

TESIS DEFENDIDA POR

Marco Antonio Sánchez Domínguez

Y aprobada por el siguiente comité:

Dr. Gustavo Olague Caballero

Director del Comité

M.C. José Luis Briseño Cervantes

Miembro del Comité

Dr. David Hilario Covarrubias Rosales

Miembro del Comité

Dr. Heriberto Márquez Becerra

Miembro del Comité

Dr. Hugo Homero Hidalgo Silva

*Coordinador del Programa en
Ciencias de la Computación*

Dr. David Hilario Covarrubias Rosales

*Director de Estudios
de Posgrado*

23 de Septiembre del 2010

TESIS

presentada por

Marco Antonio Sánchez Domínguez

para cubrir parcialmente los requisitos necesarios para
obtener el grado de **Maestro en Ciencias** en el

Centro de Investigación Científica y de Educación Superior de Ensenada

Ciencias de la Computación

Diseño de un Modelo Evolutivo de Atención Visual para la Clasificación de Objetos en Escenas Naturales Complejas

Fecha: Septiembre del 2010.

Miembros del comité: Dr. Gustavo Olague Caballero.
M.C. José Luis Briseño Cervantes.
Dr. Heriberto Márquez Becerra.
Dr. David Hilario Covarrubias Rosales

Tesis realizada en las instalaciones del Departamento de Ciencias de la Computación del
CICESE, bajo la dirección del Dr. Gustavo Olague.

RESUMEN de la tesis de **Marco Antonio Sánchez Domínguez**, presentada como requisito parcial para obtener el grado de MAESTRO EN CIENCIAS en CIENCIAS DE LA COMPUTACIÓN. Ensenada, B. C. Septiembre del 2010.

Diseño de un Modelo Evolutivo de Atención Visual para la Clasificación de Objetos en Escenas Naturales Complejas

Resumen aprobado por:

Dr. Gustavo Olague Caballero

Director de Tesis

Uno de los tópicos más investigados por la gente de visión por computadora es el reconocimiento de objetos. Sin embargo, a pesar de los avances que se han conseguido en este campo, todavía se considera de gran dificultad para la visión artificial el llevar a cabo esta tarea considerando escenas naturales. Así, las técnicas típicas de visión por computadora utilizadas para resolver este problema, se basan en el escaneo de la imagen completa para detectar objetos dentro de la escena. No obstante, el sistema visual humano separa este proceso en dos etapas: la primera llamada flujo dorsal, la cual esta basada en el mecanismo de la atención visual y se encarga de encontrar la ubicación de los objetos en la escena; y la segunda, donde el flujo ventral se encarga de la identificación del objeto una vez que la ubicación se ha realizado. Este trabajo propone un diseño novedoso para reconocer y clasificar objetos en imágenes naturales complejas emulando el proceso de visión humana. Esto se logra a través del uso de la programación genética, la cual evoluciona los parámetros del modelo de atención visual, a fin de encontrar las regiones de las imágenes que contienen los objetos pertenecientes a la misma clase. De forma paralela a este proceso, se busca optimizar el reconocimiento, evolucionando la forma de obtener el descriptor asociado a la región de interés. Esto se realiza mediante la evolución de un operador que transforma la imagen original a un espacio de parámetros y de la cual se obtiene su histograma de orientaciones. De esta forma las contribuciones se pueden enumerar como sigue: primero, se propone una implementación del modelo estandar inspirada de la ruta dorsal o ruta del ¿donde?. Segundo, nuestra metodología propone un enfoque novedoso el cual optimiza las diferentes etapas por medio de programación genética. Finalmente, para lograr dicha optimización se propone un criterio en base a un problema de reconocimiento de objetos. Nuestros resultados experimentales confirman las conclusiones de nuestro trabajo y demuestran que los resultados son comparables con aquellos del estado del arte.

Palabras clave: Atención Visual, Reconocimiento de Objetos, Búsqueda Visual, Programación Genética.

ABSTRACT of the thesis presented by **Marco Antonio Sánchez Domínguez**, as a partial requirement to obtain the MASTER SCIENCE degree in COMPUTER SCIENCES. Ensenada, B. C. september 2010.

Design of a Model of Visual Attention for the Classification of Objects in Complex Natural Scenes

Abstract approved by:

Dr. Gustavo Olague Caballero

Thesis director

One of the biggest research topics investigated by people dedicated to the study of computer vision is object recognition. However, despite of the great progress that has been achieved in this area, the solution of such tasks in the case of complex natural scenes is still considered as very difficult to be achieved; not only for computer vision systems, but for the case of untrained humans. Nevertheless, the human visual system separates such processes in two stages: the first stage called dorsal stream, which is based on the visual attention mechanism that tries to focus on the location of the object within the scene; and the second, called the ventral stream where the information attempts to identify the object once the location has been computed. This work proposes a novel model that recognize and classify objects considering complex natural images using as inspiration the human visual system. This is achieved through the use of genetic programming that modifies the parameters of the visual attention model in order to find regions of the images containing objects of the same class. In parallel to this process, we look for an optimal object recognition process that is attained by evolving the best algorithm, which computes a suitable interest region descriptor. This was achieved through the evolution of operators that transform the original image into a parameter space from which we obtain its orientation histograms. In this way, our contributions can be enumerated as follows: first, we propose an implementation of the standard model inspired from the dorsal stream or also called the "where pathway". Second, our proposed methodology gives a novel approach that optimizes all different stages through genetic programming. Finally, to achieve such optima it is defined a criterium based on a formulation of an object recognition problem. The experimental results confirm the conclusions of the work and show, that the results are comparable with those of the state-of-the-art.

Keywords: Visual Attention, Object Recognition, Visual Search, Genetic Programming.

Dedicado a mis hermanos

Güicho
y
Rafa

Agradecimientos

A Dios.

A mis padres por su eterno apoyo y creer en mi.

A mis hermanos por mostrarme el camino.

A mis cuñadas y sobrinos con cariño.

A Cinthy por quien pretendo ser mejor persona.

A todas las amistades que han formado parte de mi vida.

A mi asesor y gran maestro, Dr. Gustavo Olague Caballero.

A mi comité de tesis por su gran apoyo e interés en este trabajo

M.C. José Luis Briseño Cervantes.

Dr. Heriberto Márquez Becerra. y

Dr. David Hilario Covarrubias Rosales.

Al Centro de Investigación Científica y de Educación
Superior de Ensenada.

Al Consejo Nacional de Ciencia y Tecnología.

Tabla de Contenido

Capítulo	Página
Resumen	ii
Abstract	iv
Agradecimientos	vi
Lista de Figuras	ix
Lista de Tablas	xi
I Introducción	1
I.1 Conceptos básicos	4
I.1.1 ¿Qué es la atención?	4
I.1.2 Concepto de Atención Visual	5
I.2 Objetivo de la tesis	11
I.3 Objetivos Particulares	11
I.4 Organización de la tesis	12
II Marco Teórico	14
II.1 Reconocimiento de patrones	14
II.1.1 Reconocimiento y clasificación de objetos	18
II.2 Histograma de orientación sobre la imagen transformada	20
II.3 Máquina de soporte vectorial	22
II.4 Computación evolutiva	26
II.4.1 Programación genética	29
II.5 Trabajos previos	32
III Proceso Psicobiológico de la Atención Visual	41
III.1 Teoría de la integración de características	41
III.2 Teoría de la coherencia y ceguera al cambio	44
III.3 Organización neuronal del cerebro	45
III.4 Componentes fisiológicos involucrados en la atención visual y el re- conocimiento	48
III.4.1 La retina	54
III.4.2 Núcleo geniculado lateral (NLG)	58
III.4.3 La corteza visual	60
III.4.4 Corteza parietal	62
III.4.5 Corteza inferotemporal	64
III.4.6 Corteza prefrontal	65
III.5 Conclusiones	66
IV El Modelo de Atención Visual	67
IV.1 Cálculo preatentivo de las características visuales básicas	69
IV.1.1 Color	69
IV.1.2 Intensidad	71

Tabla de Contenido (Continuación)

Capítulo	Página
IV.1.3 Orientación	72
IV.1.4 Simetría	74
IV.2 Obtención de los mapas característicos	75
IV.3 Obtención de los mapas de notoriedad	77
IV.4 Mapas de Características Sobresalientes	78
IV.5 Selección de la región de interés	79
IV.6 Conclusión	82
V Implementación del Modelo	83
V.1 Proceso para el reconocimiento de clases de objetos	83
V.1.1 Adquisición de los datos	85
V.1.2 Selección de la región de interés	86
V.1.3 Extracción de características	88
V.1.4 Reconocimiento y clasificación de objetos	89
V.1.5 Optimización de la clasificación de objetos utilizando progra- mación genética	90
VI Resultados experimentales	97
VI.1 Herramientas de trabajo	97
VI.1.1 Unidad central de proceso	97
VI.1.2 Software	97
VI.2 Experimentos	98
VI.2.1 Prueba del algoritmo propuesto	98
VII Conclusiones y trabajo futuro	106
VII.1 Conclusiones	106
VII.1.1 Conclusiones del reconocimiento de objetos	106
VII.1.2 Conclusiones del sistema implementado	106
VII.1.3 Trabajo a futuro	107
BIBLIOGRAFIA	109

Lista de Figuras

Figura		Página
1	Ejemplo de atención ascendente	7
2	Cuadro utilizado por Yarbus para medir la trayectoria de los Ojos y sus resultados	8
3	Diagrama simplificado del proceso de reconocimiento de patrones . . .	15
4	Diferentes tipos de transformaciones sobre la misma imagen y su nueva representación después de aplicarle una ponderación gaussiana.	22
5	Obtención del histograma de orientación	23
6	Ejemplo de una Máquina de Soporte Vectorial	25
7	Proceso de cruzamiento entre dos árboles	31
8	Diagrama de Flujo de la Programación Genética	33
9	Modelo propuesto por Milanese	34
10	Modelo propuesto por Osberger	35
11	Modelo propuesto por Koch y Ullman	36
12	Ejemplo de la atención focalizada en una escena	42
13	Representación gráfica de la teoría de integración de características . .	43
14	Ejemplo de búsqueda preatentiva y búsqueda conjunta	44
15	Lobulos cerebrales	46
16	Diagrama funcional del sistema visual primario	49
17	Vista lateral y superficial del cerebro	50
18	Vista de corte coronal del cerebro	51
19	Mecanismos neuronales para el control de la atención	52
20	Rutas del procesamiento de la percepción visual	53
21	Anatomía del ojo humano	54
22	Estructura celular de la retina	56
23	Dibujo de un campo receptivo de una célula ganglionar	57
24	Campo receptivo de una célula simple en la corteza visual	61
25	Modelo de atención visual propuesto	68
26	Células concéntricas de oponencia simple	70
27	Ejemplo de células horizontales presentes en la corteza visual primaria .	73
28	Representación de una red neuronal del tipo el ganador lo toma todo .	80
29	Diferentes selecciones de la región de interés	81
30	Trayectoria realizada para la clasificación de objetos	85
31	Imagen con cuatro diferentes tipos de clases utilizadas en el diseño de nuestro algoritmo	86
32	Ejemplo de selección de la región de interés	88
33	Ejemplo de una selección de región de interés y una posible transformación de la imagen.	89

Lista de Figuras (Continuación)

Figura		Página
34	Diagrama de flujo de la fase de aprendizaje del programa genético implementado	91
35	Ejemplo del cromosoma implementado	92
36	Composición del cromosoma	92
37	Proceso de cruzamiento en el sistema implementado	95
38	Mejores individuos por cada ejecución del programa para realizar la clasificación de carros	99
39	Mejor individuo y promedio de la función de desempeño por cada generación para la clasificación de carros	100
40	Mejor individuo elegido para el reconocimiento y clasificación de carros	101
41	Ejemplo del operador de transformación de la imagen y de selección de región por el individuo ganador	102
42	Gráfica que muestra el porcentaje de intersección entre los objetos segmentados de la clase carro y el protoobjeto encontrado por nuestro sistema.	103
43	Imágenes de la clase carro con las que se entrenó el GP	104
44	Imágenes de la clase distractores con las que se entrenó el GP	105

Lista de Tablas

Tabla		Página
I	Comparativa entre Atencion Bottom - Up y Top - Down	9
II	Principales funciones de cada lobulo del cerebro	48
III	Parámetros del programa genético	96
IV	Matriz de confusión con los resultados obtenidos del mejor individuo en la clasificación de carros.	102

Capítulo I

Introducción

Imagina un escenario en donde te encuentres en una calle transitada la cual estas viendo por primera vez. Existe una gran cantidad de información visual bombardeando tus retinas, semáforos cambiando de color, anuncios publicitarios, señales de tránsito, peatones cruzando las calles, carros circulando por las vialidades, etc. ¿Cómo es que podemos procesar toda esta información?, y al mismo tiempo ser capaces de formar un modelo mental de la escena visual donde todo parece ser percibido de forma clara. Esta interrogante ha llevado a investigadores de diversas áreas a estudiar el proceso de la percepción visual humana.

El escenario imaginado muestra un ejemplo claro de lo complejo de la percepción visual humana. Una riqueza de información es percibida a cada momento, siendo ésta mucho mayor de la que puede ser procesada eficientemente por el cerebro. Estas limitantes fisiológicas provocan la necesidad de una clase de selección de lo más relevante del entorno. En general, el mecanismo que nos hace a los humanos tan efectivos en los actos de la vida diaria, es la habilidad para extraer la información más importante en etapas tempranas del procesamiento y enfocarse en atender solo esta información (Corbetta, 2002).

En visión al proceso de enfocarse solamente en la información más relevante se le llama atención visual selectiva. Aquí la información importante de la escena es extraída y entonces por fin dirigida a las áreas altas del cerebro, principalmente a la

corteza visual en donde tienen lugar procesos complejos que se enfocan en resolver el reconocimiento de objetos.

Como mencionamos, biológicamente el restringir los procesos perceptuales involucrados en la visión humana a un limitado subconjunto de datos sensoriales es lo que hace eficiente su funcionamiento. Trasladando este mismo contexto pero ahora a la visión por computadora se puede intuir que el beneficio es equivalente; utilizar un modelo de atención visual sin duda reduce el coste computacional de los procesos inherentes en las tareas de visión de alto nivel (Tsotsos, 1987).

Los humanos detectamos y reconocemos objetos importantes presentes en nuestra escena de una manera rápida y con poco esfuerzo consciente. Esta habilidad que hemos desarrollado forma parte de nuestra adaptación al medio, biológicamente podemos ver a la atención como una adaptación a nuestra estructura anatómica (Rizzolatti y Riggio, 1994), por otra parte psicológicamente, la atención puede verse como una condición aprendida que llevamos a cabo a lo largo de nuestra vida (Olson y Chun, 2001). Por ejemplo, personas que desarrollan actividades distintas, su atención no necesariamente se enfocará hacia los mismos objetos o áreas de la escena. Así para dos personas que ingresan a una casa, una dedicada a la decoración de interiores y la otra a la arquitectura; sin duda, presentarán atracción hacia distintas áreas de la casa. En el siguiente capítulo discutiremos un poco sobre el aprendizaje y la evolución de la atención. Lo que podemos afirmar es que los humanos aprendemos a atender la información visual que nos interesa.

Computacionalmente, el dotar a una máquina de la información necesaria para

aprender a enfocarse en la información importante, resulta de gran interés en la resolución de tareas como el reconocimiento y clasificación de objetos (Privitera y Stark, 2000). En el caso particular de la clasificación de objetos en escenas naturales donde la cantidad de información es excesiva, sigue siendo una tarea ardua para los investigadores del área de visión artificial, que a pesar de los grandes avances que se han conseguido, factores como la diversidad de tamaños, posiciones, y formas que los objetos pueden tomar, además del problema de iluminación y semiocultamiento que el objeto pueda tener siguen siendo aspectos que dificultan esta tarea.

Es por eso que a lo largo de este trabajo se abordará una manera novedosa para atacar el problema de clasificación tomando bases biológicamente plausibles, se parte de la evolución de un modelo de atención visual, donde el algoritmo sea capaz de enfocarse en fragmentos de la imagen donde se encuentra el objeto a clasificar. Para esto se han aumentado el número de características que ayudan a guiar nuestra atención a esas zonas, además de confinar al evolutivo la forma de combinar la diferente información para crear nuestro modelo. Al mismo tiempo hemos evolucionado la forma de obtener el descriptor del área donde se encuentra el objeto de interés, lo realizamos basandonos en el histograma de orientación de la imagen transformada. Dicha transformación se logra a través de la generación de un operador que se le aplica a la imagen, la forma de obtener este operador se lleva a cabo mediante el mismo programa genético, la utilización del descriptor de histogramas de orientación ha demostrado gran eficacia a cambios de intensidad, iluminación y escala, problemas muy frecuentes en objetos de escenas naturales (Cui y Hasler, 2009). Es por eso que con la implementación de este modelo ofrecemos un diseño que resulta robusto en el problema de clasificar objetos en escenas naturales complejas.

I.1 Conceptos básicos

En esta sección, discutiremos algunos conceptos importantes de la atención visual y de las herramientas computacionales utilizadas para llevar a cabo la clasificación de objetos. Primero definiremos lo que es la atención, los tipos de atención y la búsqueda visual, como segunda parte trataremos los conceptos teóricos relacionados con el reconocimiento de objetos, los cuales nos serán de gran ayuda para entender la implementación y funcionamiento del algoritmo implementado.

I.1.1 ¿Qué es la atención?

El concepto de atención se refiere a la habilidad de una persona o sistema para responder a los aspectos esenciales de una tarea u objetivo pasando por alto aquellos que le son irrelevantes. Se considera como una cualidad que ostenta la percepción y que hace las veces de filtro de los estímulos ambientales, evaluando cuales son los más importantes y atribuyéndoles una prioridad para luego recibir un procesamiento. Nosotros usualmente tenemos la impresión de conservar una basta representación del mundo y que cambios importantes en nuestro ambiente atraerían nuestra atención. Sin embargo, varios experimentos han revelado que nuestra habilidad para detectar cambios está altamente delimitado (Rensink, 2000), básicamente por el hecho de que nuestro organismo solo puede procesar un número reducido de estímulos perceptuales a cada momento.

La atención por lo tanto es un mecanismo muy importante para los seres vivos, ya que controla y regula los procesos perceptuales actuando como un filtro sobre los estímulos ambientales y decidiendo cuales son los más importantes dándoles prioridad para ser atendidos. El funcionamiento de la atención ha sido estudiado ya desde hace

varias décadas por investigadores de diversas disciplinas científicas como neurólogos, fisiólogos, psicólogos y en las últimas tres décadas por gente dedicada a la visión por computadora, quienes ven a la atención y en particular a la atención visual como una forma factible para disminuir la complejidad de diversos problemas en su campo de estudio.

I.1.2 Concepto de Atención Visual

Nos referimos a la atención visual como un proceso selectivo que habilitamos los seres vivos de manera efectiva para poder dirigir nuestra mirada hacia objetos de interés en nuestro ambiente visual (Itti, 2003). En términos físicos se puede referir como un haz de luz que ilumina un objeto o lugar concreto dentro de la escena visual para favorecer el tratamiento de la imagen por parte de la corteza visual. Aunque la definición es superficial podemos intuir que el proceso inicia con la recepción de un reflejo de luz emitida por un objeto y culmina con su reconocimiento en alguna parte del cerebro. Es por eso que la atención visual juega un rol esencial en el proceso de la percepción visual, el cual produce una dicotomía, donde una parte es responsable de seleccionar la región de interés, y otra de investigar más acerca de la región seleccionada (Neisser, 1967). Los mecanismos involucrados en la primera tarea son llamados preatentivos mientras los mecanismos que operan en la zona seleccionada son llamados atentivos, psicológicamente y fisiológicamente estas dos áreas han sido estudiadas como un proceso indispensable para entender la percepción visual y serán analizadas con mayor detalle en el siguiente capítulo.

Aunque fisiológicamente el proceso de la atención visual es necesario. Computacionalmente ha sido desarrollado para aplicarlo a diversas áreas de la visión artificial,

donde se ha utilizado para ayudar en tareas de alto nivel como es la clasificación de objetos, interpretación de imágenes médicas y como consecuencia en otras disciplinas que buscan controlar e interpretar el movimiento de los ojos con el fin de poder guiar la atención de los observadores, tal es el caso de los desarrolladores de videojuegos o los diseñadores de páginas web.

Atención ascendente (BU) contra atención descendente (TD)

Uno de los principales cuestionamientos en este tema, es saber ¿cuál información de una escena es la más relevante?. Simplemente no existe una respuesta general sino que depende de los conocimientos del observador y de la tarea a realizar, cuando no se tiene una meta específica sino simplemente nos encontramos explorando el escenario, aquello que sobresale por sus características físicas con respecto al ambiente es lo que llamara nuestra atención. Por ejemplo, una única manzana verde encima de una docena de naranjas atraería nuestra atención debido al color y al contraste que éste provoca con respecto a su ambiente (ver figura 1). A este tipo de atención se le llama descendente - ascendente (llamaremos BU por su acrónimo en ingles Bottom-Up). Por lo tanto la selección atencional BU esta determinada únicamente por las propiedades físicas de la imagen de entrada, este mecanismo atencional es también llamado externo, automático o reflexivo (Egeth y Yantis, 1967).

Por otra parte, existe otro tipo de atención la cual está influenciada por procesos cognitivos, por ejemplo, si estuviéramos buscando una naranja para comer, estaríamos concentrados en localizar cosas de color naranja y redondas, lo que haría dirigir nuestra mirada hacia regiones distintas que con la atención BU. En este caso a este tipo de atención se le llama descendente (llamaremos TD por su acrónimo en ingles Top-Down)



Figura 1: Ejemplo de atención ascendente

está controlada por el estado mental del sujeto, esto significa que proviene de las áreas altas del cerebro tales como el conocimiento, expectativas y metas (Corbetta, 2002).

Uno de los primeros experimentos para diferenciar estos dos mecanismos fue realizado por Yarbus (1967), en donde se le pedía a observadores fijarse en un cuadro, (ver figura 2). Su experimento consistió en registrar los movimientos oculares de los participantes al contemplar el cuadro. Éstos disponían de 3 minutos para observar una escena sobre la que previamente conocían preguntas a las que deberían responder. La idea estaba en que observaran la imagen del cuadro con el objetivo de buscar pistas que les ayudasen a conocer la respuesta. Los resultados no pudieron ser más esclarecedores. A diferentes cuestiones diferentes patrones de exploración ocular (movimiento sacádicos) sobre la obra. Con esto Yarbus demostró que la atención visual no se comporta de la misma forma siempre, sino que depende de las motivaciones u objetivos previos que el observador tenga ante una escena y esto hará decidir las estrategias de observación de las personas. Hoy en día se siguen haciendo experimentos y diseñando modelos que intentan emular completamente los movimientos sacádicos del ojo humano, para trabajos recientes ver (Chikkerur y Poggio, 2009; Elazary y Itti, 2010).

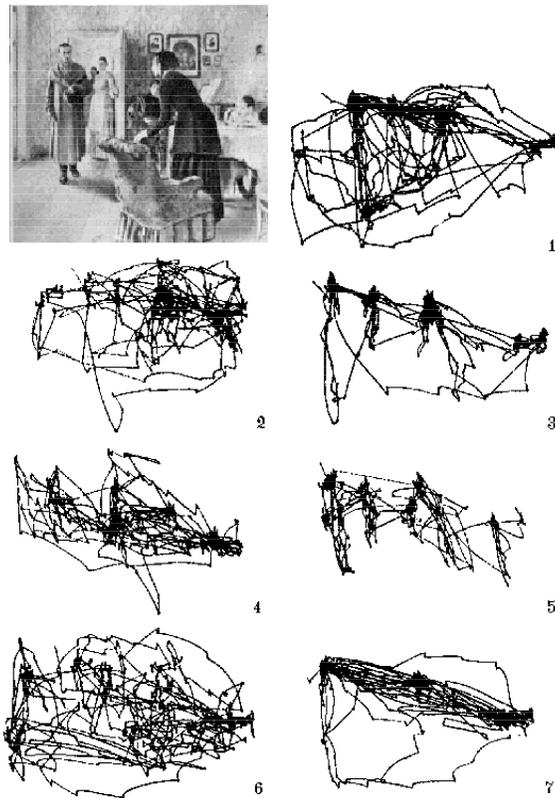


Figura 2: Cuadro utilizado por Yarbus para estudiar la trayectoria de los ojos.

En otras áreas de investigación como la psicología y la fisiología también se han logrado considerables avances. Existen teorías que explican de una manera coherente el proceso de la atención visual, biológicamente se ha llegado a conocer en buena parte del proceso de la atención visual BU, sin embargo el mecanismo TD sigue siendo hasta la fecha en muchos aspectos una incógnita. En el siguiente capítulo abordaremos con mayor detalle la información más importante e indispensable de estas dos áreas para la realización de nuestro trabajo.

Podemos resumir que la atención visual básicamente consiste de dos mecanismos: uno voluntario, llamado "Descendente" o "dependiente de tareas", y uno controlado

por estímulos, conocido como "Ascendente" o proceso "basado en imágenes" (Desimone y Duncan, 1995). Para una comparación entre ambos tipos de atención (ver tabla I). Hay que tener en cuenta que aunque existen otros factores que influyen en el proceso de la atención y el movimiento de los ojos como pueden ser sonidos, sensaciones de tacto, olores, etc. Los alcances y fines de éste trabajo se enfocarán solamente en lo relacionado al puro aspecto visual.

Tabla I: La tabla muestra una comparativa entre los mecanismos de atención

Descendente - Ascendente	Ascendente - Descendente.
Basado en Imagen	Dependiente de Tarea
Originado en la retina	Originado en la Corteza Prefrontal
Involuntario	Voluntario
Toma de 25 a 50 ms	Toma 200 ms
Es tratado antes de saber del escenario	Características visuales ajustadas acorde a la tarea

Búsqueda visual

Una de las principales tareas perceptuales de la atención visual es lo que denominamos la búsqueda visual, la cual se refiere al hecho de hallar un objeto en particular dentro de un ambiente visual. Dicho elemento recibe el nombre de objetivo, mientras que al resto de los objetos presentes en la escena se les conoce como distractores. Todos realizamos búsquedas visuales en nuestras acciones cotidianas, localizar a un amigo entre una multitud, encontrar las llaves dentro de un cuarto o hallar nuestra corbata favorita en un closet son como tales tareas de búsqueda visual. En donde la dificultad para lograr nuestro objetivo depende de varios factores como la concentración, el número de distractores o las propiedades físicas tanto del objetivo como de sus distractores.

