Instituto Nacional de Astrofísica Óptica y Electrónica

Un Enfoque Bayesiano para un Modelo Biológico del Sistema Visual

Por

Elías Ruiz Hernández

Tesis sometida como requisito parcial para obtener el grado de Maestro en Ciencias en la especialidad de Ciencias Computacionales en el Instituto Nacional de Astrofísica, Óptica y Electrónica.

Supervisada por:

Dr. Luis Enrique Sucar Succar

Investigador Titular del INAOE

Resumen

En esta tesis se expone un enfoque bayesiano de un modelo del sistema visual. Este modelo está biológicamente inspirado en el sistema visual de los monos. Estos trabajos parten de las investigaciones de Hubel y Wiesel [2], en donde se logran definir un conjunto de capas de neuronas que presentan sensibilidad a cierta orientación de la luz. Un esfuerzo por comprender mejor este modelo y llevarlo a un esquema computacional es el trabajo propuesto por Serre y Poggio [13], el cual expone el modelo del sistema visual, así como una versión computacional simplificada del mismo. Se trata de un esquema por capas en donde se emulan capas de células simples y complejas. El enfoque bayesiano presentado en esta tesis parte de este modelo a fin de lograr un modelo alternativo, que emplee estructuras como el clasificador bayesiano simple. El objetivo es comprender de una mejor manera el funcionamiento del sistema visual, desde un enfoque computacional y emplear este modelo en tareas de categorización de clases de objetos. Se presenta una etapa de entrenamiento y otra de prueba, empleando imágenes de diversas categorías de objetos.

Realizar un modelo del sistema visual basado en un enfoque bayesiano permite modelarlo con las características de una estructura formal, flexible a fin de ser modificada, y con la posibilidad de agregar información previa a fin de mejorar el reconocimiento para ciertos objetos. En esta tesis se detalla el modelo original así como el enfoque bayesiano aplicado, mostrando su evaluación experimental, con resultados similares en cuanto a su porcentaje de reconocimiento de diferentes objetos, sin embargo el modelo bayesiano propuesto es más eficiente en términos de costo computacional y flexible en cuanto a su estructura, también es más fácil de comprender y por lo tanto de modificar y extender.

Abstract

In this thesis a bayesian approach of a visual system model is presented. This model is biologically inspired in the macaque visual system. Research about the model is based in Hubel & Wiesel paper [2] where a neuron layers sensitive to certain light orientation are defined. An effort to comprehend this model and take it to a quantitative model is the work presented by Serre & Poggio [13]. Inside this work a visual system model and a simplified version of the model are exposed. They explain a layer schema where each simple cell layer and complex cell layers are emulated. Our bayesian approach is based on this model in order to create an alternative model using structures like naïve Bayes. Our goal is to understand the visual system and employ this model in categorization of objects. Our model uses training and test phases, using images from several object categories.

Constructing a model of a visual system based on a Bayesian approach allows creating with characteristic of formal structure, flexible to modifications, and possibilities to add previous information in order to improve the recognition of certain objects. In this thesis the original model and bayesian approach applied is detailed, explaining an experimental evaluation with similar results for object recognition in images, but our bayesian model proposed is more efficient in computational time of process and flexibility related to its structure, as a result is more easy to understand, modify and extend.

Agradecimientos

A mi asesor, Dr Luis E. Sucar

Por el apoyo brindado para el desarrollo de esta tesis.

A mis sinodales, Leopoldo, Miguel, Eduardo

Por evaluar y apoyar este trabajo de Tesis.

Al Consejo Nacional de Ciencia y Tecnología

Por apoyar el desarrollo de la ciencia en nuestro país.

A mis padres

Por sus sabios consejos, y por su apoyo incondicional en todas las decisiones que he tomado en mi vida.

A mi hermana Silvia

Por enseñarme desde pequeño que aprender es muy divertido.

A mis amigos

Por el gran apoyo mostrado en diversas facetas de mi vida.

A todos mis maestros

Por haber fortalecido la planta del saber que cada uno de nosotros lleva dentro.

Contenido

1 Introducción 1

1.1 Motivación 1

1.2 Objetivos de la tesis. 2

1.3 Alcances del Modelo. 3

1.4 Aportaciones. 4

1.5 Descripción del Modelo. 4

1.6 Organización de la Tesis 6

2 Modelos Bioinspirados del Sistema Visual 7

2.1 Modelo biológico de Serre y Poggio. 10

2.1.1 Capa S1 12

2.1.2 Capa C1 15

2.1.3 Capa S2 17

2.1.4 Creación de prototipos 18

2.1.5 Capa C2 19

2.1.6 Capas Adicionales 19

2.1.7 Clasificación 23

2.2 Otros modelos empleados en el reconocimiento de objetos. 23

2.3 Conclusiones. 25

3 Modelos Bayesianos 26

3.1 Clasificador Bayesiano simple 26

3.2 Redes Bayesianas. 32

3.3 Resumen 33

4 Modelo Probabilista del Sistema Visual 35

4.1 Esquema General 35

4.2 Modelado por capas 36

4.2.1 Capa S1: Banco de Filtros Gabor 36

4.2.2 Capa C1 38
4.2.3 Capa S2 44
4.2.4 Capa C2 46
4.2.5 Clasificación 47
4.3 Resumen. 49
5 Experimentos y Resultados 52
5.1 Metodología de Pruebas. 52
5.2 Experimentos Realizados 54
5.3 Resultados 54
5.3.1 Comparación con el modelo original. 54
5.3.2 Invarianza a rotación 56
5.3.3 Invarianza a escala. 57
5.4 Tiempos de Entrenamiento y prueba 59
5.5 Análisis 59
6 Conclusiones y Trabajo Futuro 61
6.1 Resumen. 61
6.2 Conclusiones del modelo probabilista del sistema visual 61
6.3 Trabajo Futuro 62

Índice de Figuras.

