



UNIVERSIDAD DE GUANAJUATO

---

---

CAMPUS IRAPUATO - SALAMANCA

DIVISIÓN DE INGENIERÍAS

*“Control de la abstracción de la imagen para mejorar la percepción visual en personas con baja visión”*

**TESIS**

QUE PARA OBTENER EL GRADO DE:

**MAESTRO EN INGENIERÍA ELÉCTRICA**

(Opción: Instrumentación y Sistemas Digitales)

PRESENTA:

***Ing. Alberto Patiño Saucedo***

DIRECTORES:

**Dr. Horacio Rostro González**

**Dr. Carlos Hugo García Capulín**

Salamanca, Gto., a 11 de mayo del 2017.

**DR. DONATO HERNÁNDEZ FUSILIER**  
**COORDINADOR DE ASUNTOS ESCOLARES**  
**PRESENTE-**


Por medio de la presente, se otorga autorización para proceder a los trámites de impresión, empastado de tesis y titulación al alumno(a) Alberto Patiño Saucedo del *Programa de Maestría en Ingeniería eléctrica* y cuyo número de *NUA* es: 145278 del cual soy director. El título de la tesis es: Control de la abstracción de la imagen para mejorar la percepción visual en personas con baja visión

Hago constar que he revisado dicho trabajo y he tenido comunicación con los sinodales asignados para la revisión de la tesis, por lo que no hay impedimento alguno para fijar la fecha de examen de titulación.

**ATENTAMENTE**

  
Horacio Rostro González

NOMBRE Y FIRMA  
**DIRECTOR DE TESIS**  
**SECRETARIO**


  
Horacio Rostro González

NOMBRE Y FIRMA  
**DIRECTOR DE TESIS**



Oscar O. Duarte Méndez

NOMBRE Y FIRMA  
**PRESIDENTE**

  
Raúl E. Sánchez Yáñez

NOMBRE Y FIRMA  
**VOCAL**

*A mi madre y mis hermanos.*

# Agradecimientos

Comienzo agradeciendo a mi asesor de tesis, Dr. Horacio Rostro, por abrirme las puertas de su laboratorio, por su orientación en el desarrollo de este trabajo, y por su apoyo invaluable en mis iniciativas.

Agradezco también a mis profesores de la División de Ingenierías del campus Irapuato-Salamanca de la Universidad de Guanajuato, por la formación recibida en mi primer año de maestría. Quisiera destacar la labor del Dr. J. Gabriel Aviña, que facilitó mi llegada a esta casa de estudios, y al Dr. Raúl Sánchez Yañez por hacerme interesar en el procesamiento digital de imágenes mediante sus clases.

Quisiera expresar también mi gratitud hacia el Dr. Pierre Kornprost del INRIA-Sophia Antipolis, al Dr. Fabio Solari y la Dra. Manuela Chessa de la Universidad de Génova, por su guía en los seis meses que trabajé junto a ellos. Así mismo a mis compañeros de laboratorio en Génova: Chiara, Damiano y Gaurvi por las interesantes charlas de café que tuvimos en inglés e italiano.

A mis compañeros y amigos de la maestría, gracias por amenizar mi estadía en México: Brayan, Avendaño, Ñe, Pepe Toño, Marco, Ernesto, Stella y todos los que dejé de nombrar.

Finalmente, debo expresar mi más profunda gratitud a mis familiares, que desde la distancia me dan ánimos y me esperan, y son mi principal motor para seguir adelante.

*Alberto Patiño Saucedo*

# Agradecimientos Institucionales

Expreso mi más sincera gratitud hacia la Universidad de Guanajuato, especialmente a la División de Ingenierías del Campus Irapuato-Salamanca por la formación y el apoyo que he recibido. A todos los profesores y personal administrativo mis agradecimientos por su duro trabajo durante estos dos años que me permitieron obtener el título de maestro en ingeniería.

Agradezco también al Grupo de Investigación de Aprendizaje Estadístico de la Universidad de Génova (Italia) y al Grupo Biovision del INRIA Sophia Antipolis (Francia), por facilitarme el acceso a sus instalaciones y a los medios de trabajo necesarios para la realización de esta tesis. Mis agradecimientos a los profesores y personal administrativo de estas instituciones por su guía y apoyo.



Este trabajo fue realizado gracias al apoyo recibido a través del Consejo Nacional de Ciencia y Tecnología de México, CONACYT, bajo la beca otorgada en la convocatoria “BECAS NACIONALES 2015 SEGUNDO PERIODO”, con el número de becario 688116.



*Alberto Patiño Saucedo*

# Resumen

En los últimos años, la tecnología de realidad virtual ha experimentado un impulso en su asequibilidad, y un creciente número de aplicaciones ha surgido, proponiendo nuevo contenido visual inmersivo en 360 grados. Para hacer este contenido accesible a personas con baja visión, se deberían adoptar las mismas estrategias usadas en pantallas tradicionales, esto es, usar métodos de mejoramiento de imágenes dedicado para facilitar su interpretación. Este trabajo introduce una aplicación de realidad virtual para dispositivos móviles que implementa contenido visual mejorado en tiempo real. Para lograr esto, se investigó el impacto de diversas transformadas de imágenes para baja visión, en colaboración con un proyecto que busca descifrar el código neuronal generado por la retina y se ideó una estrategia computacional para controlar de forma precisa la abstracción visual sin afectar la naturalidad de la imagen, a través de una reparametrización del operador Diferencia de Gaussianas. Para la aplicación en realidad virtual, nos enfocamos en la técnica propuesta por Peli et al (IOVS, 1991), la cual se basa en un filtro adaptativo que controla el contraste local como una función de la luminancia media local de una imagen. Se implementó dicho algoritmo como un *shader* sobre la GPU de un dispositivo móvil, lográndose el control del parámetro de abstracción visual en tiempo real. Diseñado como un juego, se percibe como una experiencia más agradable que los procedimientos oftalmológicos tradicionales. Este aporte puede constituir una manera en que los pacientes con baja visión ajusten el mejoramiento a sus necesidades diarias en aplicaciones de realidad virtual, y también con fines de entretenimiento.

***Palabras clave:*** abstracción visual, realidad virtual, retina, procesamiento de imágenes, baja visión, GPU, *shader*.

# Abstract

In the last years, virtual reality technology has experienced a boost in affordability, and an increasingly number of applications have emerged proposing new immersive 360 degrees visual content. To make this content accessible for low vision people, one should adopt the same strategies as in traditional displays, i.e., use dedicated image enhancement methods to facilitate their interpretation. This work introduces a virtual reality application for mobile devices that implements real-time content enhancement. In order to achieve this, the impact of various image transforms for low vision was researched, in cooperation with a project that seeks to perform a reverse engineering process of the neural code of the retina. Furthermore, it was conceived an algorithm to accurately control the visual abstraction, without compromising the natural aspect of the outcome, through a reparametrization of the Difference of Gaussians Operator. For the virtual reality application, we focused on the transform proposed by Peli et al (IOVS, 1991), which is based on an adaptive filter that controls the local contrast as a function of the local mean luminance of an image. Such algorithm was implemented as a shader on a mobile GPU, achieving the control of the visual abstraction parameter in real-time. Designed as a game, it is perceived as more enjoyable than traditional ophthalmologic experiments. This application could be a way for low vision people to adjust vision enhancements to their needs in everyday virtual reality applications, also for entertainment purposes.

**Keywords:** visual abstraction, virtual reality, retina, image, processing, low vision, GPU, shader.

# Índice general

|   |           |
|---|-----------|
| <b>1. Introducción</b>  | <b>13</b> |
| 1.1. Motivación . . . . .   | 13        |
| 1.2. Objetivos . . . . .  | 14        |
| 1.2.1. General . . . . .  | 14        |
| 1.2.2. Específicos . . . . .  | 14        |
| 1.3. Antecedentes . . . . .   | 15        |
| 1.4. Contribuciones . . . . .   | 16        |
| 1.5. Esquema de la tesis . . . . .  | 17        |
| <br>  |           |
| <b>2. Visión biológica y computacional: alternativas para baja visión</b> | <b>18</b> |
| 2.1. Generalidades de la visión biológica . . . . .                       | 18        |
| 2.1.1. Percepción visual . . . . .  | 18        |
| 2.1.2. Interacción borde-periferia . . . . .                              | 19        |
| 2.1.3. Sensibilidad al contraste y agudeza visual . . . . .               | 21        |
| 2.2. Descifrando el código neuronal en retinas biológicas . . . . .       | 22        |
| 2.2.1. Sinopsis del experimento . . . . .                                 | 22        |
| 2.2.2. Protocolo experimental de las imágenes de estímulo . . . . .       | 24        |
| 2.2.3. Algoritmos utilizados . . . . .                                    | 25        |



|   |           |
|---|-----------|
| 2.2.4. Preparación del estímulo para el experimento . . . . .                     | 31        |
| <b>3. Control de la abstracción visual utilizando la Diferencia de Gaussianas</b> | <b>34</b> |
| 3.1. Antecedentes: detección de bordes clásica . . . . .                          | 34        |
| 3.2. La Diferencia de Gaussianas . . . . .  | 36        |
| 3.3. La diferencia de Gaussianas Extendida (XDoG) . . . . .                       | 37        |
| 3.4. Reparametrización de la abstracción visual . . . . .                         | 41        |
| 3.5. Implementación . . . . .   | 45        |
| 3.5.1. En imágenes estáticas . . . . .  | 45        |
| 3.5.2. Ajuste automático del umbral e implementación en video . . . . .           | 45        |
| 3.6. Controlando la abstracción visual . . . . .                                  | 45        |
| <b>4. Aplicación en Realidad Virtual</b>  | <b>49</b> |
| 4.1. Dispositivos de Realidad Virtual . . . . .                                   | 49        |
| 4.2. Aplicaciones para baja visión en Realidad Virtual . . . . .                  | 53        |
| 4.3. Programación GPU móvil por medio de <i>shaders</i> . . . . .                 | 54        |
| 4.4. Aplicación para baja visión . . . . .  | 55        |
| 4.4.1. Diseño . . . . .   | 55        |
| 4.4.2. Implementando del algoritmo de Peli por medio de <i>shaders</i> . . . . .  | 57        |
| 4.4.3. Implementación en Unity . . . . .  | 57        |
| <b>5. Conclusiones</b>  | <b>61</b> |
| 5.1. Trabajos futuros . . . . .   | 62        |
| <b>Bibliografía</b>   | <b>63</b> |

# Índice de figuras

|  |    |
|--|----|
| 2.1. Esquema de la retina. . . . .   | 19 |
| 2.2. Mecanismo centro-periferia de la retina, con los tipos de células ON y OFF. . . . .   | 20 |
| 2.3. Tablas de verdad y circuito lógico del mecanismo de activación centro-periferia. . . . .  | 21 |
| 2.4. Sensibilidad al contraste de una retina biológica a distintas frecuencias espaciales (Adaptada de [1]). . . . .   | 22 |
| 2.5. Carta de Snellen, usada en optometría para la medición de la agudeza visual. . . . .  | 23 |
| 2.6. Superior izquierda: las elipses superpuestas a la imagen de barras representan los campos receptivos estimados de cada célula ganglionar de la retina. Inferior izquierda: Respuesta típica ( <i>raster plot</i> ) del arreglo de multielectrodos conectado a cada célula ganglionar. Derecha: Rendimiento de la tarea de clasificación a distintas frecuencias espaciales, usando cuatro métodos de decodificación neuronal. . . . . | 24 |
| 2.7. Esquema del protocolo experimental para las imágenes de estímulo. . .   | 25 |
| 2.8. Diagrama de bloques del algoritmo adaptativo de Peli. . . . .   | 27 |
| 2.9. Implementación del algoritmo adaptativo de Peli. . . . .  | 28 |
| 2.10. Implementación del algoritmo de caricaturización de la imagen. . . . .   | 29 |
| 2.11. Implementación del algoritmo de diferencia de gaussianas extendida (XDoG) a una imagen. . . . .  | 30 |
| 2.12. Imágenes de estímulo propuestas. . . . .   | 31 |

|  |    |
|--|----|
| 2.13. Valores de luminancia de las imágenes de estímulo. . . . .   | 32 |
| 2.14. Simulación de la agudeza visual de un ratón para imágenes de barras y distintas frecuencias. . . . .   | 33 |
| 2.15. Simulación de lo observado por la retina de un ratón de acuerdo a su agudeza visual promedio. . . . .  | 33 |
| 3.1. Comparación de detectores de bordes clásicos con algoritmos de mejoramiento de bordes basados en la diferencia de gaussianas. Imagen adaptada de [17] . . . . . | 38 |
| 3.2. Resultado de la variación de distintos parámetros de la XDoG. . . . .   | 40 |
| 3.3. Variación de la abstracción visual al modificarse la dispersión de las funciones gaussianas en la DoG. . . . .  | 41 |
| 3.4. Frecuencias espaciales de imágenes de barras sinusoidales con distinto grosor. . . . .  | 41 |
| 3.5. Filtro DoG. . . . .   | 43 |
| 3.6. Variación del parámetro $t$ para controlar el grosor de los bordes de la imagen resultante. . . . .   | 44 |
| (a). $t = 2$ . . . . .   | 44 |
| (b). $t = 4$ . . . . .   | 44 |
| (c). $t = 6$ . . . . .   | 44 |
| (d). $t = 8$ . . . . .   | 44 |
| (e). $t = 10$ . . . . .  | 44 |
| (f). $t = 12$ . . . . .  | 44 |
| (g). $t = 15$ . . . . .  | 44 |
| (h). $t = 20$ . . . . .  | 44 |
| 3.7. Histograma de las DoG de diversas imágenes. . . . .   | 46 |
| 3.8. Aplicación del algoritmo XDoG en video: a) umbral constante, b) umbral variable ajustado automáticamente. . . . .   | 47 |

|  |    |
|--|----|
| 3.9. Variación del nivel de abstracción visual en diversas imágenes. . . . .   | 48 |
| 4.1. Visualización de una escena artificial en una pantalla de visión estereoscópica utilizada en realidad virtual. . . . .  | 50 |
| 4.2. Componentes de un sistema de realidad virtual. . . . .  | 50 |
| 4.3. Comparación de las arquitecturas de la CPU y la GPU. Imagen adaptada de [62]. . . . .   | 54 |
| 4.4. Resultados de la implementación de <i>shaders</i> para reproducir el algoritmo de Peli, para diferentes tamaños de kernel. La imagen con tamaño de kernel igual a cero corresponde a la original. . . . . | 58 |
| 4.5. Flujo de las escenas de la aplicación en Unity. . . . .   | 59 |
| 4.6. Captura de pantalla del desarrollo del demo en Unity. . . . .   | 60 |

# Índice de tablas

|   |    |
|---|----|
| 4.1. Comparación de los dispositivos de realidad virtual en auge en 2016.<br>Datos obtenidos de [56]. . . . . | 52 |
| 4.2. Tecnologías de realidad virtual y aumentada enfocadas a baja visión. . .                                 | 53 |

# Capítulo 1

## Introducción

### 1.1. Motivación

En las últimas décadas se ha experimentado un creciente entendimiento de los mecanismos intrínsecos de la visión, tanto biológica como digital. Por parte de la visión biológica, se han hecho avances para comprender cómo la retina codifica la información visual, así como los procesos que involucran a dicha información dentro del cerebro, con especial énfasis en la corteza visual. En cuanto a la visión por computadora, continuamente surgen técnicas para obtener de forma óptima las características más relevantes de la imagen digital, mediante el análisis del color, la textura, y la detección de puntos salientes (bordes y esquinas). Con el poder de cómputo del que se dispone actualmente, también se incluyen las características espacio-temporales de imágenes dinámicas, tales como el flujo óptico. Algunos algoritmos de la visión por computadora se basan o tienen relación con lo que se conoce de la visión biológica. En particular, para la detección de bordes se ha utilizado la Diferencia de Gaussianas (DoG), cuya curva se asemeja a la curva de la sensibilidad al contraste en respuesta a distintas frecuencias espaciales observadas en retinas biológicas [1]. El avance en ambos enfoques para estudiar la visión, así como la creciente colaboración entre la neurociencia y las ciencias de la computación promete un influjo de nuevas ideas y métodos para resolver todo tipo de problemas que involucran la percepción visual.

Uno de los problemas con mayor potencial de ser resueltos mediante un enfoque dual que combine elementos de la visión biológica y digital es el de la baja visión. Ésta se define en términos de la agudeza visual, que mide la máxima frecuencia espacial que puede ser indentificada por la visión central. Según datos recientes de la Organización Mundial de la Salud, la baja visión afecta a unas 246 millones de personas, y de estas el 90 % proviene de países en desarrollo. Si bien el 80 % de los casos de baja visión se deben a errores refractivos curables, el 20 % restante, es decir, unas 49 millones de personas,

no cuenta con un tratamiento efectivo [2]. Estos casos se deben principalmente a dos condiciones que afectan la retina: la Degeneración Macular Asociada a la Edad (DMAE) y la Retinitis Pigmentosa (RP) [3]. Si se tiene en cuenta la relación entre la baja visión y la frecuencia espacial, así como el uso de nuevas tecnologías, se puede pensar en una solución asequible que, utilizando técnicas del procesamiento digital de imágenes, y su implementación en dispositivos móviles y de realidad virtual y aumentada (VR / AR por sus siglas en inglés), puedan mejorar la calidad de vida de las personas con baja visión.

El presente trabajo está motivado por una de estas recientes colaboraciones entre expertos de varias disciplinas. Se trata de un experimento sobre retinas biológicas llevado a cabo en el Departamento de Neurociencias de la Universidad de Newcastle, donde se busca evaluar la respuesta neuronal de las células de la retina a imágenes de distinta frecuencia espacial, con el fin de determinar la mejor decodificación de dicha respuesta en una tarea de discriminación visual. Los resultados confirman la incidencia de la frecuencia espacial en las tasas de reconocimiento de las imágenes de estímulo [4]. En la continuación de este experimento, del cual formamos parte, se extiende el rango del estímulo a imágenes naturales procesadas por técnicas digitales, manteniendo el mismo protocolo del primer experimento. Utilizando una de estas técnicas, basada en la Diferencia de Gaussianas, se planteó la hipótesis de que controlando la frecuencia espacial se puede mejorar la discriminación visual en personas con baja visión.

## 1.2. Objetivos

### 1.2.1. General

Diseñar, implementar y evaluar una estrategia para mejorar la percepción visual en personas con baja visión mediante el control de la abstracción visual de imágenes digitales.

### 1.2.2. Específicos

- Diseñar un protocolo experimental que incluya una serie de imágenes de estímulo para el experimento con MEARR (*Multielectrode Array Retinal Recordings*), utilizando técnicas de procesamiento de imágenes para baja visión.
- Proponer un algoritmo para controlar el nivel de abstracción visual en imágenes digitales, tomando elementos de la visión biológica y el tratamiento digital de señales.

- Implementar dicho algoritmo de forma optimizada en una aplicación móvil en realidad virtual, haciendo uso de *shaders*.

### 1.3. Antecedentes

Históricamente, los lentes ópticos han sido el pilar de la rehabilitación de personas con baja visión. Sin embargo, sus limitaciones en poder de amplificación, campo visual y percepción de la distancia han sido superados por los magnificadores electrónicos [5], los cuales son cada vez más utilizados en forma de dispositivos de mano, o implementados en pantallas, con funciones de amplificación, ajuste de brillo y contraste e inversión de los colores, mejorando así la capacidad de lectura respecto a sus contrapartes ópticas [6]. Siguiendo el desarrollo de los magnificadores digitales, y aprovechando los avances en la visión por computadora, se han propuesto diversos algoritmos de procesamiento de imágenes para mejorar la percepción visual de forma más completa y de acuerdo a cada patología. Estos algoritmos se pueden dividir en dos categorías: aquellos que modifican la estructura de la escena, como amplificación localizada [7], remapeo del texto [8], o reubicación de la escena [9], y aquellos que mejoran la imagen sin modificar su estructura, tales como la simplificación de la escena [10], mejoramiento de los bordes, atenuación del fondo [11], TRON (*Tinted Reduced Outlined Nature*), caricaturización de la imagen [12] y algoritmos de mejoramiento adaptativos [13]. En este trabajo, se hace énfasis en el procesamiento de imágenes de la segunda categoría por ser más generales y su mayor facilidad de ser adaptados a la tecnología de realidad virtual.