Han surgido diversas teorías psicológicas que intentan interpretar el proceso de la búsqueda visual como son: la teoría de la orientación visual propuesta por Posner (1984). Donde se postula que en la búsqueda de un objetivo para lograr nuestra meta, nos vamos orientando a través de pequeños objetivos visuales intermedios e inconscientes. Por otra parte surge la búsqueda guiada donde el principal propulsor es Wolfe (1994). Su modelo consta de algunas etapas parecidas al modelo que adoptamos pero con la diferencia en que él define que se requiere de un desplazamiento de la atención espacial para que el objetivo pueda ser encontrado. Y por último la que presenta mayor empatía con el proceso fisiológico conocida como la Teoría de Integración de Características (TIC) desarrollada por Anne Treisman (1980) y de la cual parte nuestro trabajo, por lo que se tratará a detalle en el siguiente capítulo.

La búsqueda visual puede ser un aspecto importante en problemas de visión por computadora, donde dado un objetivo y una imagen de prueba, lo que se busca es saber si hay una instancia del objetivo en la imagen. En gran parte la dificultad va a depender de que tan compleja sea la imagen, es decir, del número de distractores, del tamaño de la imagen, de las propiedades que presente el objeto y diversos factores externos como ruido, iluminación, etc. La complejidad computacional en la búsqueda visual ha sido investigada en (Tsotsos, 1987), donde mencionan que la búsqueda visual es ilimitada (con los objetivos no dados o sin poder ser utilizados para optimizar la búsqueda). Por ejemplo, si la orden es para encontrar semáforos sin especificarlo, esta demostrado ser NP-Completo¹. Esto es debido al hecho de que todos los subconjuntos de píxeles deben ser considerados para encontrar el objetivo en el peor de los casos.

¹NP-Completo es el subconjunto de los problemas de decisión en NP (la clase NP la forman aquellos problemas cuya solución positiva se puede verificar en un tiempo polinomial dada la información adecuada, o equivalentemente, sus soluciones pueden ser encontradas en un tiempo polinómico en una máquina no determinista) tal que todo problema en NP se puede reducir en cada uno de los problemas de NP-Completo.

En contraste, la búsqueda visual limitada (el objetivo es explícitamente conocidos de antemano) requiere tiempo lineal. Entonces, mediante experimentos psicológicos se demuestra que la naturaleza computacional del problema sugiere que influencias atencionales TD juegan un importante rol durante la búsqueda. El modelo propuesto presenta una combinación de ambas influencias BU y TD, que facilitarán la búsqueda y por ende una mejor clasificación de los objetos en las imágenes.

I.2 Objetivo de la tesis

- Generar una nueva solución para el problema de clasificación de objetos, a través del diseño de un modelo inspirado en la visión humana. Optimizando de forma conjunta la búsqueda y descripción de zonas específicas en la escena, mediante la evolución paralela de los parámetros de un modelo de atención visual y los de un operador que transforma la imagen original para facilitar su descripción.

I.3 Objetivos Particulares

Las contribuciones específicas que se pretenden alcanzar con este trabajo son:

- Una nueva representación del problema de reconocimiento de clases de objetos empleando un modelo de atención visual, un descriptor de regiones basado en el histograma de orientaciones y el uso de programación genética.
- A través del uso de la programación genética, se presenta una solución conjunta para optimizar la búsqueda de regiones de interés y la extracción de características de objetos, con el fin de decidir su pertenencia entre clases conocidas a priori.

I.4 Organización de la tesis

- *Introducción*
- *Marco Teórico.* En el segundo capítulo abordamos todos los conceptos teóricos que nos sirven de base para entender el desarrollo de este trabajo, así mismo se mencionaran los trabajos previos realizados más relacionados con nuestra implementación.
- *Proceso Psicobiológico de la Atención Visual.* En este capítulo se trata lo referente a las teorías psicológicas involucradas en el proceso de la atención visual y por otra parte, se analiza el funcionamiento fisiológico de los componentes neuronales relacionados en el proceso de la visión humana.
- *El Modelo de Atención Visual.* El objetivo de este capítulo es describir la implementación algorítmica del modelo de atención visual; también se presenta una relación biológica en cada paso del modelo de atención visual.
- *Implementación del Modelo.* Aquí procedemos a explicar el diseño de nuestro modelo evolutivo de atención visual para la clasificación de objetos. En el capítulo abordamos todas las áreas relacionadas con la clasificación de objetos, desde la selección de imágenes hasta su clasificación utilizando la SVM. Por último se explica la implementación del algoritmo evolutivo para generar la clasificación de manera automática.
- *Resultados Experimentales.* En esta sección presentamos las herramientas computacionales tanto de software como de hardware que se utilizaron en el diseño del sistema. Se muestran los experimentos realizados y hacemos un análisis detallado de los resultados obtenidos.

- *Conclusiones y Trabajo Futuro.* Por último se muestran las conclusiones y se enuncian los trabajos que pueden realizarse para dar continuidad al presente.

Capítulo II

Marco Teórico

A lo largo de este capítulo se mencionaran los conceptos más importantes que nos servirán para formular y desarrollar nuestro trabajo de tesis.

II.1 Reconocimiento de patrones

El proceso de reconocer y clasificar cosas, como sonidos, comidas, imágenes, sabores, es algo que hacemos de forma inconsciente y que nos permite adaptarnos a nuestro entorno. Esta habilidad natural que tenemos prácticamente todo ser vivo se encuentra rodeado de una gran complejidad al intentar emularlo de manera artificial. La disciplina encargada de emular este proceso se le conoce como reconocimiento de patrones.

Por lo tanto el reconocimiento de patrones tiene como objetivo la clasificación de objetos en un cierto número de categorías o clases. Dependiendo de la aplicación estos objetos pueden verse como imágenes, ondas de señales o cualquier tipo de métrica que necesitan ser clasificadas. Un sistema básico de reconocimiento de patrones consiste de tres pasos (ver figura 3):

1. Adquisición de datos, la cual se realiza a través de un sensor y cuyo objetivo es extraer la información útil, eliminando la información redundante e irrelevante.

2. Extracción de características, este paso tiene como objetivo obtener los atributos de los datos adquiridos que puedan ser utilizados en el proceso de clasificación.

3. Toma de decisiones, en este paso para tomar una decisión es necesario definir los grupos o clases en los que se van a categorizar los diferentes datos de entrada y el punto esencial que es el clasificador, el cual discrimina a través de las características obtenidas de los datos de entrada y se encarga de agruparlos en las categorías definidas.

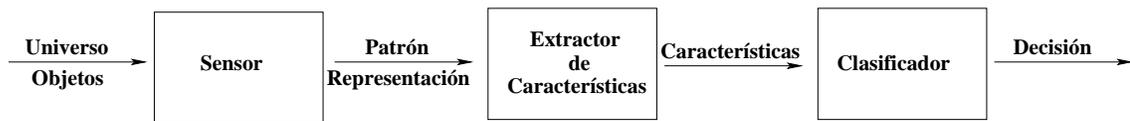


Figura 3: Diagrama simplificado del proceso de reconocimiento de patrones.

En la actualidad existen un gran número de metodologías para resolver cada uno de los pasos anteriores, aunque los dos primeros puntos van a depender mucho del tipo de patrón que se quiera reconocer. Por ejemplo, difícilmente se va a utilizar el mismo algoritmo para describir objetos en imágenes, que para segmentos de música o tonos de voz. En el caso particular de visión por computadora, el sensor y la adquisición de datos vienen dado principalmente por un dispositivo de captura de imágenes, y el proceso de extracción de características se logra con la aplicación de diversos filtros para transformar las imágenes en estructuras que permitan su identificación de forma inequívoca. Así, se pueden distinguir entre dos tipos de descriptores de imágenes, los descriptores de bordes, cuyos objetivo es la identificación de los bordes mediante el ajuste de rectas, curvas, funciones polinómicas o códigos de cadena y los descriptores de región, encaminados a obtener propiedades tales como color, textura, superficie, intensidad, etc.

Sin embargo, el último paso es más genérico, es decir una vez extraída la información

de los patrones para nuestro problema, se le pueden aplicar una gran variedad de sistemas diseñados para resolver el problema de la clasificación de dichos patrones, para lo cual no es importante si se trata de voz, de imágenes o cualquier otra información, ya que todo lo ve como si fueran datos. Estos métodos de clasificación los podemos agrupar dentro de dos principales categorías, dependiendo de si tenemos o no información previa que le permita al sistema ser entrenado y así aprender a clasificar la información.

En el primer caso, cuando no se cuenta con conocimiento a priori de los objetos a clasificar, le llamamos clasificación no supervisada, donde los datos o características que se tienen no se sabe a que clase o categoría pertenece. Entonces, el clasificador busca la forma de agrupar a los objetos que presentan características espacialmente más próximas en la misma clase. Entre los métodos más conocidos de clasificación no supervisada se encuentran: el algoritmo K-means, el cual es un método iterativo que va eligiendo los centros de la clase dependiendo de las muestras o datos que se tengan. El conocido como ISODATA donde se van formando grupos que se ajustan iterativamente usando teoría de probabilidades. Y el Simple Link y Complete Link que parten de grupos unitarios de objetos y van uniendo los grupos más parecidos en cada etapa, hasta cumplir con alguna condición de paro.

El segundo tipo de clasificación es el que más nos interesa por ser el más utilizado en el tipo de problemas que estudiaremos. Se denomina clasificación supervisada y está se presenta cuando contamos con conocimiento a priori de los objetos a clasificar, la tarea por lo tanto se reduce a clasificar un objeto dentro de una categoría que cuenta ya con objetos comunes previamente clasificados. El procedimiento para realizar este tipo de clasificación se puede dividir en dos fases. En la primera fase tenemos un conjunto de datos que nos sirven para construir un modelo o regla general para llevar a cabo

la clasificación. A esta etapa se le conoce como fase de entrenamiento o aprendizaje. En la segunda fase se lleva el proceso de clasificar objetos o muestras de las que se desconoce a la clase que pertenece y con estas poder probar la eficacia del modelo. A esta etapa se le denomina fase de evaluación o prueba. Entre los principales métodos de clasificación supervisada tenemos:

- Métodos estadísticos que se basan en la teoría de probabilidad y estadística, donde se cuenta con un conjunto de medidas numéricas con distribuciones de probabilidad conocidas y donde a partir de ellas se realiza la clasificación.
- Funciones discriminantes donde para el caso de dos clases se busca obtener una función $g(X)$, donde X contiene información perteneciente a un objeto, tal que para un nuevo objeto O , si $g(O) \geq 0$ se asigna a la clase 1 y en otro caso a la 2. Si son múltiples clases se busca un conjunto de funciones $g_i(X)$, tal que el nuevo objeto se ubica en la clase donde la función tome el mayor valor.
- El método de vecino más cercano, donde un nuevo objeto se ubica en la clase donde esté el objeto de la muestra original que más se le parezca.
- Redes neuronales artificiales, denominadas habitualmente RNA. Funcionan imitando a las redes neuronales reales en el desarrollo de tareas de aprendizaje.
- Máquinas de soporte vectorial, es un modelo que representa a los puntos de muestra en el espacio, separando las clases por un espacio lo más amplio posible. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de su proximidad pueden ser clasificadas en una u otra clase.

Entre los dos últimos métodos existen similitudes por lo que algunos autores proponen a la Máquina de Soporte Vectorial (SVM por su acrónimo en inglés Support Vector Machine) como rama de la redes neuronales, pero a pesar de estas similitudes

la SVM se encuentra mejor fundamentada en la teoría y cuenta con mejor capacidad de generalización. En la sección I.1.5 trataremos más ampliamente el tema de las SVM por ser la herramienta de clasificación que se eligió en el diseño propuesto.

II.1.1 Reconocimiento y clasificación de objetos

En visión por computadora el uso del reconocimiento de patrones es de suma utilidad en un gran número de trabajos. Su tarea primordial es el encontrar la forma correcta de reconocer y clasificar objetos en imágenes digitales. Existen dos vertientes, la primera enfocada en reconocer al mismo objeto en varias imágenes y la segunda encargada de reconocer diferentes objetos pero de la misma clase en una o varias imágenes.

La segunda vertiente es la que nos interesa y también es la más difícil de resolver, debido a la dificultad que representa encontrar atributos que nos ayuden a identificar diferentes objetos de una misma clase, por ejemplo supongamos que queremos clasificar carros, aunque por lo general todos los carros presentan algunas características similares como: llantas, puertas, volante, etc. La dificultad aumenta si nos referimos a diferentes modelos, marcas y tipos de autos, debido a que la estructura física del carro cambia drásticamente complicando su clasificación. Esto ha provocado que los investigadores vean el problema de clasificación desde varios metodologías de estudio pero con la misma convicción: encontrar una forma de describir a todos los objetos de una misma categoría de tal forma que los particularice de las demás clases de objetos.

Matemáticamente podemos ver al problema de clasificación como sigue: dada una muestra de entrada $\mathbf{x} = \{x_1, x_2, \dots, x_d\}$ que necesita ser clasificada en uno y sólo uno de los c grupos o clases c_1, c_2, \dots, c_c . La existencia de los grupos se conoce a priori (en el

caso de la clasificación supervisada). El vector de entrada \mathbf{x} representa el descriptor de la imagen y se desconoce la clase a la que pertenece. Supongamos que $y = c_j$ significa que pertenece a la clase c_j . La clasificación se refiere a la relación entre la etiqueta de pertenencia a la clase etiquetada como y y el vector de características \mathbf{x} . El objetivo es estimar la relación $x \rightarrow y$ utilizando los datos de entrenamiento $(x_i, y_i), i = 1, \dots, n$. Esta relación llamada regla de decisión es utilizada para clasificar muestras futuras.

Una definición más sencilla fue propuesta por (Russell y Norvig , 2002), el cual define el reconocimiento de clases de objetos como sigue:

- Dada una escena que contiene uno o más objetos elegidos de una colección de objetos O_1, O_2, \dots, O_n , conocidos a priori, y
- Dada una imagen de la escena tomada desde una posición y orientación desconocida; determinar lo siguiente:
¿Cuáles objetos O_1, O_2, \dots, O_n , están presentes en la escena?.

En conclusión, el reconocimiento y clasificación de objetos embebidos en imágenes digitales es un problema abierto en el campo de visión por computadora. La digitalización de un objeto en una escena real y su posible reconocimiento depende primordialmente de tres factores: las características físicas del sistema de adquisición, los parámetros o características que identifican a los objetos, y por último el clasificador que como su nombre lo indica decide a que categoría debe de pertenecer el objeto. El primer factor depende de la base de datos que se utiliza, por otra parte los puntos dos y tres han sido tratados en las últimas décadas en un sin fin de problemas de reconocimientos de patrones. En las siguientes dos subsecciones hablaremos un poco de los dos métodos elegidos para describir y clasificar los objetos en nuestro sistema.

II.2 Histograma de orientación sobre la imagen transformada

La descripción de las regiones de interés sigue siendo un paradigma actual de la visión por computadora, comúnmente es aplicado para resolver tareas difíciles como la detección y el reconocimiento de objetos, navegación de robots, recuperación de imágenes y minería de datos visuales, por nombrar algunas aplicaciones. En el caso del reconocimiento de objetos la descripción del objeto debe contener información que nos permita tener fiabilidad, es decir, poder tener cambios numéricos pequeños para objetos de la misma clase pero que al mismo tiempo, nos permita discriminar entre diferentes tipos de clases de objetos, sin dejar a un lado la rapidez y economía computacional.

Para poder obtener un descriptor con tales características, el preprocesamiento que se le hace a la imagen juega un papel crucial, debido a que la manipulación que se realice en ella mejora la obtención de información en cierto tipo de características de la imagen, por ejemplo, al aplicar un operador de textura como el de adición de píxeles la imagen procesada estará enfocada en describir regiones y un detector de texturas como la matriz de coocurrencia podría dar buenos resultados. Por otra parte, si se le aplicara a la imagen operadores diferenciales que se encargan de resaltar bordes, un descriptor como el de códigos de cadena que da información de si un punto constituye una línea recta o no, ayudaría a distinguir de mejor manera al objeto. Por lo tanto, el descriptor nos debe permitir identificar rasgos que ayuden a la formación de una hipótesis sobre la presencia de un objeto dado en la escena correspondiente. Los rasgos usados por el sistema dependen del tipo de objetos a ser reconocidos, así como de la estructura del banco de modelos utilizada.

En la literatura existe un gran número de métodos para resolver el problema de la descripción de objetos, en el caso de nuestro trabajo al aplicar primero el modelo de atención visual, estamos reduciendo considerablemente el tamaño de la imagen, por lo tanto hemos optado por dejar que un programa genético construya el modelo de pre-procesamiento óptimo que permita obtener la mejor descripción en esta región (área de atención).

En el caso del descriptor se ha optado por obtener el histograma de orientaciones sobre la imagen transformada. La esencia de dicho algoritmo es que la forma de un objeto en una imagen puede ser descrito por medio de la distribución de los bordes de la imagen procesada. El objetivo principal de esta técnica es la extracción de características en una imagen aplicandole cierta transformación. Donde en la región de interés se calcula la orientación para cada píxel en la imagen. Como un segundo paso se le aplica una ponderación con una ventana gaussiana sobrepuesta en la región para calcular los parches a una escala normalizada ver figura 4. Esta información es entonces dividida en subregiones como una malla cuadrada en orden a calcular el histograma de la orientación, ponderado por la magnitud de la imagen transformada para cada subregión.

Los valores de orientación de los píxeles en la región de interés son calculados de la siguiente forma:

$$\theta(x, y) = \arctan \left(\frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \right), \quad (1)$$

donde L = nueva locación y x, y = posición del píxel central.

Para crear el descriptor se deberá generar una imagen transformada sobre la región

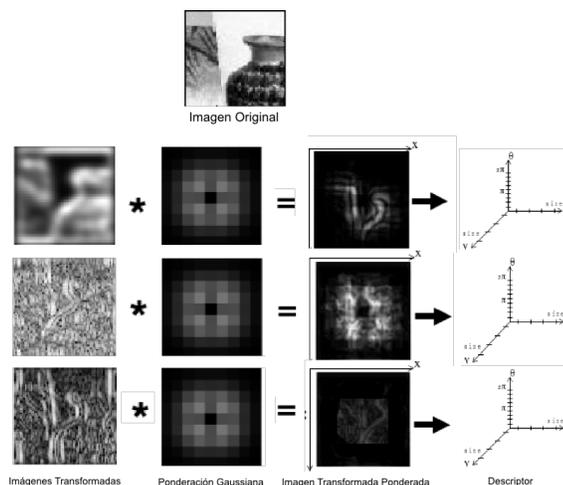


Figura 4: Diferentes tipos de transformaciones sobre la misma imagen y su nueva representación después de aplicarle una ponderación gaussiana.

de interés, la cual constará de una ventana de 4×4 subregiones, de tamaño 2×2 cada una. Así en cada región se generara un histograma de orientación considerando 8 direcciones principales ver figura 5. por lo que se obtendrá un histograma tridimensional de $4 \times 4 \times 8 = 128$ casillas, Lo que se logra con el gradiente es eliminar los cambios de brillo, la normalización de la magnitud del vector eliminará los cambios de contraste y la magnitud de la región nos ayudara a permanecer invariantes al tamaño del objeto. por lo tanto con este método obtenemos un descriptor bastante robusto para el reconocimiento de objetos.

II.3 Máquina de soporte vectorial

Si bien el punto esencial en el proceso de reconocimiento y clasificación de objetos es la descripción que se haga del objeto, la clasificación juega un papel importante y debe de estar íntimamente relacionado con la extracción de rasgos del objeto. Después de

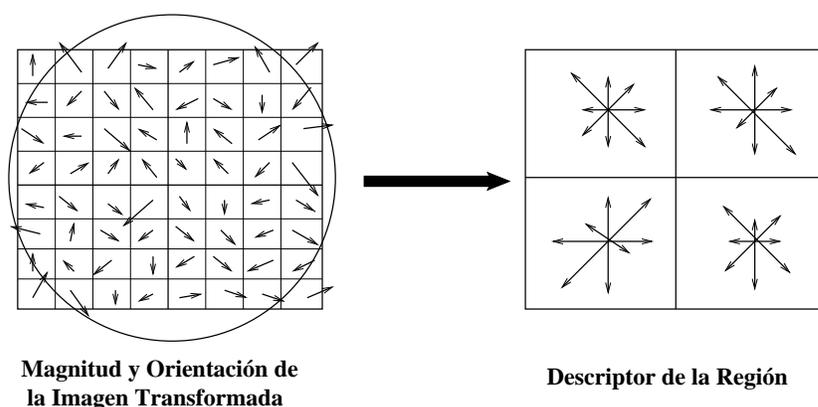


Figura 5: Obtención del histograma de orientación

obtener la descripción del objeto se procede a asignarlo a alguna clase. A lo largo de los años han surgido diferentes métodos de clasificación entre los que tenemos el método de vecinos más cercanos, métodos bayesianos y redes neuronales los cuales han mostrado cierta eficacia en el proceso de clasificación de imágenes. Sin embargo, cada uno de ellos es de mayor utilidad bajo ciertas circunstancias, por ejemplo, cuando la clasificación es supervisada el método de vecinos más cercanos es útil, ó cuando la base de imágenes es grande debido a que el cálculo puede ser diferido y es suficiente con encontrar una aproximación local. Las redes bayesianas muestran buenos resultados cuando se cuenta con conocimiento de la escena, principalmente es utilizada cuando la distribución de los objetos no es tan evidente como en el caso de los vecinos más cercanos, las redes neuronales son atractivas por su habilidad de dividir el espacio de características utilizando fronteras para las clases. Dichas fronteras se obtienen a través del entrenamiento de la red.

Sin embargo, uno de los métodos más utilizados recientemente en la clasificación de imágenes es la máquina de soporte vectorial (SVM) desarrollada por Vladimir Vapnik (Vapnik, 1995). El funcionamiento de la SVM puede dividirse en dos etapas, una fase de entrenamiento que se realiza a partir de la observación de un conjunto X de n

muestras. En teoría las salidas del sistema son dos valores simbólicos $y \in \{+1, -1\}$ de forma que el conjunto de entrenamiento está formado por los pares $(x_i, y_i), i = 1, \dots, n$, donde cada vector x_i se corresponde con un vector de entrenamiento y los valores $y_i \in \{+1, -1\}$ indican la clase a la que pertenece cada vector.

El objetivo del entrenamiento es encontrar una función de decisión capaz de separar las dos clases. En caso de no ser separables se proyectan a un espacio de dimensión mayor mediante el uso de transformaciones no lineales. En este caso, la función de decisión se sitúa en el hiperplano de esa dimensión. La función de decisión tiene la forma de la ecuación siguiente:

$$f(x) = \sum_{i=1}^n \alpha_i y_i H(x_i, x) - b \quad (2)$$

donde b es una constante. y H es el núcleo o kernel que se va a utilizar para mapear la información a un espacio de características mayor. En la figura 6 se muestra una representación gráfica del problema.

Los tipos de núcleo más comunes por mencionar algunas son:

- 1) *El lineal:* $H(x, y) = x^t y$.
- 2) *La función de base radial:* $H(x, y) = \exp\{-\|x - y\|^2 / 2\sigma^2\}$, y
- 3) *El polinomial:* $H(x, y) = (1 + (x \cdot y))^d$ con grado d .

Posterior a la fase de entrenamiento, viene una fase de decisión donde dado un vector x , se determina la clase a la que pertenece de acuerdo con el signo de (polaridad)

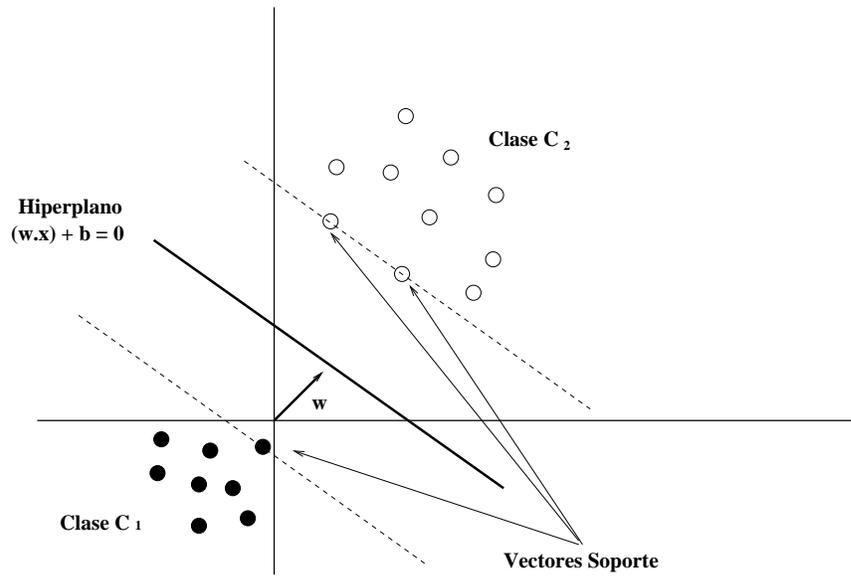


Figura 6: Entrenamiento de una máquina de soporte vectorial para encontrar el hiperplano óptimo

de $f(x)$. La magnitud puede considerarse como una medida de certidumbre sobre la decisión realizada. Los vectores de soporte son los más representativos de todos los patrones utilizados para el entrenamiento. Por lo tanto, son los patrones con mayor grado de información. La distancia mínima desde el hiperplano que separa las clases al patrón más cercano se denomina margen de separación τ . Un hiperplano de separación se dice óptimo si el margen es máximo. La distancia del hiperplano de separación al patrón z es $y_k |f(x)| / \|w\|$ donde w está dado por:

$$w = \sum_{i=1}^n \alpha_i y_i x_i = \sum_{\text{vectores soporte}} \alpha_i y_i x_i, \quad (3)$$

donde w es un vector normal al plano, x es el vector y y la clase a la que pertenece el objeto en entrenamiento y α es un vector de multiplicadores de Lagrange positivos.

