3.2 [Nielsen 1989], [vOICe], [EAV]), los patrones sonoros 3D que proponemos utilizar, al ser captados por el invidente interfieren mínimamente sus capacidades de percepción auditiva, preservando así una de las mayores fuentes de información sensorial de que dispone.

Sistemas para Rehabilitación Visual integrados en FPGA

Este capítulo se centra en describir la implementación hardware basada en circuitos reconfigurables que se ha realizado de los distintos modelos de procesamiento visual, de producción de impulsos y de transducción sensorial presentados en capítulos anteriores. En primer lugar describimos el diseño de una solución portable en hardware reconfigurable que realiza el procesamiento visual bioinspirado, basado en una combinación de filtros espaciales; lo que nos ocupa la sección 6.1. En el siguiente apartado, sección 6.2, se expone el proceso de implementación hardware de un sistema de generación de impulsos a partir de la información generada por el módulo de procesamiento visual, destinado a la estimulación nerviosa mediante neuroprótesis visuales. Finalmente, la sección 6.3 se dedica a la implementación de un modelo de localización espacial de sonidos y su integración con el módulo de la sección 6.1.

6.1 Implementación hardware de un sistema de procesamiento visual bioinspirado

Los requisitos deseados para este tipo de sistemas como la portabilidad, bajo consumo, respuesta en tiempo real y la capacidad de incorporar tareas de procesamiento que implementen diferentes modelos de procesamiento visual, hacen como una alternativa válida el uso de circuitos reconfigurables del tipo FPGA que se adecuan a las características deseadas, además de ofrecer un alto nivel de paralelismo. Una de las claves para el éxito de la implementación hardware de este tipo de sistemas se encuentra, indudablemente, en la paralelización del máximo de tareas posibles.

Como herramientas de trabajo se ha dispuesto de un lenguaje de descripción de hardware de alto nivel, conocido como **Handel-C**, que está basado en el lenguaje de programación ANSI-C e incluye las extensiones necesarias para modelar diversos aspectos relacionados directamente con el hardware, como la definición de bloques que trabajan en paralelo, interfaces con otra circuitería adicional, etc. Este lenguaje es propietario de la empresa Celoxica [Celoxica], que además provee de un entorno de diseño basado en este lenguaje y que se llama **DK**. Esta herramienta realiza un proceso de optimización de recursos y traducción a EDIF, una representación que necesita de una herramienta de *place&route* para mapear finalmente los distintos componentes sobre la circuitería específica del dispositivo que utilizemos. En nuestro caso, la herramienta es el entorno **ISE** de Xilinx [Xilinx], que es el fabricante de las FPGAs que incorporan las tarjetas utilizadas.

6.1.1 Plataformas de diseño hardware

En nuestro caso, las tarjetas utilizadas se encuentran dentro de la gama ofrecida por la empresa Celoxica. A lo largo de todo el proceso se ha hecho uso de las tarjetas RC-100, RC-200 y RC-300, cuyas características pasamos a detallar.

El uso de las distintas plataformas viene determinado por las características propias de la aplicación concreta, dando soporte a las necesidades de recursos exigidas por cada una de ellas.

6.1.1.1 RC-100

Se trata del modelo con menor cantidad de recursos de la familia. Sus características básicas se detallan a continuación, y la figura 6.1 nos ofrece una imagen de la plataforma al completo.

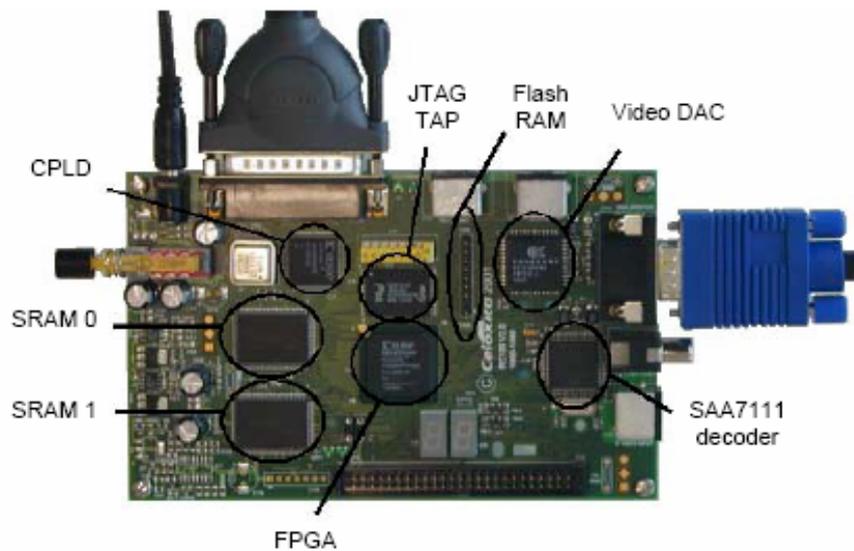


Figura 6.1: Plataforma RC-100 de Celoxica.

La tarjeta RC-100 incluye dos chips que albergan lógica programable:

- Xilinx Spartan II FPGA: con el equivalente a 200000 puertas, siendo la principal fuente de recursos programables de la tarjeta.
- Xilinx XCR3128XL CPLD: El CPLD (Complex Programmable Logic Device) es el encargado de la configuración de la Spartan II FPGA a través del puerto paralelo o de la Flash RAM.

Además la tarjeta contiene los siguientes elementos:

- Flash RAM (8 Mbytes de Flash RAM).
 - o Para almacenar múltiples configuraciones o para propósito general.
- SSRAM (Dos bancos independientes de 256k palabras x 36 bits).
 - o Usado por la FPGA para almacenar datos de propósito general.
- Sistema de salida de video (Video DAC & VGA connector).

- DAC (Convertor Digital a Analógico) con un mapa de color de 24 bits. Usado para la salida de video a un monitor VGA.
- Circuitería de decodificación de video. (Decodificador SAA7111, s-video, entrada de vídeo compuesto).
 - Acepta señales NTSC/PAL/SECAM a través de un conector RCA o S-Video y lleva las salidas de señales RGB a la FPGA.
- Puerto paralelo.
 - Usado para la programación de la FPGA y de la Flash RAM. Permite comunicación bidireccional.

Otros elementos son:

- 2 Displays de 7 segmentos.
- LEDs.
- Cuarzo para generar la señal de reloj principal.
- Programación mediante JTAG
- 2 puertos PS/2 para ratón y teclado
- Puerto de expansión para conectar con otra circuitería.

6.1.1.2 RC-200

La lógica reconfigurable que incluye está repartida en dos circuitos:

- Virtex-II 2V1000-4 FPGA de un millón de puertas.
- CPLD Xilinx XC95144XL para configuración y mantenimiento de la tarjeta SmartMedia.

Otros elementos de interfaz que incluye son los siguientes (véase figura 6.2):

- Ethernet MAC/PHY con socket 10/100baseT.
 - 2 bancos de SRAM ZBT que suman 4MB.
 - Soporte de Video:
 - Entrada/salida de video compuesto.
 - Entrada/salida de S-video.
 - Salida VGA
 - Audio AC'97 compatible que incluye:
 - Entrada de micrófono.
 - Entrada de línea (estéreo).
 - Salida de línea/Auriculares (estéreo).
 - Conector SmartMedia Flash memory para almacenar archivos BIT de configuración del dispositivo.
-

- Puerto Paralelo.
- Puerto serie RS232.
- Conectores PS/2 para teclado y ratón.
- 2 Displays de siete segmentos.
- 2 LEDs azules.
- Conector de expansión que permite hacer de interfaz con otros elementos.
- 2 pulsadores.
- Conector JTAG para la configuración y test.

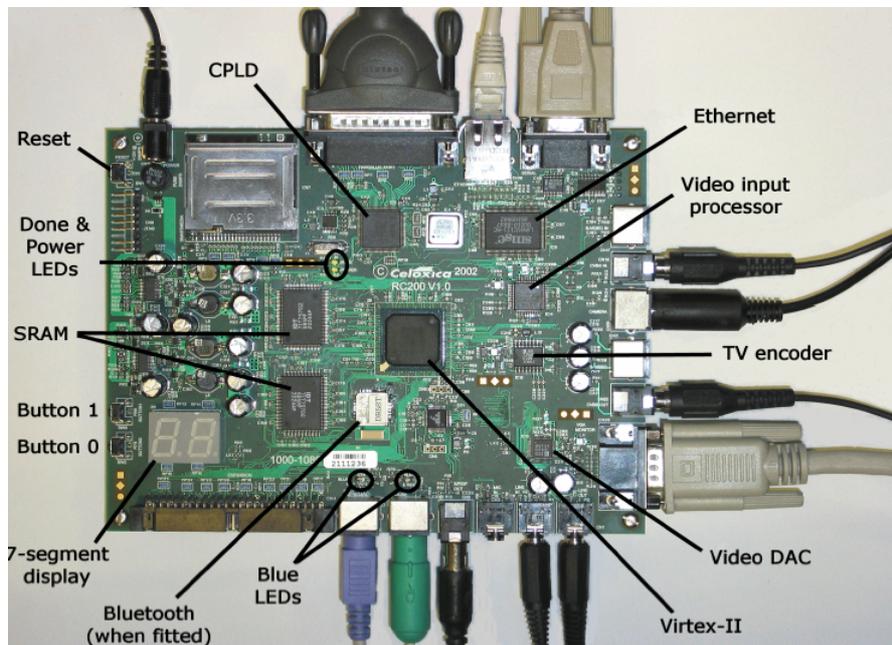


Figura 6.2: Detalle de todos los elementos incluidos en la tarjeta RC-200.

6.1.1.3 RC-300

Se trata de la plataforma portable más completa que ofrece Celoxica, y presenta las siguientes características:

- Virtex-II 2V6000-4FF1152, FPGA de 6 millones de puertas.
- Ethernet Gigabit dual implementada por un dispositivo MAC Ethernet y dos dispositivos PHY Ethernet.
- 4 bancos de ZBT SRAM, los cuales suman un total de 32MB.
- Socket DIMM para memoria DRAM.
- Soporte de video dual, incluyendo:

- Entrada/Salida DVI dual.
- Entrada de vídeo compuesto dual.
- Entrada/salida S-Video dual.
- Salida VGA dual.
- Audio AC'97 compatible que incluye:
 - Entrada de micrófono.
 - Entrada de línea (estéreo).
 - Salida de línea / Auriculares (estéreo).
 - Audio digital S/PDIF.
- Tarjeta SmartMedia preprogramada , que incluye:
 - Diversos programas de demostración.
 - Programas para hacer pruebas.
- Microcontrolador USB para:
 - Gestión del puerto USB.
 - Configuración de la FPGA.
 - Gestión de SmartMedia.
- Puertos RS-232 duales.
- Conectores PS/2 para teclado y ratón.
- Display LCD con 4 líneas de 16 caracteres.
- 8 LEDs blancos.
- 4 pulsadores.
- Cabecera de expansión de 50 pines, que incluye:
 - 33 pines de entrada/salida generales (diseño compatible con ATA UDMA-4 o más alto).
 - 3 pines de alimentación (+12V, +5V, +3.3V).
 - 2 pines de reloj.
- Conector JTAG para configuración y pruebas

En la figura 6.3 podemos observar todos estos elementos presentes en la tarjeta.

6.1.2 Dispositivo de captura de imágenes

La secuencia de imágenes de entrada al sistema era captada por una cámara de bajo coste a color, en concreto el modelo *C-Cam8* suministrado por Comedia [*Comedia*]. Sus principales características son:

- Tamaño reducido: 22 x 22 x 28 mm.
 - Lentes integradas con filtro de infrarrojos.
 - Resolución de 365k píxeles.
 - Bajo consumo.
 - Salida de video compuesto en formato PAL.
-

- Control automático de exposición.
- Compensación de luz de fondo.
- Corrección gamma.
- Control automático de ganancia.
- Balance automático de blancos.

En la figura 6.4 tenemos una imagen de la cámara utilizada.

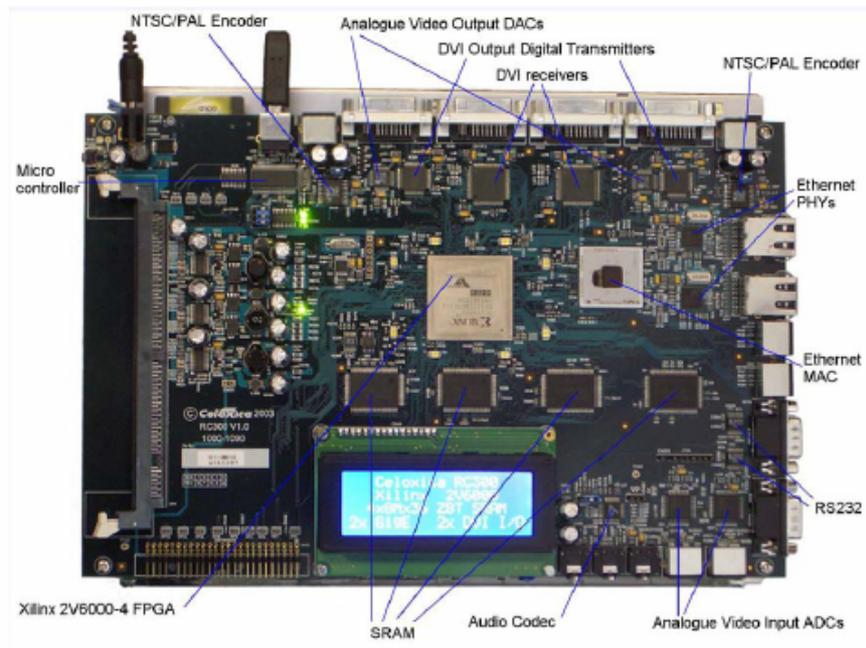


Figura 6.3: Tarjeta RC-300.



Figura 6.4: Cámara a color modelo C-Cam8 de Comedia.

6.1.3 Diagrama de bloques

La primera versión de la solución portable de nuestro sistema de procesamiento visual bioinspirado está implementada en la plataforma RC-100. Se trata de una arquitectura de procesamiento segmentada o en *pipeline*, regida por una señal de reloj común a todo el circuito. El diagrama de bloques de los distintos procesos en que se han dividido las tareas de procesamiento se presenta en la figura 6.5.

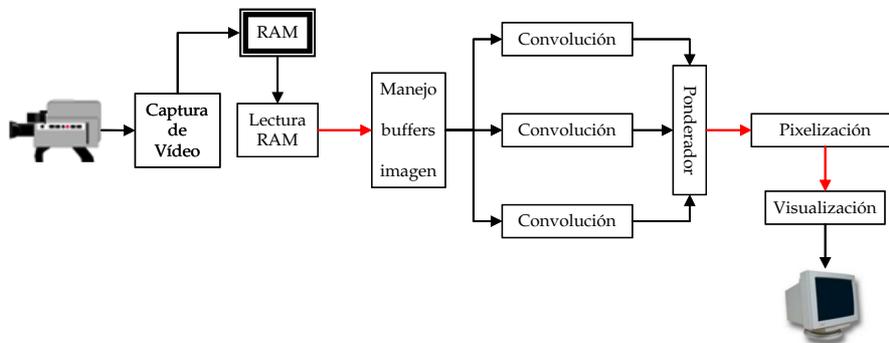


Figura 6.5: Diagrama de bloques de la implementación hardware del sistema de procesamiento visual bioinspirado.

A continuación pasamos a describir con más detalle todos los aspectos más relevantes y relativos a la implementación hardware de los diferentes procesos.

6.1.3.1 Captura de video o *frame-buffer*

El primer proceso está dedicado a la recepción de los datos que ofrece la cámara y almacenarlos en la memoria RAM, para desacoplar el funcionamiento del resto del sistema de la tasa de generación de imágenes que tenga la cámara.

La plataforma utilizada dispone de una librería de rutinas para acceder a los distintos periféricos que incorpora. En concreto, nos ofrece una función para recibir los píxeles de la imagen que genera el convertidor de video. La información que nos devuelve esta función consta de las

componentes de color de dos píxeles vecinos y las coordenadas del primero de ellos.