Fig. 1.1 Diagrama del modelo general en su fase de entrenamiento. 5
Fig. 1.2 Diagrama de la fase de prueba. 5
Fig. 2.1 Modelo general HMAX. 8
Fig. 2.2 Estructura general del *Standard Model*. 12
Fig. 2.3 Filtro Gabor con una orientación de 0 grados. 13
Fig. 2.4 Aplicación de Banco de filtros gabor 14
Fig. 2.5 Función max sobre escalas 15
Fig. 2.6 Aplicación de Max Local. 16
Fig. 2.7 Aplicación de Submuestreo. 16
Fig. 2.8 Valor de similitud y matriz de estímulos 18
Fig. 2.9 Valores de similitud entre prototipos de etapas anteriores 20
Fig. 2.10 Ejemplo de imagen usada para S4 21
Fig. 2.11 Fase de Entrenamiento del Modelo General 21
Fig. 2.12 Fase de Prueba del Modelo General 22
Fig. 2.13 Entrenamiento y prueba del clasificador. 23
Fig. 3.1 Estructura de estrella de un CBS. 27
Fig. 3.2 Ejemplo de un CBS 28
Fig. 3.3 CBS emulando una función max 31
Fig. 4.1 Esquema general del modelo probabilista propuesto 36
Fig. 4.2 Cuantizacion de los niveles de gris. 37
Fig. 4.3 histograma después de aplicar el filtro Gabor 38
Fig. 4.4 Max normal comparado con un CBS. 39
Fig. 4.5 El modelo de un CBS. 40
Fig. 4.6 CBS para vecindades de 9 nodos. 40
Fig. 4.7 grafica exponencial para la tabla del CBS. 41
Fig. 4.8 Grafica exponencial decreciente correspondiente a la clase min 42

Fig. 4.9 Resultado de aplicar el filtro max (3x3) del modelo original. 43
Fig. 4.10 CBS para la fase del submuestreo. 43
Fig. 4.11 Esquema de los filtros aplicados en la capa C1.....44
Fig. 4.12 curva gausiana de probabilidad 46
Fig. 4.13 Prototipo aprendido por un CBS 47
Fig. 4.14 Diagrama de C2. 48
Fig. 4.15 Entrenamiento y prueba en la etapa del Clasificador.....49
Fig. 4.16 Resumen general de entrenamiento del modelo bayesiano. 49
Fig. 4.17 Resumen general de la fase de prueba del modelo bayesiano. 50
Fig. 4.18 Esquema de la fase de clasificación del modelo bayesiano. 51
Fig. 5.1 Biblioteca Caltech de 101 categorías de objetos 53
Fig. 5.2 Biblioteca background.....53
Fig. 5.3 Curva Roc para un conjunto de imágenes de carros. 57

Índice de Tablas

Tabla 2.1 Banco de filtros Gabor agrupado en 8 bandas. 14

Tabla 2.2 Tamaños de ventana para aplicar max local. 16

Tabla 2.3 Cantidad de Prototipos usados.....18

Tabla 2.4 Mínimo entre bandas y mínimo local.. 19

Tabla 3.1 OR lógico tradicional y OR a partir de un CBS. 28

Tabla 4.1 Tabla de valores de probabilidad aplicados en cada nodo. 42

Tabla 4.2 Tabla del píxel a considerar para el sub-muestreo. 44

Tabla 4.3 Tabla de probabilidad para un nodo.. 46

Tabla 5.1 Resultados entre el modelo Bayesiano y el Modelo de Poggio. 55

Tabla 5.2 Resultados para imágenes en diversas orientaciones.....58

Tabla 5.3 Resultados para imágenes de rostros en diversas escalas.....58

1 Introducción

1.1 Motivación

En los últimos años se han realizado investigaciones relacionadas con la visión tanto desde enfoques biológicos como enfoques computacionales. En el lado del enfoque biológico se han hecho esfuerzos por comprender mejor el funcionamiento de la corteza visual a fin de conocer como se pueden visualizar los objetos del mundo real, así como su reconocimiento. En el lado del enfoque computacional, se ha trabajado en reconocer objetos presentes en una imagen a partir de la extracción de características de la imagen, la segmentación de regiones y la aplicación de algoritmos que permitan inferir la existencia de un objeto a partir de tal imagen, entre otras muchas áreas dentro del análisis de imágenes.