El problema de encontrar el hiperplano óptimo se reduce a encontrar el valor de w que maximice el margen τ , lo que es equivalente a minimizar la norma de w .

A manera de conclusión podemos decir que una SVM consta de dos operaciones matemáticas: la transformación no lineal de un vector de entrada a uno en el espacio de características de mayor dimensión y la construcción de un hiperplano óptimo que separe los vectores de características transformados. En la primera operación, los vectores de características originales se mapean a un espacio de mayor dimensión, donde las dos clases pueden ser linealmente separables por medio de un kernel o núcleo. Los núcleos más utilizados son el lineal, la función de base radial y el polinomial. En la segunda y tercera opción se construye un hiperplano separador óptimo para maximizar el margen de separación entre muestras positivas y negativas.

II.4 Computación evolutiva

En la naturaleza todos los seres vivos nos enfrentamos a problemas que debemos resolver para poder adaptarnos al medio, conseguir alimento, aprovechar la luz solar o reproducirnos, son tareas que llevamos a cabo para poder sobrevivir. Este sistema natural de actuar de los seres vivos, motivó que los investigadores comenzaran a ver el proceso de adaptación y evolución de las especies como un proceso de aprendizaje, mediante el cual la naturaleza dota a las especies de diferentes mecanismos, buscando hacerlas más aptas para sobrevivir. Partiendo de estos preceptos, no resulta difícil percatarse de que puedan desarrollarse algoritmos que traten de resolver problemas de búsqueda y optimización guiados por el principio de la supervivencia del más apto, que postulara Charles Darwin en su famosa teoría de la evolución de las especies. Dichos algoritmos son denominados, hoy en día, algoritmos evolutivos y su estudio conforma lo que se conoce como computación evolutiva.

Por lo tanto la computación evolutiva interpreta a la naturaleza como una inmensa máquina de resolver problemas y trata de encontrar el origen de dicha potencialidad para utilizarla en sus programas. Los orígenes de este paradigma se da en la década de los 60's, donde algunas investigaciones concluyeron que las técnicas tradicionales de Inteligencia Artificial (IA) eran inadecuadas para comprender la complejidad de sistemas adaptativos. Los sistemas de IA fundados con grandes bases de conocimiento y lógica de predicción, resultaban extremadamente difíciles de manejar y sus estructuras rígidas lo imposibilitan para adaptarse cuando las condiciones del ambiente cambian, es decir, son poco adaptables. Esta es la razón primordial por la que los investigadores optaron por buscar otras técnicas, para encontrar soluciones a problemas complejos no lineales, difíciles de analizar, entender y en algunos casos definir.

Es por eso que podemos ver a la computación evolutiva como una rama de la Inteligencia Artificial. Hacia 1960 John Holland utilizó por primera vez este principio para el desarrollo de programas informáticos capaces de automodificarse, de modo que simulasen la evolución natural. En esquema, comenzaba por un algoritmo con muchos parámetros, que modificaba ligeramente uno de ellos de forma aleatoria. Este ciclo se repetía innumerables veces eligiendo en cada caso el algoritmo padre o hijo, para modificar aquel de los dos que hubiera producido una mejor solución al problema planteado; de forma que poco a poco se consiguen algoritmos más eficaces para resolver dichos problemas. En los siguientes cincuenta años, han surgido una gran variedad de trabajos que designan a un amplio conjunto de técnicas heurísticas de resolución de problemas complejos, éstos basan su funcionamiento en un mecanismo análogo a los procesos de la evolución natural, trabajando sobre un conjunto de soluciones a un problema determinado. La metodología utilizada por estas técnicas se fundamenta en el uso de mecanismos de selección de las mejores soluciones potenciales y de construcción

de nuevas soluciones candidatas mediante recombinación de características de las soluciones seleccionadas.

El algoritmo evolutivo trabaja sobre individuos que representan potenciales soluciones al problema, codificados de acuerdo a un mecanismo prefijado. Los individuos son evaluados de acuerdo a una función objetivo¹ que toma en cuenta la adecuación de cada solución al problema que se intenta resolver. La operativa del algoritmo evolutivo comienza con una etapa de inicialización de los individuos, que puede ser completamente aleatoria, muestreando al azar diferentes secciones del espacio de soluciones, o guiada de acuerdo a características del problema a resolver. El algoritmo evolutivo podría inclusive tomar como población inicial, individuos resultantes como salida de algún otro algoritmo heurístico de manera que permita calcular buenas soluciones iniciales bien aproximadas para el problema. La evolución propiamente dicha se lleva a cabo en el ciclo que genera nuevos individuos a partir de la población actual mediante un procedimiento de aplicación de operadores estocásticos. En este ciclo se distinguen cuatro etapas:

- Evaluación: etapa que consiste en asignar un valor de adecuación (desempeño) a cada individuo en la población. Este valor evalúa que también resuelve cada individuo el problema en cuestión, y es utilizado para guiar el mecanismo evolutivo.
- Selección: proceso que determina candidatos adecuados, de acuerdo a sus valores de fitness, para la aplicación de los operadores evolutivos con el objetivo de engendrar la siguiente generación de individuos.

¹se define la función objetivo como la métrica que nos permite evaluar la eficacia de nuestro individuo.

- Aplicación de los operadores evolutivos: etapa que genera un conjunto de descendientes a partir de los individuos seleccionados en la etapa anterior.
- Reemplazo: mecanismo que realiza el recambio generacional, sustituyendo individuos de la generación anterior por descendientes creados en la etapa anterior.

Aunque en la actualidad es cada vez más difícil distinguir las diferencias entre los distintos tipos de algoritmos evolutivos existentes, por razones sobre todo históricas, se suele hablar de tres paradigmas principales.

- Programación Evolutiva.
- Estrategias Evolutivas.
- Algoritmos Genéticos.

Cada uno de estos paradigmas se originó de manera independiente y con motivaciones muy distintas, sin embargo, todos buscan de manera general el mismo objetivo, el cual consiste en obtener mejoras en la solución de problemas de búsqueda y optimización. En la siguiente sección nos enfocaremos a describir el paradigma de la programación genética por formar parte de la metodología con la que se elaboró este trabajo.

II.4.1 Programación genética

Una de las corrientes basadas en los algoritmos evolutivos que ha obtenido mejores resultados es la programación genética (PG), la cual es una metodología automatizada inspirada en la evolución biológica para encontrar programas que realicen de una mejor manera tareas definidas. La PG puede considerarse como una extensión de los algoritmos genéticos, donde los programas a evolucionar están representados en forma de

árboles analíticos aunque no solo se ha utilizado para programas, sino para cualquier otro tipo de soluciones cuya estructura sea similar a la de un programa como circuitos electrónicos o fórmulas matemáticas. Esta metodología fue propuesta por primera vez por John Koza y al igual que cualquier algoritmo de búsqueda, la PG busca en un espacio de posibles soluciones, dispone de operadores de búsqueda y de una función objetivo que la oriente.

La PG mantiene una población finita de posibles buenas soluciones o candidatos a solución (denominados individuos). Dichos candidatos a solución suelen ser programas funcionales (funciones que toman argumentos y devuelven un valor) codificados en forma de árboles. Aunque se ha utilizado a la PG para evolucionar otro tipo de estructuras como árboles de decisión, reglas de prolog, grafos y secuencias de instrucciones. A la fecha no se cuentan con análisis teóricos en favor o en contra de determinada estructura, aunque seguramente si dependan en parte de la transformación de candidatos a solución que se usen.

Los operadores de búsqueda de la PG son los denominados operadores genéticos: reproducción, cruce y mutación, aunque en general se suelen usar solo los dos primeros. El operador de reproducción simplemente crea un individuo exactamente igual al que se le pasa como argumento. El de cruce toma dos individuos, selecciona aleatoriamente un nodo en cada uno de los individuos progenitores (representados como árboles) e intercambia los dos subárboles correspondientes. La figura 7 proporciona un ejemplo del operador de cruce entre dos programas sencillos.

El operador de mutación selecciona un nodo en el árbol progenitor, corta el subárbol que pende de ese nodo, y lo substituye por un subárbol generado aleatoriamente. Sin

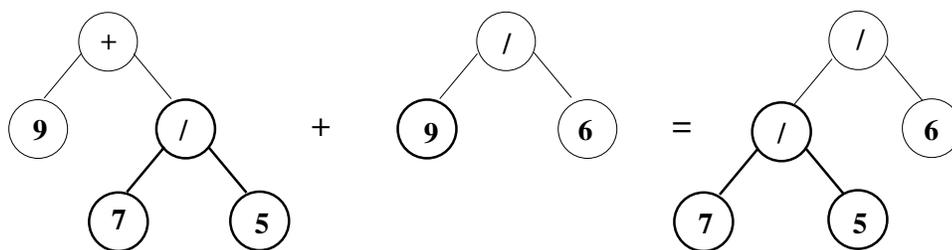


Figura 7: Proceso de cruzamiento entre dos árboles. Los subárboles intercambiados aparecen con una línea gruesa.

embargo, también en este caso se ha experimentado con otros operadores tales como encapsulación (Koza, 1992), duplicación (Koza, 1994), etc. En principio no hay razones teóricas fuertes para restringirse al conjunto estandar de operadores. Hay que hacer una salvedad: en los campos derivados de los algoritmos genéticos se le da gran importancia al operador de cruce, puesto que es capaz de combinar buenos fragmentos de dos candidatos a solución, lo que puede llevar a obtener mejores soluciones que cualquiera de los progenitores (Tackett, 1994). A la hipótesis de que un algoritmo genético funciona a base de ir combinando progresivamente buenos fragmentos se la denomina la hipótesis de los bloques de construcción (Holland, 1975).

Los árboles contienen dos tipos de información, un conjunto de terminales o ramas del árbol que está compuesto por las posibles entradas al individuo, constantes y funciones de aridad 0. Además, por el conjunto de funciones o nodos que está compuesto por los operadores constructores y funciones que pueden conformar a un individuo; las cuales pueden ser funciones booleanas, aritméticas, condicionales, etc. El conjunto de terminales y funciones elegidas para resolver un problema en particular debe ser obviamente, suficiente para representar una solución al problema; sin embargo, hay que tener en cuenta que si se usa un número grande de funciones, se aumenta el espacio de búsqueda.

Otro aspecto importante en la construcción de un programa genético es el tamaño máximo que pueda tener. Este límite puede estar impuesto por el número de nodos o por la profundidad del árbol. A final de cuentas lo que se busca son poblaciones de programas que evolucionen, transmitiendo su herencia de manera tal que los hijos se adapten mejor al medio; donde los mejores individuos tienen mayores probabilidades de reproducirse. Así, la medida de la calidad del individuo dependerá del problema.

Por último en la figura 8, mostramos un diagrama de flujo para la forma más simple del funcionamiento de un PG.

II.5 Trabajos previos

A lo largo de las últimas dos décadas han surgido varios modelos que han intentado emular el proceso de la atención visual, uno de los primeros trabajos fue propuesto por Milanese (1995). En su arquitectura (ver figura 9), le aplica diversos procesamientos a la imagen original para poder obtener nuevas variaciones de la imagen los cuales son llamados mapas característicos y son agrupados en dos categorías, una basada en regiones (perímetro, área y escala de gris) y otro basada en contornos (contraste, curvatura, longitud y orientación). Estas imágenes son mapeadas a través de un operador que obtiene una métrica de la importancia para cada píxel con respecto a sus vecinos de categoría. La fórmula del operador que se utiliza es la siguiente:

$$C_{i,j}^k = \frac{1}{\|N_{i,j}\|} \sum_{m,n \in N_{i,j}} |F_{i,j}^k - F_{m,n}^k| \quad (4)$$

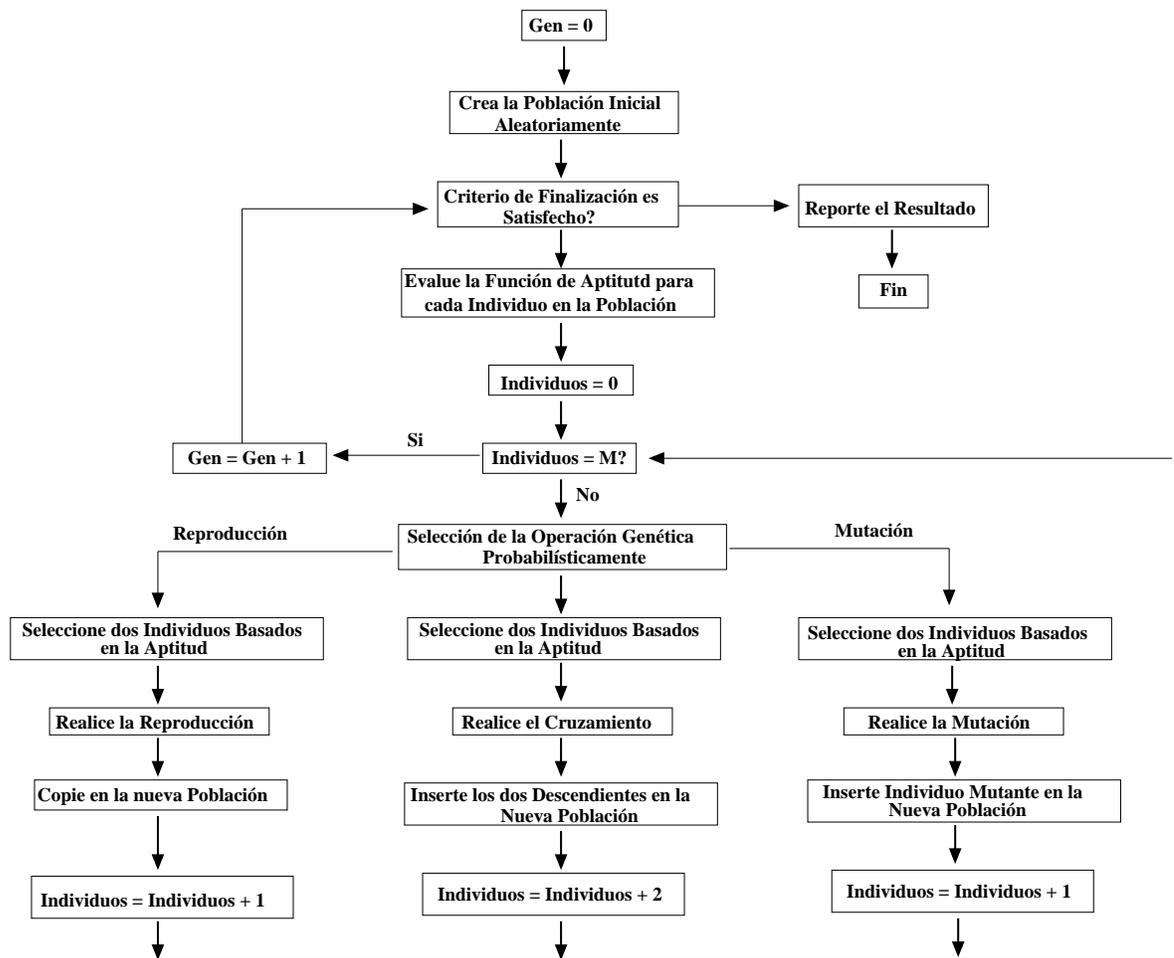


Figura 8: Diagrama de flujo que muestra los pasos más importantes en el proceso de la programación genética .

Donde F es una medida del valor para los mapas característicos y N el vecindario local del operador. Al final se obtiene como resultado un mapa de notoriedad el cual es la combinación de estos mapas característicos. El trabajo de Milanese más que enfocarse en la atención visual está muy relacionado a encontrar el objeto mejor segmentado en la imagen, el cual no forzosamente corresponde al más llamativo.

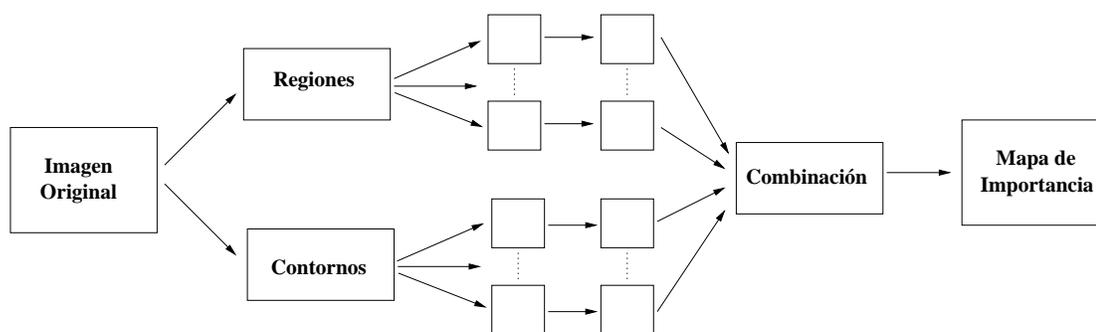


Figura 9: Un esquema básico del modelo de atención propuesto por Milanese.

Otro modelo fue el propuesto por (Osberger, 1998), los cuales presentan un enfoque en donde involucran la segmentación de la imagen utilizando división recursiva con el algoritmo Merge. Durante la segmentación, regiones menores a 16 píxeles son fusionados con el vecino más similar. A las regiones segmentadas entonces se les asigna un valor de importancia acorde a 5 criterios: una medida de contraste, el tamaño de la región, la forma del área calculada, así como el radio de píxeles que componen la región entera, la locación entre más cerca del centro la región es mejor favorecida y por último el fondo que viene siendo un valor dado por el número de píxeles en el borde con valores desfavorables.

Todas las características medidas son normalizadas a valores entre 0 y 1, donde 1 representa que el píxel es importante y 0 es el menos favorable. Estos valores son combinados a través de la suma de cuadrados derivado de las 5 características dando una única medida de importancia para cada región. El trabajo de Osberger ha sido encontrado altamente dependiente de una buena segmentación y de tener pocas bases teóricas. En la figura 10 podemos ver un diagrama de su modelo.

Tanto el trabajo de Milanese como el de Osberger carecen de un buen sustento biológico, pero en su mayoría resumen los fundamentos de los modelos de atención

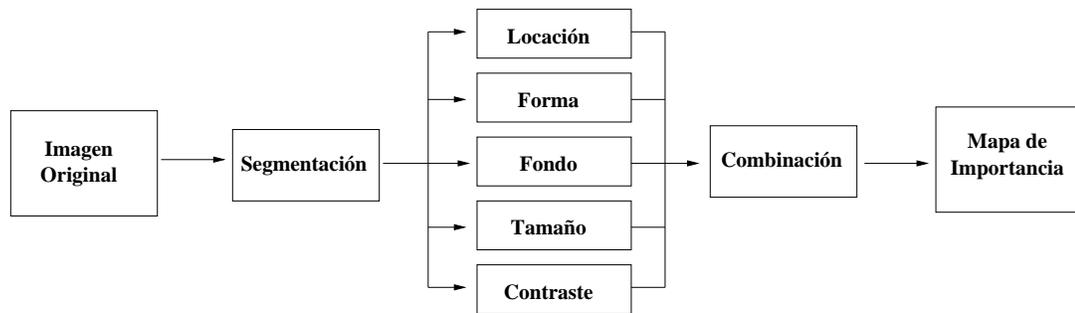


Figura 10: Un esquema básico del modelo de atención propuesto por Osberger.

visual. Una primera etapa realiza la descomposición y extracción de características de bajo nivel en la imagen, para después aplicarle ciertas transformaciones con el fin de obtener una métrica de sobresalencia en estos llamados mapas característicos y por último un método de combinación de estos mapas para obtener uno solo, el cual represente el valor de importancia para cada píxel de la imagen.

Aunque actualmente la interacción entre el reconocimiento de objetos y la atención no está del todo entendida. Existen una gran cantidad de experimentos y modelos, que sugieren que la percepción visual se divide en por lo menos dos etapas: Una etapa de procesamiento preatentiva la cual funciona de manera inmediata en la cual el campo visual entero es procesado de manera paralela; y una segunda etapa de procesamiento atentiva, donde el funcionamiento es más lento debido a que se realiza de manera serial. Así una sola región de interés de la imagen de entrada es seleccionada para realizar en ella un análisis especializado.

En este caso el primer trabajo neuronalmente creíble de la atención visual humana fue propuesto por Koch y Ullman (1985), e implementado por Itti y Koch (1999). Este modelo se enfoca en la idea de un mapa de características relevantes el cual se define como una representación topográfica de 2 dimensiones de la notoriedad de cada píxel

en la imagen. Su modelo propuesto consiste de 4 pasos claves: extracción de características de bajo nivel, diferencias centro-contorno para producir mapas característicos, la combinación de estos mapas característicos, y finalmente, la selección atencional e inhibición de retorno, ver figura 11.

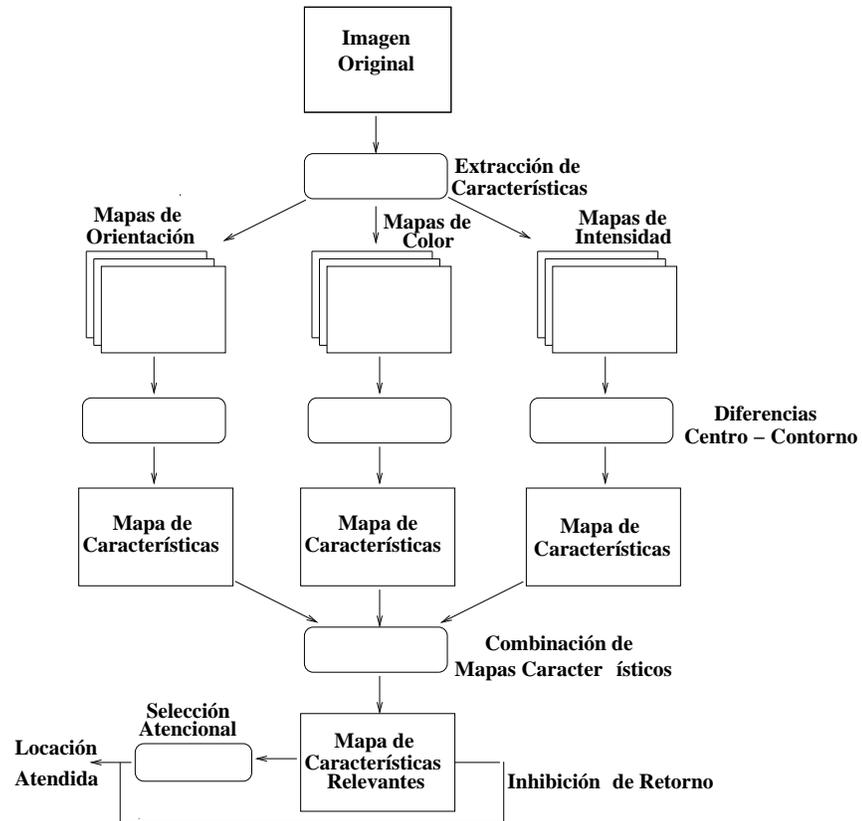


Figura 11: Un esquema básico del modelo de atención propuesto por Koch y Ullman.

El enfoque gira entorno a la extracción temprana de características, seguido por un operador que está dado por la diferencia entre la intensidad de la característica medida de cada píxel y la intensidad del contorno para producir los mapas característicos. Los mapas característicos son entonces combinados para producir un mapa

de características relevantes, que facilita la selección de regiones en la imagen para su funcionamiento posterior. Este modelo es del que estaremos hablando en nuestro trabajo, y en general existen varios modelos de atención visual basados en el modelo de Itti para ser utilizados en numerosas aplicaciones. A continuación mencionaremos algunos de los más importantes.

Ya en trabajos previos se ha demostrado que la atención visual ayuda en el proceso del reconocimiento. Tal es el caso de Rutishauser y Walther (2004) quienes mostraron que restringiendo la extracción de características y reconocimiento a regiones visualmente sobresalientes se mejora el reconocimiento. Ellos ejecutaban el algoritmo de Itti para obtener una área relevante en una imagen y como segundo paso a la zona seleccionada le aplicaron el algoritmo del SIFT (Scale-Invariant Feature Transform) propuesto por David Lowe en el 2004 para detectar puntos de interés (Lowe, 2004), sus resultados muestran un incremento considerable en el reconocimiento de objetos, en parte debido al incremento de puntos de interés asociados a los objetos y por otra parte logro una disminución considerable de falsos positivos. Navalpakkam e Itti demostraron que a través de factores TD (Top - Down) se aumentaba la precisión y la velocidad de detección de los objetos. Ellos compararon su modelo en cuatro variaciones, sin conocer ni el objetivo, ni los distractores, conociendo solo uno de los dos y conociendo ambos, obteniendo con este último sus mejores resultados, por lo que se llega a la conclusión de que en tareas de búsqueda visual el contar con información previa de los objetos y sus distractores (factores TD) mejora la precisión de reconocimiento de objetos (Navalpakkam y Itti, 2006).

Los resultados obtenidos de estos dos trabajos dan la pauta para seguir utilizando la

atención visual en trabajos relacionados con el reconocimiento y clasificación de objetos. Particularizando al caso de clasificación en imágenes naturales uno de los primeros trabajos fue propuesto por Itti y Koch. Su trabajo se basa en el entrenamiento de imágenes de señales de tránsito y latas de refresco para poder reconocerlas de manera automática. El procedimiento era muy sencillo el sistema iba modificando los pesos de los mapas de notoriedad con el fin de que se fuera especializando en contener en un radio fijo las señales de tránsito o las latas. Los resultados no fueron buenos en las señales de tránsito en parte por las diferentes formas y colores que las señales contenían. Por otra parte los resultados con las latas de refresco fueron notoriamente mejores (Itti y Koch, 2001). Walther propuso un sistema para comprobar que la atención visual selectiva permitía el reconocimiento y aprendizaje de múltiples objetos en escenas complejas, su diseño era de clasificación no supervisado, porque era capaz de encontrar regiones prominentes y obtener sus puntos de interés, si podía coincidirlo con entradas anteriores lo contaba como objeto de la misma clase. Si por el contrario no podía coincidirlo con ninguno anterior sus puntos eran tomados y generaba una nueva clase. Su modelo obtuvo buenos resultados teniendo en cuenta que la clasificación era BU es decir, sobre objetos que de por sí sobresalen por sus características (Rutishauser y Walther, 2001).