Dentro de los algoritmos de mejora de la imagen para rehabilitación en baja visión, el primero en ser propuesto y estudiado a nivel clínico fue el algoritmo de mejora adaptativa de Peli [13], que ajusta el contraste local de la imagen, basado en la información de la frecuencia espacial, de modo que se realzan ciertos detalles de la imagen que permiten al paciente percibirlos de forma más clara. Dicho algoritmo y sus variantes han sido ampliamente probados en personas con limitaciones visuales, en tareas que implican procesar texto [14], imágenes estáticas y vídeo [15]. Se reportó una mejora en el reconocimiento de caras en imágenes estáticas, así como del detalle en señales de televisión, lo cual fue preferido en general sobre la imagen no transformada. Sin embargo, un estudio independiente encontró que tenía un alcance limitado en el reconocimiento de objetos [16].

Para mejorar la imagen sin limitar el reconocimiento de objetos, Bordier et al. [11] propusieron una técnica de atenuación del fondo. Por otro lado, Al-Atabany realizó experimentos con otros algoritmos basados en el mismo principio de Peli, tales como TRON, Image Cartoonization y Edge Overlaying [12]. El común denominador de estos algoritmos es la superposición de los bordes sobre una versión suavizada de la imagen. En este caso se reportó un mejoramiento de la percepción de vídeo con los algoritmos



TRON y Edge Overlaying, así como una preferencia de imágenes caricaturizadas en imágenes estáticas.

Recientemente, Winnermoller [17] propuso una variante del algoritmo de Diferencia de Gaussianas con resultados perceptuales muy prometedores. El uso de la diferencia de Gaussianas probó ser muy versátil en la generación de imágenes de distintos estilos, mediante el control de ciertos parámetros el autor establece una hipótesis en la cual se pueden reproducir los resultados de los algoritmos para baja visión utilizados anteriormente. En particular, el presente trabajo expone una forma de controlar la abstracción visual para personas con baja visión, emulando el concepto del algoritmo adaptativo de Peli.

El uso de nuevas tecnologías para asistir a personas con limitaciones visuales incluye un amplio rango cuyo análisis detallado escapa al alcance de este proyecto. Cabe destacar, sin embargo, el reciente uso de los dispositivos de realidad virtual y aumentada, desarrollados por compañías como Google, Samsung y HTC. Usando estas plataformas o hardware propio, se están desarrollando diversas aplicaciones comerciales dirigidas tanto a pacientes como a especialistas, y que realizan principalmente magnificación, mejora del contraste, identificación de caras y objetos, lectura de textos, etc. Una revisión más detallada de las tecnologías que utilizan realidad virtual y aumentada para la mejora de la percepción visual se realiza en el capítulo 4.

## 1.4. Contribuciones

Las contribuciones de esta investigación fueron las siguientes:

1. La participación en un proyecto multidisciplinario que busca decodificar la respuesta neuronal de la retina ante diversos estímulos visuales. En este proyecto se tuvo como tarea el diseño de un conjunto de imágenes de estímulo, procesadas por medio de técnicas de imágenes dirigidas a personas con baja visión.
2. El estudio del algoritmo de Diferencia de Gaussianas para generar imágenes que permitan una percepción más clara por parte de personas con limitaciones en su agudeza visual. Se propuso una variación de este algoritmo que mediante el control de la abstracción visual puede potencialmente utilizarse en aplicaciones digitales para baja visión.
3. El diseño de una aplicación en realidad virtual dirigida a personas con baja visión, y que incluye el procesamiento de imágenes con el algoritmo de mejoramiento adaptativo de Peli. Esta aplicación fue admitida para ser presentada la conferencia Vision2017, de la *International Society for Low Vision Research and Rehabilitation (ISLRR)*, en Junio de 2017, en los Países Bajos. Se implementó el algoritmo

de Peli como un *shader* que se ejecuta en la GPU de un dispositivo móvil acoplado a un sistema de realidad virtual.

## 1.5. Esquema de la tesis

El presente trabajo se encuentra dividido en 5 capítulos. Los alcances y contenidos de cada uno se describen brevemente a continuación:

Capítulo I. INTRODUCCIÓN. Un capítulo introductorio, donde se presenta la motivación, los objetivos del trabajo, los antecedentes y la contribución de esta investigación.

Capítulo II. VISIÓN BIOLÓGICA Y COMPUTACIONAL: ALTERNATIVAS PARA BAJA VISIÓN. En este capítulo se hace una revisión de los conceptos básicos concernientes al desarrollo del trabajo: las generalidades de la visión biológica y fisiológica, una revisión de las técnicas de procesamiento de imágenes para baja visión, así como una descripción del aporte a un experimento sobre retinas biológicas, en el cual se preparó una serie de imágenes de estímulo usando diversas transformadas para baja visión, con el fin de comparar las respuestas a nivel neuronal y psicofísico.

Capítulo III. CONTROL DE LA ABSTRACCIÓN VISUAL UTILIZANDO LA DIFERENCIA DE GAUSSIANS. En este capítulo se expone en detalle el procesamiento de bordes basado en la Diferencia de Gaussianas, así como la forma de controlar el grosor de las características salientes de la imagen, con el fin de ser utilizado en aplicaciones para ayudar a personas con baja visión. Se exponen de igual forma los pasos de implementación en la CPU y algunos resultados.

Capítulo VI. APLICACIÓN EN VR. Se inicia este capítulo haciendo una revisión de las aplicaciones que utilizan realidad virtual y aumentada para mejorar la percepción visual en personas con baja visión. Luego se presenta el desarrollo de una aplicación para su uso en un dispositivo móvil acoplado a un casco de realidad virtual, utilizando una de las transformadas de imágenes para baja visión, optimizada para su uso en la tarjeta gráfica del dispositivo móvil.

Capítulos V. CONCLUSIONES. Este capítulo presenta algunas reflexiones sobre el desarrollo del trabajo, sus limitaciones, así como posibles oportunidades a futuro.

## Capítulo 2

# Visión biológica y computacional: alternativas para baja visión

En este capítulo se hace una introducción a los conceptos básicos relacionados con el procesamiento de imágenes dirigido a personas con baja visión. Primero se hace una revisión de las generalidades de la visión biológica, donde se hace énfasis en el procesamiento llevado a cabo por la retina para detectar puntos prominentes de la imagen. Se resalta a continuación la importancia de la agudeza visual en la percepción de las imágenes y finalmente se muestra uno de los intentos para descifrar la respuesta neuronal de la retina al estímulo visual, en un experimento que involucra el uso de procesamiento computacional de imágenes enfocado a baja visión, con el fin de establecer una relación entre la respuesta neuronal y la psicofísica de este problema.

### 2.1. Generalidades de la visión biológica

#### 2.1.1. Percepción visual

La capacidad de interpretar la información de la luz proyectada por el entorno se conoce como visión y es una herramienta de suma importancia para la mayoría de seres vivos. En el caso del ser humano, el ojo, órgano visual por excelencia, presenta un grado de complejidad difícilmente comparable a otros órganos del cuerpo. Dentro del ojo, el componente central de la visión es la retina. Se trata de un tejido cuya interacción con la luz inicia una cascada de eventos eléctricos y químicos, que a su vez activan una serie de impulsos nerviosos que se envían al cerebro a través del nervio óptico [18].

El tejido que conforma la retina es una estructura basada en capas de neuronas,

conectadas entre sí a través de sinapsis. Esto quiere decir que es una extensión del sistema nervioso central. En la capa más externa se encuentran dos tipos de células fotorreceptoras: los conos, que se encargan de la percepción del color y los bastones, que ayudan a ver en distintas condiciones de iluminación. La respuesta de estas células a la luz se combina en las células bipolares, que junto a las células horizontales y amacrinas modulan la información, que finalmente es recolectada por las células ganglionares, en donde se realiza una especie de preprocesamiento que es enviado en forma de impulsos eléctricos al cerebro [19].

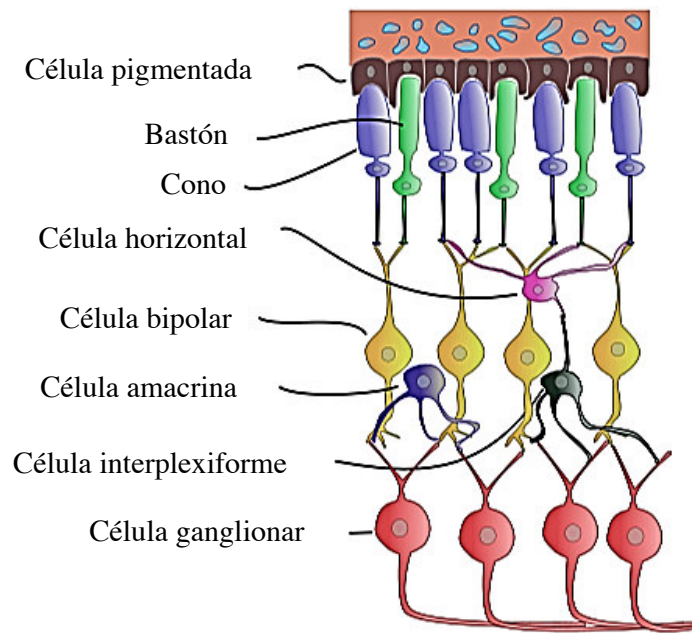


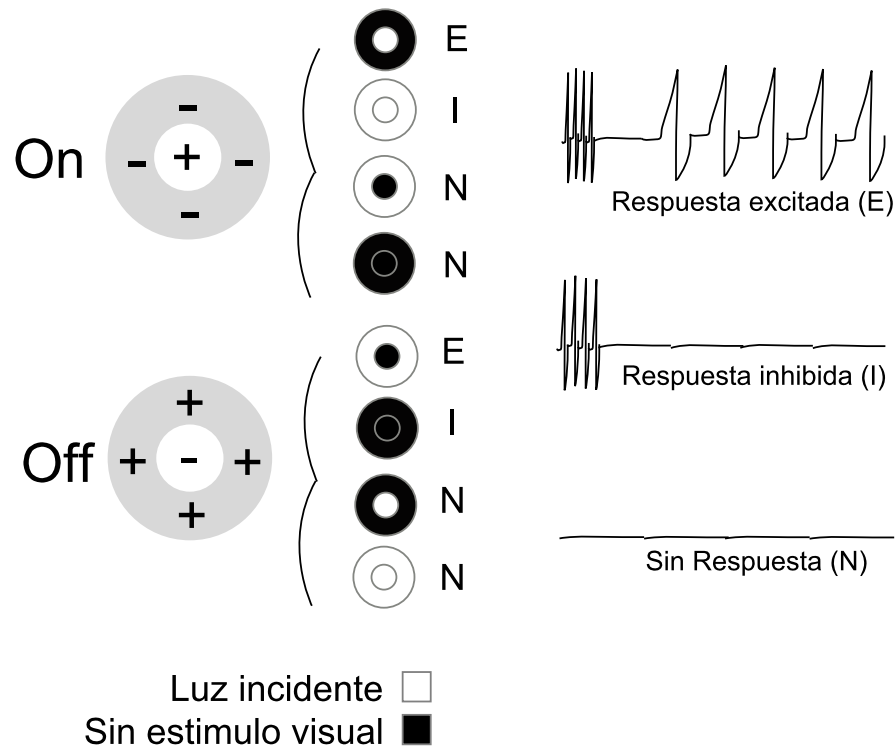
Figura 2.1: Esquema de la retina.

El área del cerebro encargada del procesamiento visual se conoce como corteza visual, localizada en el lóbulo occipital. Se ha descubierto la existencia en primates de flujos independientes de información en la corteza visual [20], así como de intrincados mecanismos para la recuperación de dicha información, incluyendo procesamiento jerárquico, paralelo y modularizado [21], tal como ocurre en las unidades de procesamiento gráfico (GPU) de los sistemas computacionales.

### 2.1.2. Interacción borde-periferia

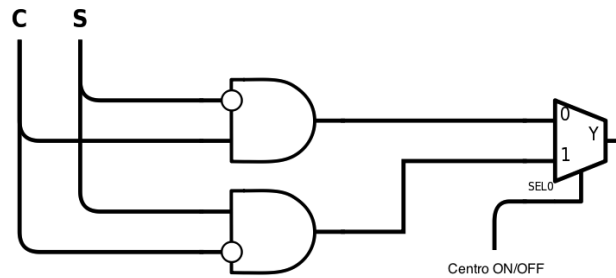
Algunas características importantes de la percepción visual, aparte de las más elementales (color e intensidad de la luz) se pueden rastrear directamente en el preprocesamiento hecho por la retina. La retina envía al cerebro información sobre objetos salientes en el espacio (contornos) y tiempo (flujo óptico) [22], aunque la codificación

neuronal de la misma es un tema que todavía es materia de investigación [4] [23]. Por otra parte, se ha establecido que cada célula ganglionar de la retina (RGC por sus siglas en inglés) cubre un área aproximadamente circular denominada campo receptivo (RF por sus siglas en inglés) [24], compuesto por una parte central y una anular que realizan tareas complementarias de excitación o inhibición de la respuesta al estímulo luminoso. Estas estructuras, denominadas *center-surround* (centro-periferia) pueden ser de dos tipos, *ON* (excitatorias) y *OFF* (inhibitorias) dependiendo de la función de la parte central de la estructura [25]. Ante un estímulo visual caracterizado por la presencia o no de luz en ambas áreas, la respuesta neuronal puede ser de tres tipos: excitatoria, inhibitoria o nula. Un esquema del mecanismo centro-periferia se presenta en la figura 2.2, donde se hace la analogía del funcionamiento de esta estructura con un sistema digital como el mostrado en la figura 2.3, el cual puede tomar una respuesta binaria excitada (1) o no excitada (0). Este mecanismo es equivalente a los algoritmos de detección de bordes usados en visión por computadora [26], ya que su respuesta es máxima ante cambios locales en el estímulo visual.



**Figura 2.2:** Mecanismo centro-periferia de la retina, con los tipos de células ON y OFF.

| CENTRO ON |               |               | CENTRO OFF |               |               |
|-----------|---------------|---------------|------------|---------------|---------------|
| Centro(C) | Periferia (S) | Respuesta (Y) | Centro(C)  | Periferia (S) | Respuesta (Y) |
| 0         | 0             | 0             | 0          | 0             | 0             |
| 0         | 1             | 1             | 0          | 1             | 0             |
| 1         | 0             | 0             | 1          | 0             | 1             |
| 1         | 1             | 0             | 1          | 1             | 0             |



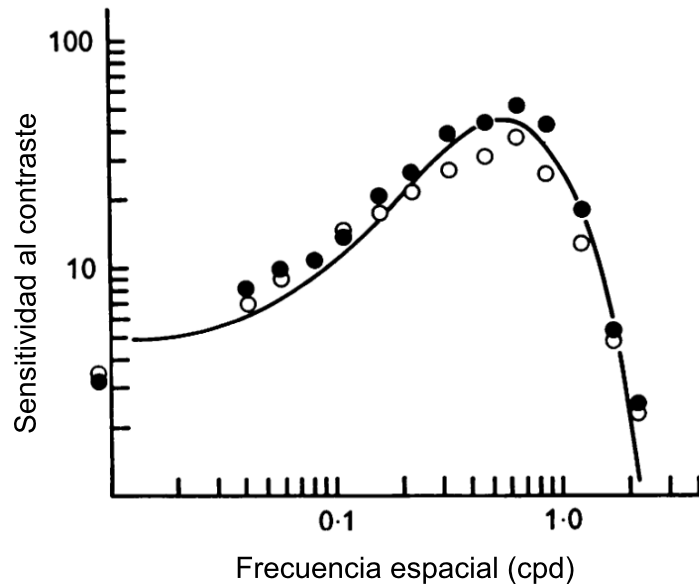
**Figura 2.3:** Tablas de verdad y circuito lógico del mecanismo de activación centro-periferia.

### 2.1.3. Sensibilidad al contraste y agudeza visual

Los campos receptivos de las células ganglionares de la retina no son idénticos. Esto quiere decir que cada célula ganglionar está adaptada a responder mejor a objetos de distintos grados de campo visual, lo cual permite que se pueda percibir en un rango de frecuencias espaciales. Para caracterizar este fenómeno se utiliza el concepto de sensibilidad al contraste, la cual es proporcional a la respuesta de las células ganglionares a un estímulo visual conformado por barras de diferentes frecuencias espaciales (modificando la separación) y contraste (modificando la intensidad de la luz de barras adyacentes). Una gráfica de un experimento de este tipo realizado sobre la retina del gato doméstico [1] se puede ver en la figura 2.4.

En este caso, una sensibilidad de 100 correspondería en una imagen de barras adyacentes de valores 0 y 1 en luminancia normalizada. La frecuencia espacial se calcula en ciclos por grado (*cpd*, cantidad de ciclos del patrón de barras en un grado del campo visual). La curva de sensibilidad al contraste puede ser modelada como una Diferencia de Gaussianas (DoG por sus siglas en inglés) y la frecuencia espacial donde esta sensibilidad es máxima se conoce como agudeza visual.

En optometría, la agudeza visual suele medirse por medio de optotipos, siendo la carta de Snellen el método más utilizado (ver figura 2.5). Las deficiencias en la agudeza visual pueden deberse a errores refractivos (aberraciones en la forma de la córnea), traumas de la corteza visual del cerebro, o factores neuronales directamente asociados a la retina, entre los que destacan los que afectan la visión central (Degeneración Macular Asociada a la Edad) y la periférica (Retinitis Pigmentosa). El valor normal de la agudeza visual es de 1 *cpd* en personas sanas que corresponde a una visión 20/20 en el



**Figura 2.4:** *Sensibilidad al contraste de una retina biológica a distintas frecuencias espaciales (Adaptada de [1]).*

estándar americano.

Con lo expuesto anteriormente se establece el rol importante de la retina en la percepción visual. Su correcto entendimiento es fundamental para comprender el funcionamiento del aparato visual de los seres vivos, y esta información probaría ser útil tanto en el diseño de sistemas de visión por computadora más eficientes, como en el de sistemas de asistencia a personas con deficiencias en su agudeza visual. El experimento que se describe a continuación hace parte de los esfuerzos realizados, por un grupo de investigadores de diversos campos del conocimiento, a fin de lograr entender el código neuronal de la retina, incluyendo la comparación de la respuesta neuronal con la respuesta a nivel psicofísico de personas con baja visión a diversos algoritmos de procesamiento de imágenes.