Recientemente se han hecho esfuerzos por conjuntar lo anterior, así que existe investigación sobre el funcionamiento de la corteza visual en los primates [14], la forma en como los primates reconocen objetos a partir de su sistema de visión. Se ha encontrado las neuronas en los primates funcionan a través de capas de neuronas que transmiten información y cada una de esas capas realiza una actividad específica en el sistema de visión para reconocer ciertas características de la imagen recibida. Estos trabajos se remontan a las investigaciones de Hubel y Weisel [2], en los cuales se describe esta diferenciación celular. Adicionalmente se han hecho investigaciones de la aplicación de modelos biológicamente inspirados para el reconocimiento de objetos. Un trabajo relacionado con ello es el modelo biológicamente inspirado de Serre y Poggio [13]. Este modelo se fundamenta en el estudio del sistema visual bajo el enfoque biológico proponiendo una estructura en capas análoga a lo expuesto por Hubel y Weisel donde se logra reconocer objetos en imágenes previo un entrenamiento. Este modelo consigue reconocer patrones cada vez más elaborados en cada capa hasta llegar a una etapa de categorización de objetos a partir de diversas imágenes. Una categoría de objetos es un conjunto de objetos que comparten características muy semejantes, por ejemplo los rostros, los cuales comparten cabello, ojos, nariz y boca, sin embargo, no por ello son iguales. De esta manera, el modelo de Poggio busca reconocer objetos a partir de una categorización mediante un modelo inspirado en el

sistema de la corteza visual de los primates. El principal objetivo es comprender mejor el sistema visual de los primates y los humanos desde un enfoque biológico tratando de aterrizarlo en un modelo computacional. En este proceso, de manera análoga al enfoque biológico, se pueden reconocer objetos mediante una categorización de imágenes a partir de un entrenamiento previo. Esta identificación de un objeto en una categoría no se corresponde con el reconocimiento de la ubicación de un objeto en el espacio de la imagen, por tanto, ubicar o seguir un objeto en el espacio queda fuera del alcance del modelo. Por el contrario, se busca determinar si en una imagen existe algún objeto que pertenezca a una categoría determinada. El modelo puede entrenarse para una categoría específica. No se considera tratar de entrenar el modelo para una cantidad múltiple de categorías. De manera alternativa, es posible tener diversos modelos entrenados, uno para cada categoría diferente, pero no contener, en un solo modelo entrenado, diversas categorías de objetos.

A partir de este modelo, tomando en cuenta que existen trabajos relacionados con el reconocimiento de objetos a partir de esquemas probabilistas ([9] y [11]) y observando que el modelo original propuesto por Serre y Poggio [13] describe el modelo desde un punto de vista biológico aterrizándolo en un modelo computacional, se tiene que este modelo no resulta tan flexible para extraer modificaciones y establecer un estudio más profundo de cada capa desde un enfoque computacional. Por tanto, se plantea construir un modelo alternativo al modelo bio-inspirado de Poggio, empleando un enfoque bayesiano. Se busca con ello tener un modelo más formal, logrando un mayor entendimiento del modelo bio-inspirado así como tener una estructura flexible que pueda modificarse de manera clara. Al ser más flexible también permite establecer una inferencia probabilista de arriba hacia abajo a fin de mejorar su desempeño en la categorización para ciertos objetos. Esta información previa representa una ventaja con respecto del modelo de Poggio.

1.2 Objetivos de la tesis

Construir bajo un enfoque bayesiano un modelo para el reconocimiento de objetos inspirado en el modelo biológico propuesto por Poggio. Se busca tener un modelo formal a partir de la teoría de Poggio, que tomando como referencia el enfoque

biológico, también permita identificar y categorizar objetos en un conjunto de imágenes previo entrenamiento.

Asimismo, se plantea aprovechar características que el modelo biológico de Poggio no posee. El modelo de Poggio no presenta una estructura flexible y su estructura resulta compleja si se desea realizar modificaciones. Con el modelo propuesto se tendrá una alternativa para resolver problemas de identificación y categorización de objetos dentro del área de visión computacional, así como ayudar a una mejor comprensión del modelo desde un punto de vista probabilista. También se busca analizar la capacidad del modelo en particular en cuanto a la invarianza a rotación y escala.

En este documento se estudia el modelo original y se presenta la propuesta del modelo alternativo, aplicando las variaciones requeridas para adaptarlo a un enfoque bayesiano.

Este nuevo enfoque es comparado con el modelo de Poggio a fin de observar los resultados en cuanto a la tarea de categorización de los objetos en las imágenes.

1.3 Alcances del Modelo

El modelo que se plantea en la tesis busca reconocer objetos en imágenes a partir de un entrenamiento con imágenes similares (no implica iguales) a aquellas con las que se desea probar el funcionamiento de la categorización correcta de los objetos. Se busca tener buenos resultados tanto en la correcta categorización del objeto en las imágenes que si lo contengan, como un resultado aceptable rechazando la existencia del objeto en imágenes que no lo posean. El modelo plantea la construcción de un enfoque bayesiano por capas. La estructura del modelo descrito en esta tesis describe un diseño jerárquico, que permitiría su reimplementación bajo el enfoque de una red bayesiana. Si bien es posible la realización de este tipo de modelo, se deja como trabajo futuro la posible implementación como un sistema monolítico de red bayesiana.

Este modelo trabaja con imágenes sin recibir información de color, por lo anterior, tanto las imágenes de entrenamiento como las imágenes de prueba parten de imágenes en escala de grises. Esto se fundamenta en los trabajos de Hubel y Weisel [2], Tsao [14] y Fukushima [4] en donde no consideran información de color tanto en el

aspecto biológico como en el computacional. De esta manera un análisis del color de las imágenes no ocupa lugar en el modelo planteado.

Por otro lado, el modelo no plantea la ubicación en el espacio del objeto encontrado en la imagen, por tanto una tarea de seguimiento de objetos queda fuera del alcance de este trabajo, puesto que estaría orientado no a indicar a que categoría pertenece un objeto, sino a ubicarlo en el espacio de una imagen y tratar de darle seguimiento a partir de una secuencia de imágenes que describa movimiento.