Como cada píxel está codificado con 16 bits en el formato RGB (5 bits para el rojo, 6 para el verde y 5 para el azul), y la memoria SSRAM en la que se almacena está organizada en 256k palabras de 36 bits cada una, podemos almacenar ambos píxeles recibidos (que ocupan un total de 32 bits) en una sola posición de la memoria. Haciendo uso de las coordenadas que nos devuelve la función podemos calcular la posición de la memoria RAM en la que quedan almacenados.

La memoria utilizada es síncrona y no permite accesos en paralelo al mismo módulo físico. Nos podemos aprovechar de que existen dos módulos físicos e implementar un sistema de captura con doble buffer, de forma que mientras almacenamos los píxeles capturados por la cámara en uno de estos módulos, los procesos que leen lo hacen del otro módulo físico, permitiendo realizar dichos accesos en paralelo, en un solo ciclo de reloj.

6.1.3.2 Lectura de la memoria RAM

Este proceso está dedicado a la tarea opuesta al anterior, es decir, leer secuencialmente los píxeles previamente almacenados en la memoria RAM y mandarlos a etapas posteriores de procesamiento.

La característica principal de este módulo es que ofrece su salida a través de un tipo de dato específico que incorpora Handel-C, y que es conocido como *canal* (*chan*). Este tipo de datos abstrae un medio de comunicación y un mecanismo de sincronización entre procesos, ya que todas las transferencias entre los procesos interconectados esperan a que ambos extremos se encuentren preparados para llevarla a cabo.

Aunque no aparezcan relacionados en el diagrama de bloques, este módulo y el de captura de imágenes ya descrito comparten una variable que es la que permite implementar la técnica de doble buffer, indicando a cada uno de los procesos el módulo de memoria físico que debe utilizar.

6.1.3.3 Filtrado espacial

En primer lugar queremos hacer resaltar que en este capítulo estamos describiendo todo el proceso llevado a cabo para conseguir una implementación concreta del sistema de rehabilitación, pero que todos los módulos desarrollados están plenamente parametrizados, de tal forma que para conseguir una versión distinta con otras características bastaría con modificar el valor de los parámetros que deseemos cambiar y realizar una compilación de todo el sistema, hasta llegar al archivo de configuración del circuito reprogramable.

El filtrado espacial de las imágenes de entrada tiene como objetivo realzar distintas características de la misma, intentando imitar la forma en que lo hace el sistema visual humano. En concreto, realizamos dos filtrados que realzan el contraste de color (magenta frente al verde y amarillo frente al azul), y uno que realza el contraste de intensidades.

Para extraer esta información necesitamos convolucionar la imagen con unas máscaras de diferencia de gaussianas, que tienen una forma conocida como '*sombrero mejicano*' y que se observa en la figura 6.6.

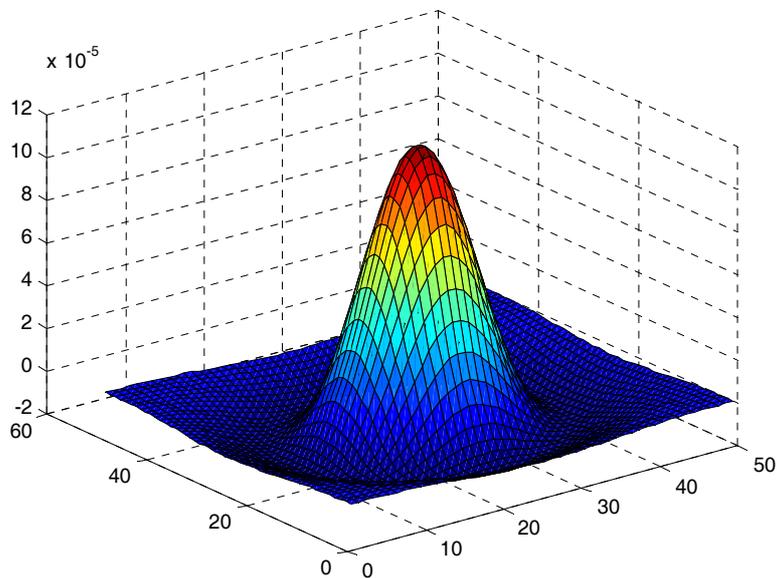


Figura 6.6: Máscara de convolución del operador diferencia de gaussianas.

La forma de obtener esta máscara es mediante la diferencia de dos máscaras gaussianas de diferente desviación típica σ . Para definir una transformación mediante una diferencia de gaussianas tendremos que definir qué valores toman las desviaciones típicas, σ_1 y σ_2 , y qué datos son los operados con cada máscara. Aprovechándonos de las propiedades de la operación de convolución podemos descomponerla en dos distintas, como nos muestra la ecuación (6.1):

$$DoG(A, \sigma_1, B, \sigma_2) = Gauss(A, \sigma_1) - Gauss(B, \sigma_2) \quad (6.1)$$

donde A es el flujo de información que será convolucionado con la máscara gaussiana con parámetro σ_1 , y B el que será operado con el núcleo definido por el parámetro σ_2 .

Aunque en principio parece que empeoramos en el número de cálculos necesarios, como veremos en la sección 6.1.3.4, permitirá compartir términos en combinaciones más complejas, reduciendo el número total de operaciones a realizar.

En la forma más general, el resultado, I' , de convolucionar una imagen I con un núcleo o máscara de convolución h de tamaño $m \times n$, se define como indica la ecuación (6.2):

$$I'(i, j) = \sum_{k=-\frac{m}{2}}^{\frac{m}{2}} \sum_{l=-\frac{n}{2}}^{\frac{n}{2}} I(i-k, j-l) h(k, l) \quad (6.2)$$

Si nos detenemos a observar el modo de funcionamiento de este operador, vemos que para calcular el valor asociado a uno de los píxeles necesitamos una vecindad del mismo que incluye puntos de filas anteriores y posteriores de la imagen, al igual que columnas anteriores y posteriores. Sin embargo, el módulo de lectura de la memoria RAM va proporcionando los píxeles de uno en uno, de izquierda a derecha y de arriba hacia abajo. Para solventar el problema utilizamos una arquitectura de cómputo basada en la que propuso Ridgeway en [Ridgeway 1994]. De esta forma podemos tener preparados los píxeles necesarios de la imagen para calcular la convolución en un momento dado introduciendo una serie de buffers tipo FIFO, que deben tener una longitud igual a la anchura de la imagen (W) y, además, tenemos que disponer de $m-1$ buffers (suponiendo un núcleo de convolución de m filas y n columnas, $m \times n$).

Adicionalmente necesitamos de m registros de desplazamiento de n etapas cada uno, que son los que nos van a proveer en todo momento de los píxeles de la imagen que están involucrados en la operación de convolución actual. La figura 6.7 nos ilustra la arquitectura resultante.

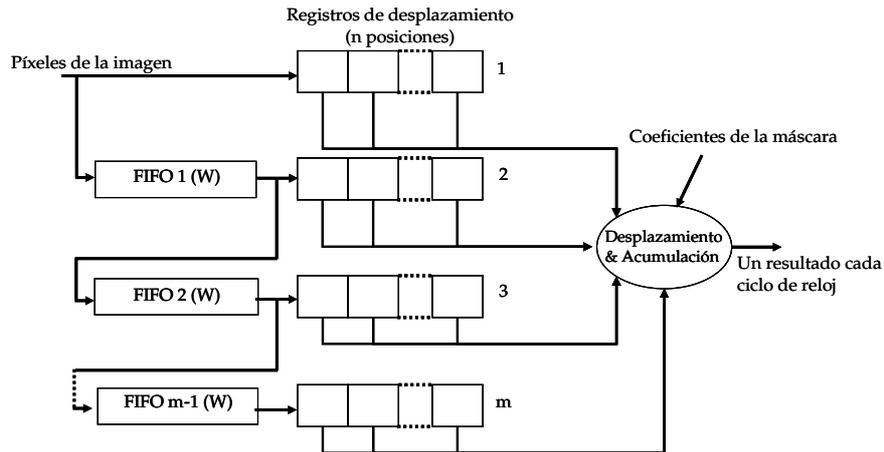


Figura 6.7: Arquitectura de cómputo de la operación de convolución [Pelayo 2003].

Debido a la limitación de recursos disponibles en la plataforma hardware que disponíamos en el momento de la implementación, se decidió aproximar las máscaras gaussianas en base a valores que sean potencia de 2, y que la suma de todos los coeficientes sea un valor lo más cercano posible a otra potencia de 2 [Pino 1999a], [Pino 1999b]. De esta forma, podemos realizar la operación de multiplicación por los coeficientes, y la división para normalizar el resultado de la operación, mediante desplazamientos a nivel de bits.

Como vemos en el diagrama mostrado en la figura 6.5, se realizan tres convoluciones en paralelo, pero con una sola estructura de FIFOs podemos realizar todo el procesamiento. No obstante, esta estructura debe ser capaz de almacenar los valores de los tres planos de color simultáneamente.

6.1.3.4 Combinación de los filtros

Este proceso se encarga de recoger los resultados producidos por los diferentes filtros y agruparlos en un solo resultado, mediante una combinación lineal de los mismos.

En nuestro caso se han calculado los siguientes filtros:

- $DoG(R + B, \sigma_1, G, \sigma_2)$: que realza el contraste magenta frente al verde.
- $DoG(R + G, \sigma_1, B, \sigma_2)$: que contrasta el amarillo frente al azul.
- $DoG(R + G + B, \sigma_1, R + G + B, \sigma_2)$: que realiza un contraste de los niveles de intensidad, produciendo resultados similares a la detección de bordes.

Y la combinación de los mismos que hemos decidido implementar es la siguiente, aunque recordamos que se puede sustituir por cualquier otra:

$$\begin{aligned} \text{Retina} = & 2 \cdot DoG(R + B, \sigma_1, G, \sigma_2) + \\ & DoG(R + G, \sigma_1, B, \sigma_2) + \\ & DoG(R + G + B, \sigma_1, R + G + B, \sigma_2) \end{aligned}$$

Podemos observar que se ha dado más peso al filtro que realza el realce del color magenta en oposición al color verde. En todas las implementaciones que vamos a realizar de filtros basados en la operación de convolución no aprovechamos de las propiedades de linealidad de esta operación para realizar optimizaciones en el espacio de los operadores que definen la combinación. La expresión anterior quedaría como sigue:

$$\begin{aligned} \text{Retina} = & 4 \cdot Gauss(R, \sigma_1) - Gauss(R, \sigma_2) + \\ & 2 \cdot Gauss(G, \sigma_1) - 3 \cdot Gauss(G, \sigma_2) + \\ & 3 \cdot Gauss(B, \sigma_1) - 2 \cdot Gauss(B, \sigma_2) \end{aligned}$$

De esta forma, todas las implementaciones que podamos imaginar, siempre que los operadores diferencia de gaussiana compartan los parámetros que definen la forma de la máscara, se pueden realizar si más que implementar estas seis convoluciones.

Otra alternativa que permite reducir el número de convoluciones sería la de realizar la optimización en el espacio de los coeficientes de las máscaras. Combinando los distintos factores y parámetros podríamos llegar a reducir el procesamiento a tan sólo tres convoluciones, una sobre cada canal de color, donde las máscaras serían el resultado de operar las gaussianas de base con los factores y combinarlas (véase para más detalle [Martínez 2005]).

Los inconvenientes que plantea este nuevo esquema de cálculo es que ya no podemos asegurar que los coeficientes de las máscaras resulten potencias de dos, ya que serán una combinación lineal basada en productos y sumas, con lo que para asegurar el correcto funcionamiento tendremos que introducir módulos multiplicadores para aplicar ponderar los píxeles con los coeficientes resultantes.

Para cada caso concreto, según sea el modelo implementado, tendremos que evaluar qué implementación es la más adecuada, ya que tendremos que ver si compensa la reducción del número de desplazadores con la inclusión de multiplicadores. En todo caso tendremos que velar por la integridad de los resultados intermedios, evitando que el rango dinámico de los mismos sobrepase las capacidades de representación de las variables que los tienen que almacenar.

6.1.3.5 Cómputo de la matriz de actividad y visualización

La siguiente tarea en el esquema consiste en la pixelización o cómputo de la matriz de actividad en base a la definición de los campos receptivos. La implementación realizada contempla una división equitativa sin solapamientos de la imagen y su mapeo a una matriz de actividad de tamaño 20x20.

Debido al limitado número de recursos de la circuitería disponible, no ha sido posible implementar la matriz de tamaño 20x20 al completo, con lo que ha sido necesario acoplar este proceso con el que realiza la visualización en pantalla de esta imagen de resolución reducida.

La forma de realizarlo ha sido manteniendo un offset fijo de la fila de la imagen de salida que estamos visualizando con respecto a la fila de la imagen perceptual generada por la combinación de los filtros que estamos acumulando para el cálculo de la matriz de actividad.

La figura 6.8 nos ilustra el resultado de acoplar las tareas de visualización y cómputo de los campos receptivos.

De esta forma que hemos comentado, basta con mantener en memoria en todo momento dos filas de la matriz de actividad, reduciendo las necesidades de almacenamiento.

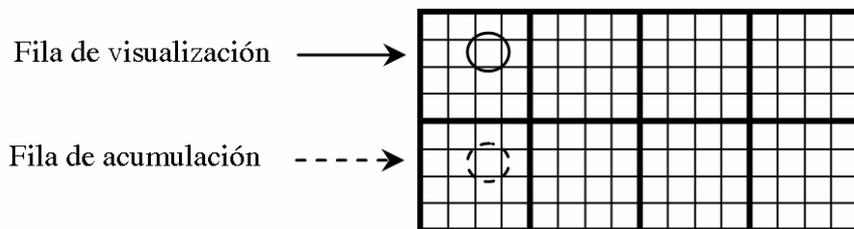


Figura 6.8: Esquema de cómputo de la matriz de actividad y de la visualización de los resultados obtenidos.

Para poder observar los resultados producidos por el circuito implementado utilizamos la salida de VGA disponible en la tarjeta RC-100, de forma que podemos conectarla a un monitor o a otros dispositivos de visualización, como unas gafas de realidad virtual, que nos permiten realizar experimentos psicofísicos con individuos sin problemas de visión (ver sección 3.3.4).

En la figura 6.9 podemos ver una foto tomada del monitor en que visualizábamos la salida generada por el sistema de procesamiento implementado.

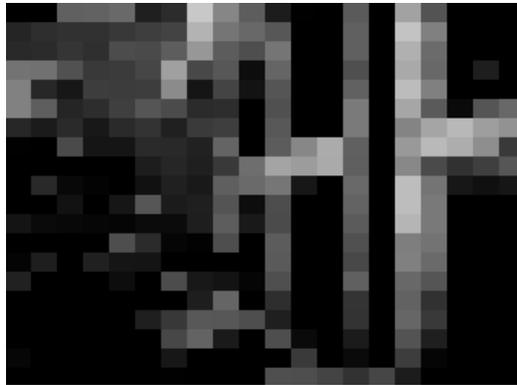


Figura 6.9: Instantánea del procesamiento visual que realiza la plataforma hardware portable.

6.1.3.6 Suavizado gaussiano

Con el fin de completar este subsistema para poder realizar experimentos psicofísicos más completos, tal y como describimos en la sección 3.3.4, se ha procedido a incorporar al sistema descrito con la posibilidad de producir imágenes suavizadas con una máscara gaussiana

(para intentar imitar la sensación producida por la neuroestimulación con los valores de la matriz de actividad).

Como se trata de una nueva operación de convolución aplicada a la matriz de actividad generada, basta con instanciar una vez más la arquitectura ilustrada en la figura 6.7. La figura 6.10 ilustra los resultados obtenidos cuando incorporamos el suavizado gaussiano. Podemos distinguir las diferencias con el resultado mostrado previamente en la figura 6.9.



Figura 6.10: Incorporación del suavizado gaussiano al procesamiento visual.

6.1.4 Resultados de síntesis de los circuitos

Tras realizar la implementación del sistema con y sin el suavizado gaussiano, los resultados de ocupación del dispositivo que incorpora la tarjeta RC-100 son lo que muestra la tabla 6.1.

	Total equivalente puertas	Ocupación (Slices)		BlockRAM	
		nº	%	nº	%
Sin suavizado	159289	1313	55	8	57
Con suavizado	229578	1535	65	12	85

Tabla 6.1: Resumen de recursos consumidos por las dos implementaciones.