## 2.2. Descifrando el código neuronal en retinas biológicas

### 2.2.1. Sinopsis del experimento

Partiendo de la red de colaboraciones del laboratorio de Sistemas Bioinspirados de la DICIS - Universidad de Guanajuato, encabezado por el Dr. Horacio Rostro, se

|                          |    |        |
|--------------------------|----|--------|
| <b>E</b>                 | 1  | 20/200 |
| <b>F P</b>               | 2  | 20/100 |
| <b>T O Z</b>             | 3  | 20/70  |
| <b>L P E D</b>           | 4  | 20/50  |
| <b>P E C F D</b>         | 5  | 20/40  |
| <b>E D F C Z P</b>       | 6  | 20/30  |
| <b>F E L O P Z D</b>     | 7  | 20/25  |
| <b>D E F P O T E C</b>   | 8  | 20/20  |
| <b>L E F O D P C T</b>   | 9  |        |
| <b>F D P L T C E O</b>   | 10 |        |
| <b>P E Z O L C F T D</b> | 11 |        |

**Figura 2.5:** Carta de Snellen, usada en optometría para la medición de la agudeza visual.

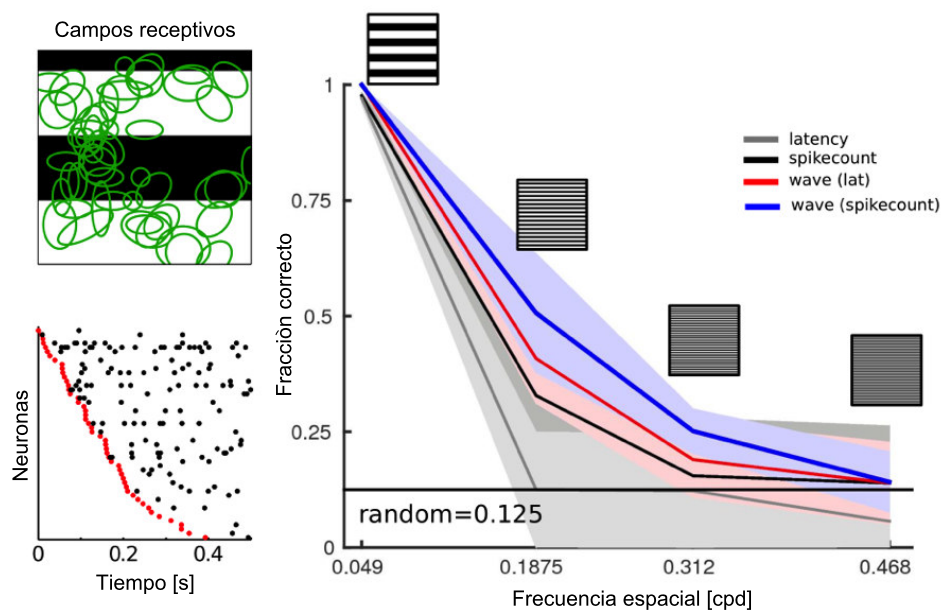
tuvo la oportunidad de participar en un proyecto multidisciplinario liderado por la Dra. Evelyne Sernagor del Departamento de Neurociencias de la Universidad de Newcastle, cuyo objetivo es entender el código neuronal generado por los estímulos visuales que inciden en las retinas biológicas. La metodología consiste en registrar las respuestas (impulsos nerviosos) de las células ganglionares de la retina de mamíferos (ratones), usando como sensor un arreglo de multielectrodos, para luego hacer un análisis de las respuestas neuronales obtenidas, midiendo las tasas de error de distintos mecanismos de decodificación en una tarea de clasificación de los estímulos visuales. En la primera parte de este proyecto [4], se corroboró la relación entre la frecuencia espacial y la capacidad de discriminar distintas fases de un mismo estímulo. Adicionalmente, se llegó a dos conclusiones importantes:

- La decodificación basada en el orden de activación (*rank order coding*) de las células ganglionares de retina (RGCs) mostró resultados más precisos que los decodificadores neuronales clásicos, basados en tasas, latencias y conteo de impulsos. Esto refuerza la hipótesis de que existe una codificación sinérgica de la respuesta al estímulo visual, dado que hay muchas interconexiones entre las RGCs.



- Las respuestas tempranas al estímulo (*first wave of stimulus-evoked spikes*) contienen la mayoría de la información relevante para la discriminación de distintas imágenes.

Un esquema del experimento así como parte de sus resultados se presenta en la figura 2.6. Se observa una reducción general del rendimiento en la detección de distintas fases a medida que la frecuencia espacial aumenta. Una descripción del protocolo experimental utilizado para los estímulos visuales se presenta en la siguiente sección.

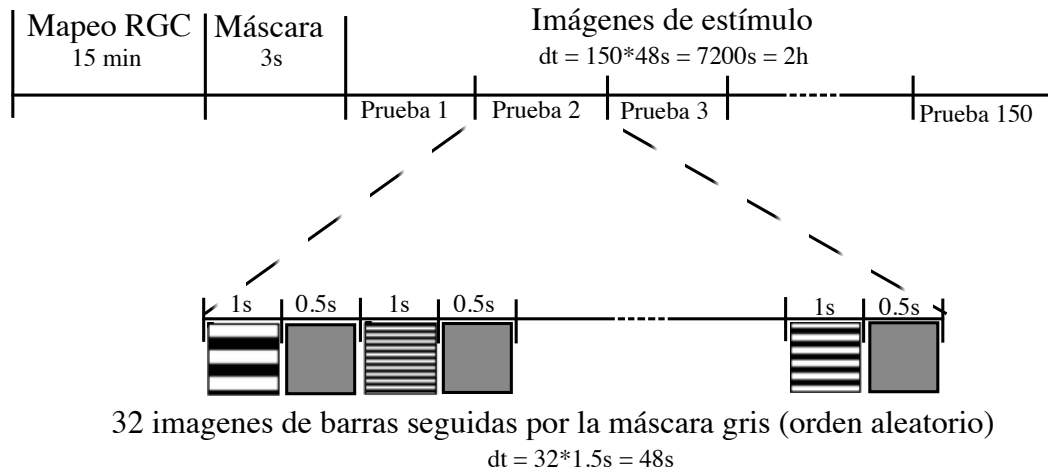


**Figura 2.6:** Superior izquierda: las elipses superpuestas a la imagen de barras representan los campos receptivos estimados de cada célula ganglionar de la retina. Inferior izquierda: Respuesta típica (raster plot) del arreglo de multielectrodos conectado a cada célula ganglionar. Derecha: Rendimiento de la tarea de clasificación a distintas frecuencias espaciales, usando cuatro métodos de decodificación neuronal.

### 2.2.2. Protocolo experimental de las imágenes de estímulo

Las imágenes de estímulo de la primera parte del experimento estuvieron basadas en otro experimento de la misma naturaleza llevado a cabo por Gollisch y Meister [27]. El estímulo consistía de 32 imágenes de barras horizontales con 4 distintas frecuencias espaciales y 8 fases. Considerando un campo receptivo de los ratones de 200-300  $\mu m$  en promedio, las barras tenían un ancho de 1600, 800, 400 y 200  $\mu m$ . Como 1 grado = 30  $\mu m$  en este tipo de retina [28], las cuatro frecuencias espaciales correspondían a

9, 18, 37 y 75 mcpd (miliciclos por grado). Para cada frecuencia espacial se definieron ocho fases  $\phi$  aplicando un desplazamiento vertical constante sobre la imagen. Cada estímulo se presentó 150 veces en bloques aleatorios de 32 estímulos, en intervalos de 0.5 s seguidos de una máscara uniforme gris, de 1 s. El tamaño de las imágenes era de 664x664 pixeles. Un esquema del protocolo se presenta en la figura 2.7.



**Figura 2.7:** Esquema del protocolo experimental para las imágenes de estímulo.

Para la segunda fase de este proyecto, se decidió hacer más complejo el estímulo, usándose 4 transformaciones de imágenes en vez de barras y distintas imágenes en lugar de fases para la tasa de discriminación, manteniendo el resto del protocolo igual. Como criterio para elegir las transformadas, se prefirió utilizar aquellas que han sido evaluadas en pacientes con baja visión, con el fin de establecer una relación entre la respuesta neuronal y la psicofísica.

### 2.2.3. Algoritmos utilizados

La visión humana tiene su máxima resolución en la parte central, llamada fovea. Esta región, aunque sólo cubre 2° del campo visual, se conecta al 10% de los axones del nervio óptico [29]. Los pacientes con daño en esta región presentan una visión borrosa. En terminología del procesamiento digital de imágenes, se produce una pérdida de los componentes de alta frecuencia de la información visual. Además, este efecto se intensifica con imágenes de bajo contraste. Es por ello que se han ideado algunos algoritmos que permiten mejorar el contraste mientras resaltan las características principales de la imagen. Esto se logra dando mayor peso a las características de alta frecuencia mientras se atenúan las de baja, análogo al preénfasis realizado en señales acústicas y de telecomunicaciones para mejorar la relación señal a ruido.

Existe una variedad de transformadas de imágenes que han sido utilizadas pa-

ra mejorar la visión residual en pacientes con baja visión. De las que han sido evaluadas a un nivel psicofísico (refiérase al trabajo de Moshtael [6] para un revisión detallada), se decidió elegir tres transformadas: el Mejoramiento Adaptativo de Peli [13], porque es una transformada clásica del procesamiento de imágenes para baja visión, la Caricaturización de la Imagen (*Image cartoonization*) porque resume la idea general de este tipo de transformadas y la Diferencia de Gaussianas Extendida (XDoG por sus siglas en inglés)[17], fundada en los principios de la visión biológica. También se tuvo en cuenta el costo computacional en la elección de estas transformadas, ya que se busca que sean implementables en tiempo real [30], para desarrollar aplicaciones dedicadas de hardware y realidad virtual. A continuación se hace una descripción de cada transformada.

### Mejoramiento Adaptativo de Peli

Esta técnica de mejoramiento de la imagen aumenta el rango dinámico del contenido de alta frecuencia. Fue propuesta por primera vez por T. Peli [31] como una técnica de mejoramiento general del contraste. Posteriormente, E. Peli [13] la propuso como una ayuda para personas con baja visión, realizando una serie de pruebas en pacientes con pérdida visual moderada, mostrando una mejoría sustancial en el reconocimiento de rostros. Desde entonces se ha convertido en una transformada estándar para evaluar la eficacia de diversas alternativas para el mejoramiento de la percepción visual en este tipo de pacientes, siendo utilizada en [32], [33], [34], [35].

Consiste en un filtro adaptativo en el que la imagen de entrada  $I(x)$  se filtra con un pasa-bajas, mediante el uso de una ventana de convolución cuadrada de tamaño  $N \times N$ . La imagen en bajas frecuencias,  $I_L(x)$  se calcula de la siguiente manera:

$$I_L(x) = \frac{1}{N^2} \sum_{x_i \in \Omega} I(x_i), \quad (2.1)$$

donde  $x = (u, v)$  son las coordenadas de cada pixel de la imagen y  $\Omega$  es la ventana cuadrada centrada en  $x$ .

Sustrayendo  $I_L(x)$  de  $I(x)$ , se obtiene el componente en altas frecuencias,  $I_H(x)$ . Luego, multiplicando  $I_H(x)$  por un factor  $k > 1$ , se incrementa el rango dinámico de la imagen. Se aplica también un procesamiento no lineal sobre la imagen de bajas frecuencias  $I_L(x)$  y el resultado es la suma de  $I'_L(x)$  (imagen pasa-bajas modificada) e  $I'_H(x)$  (imagen pasa-altas escalada).

El algoritmo específico usado para baja visión en [13] se ajusta para incrementar el contraste de las características sobre 1 cpd (agudeza visual normal del ser humano). En este caso, se usa un  $k = 5$  y la no linealidad consiste en reducir el contraste del componente de bajas frecuencias al hacerse un ajuste del histograma con factor 0.9.

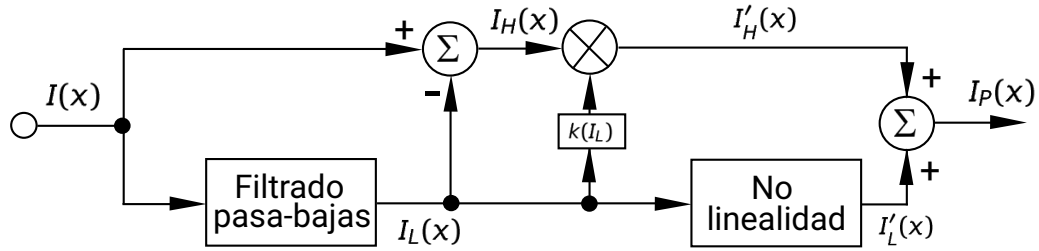


Figura 2.8: Diagrama de bloques del algoritmo adaptativo de Peli.

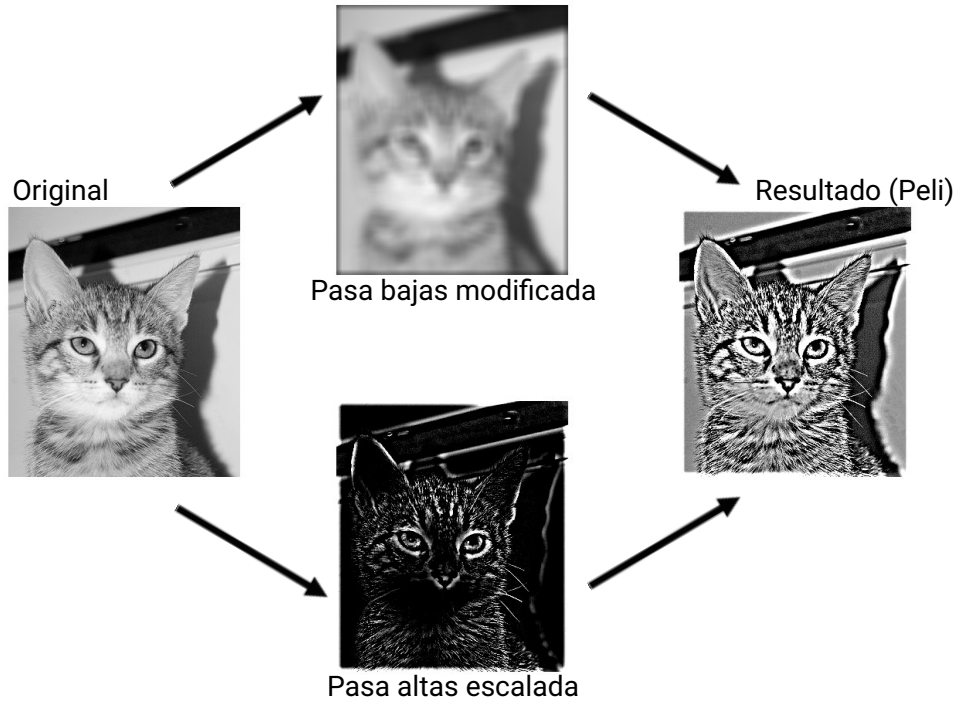
Los pasos generales de la implementación en Matlab de esta transformada son:

1. Suavizar la imagen mediante una ventana de convolución.
2. Sustraer la imagen suavizada a partir de la original (para obtener el componente de alta frecuencia).
3. Multiplicar la imagen de alta frecuencia por un escalar mayor a uno, para aumentar el rango dinámico.
4. Realizar un ajuste del histograma a la imagen de baja frecuencia.
5. Unir ambos componentes modificados para formar la imagen resultante.

De estos pasos, el único que puede suponer un costo computacional alto es el primero, especialmente si se desea suavizar mucho la imagen, lo que implicaría una ventana de convolución mayor. La figura 2.9 muestra algunos resultados de la implementación del algoritmo adaptativo de Peli. La imagen original fue obtenida de la base de datos Imagenet [36].

## Caricaturización de la Imagen

Johnson y Fairchild [37] en sus estudios sobre el mejoramiento de la imágenes de televisión encontraron que la calidad en la percepción de las caricaturas no aumentaba aún cuando se le aplicaron técnicas de mejoramiento de la imagen. Se dedujo que este tipo de imágenes era en sí una opción para su uso en personas con baja visión debido a su alto contraste, por lo que se buscaron técnicas para convertir las imágenes del mundo real en algo similar a las caricaturas. La técnica general consiste en superponer una imagen de bordes a una imagen suavizada (y opcionalmente cuantizada). Winnermøller utilizó una forma de abstracción de la imagen en tiempo real, implementada en la GPU y basada en aplicaciones sucesivas de una adaptación rápida del filtro bilateral, luego cuantizada y unida a la imagen de bordes obtenida con la Diferencia de



**Figura 2.9:** Implementación del algoritmo adaptativo de Peli.

Gaussianas (DoG) [30]. Este trabajo fue probado a nivel psicofísico [38] reportándose una mejoría en las tasas de reconocimiento y retención en memoria de imágenes de personas famosas.

La implementación de la caricaturización realizada en Matlab, basada en el filtro usado por Winnermöller, incluye los siguientes pasos:

1. Suavizar la imagen mediante la aplicación de un filtro bilateral.
2. Obtener la imagen de bordes mediante la Diferencia de Gaussianas Extendida (explicación más detallada en el capítulo 3).
3. Unir las dos imágenes para obtener el resultado.

Esta implementación requiere de una serie de optimizaciones para que pueda ser implementada en tiempo real. En específico, el filtro bilateral, que para cada pixel calcula una media ponderada de los valores de intensidad de los pixeles cercanos se define de la siguiente manera:

$$I_B(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(\|I(x_i) - I(x)\|) g_s(x_i - x), \quad (2.2)$$

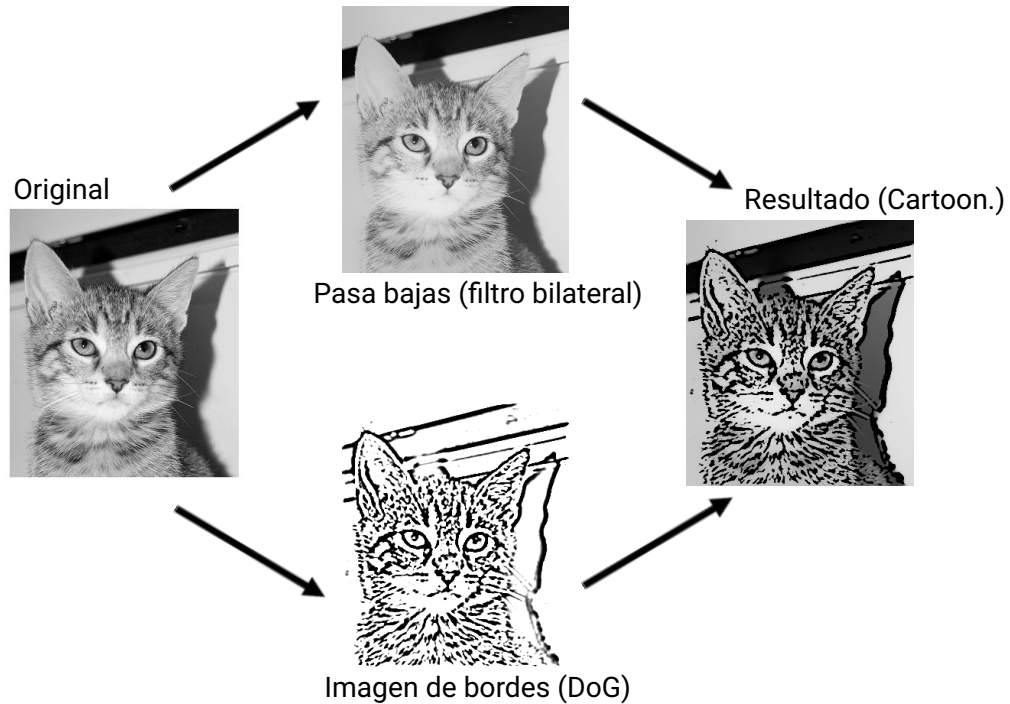
donde

$$W_p = \sum_{x_i \in \Omega} f_r(\|I(x_i) - I(x)\|)g_s(x_i - x), \quad (2.3)$$

aquí  $I_B(x)$  es la imagen filtrada,  $I(x)$  es la imagen original,  $x$  son las coordenadas del pixel actual a ser filtrado,  $\Omega$  es la ventana centrada en  $x$ ,  $f_r$  es un kernel para suavizar las diferencias en intensidad y  $g_r$  es un kernel para suavizar las diferencias en coordenadas.

Una aproximación rápida del filtro bilateral fue propuesta por Paris y Durand [39], en la que se reinterpreta el filtro bilateral como una operación de pasa bajas realizada por un kernel gaussiano. Esto permite realizar un submuestreo de la imagen de entrada con una tasa de al menos dos veces la frecuencia máxima considerada (teorema del muestreo), reduciendo el tamaño de la entrada, para posteriormente aplicar la ecuación 2.2 y sobremuestrear linealmente para obtener una imagen de salida de igual tamaño a la de entrada.

Los pasos generales del algoritmo de caricaturización de la imagen se muestran en la figura 2.10.