Otros trabajos que utilizan a la atención visual enfocados a la clasificación son el propuesto por Chikkerur en el 2009, esta investigación estudia la clasificación de imágenes naturales, basándose en un método bayesiano de atención visual que integra locaciones TD y prioridad de características en imágenes controladas por información BU. Por lo tanto, este modelo se basa en la información contextual de la escena; lo que le permite predecir el tamaño y la ubicación de los objetos a reconocer (Chikkerur y Poggio, 2009). Otro trabajo que utiliza este enfoque de probabilidad bayesiana es propuesto por Elazary. el cual esta basado en el modelo de Itti, con la diferencia de que

el modelo aprende la probabilidad de que un objeto visual aparezca teniendo un rango de valores en cada mapa característico, por lo tanto el resultado era que su modelo buscaba en los mapas con mayor probabilidad de encontrarlo, con el fin de obtener mejores y más rápidos resultados que con modelos del estado del arte que incluyen el SIFT y HMAX (Rutishauser y Poggio, 2000), hizo una comparativa con estos dos modelos, con 3 diferentes bases de datos, obteniendo en algunas ocasiones mejores resultados (Elazary y Itti, 2010).

La aplicación de GP en problemas con imágenes ha sido utilizada cada vez con mayor énfasis para resolver problemas de vision por computadora. El primer trabajo fue propuesto por Tackett (1994). El utilizó un GP para resolver una tarea de detección de objetos. Después, Johnson evolucionó rutinas visuales las cuales son habilitadas para reconocer simples patrones sobresalientes de las siluetas de personas (Johnson, 1999). Ebner introdujo una técnica basada en GP para evolucionar una tarea de operador de imagen específica, en particular para reproducir el bien conocido detector de puntos de interés Moravec (Ebner, 1998). Olague et al ha demostrado obtener buenos resultados en este rubro, se han desarrollado trabajos en donde a través de la utilización de algoritmos genéticos se puede seleccionar la región óptima en imágenes para su clasificación; a través de la utilización de un descriptor de texturas y de SVM. El programa era entrenado para encontrar el área mínima necesaria y el número de descriptores necesarios para reconocer y clasificar de manera eficiente objetos diversos tales como vacas, casas, carros, etc (Olague y Romero, 2006) y expresiones faciales en imágenes térmicas (Hernandez y Olague, 2007).

Otro trabajo que ha utilizado algún tipo de evolución o de enfoque genético para tratar el problema de la atención visual es el trabajo de Stentiford (2001), Él contaba

con la hipótesis de que lo diferente es lo que atrae la atención, por lo cual el presentó una estrategia con mapas de píxeles, donde el genético hacia una comparativa entre cromosomas o píxeles y sus vecindario. Si los píxeles contaban con distinto vecindario sobrevivían, si eran iguales se desechaban. Los resultados de su trabajo fueron buenos con imágenes con pocas características visuales pero a la hora de complicar la imagen los resultados de su trabajo eran más parecidos a los logrados por un segmentador de bordes.

Sin duda el trabajo con características de diseño más parecidas al modelo propuesto en esta tesis es el trabajo publicado por Pereira y Gomes (2006). Su modelo también está basado en el modelo de Itti, él se apoya en la construcción de un algoritmo genético que pueda modificar los pesos de los 27 mapas característicos que influyen en la obtención de los píxeles de interés de la imagen, su objetivo era que el modelo combinara estos pesos para enfocarse en los puntos de interés dentro de las regiones sobresalientes previamente etiquetadas. La función objetivo se basaba en obtener el mayor porcentaje de puntos dentro de las zonas etiquetadas. Los resultados del modelo de Pereira logran una mejora del 20 por ciento en la concentración de puntos de interés sobre regiones importantes en las imágenes que aplicando el mismo modelo sin evolucionar. El modelo de Pereira no fue llevado a cabo en tareas de visión de alto nivel.

Capítulo III

Proceso Psicobiológico de la Atención Visual

Como vimos anteriormente la atención visual selectiva se refiere al mecanismo por el cual los humanos podemos dirigir nuestra mirada hacia los lugares de interés en el ambiente visual. Este mecanismo permite solamente a una pequeña parte de la información sensorial de entrada llegar a la memoria de corto plazo y a la conciencia visual; permitiendo en este sentido, dividir el complejo problema del entendimiento de la escena, a una forma serial rápida y computacionalmente menos demandante en lo referente a las tareas de análisis visual focalizado (ver figura 12). Sin embargo, la compleja tarea de la visión no es algo puramente atencional, se puede derivar que debido al corto campo visual y al diminuto tiempo de fijación, habría un pobre entendimiento de la escena visual. Es por eso que la visión parece contar con una elaborada cooperación entre procesos como un: masivamente paralelo sistema preatentivo de análisis del campo completo y otro más detallado circunscrito sistema llamado análisis secuencial. Por lo tanto, en este capítulo se explicará de forma detallada este mecanismo de atención visual desde sus bases psicológicas y fisiológicas.

III.1 Teoría de la integración de características

Como mencionamos en la introducción existen varios modelos psicológicos propuestos para explicar la atención visual humana, pero ninguno tan influyente como el que propusieron Anne Treisman y Garry Gelade (1980). Ellos postularon que los diferentes



Figura 12: Ejemplo de la atención focalizada en una escena.

tipos de atención son responsables para vincular las diferentes características visuales dentro un todo consciente. El mecanismo básico de su modelo de procesamiento visual consiste en descomponer la imagen en algunas características visuales primarias, las cuales son presentadas como mapas independientes que más tarde son integrados en un único mapa maestro que es accedido en orden a dirigir la atención a las regiones más notorias. Por lo tanto, nuestra atención enfocada puede ocurrir únicamente después de que las características han sido asociadas en esa pequeña región del mapa maestro. En la figura 13 se muestra una representación de esta teoría.

En la realización de experimentos Treisman, distinguió dos tipos de búsqueda visual, la búsqueda de características y la búsqueda conjunta. La primera es realizada de forma rápida y de manera preatentiva, donde los objetivos son definidos por una sola característica primaria. La búsqueda conjunta se realiza de forma serial hacia objetivos definidos por un conjunto de características primarias. El proceso es mucho más lento y necesita forzosamente de la concentración (ver figura 14, lado derecho). Así concluyó después de varios experimentos que el color, la orientación y la intensidad son las características visuales primarias con las cuales la búsqueda visual puede llevarse a efecto afin de encontrar objetos predominantes.

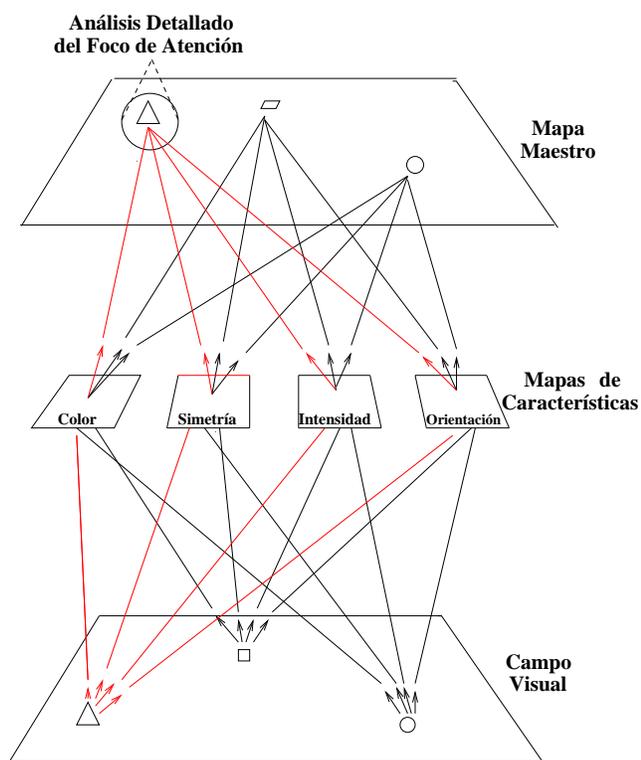


Figura 13: Representación gráfica de la teoría de integración de características.

Esta teoría también sugiere que la atención puede facilitar la localización de las características mediante la presencia física del estímulo y la información sensorial almacenada durante un corto plazo de tiempo. Por lo tanto, la percepción de un objeto genera una representación temporal en un "fichero - objeto" que, de forma general, recoge, integra y actualiza la información acerca de las características del estímulo. Las características del "fichero - objeto" se puede almacenar como una señal y se puede recordar en las próximas veces que aparezca el objeto visual.

Por lo tanto, podemos concluir que para hacer la búsqueda de objetos, primero se produce una integración de las características con lo cual se puede saber que áreas son las más llamativas y esto varía dependiendo de la tarea a seguir. Después se realiza

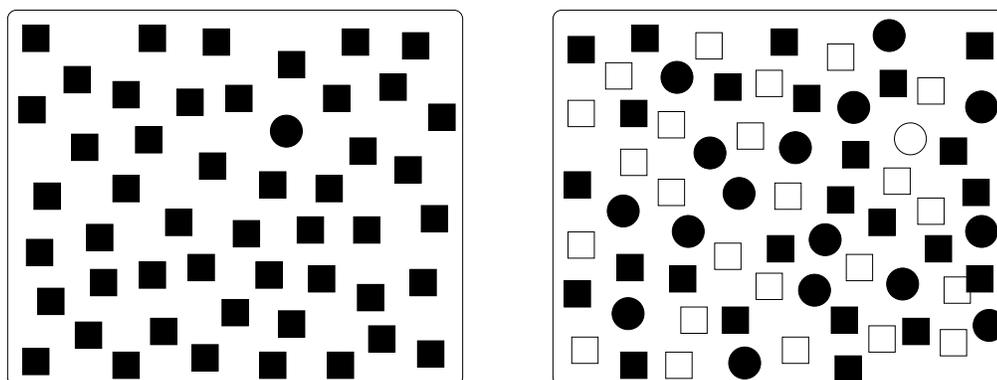


Figura 14: Ejemplo de búsqueda preatentiva (lado izquierdo) y búsqueda conjunta (lado derecho).

un proceso de fijación y atención en las áreas prominentes, para posteriormente hacer una comparativa con memoria a fin de poder asociar el objeto con uno ya conocido.

Experimentos posteriores (Kanizsa, 1988), han demostrado que utilizando la similitud de objetos como una característica primaria, se obtiene un menor tiempo de reacción y por lo tanto un menor tiempo de procesamiento serial. Más adelante se dará una posible relación entre el proceso de la TIC y el proceso fisiológico; pero primero veremos la biología del sistema visual humano.

III.2 Teoría de la coherencia y ceguera al cambio

El poner atención solamente en una pequeña región de la escena a la vez, provoca que no podamos percibir cambios surgidos en entornos diferentes al área atendida. A este tipo de fallas de los observadores se llama ceguera al cambio y fue propuesto por McConkie y Currie (1996). Existen varios factores que provocan este fenómeno como es el parpadeo, la concentración o el mismo movimiento del ojo. Obviamente acompañado de cambios en la escena, este tipo de fenomenos es muy utilizado para realizar trucos

de magia. Pero, ¿cómo es que realmente hacemos caso omiso a estos cambios?.

La teoría de la coherencia propuesta por Rensink (1997), menciona que entre las etapas preatentiva y la atenta propuestas en TIC. Existe además, otra etapa donde se realiza un preprocesamiento rápido de la imagen, involucrando propiedades geométricas y fotométricas de forma paralela a través de todo el campo visual. El resultado son estructuras que proveen una descripción local de la estructura, tal como su orientación tridimensional y agrupación de fragmentos dados por sus bordes relacionados. A estas estructuras se les llama protoobjetos; éstos pueden llegar a ser bastante complejos, y son coherentes solo en una región de la imagen. Estos perduran hasta que vuelvan a ser reemplazados por un nuevo estímulo de la retina en la misma zona.

Los protoobjetos son lo que nos permiten mantener la idea de que todo el alrededor de nuestro mundo permanece coherente. Ahora se sabe que la única forma de notar cambios es enfocar nuestra atención en la zona durante el momento donde se está llevando a cabo el cambio. Sino se refutara este proceso estaríamos enfocados en nuestra tarea pensando que todo lo demás sigue sin novedad. Así simplemente percibimos manchas de color uniforme en el resto del ambiente. Cabe mencionar que estos protoobjetos no corresponden exactamente con los objetos reconocibles, así pueden ser desde un fragmento hasta varios objetos en un solo protoobjeto.

III.3 Organización neuronal del cerebro

El cerebro es el órgano más sofisticado del cuerpo humano, su tarea primordial es controlar y administrar las actividades que realizan los órganos sensoriales de nuestro cuerpo. Su funcionamiento se lleva a cabo por medio de las conexiones sinápticas de las

millones de neuronas con las que cuenta. Para su estudio los neurólogos han dividido al cerebro humano en cuatro lóbulos: frontal, temporal, parietal y occipital (ver figura 15). El último es de especial interés por la comunidad dedicada al estudio de la visión por ser el lóbulo donde se encuentra la corteza visual.

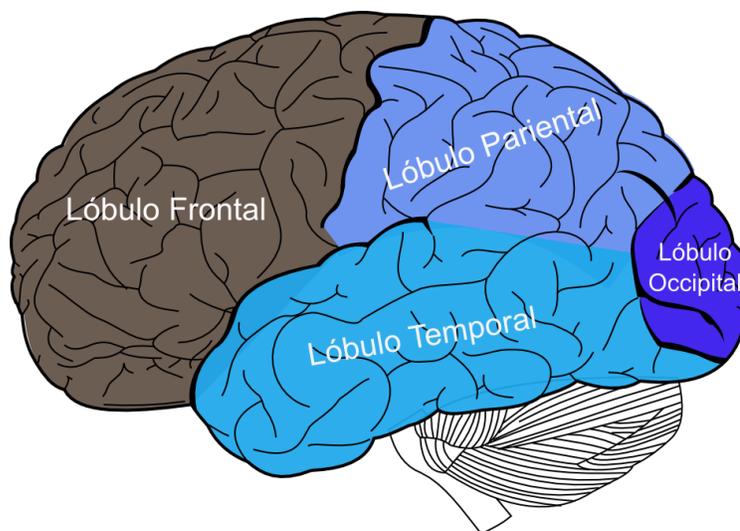


Figura 15: Vista lateral de los lobulos cerebrales

El lóbulo frontal está relacionado con el movimiento de las partes de todo el cuerpo desde los pies hasta los ojos, este último también de gran apreciación por ser importante en la elección de la región de interés, también realiza tareas correspondientes al comportamiento, orientación espacial, solución de problemas e imaginación. Se ha comprobado que daños en esta región del cerebro ocasiona pérdida de movilidad en partes del cuerpo, problemas de atención, cambios de comportamiento así como dificultad para solucionar problemas y expresarse de manera correcta.

El lóbulo parietal se encarga de las sensaciones y de la percepción. Este lóbulo recibe información sensorial de todo el cuerpo como son los receptores sensoriales de

la piel, los músculos, y las articulaciones. Sus funciones principales son: proveer información para orientarnos espacialmente, procesa los impulsos nerviosos producidos a partir de sensaciones como el dolor, la temperatura, el tacto; y también procesa funciones de lenguaje. Investigadores han relacionado que las lesiones sufridas en este lóbulo alteran la comprensión del lenguaje y aquellas tareas que se estudian en este trabajo como el reconocimiento de objetos. Es por esto que resulta muy interesante estudiar las cuestiones anatómicas y fisiológicas.

Por su parte, el lóbulo temporal tiene como funciones la comprensión del lenguaje hablado, la regulación de emociones y motivaciones como son la ansiedad, el placer y la ira. También está encargado de la percepción de la forma visual y del color; además, de regular funciones del oído así como el equilibrio y el balance. Cualquier daño presentado en estos lóbulos se refleja en pérdida de la memoria de corto plazo, dificultad en el reconocimiento de rostros y en la comprensión de palabras.

Por último, en la parte posterior del hemisferio cerebral encontramos al lóbulo occipital, donde se encuentra la corteza visual primaria y las áreas visuales secundarias especializadas en el procesamiento de las imágenes. Se puede intuir que esta área representa el centro de la percepción visual. Por su ubicación, el lóbulo occipital es uno de los más factibles a ser dañados. Alteraciones en esta zona del cerebro se convierten en problemas de visión en cuanto a localización de objetos, dificultad para detectar movimientos y la incapacidad para reconocer objetos en dibujos o fotos.

Cada uno de los lóbulos se encarga de actividades precisas y algunos de ellos tienen en común algunas tareas como es el caso del procesamiento visual. De esta misma forma, dentro de ellos, las neuronas trabajan en asociaciones especializándose en tareas

Tabla II: Funciones principales de los lóbulos pertenecientes al cerebro humano.

Lóbulo	Funciones.
Frontal	Orientación espacial del cuerpo, movimiento de extremidades, solución a problemas, creatividad y planeación.
Parietal	Detección de movimiento, funciones de lenguaje y procesamiento de dolor, temperatura y tacto.
Temporal	Percepción de la forma visual y el color, reconocimiento de objetos y rostros y regulación de emociones.
Occipital	Localización de objetos, estimación de dirección, velocidad y trayectorias de los objetos.

como lo veremos más adelante. En la tabla II se muestra un resumen de las funciones principales de cada lóbulo.

III.4 Componentes fisiológicos involucrados en la atención visual y el reconocimiento

El sistema visual humano sin duda ha evolucionado para adaptarse al medio ambiente, como se vio en la sección anterior ha logrado una estructura física que le permite ser altamente eficiente en sus tareas diarias. El gran rendimiento alcanzado por nuestro sistema visual se basa no solo por la eficacia de elegir la información relevante sino también por el procesamiento paralelo de diferentes tipos de información en múltiples áreas corticales. Por ejemplo, en el caso de la visión como se verá más adelante que existen áreas del cerebro especializadas e independientes para tratar cierta información de la imagen como es el color, la orientación y el movimiento. Por lo tanto, aunque la parte principal encargada del procesamiento visual de las imágenes es la corteza visual primaria, existen otras áreas necesarias para llevar a cabo este proceso. La estructura principal y la relación de los componentes del sistema visual humano se muestran en el

diagrama funcional de la siguiente figura 16, donde podemos ver que está constituido por la retina, los núcleos laterales geniculados (NLGs), y las áreas V1, V2, V3, V4 y MT (V5) de la corteza visual.

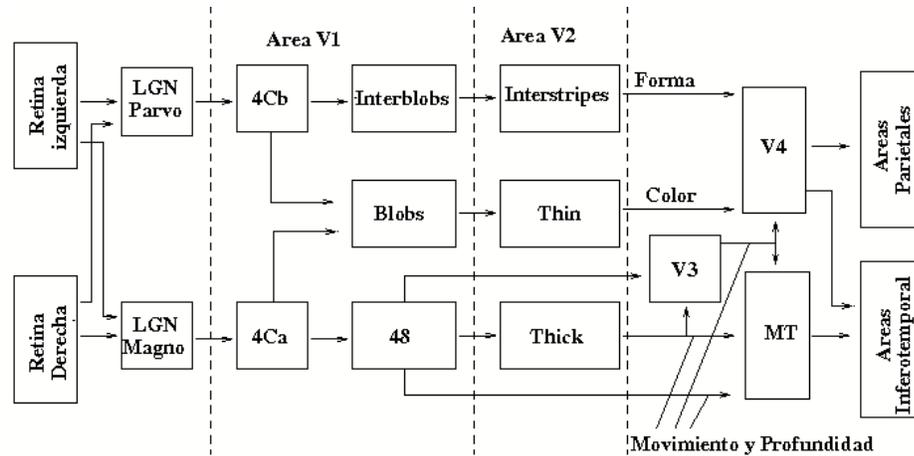


Figura 16: Diagrama que muestra las áreas más importantes involucradas en el proceso de la visión.

La representación anatómica de cada uno de estos componentes de la visión se muestran en la figura 17, donde podemos ver un bosquejo de como es el proceso de la percepción visual. Sin entrar en detalles de los mecanismos más importantes involucrados en el procesamiento de la visión, es preciso resumir brevemente el proceso de la visión en pocas líneas. Todo comienza con la luz que reflejan los objetos la cual ingresa por los ojos y es proyectada dentro de la retina, desde aquí se propaga la información visual a través del nervio óptico hasta llegar al quiasma óptico, lugar donde la información sigue dos rutas para cada hemisferio del cerebro: la ruta colicular que se dirige hacia el colículo superior, y la más importante, la ruta retino-geniculado, la cual transmite cerca del 90% de la información visual y se dirige hacia el núcleo geniculado lateral (NGL). Desde el NGL, la información es transmitida a la corteza visual primaria

(V1). Hasta aquí, el flujo de procesamiento es también llamada ruta visual primaria. Desde V1, la información es transmitida a las áreas "altas" del cerebro V2-V5, a la corteza inferotemporal (IT), el área medio temporal (MT o V5) y la corteza parietal posterior (PP). De esta forma se llevan a cabo tareas complejas como el reconocimiento de objetos.

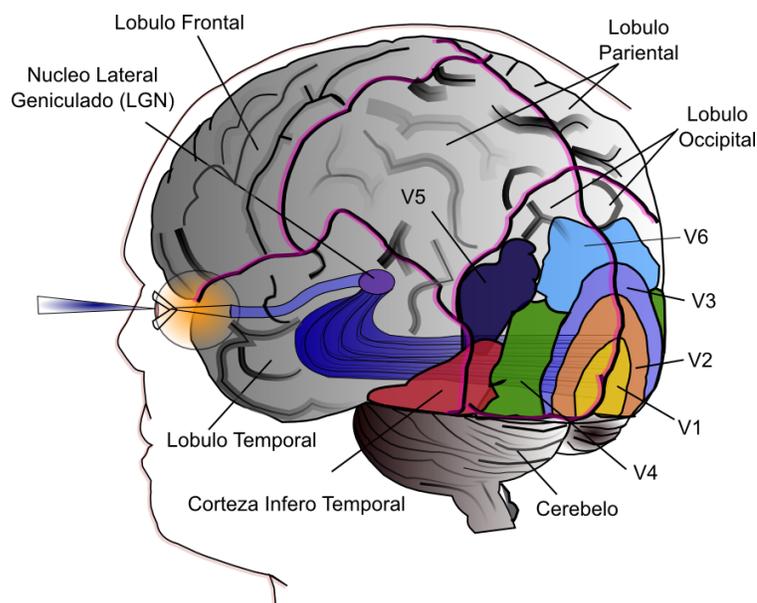


Figura 17: Vista lateral y superficial del cerebro, donde se muestra la trayectoria visual del cerebro.

En la figura 17 se pueden ver algunas de las áreas más importantes involucradas en el proceso de la visión; sin embargo áreas como el quiasma óptico no pueden ser observados. En la figura 18, podemos ver un corte coronal el cual nos permite observar las partes importantes del cerebro en relación al proceso visual desde otra perspectiva. En este esquema podemos ver como el sentido provisto por los dos ojos no solo deriva en una visión estereoscópica, sino también tiene como objetivo, a través del quiasma óptico repartir la información en cada uno de los dos hemisferios, lo cual permite un desarrollo eficiente del proceso de reconocimiento.

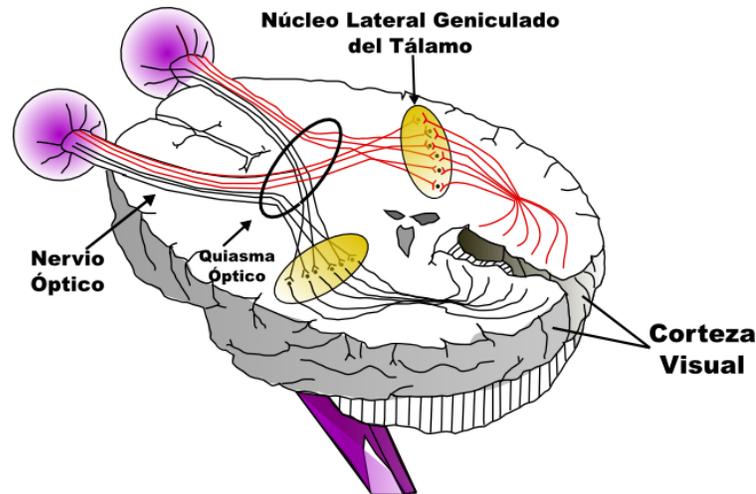


Figura 18: Corte coronal del cerebro, el cual proporciona otra vista de la trayectoria visual en el cerebro. Solo se muestra la corteza visual primaria (V1)

Una vez mencionados de manera superficial los conceptos básicos del proceso de la visión, nos enfocaremos en las regiones del cerebro que participan en el desarrollo de la atención visual; las cuales incluyen la mayoría de las áreas del procesamiento visual temprano. Sin embargo, aunque los dos métodos de atención BU y TD trabajan de forma paralela, ambos procesos pueden ser encontrados en áreas específicas del cerebro. Las áreas del cerebro más importantes involucradas en el proceso BU son principalmente la retina, el núcleo lateral geniculado (NLG), el colículo superior (CS), y el área V1. Desde este nivel, las áreas del cerebro están altamente enfocadas a los procesos TD. Desde V1 el proceso se descompone en dos principales caminos: uno es el ¿dónde ver?, el cual viene desde la periferia de la retina y se dirige a la corteza parietal posterior, y la otra es el ¿qué ver?, el cual va desde la fovea y se dirige a la corteza inferotemporal. Mientras que el primero participa en el movimiento ocular

y procesamiento de la información espacial, el segundo contribuye al proceso del reconocimiento de objetos, sin duda lo realizan de una manera cooperativa, ver figura 19.