Si recordamos, en la sección 6.1.1.1, cuando se detallaron las características de esta plataforma, se dijo que el equivalente en puertas lógicas de la FPGA que incorpora es de 200000. En este cómputo no se

incluyen los recursos de memoria centralizada o BlockRAM que incorpora. Sin embargo, la herramienta de síntesis sí incluye el consumo de estos recursos de almacenamiento en el cálculo del total de puertas equivalentes. Por esta razón, para la implementación del sistema que incorpora el suavizado se obtiene un valor total de puertas equivalentes que supera los que tendríamos disponibles.

En cuanto a las características temporales del sistema mencionaremos que, salvo el tiempo de latencia inicial en el que se están llenando todas las etapas del pipeline, se obtiene un píxel resultado por cada ciclo de reloj.

El mínimo periodo conseguido para la señal de reloj es de 75ns, con lo que podemos afirmar que éste es el tiempo de procesamiento asociado a cada píxel. Para procesar una imagen completa de entrada, proporcionada en formato PAL de 572 x 720, se consumen algo más de 0.03 segundos, con lo que la tasa de procesamiento es superior a 30 imágenes o frames por segundo. Suficiente para generar imágenes que produzcan una sensación de continuidad para el ojo humano.

6.2 Retina artificial

Además de la realización de experimentos psicofísicos para caracterizar las posibilidades del sistema de procesamiento visual que hemos implementado, podemos acoplarlo con la implementación hardware de un modelo de producción de impulsos destinados a la neuroestimulación.

En esta sección vamos a describir la arquitectura que nos permite implementar dicho modelo, que se ha descrito en la sección 3.3.3 de esta memoria de tesis doctoral. En concreto se ha realizado la implementación del modelo más básico del tipo 'integra y dispara' que incluye un término de pérdidas constante.

La estructura en bloques de un módulo básico es la que nos presenta la figura 6.11.

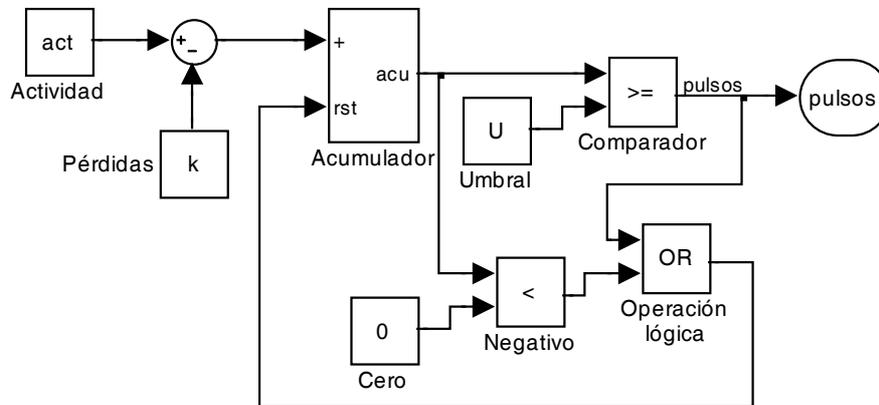


Figura 6.11: Esquema de procesamiento para la generación de eventos de estimulación para un electrodo individual.

En este punto se nos plantean algunas cuestiones, como la necesidad de incluir tantos módulos de generación de impulsos como elementos en la matriz de actividad tengamos. En la implementación hardware bastaría con replicar el proceso que realiza esta función básica tantas veces como sea necesario.

No obstante, los requisitos temporales de esta aplicación no hacen necesario el consumo de tantos recursos. Nuestro sistema visual entiende como simultáneos eventos producidos dentro de una ventana de duración del orden del milisegundo. La velocidad de procesamiento conseguida nos permite realizar todo el procesamiento de forma secuencial, manteniéndonos dentro de los límites temporales exigidos por la aplicación. Por tanto, nos basta con incorporar un solo módulo hardware para realizar la acumulación, comprobación y generación de evento cuando sea necesario, para todos los canales de estimulación que hemos introducido en nuestro sistema.

La arquitectura de cómputo secuencial final, incluyendo los elementos de memoria necesarios para su implementación es la que muestra la figura 6.12.

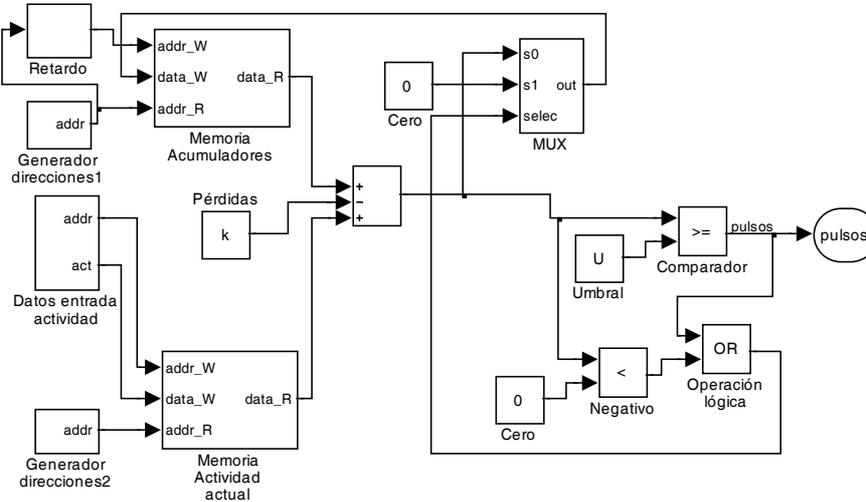


Figura 6.12: Arquitectura para la generación de impulsos para diferentes electrodos, utilizando un módulo individual, memorias de almacenamiento y esquema de cómputo secuencial.

6.3 VIS2SOUND: Sistema completo de transducción sensorial

Esta sección presenta otro ejemplo de aplicación práctica del módulo de procesamiento visual cuya implementación se ha tratado en la sección 6.1. Se trata de un sistema de transducción sensorial que transforma niveles de actividad extraídos de las imágenes capturadas del entorno en sonidos que incorporan información sobre localización espacial.

El contenido de este apartado se estructura en dos partes: la primera dedicada a la implementación hardware del sistema de localización de señales acústicas, y la segunda, que presenta el proceso de integración de este sistema con una ampliación de la arquitectura de procesamiento visual que incluye como novedad la estereovisión.

6.3.1 Subsistema de generación de sonidos 3D

El desarrollo de este módulo se ha realizado de una forma incremental, cambiando de plataforma hardware conforme aumentaban los recursos necesarios para incorporar mayor funcionalidad. A lo largo de este apartado iremos mostrando las distintas versiones que se han ido realizando, incluidas las herramientas software para el ajuste de los parámetros para cada usuario, hasta llegar a la implementación del sistema completo orientado a una solución portable.

6.3.1.1 Implementación en la RC-100

El desarrollo de este sistema lo hemos iniciado con la tarjeta RC-100 descrita con anterioridad. Dado que esta tarjeta no dispone de circuitería dedicada a la interfaz con señales de audio, ni de conector de auriculares, construimos un módulo para conectar los auriculares (ver figura 6.13 izquierda). Este módulo se conecta con la tarjeta RC-100 a dos líneas de señal digital, correspondientes a los canales izquierdo y derecho. Además consta de un potenciómetro que permite modificar el volumen de salida (esquema en figura 6.13 derecha).

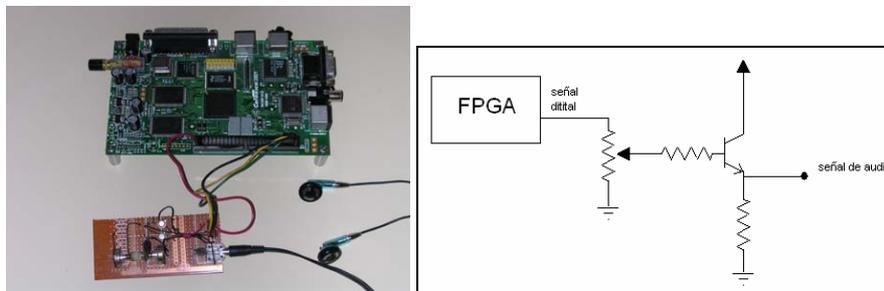


Figura 6.13: Izquierda: Detalle del sistema inicial con la RC-100 de Celoxica. Derecha: esquema eléctrico que implementa la interfaz con las señales de audio, para un solo canal.

Como sonidos de referencia se podrían seleccionar frecuencias distintas para distinguir distintos objetos de la escena, localizados en distintas posiciones. En nuestro caso, como punto de partida hemos decidido incluir sólo un sonido, asociándolo al objeto o ubicación espacial de mayor relevancia según el procesamiento visual realizado. Una vez fijada una frecuencia de reproducción, para generar las señales audibles usamos PWM (Pulse Width Modulation).

Como ya hemos dicho, la tarjeta no dispone de circuitería que ofrezca una reproducción a una determinada frecuencia. Para implementar esto simplemente usamos un divisor de frecuencia que ajusta la frecuencia de funcionamiento de la placa, 25 MHz, a la de reproducción deseada, 11 kHz.

En cada periodo de reproducción de la señal PWM tendremos la salida activa en alta el número de intervalos que indique el valor de la muestra actual (codificada en 8 bits) y el resto de intervalos con la salida a cero.

La **primera versión** del sistema implementada almacena en el interior de la FPGA un total de 1103 muestras de la señal seno, que corresponden a un sonido de duración 0.1 s, dado que la frecuencia de muestreo de la señal reproducida es de 11025Hz.

Por tanto, este primer modelo construido filtra las 1103 muestras del seno con uno de los filtros almacenado también en el interior del dispositivo. Si recordamos, los filtros utilizados constan de 128 coeficientes por cada canal auditivo.

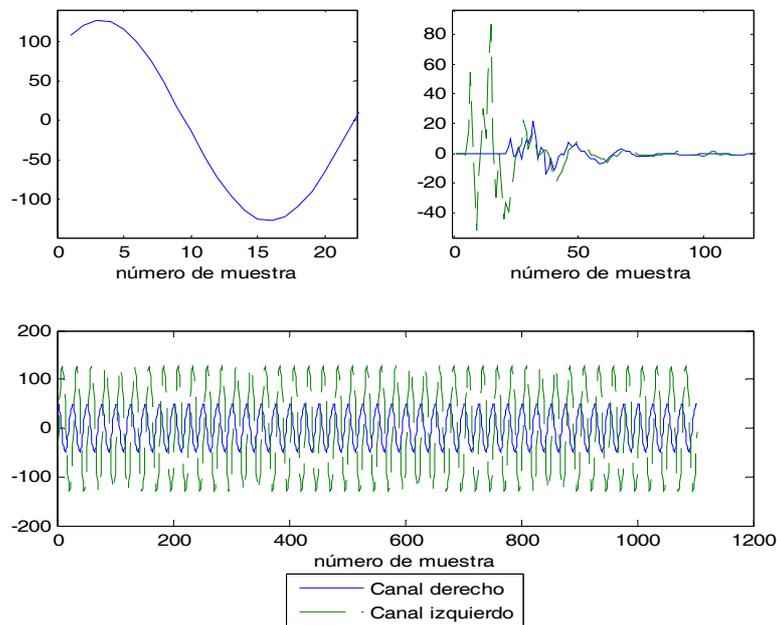


Figura 6.14: Arriba: A la izquierda tenemos parte de la señal original a reproducir, y a la derecha los coeficientes de un filtro (muestreados con 8 bits). Abajo: Resultado de filtrar la señal original con el filtro mostrado.

Esta primera aproximación es capaz de filtrar en tiempo real las 1103 muestras con el filtro almacenado, pero debido a la escasez de recursos no se pueden almacenar más filtros de otras localizaciones espaciales. Para reducir la ocupación del circuito se ha reducido el tamaño en bits de su representación de 16 a 8 y así ahorrar la mitad de espacio. La figura 6.14 nos muestra en la parte superior izquierda la representación de los coeficientes de los filtros, y en la parte inferior el resultado de su convolución con la señal seno.

Teniendo en cuenta todo lo dicho, esta primera versión basa su funcionamiento en un código Handel-C tan sencillo como el siguiente:

```
while (1){
    par{
        Muestra(&i);
        Filtro (sin,i,der,salidader);
        PWM(salidader, right);
        Filtro (sin,i,izq,salidaizq);
        PWM(salidaizq, left);
    }
}
```

donde :

- **i** representa al número de muestra actual.
 - **sin** es el vector que contiene las 1103 muestras de la señal seno.
 - **der** e **izq** son los vectores que contienen las muestras de cada uno de los canales del filtro. Derecha e izquierda respectivamente.
 - **salidader** y **salidaizq** son variables temporales que almacenan el resultado del filtrado. En cada ciclo contienen el resultado del ciclo anterior.
 - **right** y **left** son los buses que están ligados a los terminales de la FPGA que se conectan al módulo de los auriculares.
 - La función **Filtro** simplemente almacena en una variable temporal, pasada por referencia, el resultado de filtrar la muestra número **i** del vector que contiene a la señal seno con uno de los vectores que contiene uno de los canales del filtro.
 - La función **PWM** aplicará codificación PWM para poner en uno de los buses el resultado del filtrado (**salidader** y **salidaizq**).
 - La función **Muestra** temporiza la reproducción según la frecuencia de muestreo y actualiza el número de muestra, **i**, es decir, marca el periodo de reproducción.
-

En resumen, en cada ciclo de reproducción se filtra la muestra actual de los dos canales y el resultado se almacena en dos variables temporales. A la misma vez se reproduce a través de cada una de las salidas el resultado del filtrado del ciclo anterior.

A partir de los resultados observados pasamos a intentar mejorar las características del sistema en una **segunda implementación**. Aún habiendo reducido el espacio que ocupaban los filtros a la mitad, necesitamos aprovechar mejor el espacio para poder almacenar más filtros.

Basándonos en la periodicidad de la señal seno, la nueva versión solo almacenaría un periodo de la misma, en nuestro caso 25 muestras, según la frecuencia de muestreo ya establecida en la primera versión. De esta forma pasamos de las 1103 muestras almacenadas a 25, que son recuperadas siguiendo un esquema de buffer circular.

El código que implementa este tipo de filtrado para cada muestra a partir de un solo periodo se puede ver a continuación:

```
k = (i % lon_sin);
temp = 0;
for(j=0;j<=129;j++){
    par{
        k = ((k<1) ? (24) : (k-1));

        op1 = adjs(entrada[(k)<-5], width(resultado));
        op2 = adjs(filtro[(j)<-7], width(resultado));

        resultado = op1 * op2;

        if (j>2)
            temp += adjs(resultado, width(temp));
        else
            delay;
    }
}
```

donde:

- **entrada** y **filtro** son los vectores que contienen las 25 muestras de la señal seno y los 128 coeficientes de uno de los canales del filtro.
 - **i** es el índice virtual de la muestra actual.
 - **k** es la posición dentro del vector que almacena un solo periodo y que implementa el buffer circular.
-

En este punto de desarrollo, hemos decidido implementar una **tercera versión** que pueda almacenar y reproducir varios filtros para comprobar en hardware el efecto auditivo esperado de localización espacial.

Así que almacenamos en la tarjeta cuatro filtros distintos correspondientes a la elevación 0, y los azimut que podemos ver en la figura 6.15. Este sistema de prueba reproduce cada uno de los resultados de filtrar la señal seno con los cuatro filtros de forma secuencial, creando un efecto de fuente de sonido que se mueve alrededor de la cabeza del usuario siempre a igual distancia, formando una circunferencia (véase figura 6.15). Se ha comprobado que el efecto auditivo resultante era el esperado.

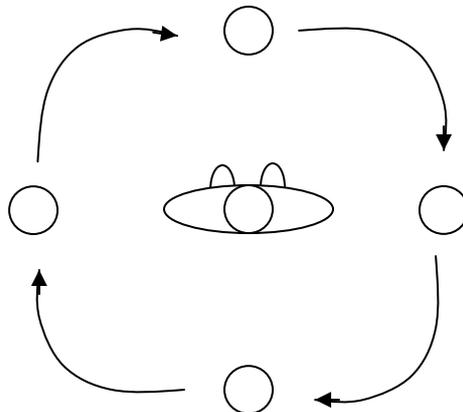


Figura 6.15: Representación del experimento implementado en la tercera versión de sistema.