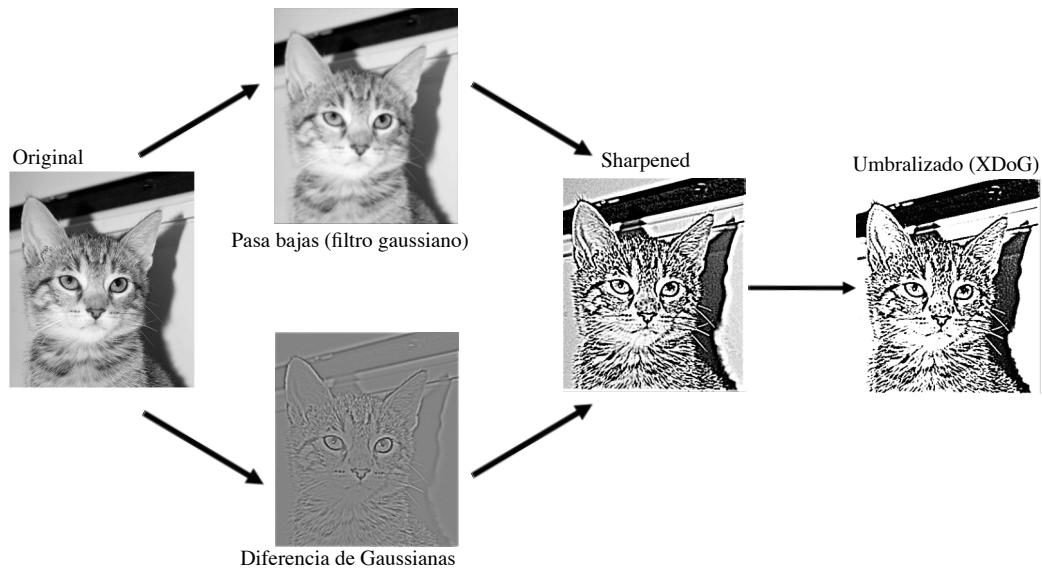


**Figura 2.10:** Implementación del algoritmo de caricaturización de la imagen.

### Diferencia de Gaussianas (DoG)

Esta técnica se basa en el algoritmo de detección de bordes propuesto por Marr y Hildreth [26], que utiliza una detección de cruces por cero de la segunda derivada de la función de luminancia de la imagen, y que a su vez se basa en el mecanismo biológico descubierto por Young [40], la interacción borde-centro que se introdujo en la sección 2.1.2. Este algoritmo se puede aproximar como una máscara de diferencia de gaussianas. Adicionalmente, Winnermöller [41] propuso la inclusión de una binarización “suave” en la que utiliza una función tangente hiperbólica en lugar de una función escalón, esto para agregar consistencia temporal en aplicaciones de video. La diferencia de gaussianas puede ser vista como un filtrado pasa bandas que resalta las características deseadas de la imagen, aumentando el contraste respecto a la información menos relevante. Algunas técnicas basadas en la diferencia de Gaussianas han sido evaluadas en estudios sobre la baja visión [12] [32] .

Los pasos de implementación de la diferencia de Gaussianas Extendida, propuesta por Winermoller, son:



**Figura 2.11:** Implementación del algoritmo de diferencia de gaussianas extendida (XDoG) a una imagen.

1. Crear una imagen suavizada mediante la convolución con una máscara gaussiana de desviación estándar  $\sigma_1$
2. Crear otra imagen suavizada, utilizando una máscara gaussiana con desviación estándar  $\sigma_2 \neq \sigma_1$
3. Obtener la respuesta de la diferencia de gaussianas restando una imagen a otra.

4. Escalar la respuesta obtenida, multiplicándola por un valor.
5. Unir la respuesta escalada con la primera imagen suavizada.
6. Realizar una umbralización con la función tangente hiperbólica para obtener la respuesta final.

La técnica de la diferencia de gaussianas tiene la ventaja de ser separable, lo cual permite convertir la convolución de dos dimensiones en dos convoluciones de una dimensión. En el capítulo 3 se discuten los detalles de este tipo de transformadas, así como una reparametrización para controlar mejor la abstracción visual con la misma.

#### 2.2.4. Preparación del estímulo para el experimento

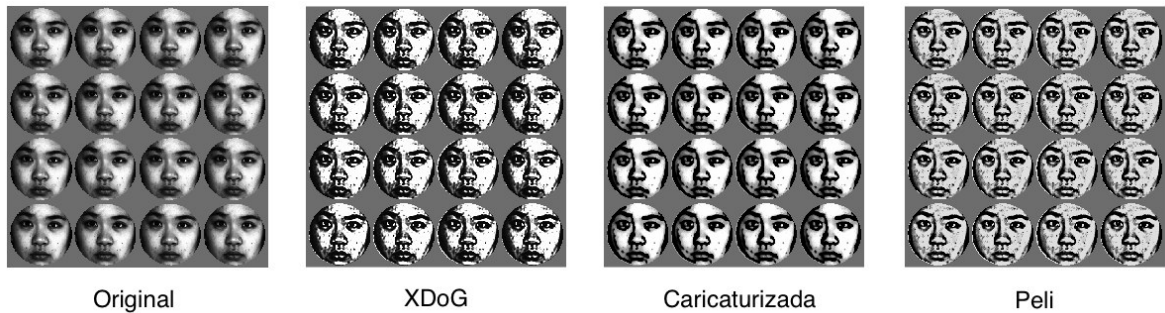
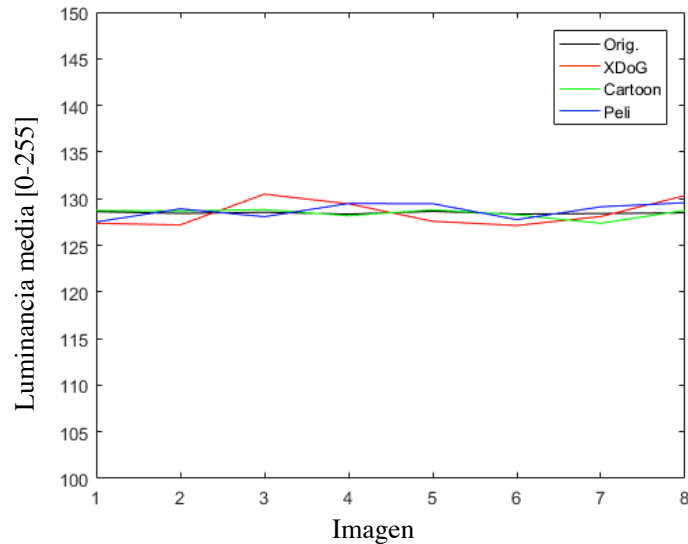


Figura 2.12: Imágenes de estímulo propuestas.

Para el estímulo del nuevo experimento de retinas biológicas, se decidió utilizar imágenes de caras, ya que la mayoría de los estudios psicofísicos de procesamiento de imágenes para baja visión se enfocan en el reconocimiento de caras [12] [32] [42] [43]. Es discutible el hecho de utilizar caras cuando la vista de los ratones no está optimizada para el reconocimiento de las mismas, pero se puede argumentar que una baja agudeza visual en humanos es comparable a la de los ratones. El protocolo exigía una luminancia promedio de 128 para cada imagen, por lo cual se optó por incluir varias caras repetidas en la misma imagen, enmarcadas en un patrón de formas ovaladas. Esto con el fin de tener un balance entre las características realzadas de la imagen (que tienden a tener luminancia cero) con las atenuadas (que tienden a tener luminancia máxima). El hecho de tener una luminancia gris en promedio se hace para reducir la variabilidad, dado que la luminancia puede afectar a la tasa de discriminación.

Ocho imágenes de las caras fueron tomadas de la base de datos *Face Recognition Database* del *Center for Biological and Computational Learning* del MIT. El fondo de las imágenes originales se removió, dejando sólo un óvalo correspondiente a la cara.



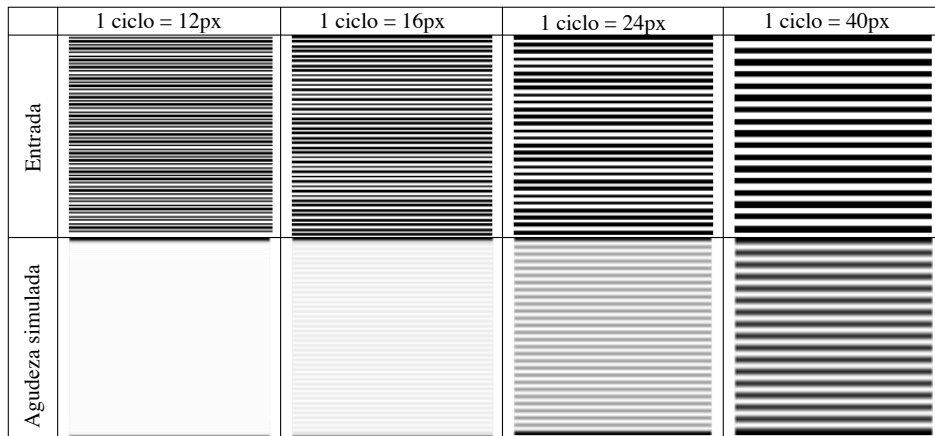


**Figura 2.13:** Valores de luminancia de las imágenes de estímulo.

En total se generaron 16 óvalos faciales por imagen, tal como se observa en la figura 2.12. Todas las imágenes fueron convertidas a la escala de grises y ecualizadas por el método del histograma, para hacer su intensidad promedio cercana al gris. A estas imágenes se les aplicaron las tres transformadas elegidas: el mejoramiento de Peli, la caricaturización y la Diferencia de Gaussinas (XDoG), y junto al original conformaron los 4 conjuntos de datos del protocolo. La figura 2.12 muestra el conjunto de transformadas de una de las imágenes de caras, mientras que la figura 2.13 muestra una gráfica de los valores medios de luminancia de las 8 caras, de acuerdo a cada transformada. Se observa una consistencia en los valores promedios de intensidad con valores cercanos a 128.

Adicionalmente, se hizo una simulación de lo que un organismo con una agudeza visual de 0.5 cpd, tal como el ratón, podría observar. Esto se hizo obteniendo el contenido frecuencial de la imagen por medio de la Transformada Rápida de Fourier y multiplicándolo por una máscara gaussiana bidimensional cuya tercera desviación estándar correspondiese a la frecuencia 0.5 cpd (ciclos de 15 píxeles). La imagen con la visión simulada se obtuvo aplicando la Transformada Inversa Rápida de Fourier a este resultado. Los resultados de la simulación de la agudeza visual del ratón para imágenes de barras y caras se observan en las figuras 2.14 y 2.15 respectivamente.

De lo anteriormente expuesto se deduce que la agudeza visual juega un rol importante en la percepción visual. Cabe plantearse si hay alguna forma de mejorar esta percepción mediante el control de la agudeza visual de las imágenes percibidas en personas con baja visión. Una técnica basada en el procesamiento hecho por la diferencia de gaussianas provee la forma de controlar la abstracción visual, lo cual se muestra en el capítulo siguiente.



**Figura 2.14:** Simulación de la agudeza visual de un ratón para imágenes de barras y distintas frecuencias.



**Figura 2.15:** Simulación de lo observado por la retina de un ratón de acuerdo a su agudeza visual promedio.

## Capítulo 3

# Control de la abstracción visual utilizando la Diferencia de Gaussianas

En este capítulo se presenta una introducción a la diferencia de Gaussianas Extendida, que fue propuesta por primera vez por Winnermöller ([41]), de Adobe Systems en 2011. Se trata de un procesamiento no fotorrealista de imágenes, basado en la diferencia de gaussianas que fue concebida para realizar efectos artísticos. Se plantea cómo esta transformada puede ser utilizada no sólo para mejorar la apariencia visual de la imagen, sino también para aumentar la percepción visual en personas con baja visión. La agudeza visual es uno de los indicadores más importantes de la visión, el cual mide la resolución espacial del sistema de procesamiento visual. Basado en este concepto, se dedujo una forma para controlar la abstracción visual de una imagen sin afectar su naturalidad. El resultado es parecido a un detector de bordes con grosor de bordes ajustable. Mediante el control de la abstracción visual, se puede pensar en un algoritmo adaptativo de imágenes que permita mejorar la percepción visual en personas con limitaciones visuales.

### 3.1. Antecedentes: detección de bordes clásica

La detección de bordes juega un papel primordial en la visión biológica y computacional. Por parte de la visión biológica, se han propuesto diversos mecanismos para entender el proceso que lleva a diferenciar las características salientes o prominentes de la imagen, dependiendo del tipo de prominencia: en intensidad, color, textura, orientación y movimiento [44] [45]. Para la saliencia en intensidad se ha mostrado la

existencia de un mecanismo de activación centro-periferia de las células ganglionares de la retina [46], donde un estímulo dentro del campo receptivo es inhibido en el centro y excitado en la periferia, contribuyendo a la primera etapa de procesamiento neuronal de bordes y esquinas. Este mecanismo es análogo al procesamiento típico que se hace de forma computacional, donde se suele definir el borde como un punto donde la intensidad es marcadamente diferente respecto al entorno.

Los primeros intentos para la detección digital de bordes se enfocaban en identificar las coordenadas donde el gradiente de la imagen fuera mayor. Esto se hacía mediante la convolución con un kernel pequeño, tal como los de Roberts, Prewitt y Sobel [47]. Aunque computacionalmente eficientes, eran muy sensibles al ruido y tenían poca precisión, además de ser común la identificación de bordes duplicados. Las alternativas para solucionar estos problemas tomaron dos rumbos distintos: por un lado Canny [48] propuso un suavizado de la imagen para reducir el ruido y también el algoritmo de supresión de no máximos, con el fin de reducir al mínimo el espesor de los bordes. Por otro lado, Marr y Hildreth [26] sugirieron el uso de la segunda derivada, o Laplaciano, para identificar los cruces por cero de la imagen de bordes, pero este enfoque hacía al procesamiento aún más sensible al ruido, lo cual fue resuelto al agregarse un suavizado gaussiano previo, resultando en un operador llamado el Laplaciano de la Gaussiana (LoG por sus siglas en inglés). En ambas propuestas se realizaba una umbralización binaria para entregar la imagen de bordes final. Una característica del LoG era su limitada eficiencia, dado que no era separable. Por esta razón se propuso aproximarlos mediante la diferencia de Gaussianas, la cual es una operación separable y además es consistente con el enfoque biológico dado por Young [40] y Lindeberg [25] donde se muestra que la sensibilidad al contraste en función de la frecuencia espacial se puede modelar como una diferencia de funciones gaussianas.

Aunque la teoría de Marr y Hildreth es más consistente con la visión biológica, el detector de bordes que adquirió más popularidad en distintas aplicaciones, especialmente aquellas en tiempo real (robótica) fue el detector de Canny. Su teoría de la detección de bordes aún se considera un pilar fundamental en la visión por computadora, y en particular sus tres principios de un buen algoritmo de detección de bordes:

- Buena detección. El algoritmo debe localizar tantos bordes como aparezcan en la imagen.
- Buena localización. Los bordes detectados deben estar tan cerca como los bordes reales.
- Respuesta mínima. El borde debe ser detectado sólo una vez, y el ruido no debería crear falsos bordes.

Analizando estos principios, se puede argumentar que no existe una manera objetiva de determinar los bordes de una manera exacta, sino que esto depende de

la aplicación final, y de la perspectiva del observador. En especial, el tercer principio plantea algunos cuestionamientos ya que favorece la detección de bordes de un pixel de grosor, lo que es conveniente en aplicaciones de robótica, pero no tanto en aplicaciones que buscan mejorar percepción psicofísica de la imagen, lo cual se puede definir como mejoramiento de los bordes. Mientras la detección de bordes clásica se ocupa de la localización exacta de los bordes y su extensión, en el mejoramiento de los bordes lo que prima es el peso (grosor) y la estructura (forma) de los bordes. Es en este caso donde el algoritmo de diferencia de Gaussianas tiene un rol más relevante, como se muestra en las siguientes secciones.

## 3.2. La Diferencia de Gaussianas

La Diferencia de Gaussianas (DoG por sus siglas en inglés) es un operador que puede ser utilizado para enfatizar o detectar bordes, cuando se aplica a una imagen. Aunque no es tan preciso como el algoritmo de Canny en la localización de bordes a diversas escalas [49], logra resultados estéticamente más agradables sin post-procesamiento. Su uso se remonta a la propuesta de Marr y Hildreth de utilizar la segunda derivada, para detectar los cruces por cero y así limitar los resultados de un detector de bordes. A pesar de que en funciones unidimensionales esto se logra con facilidad, para más dimensiones se requiere que la segunda derivada sea tomada en dirección perpendicular al cruce por cero, cuya ubicación en primer lugar es lo que se quiere buscar. Para resolver esto, se propuso utilizar el operador Laplaciano,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}, \quad (3.1)$$

la desventaja de éste es su alta sensibilidad al ruido. Por ello se propuso utilizar un suavizado mediante la convolución de la imagen de entrada con un filtro gaussiano,

$$G(x) = \frac{1}{2\pi\sigma} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right), \quad (3.2)$$

un filtro con propiedades deseables en frecuencia y escala [50]. La ecuación 3.2 describe una función gaussiana bidimensional, centrada en el origen, donde  $x$  es una coordenada de dos valores y  $\sigma$  es la desviación estándar de la distribución en el dominio espacial (que es inversamente proporcional a la frecuencia de corte de un filtro pasa-bajas en el dominio de la frecuencia). Dado que la propiedad conmutativa aplica entre los operadores Laplaciano y convolución, se tiene que:

$$\nabla^2(G(x) * I(x)) = (\nabla^2 G(x)) * I(x), \quad (3.3)$$

donde  $*$  representa el operador de la convolución. De esta forma, en vez de aplicarse el suavizado gaussiano y posteriormente la diferenciación, se pueden combinar ambas

operaciones en una sola, conocida como el Laplaciano de la Gaussiana (LoG). Para extraer los bordes de una imagen filtrada con el LoG, se utiliza un algoritmo que detecta los cruces por ceros, resultando en una respuesta como la que se muestra en la figura 3.1d. Sin embargo, para efecto de mejoramiento de los bordes, el algoritmo de cruces por ceros debe omitirse, y se hace una aproximación que permita la separabilidad de la operación sin comprometer mucho el resultado final. Se aproxima el Laplaciano de la Gaussiana con una Diferencia de Gaussianas. Marr y Hildreth encontraron que esta aproximación es más exacta cuando la relación entre las dos desviaciones estándar es de 1.6, es decir:

$$G(x, \sigma) - G(x, k\sigma) \approx \nabla^2(G(x)), \quad (3.4)$$

donde  $k=1.6$ .