1.4 Aportaciones

En este trabajo se busca desarrollar un modelo bayesiano, el cual no había sido considerado para la elaboración de un modelo del sistema visual. Se busca construir un enfoque que, basado en el modelo propuesto tenga una mayor comprensión que el modelo del sistema visual, comparándolo con el modelo original, así como hacer un análisis de la invarianza del modelo a rotación y escala. Asimismo se crea un precedente de un enfoque bayesiano el cual queda abierto a realizar modificaciones y extensiones al mismo, a fin de mejorarlo.

1.5 Descripción del modelo

Visto desde el enfoque computacional, El modelo de Poggio [13] parte de la aplicación de una serie de filtros evaluados por capas, donde cada capa semeja un determinado tipo de neuronas del modelo biológico del sistema visual. Presenta 2 fases, una de entrenamiento y otra de prueba. Un diagrama de bloques general del modelo en la fase de entrenamiento se ilustra en la Fig. 1.1

Esencialmente se aplican 3 capas en la fase de entrenamiento únicamente. La capa S1 consiste en la aplicación de un banco de filtros Gabor en diferentes orientaciones y escalas. La Capa C1 en un conjunto de filtros max. Finalmente en la capa S2 se extraen patrones aleatorios de los conjuntos de imágenes generados por las dos capas anteriores.

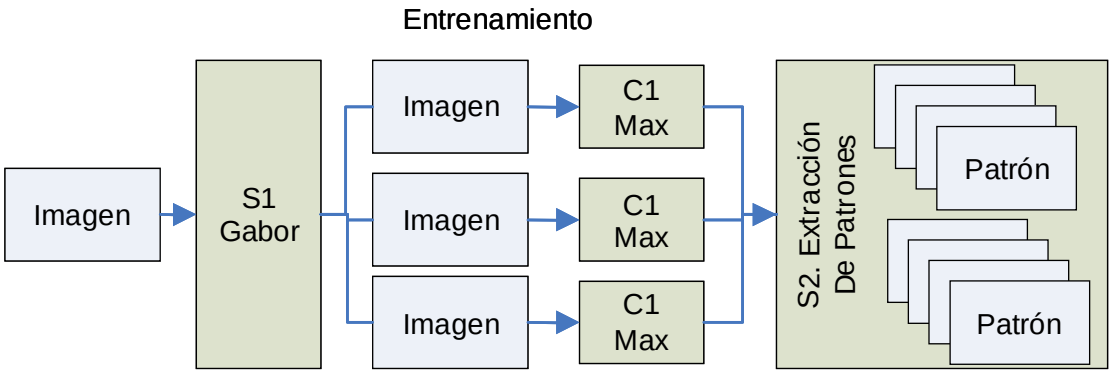


Fig. 1.1 Diagrama del modelo general en su fase de entrenamiento. Al final se obtienen patrones que serán evaluados en la fase de prueba.

En la fase de prueba se aplican 4 capas y una etapa de clasificación. Un diagrama de esta fase de prueba se ilustra en la Fig. 1.2

En esta fase de prueba se aplican las mismas capas S1 y C1 (bancos de filtros Gabor y filtros Max) en la capa S2 se evalúan los patrones obtenidos en las imágenes con la finalidad de obtener una métrica. Los resultados se someten a filtros Max y finalmente se clasifican los datos obtenidos.

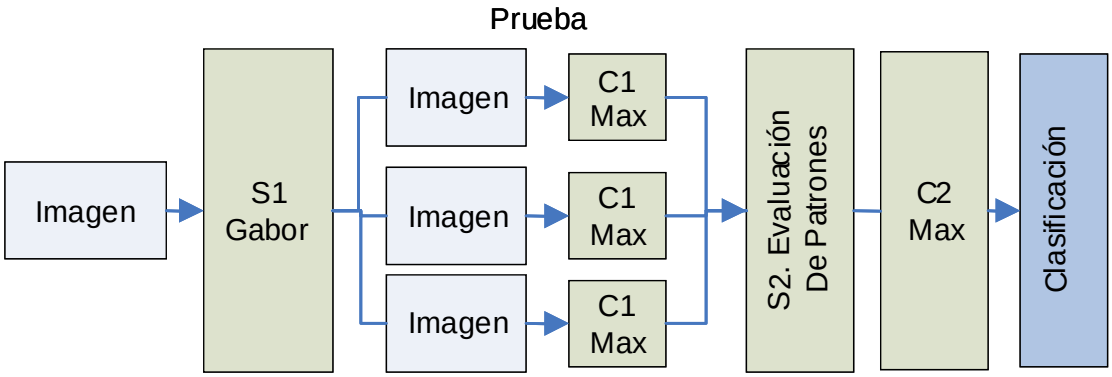


Fig. 1.2 Diagrama de la fase de prueba. La capa S2 ahora evalúa los patrones extraídos del conjunto de entrenamiento. Se agregan las capas C2 y una etapa de clasificación.