Basándonos en las propiedades de periodicidad de las señales, y en las propiedades de la operación de filtrado cuando trabajamos con señales periódicas (ver figura 6.16), podemos reducir la complejidad de nuestro sistema almacenando en el circuito el resultado de filtrar las señales previamente. Esta es la novedad de la **cuarta versión**.

Cuando filtramos una señal periódica cualquiera, el resultado es otra señal periódica, con igual periodo que la original. Por tanto, podemos quedarnos con un periodo resultante de filtrar nuestra señal seno con los diferentes filtros de localización, y podemos asegurar que el periodo de estas señales resultantes será de 25 muestras, al igual que la señal original. Esta afirmación puede comprobarse analizando las gráficas presentadas en la figura 6.16.

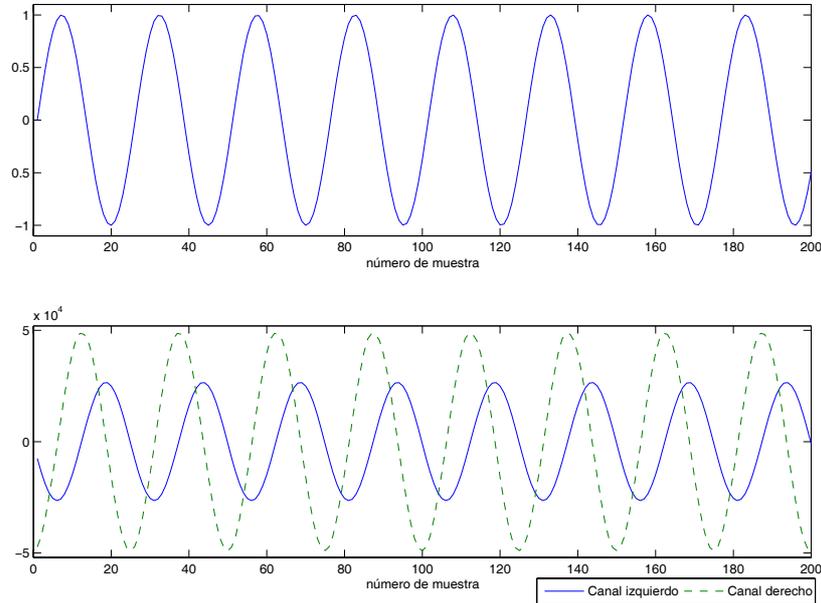


Figura 6.16: Representación de la señal seno (arriba) y el resultado de filtrarla con ambos canales de uno de los filtros (abajo), resultando una señal con el mismo periodo que la original.

A continuación hacemos un estudio sobre los recursos necesarios para cada aproximación descrita.

Debido al reducido número de recursos disponibles en la tarjeta, todas las variantes (salvo la n° 3) se han implementado almacenando los dos canales de un único filtro. También tienen en común la representación de las señales y los filtros como enteros con signo de 8 bits.

La variante 1 consiste en la implementación inicial, en la que se almacenan 1103 muestras de la señal seno y se calcula el filtrado de dicha señal en tiempo de ejecución. La variante 2 difiere de la primera en que sólo se almacena un periodo de la señal a filtrar (25 muestras), reduciéndose los recursos necesarios. La tercera variante se realizó a modo de prueba (a partir de la segunda), e incorpora cuatro filtros de igual elevación y diferente azimut que se reproducen secuencialmente, comprobando la localización espacial. La cuarta ya contempla el almacenamiento del resultado de filtrar previamente el sonido de referencia, y seleccionando un solo periodo de la señal resultado.

La tabla 6.2 recoge las características de los circuitos sintetizados para las distintas variantes de esta primera versión.

Versión	Ocupación (%)		Periodo mínimo (ns)
	Slices	BlockRAM	
1	17	42	41.32
2	21	21	49.29
3	23	64	49.62
4	6	7	21.68

Tabla 6.2: Resultados de síntesis para distintas versiones del circuito en la tarjeta RC-100.

6.3.1.2 Implementación en la RC-200

Como se ha visto en los resultados resumidos en la tabla 6.2, con tan solo un filtro hemos ocupado el 7% de BlockRAM, por lo que deducimos que no existe la posibilidad de almacenar los 191 filtros que hay que incluir en el sistema.

Pos estas razones, decidimos migrar nuestro trabajo a la plataforma RC-200 porque además contiene circuitería específica para reproducción de sonido sin más que especificarle una frecuencia de muestreo y proporcionarle las muestras de la señal. La nueva tarjeta también ofrece la posibilidad de ajustar el volumen de reproducción lo que nos permite dar un efecto de distancia basado en la atenuación en amplitud de la señal. Además se puede configurar el volumen de la salida para cada canal por separado, permitiendo adaptar el sistema a las particularidades de cada usuario.

El proceso de configuración y pruebas se lleva a cabo mediante un software que se ejecuta en PC y, mediante una interfaz de comunicación serie RS-232, se transfiere a la plataforma que contiene la FPGA en la que se implementa el módulo de producción de sonidos 3D.

El objetivo de la **primera versión** es lograr almacenar en la tarjeta un periodo por canal de todas las señales resultantes de filtrar la señal seno con cada uno de los filtros.

En esta primera implementación, el almacenamiento de las señales filtradas se realiza en la RAM embebida que dispone la FPGA. Esta RAM

tiene una capacidad muy limitada, por lo que se ha tenido que elegir para cada muestra una representación con signo de tan sólo 10 bits, aún así no se ha podido almacenar ni el 25% del total de las señales filtradas. El resto de señales se han tenido que guardar utilizando los Slices de la FPGA, lo cual tiene como consecuencia la utilización de prácticamente la totalidad de la FPGA como unidad de almacenamiento, en lugar de unidad de procesamiento. Esto supone una gran limitación a la hora de implementar la funcionalidad que necesitamos para nuestro sistema.

Es importante decir que las señales filtradas se almacenarán en la tarjeta a la vez que se programa la FPGA en una estructura lógica como la de la figura 6.17.

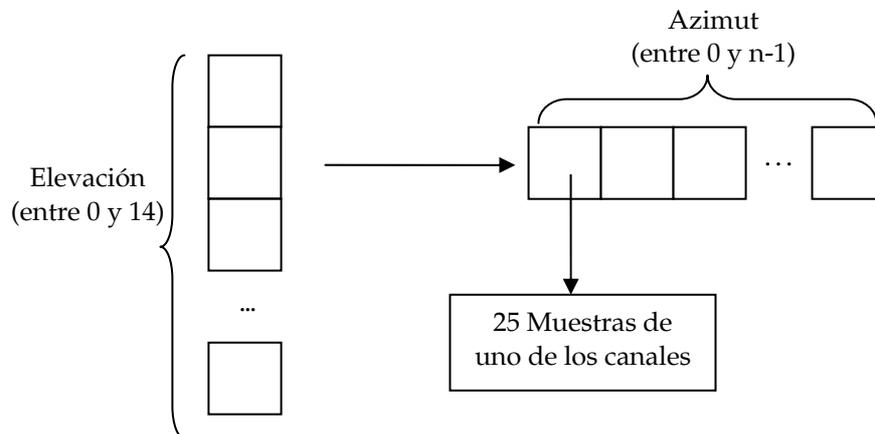


Figura 6.17: Estructura diseñada para almacenar las señales filtradas. Presenta dos niveles de indexación, el primero determina la elevación, y el segundo el azimut.

Necesitaremos dos estructuras de almacenamiento como la de la figura 6.21, una para cada canal auditivo. A partir de esta organización solamente necesitamos 3 parámetros para determinar una señal del conjunto a reproducir: elevación, desplazamiento y hemisferio o lado (recordemos que solo se almacenan las señales procedentes del hemisferio derecho, y si el hemisferio deseado es el izquierdo simplemente se invierten los canales). De modo que con esta disposición podemos acceder a la muestra i de cualquier señal de forma directa y rápida:

`(vector_muestras[elevacion][desplazamiento])[i]`

Para configurar y evaluar el comportamiento del sistema, como primera aproximación, hemos desarrollado un software externo ejecutado

en un PC, y que se comunica con la tarjeta mediante RS232. En cada momento, envía al módulo de sonido los valores de elevación, desplazamiento y hemisferio.

La tarjeta RC-200 cuenta con librerías propias para la implementación de la comunicación aportando funciones para leer y escribir un byte en el puerto serie, así como para configurar el protocolo de transmisión. Realizamos un software en el que se puede elegir mediante una interfaz gráfica qué señal filtrada reproducir (figura 6.18).

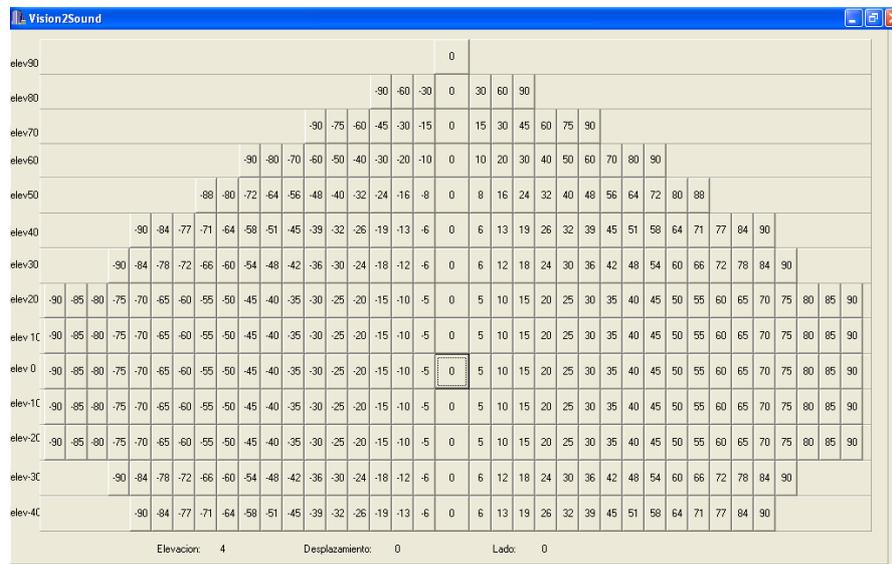


Figura 6.18: Interfaz del entorno software para test del sistema.

El protocolo de comunicación utilizado en esta primera versión es bastante sencillo, consiste en que la tarjeta siempre está en escucha del puerto para leer tres valores consecutivos. Cada valor se corresponde con la elevación, desplazamiento y lado, respectivamente, de la señal a reproducir. En el PC simplemente se escriben en el puerto esos datos correspondientes a la posición seleccionada en el interfaz.

Resumiendo, la funcionalidad del sistema por lo tanto es la siguiente: El sistema recibe por medio del puerto serie tres valores consecutivos – elevación, desplazamiento y lado – y reproduce la señal filtrada correspondiente a esos valores. La duración del sonido es ilimitada, se está reproduciendo la señal hasta recibir una nueva posición.

Tras múltiples pruebas con la versión disponible se detectó la dificultad para localizar un sonido emitido de forma independiente, sin relacionarlo con ningún otro, y consideramos la posibilidad de incluir un sonido de una duración menor, y localizado en una posición de referencia fija, que se reproduzca justo antes del sonido asociado a la posición elegida. Éstas son las características de la **segunda versión**.

Este sonido de referencia es el correspondiente a elevación cero grados y desplazamiento también cero grados, es decir, justo delante del usuario. De este modo, el usuario al escuchar primero un sonido siempre fijo le es más fácil apreciar la diferencia con el sonido siguiente, y por tanto, localizar con más exactitud la procedencia de este último.

El software del PC (figura 6.18) no fue necesario modificarlo para esta versión. En la tarjeta, una vez recibida cada petición de reproducción mediante los tres parámetros comentados, se reproduce el sonido de la posición de referencia durante 0.5s, y después la señal de localización correspondiente.

Los resultados de las pruebas realizadas han sido satisfactorios, incrementándose la capacidad para discriminar la localización espacial del sonido generado por el sistema. Como se comentaba en la primera versión, hemos tenido que utilizar toda la BlockRAM y además SLICES de la FPGA hasta llegar a un 97% de ocupación (ver más adelante la tabla 6.3) para poder almacenar las señales filtradas, y por tanto ahora no disponemos de espacio para aumentar la funcionalidad del sistema.

Para solucionarlo, en la **tercera versión** vamos a almacenar las señales filtradas en la memoria SRAM adicional de la placa, sin que se ocupen recursos internos de la FPGA, añadiendo un procedimiento de transmisión de las mismas desde el PC hasta la memoria disponible en la placa.

El inconveniente que presenta esta implementación es que el almacenamiento se realiza en tiempo de ejecución. Esto quiere decir que no se puede programar la memoria SRAM antes de arrancar el sistema para que contenga las señales filtradas, sino que una vez arrancado el sistema las señales son transferidas mediante comunicación RS232 a la tarjeta y almacenadas en la SRAM. Lógicamente esta operación requiere un tiempo que dependerá del número de señales filtradas a almacenar.

Teniendo presente que la estructura de un banco de la memoria SRAM de la RC-200 consta de 512k palabras de 36 bits, ideamos que la mejor forma de almacenar la señales era situando en cada posición 2

muestras, una de cada canal. Y en posiciones consecutivas de memoria las 25 muestras de cada señal. Esta idea permite que a la hora de reproducir una señal podamos acceder a la vez a las dos muestras que también se reproducirán a la vez, no perdiendo así ciclos en esperar a completar lecturas.

En la tarjeta se establece una estructura lógica con el mismo orden de llegada, resultando la organización de la figura 6.19.

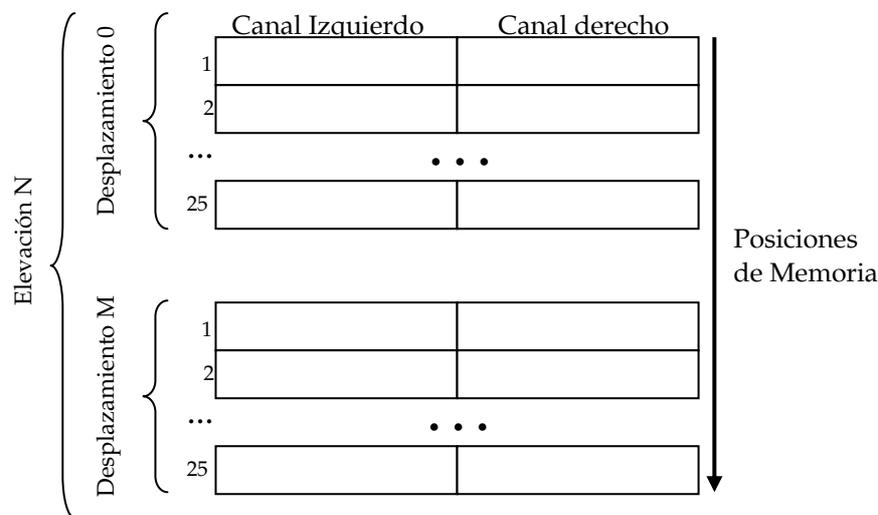


Figura 6.19: Estructura lógica en memoria para el almacenamiento de las señales.

Para acceder a la SRAM, ya sea para escribir como para leer, se necesitan dos ciclos: en un ciclo se establece la dirección de memoria a la que se quiere acceder y en el siguiente ciclo se accede. Aunque estas operaciones se pueden realizar en paralelo, es decir, a la misma vez que leemos de una dirección, establecemos la dirección de la siguiente lectura.

La interfaz para elegir la señal a reproducir correspondiente a una posición es la misma de la versión anterior (mostrada en la figura 6.18). Es decir, se recibirán los 3 parámetros necesarios – elevación, desplazamiento y lado – y a partir de estos valores habrá que acceder a memoria a la posición adecuada.

Como para todas las elevaciones no existen el mismo número de desplazamientos, no todas ellas ocupan el mismo número de posiciones en memoria. Así que calculamos las direcciones de comienzo de cada elevación y las almacenamos en un vector. De esta forma, para acceder a la

primera dirección de memoria donde se encuentra almacenada una posición cualquiera, bastará sumar a la dirección de comienzo de la elevación seleccionada el número de desplazamiento multiplicado por 25. Esto queda resumido en la siguiente fórmula:

$$\text{Dirección_comienzo} = \text{comienzo}[\text{elevación}] + \text{desplazamiento} * 25$$

donde **comienzo** es el vector que almacena las direcciones de comienzo para cada elevación.