El operador Diferencia de Gaussianas está entonces dado por:

$$D(x) = \frac{1}{2\pi\sigma} \left( \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) - \exp\left(-\frac{\|x\|^2}{2(k\sigma)^2}\right) \right). \quad (3.5)$$

La respuesta de una imagen a la que se le aplica la diferencia de gaussianas será  $T(D(x) * I(x))$ , donde  $T(u)$  representa una función de umbralización binaria, definida como:

$$T(u) = \begin{cases} 1, & u \geq \varepsilon, \\ 0, & u < \varepsilon. \end{cases} \quad (3.6)$$

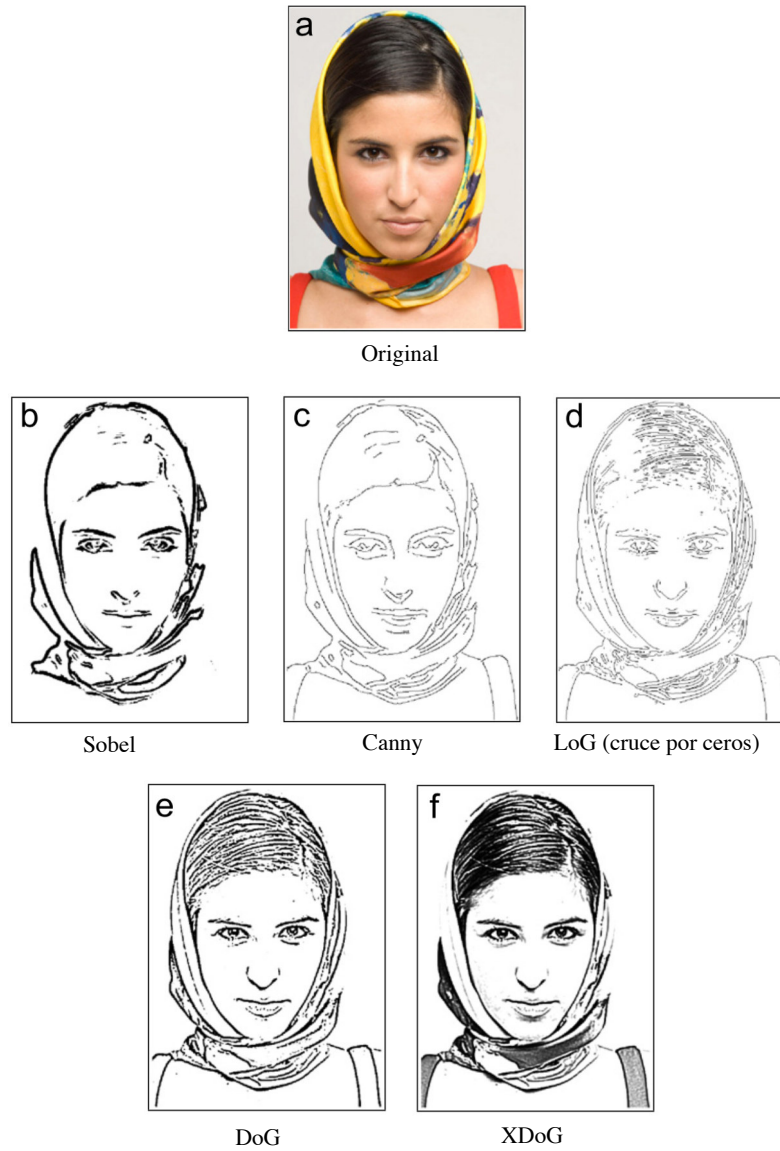
donde  $u$  representa el valor de luminancia de la imagen en determinada posición y  $\varepsilon$  representa el umbral de la binarización. La figura 3.1e muestra la respuesta de  $T(u)$  a la diferencia de gaussianas. Se observa una apariencia más estilizada respecto a los detectores de bordes clásicos.

### 3.3. La diferencia de Gaussianas Extendida (XDoG)

La imagen de bordes generada por la simple umbralización binaria guarda relación con los modelos biológicos propuestos por Young [40]. Tomando como inspiración esos modelos, Winnermöller [41] generó una imagen de bordes usando una variante de la DoG en la cual el peso del efecto inhibitorio de la función gaussiana más ancha puede variar, resultando en la siguiente ecuación:

$$D_x(x) = G(x, \sigma) - \tau G(x, k\sigma), \quad (3.7)$$

donde  $\tau$  es el peso del efecto inhibitorio de la segunda gaussiana.



**Figura 3.1:** Comparación de detectores de bordes clásicos con algoritmos de mejoramiento de bordes basados en la diferencia de gaussianas. Imagen adaptada de [17]

Adicionalmente, se reemplazó la umbralización binaria con una rampa continua:

$$T_x(u) = \begin{cases} 1, & u \geq \varepsilon, \\ 1 + \tanh(\varphi(u - \varepsilon)), & u < \varepsilon. \end{cases} \quad (3.8)$$

donde  $\varphi$  controla la pendiente de la transición entre valores claros y oscuros de la luminancia.

De esta forma  $T_x(D_x(x) * I(x))$  se refiere al filtro XDoG (*eXtended Difference*

of Gaussians) de una imagen de entrada  $I$ . La figura 3.1f muestra cómo la salida de la transformada DoG se modifica como resultado de una umbralización suave y un efecto inhibitorio variable. Combinando los parámetros  $\sigma, \tau, \varphi$  y  $\varepsilon$  se pueden lograr una gran cantidad de efectos sobre la imagen resultante.

Se encontró, sin embargo, que la XDoG es difícil de controlar, dado que una variación de la sensibilidad del filtro a los bordes implica ajustar los parámetros de forma coordinada. Si se hace la descomposición de  $D_x(x)$  como sigue:

$$D_x(x) = G(x, \sigma) - \tau G(x, k\sigma) = (1 - \tau)G(x) + \tau D(x). \quad (3.9)$$

Se puede observar que dicha descomposición es equivalente a un promedio ponderado de la imagen suavizada y la DoG estándar. Con el fin de simplificar el control de los efectos visuales de la transformada, se realiza una reparametrización con las siguientes características deseadas:

- Remoción de la ajustada interdependencia entre parámetros.
- Presencia de parámetros más intuitivos al realizarse el mapeo de algunas operaciones conocidas del procesamiento de imágenes, tales como el suavizado y aumento de la nitidez (*sharpening*).
- Invertibilidad, esto es, que sea posible una conversión completa entre el antiguo y el nuevo conjunto de parámetros y viceversa.

Esta reparametrización deseada se logra al dividir la ecuación 3.9 sobre  $\tau - 1$ , resultando en una representación de la transformada XDoG como un operador de nitidez ajustado  $S(x)$ :

$$S(x) = \frac{D_x(x)}{\tau - 1} = \frac{(1 - \tau)G(x, \sigma) + \tau D(x, \sigma)}{\tau - 1}, \quad (3.10)$$

$$S(x) = \frac{(1 - \tau)G(x, \sigma) + \tau G(x, \sigma) - \tau G(x, k\sigma)}{\tau - 1}, \quad (3.11)$$

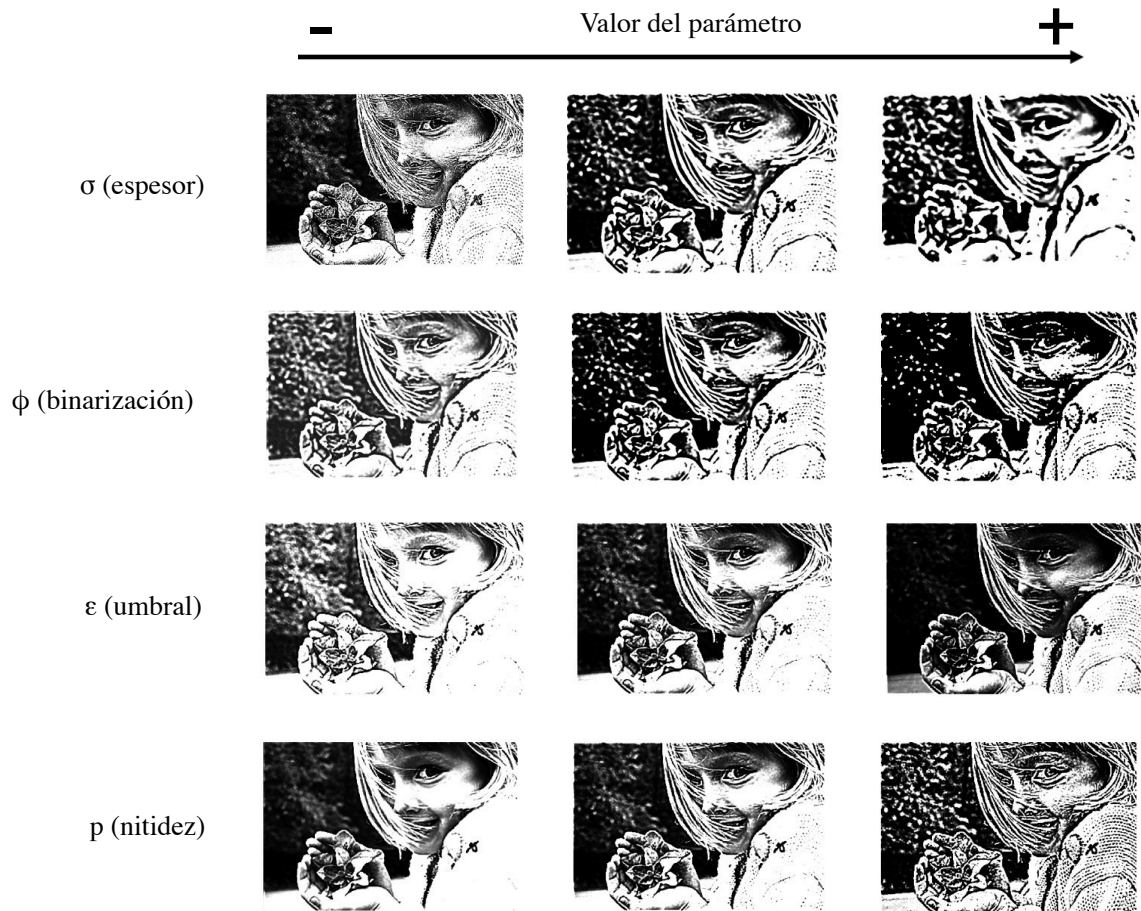
$$S(x) = \frac{G(x, \sigma) - \tau G(x, k\sigma)}{\tau - 1}. \quad (3.12)$$

Haciendo  $p = \tau/(\tau - 1)$ :

$$S(x) = (1 + p)G(x, \sigma) - pG(x, k\sigma). \quad (3.13)$$

Esta nueva fórmula permite obtener el mismo rango de imágenes de salida de la fórmula original, con la ventaja de que ahora el parámetro  $p$  hace posible controlar

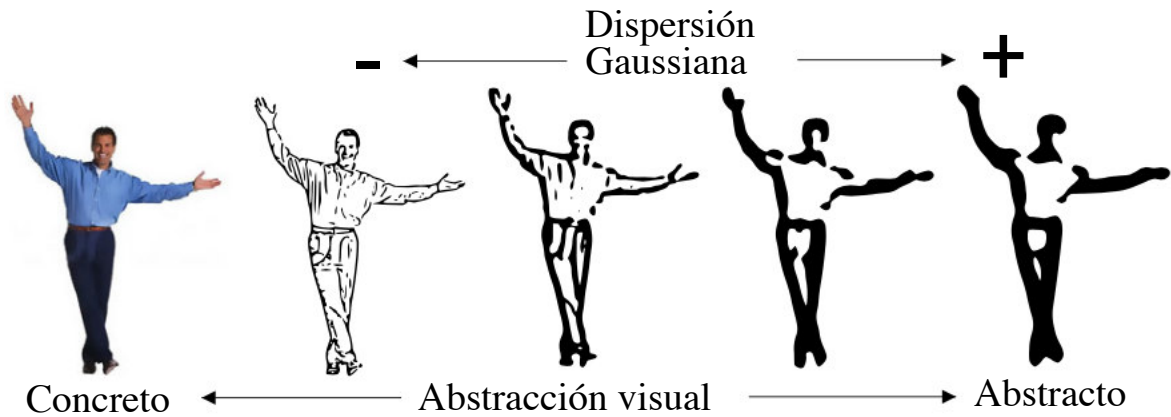




**Figura 3.2:** Resultado de la variación de distintos parámetros de la XDoG.

el peso del efecto de nitidez (*sharpening*) sin afectar otros aspectos del filtro. Adicionalmente,  $\epsilon$  se mide ahora de forma proporcional a la intensidad de la imagen, lo que significa un desacople efectivo de los parámetros.

Es posible obtener un rango amplio de efectos visuales por medio de la XDoG (ver figura 3.2). El parámetro  $\varphi$  controla la agudeza de las transiciones blanco-negro en la imagen y por tanto el contraste,  $\epsilon$  controla la sensibilidad de la umbralización  $p$  controla la nitidez y una variación de  $\sigma$  hace que varíe el espesor de las características detectadas. Este último parámetro está también relacionado con la frecuencia espacial de la imagen resultante, lo cual a su vez incide en la abstracción visual, pensando en aplicaciones de imágenes dirigidas a personas con baja agudeza visual. La figura 3.3 muestra cómo a medida que el kernel de la diferencia de gaussiana está más disperso, la abstracción visual de la imagen aumenta. Una forma de controlar la abstracción visual de una forma intuitiva es modificando el grosor de los bordes de la imagen resultante, mediante una manipulación de las desviaciones estándar de la DoG, lo cual se muestra en la sección que viene a continuación.



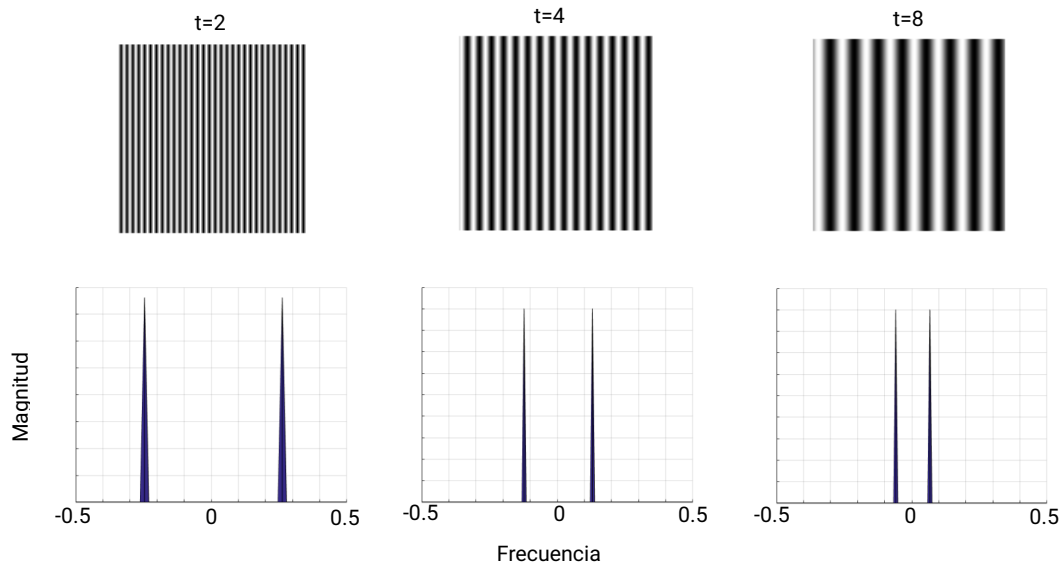
**Figura 3.3:** Variación de la abstracción visual al modificarse la dispersión de las funciones gaussianas en la DoG.

### 3.4. Reparametrización de la abstracción visual

Analizando los contenidos frecuenciales de diferentes grosores de barras sinusoidales, se tiene que cada pico de frecuencia corresponde a diferentes grosores de borde, con la relación:

$$t = \frac{1}{2w}, \tag{3.14}$$

donde  $t$  es el grosor y  $w$  es la distancia entre la frecuencia cero y el pico de frecuencias. Se analiza esto normalizando la longitud del espectro frecuencial.



**Figura 3.4:** Frecuencias espaciales de imágenes de barras sinusoidales con distinto grosor.

Dada una imagen de entrada, se aplica un filtro definido como la diferencia de

dos funciones gaussianas bidimensionales. Para simplificar el análisis, se toma el caso unidimensional, pero se puede extender fácilmente a dos dimensiones. En primer lugar, se aprovecha el hecho de que la transformada de Fourier de una función gaussiana  $g(x)$  con desviación estándar  $\sigma$  es otra función gaussiana. Explícitamente:

$$g(x) = \exp(-x^2/2\sigma^2), \quad (3.15)$$

$$\mathfrak{F}(g(x)) = G(w) = A \cdot \exp(-w^2/2\sigma_f^2), \quad (3.16)$$

donde  $A$  es la relación entre amplitudes de ambas gaussianas:

$$A = \sigma\sqrt{2\pi}, \quad (3.17)$$

y  $\sigma_f$  es la desviación estándar de la gaussiana en el espacio de Fourier, en términos de la desviación estándar espacial:

$$\sigma_f^2 = \frac{1}{4\pi^2\sigma^2}. \quad (3.18)$$

Esto significa que la transformada de Fourier de una función es la misma función pero escalada en amplitud y desviación estándar, de modo que es posible construir diferentes filtros pasa bajas ajustando sólo un parámetro (el  $\sigma$  de la función gaussiana). A partir de dos filtros pasabajas con distintas frecuencias de corte se pueden construir filtros que favorezcan el paso de sólo un rango de frecuencias. Se prefieren filtros pasa bandas porque se quiere que las frecuencias espaciales de la imagen de salida estén centradas alrededor de un valor (no distribuidas como en la imagen de entrada). De esta forma la amplitud de los bordes y en consecuencia la abstracción visual pueden ser controladas.

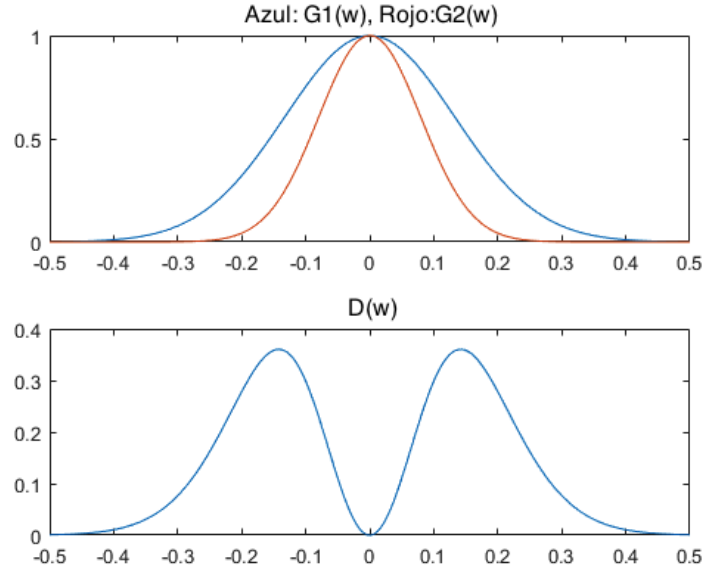
Se podrían utilizar funciones lineales para diseñar un filtro que permita el paso de las frecuencias deseadas, pero la imagen resultante tendría artefactos que afectarían su naturalidad. En vez de esto se utiliza la diferencia de gaussianas, definida como:

$$d(x, \sigma) = g_1(x, \sigma) - \tau g_2(x, k\sigma). \quad (3.19)$$

En este caso  $\tau$  es el factor inhibitorio de la segunda gaussiana y  $k$  es el factor de escalamiento entre desviaciones estándar (con valor de 1.6 según lo sugerido por Marr y Hildreth). La Transformada de Fourier de esto es:

$$\mathfrak{F}(d(x, \sigma)) = D(x, \sigma) = \sigma\sqrt{2\pi} \cdot \exp(-w^2/2\sigma_{f1}^2) - k\tau\sigma\sqrt{2\pi} \cdot \exp(-w^2/2\sigma_{f2}^2). \quad (3.20)$$

Para obtener un cero en la frecuencia central, se requiere que la amplitud de ambas gaussianas en el espacio de Fourier sea la misma. Esto se logra haciendo  $\tau$  igual a  $1/k$ , lo cual resulta en que  $\sigma_{f2} = \sigma_{f1}/k$ .