Este modelo se desarrolló empleando un enfoque bayesiano y para su realización se basó en el modelo visual de Poggio [13]. Así, este modelo se prueba con un conjunto de imágenes para probar su categorización. Se evalúa su funcionamiento en la tarea de reconocimiento de una categoría de objetos determinada (rostros, carros, etc) con un conjunto de entrenamiento previo. Se encontraron interesantes resultados para la categorización de un objeto, aunque algunos resultados más limitados para ejemplos

donde se involucraron imágenes escaladas y rotadas, tanto en el modelo original como en la propuesta bayesiana que se desarrolla en esta tesis.

1.6 Organización de la tesis

En el capítulo 2 se describen modelos biológicamente inspirados, así como algunos otros relacionados con el reconocimiento de objetos en imágenes. De ellos se detalla con mayor profundidad el modelo biológico de Poggio, su fundamento en el sistema visual, así como su estructura basada en capas. Este modelo a efectos de llevarlo a una implementación computacional ha sido simplificado, reduciendo el número de capas del modelo propuesto con la finalidad de evitar un excesivo costo computacional. Es precisamente el modelo simplificado el que se estudia a detalle.

Posteriormente en el capítulo 3 se explica de manera detallada el funcionamiento de los clasificadores bayesianos y la información que pueden aportar, así como estrategias para entrenarlos y una introducción a la inferencia probabilista. También se comentan algunos trabajos de visión computacional empleando enfoques probabilistas.

En el capítulo 4 se expone el modelo propuesto. El enfoque bayesiano, su estructura por capas, la elaboración de modelos bayesianos análogos al modelo propuesto por Poggio, que van desde ir describiendo la información del conjunto de imágenes tanto de entrenamiento como prueba, hasta la clasificación que permite categorizar si una imagen contiene el objeto perteneciente a aquella categoría para la cual se entrenó el modelo.

En el capítulo 5 se esquematizan los resultados obtenidos de aplicar ambos modelos a bases de datos de imágenes, entre ellas la base de datos Caltech de 101 categorías de imágenes [3]. Se analizan las variaciones de los resultados, empleando diversos parámetros de entrenamiento.

Finalmente, en el capítulo 6 se exponen las conclusiones del trabajo, el posible trabajo futuro así como las extensiones aplicables al modelo propuesto en la tesis.

2 Modelos Bioinspirados del Sistema Visual

El sistema visual ha sido estudiado desde el punto de vista fisiológico a fin de comprender mejor su funcionamiento. En este sentido se han realizado investigaciones en diversos animales, aunque hay un interés especial por el sistema de visión de mamíferos, en particular, los monos. Se cree que si se parte de estudiar este sistema de visión, se puede tener una mejor comprensión del sistema visual del ser humano, considerando la analogía existente de los monos y el hombre. En este sentido, hay trabajos para tratar de comprender el funcionamiento del sistema visual en los monos. Los trabajos de Hubel y Wiesel [2] desarrollaron una teoría del sistema visual a partir de estudiar el comportamiento fisiológico de las neuronas en los monos. Esta teoría sostiene principalmente que:

- Existen diversas activaciones en las neuronas a partir de estimular la retina de los monos con patrones de luz determinados.
- Hay una tendencia de las células de agruparse y presentar igual activación frente a un patrón de luz determinado.
- Las neuronas pueden clasificarse en 2 tipos: células simples y complejas, de acuerdo a presentar una selectividad a la orientación de los patrones de luz, o bien, selectividad a la invarianza del patrón presentado.

Otros trabajos como el de Doris Tsao [14], tratan de comprender el reconocimiento de rostros y objetos dentro de la corteza visual en los macacos, donde sugieren que tanto humanos como macacos comparten una arquitectura similar para el procesamiento de objetos en la corteza visual.

En el trabajo de Riesenhuber y Poggio [10] se elabora un modelo bajo una concepción más cuantitativa del modelo del sistema visual propuesto por Hubbel y Wiesel de modo que se tenga una comprensión más clara desde un enfoque computacional en vez del enfoque fisiológico, aunque consistente con éste último. En este trabajo se toma como base un modelo jerárquico a partir del cual se concluye que las neuronas complejas se comportan como un filtro Max (estímulo máximo de un vecindario local en una imagen), presentando cierta invarianza a la oclusión y escala de

los objetos. En este mismo trabajo, también se compara el filtro sum (promedio de los valores de los píxeles) con un filtro Max (píxel de mayor valor) en donde éste último logra una respuesta más robusta al estímulo que el filtro sum al variar la escala de un objeto. En la elaboración de este enfoque se considera un “modelo de retina” de 160 píxeles. Este modelo propuesto se le denominó HMAX y aparece descrito en la Fig. 2.1