En esta versión se ha almacenado el sonido de referencia en la BlockRAM y además, los coeficientes de los canales auditivos son iguales, con lo que sólo es necesario almacenar uno de ellos y enviar la misma muestra a reproducir.

Como hemos ahorrado bastante espacio alojando las señales en la SRAM se han implementado nuevas funcionalidades en nuestro sistema, como por ejemplo, cambiar la duración de las señales, tanto de las relacionadas con una posición como la del sonido de referencia, así como el tiempo de silencio entre una reproducción y la siguiente. En detalle estos tiempos son:

- Tiempo de reproducción total: Determina el tiempo durante el que se reproduce el sonido, en milisegundos. Hay una opción especial que si se activa, el sonido se reproducirá sin limitación de tiempo.
 - Tiempo de reproducción del sonido de una posición: Determina la duración que tendrá la reproducción de la señal filtrada, en milisegundos.
 - Tiempo de reproducción de la señal de referencia: tiempo en milisegundos que dura la reproducción de la señal de la posición de referencia.
 - Tiempo de silencio: tiempo de espera entre la finalización de la reproducción de la señal filtrada y el comienzo de una nueva reproducción de la señal de referencia.
 - Periodo de repetición de la señal. Este parámetro no es configurable, sino que se obtiene como la suma de la duración de la señal del filtro, la duración de la señal de referencia y el intervalo sin sonido.
-

La figura 6.24 muestra un esquema de la relación que guardan los distintos tiempos.

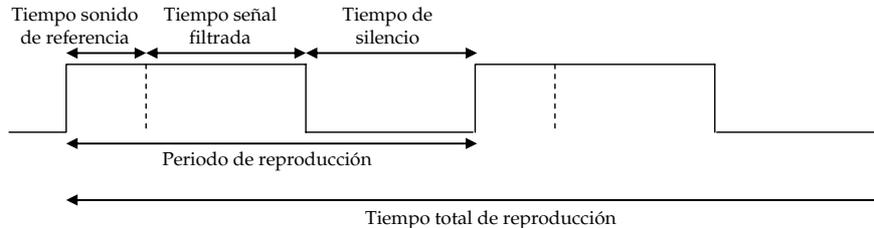


Figura 6.20: Representación de las distintas duraciones configurables.

La funcionalidad más importante que introducimos es la posibilidad de crear un espacio de posiciones distinto del hasta ahora utilizado y totalmente configurable, es decir, que podamos asociar a cada nueva posición la señal filtrada deseada. El nuevo espacio de posiciones es de forma rectangular con igual número de desplazamientos para todas las elevaciones. A este tipo de espacio de posiciones en adelante lo llamaremos malla. Y cada malla se corresponde en realidad con la resolución del campo visual con el que se dota al sistema global. Por supuesto, se dará la posibilidad de en todo momento configurar dicha malla, con un número de filas (elevaciones) y columnas (desplazamientos) deseado.

Como acabamos de decir, a cada posición de la malla se podrá asignar una señal cualquiera del repertorio, sea cual sea su elevación y desplazamiento, incluso su lado. La intención es hacer el sistema lo máximo configurable posible, ya que, un usuario puede considerar en su nueva malla un filtro arbitrario para cada una de las posiciones (identificado por sus parámetros elevación, azimut y hemisferio).

Para poder configurar la tarjeta con todos estos parámetros en el PC se implementó un software que permite crear una malla de forma rápida y sencilla, así como asignarle a cada posición de ésta una señal cualquiera del repertorio. Por supuesto, es necesario transmitir todas estas configuraciones a la tarjeta, por lo que sobre RS232 fue necesario crear un protocolo de configuración.

Para poder realizar todas las tareas del sistema, en la plataforma hardware definimos dos modos de funcionamiento:

- Modo normal: el sistema se encarga de reproducir la señal filtrada cuyos parámetros de elevación, azimut y lado recibe por el puerto serie.

- Modo malla: el sistema debe de tener almacenada una malla, y reproduce la señal filtrada cuyos parámetros de elevación, azimut y lado se obtendrán de una posición de la malla previamente definida.

Para poder configurar todos los parámetros, la tarjeta siempre se encontrará a la espera de recibir comandos. Si se reconoce el comando recibido como válido se devuelve una confirmación positiva, y si por el contrario no se reconoció el comando, o no se puede ejecutar, se contesta con una confirmación negativa.

Los distintos comandos que forman parte del protocolo que hemos definido son los siguientes:

- *SETDURACIONREF*: configura la duración de la señal de referencia.
 - *SETDURACIONPITIDO*: establece la duración de la señal con localización espacial.
 - *SETDURACIONNADA*: modifica la duración del intervalo en el que se mantiene la salida en silencio.
 - *SETDURACIONTOTAL*: permite cambiar la duración del periodo de reproducción.
 - *SETMODO*: cambia el modo de funcionamiento del sistema entre modo normal y modo malla.
 - *PLAY*: ordena a la plataforma que reproduzca el sonido cuyos parámetros tenga definidos.
 - *STOP*: detiene la reproducción de sonidos.
 - *CONFIG*: establece los parámetros de una señal espacial (elevación, desplazamiento y hemisferio).
 - *SETDESP*: envía el parámetro de azimut de una señal.
 - *SETELEV*: envía el parámetro de elevación de una señal.
 - *SETLADO*: envía el parámetro de hemisferio de una señal.
 - *GETMALLACONFIGURADA*: pregunta si la malla ha sido configurada o no en hardware.
 - *SETFILTROS*: indica el comienzo del envío de todas las señales filtradas que se deben almacenar en la SRAM.
 - *SETFILTROSACTIVADO*: sondea a la tarjeta para saber si tiene almacenadas las señales filtradas en memoria.
 - *SETMALLA*: envía la configuración de la malla: tamaño y los parámetros de las señales asociadas a cada posición de la misma.
 - *SETPOSICIONMALLA*: comunica la posición de la malla a reproducir.
-

Para implementar la malla, en los SLICES de la FPGA tenemos almacenada una matriz de tamaño igual a la máxima malla posible (10x25). Cada componente de la matriz estará formado por una estructura donde se almacenará la elevación, el desplazamiento y el hemisferio correspondiente a una determinada señal de posición.

Se ha modificado el software que se ejecuta en el PC, para que realice todas las funciones antes descritas. Pero haremos una descripción más exhaustiva cuando hablemos de la cuarta versión que contendrá éstas y otras funcionalidades más.

En la **cuarta versión** se ha afrontado el tema del volumen y cómo poder obtener la configuración de malla que almacena la tarjeta y transferirla al PC. Otra nueva función en el PC, será poder guardar configuraciones en archivos, y poder recuperarlas desde el mismo entorno posteriormente.

Pero se plantea el problema de que no podemos seguir desarrollando sobre ella ya que tenemos ocupado el 99% de los SLICES, como se puede ver en la tabla 6.3. Esta elevada ocupación se debe a que el almacenamiento de la malla, que si recordamos era una matriz de struct, se ha hecho mediante SLICES. No se almacenó en BlockRAM porque no se admite en memoria matrices de struct. Está claro pues que es necesario cambiar la estrategia de almacenamiento de la malla para poder seguir con la implementación del sistema.

En esta nueva versión la matriz que almacena la asignación de señales contiene un valor de 14 bits. El bit más significativo contiene el hemisferio de la señal filtrada y el resto de bits contiene la dirección en memoria SRAM de la señal filtrada. Pero lo más importante es que la matriz se ha almacenado en los bloques de la BlockRAM, lo que nos ha permitido aumentar el tamaño máximo de la matriz a 25x25, a la vez que liberar recursos de la FPGA.

La configuración de la matriz se ha hecho calculando para cada posición de la malla la dirección de la señal que se quiere asignar (el cálculo es el mismo que se hace para acceder a la memoria SRAM) y concatenando dicha dirección con el hemisferio de la señal filtrada. Al igual que en la anterior aproximación, la información de qué señal filtrada le corresponde a cada posición de la malla se la comunicará el PC a la tarjeta con el comando *SETMALLA*.

Las nuevas funcionalidades que incluye esta versión son las siguientes:

- Volumen de cada canal: se puede modificar el volumen de cada canal (izquierdo y derecho) de forma independiente, con el objetivo de corregir posibles deficiencias en los oídos del usuario (ver figura 6.21).
- Obtener configuración: permite recuperar la configuración de la malla que haya en la tarjeta y llevarla hasta el entorno software, para poder trabajar con ella.
- Guardar configuración: almacena en un fichero la configuración de malla que tenemos en el entorno software.
- Cargar configuración: cargar al entorno software una configuración desde un archivo.

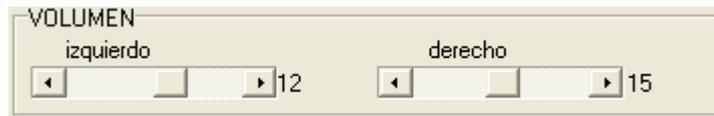


Figura 6.21: Detalle de la interfaz para configurar el volumen de los canales de audio.

Se han añadido nuevos comandos de configuración posibles para el sistema como son:

- *SETVOLUMEN*: modifica el volumen de reproducción de los canales auditivos.
- *GETCONFIGURACION*: solicita a la plataforma hardware todos los parámetros de configuración que tenga establecidos.

En la fase de arranque, después de almacenar las señales filtradas en la SRAM, si se detecta que la malla está configurada en la tarjeta, el entorno se actualizará automáticamente con esta configuración.

Los pasos seguidos por la herramienta en su funcionamiento son los siguientes:

1) Fase de arranque

Al ejecutar el software tendremos una interfaz como la de la figura 6.22, que nos permite elegir el puerto serie de salida del PC (COM1, COM2 o COM3).



Figura 6.22: Entorno software para la conexión y transmisión de señales.

Lo primero que hace el software es intentar conectar con la tarjeta para comprobar si están almacenadas las señales filtradas en la memoria SRAM del sistema, las cuales son indispensables para llevar a cabo la configuración. Si se recibe confirmación positiva, quiere decir que las señales ya están almacenadas y no tenemos que volver a enviarlas.

A continuación, se comprueba si la malla ya está configurada. En caso negativo se pasa a la pantalla de configuración, con los parámetros de configuración puestos a sus valores por defecto y se muestra en pantalla una malla sin configurar, como la de la figura 6.23. Si la confirmación es positiva, se accede a la configuración de la tarjeta y se obtienen todos los parámetros de configuración, y se actualiza el software para que muestre la malla en pantalla tal y como está configurada en la tarjeta.

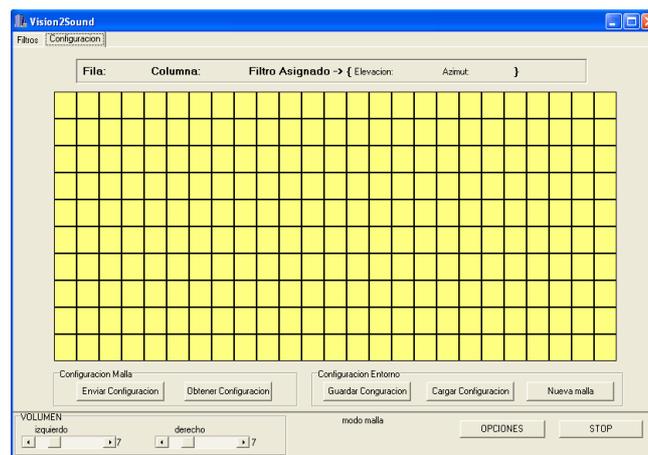


Figura 6.23: Detalle del interfaz con una malla sin configurar.

2) Procedimiento para la configuración del sistema

Como ya se ha comentado, en caso de que la malla no estuviese configurada en la tarjeta, después de la fase de arranque aparecería una pantalla como la que se muestra en la figura 6.23 con una malla de tamaño 10x25. A continuación vamos a ver los pasos que hay que seguir para configurar la malla en el entorno software y, una vez configurada, enviarla a la tarjeta.

El primer paso sería elegir el tamaño de malla que se desea, (el número de filas debe de estar comprendido entre 1 y 25, y el de columnas igual). Una vez que tenemos la malla con la resolución deseada, el próximo paso es asignarle una señal a las posiciones que queramos de la malla. Para ellos se hace doble click sobre la posición a la que se le quiere asignar una señal y aparecerá la pantalla con el repertorio completo de señales de donde seleccionaremos la señal deseada.

Una vez que hemos asignado una señal de posición a una casilla de la malla volveremos a la pantalla de configuración de malla y la posición a la que se le ha asignado el filtro aparecerá en color rojo, además si pinchamos sobre la posición con solo un clic aparecerán los parámetros del filtro que tiene asignado. Cuando tenemos hecha la asignación de filtros a las posiciones que queremos se envía la configuración de la malla a la plataforma hardware.

Cuando hay una configuración de malla en el entorno software, ésta puede ser modificada cambiando la asignación de señales a las posiciones que queramos. Para que la modificación sea válida en la tarjeta hay que actualizar de nuevo los datos en el circuito.

La tabla 6.3 presenta los resultados de síntesis de los distintos circuitos realizados en esta plataforma hardware:

Versión	Ocupación (%)		Periodo mínimo (ns)
	Slices	BlockRAM	
1	94	97	29.71
2	94	97	32.92
3	99	2	27.18
4	69	65	40.65

Tabla 6.3: Resultados de síntesis de las distintas versiones implementadas en la RC-200.

Como se puede ver los recursos ocupados en las dos primeras versiones son elevados debido a que las señales filtradas se almacenan en la FPGA. A partir de la versión 3, las señales filtradas se guardan en la memoria SRAM externa permitiendo reducir considerablemente los recursos y destinarlos a otros fines. En las dos primeras versiones las señales están representadas con 10 bits, mientras que en la tercera y cuarta se han utilizado 16 bits, esto no influye en la ocupación ya que en estas dos últimas se usó la SRAM para almacenar las señales.

También se puede observar que el número de SLICES utilizados en la versión 3 es muy elevado. Esto se debe a que en esta versión se incluye el protocolo de configuración, pero la principal razón es que se almacena en los SLICES la matriz que representa en la tarjeta la malla del campo visual. Esto se resuelve en la versión 4 almacenando la matriz comentada en la memoria RAM embebida, a esto se debe la subida del porcentaje de utilización de los bloques de memoria RAM embebida reflejada en la tabla.

Respecto al comportamiento temporal, se observa que en las tres primeras versiones los retardos son muy parecidos y relativamente bajos, en cambio, en la versión 4 el retardo empeora mucho. Esto se debe a que la versión 4 incorpora en el protocolo de comunicación unas opciones para obtener la configuración de la tarjeta y para implementar estas opciones se han tenido que utilizar divisiones, que son las operaciones que más retardo provocan.

6.3.2 Integración del sistema final

En este último apartado nos ocupamos del proceso de integración de un sistema de procesamiento visual con el de generación de sonidos 3D con la finalidad de construir un sistema de ayuda para invidentes basado en la transducción sensorial.

En primer lugar, hemos decidido ampliar el sistema descrito en la sección 6.1 para que incluya otras características como el procesamiento estéreo. En concreto, el tipo de procesamiento implementado sigue la arquitectura mostrada en la parte inferior de la figura 4.11. Si recordamos, se basa en el uso de dos canales de entrada de video, cada uno de los cuales es procesado por separado mediante estructuras de cómputo similares, hasta conseguir la matriz de actividad de ambos.

A partir de este punto se hace una estimación de la profundidad tomando como un par estéreo las dos matrices de actividad obtenidas. Al tratarse de imágenes de una resolución reducida, el tiempo de cómputo de las disparidades es mínimo, y podemos hacer una búsqueda en el rango máximo de disparidades (que se corresponde con el número de columnas presentes en la matriz de actividad).

Se ha decidido incluir un nuevo elemento hardware con la finalidad de asistir a nuestro sistema en varias tareas. Se trata de un sensor de *ultrasonidos* situado en la zona central entre las dos cámaras (como puede verse en la figura 6.24).