**Figura 3.5:** *Filtro DoG.*

La diferencia de gaussianas normalizada en el espacio de Fourier, tomando  $\sigma_f = \sigma_{f1}$  será:

$$D(w) = \exp(-w^2/2\sigma_f^2) - \exp(-w^2k^2/2\sigma_f^2). \quad (3.21)$$

Tomando la ecuación 3.21, se encuentra que la frecuencia a la cual la DoG tiene un pico es:

$$w_m = \arg \max_w D(w). \quad (3.22)$$

Tomando la derivada de  $D(w)$  e igualando a cero, tenemos:

$$-\frac{w_m}{\sigma_f^2} \exp(-w_m^2/2\sigma_f^2) + \frac{w_mk^2}{\sigma_f^2} \exp(-w_m^2k^2/2\sigma_f^2) = 0, \quad (3.23)$$

lo cual lleva a:

$$\exp(-w_m^2/2\sigma_f^2) = k^2 \exp(-w_m^2k^2/2\sigma_f^2). \quad (3.24)$$

Aplicando el logaritmo natural a ambos lados y despejando  $w_m$  se tiene:

$$w_m = \sqrt{\frac{4\sigma_f^2 \ln(k)}{k^2 - 1}}. \quad (3.25)$$

Tomando  $\sigma_f^2 = \frac{1}{4\pi^2\sigma^2}$ , para dejar  $w_m$  en términos de la desviación estándar

de la gaussiana en frecuencia espacial, se tiene:

$$w_m = \sqrt{\frac{\ln(k)}{\pi^2 \sigma^2 (k^2 - 1)}}. \quad (3.26)$$

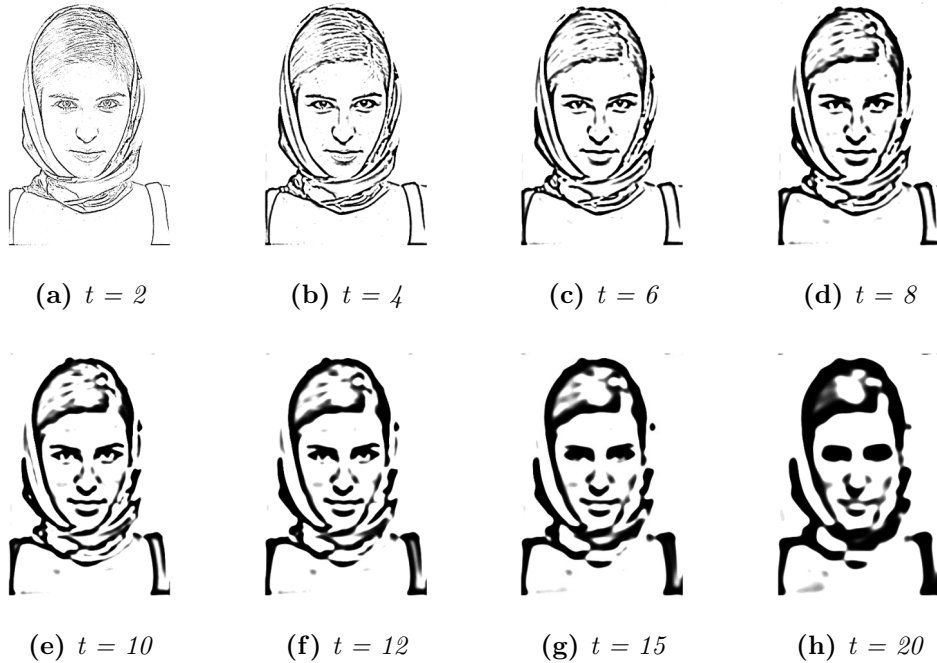
Ahora, tomando la relación descrita en 3.14, el grosor aproximado del borde en función del  $\sigma$  espacial de la primera gaussiana puede ser calculado como:

$$t(\sigma) = \sqrt{\frac{\pi^2 \sigma^2 (k^2 - 1)}{4 \ln(k)}}, \quad (3.27)$$

o si se quiere calcular el  $\sigma$  para un grosor específico:

$$\sigma(t) = \sqrt{\frac{4t^2 \ln(k)}{\pi^2 (k^2 - 1)}}. \quad (3.28)$$

Aplicando la fórmula descrita en la Eq. 3.28 para obtener el  $\sigma$  del operador gaussiano, a diferentes grosores de borde, y después de aplicar la umbralización, se obtienen resultados como los mostrados en la figura 3.6.



**Figura 3.6:** Variación del parámetro  $t$  para controlar el grosor de los bordes de la imagen resultante.

## 3.5. Implementación

### 3.5.1. En imágenes estáticas

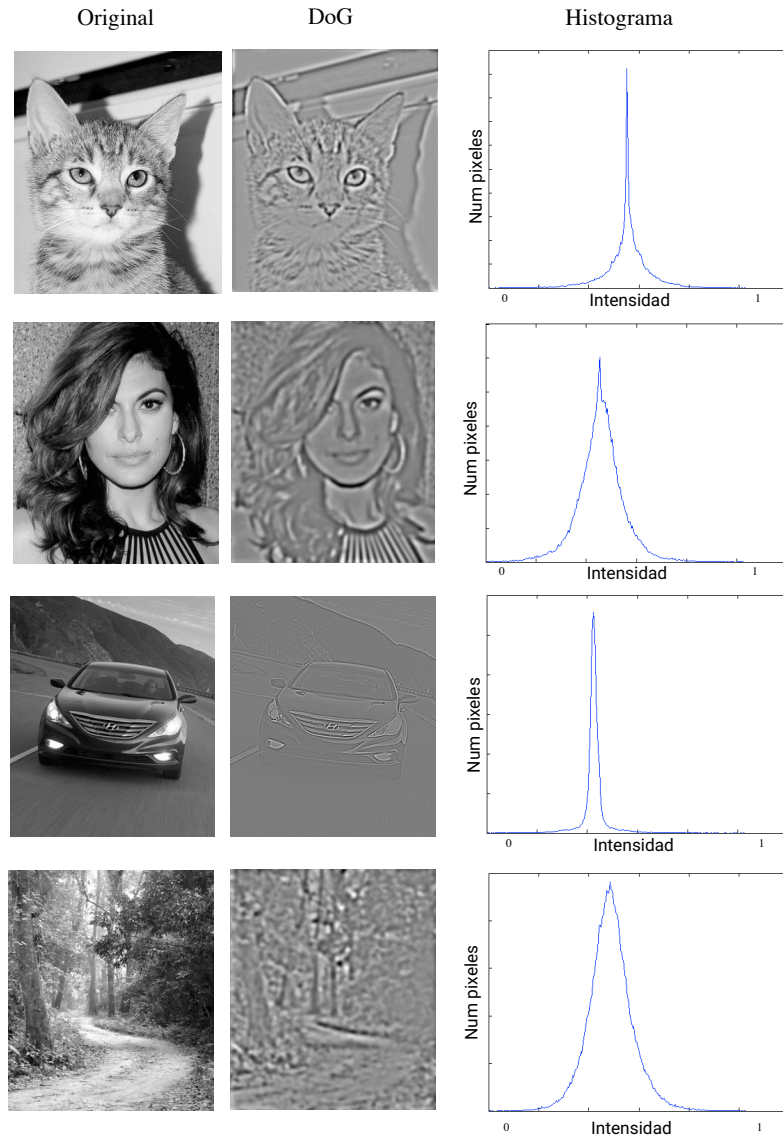
El operador gaussiano bidimensional definido en la ecuación 3.2 es linealmente separable y por tanto se puede implementar eficientemente como dos operadores gaussianos unidimensionales aplicados de forma sucesiva (uno en la dirección horizontal y otro en la vertical). El cálculo se realiza de la siguiente forma: dada una imagen de entrada  $I$ , se calcula  $I_{G1_x}$  e  $I_{G2_x}$  como las respuestas a una convolución de  $I$  con dos gaussianas, de desviaciones estándar  $\sigma$  y  $1.6\sigma$ , donde  $\sigma$  está dado por la ecuación 3.28 si se quiere dar a la imagen de salida bordes de un espesor  $t$ . Luego se calculan  $I_{G1_y}$  e  $I_{G2_y}$  a partir de  $I_{G1_x}$  e  $I_{G2_x}$  al aplicarse los mismos operadores gaussianos en dirección perpendicular. La diferencia de gaussianas se calcula entonces como  $D_G = I_{G1_y} - I_{G2_y}$ . La imagen  $D_G$  es normalizada y umbralizada con la función definida por la ecuación 3.8.

### 3.5.2. Ajuste automático del umbral e implementación en vídeo

Al analizarse la imagen  $D_G$  y su histograma 3.7 se observa que hay un valor de intensidad dominante  $\gamma$ . Es conveniente elegir un valor del umbral  $\varepsilon$  que sea menor que  $\gamma$ , para que los valores correspondientes a la intensidad dominante adquieran la intensidad máxima en la imagen resultante, generándose una imagen de bordes oscuros. En las imágenes umbralizadas de la figura 3.9 se toma un  $\varepsilon = \gamma - 0,02$ . Se observa una consistencia en la sensibilidad al ruido aún comparando distintas imágenes y diferentes grosores. El ajuste automático del umbral resultó útil al hacerse pruebas de este algoritmo en fotogramas sucesivos de vídeo, ya que el resultado mostró una mejor coherencia temporal comparado con el resultado obtenido con un umbral estático, tal como se muestra en la figura 3.8. Una demostración del algoritmo con ajuste del umbral automático aplicado a un vídeo puede observarse en [51].

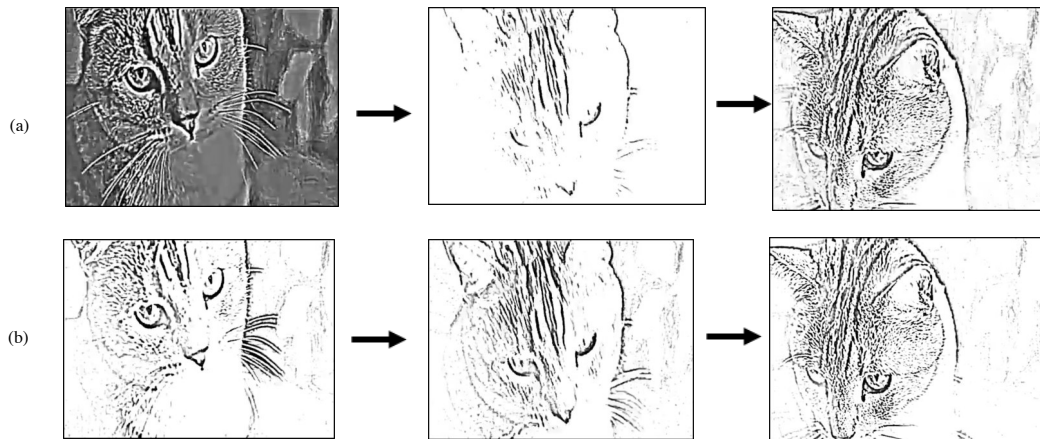
## 3.6. Controlando la abstracción visual

Las imágenes de la figura 3.9 muestran que es posible controlar el nivel de detalle de la imagen resultante por medio de la Diferencia de Gaussianas. También se mostró como es posible implementar este algoritmo en imágenes continuas sucesivas sin perder coherencia temporal. Tal como se expuso en el capítulo anterior, el nivel de detalle o abstracción visual guarda una amplia relación con la percepción visual al



**Figura 3.7:** *Histograma de las DoG de diversas imágenes.*

nivel biológico y psicofísico. Diversos algoritmos aprovechan esto para resaltar las características de cierta banda de frecuencias espaciales y así generar una mejor percepción de las imágenes, en personas con diversos niveles de agudeza visual. En este trabajo se hizo un esfuerzo para controlar una forma precisa el nivel de abstracción, sin comprometer demasiado la estética del resultado visual. El algoritmo DoG modificado que se presentó en las secciones precedentes se puede mapear directamente en los valores de frecuencia espacial deseados, y estos a su vez guardan una estrecha relación con el valor de la agudeza visual. Resta hacer una evaluación de esta variación del algoritmo DoG en la percepción de personas con agudeza visual limitada, pero esto queda para un trabajo posterior. La hipótesis es que si se resaltan las frecuencias justo por debajo



**Figura 3.8:** Aplicación del algoritmo XDoG en video: a) umbral constante, b) umbral variable ajustado automáticamente.

del límite de agudeza visual, se obtienen los mejores resultados en tareas de detección y clasificación.

Existen implementaciones rápidas de filtros gaussianos en tanto en CPU [52] como en GPU [53] [54], con lo cual se abre la posibilidad de realizar una implementación rápida de estos algoritmos pensando en aplicaciones en tiempo real. En efecto, en el capítulo siguiente se describe la utilización de una implementación rápida del suavizado gaussiano para reproducir el algoritmo adaptativo de Peli, usado en una aplicación dirigida a personas con baja visión.



CAPÍTULO 3. CONTROL DE LA ABSTRACCIÓN VISUAL UTILIZANDO LA DIFERENCIA DE GAUSSIANAS

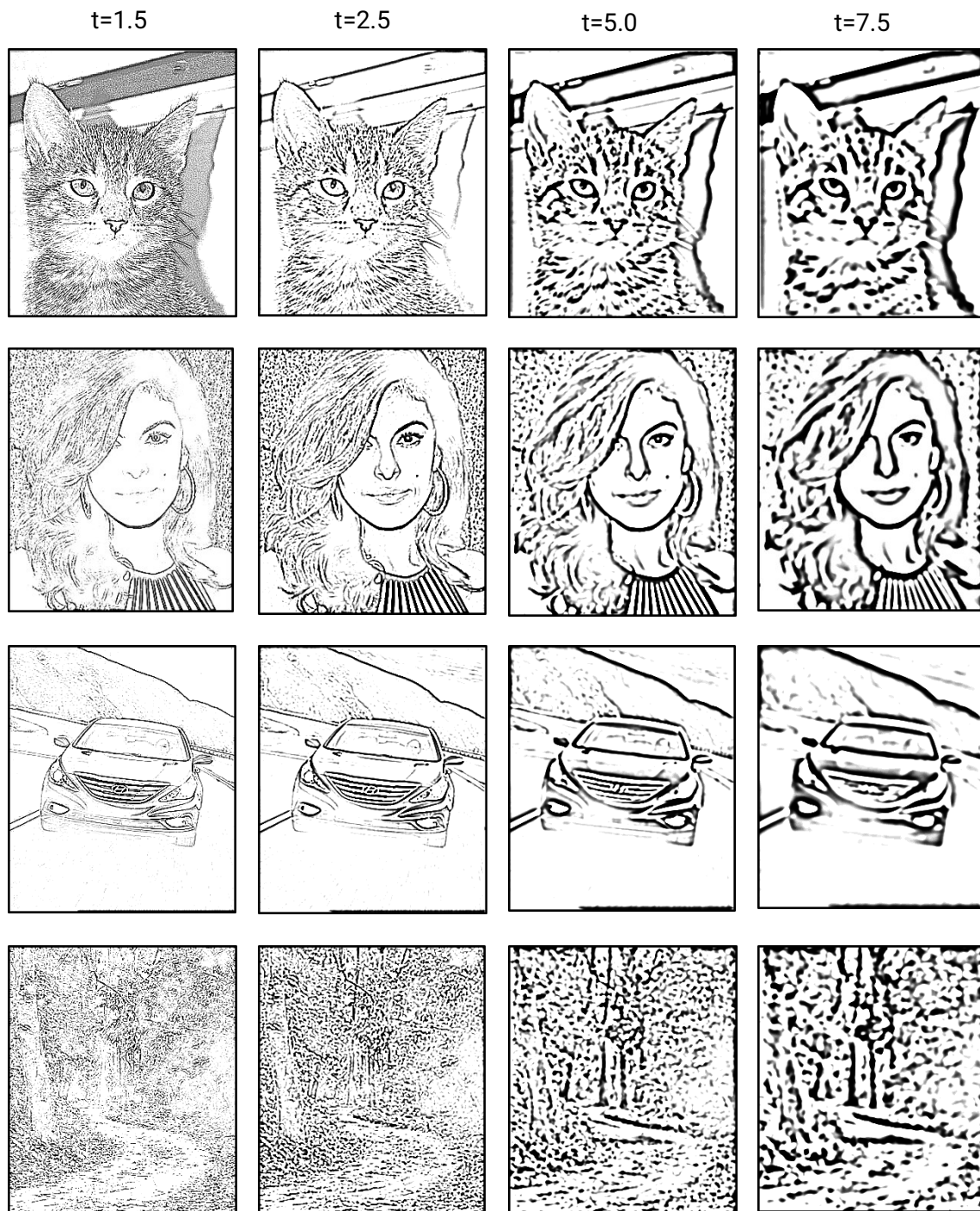


Figura 3.9: Variación del nivel de abstracción visual en diversas imágenes.

# Capítulo 4

## Aplicación en Realidad Virtual

En este capítulo se presenta la revisión de una tecnología que ha ganado creciente ímpetu en los últimos años, la realidad virtual. Se muestra el desarrollo reciente de aplicaciones para baja visión en este tipo de tecnología y se propone la utilización de un casco de realidad virtual acoplado a un dispositivo móvil para el desarrollo de una aplicación propia dirigida a personas con baja visión, utilizando una técnica de mejoramiento del contraste. Se muestra el desarrollo preliminar de esta aplicación, que incluye la implementación de un *shader*, pieza de código sobre la GPU que controla el renderizado de gráficos.

### 4.1. Dispositivos de Realidad Virtual

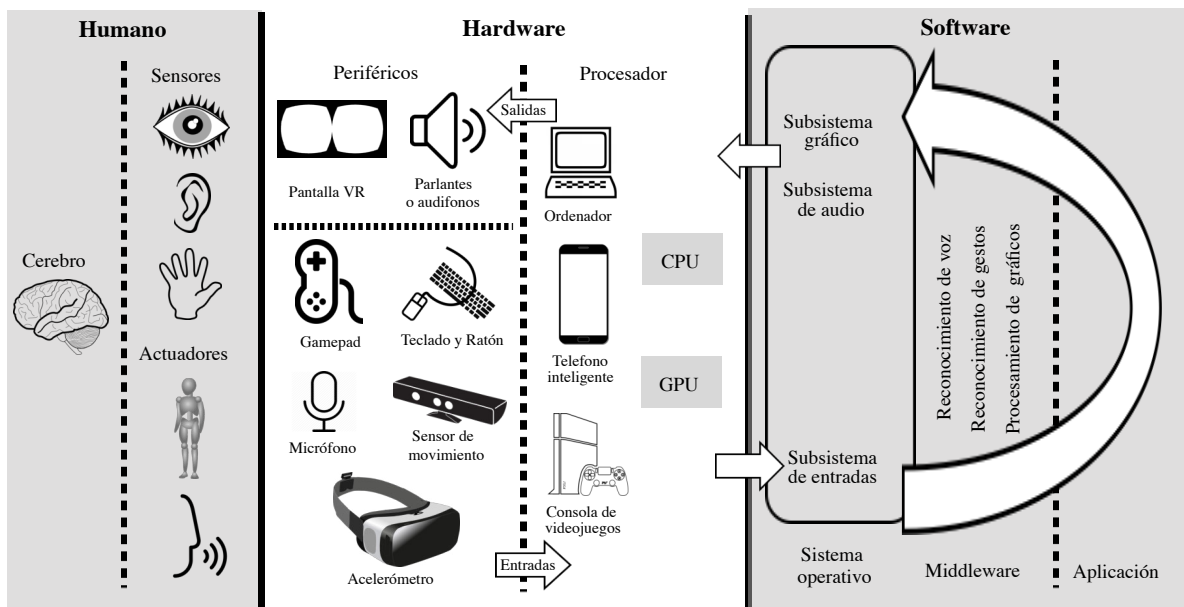
La realidad virtual (VR por sus siglas en inglés) hace referencia a la tecnología que utiliza software y hardware interactivo para generar imágenes realistas, con el objetivo de replicar un entorno real o generar uno imaginario. Se define como “la ilusión de la participación en un ambiente sintético en lugar de la observación externa de dicho ambiente” [55].

Una persona que utiliza un equipo de realidad virtual es capaz de mirar alrededor dentro del mundo artificial, e incluso moverse e interactuar con objetos virtuales colocados en el ambiente. El hardware de un dispositivo de realidad virtual consiste en un casco o marco montado sobre la cabeza de tal forma que una pantalla se coloque en frente de los ojos. Esta pantalla proyecta imágenes estereoscópicas en tres dimensiones. Adicionalmente puede contar con sensores de rastreo del movimiento de la mano y/o cuerpo, un sistema de sonido binaural así como de sensores hápticos que permiten la interacción mediante el tacto. El procesamiento de la información de los sensores y periféricos de salida se realiza en un ordenador, dispositivo móvil o consola de videojuegos.



**Figura 4.1:** Visualización de una escena artificial en una pantalla de visión estereoscópica utilizada en realidad virtual.

La figura 4.2 muestra los principales componentes de un sistema de realidad virtual.