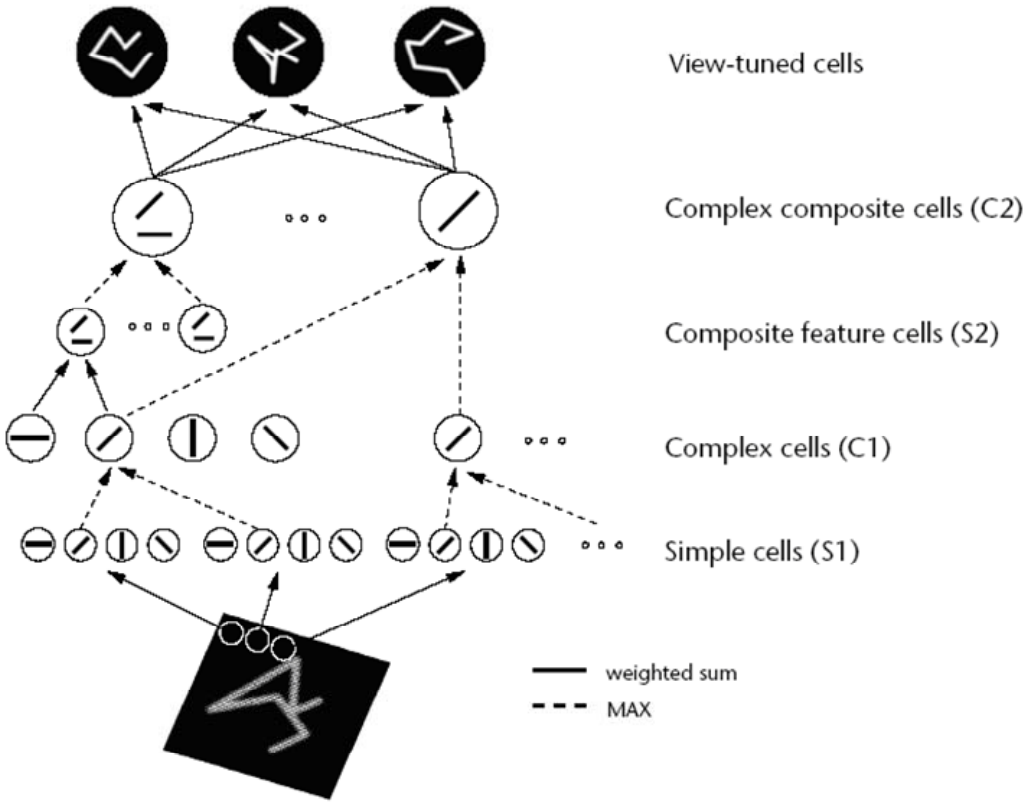


Fig. 2.1 La figura muestra el modelo general HMAX de células simples y complejas. El objetivo es ayudar a construir la invarianza en el modelo en cuanto a posición y escala (De Riesenhuber y Poggio [10]).

En el modelo HMAX cada capa consiste en la aplicación de diversos filtros, los cuales semejan las capas de neuronas simples y complejas. Esencialmente en el modelo HMAX se discute la aplicación de filtros Gabor en S1, la aplicación de un filtro Max en C1, una extracción de patrones en S2 y una capa análoga de filtros Max en C2. Este modelo sirve de base para la comprensión del sistema visual.

A partir de estos trabajos se realizaron esfuerzos por comprender este sistema de visión y tratar de realizar un enfoque computacional de manera que emule el sistema visual. Existen diversos trabajos que se acercan a tener un enfoque computacional biológicamente inspirado. Entre ellos está el propuesto por Fukushima [4], en el cual se plantea un modelo basado en redes neuronales para semejar el sistema de capas de

neuronas simples y complejas que pudieran reconocer números y letras en diferentes posiciones en una imagen monocromática. En este modelo, las capas se denominan como S para las capas de neuronas simples y C para las complejas. A partir de esta nomenclatura, otros trabajos las han nombrado de manera equivalente. El aprendizaje del modelo es no supervisado de modo que no requiere un experto en el dominio para ser entrenada la red. El trabajo de Fukushima parte de la aplicación de unos filtros de contraste y la extracción de características de la imagen que sean relevantes. Para características no relevantes se aplica una máscara a fin de bloquearlas.

En el trabajo de Wersing y Korner [15] se desarrolla un modelo jerárquico neuronal en donde muestran que tanto para las tareas de aprendizaje y reconocimiento de objetos no requieren de un proceso de segmentación previo de los objetos contenidos en las imágenes.

En el trabajo de Yann LeCun [6] se aplican métodos de aprendizaje a un modelo biológicamente inspirado. Estos métodos de aprendizaje incluyen clasificadores basados en vecinos más cercanos así como máquinas de soporte vectorial y redes de convolución. El objetivo es poder reconocer objetos naturales como animales, figuras humanas, carros y aviones. En su trabajo se concluye que se ha desarrollado un modelo que presenta invarianza a la posición y la intensidad de luz (contraste y brillo).

En el trabajo de Serre y Poggio [13] se construye una teoría biológicamente inspirada en el sistema visual que permite reconocer objetos en imágenes. El objetivo de esta teoría es comprender de manera más profunda el sistema visual, apoyándose en el enfoque cuantitativo propuesto por Riesenhuber [10] para construir una teoría construida sobre el clásico modelo de células simples y complejas propuesto por Hubel y Weisel. De esta manera, el modelo de Serre y Poggio sostiene los siguientes puntos:

- Construir una arquitectura jerárquica que resulte invariante a transformaciones de los objetos (como posición y escala) y que extraiga características específicas inspiradas en el modelo biológico.
- Describir bajo un enfoque computacional las funciones principales de los dos tipos de células (simples y complejas) representadas bajo un esquema de dos operaciones análogas a la selectividad a la orientación (células simples) y a la invarianza (células complejas).

Este modelo ha presentado resultados interesantes en la categorización de imágenes previo entrenamiento. De este modo, el modelo además de presentar una mejor comprensión del sistema visual de los macacos, ha funcionado para indicar si en una imagen existe un objeto de una clase para la cual ha sido entrenado. Asimismo, presenta tolerancia a los cambios de posición y escala del objeto, así como cierta tolerancia a la oclusión parcial del objeto. De manera más limitada, presenta cierta tolerancia a la rotación del objeto. Resulta interesante resaltar, que este modelo no parte de una segmentación previa de las imágenes con las que se entrena o con las que se prueba, además de no ser supervisada su fase de entrenamiento. Es posible entrenarlo con un conjunto de imágenes tanto naturales como sintéticas.