Figura 6.24: Dispositivo de captura de imágenes utilizado en VISION2SOUND. En la zona central podemos distinguir el sensor de ultrasonidos incorporado.

Este componente nos suministra información complementaria del entorno, que es de vital importancia cuando nos enfrentamos a escenas que presentan zonas homogéneas y no son detectadas por los bancos de filtrado espacio-temporal. Su uso nos puede ayudar a aislar o ponderar zonas que se encuentren a menos de una distancia determinada, además de permitirnos implementar una adaptación dinámica a las condiciones del entorno.

Otra posibilidad que podemos contemplar es la intentar imitar otro tipo de características del sistema visual humano como la convergencia o fijación de los ojos en un punto determinado. La forma de implementarlo en nuestro sistema sería fijar un punto situado a la distancia que nos informa el módulo de ultrasonidos y simular la convergencia ocular con desplazamientos de las imágenes de ambas cámaras, de forma que anulemos las disparidades a esa distancia, y calcular la distancia relativa a este punto en base a los nuevos valores de disparidad.

Con la finalidad de poder integrar ambos sistemas, y debido a la necesidad de capturar imágenes de dos canales de video, decidimos realizar todo el trabajo en la plataforma RC-300 de Celoxica, descrita en la sección 6.1.1.3, que tienes 6 veces más recursos que la RC-200 y ya incluye la circuitería para conectar dos cámaras.

También se ha realizado una mejora de las etapas de procesamiento que se han implementado. Si bien la combinación de filtros implementada es la misma que se describió en la sección 6.1.3.4, los módulos de convolución se han extendido para trabajar con máscaras de tamaño 7x7, lo que mejora los resultados de este proceso. Por el contrario, se ha decidido una reducción de tamaño de las matrices de actividad, quedando finalmente con 12 filas y 15 columnas (aunque son dos parámetros que podemos modificar si lo deseamos y reconfigurar el circuito).

A partir de la información aportada por las matrices de actividad, el mapa de disparidades calculado a partir de ellas, y la información sobre la distancia que nos dice el sensor de ultrasonidos, se construye una figura de información que incluye como caso especial la detección por parte del módulo de ultrasonidos de algún objeto a una distancia menor de 40 cm, en la que informaremos al usuario de la existencia de un obstáculo en la zona central de su campo visual.

A partir de esta figura de información construida, en nuestra aproximación, hemos decidido tomar la posición del máximo valor y enviar su fila, columna y disparidad medida al sistema de generación de sonidos (siempre que supere un determinado umbral definido por el usuario) para que reproduzca el sonido configurado previamente como hemos descrito en la sección 6.3.1.2. Cabe mencionar que se ha realizado una modificación al sistema de generación de sonidos que elimina la necesidad de transferir por software todos los filtros hacia el circuito, de forma que se almacenan una sola vez en la tarjeta de memoria del tipo SmartMedia™ incluida en la plataforma.

La figura 6.26 nos muestra el aspecto final de la plataforma en la que hemos implementado el sistema completo:

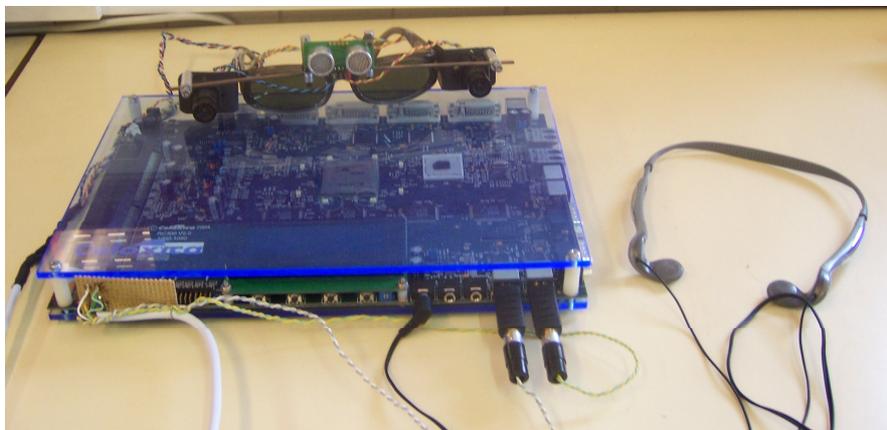


Figura 6.26: Prototipo del sistema de transducción sensorial implementado en una plataforma portable.

Para completar la funcionalidad y permitir la configuración parcial del sistema durante su utilización, se ha implementado un sistema menú visible mediante una pantalla LCD que incorpora la tarjeta, y mediante los botones que tiene se pueden modificar algunos valores. En concreto podemos seleccionar si se reproduce o no tanto el sonido de referencia como el sonido posicionado, por si en un momento determinado queremos dejar de recibir información del sistema. Para solventar el problema de diferencia de ganancia de los sensores de imagen que incorporan las cámaras, podemos ponderar los valores capturados por las mismas por un factor entre 0 y 1 con una resolución de 7 bits. Y finalmente, podemos modificar dos parámetros más como son el volumen global de reproducción de los sonidos y el umbral utilizado para generar sonidos 3D.

Si hablamos de los resultados de ocupación del circuito que generan la versión completa, vemos que son necesarios el 70% de los SLICES, el 18% del total de BlockRAM disponibles, el 8% de los multiplicadores embebidos en la FPGA (que permiten multiplicar dos valores de 18 bits). El total de puertas equivalentes que calcula la herramienta de síntesis se eleva aproximadamente a 2200000. En cuanto a las características temporales, el circuito puede funcionar a una frecuencia de 40MHz. Aunque la tasa de imágenes por segundo que podemos procesar es superior a 60, no tiene sentido procesar más del número de imágenes que determina la duración de los sonidos reproducidos por el sistema de localización espacial. Si suponemos una duración de 0.1 s, no hace falta procesar más de 10 imágenes por segundo.

6.4 Conclusiones

En este capítulo hemos descrito en primer lugar todos los aspectos relacionados con la implementación en chips reconfigurables de los distintos módulos de procesamiento propuestos en los capítulos anteriores. En todos los casos, tras un proceso de adaptación de los cálculos y los tipos de datos, hemos llegado a soluciones integradas en FPGA cuyas velocidades de procesamiento y latencia son adecuadas para ser utilizadas por el individuo incluso en movimiento.

En primer lugar hemos introducido el diseño en hardware del sistema de procesamiento visual basado en el filtrado espacial de las imágenes de entrada con máscaras de realce del contraste de color. También incorpora la reducción de resolución basada en la definición de campos receptivos.

Además hemos presentado la posible fusión de este sistema de procesamiento biospirado con un sistema de generación de eventos asociados a diferentes canales de estimulación en el ámbito de las neuroprótesis visuales.

Finalmente hemos descrito la implementación hardware del sistema de localización espacial de sonidos descrito en el capítulo 5, así como la posibilidad de integrarlo con un sistema mejorado de procesamiento visual que incluye estereovisión y es asistido por un sensor de ultrasonidos. Esta aproximación constituye una nueva propuesta de sistema de ayuda para invidentes, con marcadas diferencias de uso frente a las aproximaciones propuestas con anterioridad [Nielsen 1989], [EAV], [vOICe].

Conclusiones y principales aportaciones

En los distintos capítulos de esta memoria se recoge todo el trabajo de tesis doctoral realizado en el marco de los sistemas de procesamiento de información orientados a los sistemas de ayuda para invidentes. Los resultados presentados se pueden englobar en dos vertientes, los relacionados con los sistemas de rehabilitación visual mediante neuroestimulación, y los destinados a los sistemas electrónico-informáticos de transducción sensorial, como alternativa no invasiva (sin necesidad de cirugía).

En todos los sistemas de este tipo se imponen las necesidades de portabilidad y de funcionamiento en tiempo real, entendiendo por tiempo real el necesario para evocar percepciones visuales continuas, además de un desarrollo totalmente asistido por los pacientes con deficiencia visual.

Cuando nos referimos a las neuroprótesis visuales, uno de los grandes retos es llevar a cabo una codificación neuromórfica de los estímulos a aplicar, de manera que permitan la percepción de formas y de movimiento en lugar de fosfenos aislados y separados en el tiempo. Respecto a los patrones generados, es deseable que sean específicos, significativos y fáciles de interpretar por el paciente, sin que el proceso de entrenamiento-aprendizaje presente muchas dificultades.

Para que estos sistemas puedan extraer la información más relevante, consideramos que deben estar inspirados en las características perceptivas del sistema visual humano. En este sentido, consideramos que deben atender a aspectos como el procesamiento basado en imágenes en **color**, la detección de **cambios** temporales en las secuencias de imágenes capturadas, la extracción de información sobre la **profundidad** en base a procedimientos que combinen medidas absolutas con medidas relativas que permitan modelar el funcionamiento del sistema visual humano. Éste extrae dicha información de la convergencia ocular y las neuronas sintonizadas por la disparidad.

Entre las aportaciones de esta tesis figuran algunos de los principales módulos que componen Retiner, un entorno de diseño de sistemas de procesamiento visuales, cuyas capacidades se extienden desde el modelado y la simulación de los mismos, hasta su ajuste y comparación de resultados con datos biológicos obtenidos en experimentos de registro de actividad neuronal. En concreto, se ha trabajado en los módulos de entrada del sistema, tanto en la extensión de la captura desde una cámara a dos cámaras haciendo uso del toolbox de adquisición de imágenes de Matlab™, como en el que permite procesar secuencias previamente almacenadas en archivos de vídeo en formato AVI. Esta característica se hizo necesaria cuando el tiempo de respuesta en el procesamiento on-line, con captura de cámara, creció por el aumento en la complejidad de los sistemas modelados.

El módulo de cálculo de disparidades y estimación de la profundidad es una de las principales contribuciones de esta tesis al sistema completo. Los resultados obtenidos en esta etapa nos permiten mejorar el proceso de selección de información visual, incorporando otra fuente de información que está presente en el sistema visual humano. Otro módulo que ha requerido nuestro esfuerzo ha sido el que permite procesar archivos de vídeo con una configuración temporal personalizada por el usuario, afectando también al filtrado temporal, para el que se permite seleccionar un número variable de imágenes previas con las que computar las diferencias. La resolución temporal configurable nos permite por un lado, afinar el modelo incluso a los tiempos de respuesta de las neuronas observados en experimentos biológicos, y por otro, eliminar las discontinuidades en el procesamiento de imágenes en vivo capturadas directamente de la cámara.

También recogemos entre las aportaciones propias el módulo software encargado de la definición de los distintos tipos de campos receptivos (junto con el formato del archivo de salida de esta información, que permite su modificación manual a gusto del usuario). Gracias a este

módulo, Retiner nos permite modelar la diferencia de resolución dependiente de la excentricidad que se observa en la retina humana, además de las diferentes formas de los campos receptivos que presentan otras especies animales. Al mismo tiempo provee a los sistemas de ayuda a la baja visión de la posibilidad de definir transformaciones de la imagen de entrada al objeto de aprovechar el resto visual de los pacientes, permitiendo la derivación de información visual hacia las zonas del campo visual que conservan su funcionalidad.

Nuestra contribución también se extiende hasta la generación de la matriz de actividad, teniendo en cuenta los campos receptivos previamente determinados, y de la salida graduada, que pueden ser utilizadas mediante visualizadores para la realización de experimentos psicofísicos. Estos experimentos, realizados con personas que presentan visión normal, nos van a permitir determinar la capacidad visual máxima que se va a poder recuperar, para una determinada configuración del sistema, así como valorar qué operaciones de filtrado mejoran la capacidad de percepción para una resolución espacial dada.

Además hemos descrito la forma en que la generación de sonidos localizados espacialmente extiende a nuestro sistema de procesamiento visual bioinspirado, Retiner. Esta extensión está orientada a los sistemas de rehabilitación visual basados en la transducción sensorial, usando como medio de transmisión de la información el sistema auditivo. Cabe destacar el alto grado de adaptabilidad a cada paciente que permite la implementación realizada, así como su adecuación a la implementación en hardware configurable (FPGA). El objetivo principal de esta implementación software del modelo responde a las necesidades de evaluar la generación de sonidos 3D y llevar a cabo el ajuste / configuración de todos los parámetros para cada sujeto.

La efectividad de un sistema de conversión de imágenes a sonidos 3D va a depender en gran medida de la calidad y selectividad del sistema de procesamiento visual. Los patrones sonoros 3D que proponemos utilizar, al ser captados por el invidente interfieren mínimamente sus capacidades de percepción auditiva, preservando así una de las mayores fuentes de información sensorial de que dispone.

Finalmente hemos descrito todos los aspectos relacionados con la implementación en circuitería reconfigurable de un sistema de procesamiento visual basado en el filtrado espacial de las imágenes de entrada con máscaras de realce del contraste de color, que además incorpora la reducción de resolución basada en la definición de campos receptivos.

Como dos soluciones posibles que conforman un sistema completo, hemos presentado la posible fusión de este sistema de procesamiento bioinspirado con un sistema de generación de eventos asociados a diferentes canales de estimulación en el ámbito de las neuroprótesis, y hemos descrito la implementación hardware de un sistema de localización espacial de sonidos integrado con un sistema mejorado de procesamiento visual que incluye estereovisión y es asistido por un sensor de ultrasonidos.

Más concretamente, el trabajo detallado en esta memoria, presenta las siguientes aportaciones:

1. Se ha definido un modelo de procesamiento de información visual acorde con una arquitectura que abstrae las características funcionales de distintos elementos del sistema visual, tales como el realce espacial y temporal locales, el contraste de color, las diferencias de comportamiento centro-periferia de la retina y la incorporación de otro tipo de información como la profundidad. La arquitectura consiste en una combinación configurable de filtros, cuya salida es codificada en impulsos siguiendo un esquema 'íntegra y dispara'.
 2. En colaboración con el resto de personal investigador del proyecto CORTIVIS se ha implementado el entorno software Retiner, una aplicación definida sobre Matlab™, que permite la especificación y simulación de modelos de procesamiento visual con un elevado número de parámetros, así como la comparación de los resultados de simulación con registros obtenidos de experimentación con animales.
 3. Se ha modelado un sistema para la generación de señales acústicas que incorporen información sobre localización espacial, que nos permite evocar en el sujeto que las perciba la sensación de que dicho sonido procede de una posición tridimensional arbitraria.
 4. Se han desarrollado en lenguaje de descripción de hardware (HDL) de alto nivel la codificación de los distintos módulos componentes del sistema de procesamiento de información visual que incorpora el entorno software Retiner.
-

5. Se ha propuesto la integración de diversos módulos hardware de procesamiento visual con un módulo de generación de eventos asociados a distintos elementos para estimulación, que puede ser incorporado a sistemas completos que incluyan una neuroprótesis como plataforma de estimulación.
6. Se ha diseñado un sistema completo de ayuda a invidentes, basado en la transducción sensorial, que transforma la información visual del entorno en patrones sonoros que incluyen información sobre la localización de los elementos más relevantes en el entorno del sujeto, con las máximas posibilidades para su particularización a nivel personal.
7. Se ha llevado a cabo la validación experimental de los distintos modelos e implementaciones propuestas, tanto en software como en plataformas hardware basadas en FPGA, demostrando la viabilidad de estos sistemas para su aplicación en el campo de la rehabilitación visual.