**Figura 4.2:** Componentes de un sistema de realidad virtual.

El desarrollo de sistemas de realidad virtual no es nuevo, ya que se remonta a la presentación del *Sensorama* en 1962, un simulador multisensorial que incluía la visualización de imágenes, simulación de la sensación de viento, sonido e incluso olor, que presentaba todas las características de un sistema de realidad virtual moderno, excepto en que no era interactivo. Posteriormente la necesidad de tener simuladores de vuelo le dio un impulso al desarrollo de los gráficos en realidad virtual, pero no

fue sino hasta hace pocos años, con el aumento en las capacidades de cómputo de los ordenadores personales y dispositivos móviles, que se empezó a masificar esta tecnología, siendo actualmente ofrecida por multitud de compañías. La tabla 4.1 muestra una selección de dispositivos de realidad que han sido liberados en el 2016, con sus principales especificaciones. Entre las aplicaciones donde se explora el uso de la realidad virtual se encuentran:

- Operación remota en ambientes peligrosos. Esto se hace mediante la teleoperación, en donde el trabajador controla un avatar robótico desde un lugar seguro.
- Visualización científica. Provee herramientas para la visualización de datos dinámicos tridimensionales, dando la posibilidad de interactuar con el modelo computacional del fenómeno de estudio, para ofrecer nuevas soluciones que de otra forma serían costosas computacionalmente.
- Visualización arquitectónica. Provee una forma de simular la sensación de espacio dentro de un diseño arquitectónico.
- Diseño. Las herramientas de diseño como Autocad trabajan con dispositivos de entrada en 2D, lo cual dificulta la creación de modelos en 3D, lo cual se solucionaría si el diseñador utiliza un entorno de realidad virtual.
- Educación y entrenamiento. Los simuladores de vuelo y de vehículos militares, así como la simulación de procedimientos quirúrgicos son ejemplos de aplicaciones de la realidad virtual en esta categoría.
- Exploración espacial. Por medio de telepresencia, los geólogos podrían ser capaces de analizar remotamente la superficie de un planeta con los datos de un vehículo no tripulado.
- Entretenimiento. Incluye principalmente videojuegos y películas inmersivas. Es el área que está dando más impulso a la realidad virtual en la actualidad. La masificación de productos de entretenimiento permite la reducción de los costos de esta tecnología.
- Tratamientos médicos. Algunas patologías como el trastorno de ansiedad están siendo tratadas con terapias donde el paciente se adentra en la realidad virtual.

En este trabajo se hace uso de un dispositivo Gear VR, desarrollado en conjunto entre Samsung y Oculus, y cuyo centro de procesamiento y pantalla lo provee un teléfono inteligente Galaxy acoplado al casco. Se busca desarrollar una aplicación que permita mejorar la percepción en personas con baja visión, en concordancia con el desarrollo tecnológico reciente, y basado en el procesamiento de imágenes presentado en los capítulos precedentes. En la siguiente sección se hace una revisión de los desarrollos comerciales para baja visión en realidad virtual, con el fin de establecer oportunidades de innovación en este ámbito.

|   |  |  |  |  |  |
|---|---|--|---|---|---|
| <b>Nombre del dispositivo</b>                 | PlayStation VR  | Oculus Rift  | HTC Vive  | Gear VR   | Daydream View   |
| <b>Material</b>                               | Plástico  | Plástico / Fábrica   | Plástico  | Plástico  | Fábrica   |
| <b>Estación base</b>                          | PS4 (consola)   | PC   | PC  | Galaxy (smartphone)   | Daydream (smartphone)   |
| <b>Rastreo posicional</b>                     | ✓   | ✓  | ✓   | X   | X   |
| <b>Controles de movimiento</b>                | ✓   | ✓  | ✓   | X   | X   |
| <b>Rastreo en 360 grados</b>                  | ✓   | ✓  | ✓   | X   | X   |
| <b>Cámara incorporada</b>                     | X   | X  | ✓   | ✓   | X   |
| <b>Soporte de gamepad/joystick</b>            | ✓   | ✓  | ✓   | ✓   | X   |
| <b>Resolución de pantalla / ojo</b>           | 1080x960  | 1200x1080  | 1200x1080   | 1440x1280   | 1440x1280   |
| <b>Tipo de pantalla</b>                       | OLED  | OLED   | OLED  | AMOLED  | AMOLED  |
| <b>Campo visual</b>                           | 100°  | 110°   | 110°  | ~101°   | No disponible   |
| <b>Frecuencia de actualización</b>            | 90-120Hz  | 90Hz   | 90Hz  | 60Hz  | 60Hz  |
| <b>Inhalámbrico</b>                           | X   | X  | X   | ✓   | ✓   |
| <b>Soporte de uso de lentes</b>               | ✓   | ✓  | ✓   | ✓   | ✓   |
| <b>Ajuste de foco</b>                         | X   | X  | X   | ✓   | X   |
| <b>Ajuste de distancia a los ojos</b>         | ✓   | X  | ✓   | X   | X   |
| <b>Audífonos incorporados</b>                 | X   | ✓  | X   | X   | X   |
| <b>Software</b>                               | PS4   | Windows/Oculus   | Windows/Vibe  | Android/Oculus  | Android/Daydream  |
| <b>Fecha de lanzamiento</b>                   | Octubre 2016  | Marzo 2016   | Abril 2016  | Agosto 2016   | Noviembre 2016  |
| <b>Precio (USD) del casco + estación base</b> | \$800   | ~ \$1400   | ~ \$1600  | \$650   | \$730   |

**Tabla 4.1:** Comparación de los dispositivos de realidad virtual en auge en 2016. Datos obtenidos de [56].

## 4.2. Aplicaciones para baja visión en Realidad Virtual

Las nuevas tecnologías para asistir a personas baja visión, se pueden clasificar en dos grupos: las que mejoran la imagen y las que la transforman en otro tipo de información sensorial como señal táctil o acústica (estos más enfocados a personas con pérdida total de la visión). Haciendo énfasis en el primer grupo, los dispositivos de realidad virtual (VR) y aumentada (AR) comienzan a ser utilizados en aplicaciones comerciales (ver tabla 4.2) y de investigación [57].

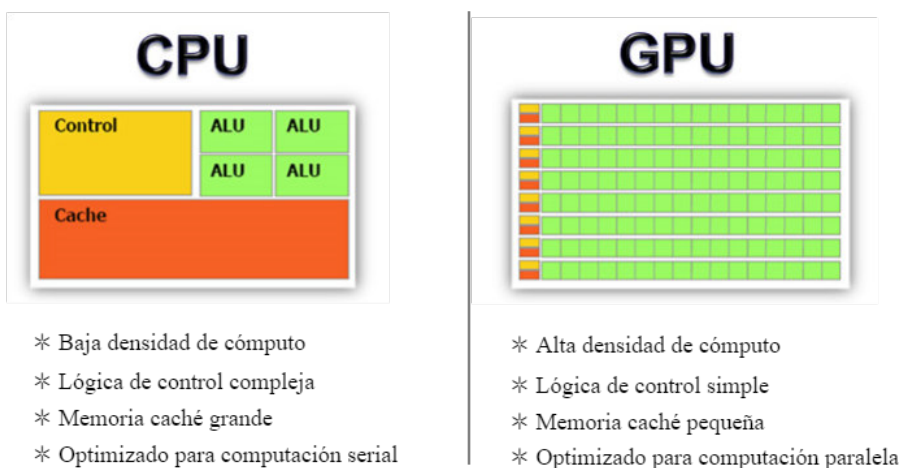
| Producto                   | Descripción   | Características técnicas                             |
|----------------------------|---|--|
| eSight[58]                 | Mejoramiento general de la imagen. Ajuste de color, contraste, brillo y magnificación (hasta 14x). Captura y almacenamiento de fotos.   | Hardware dedicado.                                   |
| IrisVision (Visionize)[59] | Dirigido a pacientes con degeneración macular. Magnifica el centro del campo visual y permite el control del contraste, el brillo, la forma y la posición del magnificador.   | Usa teléfono Samsung S6 + Gear VR.                   |
| Vivid Vision[60]           | Herramienta de rehabilitación, dirigida a profesionales de la visión. Permite un diagnóstico interactivo de la dominancia del ojo, ángulo de desviación, sensibilidad del contraste. Permite almacenar y analizar datos diagnósticos. | Usa Oculus Rift y rastreo del movimiento de la mano. |
| Ox Sight[61]               | Dirigido a personas con visión túnel y discapacidad visual severa. Simplifica la escena al convertirla en sombras y formas de tal manera que el usuario pueda distinguir objetos dentro de su ambiente físico.                        | Hardware dedicado.                                   |

**Tabla 4.2:** *Tecnologías de realidad virtual y aumentada enfocadas a baja visión.*

Estas aplicaciones no están utilizando un procesamiento complejo de la imagen ni tienen en cuenta las transformadas diseñadas para baja visión que se mencionaron en los capítulos precedentes. Tampoco se ha hecho un esfuerzo por aplicar este tipo de transformadas en la GPU de dispositivos móviles, por lo cual el trabajo propuesto puede constituir un aporte significativo y original en dos sentidos, tanto en el procesamiento en paralelo de imágenes digitales para dispositivos móviles como en la asistencia a personas con baja visión. En la siguiente sección se exponen los conceptos básicos de la programación de gráficos en la GPU de dispositivos móviles.

### 4.3. Programación GPU móvil por medio de *shaders*

La GPU (unidad de procesamiento gráfico, por sus siglas en inglés) de un dispositivo móvil es un co-procesador dedicado, diseñado con el fin de acelerar aplicaciones gráficas, interfaces de usuario y contenidos 3D de los teléfonos inteligentes. Su arquitectura es de instrucción única, múltiples datos (SIMD, *single instruction-multiple data*) para procesar cargas de trabajo en forma paralelizada. El procesamiento de gráficos en 3D es uno de los mejores ejemplos de procesamiento en paralelo, ya que permite el procesamiento de millones de píxeles/vértices u operaciones de punto flotante (GFLOPS) por segundo. La unidad básica de cada GPU la constituyen los *shaders* (unidades SIMD) que procesan de forma independiente vértices, primitivas y fragmentos (píxeles). La figura 4.3 muestra algunas de las diferencias entre las arquitecturas de la CPU y la GPU. Cada una tiene sus ventajas y desventajas y deben usarse en conjunto para alcanzar una solución óptima a aplicaciones específicas.



**Figura 4.3:** Comparación de las arquitecturas de la CPU y la GPU. Imagen adaptada de [62].

La interfaz de programación (API) más utilizada para el desarrollo de *shaders* en dispositivos móviles es *OpenGL ES* del grupo *Khronos* y se puede utilizar en la mayoría de los teléfonos inteligentes a través de los sistemas operativos de los mismos, incluyendo Android, iOS y Windows. Otras interfaces incluyen OpenVG (también del grupo *Khronos*, optimizado para gráficos en 2D), CG (Desarrollado en conjunto por nVidia y Microsoft, y usado por el motor gráfico Unity), HLSL (Desarrollado por Microsoft para aplicaciones en DirectX).

Muchos problemas computacionales incluyendo procesamiento de imágenes, cálculos matemáticos y otros algoritmos paralelos se pueden resolver de manera eficaz

en una arquitectura GPU SIMD. También es posible hacer una división de las tareas entre varios núcleos de la GPU con el fin de mejorar el rendimiento. Entre los problemas que están siendo atacados, especialmente en el desarrollo de gráficos para móviles con el uso de procesamiento en la GPU están:

- Realidad virtual/aumentada: La realidad aumentada sobrepone imágenes sobre lo que se percibe del mundo real, mientras que la realidad virtual despliega toda la escena sobre la retina, dando una sensación de inmersión en otra realidad. La generación de estas imágenes se puede ver ampliamente optimizada por la GPU.
- Extracción de características: Un aspecto vital para muchos algoritmos de visión por computadora, los descriptores de puntos de interés como SIFT y SURF son ejemplos de algoritmos que pueden ser paralelizados en la GPU.
- Sistemas de asistencia al conductor: Con la extracción de características y la detección de bordes, se puede generar información adicional de ayuda a los conductores.
- Seguridad y vigilancia: Incluye el reconocimiento facial y de objetos con fines de seguridad.
- Procesamiento del movimiento: Interfaces naturales de usuario tales como reconocimiento de gestos pueden ser optimizadas en la GPU.
- Fusión de sensores: Combina el procesamiento de la visión con otros sensores presentes en los dispositivos móviles tales como acelerómetros, infrarrojos, datos de ubicación, entre otros, con el fin de resolver problemas más complejos.

A este conjunto de problemas, el presente trabajo busca sumar la asistencia a personas con baja visión.

## 4.4. Aplicación para baja visión

### 4.4.1. Diseño

Una propuesta de aplicación para baja visión en realidad virtual se presentará en la conferencia Vision2017, de la *International Society for Low Vision Research and Rehabilitation (ISLRR)*, a llevarse a cabo en Junio de 2017 en los países bajos, bajo el título: *Real-time image enhancement in virtual reality applications for low vision people* (Mejoramiento de la imagen en tiempo real en aplicaciones de realidad virtual para



personas con baja visión). En el resumen de la propuesta, que se presenta a continuación, se pueden ver los alcances de la misma:

“En los últimos años, la tecnología de realidad virtual ha experimentado un impulso en su asequibilidad, y un creciente número de aplicaciones ha surgido, proponiendo nuevo contenido visual inmersivo en 360 grados. Para hacer este contenido accesible a personas con baja visión, se deberían adoptar las mismas estrategias usadas en pantallas tradicionales, esto es, usar métodos de mejoramiento de imágenes dedicado para facilitar su interpretación. Este trabajo introduce una aplicación de realidad virtual para dispositivos móviles que implementa contenido visual mejorado en tiempo real. Está implementado como una tarea de búsqueda visual en un conjunto de imágenes estáticas de 360 grados: el usuario inmerso puede manipular los parámetros del algoritmo de mejoramiento en una forma intuitiva, usando un controlador externo. En particular, nos enfocamos en la técnica propuesta por Peli et al (IOVS, 1991), la cual se basa en un filtro adaptativo que controla el contraste local como una función de la luminancia media local de una imagen. Se ha mostrado que dicha transformada mejora las tasas de reconocimiento en pacientes con pérdida visual moderada, escotoma central o cataratas. Nuestra aplicación es, en nuestro conocimiento, el primer intento para evaluar el impacto de este tipo de mejoramiento de la imagen en un entorno inmersivo de realidad virtual. Específicamente, nuestro sistema permite el ajuste en tiempo real de la transformada y provee todos los datos cuantitativos para analizar a posteriori el comportamiento de los usuarios en como los parámetros afectan el rendimiento de la tarea de búsqueda visual. Diseñado como un juego, se percibe como una experiencia más agradable que los procedimientos oftalmológicos tradicionales. Más generalmente, este aporte puede constituir una manera en que los pacientes con baja visión ajusten el mejoramiento a sus necesidades diarias en aplicaciones de realidad virtual, y también con fines de entretenimiento.”

En la propuesta anterior, la tarea más significativa consistía en implementar el algoritmo de Peli en tiempo real. Se consideró inicialmente apoyarse en el uso de librerías de imágenes como OpenCV, pero fue descartado debido a la falta de compatibilidad con el motor gráfico utilizado en el desarrollo de la aplicación, Unity 3D. Además el tamaño de la pantalla del dispositivo móvil supondría un problema por el limitado poder de la CPU móvil. Por consiguiente, aprovechando que Unity cuenta con la posibilidad de crear y modificar *shaders* sobre la GPU, se decidió utilizar esta alternativa, la cual se describe en la siguiente sección.

#### 4.4.2. Implementando del algoritmo de Peli por medio de *shaders*

Como un ejemplo de aplicación de procesamiento sobre la GPU de un dispositivo móvil, se implementó un algoritmo de mejoramiento de la imagen, adaptable en tiempo real, haciendo uso de la programación de *shaders* sobre el motor gráfico Unity 3D, compilado para un dispositivo Android.

En la plataforma gráfica de Unity, los *shaders* se escriben en el lenguaje CG, y existen diversos efectos visuales disponibles para usarse dentro de las propiedades de la cámara. Entre estos efectos, el más útil para la implementación del algoritmo de Peli resulta ser el de Blur o suavizado de la imagen, que es equivalente al filtrado pasa-bajas. Modificando el *shader* de Blur para convertirlo en la transformación de mejoramiento adaptativo de Peli, se llega a la siguiente ecuación, implementada dentro del *shader* que renderiza fragmentos (o píxeles):

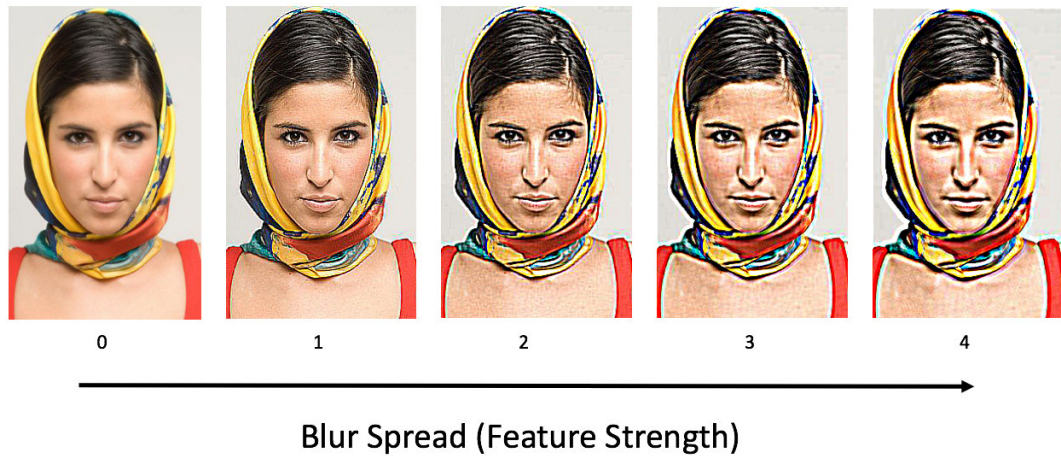
$$g = k * (f - f_L) + m + n * (f_L - m); \quad (4.1)$$

donde  $g$  es la salida del *shader*,  $f_L$  es la salida del *shader* de fragmento del blur,  $k = 5$  es la escala para el rango dinámico,  $m$  es la media de la imagen y  $n = 0.9$  es el factor usado para ajustar el histograma del componente pasa-bajas.

Finalmente, para controlar el resultado, se controla un parámetro del blur original, el tamaño del kernel. Un tamaño pequeño del mismo se traduce en características finas o débiles, mientras que un tamaño grande representa en la imagen final características fuertes, o mayor abstracción visual. Dependiendo de la agudeza visual del paciente, diferentes niveles de abstracción permiten reconocer las formas de diferentes maneras. El hecho de usar procesamiento en la GPU hace responder al sistema en tiempo real, incluso para las imágenes de gran resolución que se manejan en realidad virtual, como se observa en la implementación mostrada en [63].

#### 4.4.3. Implementación en Unity

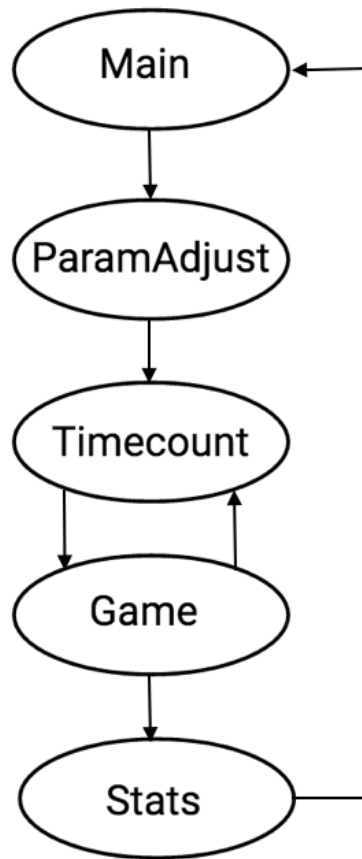
Un demo de la implementación del juego para baja visión se construyó utilizando la plataforma Unity y el *shader* descrito en la sección anterior. Este demo se construyó como una aplicación de Android para el teléfono Galaxy acoplado al Gear VR, con el apoyo de *scripts* escritos en el lenguaje C#. Consiste en una serie de escenas estáticas donde diversos elementos deben ser encontrados, colocando un cursor fijo a la vista, que se mueve siguiendo la rotación de la cabeza. Las imágenes fueron obtenidas por medio de una cámara 360 (Samsung Gear 360). Las escenas que conforman el juego son las siguientes:



**Figura 4.4:** Resultados de la implementación de *shaders* para reproducir el algoritmo de Peli, para diferentes tamaños de *kernel*. La imagen con tamaño de *kernel* igual a cero corresponde a la original.