En el lado de las limitaciones del modelo, existe cierta tendencia a la generación de falsos positivos en la categorización de las imágenes (predecir si está contenido el objeto en una imagen). También a pesar de representar un gran esfuerzo en el sentido de comprender mejor el sistema visual, presenta un aspecto poco formal así como una estructura que no es flexible en el caso de que se desee hacer modificaciones en alguna de sus capas.

De esta manera en el presente capítulo, se describe el modelo propuesto por Serre y Poggio a fin de lograr una mejor comprensión de su enfoque computacional. A partir de este, en capítulos posteriores, se desarrolla un modelo formal, basado en un enfoque bayesiano, cuya finalidad es presentar una estructura más comprensible y flexible de realizar modificaciones en cada una de sus capas, así como poder incorporar conocimiento previo al modelo; lo cual es una característica que el modelo de Serre y Poggio no posee.

2.1 Modelo biológico de Serre y Poggio

El modelo propuesto por T. Serre y T. Poggio también denominado “Standard Model” [13] establece capas inspiradas en el sistema biológico visual. Considera únicamente los primeros 150 milisegundos del flujo de información en el sistema visual de los macacos. En este periodo de tiempo, ocurre la tarea de reconocimiento y categorización en el sistema visual. Partiendo del hecho que mientras en el sistema biológico se hace una diferenciación entre neuronas simples y complejas, las cuales se

alternan hasta llegar a una etapa de categorización de la imagen recibida en la retina, el modelo propuesto tiene dos tipos de capas (abreviándose capa S y capa C) que se inspiran en estos dos tipos de neuronas. De manera análoga al modelo biológico, estas capas se van alternando conformando una red jerárquica hasta lograr cierta invarianza a la posición y escala. Por el lado de las neuronas simples, éstas son sensibles a una cierta orientación de la luz que incide sobre una imagen. De esta manera son selectivas a un tipo de estímulo (orientación de la luz) con un mismo vecindario local del espacio visual, es decir, el vecindario de píxeles de una región de la imagen. En el caso de las neuronas complejas, éstas son sensibles al estímulo más fuerte obtenido dentro de una vecindad local de un campo visual, es decir, una pequeña región de un área observada. Al ser selectivas al impulso más fuerte de una vecindad local, presentan cierta tolerancia a pequeños cambios de posición y escala, por tanto se incrementa la invarianza a la posición y escala al observar objetos semejantes pero de diferente tamaño y/o en diferente posición, aunque esto no incluye una rotación considerable del objeto, por ejemplo, una rotación de 90 grados. El enfoque del modelo de Poggio, al estar biológicamente inspirado en el sistema visual, no contempla la invarianza a la rotación. Partiendo de este esquema principal, es posible continuar añadiendo capas alternadas de neuronas S y C con las propiedades antes descritas. Se considera que una mayor cantidad de capas podría traer mejores resultados en cuanto a la categorización correcta de una clase de objetos en imágenes, considerando el entrenamiento previo.

Esta categorización permite identificar objetos como: rostros, autos, hojas, aviones, etc. De esta manera el modelo bio-inspirado de Poggio busca identificar objetos en imágenes previo un entrenamiento con una categoría de tales objetos. Este entrenamiento es análogo al entrenamiento que reciben los primates cuando observan un objeto por primera vez. Posteriormente su sistema visual aprende a categorizar e identificar tales objetos.

Se parte de imágenes en escala de grises, de modo que la información del color no se considera. Las imágenes son agrupadas de manera que permitan trabajar primero con las imágenes de entrenamiento y posteriormente con las imágenes de prueba.

El modelo básico (*Standard Model*) contiene las capas S1, C1, S2, C2; aunque en Poggio y Serre definen más capas adicionales análogas a S2 y C2 (S3, C3, S4) con el objetivo de presentar una estructura más robusta, es decir, un modelo extendido al modelo básico. El modelo se ilustra en la Fig. 2.2 .

El modelo se resume en establecer una categorización, a partir de la extracción de ciertas características en las imágenes basándose en el modelo del sistema visual, para posteriormente encontrar objetos en imágenes de prueba que coincidan con la categoría de objetos en las imágenes de entrenamiento. En el resto del capítulo se describen más a detalle cada una de las capas del modelo.