A lo largo de la realización de todo el trabajo presentado en esta memoria de tesis doctoral, y en relación a los resultados y aportaciones, se han generado, total o parcialmente, las siguientes publicaciones:

- Comunicaciones a congresos Nacionales e Internacionales:

- C. Morillas, S. Morillas, E. Ros, A.F. Díaz, F.J. Pelayo, Hardware reconfigurable para movimiento guiado por visión. Actas de las JCRA 2002, "II Jornadas sobre Computación Reconfigurable y Aplicaciones", Almuñécar – Granada (España), 18-20 Septiembre 2002. ISBN: 84-699-9448-4, pp. 245-248.
 - C. Morillas, S. Morillas, E. Ros, A.F. Díaz, B. Pino, F.J. Pelayo, Real-time vision guided movement with reconfigurable Hardware. Actas del Dynamic Perception, University Bochum, Ruhr (Alemania), Noviembre 2002. ISBN: 3-89838-032-7; ISBN: 1-58603-305-0, pp. 53-58.
 - F.J. Pelayo, A. Martínez, S. Romero, C. Morillas, E. Ros, E. Fernández, Cortical Visual Neuro-prosthesis for the Blind: Retina-like Software/Hardware Preprocessor. Actas de "The 1st International IEEE EMBS Conference on Neural Engineering", Isla de Capri (Italia), Marzo 2003. ISBN: 0-7803-7819-9.
-

- F.J. Pelayo, A. Martínez, C. Morillas, S. Romero, L. Sousa, P. Tomás, Retina-Like Processing and Coding Platform for Cortical Neuro-stimulation. Actas de "25th Annual International Conference of the IEEE EMBS", Cancún (México), 17-21 de Septiembre de 2003. ISBN: 0-7803-7790-7.
 - F.J. Pelayo, C. Morillas, A. Martínez, S. Romero, E. Ros, B. Pino, A reconfigurable machine to model the first stages of the human visual pathway. Actas de las JCRA 2003, "III Jornadas sobre Computación Reconfigurable y Aplicaciones", Madrid (España), 10-12 de Septiembre de 2003. ISBN: 84-60099-28-8, pp. 173-182.
 - A. Martínez, F.J. Pelayo, C. Morillas, S. Romero, R. Carrillo, B. Pino, Generador automático de sistemas bioinspirados de visión en hardware reconfigurable. Actas de las JCRA 2004, "IV Jornadas sobre Computación Reconfigurable y Aplicaciones", Barcelona (España), 13-16 de Septiembre de 2004. ISBN: 84-688-7667-4, pp. 597-603.
 - S. Romero, C. Morillas, A. Martínez, F. Pelayo, E. Fernández, A Research Platform for Visual Neuroprostheses. Actas del SICO 2005 (CEDI 2005), "Simposio de Inteligencia Computacional" ("I Congreso Español de Informática"), Granada, 13-16 de Septiembre de 2005. ISBN: 84-9732-444-7, pp. 357-362.
 - Antonio Martínez, Francisco J. Pelayo, Christian A. Morillas, Eduardo Ros, Leonardo M. Reyneri, Samuel Romero, Automatic synthesis of vision processors using reconfigurable hardware. Actas del JCRA 2005 (CEDI 2005), "V Jornadas de Computación Reconfigurable y Aplicaciones" ("I Congreso Español de Informática"), Granada, 13-16 de Septiembre de 2005. ISBN: 84-9732-439-0, pp. 179-184.
 - Christian A. Morillas Gutiérrez, Juan P. Cobos Carrascosa, Luís F. Castillo Tarifa, Rodrigo Agís Melero, Francisco J. Pelayo Valle, Módulo de sonido 3D en hardware reconfigurable. Actas del JCRA 2006, "VI Jornadas de Computación Reconfigurable y Aplicaciones", Cáceres, 12-14 de Septiembre de 2006.
 - Rodrigo Agís Melero, Francisco F. Franco Ramón, Francisco J. Pelayo Valle, Eduardo Ros Vidal, Christian A. Morillas Gutiérrez, WellMerit: Robot experimental con visión estéreo. Actas del JCRA
-

2006, "VI Jornadas de Computación Reconfigurable y Aplicaciones", Cáceres, 12-14 de Septiembre de 2006.

- S. Romero, C. Morillas, A. Martínez, F.J. Pelayo, E. Fernández, Models and Tools for testing visual neuroprostheses. Presentado en el FENS 2004, "4th Forum of European NeuroScience", Lisboa (Portugal), 10-14 de Julio de 2004. FENS Abstr., vol.2, A085.17, 2004.

- E. Fernández, C. Morillas, S. Romero, A. Martínez, F. Pelayo, Neuroengineering Tools for the Design and Test of Visual Neuroprostheses. Presentado en el ARVO 2005, "2005 Annual Meeting of The Association for Research in Vision and Ophthalmology", Fort Lauderdale – Florida (EEUU), 1-5 de Mayo de 2005. Publicado en un número electrónico de Investigative Ophthalmology & Visual Science, Volumen 46 (Mayo 2005), E-Abstract 1483, ISSN: 1552-5783.

- Revistas y otras publicaciones con índice de impacto (SCI):

- F.J. Pelayo, S. Romero, C. Morillas, A. Martínez, E. Ros, E. Fernández, Translating Image Sequences into Spike Patterns for Cortical Neuro-stimulation. Neurocomputing, Volumen 58-60 (Junio 2004), Computational Neuroscience: Trends in Research 2004, pp. 885-892, ISSN: 0925-2312

- L. Sousa, P. Tomás, F.J. Pelayo, A. Martínez, C. Morillas, S. Romero, A FPL bioinspired visual encoding system to stimulate cortical neurons in real-time. Lecture Notes in Computer Science, vol. 2778. ISSN: 0302-9743; ISBN: 3-540-40822-3, pp. 691-700.

- Christian A. Morillas, Samuel F. Romero, Antonio Martínez, Francisco J. Pelayo, Eduardo Fernández, A Computational Tool to test Neuromorphic Encoding Schemes for Visual Neuroprostheses. Lecture Notes in Computer Science, vol. 3512. ISSN: 0302-9743; ISBN: 3-540-26208-3, pp. 510-517.

- Antonio Martínez, Leonardo M. Reyneri, Francisco J. Pelayo, Samuel F. Romero, Christian A. Morillas, Begoña Pino, Automatic generation of bio-inspired retina-like processing hardware. Lecture Notes in Computer Science, vol. 3512. ISSN: 0302-9743; ISBN: 3-540-26208-3, pp. 527-533.

- Samuel Romero, Antonio Martínez, Christian Morillas, Francisco Pelayo, Eduardo Fernández, Sistema de Pre-Procesamiento de la Información Visual para Neuroprótesis Visuales. Revista de Neurología, Volumen 41 Suplemento 2 (Septiembre 2005), Abstract P447, ISSN: 0210-0010.
- Christian Morillas, Samuel Romero, Antonio Martínez, Francisco Pelayo, Leonardo Reyneri, Markus Bongard, Eduardo Fernández, A Neuroengineering suite of Computational Tools for Visual Prostheses. Aceptado para ser publicado en Neurocomputing.
- Christian A. Morillas, Samuel F. Romero, Antonio Martínez, Francisco J. Pelayo, Eduardo Ros, Eduardo Fernández, A Design Framework to Model Retinas. Aceptado para su publicación en Biosystems.

- Capítulos de libro:

- L. Sousa, P. Tomás, F. Pelayo, A. Martínez, C. A. Morillas, S. Romero, Bioinspired Stimulus Encoder for Cortical Visual Neuroprostheses. Capítulo del libro New Algorithms, Architectures, and Applications for Reconfigurable Computing. Editan: P. Lysaght and W. Rosenstiel. Springer. Abril 2005. ISBN: 1-4020-3127-0, Capítulo 22, pp. 279-290.
 - Samuel Romero, Francisco J. Pelayo, Christian A. Morillas, Antonio Martínez, Eduardo Fernández, Reconfigurable Retina-like Preprocessing Platform for Cortical Visual Neuroprostheses, Capítulo del libro Neural Engineering, Vol. 3 Neuro-Nanotechnology – Biorobotics, Artificial Implants and Neural Prosthesis. Edita: Metin Akay. IEEE Press Series on Biomedical Engineering. En imprenta.
-

Referencias bibliográficas

- [AFB 2003] American Foundation for the Blind. 2003. http://www.afb.org/info_document_view.asp?documentid=1367
- [Algazi 2001] V. R. Algazi, R. O. Duda, D. M. Thompson, C. Avendano, *The CIPIC HRTF Database*. Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, pp. 99-102. 2001.
- [Ayache 1987a] N. Ayache, B. Faverjon, *Efficient registration of stereo images by matching graph descriptions of edge segments*. The International Journal of Computer Vision. 1987.
- [Ayache 1987b] N. Ayache, F. Lustman, *Fast and reliable passive trinocular stereovision*. Proc. IEEE ICCV'87. 1987.
- [Ayache 1988] N. Ayache, C. Hansen, *Rectification of images for binocular and trinocular stereovision*. Ninth International Conference on Pattern Recognition. 1988.
- [Baker 1981] H. H. Baker, T. O. Binford, *Depth from edge -and intensity-based stereo*. Proceedings of the 7th Joint Conference on Artificial Intelligence. 1981.
- [Barnard 1982] S. Barnard, M. Fischler, *Computational stereo*. ACM Comput. Surveys. 1982.
- [Beaudot 1994] William H. A. Beaudot, *Le Traitement Neuronal de l'Information dans la Rétine des Vertébrés : Un creuset d'idées pour la vision artificielle*, Tesis doctoral, Diciembre 1994.
- [Belhumeur 1996] P. N. Belhumeur, *A Bayesian Approach to Binocular Stereopsis*. Int'l J. Computer Vision, 19(3): 237-260. 1996.
- [Birchfield 1996] S. Birchfield, C. Tomasi, *Depth Discontinuities by Pixel-to-Pixel Stereo*. Technical Report STAN-CS-TR-96-1573. Stanford University. 1996.

- [Birchfield 1998] S. Birchfield, C. Tomasi, *Depth Discontinuities by Pixel-to-Pixel Stereo*. Proc. IEEE Int'l Conf. Computer Vision, pp. 1073-1080. 1998.
- [Blake 1987] A. Blake, A. Zisserman, *Visual reconstruction*. MIT Press. 1987.
- [Bloom 1977a] P. J. Bloom, *Determination of monaural sensitivity changes due to the pinna by use of minimum-audible-field measurements in the lateral vertical plane*. Journal of the Acoustical Society of America, 61: 820. 1977.
- [Bloom 1977b] P. J. Bloom, *Creating source elevation illusions by spectral manipulation*. Journal of the Audio Engineering Society, 25:560. 1977
- [Boahen 2000] K. A. Boahen, *Point-to-point connectivity between neuromorphic chips using address events*, IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 47(5): 416-434. 2000.
- [Bolles 1993] R. C. Bolles, H. H. Baker, M. J. Hannah, *The JISCT stereo evaluation*. Proc. of the Image Understanding Workshop. 1993.
- [Bouguet] Camera Calibration Toolbox for Matlab, by Jean-Yves Bouguet.
http://www.vision.caltech.edu/bouguetj/calib_doc/index.html
- [Brindley 1968] G. S. Brindley, W. S. Lewin, *The sensations produced by electrical stimulation of the visual cortex*. J. Physiol. 196: 479-493. 1968.
- [Burt 1982] P. J. Burt, X. Xu, *Local correlation measures for motion analysis*. IEEE PRIP Conference. 1982.
- [Campbell 1991] P. K. Campbell, K. E. Jones, R. J. Huber, K. W. Horch, R. A. Normann, *A Silicon-based, Three-dimensional Neural Interface: Manufacturing Processes for an Intracortical Electrode Array*. IEEE Transactions on Biomedical Engineering, 38(8): 758-768. 1991.
- [Celoxica] Sitio web de Celoxica: <http://www.celoxica.com>
-

-
- [Chow 2001] A. Y. Chow, M. T. Pardue, V. Y. Chow, G. A. Peyman, C. Liang, J. I. Perlman, *Implantation of silicon chip microphotodiode arrays into the cat subretinal space*. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 9: 86-95. 2001.
- [Chow 2002] A. Y. Chow, M. T. Pardue, J. I. Perlman, S. L. Ball, V. Y. Chow, J. R. Hetling, *Subretinal implantation of semiconductor-based photodiodes: Durability of novel implant designs*. Journal of Rehabilitation Research and Development, 39: 313-321. 2002.
- [Cochran 1992] S. D. Cochran, *3-D surface description from binocular stereo*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1992.
- [Comedia] Sitio web: <http://www.pacific.net.hk/~comedia>
- [Cortivis] Proyecto europeo del V Programa Marco "CORTIVIS: Cortical Visual Neuroprosthesis for the Blind". Ref: QLK6-CT-2001-00279. Sitio web: <http://cortivis.umh.es>
- [Cox 1996] I. J. Cox, S. L. Hingorani, S. B. Rao, B. M. Maggs, *A Maximum Likelihood StereoAlgorithm*. Computer Vision and Image Understanding, 63: 542-567. 1996.
- [Cudeiro 2005] J. Cudeiro, *Maestría en Neurociencia y Biología del comportamiento*. Viguera editores S.L. 2005.
- [Cyberkinetics] <http://www.cyberkineticsinc.com>
- [Dobelle 2000] W. H. Dobelle, *Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex*. ASAIO Journal, 46: 3-9. 2000.
- [Dowling 1987] John E. Dowling, *The Retina: An approachable part of the brain*. The Belknap Press of Harvard University Press. 1987. ISBN: 0-674-76680-6.
- [Duke-Elder 1963] S. Duke-Elder, *System of Ophthalmology: Normal and Abnormal Development*, vol. 3. London: Klimpton. 1963.
- [EAV] <http://www.iac.es/proyect/eavi/index.html>
-

- [Eckmiller 1997] R. Eckmiller, *Learning retina implants with epiretinal contacts*. *Ophthalmic Research*, 29: 281-289. 1997.
- [Enhanced Vision] <http://www.enhancedvision.com>
- [Faugeras 1993a] O. Faugeras, *Three-Dimensional Computer Vision*. MIT Press. 1993.
- [Faugeras 1993b] O. Faugeras, *Real time correlation-based stereo: algorithm, implementations and applications*. INRIA research rapport 2013. 1993.
- [Fua 1991] P. Fua, *Combining stereo and monocular information to compute dense depth maps that preserve depth information*. Proc. of the 12th International Joint Conference on Artificial Intelligence. 1991.
- [Fua 1993] P. Fua, *A parallel stereo algorithm that produces dense depth maps and preserves image features*. *Machine Vision and Applications*, 6. 1993.
- [Gardner 1994] B. Gardner, K. Martin, *HRTF Measurements of a KEMAR Dummy-Head Microphone*. MIT Media Lab Perceptual Computing Technical Report #280. 1994.
- [Gerstner 2002] W. Gerstner, W. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press. 2002. ISBN: 0-521-89079-9.
- [Gil 1995] M. A. Gil, L. Ortigosa, E. Civancos, M. C. Mesa, J. L. Delgado, P. Abreu, *Mácula Rojo Cereza: a propósito de un caso*. II Congreso de la Sociedad Española de Retina y Vítreo, Madrid, Febrero de 1995.
- [Goldstein 1999] E. B. Goldstein, *Sensation and Perception*, 5ª edición. Brooks/Cole. 1999. ISBN: 0534346804.
- [Gouras 1991] P. Gouras, *Color Vision*, in *Principles of Neural Science*. Capítulo 31. Editado por E. R. Kandel, J. H. Schwartz, T. M. Jessell. 1991. ISBN: 0-444-01562-0.
- [Graw 2004] Informe coordinado por Xavier Graw: *Tecnología y discapacidad visual. Necesidades tecnológicas y aplicaciones en la vida diaria de las personas con ceguera y deficiencia visual*.
-

-
- ONCE, 2004. ISBN: 84-484-0133-6. Documento disponible en formato electrónico en <http://www.once.es>
- [Grimson 1981a] W. E. L. Grimson, *A computer implementation of a theory of human stereo vision*. Philosophical Transactions of the Royal Society of London. 1981.
- [Grimson 1981b] W. E. L. Grimson, *From images to surfaces: a computational study of the human early visual system*. MIT Press. 1981.
- [Grimson 1984] W. E. L. Grimson, *Binocular shading and visual surface reconstruction*. Computer Vision Graphics and Image Processing, 28(1). 1984.
- [Grimson 1985] W. E. L. Grimson, *Computational experiments with a feature based stereo algorithm*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7. 1985.
- [Hubel 1988] David H. Hubel, *Eye, Brain and Vision*. Scientific American Library, A division of HPHLP. 1988. ISBN: 0-7167-5020-1.
- [Humayun 1994] M. S. Humayun, R. H. Probst, E. de Juan, K. McCormick, D. Hickingbotham, *Bipolar surface electrical stimulation of the vertebrate retina*. Archive of Ophthalmology, 112: 110-116. 1994.
- [Humayun 1999] M. S. Humayun, E. de Juan Jr., J. D. Weiland, G. Dagnelie, S. Katona, R. Greenberg, S. Suzuki, *Pattern electrical stimulation of the human retina*. Vision Research 39(15): 2569–2576. 1999.
- [Ims Chips] <http://www.ims-chips.de>
- [Intille 1994] S. S. Intille, A. F. Bobick, *Incorporating Intensity Edges in the Recovery of Occlusion Regions*. Proc. Int'l Conf. Pattern Recognition, 1: 674-677. 1994.
- [Kanade 1992] T. Kanade, M. Okutomi, T. Nakahara, *A multiple-baseline stereo method*. Proceedings of the Image Understanding Workshop. 1992.
- [Kanade 1994] T. Kanade, O. Okutomi, *A stereo matching algorithm with an adaptive window: theory and experiment*. IEEE
-