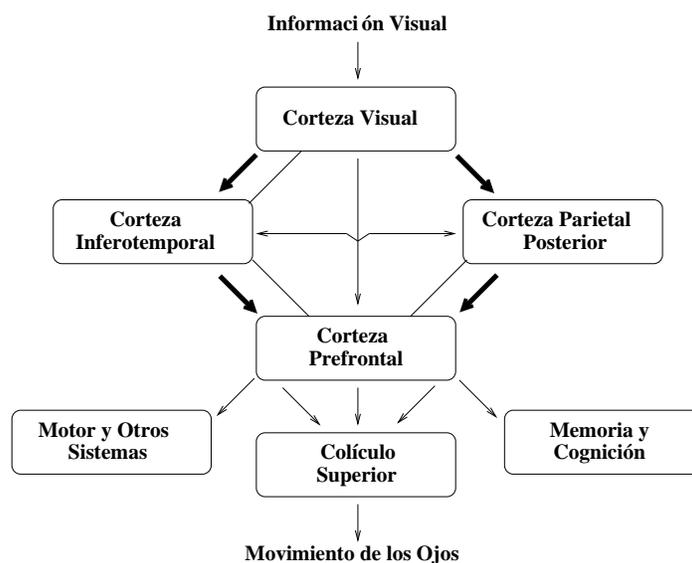


Figura 19: Mecanismos neuronales para el control de la atención.

Como se mencionó la corteza visual primaria (V1) envía la información por dos rutas, la primera llamada franja dorsal o el camino de ¿dónde ver?, la cual cruza el área visual V2, el área dorsomedial y el área visual medio temporal también conocida como V5 para finalmente llegar a la corteza parietal posterior. Como su nombre lo indica está enfocada en la percepción del movimiento y así está dirige la mirada hacia las zonas de interés. Por otra parte la ruta ventral que también atraviesa el área visual V2, luego pasa a través del área visual v4 y llega finalmente a la corteza inferotemporal. La ruta del camino inferotemporal es llamada a menudo ventra y define el ¿qué ver?, y está asociado con el reconocimiento y representación del objeto. En la figura 20 se

muestra una representación anatómica de las dos rutas.

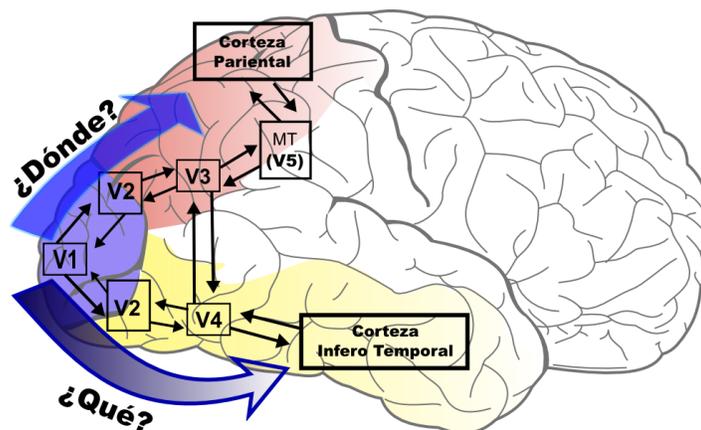


Figura 20: Rutas del procesamiento de la percepción visual.

Aquí es donde se produce la dicotomía mencionada en la introducción entre las rutas dorsal/ventral, la cual fue definida por primera vez por Ungerleider y Mishkin (1982) y sigue siendo discutible entre científicos de la visión y psicólogos. Aunque a final de cuentas, se puede ver como una simplificación del verdadero estado de las acciones dentro de la corteza visual.

El funcionamiento básico de las neuronas en la corteza visual es disparando impulsos eléctricos cuando el estímulo visual aparece dentro de su campo receptivo. Un campo receptivo es la región dentro del campo visual el cual produce un impulso eléctrico. Cada neurona puede responder a un subconjunto de estímulos dentro de su campo receptivo. Esta propiedad es llamada sintonía neuronal y en las áreas visuales tempranas, las neuronas tienen una sintonía simple. Por ejemplo, una neurona V1 dispara estímulos verticales en su campo receptivo. En las áreas visuales altas, la sintonía neuronal se vuelve más compleja. Por ejemplo, en la corteza inferotemporal, una neurona

puede dispararse solamente cuando cierto objeto aparece en su campo receptivo.

Una vez explicados los procesos básicos del proceso de la atención visual y lejos de dar una exhaustiva explicación de los mecanismos en el cerebro humano, nos enfocaremos en describir las partes que son necesarias para entender el procesamiento visual involucrado en la atención selectiva y que tienen un rol importante en el desarrollo de este trabajo.

III.4.1 La retina

El proceso visual comienza a través del ojo, más específicamente en la retina. Ella es la encargada de captar los rayos de luz que llegan al ojo por medio de la cornea, el cristalino y las cámaras anterior y posterior, también se integran los bastones y los conos. En la figura 21 se muestra una figura con los componentes principales del ojo humano

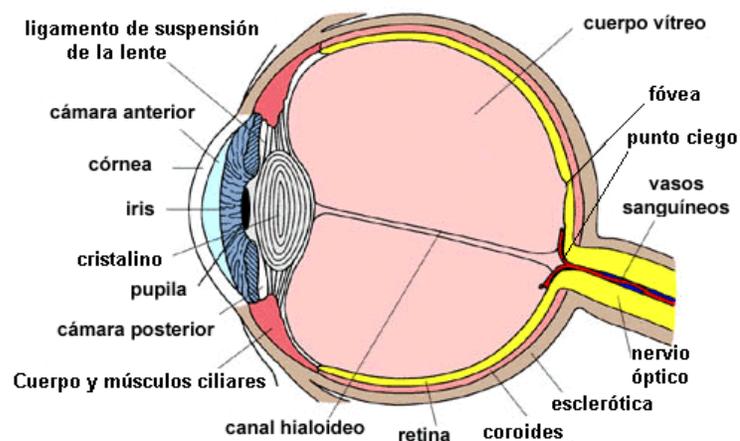


Figura 21: Anatomía del ojo humano

En el ojo humano existen 100 millones de bastones y de seis a siete millones de conos aproximadamente. Tanto los conos como los bastones son células fotorreceptoras sensibles a la luz. La distribución de conos y bastones es diferente en el ojo. Existe una parte en la retina llamada fovea en la que sólo hay conos y es el área de mayor agudeza visual. La estructura de los bastones y los conos es la misma: están formados por un cuerpo celular, un axón y un proceso fotosensible.

Dentro de la estructura de los bastones, el proceso fotosensible se divide en externo e interno. El proceso externo está encargado de atrapar la luz que llega a la retina; dicho de otra forma, es la parte de los bastones que funcionan como receptores, mientras que en el interno, se lleva a cabo la actividad para producir energía. Por otro lado, el proceso interno tiene mitocondrias, aparato de golgi y es el encargado de generar la sustancia escotopsina, que más tarde es enviada al segmento externo. Los bastones son altamente sensibles a la luz y ellos son los que se utilizan cuando la intensidad de ésta disminuye, como puede ser en la noche.

Los conos tienen una estructura similar a la de los bastones y de la misma manera, un cono es una célula fotorreceptora y es sensible a la luz. Contrarios a los bastones, los conos son los utilizados cuando la intensidad de la luz es más alta, es decir, durante el día. Por esta razón es que al haber poca luz, no somos capaces de distinguir colores.

Para enfocar las imágenes en la fovea, los globos oculares realizan una serie de movimientos. Las señales de luz captadas por la retina, pasan por los conos y bastones. Para esto, la luz debió pasar por la córnea, el cristalino y el humor vítreo, ver figura 22, hasta llegar a los conos y bastones. Conectadas a los bastones encontramos las células bipolares.

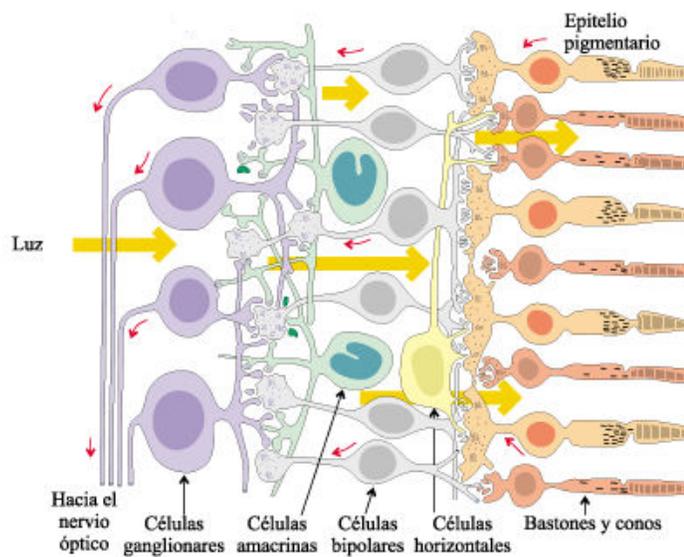


Figura 22: Estructura celular de la retina.

Al tratarse de neuronas, y recordando aspectos básicos de su estructura, las células amacrinas y horizontales son las responsables de que suceda la excitación e inhibición en la retina, procesos que permiten que la retina descomponga la imagen en campos receptores circulares para las células ganglionares. Las células amacrinas son interneuronas que contribuyen a la percepción del movimiento.

Los campos receptores son más o menos circulares y antagónicos; ésto significa que cada campo receptor tiene un centro y una periferia concéntrica. Las señales que caen fuera del centro, en la periferia, producen una respuesta contraria a la de las señales que caen en el centro, ver figura 23.

Cuando una neurona ganglionar se encuentra excitada, el centro de su campo receptor se activa (centro encendido) en cuyo caso las señales cayeron. En caso contrario, se trata de un campo con centro apagado, que sucede cuando las señales de luz caen en la periferia. La respuesta de la periferia es opuesta. Y a ésto se le conoce como

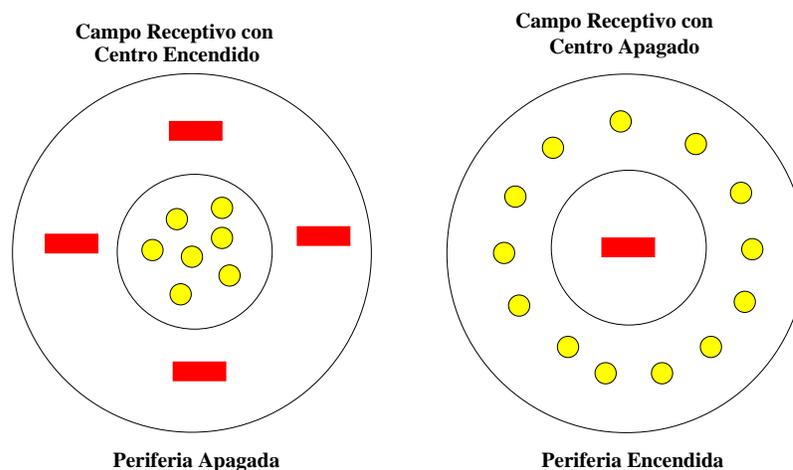


Figura 23: Campos receptores circulares de las neuronas ganglionares con centro activado y con centro desactivado respectivamente..

proceso antagónico (Ojeda, 2004).

El comportamiento de excitación e inhibición permite que se detecten los contrastes débiles y también los cambios rápidos de la intensidad de la luz. Desde este punto, la información continúa su recorrido por medio del nervio óptico a través de los axones de las células ganglionares (Ojeda, 2004) hasta llegar al quiasma óptico, ver figura 18.

Los movimientos sacádicos

La gran concentración de células fotorreceptoras, en particular los conos se encuentran en tan solo una pequeña área de la retina, este fenómeno ha hecho necesario que surga el proceso de la atención visual, a fin de contrarrestar este problema anatómico y poder mantenernos al tanto de la basta información visual que existe en el ambiente. De esta forma surge el movimiento ocular (movimientos sacádicos o simplemente sacádicos) los cuales recorren todas las regiones de interés en el ambiente.

Los sacádicos constituyen uno de los movimientos más característicos de los ojos. Son movimientos fundamentalmente voluntarios, aunque también los hay involuntarios y su objetivo no es otro que el de disponer la imagen visual en la fóvea que es la región de la retina que dispone de mayor agudeza visual. Los ojos sólo permanecen relativamente quietos para enfocar una zona concreta de la escena durante periodos de tiempo muy breves, frecuentemente, entre 200 a 350 milisegundos de duración. Sin embargo, una fijación es un complejo proceso, en el que se han identificado dos componentes que pueden estar más o menos relacionados (Viviani y Mounoud, 1990). Un primer componente queda definido por el periodo refractario motor entre movimientos sacádicos (100-200 ms.), un segundo componente está vinculado al procesamiento cognitivo (con una duración mínima de unos 50 ms.), el cual está influido por numerosos factores, pero en el que se determina qué zonas del estímulo se atienden, cómo se integra la información anterior y venidera, a qué zonas de la periferia visual se debe prestar atención y a qué zonas de la escena se dirigirá la siguiente fijación ocular. Como vimos en la introducción Yarbus realizó experimentos y comprobó que el movimiento de los ojos depende de la información cognitiva del observador.

III.4.2 Núcleo geniculado lateral (NLG)

El NLG es el lugar de terminación de los impulsos dirigidos a la corteza visual, se ubica en el Tálamo y se compone de 6 capas neuronales. Las capas 1 y 2 son llamadas magnocelulares debido a que son más grandes que las neuronas de las otras capas. Las neuronas de las capas 3, 4, 5 y 6 son conocidas como parvocelulares. Está formado por dos vías principales: una ventral y una dorsal.

Una vez que a las señales luminosas se les dio un tratamiento en la retina, la información viaja a través de dos caminos paralelos hacia el cerebro: la vía parvocelular, que contiene información del color y la forma del objeto, es decir, transporta información sobre qué es el objeto, y la vía magnocelular, cuya información es acerca del movimiento, el contraste (Capilar, 2004) y del espacio (Sánchez. 2005), indicando así dónde está el objeto.

Desde el núcleo geniculado ya se comienza a distinguir una organización en columnas de las 6 capas que lo conforman. Esto se observa cuando la información captada en un punto del campo receptor es transportada al LGN por vías cruzadas (capas 1, 4 y 6) y directas (capas 2, 3 y 5).

El siguiente trayecto de la información es del LGN a V1 ó área 17 de Brodmann (Ojeda, 2004) o área estriada). Es en ésta área en donde la cantidad de neuronas es 500 veces más grande que en LGN. El área V1 es una de las partes más importantes del procesamiento de imágenes y es una de las áreas de la corteza visual.

La poca información visual (cerca del 10%) que no pasa por el NLG se va directo al colículo superior donde ayuda a la orientación de los ojos y la cabeza hacia los principales estímulos del entorno. Por lo tanto, el colículo contiene parte del aparato neuronal necesario para el control de los músculos de los ojos y del cuello. Existen dos proyecciones principales desde el colículo superior, una hacia las regiones de la formación reticular, que controlan específicamente los movimientos verticales u horizontales de los ojos, y la otra hacia la medula espinal cervical, para controlar los músculos del cuello.

III.4.3 La corteza visual

La región principal donde se lleva a cabo la visión es en la corteza visual, la cual se puede dividir en dos áreas: la corteza visual primaria o V1 y las áreas visuales corticales extra estriadas también conocidas como V2, V3, V4 y V5.

El área V1, es la primera área cortical que recibe información directa del NLG. Al no ser la única región donde se lleva a cabo el proceso de la visión, sino que la información se transmiten directa e indirectamente a otra áreas especializadas en atributos como color, movimiento, reconocimiento de objetos, etc. Por esto se le conoce a V1 como una estación central, pero el área V1 no solo funciona como una estación que recibe y reparte la información, sino que también realiza tareas de procesamiento de la información visual, así se sabe que sus células son selectivas a la orientación y al movimiento local en una orientación. También se ha descubierto que las células V1 observan el mundo a través de una pequeña ventana lo que hace que su información sobre el movimiento local sea ambiguo.

La corteza visual primaria está formada por varios tipos de células entre las que se encuentran las células simples, las complejas e hipercomplejas. Las células conocidas como células simples o células S, fueron descubiertas en la corteza cerebral del gato por Hubel y Wiesel (). Estas se localizan en V1 y responden con mayor intensidad a las líneas rectas en una orientación determinada. Cuando la orientación de la línea cambia, la respuesta es menos intensa. En la figura 24 se da una explicación gráfica.

Dichas células se encargan de realizar el primer análisis cortical a la información recién llegada de LGN a V1. Una vez que analizan la imagen, la información obtenida se da en términos de orientaciones y de frecuencia espacial. La mayor parte de las

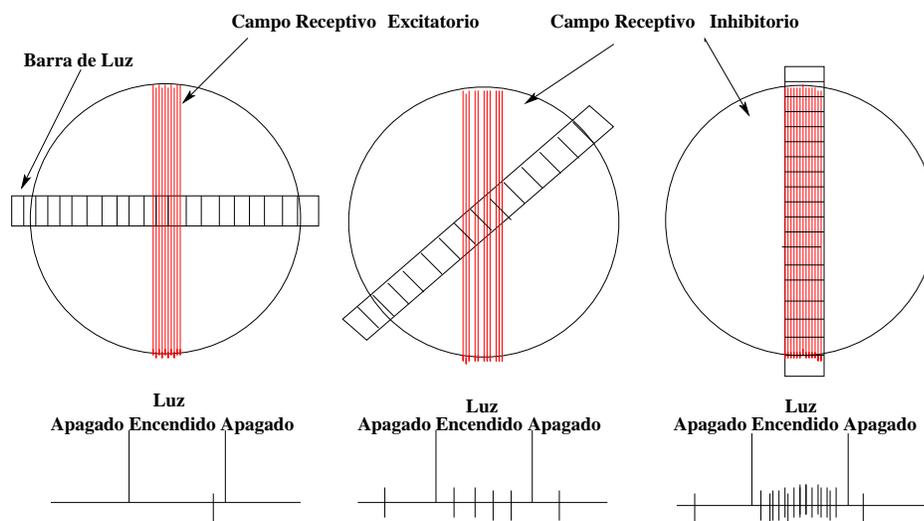


Figura 24: Campo receptivo de varias células simples en la corteza visual.

células simples se concentran en el área 17. Por su parte, las células complejas dan su mejor respuesta cuando se les presenta un estímulo en forma de barra en movimiento y con una orientación determinada (Hubel y Wiesel, 1957). Estas neuronas predominan más en el área 18.

Las células hipercomplejas están encargadas de realizar un análisis de discontinuidad de ángulos y esquinas, así como de movimiento, posición y orientación. Estas células logran definir formas geométricas, por lo que cuando se presentan figuras incompletas, el cerebro se encarga de "rellenar los huecos". A ello se deben las ilusiones ópticas que nos hace ver cosas que realmente no existen. El área 19 es la que presenta mayor cantidad de células hipercomplejas.

De forma general, los campos receptores de las células de V1 responden mejor a líneas, barras, hendiduras, bordes y ángulos con una orientación específica. En la corteza visual el procesamiento de imágenes es jerárquico. Comienza en las células simples y pasa a las complejas. Varias células simples están conectadas a una compleja

y el campo receptor de esta interacción corresponde al de las neuronas complejas. De las células complejas, las siguientes son las hipercomplejas.

Como vimos la información pasa de V1 hacia otras áreas a través de dos vías: la vía dorsal y la vía ventral. A las áreas V2 y V3 se les asocia con actividades de detección de movimiento, percepción del color, agudeza visual y percepción de la profundidad. V4 (en la corteza temporal) interviene en la percepción del color y V5 (en la corteza parietal), en la detección de movimiento. En la siguiente sección trataremos de manera superficial el funcionamiento de estas áreas.

Las áreas visuales secundarias, llamadas también extraestriadas están conformadas por las áreas V2, V3, V4 y V5, ocupando a grosso modo, las áreas 18 y 19 de Brodmann. Al igual que con la corteza visual primaria. Las áreas V2 y V3 se pueden dividir en una parte dorsal y una ventral. Las conexiones entre las áreas visuales secundarias son múltiples y recíprocas; lo que a menudo resulta incongruente con la hipótesis jerárquica. Pero se pueden distinguir tres estaciones de análisis de información específica. Por un lado, un camino que conecta el área V1 y V2 con el área V3, donde se procesa la información relativa a las formas en movimiento. Además la que viaja directamente de V1 a V4 la cual se encarga de ver el color de los objetos. Por último aquella que viaja de V1 y V2 a V5 relacionadas con el movimiento de los objetos.

III.4.4 Corteza parietal

Las áreas V3, V4 y V5 envían información además de la circunvolución temporal, a la corteza parietal posterior. Esta región también está involucrada en la percepción visual. El lóbulo parietal se ubica el área MT, media temporal o V5 y esta especializada

en procesar el movimiento líneal uniforme, además de que las neuronas de ésta área son selectivas a la velocidad. El lóbulo parietal parece tener relación con la percepción de los aspectos espaciales de la visión y de los lugares en los que se encuentran los objetos ver (Ojeda, 2004) dentro del campo visual.

En MT, la información está organizada de forma similar a V1, es decir, en columnas de orientación o dirección y son selectivas a la velocidad. Los campos receptores de algunas neuronas de MT poseen periferias antagonistas, que cuando reciben el estímulo de movimiento, disminuye la respuesta del movimiento en el centro del campo receptivo. Los campos de otras neuronas refuerzan la respuesta del centro facilitando la detección de movimiento global.

Como hemos visto la corteza parietal se encuentra dentro del área medio temporal y su tarea es aportar las claves sensoriales (especialmente visuales) para los movimientos dirigidos a un blanco. Por ejemplo, pacientes con lesiones en esta zona tienen problemas para tomar objetos sin que exista ningún problema motor. En estudios de registros de monos se ha puesto de manifiesto que las neuronas de esta área aumentan la tasa de descarga cuando el animal intenta agarrar un objeto que desea, pero no cuando realiza el mismo movimiento sin estar presente el objeto. Es por esto que se puede concluir que esta parte del cerebro es fundamental en la ruta dorsal, es decir en el proceso de detectar el lugar donde se encuentra el objeto de interés.

III.4.5 Corteza inferotemporal

El lugar más importante donde se lleva a cabo el reconocimiento de patrones visuales complejos y por ende la identificación de objetos tiene así lugar en la corteza temporal inferior o inferotemporal. Esta se encuentra localizada en la mitad ventral del lóbulo temporal. Esta área recibe influjo de la corteza prestriada y de varios núcleos talámicos. En esta región es donde convergen los análisis de forma, color, movimiento y profundidad. Se sabe además que sus neuronas responden mejor a figuras tridimensionales que a los estímulos simples, tales como puntos, líneas y rejillas sinusoidales.

En los ochentas se observó que algunas células en el lóbulo temporal respondían selectivamente a caras o elementos faciales (Perret y Mistlin, 2003), con independencia de la luminancia, matiz o tamaño de la imagen (Jeffreys, 1992). Desde entonces han sido cientos los datos recabados respecto a las peculiares características de este procesamiento. Se sabe, por ejemplo, que la actividad neuronal temporal, responde mejor ante estímulos faciales completos, disminuyendo su respuesta ante los ojos, nariz, boca, etc, cuando se presentan de forma aislada.

Sin duda la Corteza Inferotemporal (IT), es la estructura más importante para el procesamiento de imágenes. La integración procedente de las distintas áreas visuales le permite responder a estímulos más complejos, tales como barras cromáticas, patrones o escenas.

Podemos resumir que la corteza inferotemporal se encarga de identificar y clasificar en categorías los objetos, para acto seguido mandar esa información a otras regiones del cerebro principalmente a la Corteza Prefrontal.

III.4.6 Corteza prefrontal

Finalmente debe existir un lugar donde lleve acabo la unión de la información de la ruta dorsal y la ruta ventral, donde puedan interactuar en un entendimiento de escena que involucre las funciones de ambas, el reconocimiento y el desplazamiento espacial de la atención. Una región donde la interacción ha sido extensivamente estudiada es la Corteza Prefrontal (CPF) quien está bidireccionalmente conectada a ambas cortezas. Es el lugar principal donde llega la mayoría de la información de las estructuras del cerebro, La CPF también es conocida como una corteza de asociación y está implicada en una gran cantidad de procesos cognitivos como son la memoria operativa o de trabajo, las funciones ejecutivas, la toma de decisiones, la planificación del comportamiento y el procesamiento de señales emocionales por mencionar solo algunos de los aspectos en los que participa esta estructura.

La corteza prefrontal entonces controla varias etapas del procesamiento visual consciente de los humanos, integra la información proveniente de la corteza parietal y de la inferotemporal. Concluimos que esta corteza controla nuestra visión, y nos permite seleccionar lo importante de lo irrelevante de una escena visual. También interviene en la capacidad de mantener objetivos en la mente y al mismo tiempo procesar subobjetivos secundarios (Koechlin , 2003). Por último, se sabe que tiene un gran número de conexiones con estructuras motoras, así como con el colículo superior el cual esta encargado del movimiento de los ojos.

III.5 Conclusiones

En este capítulo, se ha descrito de manera general la fisiología de algunos de los componentes más importantes en el proceso de la atención visual. Los cuales abarcan desde etapas de procesamiento temprano en la retina hasta su conjunción en áreas especializadas de la corteza visual en el cerebro, cada una de estas áreas están involucradas en el procesamiento de la localización o del reconocimiento de ciertos patrones de los objetos. Este funcionamiento biológico ha servido como base para el diseño de algoritmos matemáticos que intentan emular el procesamiento de la información visual por parte de los humanos. En el siguiente capítulo se explicarán las bases matemáticas de como se llevo a cabo el diseño de algunos de los aspectos biológicos vistos.