- *Main*. Es la escena de bienvenida del juego, contiene un menú para entrar al juego y las instrucciones.
- *ParamAdjust*. En esta escena se ajusta el parámetro de mejoramiento de la imagen, implementada como un *shader* superpuesto cámara principal del juego.
- *TimeCount*. Muestra un conteo regresivo hacia la siguiente escena.
- *Game*. Esta escena muestra diversas imágenes según el valor del nivel que se esté jugando. Si se encuentra el objeto de búsqueda, se continúa hacia el siguiente nivel, hasta que se agota el número de niveles.
- *Stats*. Es la escena de salida del juego, donde se muestran datos del rendimiento del jugador y se da la opción de reiniciar el juego.

Un diagrama del flujo de las escenas del juego se presenta en la figura 4.5.



**Figura 4.5:** *Flujo de las escenas de la aplicación en Unity.*

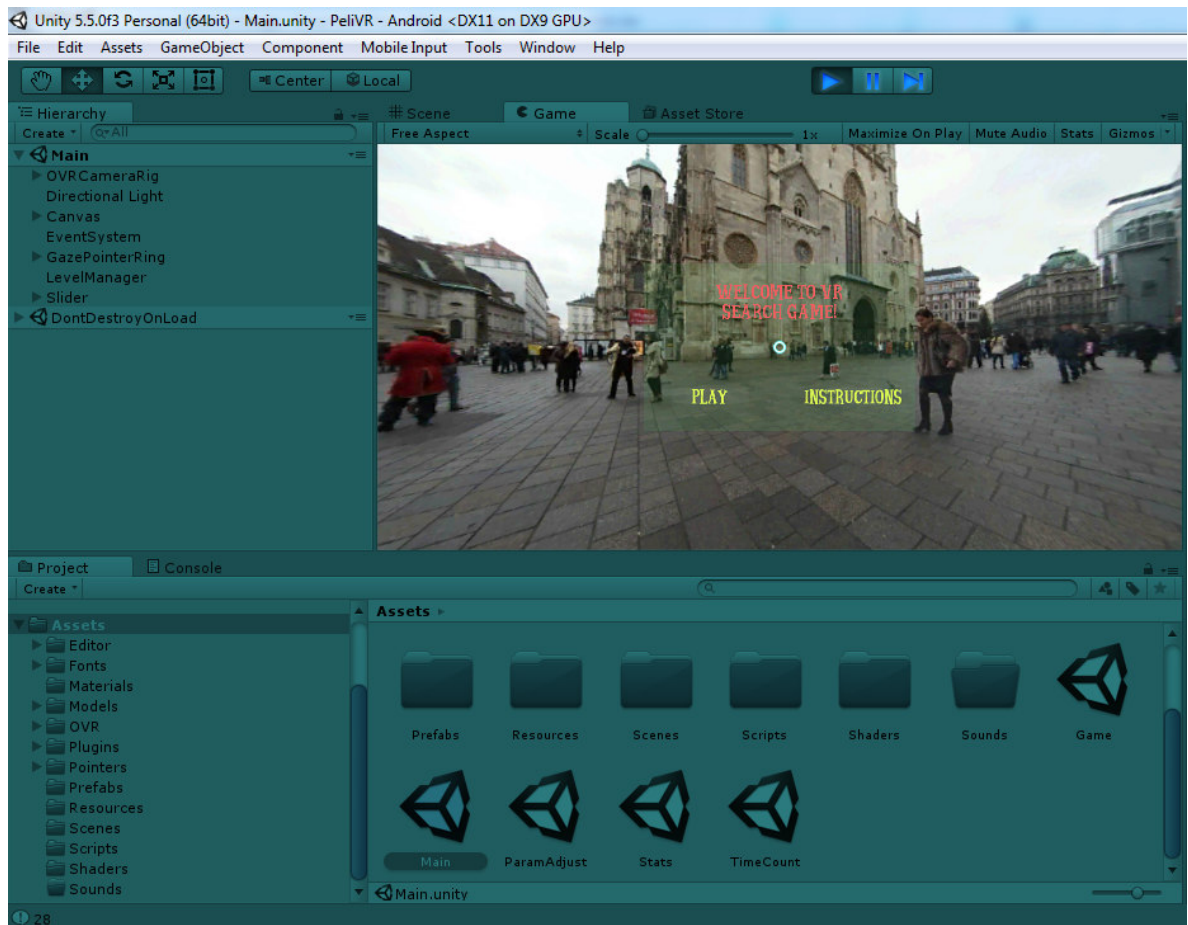


Figura 4.6: Captura de pantalla del desarrollo del demo en Unity.

# Capítulo 5

## Conclusiones

En esta tesis se presentó una serie de trabajos cuyo eje central es el concepto de la abstracción visual, su representación y control de forma computacional y su posible aplicación en el desarrollo de ayudas tecnológicas para pacientes con baja visión.

El primer aporte fue la participación en un proyecto multidisciplinario cuyo objetivo es decodificar la respuesta neuronal de la retina ante diversos estímulos visuales. Dicha participación consistía en diseñar un conjunto de imágenes de estímulo, procesadas por medio de técnicas de imágenes dirigidas a personas con baja visión. Para llevar a cabo este trabajo se realizó una revisión de los conceptos biológicos asociados con la activación neuronal de la retina ante un estímulo luminoso y su relación con el enfoque computacional de detección de características espaciales en la imagen digital. Además se eligieron e implementaron tres transformaciones de mejoramiento de imágenes cuyo impacto ya ha sido medido a nivel psicofísico en pacientes con baja visión, con el fin de establecer una relación más precisa entre la respuesta neuronal y la perceptual. Estas transformaciones fueron aplicadas a un conjunto de imágenes, de acuerdo a un protocolo experimental previamente establecido, y enviadas al Departamento de Neurociencia de la Universidad de Newcastle, donde actualmente están siendo usadas como entrada del experimento.

El segundo aporte surgió como consecuencia de una observación derivada del primer trabajo. Se observó que había una relación entre la frecuencia espacial y la tasa de reconocimiento de la retina ante distintos estímulos que no estaba siendo explotada de forma adecuada desde el punto de vista computacional. Además, teniendo presente que el principal indicador de pérdida en la visión, la agudeza visual, se define en términos de la frecuencia espacial, se buscó una estrategia computacional que permitiera un control del contenido frecuencial de la imagen, sin afectar la naturalidad de la misma. Se le llamó a esto control de la abstracción de la imagen, ya que se puede elegir el nivel de detalle de la respuesta. El algoritmo utilizado fue el de Diferencia de Gaussianas, donde

se propuso tomar una relación fija entre las desviaciones estándar de dos funciones gaussianas de igual magnitud, y luego reparametrizar dichas desviaciones en término del grosor deseado de las características a resaltar de la imagen. Los resultados muestran un control efectivo de la abstracción visual, con resultados estéticamente agradables, aunque hace falta realizar un estudio más profundo de las implicaciones para pacientes con baja visión.

El último aporte consistió en participar en el diseño de una aplicación en realidad virtual dirigida a personas con baja visión, en conjunto con investigadores de la Universidad de Génova (Italia) e INRIA-Sophia Antipolis (Francia) y que incluye el procesamiento de imágenes con el algoritmo de mejoramiento adaptativo de Peli. Esta aplicación fue admitida para ser presentada la conferencia Vision2017, de la *International Society for Low Vision Research and Rehabilitation (ISLRR)*, en Junio de 2017, en los Países Bajos. Se implementó el algoritmo de Peli como un *shader* que se ejecuta en la GPU de un dispositivo móvil acoplado a un casco de realidad virtual. Este trabajo representa uno de los primeros intentos para aplicar transformaciones de imágenes para bajao visión tanto en la GPU móvil como en un entorno de realidad virtual.

## 5.1. Trabajos futuros

Como continuación de los aportes presentados en esta tesis, se proponen los siguientes trabajos a futuro:

- La evaluación de algoritmo propuesto de control de la abstracción visual a nivel psicofísico y neuronal.
- La implementación de diversos algoritmos para baja visión en la GPU de dispositivos móviles, por ejemplo algoritmos de mejoramiento de la imagen, reposicionamiento de la escena, lectura y remapeo de la información.
- El desarrollo de aplicaciones en realidad virtual o aumentada dirigidos a personas con baja visión y adaptables a diversas patologías.

# Bibliografía

- [1] C. Enroth-Cugell and J. G. Robson, “The contrast sensitivity of retinal ganglion cells of the cat,” *The Journal of physiology*, vol. 187, no. 3, pp. 517–552, 1966.
- [2] “Visual impairment and blindness. Fact Sheet N 282,” *World Health Organization*, Aug 2014.
- [3] C. Lok, “Curing blindness: vision quest,” *Nature News*, Sep 10 2014.
- [4] G. Portelli, J. M. Barrett, G. Hilgen, T. Masquelier, A. Maccione, S. Di Marco, L. Berdondini, P. Kornprobst, and E. Sernagor, “Rank order coding: a retinal information decoding strategy revealed by large-scale multielectrode array retinal recordings,” *eneuro*, vol. 3, no. 3, pp. eNEURO–0134, 2016.
- [5] J. S. Wolffsohn and R. C. Peterson, “A review of current knowledge on electronic vision enhancement systems for the visually impaired,” *Ophthalmic and Physiological Optics*, vol. 23, no. 1, pp. 35–42, 2003.
- [6] H. Moshtael, T. Aslam, I. Underwood, and B. Dhillon, “High tech aids low vision: a review of image processing for the visually impaired,” *Translational vision science & technology*, vol. 4, no. 4, pp. 6–6, 2015.
- [7] J. M. Wensveen, H. E. Bedell, and D. S. Loshin, “Reading rates with artificial central scotomata with and without spatial remapping of print,” *Optometry & Vision Science*, vol. 72, no. 2, pp. 100–114, 1995.
- [8] A.-C. Scherlen and V. Gautier, “A new concept for visual aids:” visar” visual signal adaptive restitution,” in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pp. 1976–1979, IEEE, 2006.
- [9] W. I. Al-Atabany, T. Tong, and P. A. Degenaar, “Improved content aware scene retargeting for retinitis pigmentosa patients,” *Biomedical engineering online*, vol. 9, no. 1, p. 52, 2010.



- 
- [10] W. Atabany and P. Degenaar, “A robust edge enhancement approach for low vision patients using scene simplification,” in *Biomedical Engineering Conference, 2008. CIBEC 2008. Cairo International*, pp. 1–4, IEEE, 2008.
- [11] C. Bordier, J. Petra, C. Dauxerre, F. Vital-Durand, and K. Knoblauch, “Influence of background on image recognition in normal vision and age-related macular degeneration,” *Ophthalmic and Physiological Optics*, vol. 31, no. 3, pp. 203–215, 2011.
- [12] W. I. Al-Atabany, M. A. Memon, S. M. Downes, and P. A. Degenaar, “Designing and testing scene enhancement algorithms for patients with retina degenerative disorders,” *Biomedical engineering online*, vol. 9, no. 1, p. 27, 2010.
- [13] E. Peli, R. B. Goldstein, G. M. Young, C. L. Trempe, and S. M. Buzney, “Image enhancement for the visually impaired. simulations and experimental results.,” *Investigative Ophthalmology & Visual Science*, vol. 32, no. 8, pp. 2337–2350, 1991.
- [14] E. M. Fine and E. Peli, “Enhancement of text for the visually impaired,” *JOSA A*, vol. 12, no. 7, pp. 1439–1447, 1995.
- [15] E. Peli, “Recognition performance and perceived quality of video enhanced for the visually impaired,” *Ophthalmic and Physiological Optics*, vol. 25, no. 6, pp. 543–555, 2005.
- [16] M. Everingham, B. Thomas, and T. Troscianko, “Head-mounted mobility aid for low vision using scene classification techniques,” *Int. J. of Virt. Reality*, vol. 3, no. 4, 1999.
- [17] H. Winnemöller, J. E. Kyprianidis, and S. C. Olsen, “Xdog: an extended difference-of-gaussians compendium including advanced image stylization,” *Computers & Graphics*, vol. 36, no. 6, pp. 740–753, 2012.
- [18] E. Addo, O. A. Bamiro, and R. Siwale, “Anatomy of the eye and common diseases affecting the eye,” in *Ocular Drug Delivery: Advances, Challenges and Applications*, pp. 11–25, Springer, 2016.
- [19] S. Reception, “Human vision: Structure and function of the human eye-vol. 27,” *Encyclopaedia Britannica*, p. 179, 1987.
- [20] W. Merigan, A. Freeman, and S. P. Meyers, “Parallel processing streams in human visual cortex,” *Neuroreport*, vol. 8, no. 18, pp. 3985–3991, 1997.
- [21] J. J. Nassi and E. M. Callaway, “Parallel processing strategies of the primate visual system,” *Nature Reviews Neuroscience*, vol. 10, no. 5, pp. 360–372, 2009.
- [22] H. Wässle, “Parallel processing in the mammalian retina,” *Nature Reviews Neuroscience*, vol. 5, no. 10, pp. 747–757, 2004.

- 
- [23] G. D. Field and E. Chichilnisky, “Information processing in the primate retina: circuitry and coding,” *Annu. Rev. Neurosci.*, vol. 30, pp. 1–30, 2007.
- [24] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat,” *Journal of neurophysiology*, vol. 28, no. 2, pp. 229–289, 1965.
- [25] T. Lindeberg, “A computational theory of visual receptive fields,” *Biological cybernetics*, vol. 107, no. 6, pp. 589–635, 2013.
- [26] D. Marr and E. Hildreth, “Theory of edge detection,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.
- [27] T. Gollisch and M. Meister, “Rapid neural coding in the retina with relative spike latencies,” *science*, vol. 319, no. 5866, pp. 1108–1111, 2008.
- [28] S. Remtulla and P. Hallett, “A schematic eye for the mouse, and comparisons with the rat,” *Vision research*, vol. 25, no. 1, pp. 21–31, 1985.
- [29] R. Acharya, Y. E. Ng, and J. S. Suri, *Image modeling of the human eye*. Artech House, 2008.
- [30] H. Winnemöller, S. C. Olsen, and B. Gooch, “Real-time video abstraction,” in *ACM Transactions On Graphics (TOG)*, vol. 25, pp. 1221–1226, ACM, 2006.
- [31] T. Peli and J. S. Lim, “Adaptive filtering for image enhancement,” *Optical Engineering*, vol. 21, no. 1, pp. 211108–211108, 1982.
- [32] S. J. Leat and M. Mei, “Custom-devised and generic digital enhancement of images for people with maculopathy,” *Ophthalmic and Physiological Optics*, vol. 29, no. 4, pp. 397–415, 2009.
- [33] M. Fullerton and E. Peli, “Digital enhancement of television signals for people with visual impairments: Evaluation of a consumer product,” *Journal of the Society for Information Display*, vol. 16, no. 3, pp. 493–500, 2008.
- [34] M. Fullerton, R. L. Woods, F. A. Vera-Diaz, and E. Peli, “Measuring perceived video quality of mpeg enhancement by people with impaired vision,” *JOSA A*, vol. 24, no. 12, pp. B174–B187, 2007.
- [35] P. Satgunam, R. L. Woods, G. Luo, P. M. Bronstad, Z. Reynolds, C. Ramachandra, B. W. Mel, and E. Peli, “Effects of contour enhancement on low-vision preference and visual search,” *Optometry and vision science: official publication of the American Academy of Optometry*, vol. 89, no. 9, p. E1364, 2012.
- [36] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.

- 
- [37] G. M. Johnson and M. D. Fairchild, “Sharpness rules,” in *Color and Imaging Conference*, vol. 2000, pp. 24–30, Society for Imaging Science and Technology, 2000.
- [38] H. Winnemöller, *Perceptually-motivated non-photorealistic graphics*. PhD thesis, Citeseer, 2006.
- [39] S. Paris and F. Durand, “A fast approximation of the bilateral filter using a signal processing approach,” *Computer Vision–ECCV 2006*, pp. 568–580, 2006.
- [40] R. A. Young, “The gaussian derivative model for spatial vision: I. retinal mechanisms,” *Spatial vision*, vol. 2, no. 4, pp. 273–293, 1987.
- [41] H. Winnemöller, “Xdog: advanced image stylization with extended difference-of-gaussians,” in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, pp. 147–156, ACM, 2011.
- [42] G. Luo, P. Satgunam, and E. Peli, “Visual search performance of patients with vision impairment: effect of jpeg image enhancement,” *Ophthalmic and Physiological Optics*, vol. 32, no. 5, pp. 421–428, 2012.
- [43] E. Wiecek, M. L. Jackson, S. C. Dakin, and P. Bex, “Visual search with image modification in age-related macular degeneration visual search in amd,” *Investigative ophthalmology & visual science*, vol. 53, no. 10, pp. 6600–6609, 2012.
- [44] H. Nothdurft, “Texture segmentation and pop-out from orientation contrast,” *Vision research*, vol. 31, no. 6, pp. 1073–1078, 1991.
- [45] R. B. Ivry and A. Cohen, “Asymmetry in visual search for targets defined by differences in movement speed,” *JOURNAL OF EXPERIMENTAL PSYCHOLOGY HUMAN PERCEPTION AND PERFORMANCE*, vol. 18, pp. 1045–1045, 1992.
- [46] J. R. Cavanaugh, W. Bair, and J. A. Movshon, “Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons,” *Journal of neurophysiology*, vol. 88, no. 5, pp. 2530–2546, 2002.
- [47] N. Senthilkumaran and R. Rajesh, “Edge detection techniques for image segmentation—a survey of soft computing approaches,” *International journal of recent trends in engineering*, vol. 1, no. 2, pp. 250–254, 2009.
- [48] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.
- [49] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

- [50] T. Lindeberg, "Scale-space theory: A basic tool for analyzing structures at different scales," *Journal of applied statistics*, vol. 21, no. 1-2, pp. 225–270, 1994.
- [51] A. Patiño, "Silvercat - xdog with adaptive threshold." <https://vimeo.com/170214816>. Accedido el 31-03-2017.
- [52] R. A. Haddad and A. N. Akansu, "A class of fast gaussian binomial filters for speech and image processing," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 723–727, 1991.
- [53] R. Brunner, K. Chen, and J. Harper, "Gaussian blur approximation suitable for gpu," July 8 2008. US Patent 7,397,964.
- [54] S. C. Olsen, H. Winnemöller, and B. Gooch, "Implementing real-time video abstraction," in *ACM SIGGRAPH 2006 Sketches*, p. 133, ACM, 2006.
- [55] R. A. Earnshaw, *Virtual reality systems*. Academic press, 2014.
- [56] W. Shanklin, "2016 vr comparison guide.." <http://newatlas.com/best-vr-headsets-comparison-2016/45984/>. Accedido el 31-03-2017.
- [57] A. Diniz-Filho, E. R. Boer, C. P. Gracitelli, R. Y. Abe, N. van Driel, Z. Yang, and F. A. Medeiros, "Evaluation of postural control in patients with glaucoma using a virtual reality environment," *Ophthalmology*, vol. 122, no. 6, pp. 1131–1138, 2015.
- [58] "Página principal de esight." <http://www.esighteyewear.com/>. Accedido el 31-03-2017.
- [59] "Página principal de visionize." <http://www.visionizellc.com/>. Accedido el 31-03-2017.
- [60] "Página principal de vivid vision." <http://www.seevividly.com/>. Accedido el 31-03-2017.
- [61] "Página principal de ox sight." <http://oxsight.co/>. Accedido el 31-03-2017.
- [62] B. Tao, "Understand the mobile graphics processing unit." <http://embedded-computing.com/articles/understand-the-mobile-graphics-processing-unit/>. Accedido el 31-03-2017.
- [63] A. Patiño, "Peli transform in unity.." <https://vimeo.com/210280947>. Accedido el 31-03-2017.