-
- Transactions on Pattern Analysis and Machine Intelligence. 1994.
- [Kandel 1991] E. R. Kandel, C. Mason, *Central Visual Pathways*, in *Principles of Neural Science*. Capítulo 29. Editado por E. R. Kandel, J. H. Schwartz, T. M. Jessell. 1991. ISBN: 0-444-01562-0.
- [Kaplan 1986] E. Kaplan, R. Shapley, *The primate retina contains two types of ganglion cells, with high and low contrast sensitivity*. Proc. Natl. Acad. Sci. USA, 83: 2755-2757. 1986.
- [Kendall 1984] G. S. Kendall, W. L. Martens, *Simulating the Cues of Spatial Hearing in Natural Environments*. Proceedings of the 1984 International Computer Music Conference. 1984.
- [Kimura 2001] M. Kimura, H. Saito, *3D Reconstruction based on Epipolar Geometry*. IEICE Transactions on Information and Systems, E84-D. 2001.
- [Kitamura 1990] Y. Kitamura, M. Yachida, *Three-dimensional data acquisition by trinocular vision*. Advanced Robotics, 4. 1990.
- [Kolb 1984] H. Kolb, R. Nelson, *Neural architecture of the cat retina*. Progress in Retinal Research, editado por N. Osborne y G. Chader. Oxford, UK: Pergamon, 3: 21-60. 1984.
- [Lane 1994] R. Lane, N. A. Thacker, N. L. Seed, *Stretch-correlation as a real-time alternative to feature-based stereo matching algorithms*. Image and Vision Computing. 1994.
- [Lazzaro 1993] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti, D. Gillespie, *Silicon Auditory Processors as computer peripherals*. Advances in Neural Information Processing Systems, 5: 820-827.1993.
- [Lenz 1987] R. K. Lenz, R. Y. Tsai, *Techniques for calibration of the scale facto rand image center for high accuracy 3D Machine Vision metrology*. Proceedings IEEE Int. Conf. Robotics and Automation. 1987.
- [Leventhal 1981] A. G. Leventhal, R. W. Rodieck, B. Dreher, *Retinal Ganglion Cell classes in the old world monkey: morphology*
-

-
- and central projections*. Science Wash. DC, 213: 1139-1142. 1981.
- [Lorente 1933] R. Lorente de N6, *Studies on the structure of the cerebral cortex I. The area entorhinalis*. J Psychol Neurol 1933; 45: 381-438.
- [LVES] LVES, Universidad John Hopkins de Baltimore en colaboración con la NASA.
<http://vesuvius.jsc.nasa.gov/er/seh/pg66s95.html>
<http://www.hopkinsmedicine.org/press/1994/JUNE/199421.HTM>
- [Mahowald 1994] M. Mahowald, *An Analog VLSI System for Stereoscopic Vision*. Editorial Kluwer Academic Publishers. 1994. ISBN: 0-7923-9444-5.
- [Mann 1964] I. C. Mann, *The Development of the Human Eye*. London: British Medical Association. 1964.
- [Marr 1979] D. Marr, T. Poggio, *A computational theory of human stereo vision*. Proceedings of the Royal Society of London. 1979.
- [Martínez 2005] A. Martínez, F. J. Pelayo, C. A. Morillas, E. Ros, L. M. Reyneri, S. Romero, *Automatic synthesis of vision processors on reconfigurable hardware*. V Jornadas de Computación Reconfigurable y Aplicaciones (JCRA 2005). Granada. 2005.
- [Martínez 2006] Antonio Martínez Álvarez, *Generación automática de sistemas bioinspirados de visión en hardware reconfigurable*. Tesis Doctoral. Departamento de Arquitectura y Tecnología de Computadores. Mayo 2006.
- [Matthies 1989] L. Matthies, M. Okutomi, *A bayesian foundation for active stereo vision*. SPIE, Sensor Fusion II: Human and machine strategies. 1989.
- [Medioni 1985] G. Medioni, R. Nevatia, *Segment-based stereo matching*. Computer Vision, Graphics and Image Processing, 31. 1985.
- [Meijer 1992] P. B. L. Meijer, *An Experimental System for Auditory Image Representations*. IEEE Transactions on Biomedical Engineering, 39(2): 112-121. 1992.
-

- [Milenkovic 1985] V. J. Milenkovic, T. Kanade, *Trinocular vision using photometric and edge orientation constraints*. Proceedings of the Image Understanding Workshop. 1985.
- [Morillas 2005] C. A. Morillas, S. F. Romero, A. Martínez, F. J. Pelayo, E. Fernández, *A Computational Tool to test Neuromorphic Encoding Schemes for Visual Neuroprostheses*. Lecture Notes in Computer Science, 3512: 510-517. 2005.
- [Morillas 2006a] C. Morillas, S. Romero, A. Martínez, F. Pelayo, L. Reyneri, M. Bongard, E. Fernández, *A Neuroengineering suite of Computational Tools for Visual Prostheses*. Aceptado para su publicación en número especial de Neurocomputing. En imprenta. 2006.
- [Morillas 2006b] C. A. Morillas, S. F. Romero, A. Martínez, F. J. Pelayo, E. Ros, E. Fernández, *A Design Framework to Model Retinas*. Aceptado para su publicación en número especial de Biosystems. En imprenta. 2006.
- [Morillas 2006c] C. A. Morillas, J. P. Cobos, L. F. Castillo, R. Agís, F. J. Pelayo, *Módulo de sonido 3D en hardware reconfigurable*. VI Jornadas de Computación Reconfigurable y Aplicaciones (JCRA 2006). Cáceres. En imprenta. 2006.
- [Mortara 1995] A. Mortara, A. E. Vittoz, Ph. Venier, *A communication scheme for analog VLSI perceptive systems*. IEEE Journal of Solid State Circuits, 30(6): 660–669. 1995.
- [Mountcastle 1957] V. B. Mountcastle, *Modality and topographic properties of single neurons of cat's somatic sensory cortex*. Journal of Neurophysiology 1957; 20: 408-434.
- [Newman 1979] W. R. Newman, R. F. Sproull, *Principles of Interactive Computer Graphics*. McGraw – Hill. 1979.
- [NFB 2003] National Federation of the Blind. 2003. <http://www.nfb.org/stats.htm>
- [Nielsen 1989] L. Nielsen, M. Mahowald, and C. Mead, *SeeHear*, in *Analog VLSI and Neural Systems*, by C. Mead. Reading: Addison-Wesley. Capítulo 13, pp. 207-227, 1989. ISBN: 0-201-05992-4.
-

-
- [Nishihara 1984] H. K. Nishihara, *Practical real-time imaging stereo matcher*. Optical Engineering. 1984.
- [Normann 1996] R. A. Normann, E. M. Maynard, K. S. Guillory, D. J. Warren, *Cortical implants for the blind*. IEEE Spectrum, 33(5): 54-59. 1996.
- [Normann 1999] R. A. Normann, E. M. Maynard, P. J. Rousche, D. J. Warren, *A neural interface for a cortical vision prosthesis*. Vision Research, 39: 2577-2587. 1999.
- [Ohta 1985] Y. Ohta, T. Kanade, *Stereo by intra -and inter- scanline search*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2. 1985.
- [Okutomi 1990] M. Okutomi, T. Kanade, *A locally adaptive window for signal matching*. Proc. of the ICCV. 1990.
- [Pelayo 1997] F. J. Pelayo, E. Ros, X. Arreguit, A. Prieto, *VLSI Implementation of a Neural Model using Spikes*. Analog Integrated Circuits and Signal Processing, 13(1-2): 111-121. 1997.
- [Pelayo 2003] F. J. Pelayo, C. Morillas, A. Martínez, S. Romero, E. Ros, B. Pino, *A Reconfigurable Machine to Model the First Stages of the Human Visual Pathway*. III Jornadas de Computación Reconfigurable y Aplicaciones (JCRA 2003). Madrid. 2003.
- [Pelayo 2004] F.J. Pelayo, S. Romero, C. Morillas, A. Martínez, E. Ros, E. Fernández, *Translating Image Sequences into Spike Patterns for Cortical Neuro-stimulation*. Neurocomputing, Volumen 58-60 (Junio 2004), Computational Neuroscience: Trends in Research 2004, pp. 885-892, ISSN: 0925-2312
- [Peli 2000] E. Peli, *Augmented Vision for Central Scotoma and Peripheral field loss*. Vision Rehabilitation: Assessment, Intervention and Outcomes. Selected papers from Vision '99: International Conference on Low Vision. pp. 70-74. 2000.
- [Perry 1981] V. H. Perry, A. Cowey, *The morphological correlates of X- and Y-like retinal ganglion cells in the retina of monkeys*. Experimental Brain Research, 43: 226-228. 1981.
-

- [Pino 1999a] B. Pino, F. J. Pelayo, J. Ortega, A. Prieto, *Design and Evaluation of a Reconfigurable Digital Architecture for Self-Organizing Maps*. Proc. of The fourth Int. Conf. on Microelectronics for Neural Networks and Fuzzy Systems (MICRONEURO'99), IEEE Comp. Soc. Press. Granada. 1999.
- [Pino 1999b] B. Pino, *Una Arquitectura Digital VLSI para implementar Mapas Auto-Organizativos*. Tesis Doctoral. Departamento de Arquitectura y Tecnología de Computadores. Diciembre 1999.
- [Pollard 1985] S. B. Pollard, J. E. W. Mayhew, J. P. Frisby, *PMF: A stereo correspondence algorithm using a disparity gradient limit*. Perception, 14: 449-470. 1985.
- [Press 1992] W. H. Press, *Numerical recipes in C: The art of scientific computing*. Cambridge University Press. 1992.
- [Ramón y Cajal 1892] S. Ramón y Cajal, *La rétine des vertébrés. La cellule*. Tomo 9, primer fascículo.
- [Ridgeway 1994] D. Ridgeway, *Designing Complex 2-Dimensional Convolution filters*. The programmable Logic DataBook, Xilinx. 1994.
- [Rizzo 1999] J. F. Rizzo, J. Loewenstein, S. Kelly, D. Shire, T. Herndon, J. L. Wyatt, *Electrical stimulation of human retina with a microfabricated electrode array*. Investigative Ophthalmology and Visual Science, 40, S783. 1999.
- [Robert 1991] L. Robert, O. D. Faugeras, *Curve-based stereo: figural continuity and curvature*. IEEE CVPR'91. 1991.
- [Romero 2004] S. Romero, C. Morillas, A. Martínez, F. J. Pelayo, E. Fernández, *Models and Tools for testing visual neuroprostheses*. FENS Abstr., vol.2, A085.17. 2004.
- [Romero 2006] Samuel F. Romero García, *Entorno Hardware/Software para experimentación con Neuroprótesis Visuales*. Tesis Doctoral. Departamento de Arquitectura y Tecnología de Computadores. Septiembre 2006.
-

-
- [Ross 1993] B. Ross, *A practical stereo vision system*. Proc. IEEE CVPR'93. 1993.
- [Saint-Marc 1991] P. Sint-Marc, J. S. Chen, G. Medioni, *Adaptive smoothing: a general tool for early vision*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1991.
- [Schmidt 1996] E. M. Schmidt, M. J. Bak, F. T. Hambrecht, C. V. Kufta, D. K. O'Rourke, P. Vallabhanath, *Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex*. Brain, 119(2): 507-522.1996.
- [Shi 1993] S. Shi, Y. Hung, W. Lin, *Accurate linear technique for camera calibration considering lens distortion by solving an eigenvalue problem*. Optical Engineering. 1993.
- [Shirai 1987] Y. Shirai, *Tree-Dimensional Computer Vision*. Springer – Verlag. 1987.
- [Sid-Ahmed 1990] M. A. Sid-Ahmed, M. T. Boraie, *Dual camera calibration for 3D Machine Vision metrology*. IEEE Transactions on Instrumentation and Measurement. 1990.
- [Sinha 1989] S. S. Sinha, B. G. Schunck, *Discontinuity preserving using surface reconstruction*. Proc. Conf. Comput. Vision Patt. Recogn. 1989.
- [Smith 1986] G. B. Smith, *Stereo integral equation*. Proc. of the A.A.A.I. 1986.
- [Sonka 1993] D. Sonka, H. Hlavac, R. Boyle, *Image Processing. Analysis and Machine Vision*. Chapman & Hall. 1993.
- [Stella 1992] F. Stella, *A modified stereo matching algorithm suitable for implementation on a convolution specialized hardware*. Pattern Recognition Letters. 1992.
- [Sterling 1986] P. Sterling, M. Freed, R. G. Smith, *Microcircuit and functional architecture of the cat retina*. Trends in Neuroscience, 9: 186-192. 1986.
- [Sterling 1990] Peter Sterling, *Retina*, in *The Synaptic Organization of the Brain*, by Gordon M. Shepherd. Oxford University Press. Capítulo 6, pp. 170-213, 1989. ISBN: 0-19-506256-6.
-

- [Terzopoulos 1988] D. Terzopoulos, *The computation of visible-surface representations*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1988.
- [Thacker 1992] N. A. Thacker, *Statistical analysis of a stereo matching algorithm*. Proc. British Machine Vision Conference. 1992.
- [Troyk 2003] P. Troyk, M. Bak, J. Berg, D. Bradley, S. Cogan, R. Erickson, C. Kufta, D. McCreery, E. Schmidt, V. Towle, *A Model for Intracortical Visual Prosthesis Research*. Artificial Organs, 27(11): 1005-1015. 2003.
- [Tsai 1986] R. Y. Tsai, *An efficient and accurate camera calibration technique for 3D Machine Vision*. Proceedings CVPR'86, IEEE. 1986.
- [Tsai 1987] R. Y. Tsai, *A versatile camera calibration technique for high-accuracy 3D Machine Vision metrology using off-the-shelf TV cameras and lenses*. IEEE Journal of Robotics and Automation. 1987.
- [Tsai 1989] R. Y. Tsai, *Synopsis of recent progress on camera calibration for 3D Machine Vision*. The Robotics Review, MIT Press. 1989.
- [van Essen 1981] D. C. van Essen, J. H. R. Maunsell, J. L. Bixby, *The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization*. J. camp. Neural. 199: 293-326. 1981.
- [Veraart 1998] C. Veraart, *Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode*. Brain Research, 813: 181-186. 1998.
- [vOICe] <http://www.seeingwithsound.com>
- [Warren 2001] D. J. Warren, E. Fernández, R. A. Normann, *High-resolution two-dimensional spatial mapping of cat striate cortex using a 100-microelectrode array*. Neuroscience, 105(1): 19-31. 2001.
- [Wolff 1993] L. B. Wolff, E. Angelopoulou, *3-D stereo using photometric ratios*. Proc. of the Image Understanding Workshop. 1993.
-

- [Xilinx] Sitio web de Xilinx: <http://www.xilinx.com>
- [Yachida 1986] M. Yachida, Y. Kitamura, M. Kimachi, *Trinocular vision: New approach for correspondence problem*. Proc. ICPR. 1986.
- [Zhang 1998] Z. Zhang, *Determining the epipolar geometry and its uncertainty: A review*. International Journal of Computer Vision. 1998.
- [Zrenner 2000] E. Zrenner, A. Stett, W. Barth, S. Weiss, H. Haemmerle, *Electrical multisite stimulation of the isolated chicken retina*. Vision Research, 40: 1785-1795. 2000.
- [Zrenner 2001] E. Zrenner, F. Gekeler, H. Schwahn, A. Stett, K. Kohler, *Subretinal microphotodiodes to replace photoreceptor-function. A review of the current state*. Les Seminaires Ophthalmologiques d'IPSEN, tome 12: 77-95. 2001.
-