Capítulo IV

El Modelo de Atención Visual

Nuestro sistema de atención esta basado en la implementación de (Itti y Koch , 1998) del modelo de atención Ascendente - Descendente (BU) introducido por (Koch y Ullman , 1985). Esta arquitectura BU esta diseñada para detectar regiones sobresalientes de imágenes utilizando tres características básicas que son, color, orientación e intensidad. En nuestro algoritmo incrementamos una característica más que es la simetría la cual como explicaremos más adelante tiene su motivación biológica y al mismo tiempo su uso muestra una mejoría en la clasificación de objetos.

El modelo esta basado en la Teoría de Integración de Características (ver sección II.1). De manera general lo que hace el modelo es por cada imagen de entrada calcular un solo mapa de características sobresalientes (SM por su acrónimo en inglés Saliency Map) a partir de las sumas de los mapas para color, orientación, intensidad y simetría todos ellos a diferentes escalas (ver figura 25). Una red neuronal del tipo el ganador lo toma todo compara las locaciones sobresalientes y regresa las coordenadas de la locación más prominente. Finalmente, la inhibición de retorno (IOR) es aplicada a una región de radio fijo alrededor de la región atendida en el mapa de características sobresalientes. Iterativamente se vuelve a encontrar la siguiente área más prominente, todo esto emulando el funcionamiento de las sácadas. Este modelo es considerado biológicamente como el más plausible y ha sido verificado en experimentos psicofísicos con humanos (Peters, 2003), y utilizado en problemas de reconocimiento de objetos (Bonaiuto, 2006; Walther, 2006) y problemas de navegación (Chung, 2004).

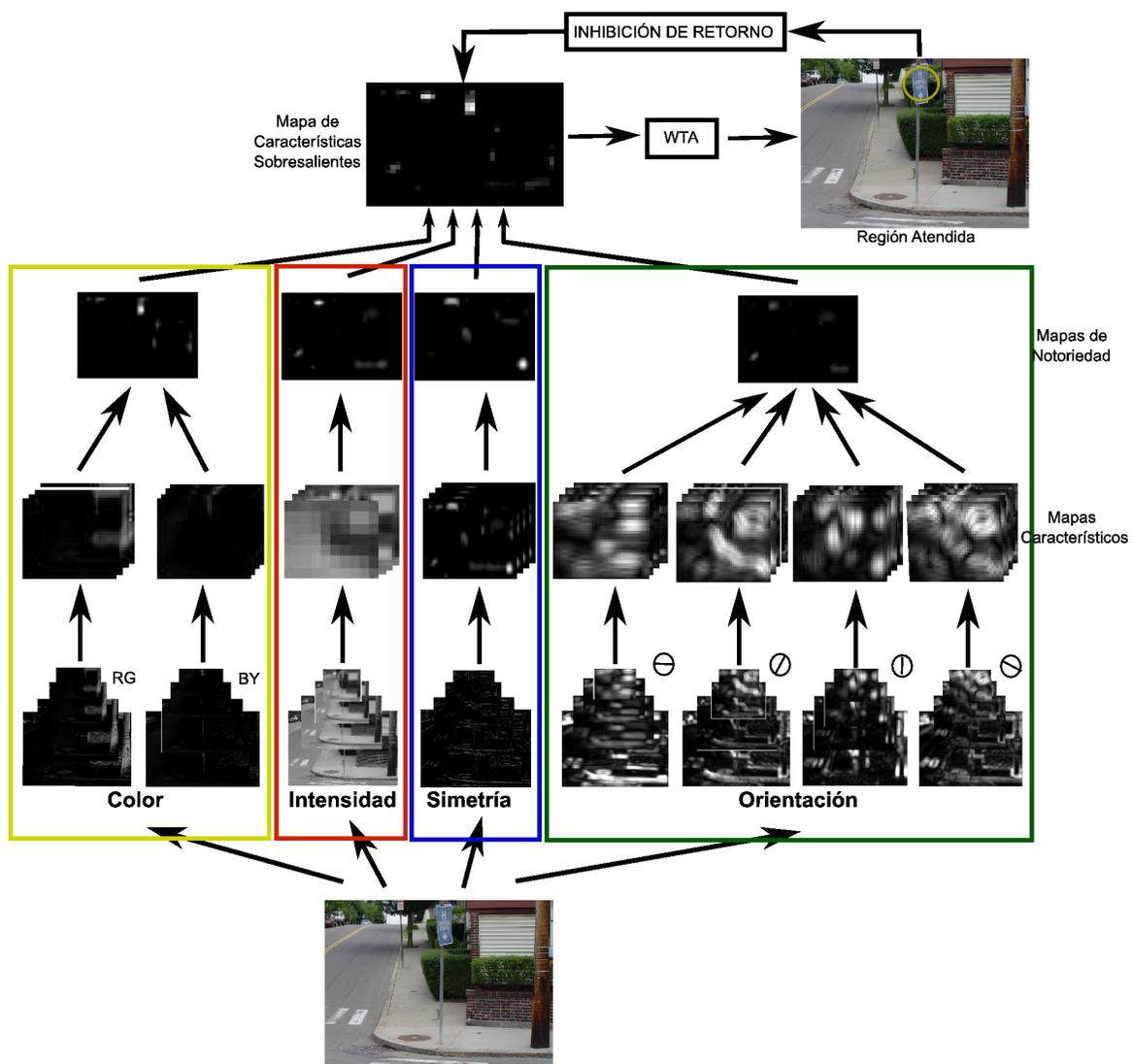


Figura 25: Modelo de atención visual propuesto

En las secciones siguientes revisaremos los detalles del modelo en orden a explicar nuestras incorporaciones en el mismo marco de trabajo. En la sección III.5 nosotros describimos nuestro método de selección de regiones sobresalientes y por último nos enfocamos a tratar las modificaciones y adaptaciones realizadas al modelo

IV.1 Cálculo preatentivo de las características visuales básicas

La primera etapa en el modelo de atención es el cálculo de las características visuales tempranas. En la visión biológica, las características visuales son calculadas en la retina, colículo superior, núcleo lateral geniculado y principios de las áreas corticales visuales (Suder y Worgother, 2000). Las neuronas en las etapas tempranas son sintonizadas para simples atributos visuales tales como intensidad, oponencia de color y orientación. Las características visuales son calculadas preatentivamente de una manera paralela en todo el campo visual tal y como lo menciona Treisman. En el algoritmo se utilizan las características de color, orientación, intensidad y aumentamos con respecto a la mayoría de los trabajos enfocados al tema la simetría. Aunque biológicamente existen otras características visuales que guían nuestra atención (Friedman y Wolfe, 1995), computacionalmente los resultados muestran que no se obtienen mejoras en el proceso de atención, utilizando otras características a las que mencionaremos a continuación.

IV.1.1 Color

La imagen de entrada (I) es descompuesta en dos oponencias de color rojo-verde(RG, Red-Green) y azul-amarillo(BY, Blue-Yellow), En 1957 Hurvich y Jameson demostraron que estas dos oponencias podían cubrir la luz entera visible (Hurvich, 1957), en otras palabras con esto se logra eliminar la influencia del brillo.

Biológicamente la codificación del color se presenta como vimos anteriormente desde

las células fotorreceptoras (conos) preparadas para recibir las longitudes de onda de los colores. Aquí se dispone de un pigmento especial que se rompe cuando percibimos el color y provoca así el potencial de acción. El color amarillo es un caso especial porque no es percibido en los conos sino en las células ganglionares de la retina.

Estas células ganglionares responden de forma oponente al rojo y verde, así como al amarillo y azul, lo hacen a través de células oponentes simples que lo que hacen es excitar un color en el centro e inhibir el color opuesto en la periferia, ver figura 26. En V1 las células que responden al color están agrupados en n glóbulos. En varios módulos de la corteza hay células que responden al color de esa parte del campo visual. De esta forma se han encontrado células oponentes en V1, V2 y V4.

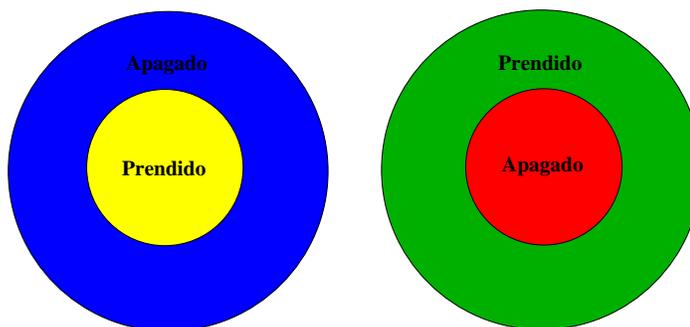


Figura 26: Células oponentes simples en la retina y en el cuerpo geniculado, para diferenciar contrastes cromáticos.

Algorítmicamente el proceso para la obtención de la oponencia de color es de la siguiente manera. Para obtener la oponencia rojo-verde se trabaja sobre el dominio espacial de la imagen, donde para cada píxel restamos los valores de los 2 colores involucrados y lo dividimos por el valor máximo entre el rojo, verde y azul del mismo píxel, como se muestra en la siguiente fórmula:

$$M_{RG} = \frac{r - g}{\max(r, g, b)}, \quad (5)$$

donde r , g y b son el valor correspondiente al color rojo, verde y azul respectivamente del píxel.

Para el caso de la oponencia amarillo-azul el caso es diferente debido a que el amarillo no es un color básico para el modelo rgb , sin embargo su obtención aparece anulando el valor correspondiente del azul, por lo que se obtiene restando el valor del píxel azul menos el valor mínimo entre el rojo-verde y dividiendo por el valor máximo entre el rojo, verde y azul.

$$M_{BY} = \frac{b - \min(r, g)}{\max(r, g, b)} \quad (6)$$

donde r , g y b son el valor correspondiente al color rojo, verde y azul respectivamente del píxel.

Por último cabe mencionar que para evitar grandes fluctuaciones de los valores de oponencia de color, las bajas luminancias, M_{RG} y M_{BY} son enviadas a cero para locaciones con valor $\max(r, g, b) < \frac{1}{10}$, asumiendo un rango dinámico entre $[0, 1]$.

IV.1.2 Intensidad

La intensidad es una medida que permite distinguir la cantidad de luz que incide en un dispositivo fotosensible. Fisiológicamente, al igual que con el color tenemos células ganglionares especializadas que tienen una forma de responder a la luz que no es homogénea. Estas responden de forma oponente a la luz y la oscuridad, de manera que cada célula ganglionar tiene un campo receptivo circular (parte del espacio visual al

que responde esa célula). Además poseen características opuestas en el centro y en la periferia, es decir, que si la iluminación provoca excitación en una célula cuando es proyectada en el centro, esa misma iluminación en esa misma célula provoca inhibición proyectada en la periferia. Por lo tanto, las células ganglionares no responden a la luz en general sino a la intensidad de luz que se proyecta sobre el centro o sobre la periferia.

El ritmo de activación de las células ganglionares es baja con iluminación débil, por lo tanto un aumento rápido de la activación indica un aumento rápido de la intensidad de la luz en el centro.

Llevarlo al diseño resulta facil, para obtener la intensidad de la imagen promediamos por cada pixel su valor rojo, verde y azul. La fórmula queda desarrollada de la siguiente forma:

$$M_I = \frac{r + g + b}{3} \quad (7)$$

donde r, g y b son el valor correspondiente al color rojo, verde y azul respectivamente del píxel.

IV.1.3 Orientación

Otra característica importante es la orientación, la cual nos permite obtener bordes con cierta inclinación angular de los objetos importantes en la imagen, la forma en que algorítmicamente está diseñada esta característica es a través de la utilización de filtros de gabor. Matemáticamente se define por el producto de un filtro pasabaja con una onda sinusoidal. Lo que al final se obtiene es una imagen segmentada donde lo que sobresale son los bordes con la orientación previamente definida. Para el caso de

nuestro modelo las orientaciones utilizadas son $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

La presencia de orientaciones en el proceso visual humano se da con la presencia de células simples y células complejas en la corteza visual primaria V1 y en la capa V2 respectivamente, y gracias a su sensibilidad hacia estímulos con cierta orientación estas permiten descomponer los perfiles de la imagen en segmentos lineales pequeños de diferentes orientaciones.

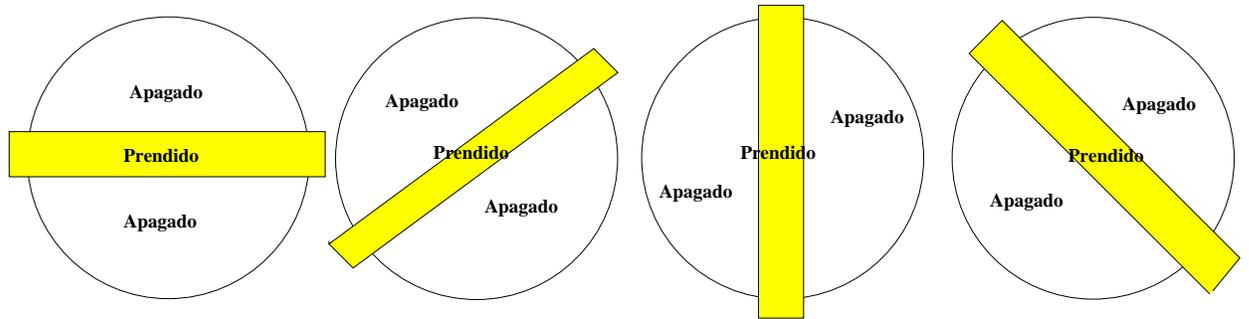


Figura 27: Células horizontales simples en V1 y V2, para estimular bordes con diferentes orientaciones.

El programa para emular esta característica realiza el cálculo de la convolución de la característica de intensidad multiplicado por el filtro de gabor; esto se hace para cada una de las orientaciones implementadas, como se muestra en la siguiente formula:

$$M_\theta = ||M_I(\alpha) * G_0(\theta)|| + ||M_I(\alpha) * G_{\pi/2}(\theta)||. \quad (8)$$

donde,

$$G_\psi(x, y, \theta) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\delta^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right).. \quad (9)$$

Esto es un filtro de gabor con relación de aspecto γ , desviación estándar δ , longitud de onda λ , fase ψ y coordenadas (x', y') transformada con respecto a la orientación θ :

$$x' = x\cos(\theta) + y\sin(\theta), y' = -x\sin(\theta) + y\cos(\theta) \quad (10)$$

Con los siguientes valores $\gamma = 1$, $\delta = 7/3$ pixeles, $\lambda = 7$ pixeles, y $\psi \in \{0, \pi/2\}$. Los filtros son truncados a 19 x 19 pixeles.

IV.1.4 Simetría

Nuestra experiencia visual es muy rica en visiones simétricas. La simetría es un importante mecanismo que ayuda a identificar la estructura de los objetos. Por lo regular los objetos que pueden ser más relevantes como personas, animales, plantas, carros, etc., tienden a tener una pronunciada simetría, por lo tanto esto es un importante aspecto en la detección BU.

Para identificar la simetría en imágenes, las frecuencias locales son analizadas a fin de determinar las simetrías y asimetrías locales. Para lograr este efecto se trabaja sobre la frecuencia de la imagen utilizando un filtro de Gabor con dos funciones periódicas diferentes. Una onda coseno se utiliza para identificar simetrías locales y por otra parte con una onda sinusoidal que encuentra las asimetrías locales. Matemáticamente, se aplica una Gaussiana en una escala logarítmica conocida como Log-Gabor. Las diferencias de estos filtros pares e impares son tomadas en algunas orientaciones, a fin de obtener un mapa de simetría.

$$M_s = Sym(x) = \frac{\sum_{k=n} ||e_n(x) - o_n(x)| - T|}{\sum_{k=n} A_n(x) + \varepsilon} \quad (11)$$

Donde e_n es la función coseno par, o_n es la función seno impar, A_n es la magnitud del filtro vector responsable, ε es un término para prevenir la división entre cero, y T es un término de compensación de ruido.

Biológicamente este proceso de simetría estaría localizado en el NLG, el cual juega un papel importante en la detección de formas e información de patrones tales como la simetría. por lo tanto, esto funciona como un preprocesador para la corteza visual la cual se encarga de encontrar una región importante (Park , 2001).

IV.2 Obtención de los mapas característicos

Una vez hecha la descomposición de la imagen original (I), en 8 imágenes de características básicas (1 para intensidad, 2 para color, 1 para simetría y 4 para orientación), cada una de ellas es submuestreada dentro de una pirámide gaussiana diádica mediante convolución con un filtro gaussiano linealmente separable y con reducción a dos. Matemáticamente lo que se hace es aplicar un filtro de convolución $k_f = [1 \ 5 \ 10 \ 10 \ 5 \ 1]/32$ sobre el dominio espacial de la imagen, en el cual ya existe una reducción a factor de 2, este procedimiento se repite 8 veces y por lo tanto obtenemos 8 imágenes de cada una de las características a diferentes niveles.

$M_I(\alpha)$, $M_{C(\Omega)}(\alpha)$, $M_S(\alpha)$, $M_{O(\theta)}(\alpha)$, donde $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, $\alpha = [0, \dots, 8]$ y

$\Omega = BY$ y RG (12)

Este proceso es repetido para obtener los siguientes niveles $\alpha = [0, \dots, 8]$ de la pirámide (Burt 1983). Por lo tanto, la resolución del nivel α es $1/2^\alpha$ veces la resolución de la imagen original. Por ejemplo, los 8 niveles tienen una resolución de $1/256$ de la imagen de entrada I y $(1/256)^2$ del número total de píxeles.

Una vez obtenidos todas las imágenes con sus diferentes escalas procedemos a realizar lo que se conoce como Diferencias Centro-Contorno, con el fin de encontrar píxeles sobresalientes con respecto de sus vecinos a diferentes escalas. El procedimiento para llevarlo a cabo es el siguiente: como primer paso al trabajar con diferentes escalas se tienen que igualar los tamaños de las imágenes, el procedimiento para igualarlos es una interpolación a escala fina, lo cual consiste en aumentar la resolución de la imagen con los píxeles que se tienen por medio de un promedio.

Una vez teniendo las imágenes del mismo tamaño vamos a calcular lo que se conocen como los mapas característicos para lo cual utilizamos la mencionada diferencias Centro-Contorno (θ), entre un centro a escala fina c y un contorno de escala gruesa s . por lo tanto, seis diferentes pares de escalas espaciales centro-contorno son utilizadas para calcular los mapas característicos de cada imagen característica, con $c \in \{2, 3, 4\}$ y $s = \{3, 4\}$:

$$F_{l,c,s} = N(|M_l(c) \theta M_l(s)|) \quad \forall l \in L = L_I \cup L_C \cup L_O \cup L_S. \quad (13)$$

con

$$L_I = \{I\}, L_C = \{RG, BY\}, L_S = \{S\}, L_O = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\} \quad (14)$$

Por lo tanto, al final tenemos un total de 48 mapas característicos calculados: 6 para la intensidad, 6 para la simetra, 12 para el color y 24 para su orientación. Los mapas obtenidos tienen la forma de píxeles de intensidad, donde su valor va a aumentar conforme sea más contrastante con respecto a sus vecinos.

El papel de los campos receptivos Centro-Contorno no solo es permitir a las células ganglionares transmitir información de si las células fotorreceptoras están expuestas a la luz. Su tarea principal es medir las diferencias en las tasas de disparo de las células en el centro y su contorno. Lo cual produce que se pueda transmitir información acerca del contraste de las características en la imagen

IV.3 Obtención de los mapas de notoriedad

Una vez obtenidos los 48 mapas se normalizan los mapas característicos. $N(\cdot)$ es un operador iterativo de normalización no lineal, que simula la competición local entre vecinos de locaciones sobresalientes. Cada paso iterativo consiste de la excitación a sí mismo e inhibición inducida por sus vecinos. Esto es implementado mediante el uso de convoluciones utilizando un filtro de "diferencias de Gaussianas", seguida por una rectificación. En el modelo utilizado el número de iteraciones es de cinco. Para más detalles ver (Itti , 2001).

Los mapas característicos son sumados mediante las combinaciones de centro-contorno utilizando sumas a través de escalas \oplus , y las sumas son normalizadas de nuevo:

$$\bar{L}_1 = N(\oplus_{c=2}^4 \oplus), \quad (15)$$

donde N es la normalización realizada y c el nivel al que se encuentra el mapa

Para las características generales de color y orientación, las contribuciones de las subcaracterísticas son sumadas y normalizadas una vez más para producir "mapas de notoriedad". Para la intensidad y la simetría, el mapa de notoriedad es el mismo como en F_I y F_S respectivamente, el cual es obtenido a partir de la ecuación 12.

$$C_I = \bar{F}_I, C_S = \bar{F}_S, C_C = N\left(\sum_{l \in L_C} \bar{F}_I\right), C_O = N\left(\sum_{l \in L_O} \bar{F}_I\right) \quad (16)$$

Todos los mapas de notoriedad son combinados dentro de un único mapa de sobresalencia como se muestra en la siguiente sección.

IV.4 Mapas de Características Sobresalientes

El mapa de características sobresalientes (SM por el acrónimo en inglés de Saliency Map) define el lugar de las locaciones de la imagen más importantes y por ende en donde la atención podría ser dirigida en algún momento dado. La obtención de este mapa es propuesto como un simple promediado de los mapas de notoriedad previamente calculados.

$$S = \frac{1}{4} \sum_{k \in \{I, C, O, S\}} C_K. \quad (17)$$

Cuando se definió el concepto de SM por Koch y Ullman, se realizó en términos de procesos neuronales más que de procesos cognitivos. Anatómicamente no se conoce con precisión en que parte del cerebro esta localizado el SM. Se sabe que no es lógicamente necesario que surja en una locación particular y podría ser entendido como un mapa funcional donde sus componentes podrían estar distribuidos en varias áreas del cerebro. Así Koch y Ullman propusieron que el SM se encontraba en el NLG. Otro núcleo talámico, el pulvinar, es conocido por estar involucrado en la atención (Robinson, 1992) y también se ha sugerido como un candidato para albergar al SM. Otra posibilidad es el colículo superior igualmente conocido por estar involucrado en el control de la atención. También algunas áreas neocorticales han sido sugeridas, incluyendo V1 (Li, 2002), V4 (Mazer, 2003) y la corteza parietal posterior (Gottlieb, 2007).

IV.5 Selección de la región de interés

Una vez contando con el SM, se procede a obtener el pixel con el valor más alto de sobresaliencia. La forma más obvia de obtenerlo sería con una simple búsqueda del valor más alto sobre la matriz del SM. Sin embargo, Koch prefirió usar una red neuronal del tipo el ganador lo toma todo (WTA), la cual es una red neuronal del tipo no supervisado, donde no se conoce la salida. El objetivo de usar la red neuronal WTA, es porque los tiempos de reacción son muy semejantes a los que maneja el ojo humano y segundo porque emulan la competición de las neuronas por recibir la atención,

o región de interés se logra mediante un proceso de retroceso al píxel (K_w) en el mapa característico ya sea de intensidad, contraste, orientación o color, que haya sido el más prominente, en otras palabras que tenga el valor más alto.

$$K_w = \operatorname{argmax}_{k \in \{I, C, O, S\}} C_K(x_w, y_w). \quad (18)$$

Una vez que nos encontramos en este mapa, el píxel se empieza a unir con los píxeles vecinos que tengan algún valor, es decir que no esten de color negro. Una vez logrado esto se umbraliza el mapa para borrar otras áreas donde pudiera haber valores diferentes de cero y por último se procede a la aplicación de una convolución gaussiana para esparcir el denominado protoobjeto, en la figura 29, mostramos los 3 casos de selección de región de interés



Figura 29: Diferentes modelos simbolicos de la obtención de la región de interés.

Para los fines de este trabajo ha sido modificada la región de interés la cual va a corresponder a un cuadrado que cubra el área del protoobjeto. Para realizarlo se selecciona el punto más lejano del protoobjeto al píxel prominente k_w , y se calcula la distancia (d), con esta información se genera un cuadro con área d^2 y centro en k_w . En

la figura 32 se muestra una explicación gráfica.

Por último el área atendida es inhibida. Volviendo al SM donde se produce de nuevo la competición de la red neuronal WTA por la segunda locación más sobresaliente, la cual es atendida y subsecuentemente inhibida, esto se sigue haciendo de manera iterativa hasta terminar las zonas sobresalientes, esto permite al modelo simular el movimiento sacádico del ojo en la imagen en orden a decrementar la importancia de las locaciones atendidas.

IV.6 Conclusión

En este capítulo se abordaron los conceptos matemáticos del desarrollo del modelo de atención visual, se dio una explicación biológicamente plausible para cada paso que se llevo a cabo en el diseño del algoritmo, además de explicar algunos aspectos de como ha evolucionado dicho modelo, con respecto a la región de interés mostramos un mecanismo para extraer una región de la imagen cerca del foco de atención (FOA) que corresponde a una extensión aproximada de un objeto en esa locación.

Capítulo V

Implementación del Modelo

Recordando el objetivo de esta tesis, el cual es:

Generar una nueva solución al problema de reconocimiento de clases de objetos evolucionando de manera conjunta tanto la obtención de la región de interés como la de su descriptor asociado. Mediante el uso de un modelo de atención visual y un descriptor basado en el histograma de orientaciones.

En este capítulo se muestra el proceso que se llevó a cabo para cumplir con dicho objetivo. Una vez realizada la implementación del sistema, se presentan las imágenes de prueba donde se realiza el reconocimiento de distintos objetos como peatones, carros, señales de tránsito, bicicletas, etc. Donde lo que se pretende es que el sistema sea entrenado con la finalidad de que sea capaz de hacer la búsqueda y poder detectar de manera independiente la presencia o no de cada una de estas clases de objetos en imágenes distintas a las utilizadas en la fase de entrenamiento.

V.1 Proceso para el reconocimiento de clases de objetos

El proceso general que sigue nuestro modelo para realizar la clasificación de los objetos en imágenes está compuesto de cinco etapas (ver figura 30):

- 1. Adquisición de los datos
- 2. Selección de la región de interés.
- 3. Segmentación de la imagen o región de interés.
- 4. Extracción de rasgos o características (Descripción).
- 5. Reconocimiento y clasificación de objetos.

De manera general hablaremos del procedimiento de nuestro diseño. Además en las secciones posteriores se describirán con detalle cada uno de los pasos que se han llevado a cabo para realizar este trabajo. Las imágenes utilizadas fueron obtenidas del MIT, éstas corresponden a las calles de Boston y las utilizaremos para reconocer de manera independiente objetos como peatones, carros, señales de tránsito y bicicletas. La selección de la región de interés la llevamos a cabo mediante el modelo de atención visual descrito en el capítulo anterior, el modelo ha sido evolucionado para realizar la búsqueda de objetos de la misma clase que deseamos clasificar, con lo que obtenemos una región de interés en la imagen. El tratamiento que se le da al área enfocada va a depender del resultado del mismo programa genético el cual al mismo tiempo ha sido evolucionado para encontrar el mejor descriptor. La obtención de dicho descriptor se realiza mediante la aplicación de operadores sobre la imagen original de la cual podamos obtener un descriptor basado en su histograma de orientaciones de los píxeles de la región, de tal suerte que nos permita de mejor manera separar y clasificar a los objetos. Por último, el proceso de clasificación lo realizamos con el uso de una máquina de soporte vectorial, la cual además tiene un papel importante en el proceso de la evolución de nuestro diseño al jugar el rol de la función objetivo de acuerdo a su porcentaje de clasificación. En las siguientes líneas veremos con mayor énfasis la

implementación del sistema.

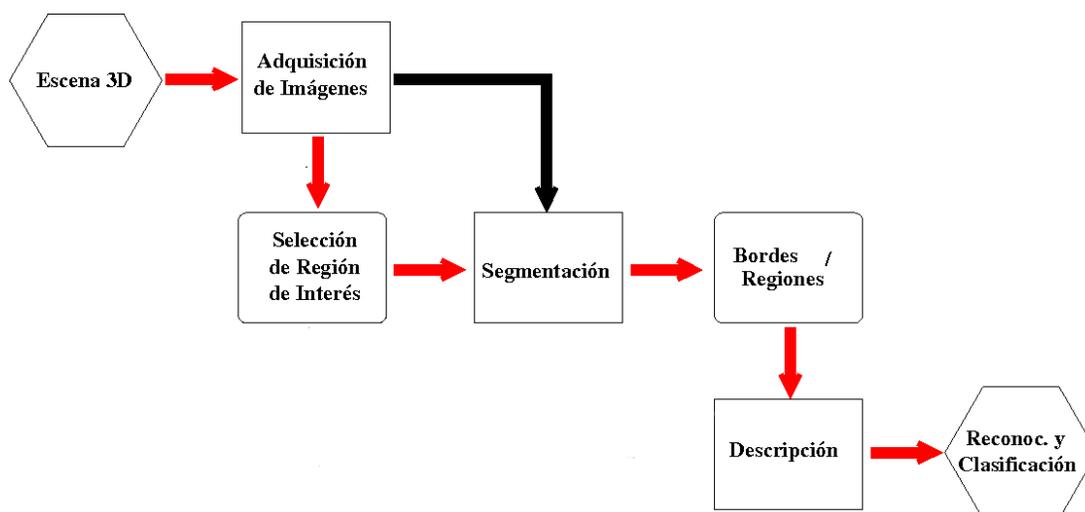


Figura 30: Trayectoria realizada para la clasificación de objetos.

V.1.1 Adquisición de los datos

Las imágenes elegidas pertenecen al Center for Biological Computational Learning (CBCL) un laboratorio correspondiente al MIT. Ellos crearon una base de datos de las calles de Boston con una cámara DSC-F717. Estas imágenes fueron utilizadas para un trabajo donde probaban varios tipos de segmentación. Aunque también han sido utilizados en proceso específicos de reconocimiento de objetos como el reporte técnico de Chikkerur (2009) donde a través de redes bayesianas y características basadas en forma predicen la localización de los objetos.

En nuestro trabajo partimos de esta base de datos, tomando como entrada 50 imágenes para cada clase utilizada. Las dimensiones de las imágenes son de 640 x 480 y el formato en el que se encuentran es png. En la figura 32, mostramos algunas de las

imágenes utilizadas.

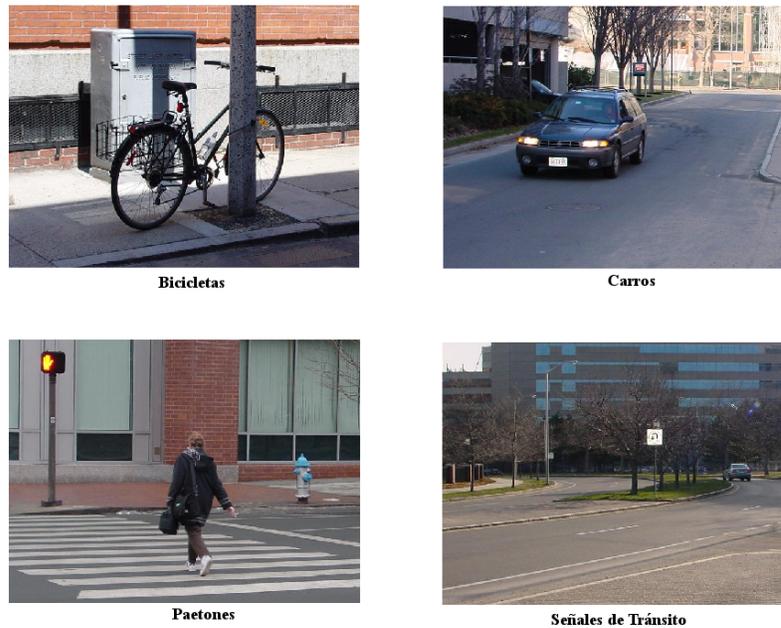


Figura 31: Ejemplo de cuatro diferentes tipos de clases utilizadas en el diseño de nuestro algoritmo.

V.1.2 Selección de la región de interés

Una vez teniendo las imágenes se lleva a cabo la selección de la región en donde se encuentra el objeto de interés. En nuestro caso el objeto que queremos clasificar. La selección del área la realizamos mediante el modelo de atención visual descrito en el capítulo anterior, el modelo ha sido evolucionado modificando la forma de combinar los mapas de notoriedad para formar un nuevo SM. Anteriormente la forma de combinar

los mapas se realizaba mediante el promedio de estos mapas como se muestra en la Ecuación 17, en este trabajo se propone una nueva forma de combinar estos mapas para dirigir la atención hacia las zonas de interés. Dicha ecuación cambiará dependiendo de la clase de objetos en la que queramos especializar la búsqueda. La mejor forma de combinar estos mapas la eligirá un programa genético el cual trabajará con la combinación de los 4 mapas de notoriedad $\{C_I, C_C, C_O, C_S\}$ y con algunas operaciones extra aparte de la suma los cuales son: $\{+, -, *, \parallel, \sqrt{I_t}, \frac{I_t}{2}, \log_2(I_t)\}$. Sin duda, este simple cambio permitió dirigirse a zonas distintas de las que se lograrían con el simple promedio. En la sección IV.1.5 se darán mas detalles de la forma de obtener esta ecuación.

Una vez contando con el píxel ganador $P_w = (x_w, y_w)$ al igual que en el modelo de Walther se procede a obtener el protoobjeto, (ver capítulo III.5). Una vez obteniendo el protoobjeto generado del SM, se procede a encerrarlo dentro de un cuadrado. La forma de obtener este cuadrado es mediante la localización del punto más lejano del perímetro del protoobjeto nombrado como P_{po} al píxel prominente P_w . Una vez obtenido ese punto se calcula su distancia (d) a P_w . Esa distancia nos servirá de radio para generar un cuadrado de lado $2d$ y con centro en P_w . En la siguiente figura veremos una explicación gráfica de como obtener esta región.

La obtención del cuadrado no genera el área mínima, ni garantiza que se cubra toda la región que engloba al objeto. Lo que se busca con este cuadro es cubrir la zona a fin de que baste para que el descriptor pueda obtener la información suficiente que le permita separar de forma correcta las clases de objetos. En el siguiente capítulo abordaremos la forma de obtener el descriptor y veremos porque es necesario que la región tenga forma rectangular.



Figura 32: Se muestra la forma de obtener la región de interés, primero se calcula la distancia (d) y posteriormente se calcula el cuadrado que circunscribe al protoobjeto.

V.1.3 Extracción de características

La extracción de características se realiza mediante la utilización del histograma de orientación de una imagen transformada que como hemos mencionado anteriormente permanece robusto ante cambios de intensidad, escala y orientación de la región de la imagen; fenómeno muy frecuente en las imágenes naturales. Como vimos en la introducción este descriptor regresa un vector de 128 valores que viene dado de las 16 subregiones de tamaño 4×4 en las que se divide la región, así como de las 8 orientaciones posibles a la que se puede asignar un valor.

Lo interesante en el método de extracción de características es la aplicación a través de la programación genética de múltiples operadores sobre la región seleccionada que nos permitirá obtener una infinidad de imágenes procesadas en donde el descriptor

podrá basar su descripción sobre la imagen transformada que le ofrezca mejores resultados. En la figura 35, se muestran los resultados de la aplicación de dos posibles operadores de la misma imagen. La obtención de estos operadores se explicara a más detalle en la sección de la implementación del GP.

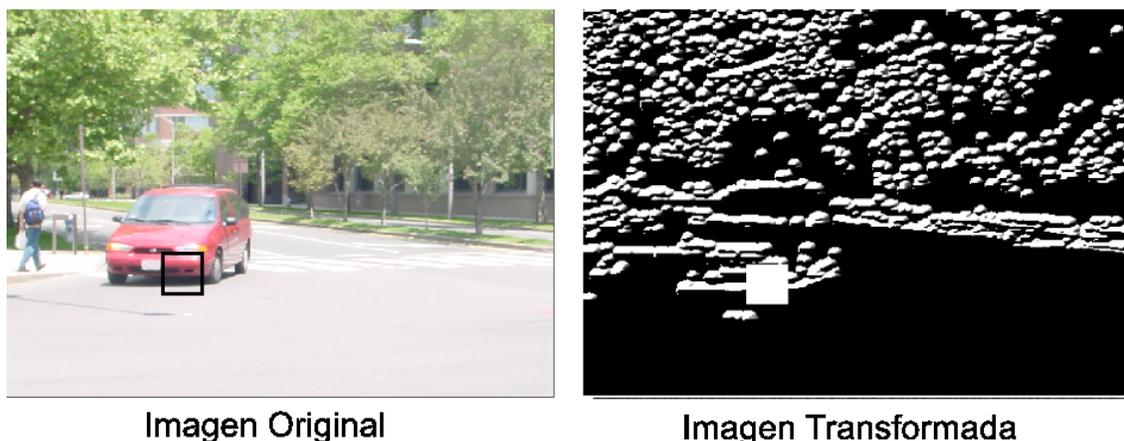


Figura 33: Ejemplo de una selección de región de interés y una posible transformación de la imagen.

Una vez seleccionada la región de interés y después de haber obtenido la imagen transformada de la región de interés se procede a aplicarle una ponderación Gaussiana con el fin de darle mayor importancia a los píxeles más cercanos y demeritar en cierta manera a los que se encuentran en la periferia. Por último, se procede a obtener el descriptor del histograma de orientación sobre la región seleccionada como se muestra en la sección II.2.

V.1.4 Reconocimiento y clasificación de objetos

Una vez realizado el proceso de detección del objeto y de la extracción de características, lo que obtenemos es un vector unidimensional de 128 valores por cada objeto

perteneciente a nuestra clase. El siguiente paso consiste en recabar la suficiente información. En este caso la colección de varios vectores de diferentes objetos de la misma clase con lo cual podamos entrenar nuestra máquina de soporte vectorial. Esto permite de manera automática clasificar otro descriptor perteneciente a otro objeto por lo que se puede decidir si pertenece o no a la clase.

El primer proceso de entrenamiento de la SVM consistió en la recolección de 50 objetos de la misma clase, los cuales fueron optimizados con el uso de un programa genético como se vera en el siguiente capítulo para describir la mejor área con la que se obtuvieran los mejores resultados. Como segundo paso se describieron 50 objetos no pertenecientes a la clase pero que si pertenecian a las demás con lo cual lo que se pretende es lograr un clasificador más robusto que evite confundirse con distractores.

Por lo tanto podemos resumir que el reconocimiento y la clasificación lo realizamos con la conjunción de un modelo de atención visual, un descriptor del histograma de orientaciones y la máquina de soporte vectorial. Sin embargo, la utilización y optimización de cada uno de estos pasos se explicará en el siguiente sección.

V.1.5 Optimización de la clasificación de objetos utilizando programación genética

Como se ha visto a lo largo de este trabajo, dos puntos son fundamentales en el proceso de clasificación de objetos. Por una parte, el encontrar la región de interés donde se encuentra el objeto a clasificar, y el segundo punto es encontrar un descriptor robusto de esa región de interés con el fin de que nos facilite la clasificación de manera correcta de los objetos de la misma clase.

Es por eso que en el diseño de nuestro algoritmo hemos optado por la implementación de un programa genético que nos optimice estos dos puntos, es decir, que evolucione hasta encontrar la mejor forma de combinar las características de nuestro modelo de atención visual a fin de enfocarse en la búsqueda de la región del objeto que queremos clasificar. Al mismo tiempo, queremos evolucionar el operador que se le aplicará a la imagen con el fin de encontrar el histograma de orientación que mejor nos permita identificar a los objetos. Para realizar este proceso hemos implementado un programa genético que nos proporcione una manera automática de encontrar estos parámetros.

En la fase de entrenamiento se pasan las imágenes a clasificar pertenecientes a las dos clases y el programa evolutivo va aprendiendo a clasificar hasta poder obtener los mejores resultados, que en nuestro algoritmo va a ser los operadores para la transformación de la imagen y la localización del área de interés. en la siguiente figura se muestra un diagrama del proceso de entrenamiento.

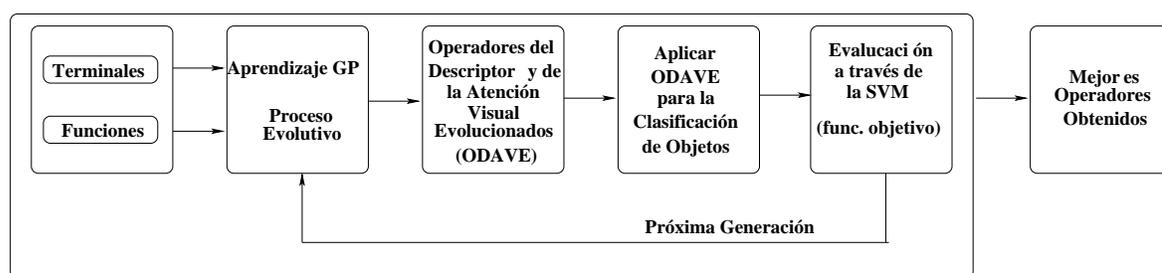


Figura 34: Diagrama de flujo de la fase de aprendizaje del programa genético implementado.

La implementación del algoritmo se realizó de la siguiente forma, como primer paso se propone la evolución de manera conjunta tanto de la selección de la región de interés

como del descriptor de la imagen. Por lo tanto, se generaran 2 genes en el cromosoma el primero viene dado por la información perteneciente a la selección del área y el segundo para aplicarles operaciones a la imagen para su futura extracción de información. En la siguiente figura se muestra un ejemplo de la composición del cromosoma.

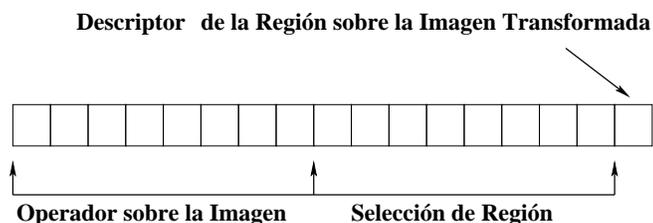


Figura 35: Ejemplo del cromosoma implementado.

Representación, espacio de búsqueda y operaciones genéticas

Como se explicó en la sección anterior, cada gen esta constituido por un árbol que representa la fórmula que se va a aplicar ya sea para obtener el operador o la región de interés. Se puede intuir que el conjunto de funciones y terminales va a ser distinto en cada árbol (ver figura 37). Entonces por lo único que van a estar unidos es por la operación final que es la descripción de la zona de interés.

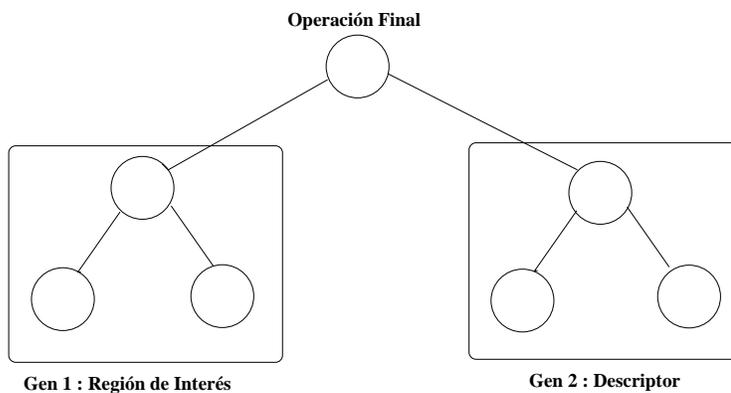


Figura 36: Los dos árboles con los que cuenta nuestro algoritmo genético.

La representación de la estructura a la que podrá evolucionar nuestro programa dependerá de las funciones y terminales con las que se entrena el GP, en el caso del primer gen o del operador sobre la imagen se tienen los antecedentes de (Perez y Olague, 2009) quiénes han logrado con las funciones y terminales que a continuación se describirán grandes mejoras en la detección de puntos de interés a comparación de los logrados por detectores del estado de arte como el SIFT o el SURF, Perez y Olague demostraron que con la utilización de filtros diferenciables, convoluciones con Gaussianas y el uso de las operaciones básicas las mejoras con respecto a descriptores del estado del arte son mayores. Nosotros decidimos entonces usar tales ideas para establecer nuestros conjuntos de funciones y terminales como sigue:

$$F = \{ +, -, /, *, \| + \|, \| - \|, \sqrt{I_t}, \frac{I_t}{2}, \log_2(I_t), D_x G_\sigma, D_y G_\sigma, G_\sigma \}$$

$$T = \{ I, D_x(I), D_y(I), D_{xx}(I), D_{yy}(I), D_{xy}(I) \}$$

Donde I es la imagen de entrada e I_t puede ser alguna de las terminales en T , como la salida de cualquiera de las funciones en F ; D_u simboliza la derivada de la imagen a lo largo de la dirección u . $D_u = I * G_{u(\sigma=1)}$; G_σ son filtros de alisamiento gaussiano con $\sigma = 1$ o 2 ; $D_u G_\sigma$ representa la derivada de un filtro Gaussiano con una imagen borrosa σ . De manera general estas son las principales funciones que se les puede realizar a la imagen para obtener un nuevo operador.

En la segunda parte el caso es distinto, las terminales y funciones van relacionadas con la elección de la región de interés. En este caso las terminales vienen siendo los 4 mapas de notoriedad. Así para las funciones utilizadas tenemos las operaciones básicas como sumas, restas, multiplicación, etc. En las siguientes líneas definimos las funciones y terminales para esta parte del gen.

$$F = \{ +, -, /, *, \| + \|, \| - \|, \sqrt{(I_t)}, \frac{I_t}{2}, \log_2(I_t) \}$$

$$T = \{ C_I, C_S, C_C, C_O \}$$

Función objetivo

Para saber que tanto va evolucionando nuestro modelo necesitamos algo que nos ayude a medir o saber como se esta comportando nuestro sistema, esta es la labor de la función objetivo. En nuestra implementación para evaluar a nuestro modelo la función objetivo viene dada por el porcentaje de clasificación que nos arroje la máquina de soporte vectorial en la fase de entrenamiento. Esta representa el número de objetos que clasificó de manera correcta del total de objetos que se estan clasificando.

Función objetivo = número de objetos clasificados de manera correcta / número total de objetos (19)

Inicialización y parámetros del GP

Una vez definidos el espacio de búsqueda y la función objetivo, el siguiente paso es iniciar el proceso de evolución aleatoriamente. El método de inicialización utilizado es el conocido como "Ramped Half and Half", en el cual se define una probabilidad por cada tamaño, en nuestro caso entre 2 y 5 por lo que se crearan un número igual de árboles con una profundidad especifica. Por lo tanto, el 25% de la población inicial tendrá profundidad 2, el 25% 3, el 25% 4, y el 25% tendrá profundidad 5.

El número de generaciones es igual a 20, y el número de individuos por cada generación es de 50. Los parámetros de cruzamiento y mutación vienen dados por los valores de

0.90 y 0.10 respectivamente. Un detalle a tomar en cuenta en estos dos parámetros es el hecho de que el cruzamiento y mutación solo se puede generar sobre un solo gen, es decir, no se pueden cruzar 2 individuos intercambiando valores del gen 1 con los del gen 2 debido a que nuestro cromosoma podría contar solo con información del operador de la imagen o en el otro caso pura información para encontrar la región de interés, en la siguiente figura se muestra un ejemplo de como se lleva acabo este proceso en nuestra implementación.

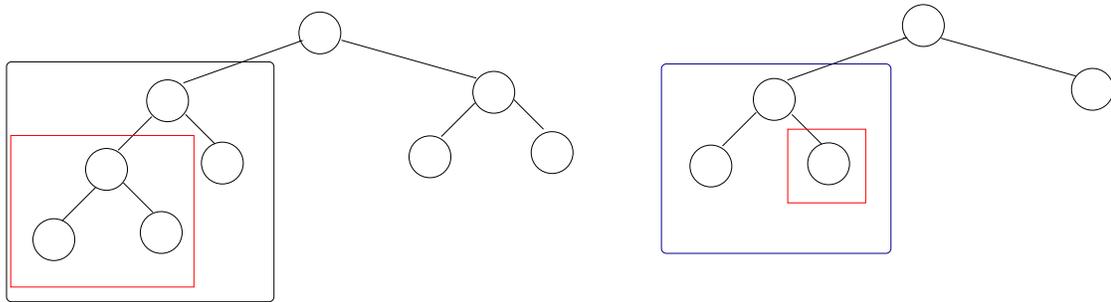


Figura 37: Se muestra el proceso de cruzamiento, donde se muestra que la combinación solo se puede llevar a cabo entre nodos del mismo lado de la raíz.

Otros valores de nuestro GP son la profundidad máxima del árbol la cual es de 9 niveles. Para la selección de los individuos se eligió el Muestreo universal estocástico, que funciona generando las copias necesarias en una sola selección; para esto se ayuda de apuntadores. Por último, en la tabla III hacemos un resumen de los parámetros del GP implementado.

Tabla III: La tabla muestra los parámetros utilizados en la implementación del Programa Genético

Parámetros	Descripción.
Generaciones	25
Tamaño de la Población	30 individuos
Inicialización	Ramped Half and Half
Cruzamiento	0.90
Mutación	0.10
Profundidad del Árbol	Selección de Profundidad Dinámica
Profundidad Máxima Dinámica	7 niveles
Profundidad Máxima Real	9 niveles
Selección	Muestreo Universal Estocástico
Elitismo	1/50

Capítulo VI

Resultados experimentales

VI.1 Herramientas de trabajo

En esta sección se mencionan las características de la computadora y del software utilizado para la realización de la tesis.

VI.1.1 Unidad central de proceso

- Computadora de Escritorio, marca DELL, memoria ram de 4 GB, Procesador Quad Core, disco duro de 120 Gb,
- Tarjeta de Video Geforce Fx 5200.
- Sistema Operativo: Linux (distribucion suse 11.2 de 32 bits).

VI.1.2 Software

Para la implementación nos basamos de varios programas computacionales los cuales seran descritos a continuación:

- Matlab: Software especializado en matemáticas que cuenta con su propio lenguaje de programación, todo el diseño del trabajo esta implementado en este software.
- Image Processing Toolbox: Es una colección de funciones de matlab que contiene una variedad de algoritmos y herramientas gráficas enfocadas en el procesamiento digital de imágenes.

- Saliency Toolbox: Es una colección de funciones en Matlab y scripts para calcular el mapa de características sobresalientes de una imagen. El saliency toolbox es una reimplementación del toolkit de iNVIT desarrollado en el laboratorio de Laurent Itti en la Universidad del Sur de California. El código fue originalmente desarrollado por Dirk Walther en su tesis doctoral en el laboratorio de Koch en el Instituto tecnológico de California y ha sido citado en más de cien publicaciones.
- GPLAB Toolbox: Este software esta conformado por una serie de funciones usadas en programación genética.
- Support Vector Machine Toolbox: Contiene un conjunto de rutinas para la clasificación utilizando máquinas de soporte vectorial.

VI.2 Experimentos

En esta sección se muestran los resultados de aplicar nuestro algoritmo en problemas concretos.

VI.2.1 Prueba del algoritmo propuesto

Fase de entrenamiento para la clase carro

Como primer paso en la evaluación de nuestro modelo se probó con la clase (carro), el experimento consistió en el entrenamiento de nuestro programa evolutivo con 100 imágenes, 50 en donde se incluían a carros y 50 donde no los había, en estas últimas 50 imágenes se incluyeron objetos de otras clases (distractores) con lo cual se busca que el programa sea más elitista. Es decir que pueda enfocarse en buscar coches aún cuando se tengan objetos sobresalientes de otras clases.

El programa genético se inicio con los parámetros mostrados en la última sección del capítulo anterior; 25 generaciones, 50 individuos, con cambio generacional guardando solo al mejor individuo. La SVM que se utilizó fue utilizando un kernel lineal para lograr separar de mejor manera nuestras clases, con una mutación del 10% y un cruceamiento del 90%. Así los resultados fueron los siguientes:

Como primer paso se realizaron 5 corridas del programa genético con los mismos parámetros de entrada, lo que se buscó con esto es obtener al mejor individuo en diferentes evoluciones, descartando desventajas que se pudieran tener al hacer una sola corrida del programa genético; como es el estar condicionado la calidad en función de la población inicial. En la siguiente gráfica se muestran los resultados obtenidos en las 5 ejecuciones, donde lo que se muestra es al mejor individuo obtenido, podemos observar que el mejor individuo fue encontrado en la tercera corrida como se muestra en la gráfica 38.

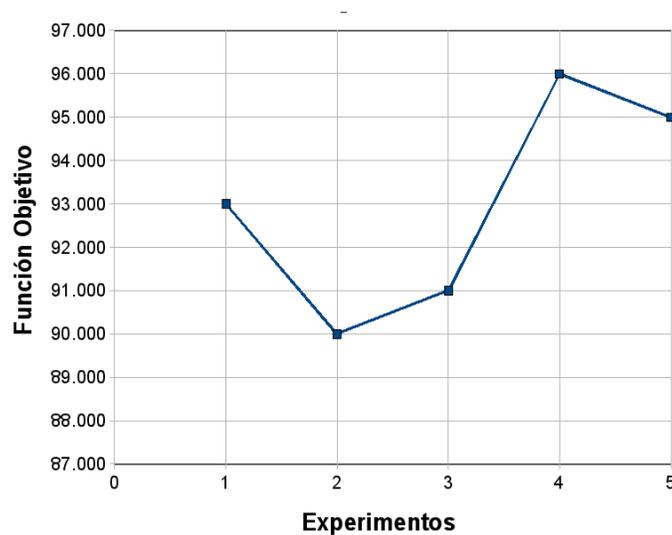


Figura 38: Mejores individuos por cada ejecución del programa para realizar la clasificación de carros

En la gráfica 39 se muestra la evolución del mejor individuo obtenido a través de las generaciones; también podemos ver el promedio en cada una de las mismas. Con el fin de conocer la generación donde ya no se obtiene una mejora en el desempeño de nuestros individuos.

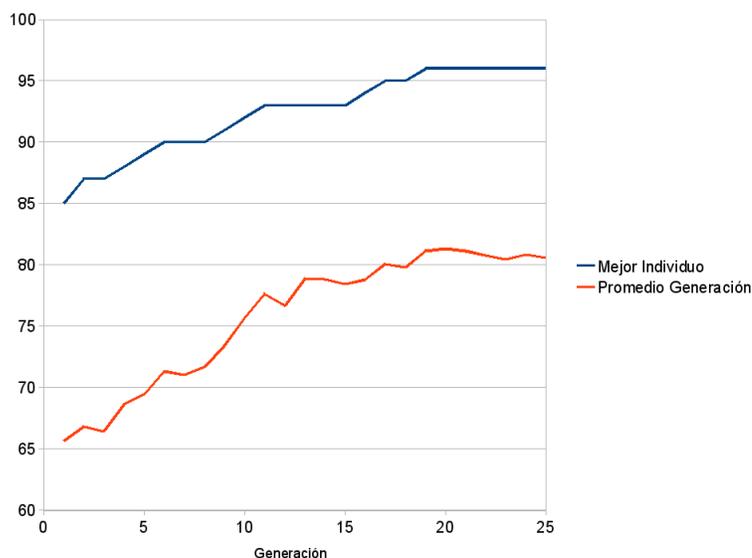


Figura 39: Mejor individuo y promedio de la función de desempeño por cada generación para la clasificación de carros.

El mejor individuo encontrado para la solución de nuestro problema de búsqueda y clasificación de carros, esta compuesto por el diagrama de árbol que se muestra en la siguiente figura 40.

Por lo tanto se puede concluir que la mejor forma de combinar los mapas de notoriedad se consiguio con la siguiente formula:

$$SM = \log(\sqrt{(C_S)} + C_C) * \sqrt{(C_O)} \quad (19)$$

donde C_s , C_c y C_o , corresponden a los mapas de notoriedad en cuanto a simetría, color y orientación respectivamente.

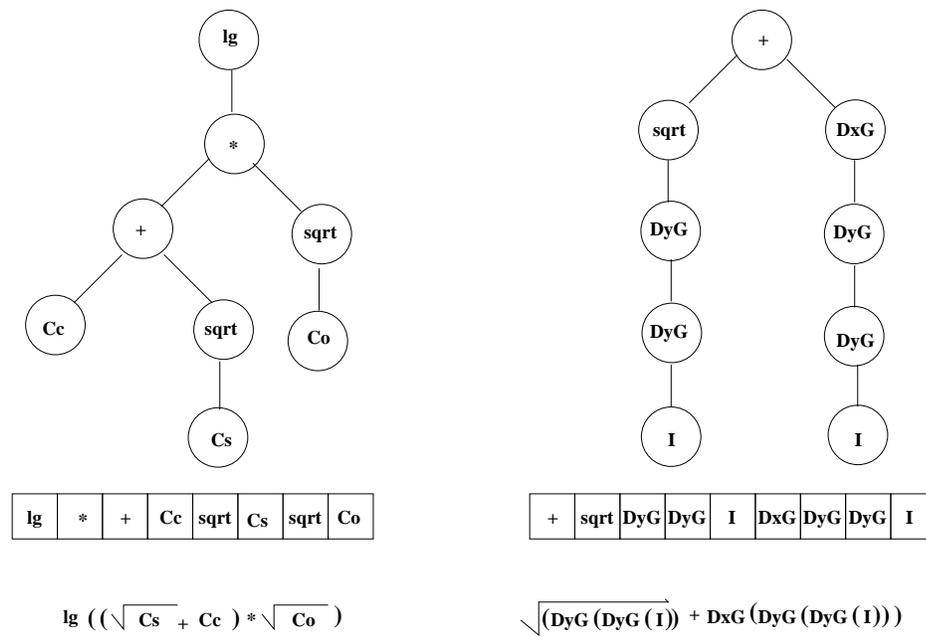


Figura 40: Mejor individuo encontrado para el reconocimiento y clasificación de carros

Y por otra parte, el mejor individuo arrojó el siguiente operador como resultado:

$$Op = \sqrt{(DyG(DyG(I)))} + DxG(DyG(DyG(I))) \quad (20)$$

donde I, es la imagen de entrada.

Por último, se muestra un ejemplo de la aplicación del operador de transformación de la imagen y el generador de la región de interés sobre una de las imágenes con las que se entrenó. ver figura 41.

Fase de prueba para la clase de carro

La fase de prueba consiste en examinar al individuo ganador en imágenes distintas que con las que se realizó el entrenamiento y corroborar que tan efectivo es nuestro nuevo

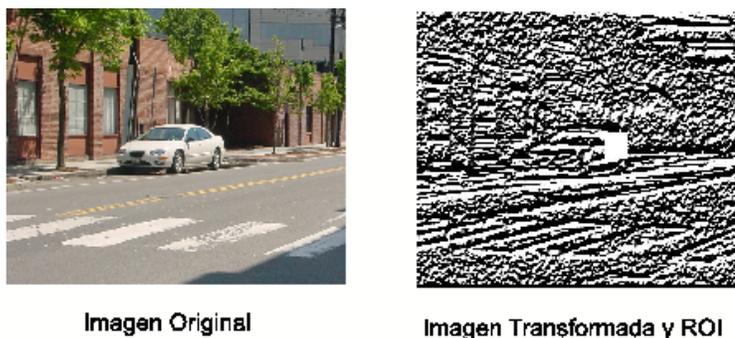


Figura 41: Se presenta un ejemplo de la aplicación del individuo ganador sobre una de las imágenes de entrenamiento.

operador de descripción y región de interés. Para este proceso se volvieron a usar 100 imágenes 50 en donde se encontraban automoviles y 50 en donde no los había y se ejecutó el programa, los resultados finales en el proceso de reconocimiento de carros se muestra en la siguiente tabla de confusión.

Tabla IV: Matriz de confusión con los resultados obtenidos del mejor individuo en la clasificación de carros.

	Carros	No Carros.
Carros	33	17
No Carros	12	38

Por lo tanto en la fase de prueba se obtuvo una clasificación del 71%. Como prueba final, se ha corroborado que la región de interés encontrada por nuestro algoritmo evolucionado corresponda con la ubicación del objeto, para realizar este procedimiento se segmento de manera manual la región ocupada por los objetos de prueba. Se llevo a cabo la prueba en primera instancia como la simple intersección entre el área del protoobjeto y el objeto. Dando como resultado que los objetos que interceptaban más del 40% del protoobjeto eran reconocidos y clasificados de manera satisfactoria (ver

figura 42).

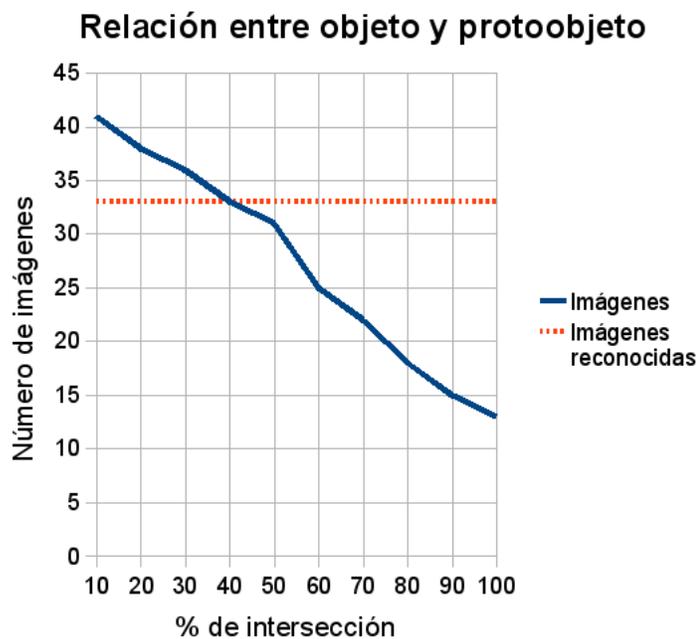


Figura 42: Gráfica con el porcentaje de intersección entre los objetos segmentados de la clase carro y el protoobjeto encontrado por nuestro sistema.

Por último en la figura 43 y 44 se muestran algunas de las imágenes utilizadas para entrenar y probar la búsqueda y reconocimiento de carros en nuestro modelo.

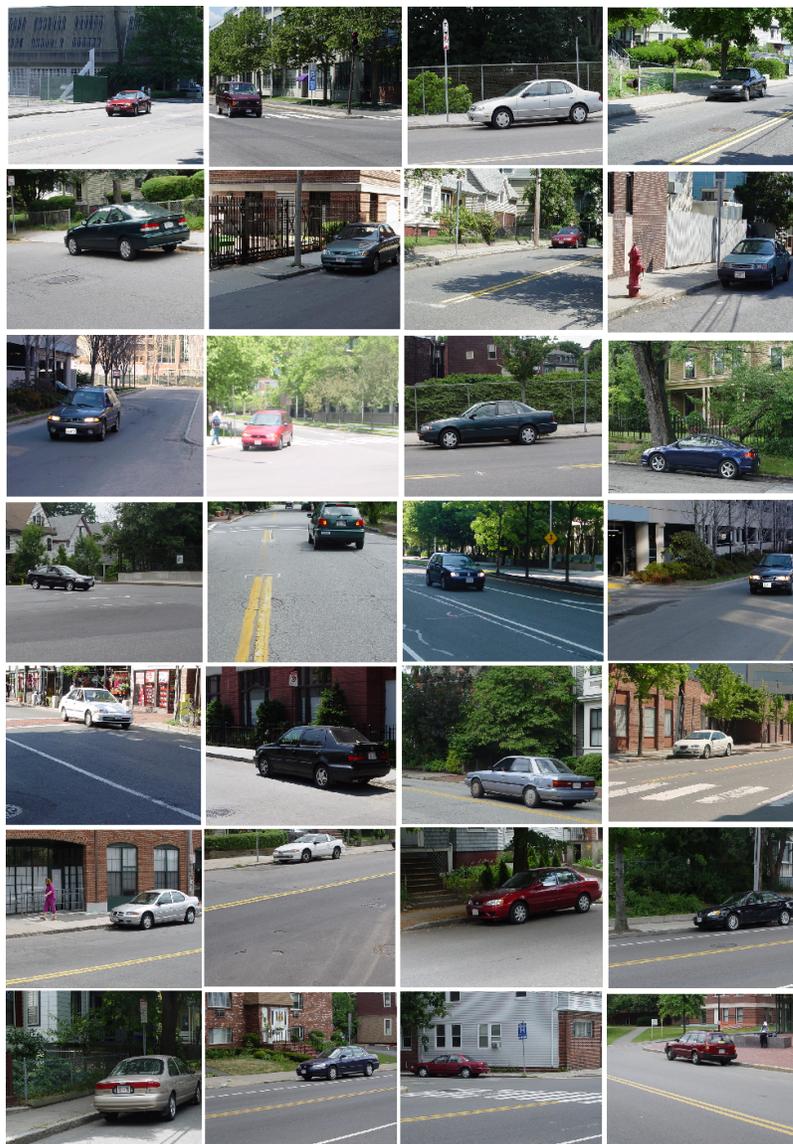


Figura 43: Imágenes pertenecientes a la clase carro con las que se entrenó el programa genético

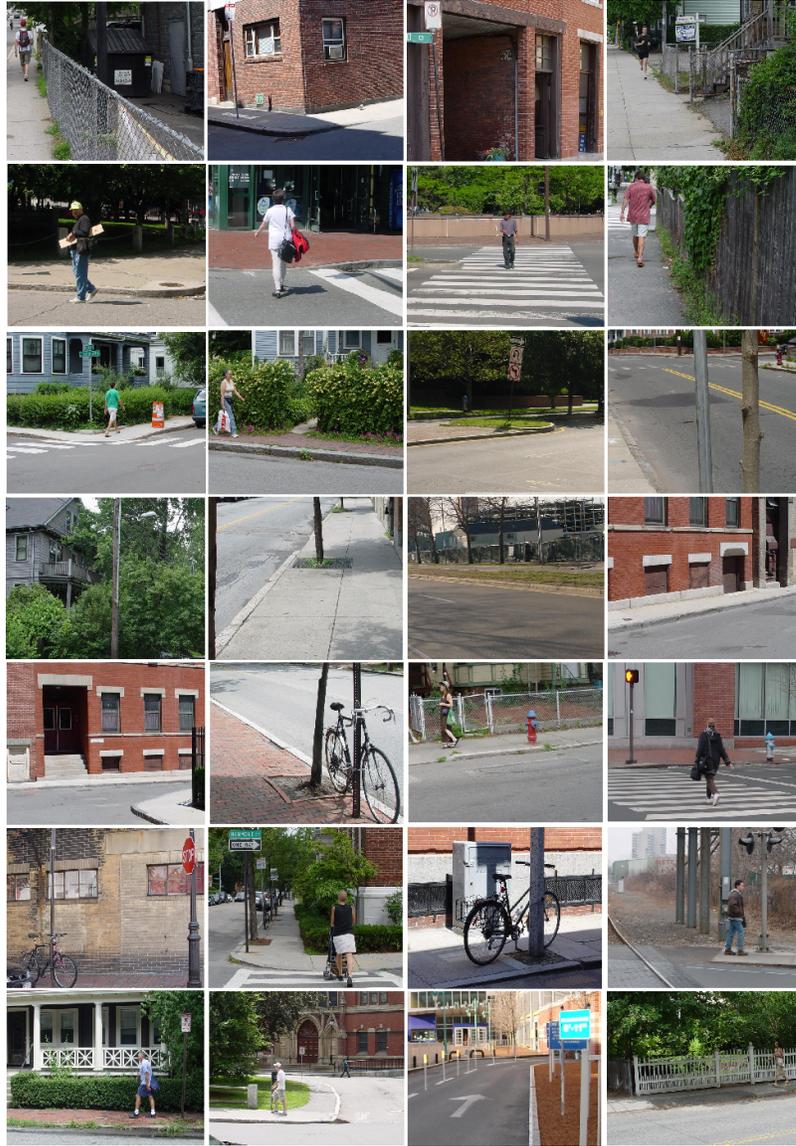


Figura 44: Imágenes pertenecientes a la clase distractores con las que se entrenó el programa genético

Capítulo VII

Conclusiones y trabajo futuro

VII.1 Conclusiones

En este capítulo se presentan las conclusiones del trabajo realizado. Estas se presentan en dos partes, aquellas que se relacionan con el problema de reconocimiento de objetos y otra donde se muestra el trabajo a futuro.

VII.1.1 Conclusiones del reconocimiento de objetos

El problema de reconocimiento de objetos es muy amplio. Desafortunadamente, aún cuando existen diferentes métodos que muestran buenos resultados, no existe un método que pueda manejar el problema de reconocimiento de objetos en el mundo real. Toda la investigación realizada permite solo un número limitado de clases de objetos debido a la poca disponibilidad de los conjuntos de datos de entrenamiento los cuales consumen mucho tiempo de recolección. Además los modelos tienen que ser fáciles de manipular a fin de hacer inferencias de manera más directa.

VII.1.2 Conclusiones del sistema implementado

Con la realización de este trabajo hemos llegado a diversas conclusiones. Una de las principales es el hecho de ver que si es posible evolucionar un modelo de atención visual para enfocarse en la búsqueda de una determinada clase de objetos. De la misma forma la utilización de un descriptor basado en el histograma de orientaciones es bastante

util en la extracción de rasgos en los objetos, en parte por permanecer robusto ante cambios de orientación y luminocidad. La conjunción de estos dos procesos en un programa genético nos dio la oportunidad de evolucionar dos rubros importantes en el reconocimiento de objetos al mismo tiempo, permitiendo obtener buenos resultados.

Se mostró que los resultados son mejores cuando se trata de objetos que tienen la misma forma o cuando se realiza la búsqueda del mismo objeto. Sin embargo, al buscar diferentes objetos los resultados obtenidos siguen siendo bastante buenos en comparativa con respecto a modelos actuales. Uno de los problemas que se presentaron fueron los conocidos como falsos positivos, como medida de erradicación se pretende ampliar la función objetivo para que tome en cuenta si la localización está o no dentro del área del protoobjeto.

VII.1.3 Trabajo a futuro

El trabajo realizado abre pautas para seguir avanzando hacia varias líneas de trabajo. Uno de los problemas que se ha planteado es utilizar este modelo pero en fragmentos de video donde se pueda llevar a cabo el seguimiento hacia objetos de la misma clase, en primera instancia se pensaría que el programa no cambiará mucho al tener ya los parámetros optimizados para buscar cierto tipo de clases de objetos, pero se deben tener en cuenta otras consideraciones básicas como una característica primaria de movimiento que permita seguir al objeto si este se está trasladando. Otra consideración es el tiempo de respuesta del sistema en video, por lo que ciertas partes del código podrían ser implementadas en lenguaje de programación CUDA (Compute Unified Device Architecture), que es una herramienta que permite codificar algoritmos en sus tarjetas NVidia de manera paralela a través de sus múltiples procesadores gráficos

(gpu's), lo que lograría reducir los tiempos de procesamiento del sistema.

Otra aplicación importante es llevar el proceso de la atención visual para el reconocimiento de imágenes pero de manera estereoscópica (cámaras capaces de capturar imágenes y video en tres dimensiones), donde a través de sus dos lentes podamos obtener imágenes 3D. En este aspecto se deben considerar características primarias como la profundidad. El proyecto se vuelve más ambicioso cuando se quiere hacer biológicamente plausible, donde habría que relacionar otros aspectos no considerados en este trabajo como la función que realiza el quiasma óptico, la forma de dividirse la información entrante en los NLG y principalmente emular una serie de complejos procesos fisiológicos en el cerebro los cuales reconstruyen la escena 3D.

Como se puede ver el tema de la atención visual tiene muchas aplicaciones ya que a final de cuentas en muchos problemas de visión por computadora al reducir el tamaño de procesamiento a solo la parte relevante presenta muchas ventajas; por lo que las aplicaciones en problemas de visión de alto nivel son muchas y variadas. Lo que se logra con este trabajo es dar la pauta para en posteriores tesis tomar en cuenta al modelo desarrollado como una excelente alternativa y que de manera particular, me servirá de base para una propuesta doctoral bajo la guía del Dr. Gustavo Olague.

Bibliografía

- and, R. R. 1997. "To see or not to see: the need for attention to perceive changes in scenes." *Journal of Experimental Psychology*. 5(8):368-373 p.
- Chikkerur, S. y Poggio, T. 2009. "An integrated model of visual attention using shape-based features". Technical report, Massachusetts Institute of Technology. Tech Report, CBCL-278, 16 p.
- Corbetta, M. y Shulman, G. 2002. "Control of goal-directed and stimulus-driven attention in the brain." *Nat Rev Neurosci*. (3).
- Desimone, R. y Duncan, J. 1995. "Neural mechanisms of selective visual attention." *Neurosci Journal*. (18).
- Egeth, H. y Yantis, S. 1997. "Visual attention: Control, representation, and time course." *Annual Review of Psychology*. (48).
- Elazary, L. y Itti, L. 2010. "A bayesian model for efficient visual search and recognition." *Vision Research*. 23(5):1338-1352 p.
- Friedman, S. y Wolfe, J. 1995. "Second-order parallel processing: Visual search for the odd item in a subset." *Journal of Experimental Psychology: Human Perception and Performance*. 3(21).
- Hernandez, B. y Olague, G. 2007. "Visual learning of texture descriptors for facial expression recognition in thermal imagery." *Computer Vision and Image Understanding*. 2(106).
- Holland, J. 1975. "Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence". University of Michigan Press. First edición. Michigan, USA. 183 p.

- Itti, L. 2003. "Modeling primate visual attention." Computational Neuroscience: A Comprehensive Approach.
- Itti, L. y Koch, C. 1998. "A model of saliency-based visual attention for rapid scene analysis." IEEE Transactions on Pattern Analysis and Machine Intelligence. 20(11):1254-1259 p.
- Itti, L. y Koch, C. 1999. "Hierarchical models of object recognition in cortex." Nature Neuroscience. (24).
- Joseph, R. 1990. "Neuropsychology, neuropsychiatry, and behavioral neurology". Springer. First edición. USA. 412 p.
- Koch, C. y Ullman, S. 1985. "Shifts in selective visual attention: Towards the underlying neural circuitry." Human Neurobiology. (4).
- Koza, J. 1992. "Genetic programming: On the programming of computers by means of natural selection". The MIT Press. First edición. Boston, USA. 819 p.
- Lowe, D. 2004. "Distinctive image features from scale-invariant keypoints". En: "International Journal of Computer Vision". 91-110 p. 60, 2 (2004).
- McConkie, G. y Currie, C. 1996. "Visual stability across saccades while viewing complex pictures." Journal of Experimental Psychology. 3(22):369-378 p.
- Milanese, R. y Bost, J. M. 1992. "A bottom-up attention system for active vision". En: "10th European Conference on Artificial Intelligence". 808-810 p. Septiembre 7 - Septiembre 10, Berlín, Alemania.
- Milanese, R. y Gil, S. 1995. "Attentive mechanisms for dynamic and static scene analysis". Optical Engineering Journal. (34).
- Navalpakkam, V. y Itti, L. 2006. "An integrated model of top-down and bottom-up attention for optimal object detection". En: "IEEE Conference on Computer Vision and Pattern Recognition". 2049-2056 p. Junio 2006.

- Neisser, U. 1967. "Numerical recipes in c". Appleton Century-Crofts. First edición. New York, USA. 351 p.
- nuela Capilar, F. V. 2004. "Neurorradiología diagnostica y terapéutica". Masson. First edición. Barcelona, España. 351 p.
- Ojeda, J. L. y Icardo, J. M. 2004. "Neuroanatomía humana". Masson. First edición. España. 324 p.
- Olague, G. y Romero, E. 2006. "Multiclass object recognition based on texture linear genetic programming.". En: "9 th European Workshop on Evolutionary Computation in Image Analysis and Signal Processing". 291-300 p. Septiembre 7 - Septiembre 10, Berlín, Alemania.
- Osberger, W. y Maeder, J. 1998. "An automatic image quality assesment technique incorporating higher level perceptual factors.". En: "Proceedings International Conference on Image Processing". 414-418 p. no. 3.
- Pereira, E. y Gomes, H. 2006. "Guiding a bottom-up visual attention mechanism to locate specific image regions using a distributed genetic optimization .". IEEE Transactions on Pattern Analysis and Machine Intelligence. 4225(11):257-266 p.
- Perrett, D. y Mistlin, A. 2003. "Visual neurones responsive to faces.". Trends in Neurosciences. 9(10):358-364 p.
- Rusell, S. y Norvig, P. 2002. "Artificial intelligence: A modern approach". Prentice Hall, Inc. Second edición. California, USA. 1132 p.
- Rutishauser, U. y Walther, D. 2004. "On the usefulness of attention for object recognition". En: "Second Workshop on Attention and Performance in Computational Vision". 96-103 p. the European Computer Vision Conference (ECCV04).
- Sánchez, C. C. 2005. "Modelo conexionista neuromimético para la percepción". *Tesis de Maestría, Université Henri Poincaré, Nancy, Francia.*

- Stentiford, F. 2001. "An evolutionary programming approach to the simulation of visual attention.". En: "Proceedings International Conference on Image Processing". 114-118 p. no. 3.
- Suder, K. y Worgother, F. 2000. "The control of low-level information flow in the visual system". *Neurosci Journal*. (11).
- Tackett, W. 1994. "Recombination, selection, and the genetic construction of computer programs.". Tesis de Maestría, University of Southern California, California, USA.
- Treisman, A. y Gelade, G. 1980. "A feature integration theory of attention.". *Cognitive Psychology*. (12).
- Vapnik, V. 1995. "The nature of statistical learning theory". Springer-Verlag. First edición. New York, USA. 188 p.
- Vapnik, V. 1997. "Neurobiología de la visión". UPC.
- Viviani, P. y Mounoud, P. 1990. "Perceptuomotor compatibility in pursuit tracking of two-dimensional movements.". *J. Mot. Behav.* (22).
- Wolfe, J. 1994. "Guided search 2.0: A revised model of visual search.". *Psychonomic Bulletin Review*. 2(1).
- Yarbus, A. 1967. "Eye movements and vision". Plenum Press. First edición. New York, USA. 222 p.
- Ye, Y. y Tsotsos, J. 2001. "A complexity level analysis of the sensor planning task for object search.". *Computational Intelligence*. 17(4).