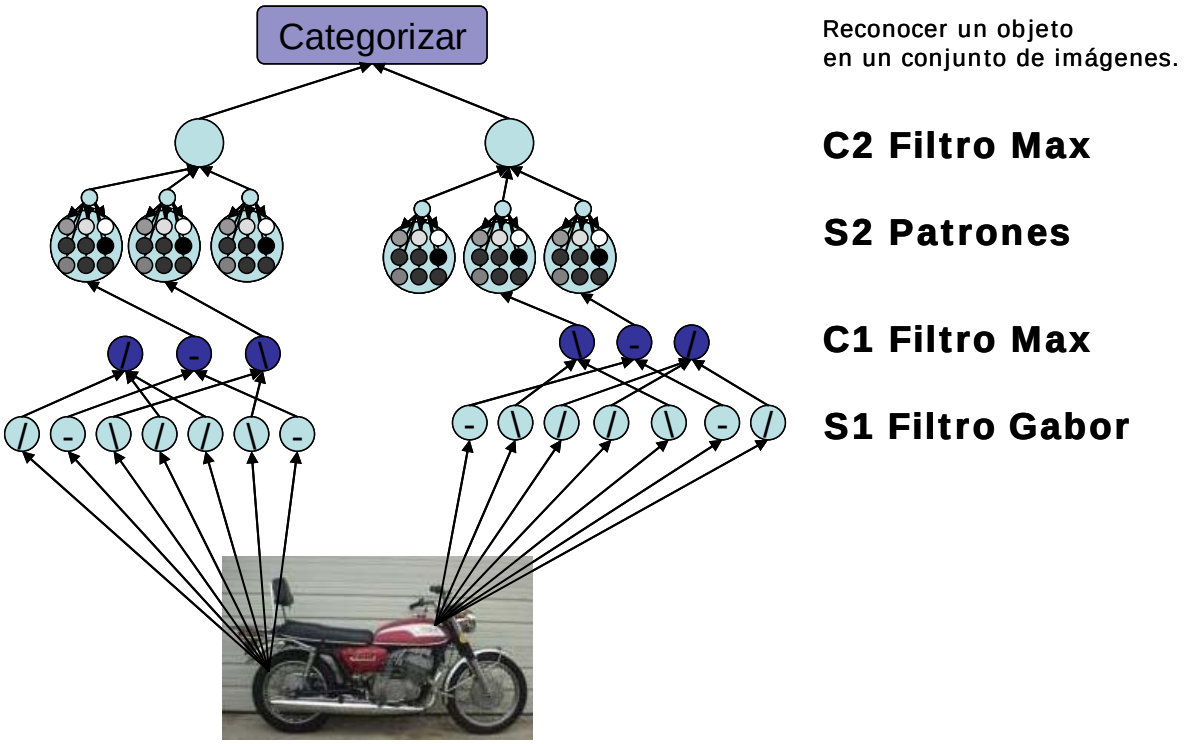


Fig. 2.2 Estructura general del *Standard Model*. Tiene 4 capas, dos simples y dos complejas y una capa de clasificación en la parte superior.

2.1.1 Capa S1

Esta capa contiene un banco de filtros Gabor que si bien aparece en [13], fue propuesto por Jones y Palmer [5]. De acuerdo al modelo, las neuronas que corresponden con S1 son sensibles a distintas orientaciones de un objeto en una imagen, es decir, no ocurre la misma excitación al ver una imagen de una barra a 0º que verla a 90º, diferentes neuronas se activan en diferentes orientaciones.

El filtro Gabor por su parte, representa una alternativa para extraer características de bordes de una imagen, en donde uno de los parámetros es la orientación. De modo que es posible asignar mayor importancia a una orientación determinada (0º o 90º por ejemplo). Con ello se extrae información relevante sobre los bordes orientados en un ángulo especificado.

El filtro Gabor es un tipo de filtro Gaussiano que está descrito por la ecuación:

$$F(u_1, u_2) = e^{\left(-\frac{v_1^2+\gamma^2 v_2^2}{2\sigma^2}\right)} \cos\left(\frac{2\pi}{\lambda} v_1\right) \quad \textbf{(2.1)}$$

Donde:

$$v_1 = u_1 \cos\theta + u_2 sen\theta \quad \textbf{(2.2)}$$

$$v_2 = -u_1 sen\theta + u_2 \cos\theta \quad \textbf{(2.3)}$$

(u_1, u_2) son los valores del filtro en el rango [-η,η] donde η es el tamaño de ventana del filtro) en un sistema coordenado, θ es la orientación, γ un coeficiente constante de aspecto, σ es un ancho efectivo del filtro dependiente de la escala, λ una longitud de onda dependiente de la escala. La escala del filtro Gabor define el tamaño de filtro (η) y con ello define a σ y λ. En la Fig. 2.3 se muestra la aplicación de un filtro Gabor de orientación 0º. Como resultado, sólo se aprecian los cambios muy cercanos a cero grados, otros bordes se pierden.

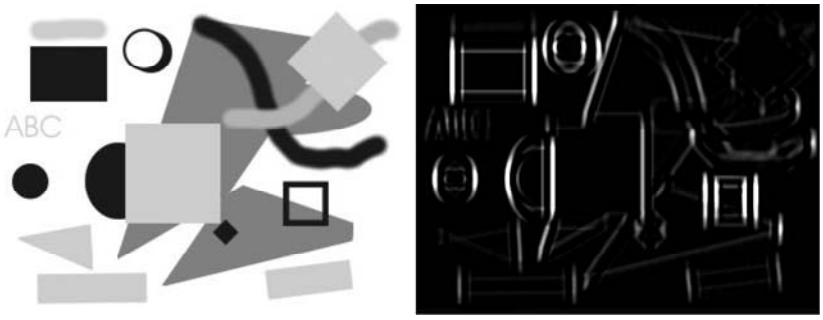


Fig. 2.3 Izquierda: la imagen original. Derecha: la imagen después de aplicar el filtro Gabor con una orientación de 0 grados.

Considerando que la selectividad a la orientación de las neuronas ocurre a distintas orientaciones, el filtro Gabor puede aplicarse en diferente orientación y escala, dando como resultado un conjunto de imágenes por cada imagen evaluada. En la Fig. 2.4 se aplican 3 orientaciones y 2 escalas.

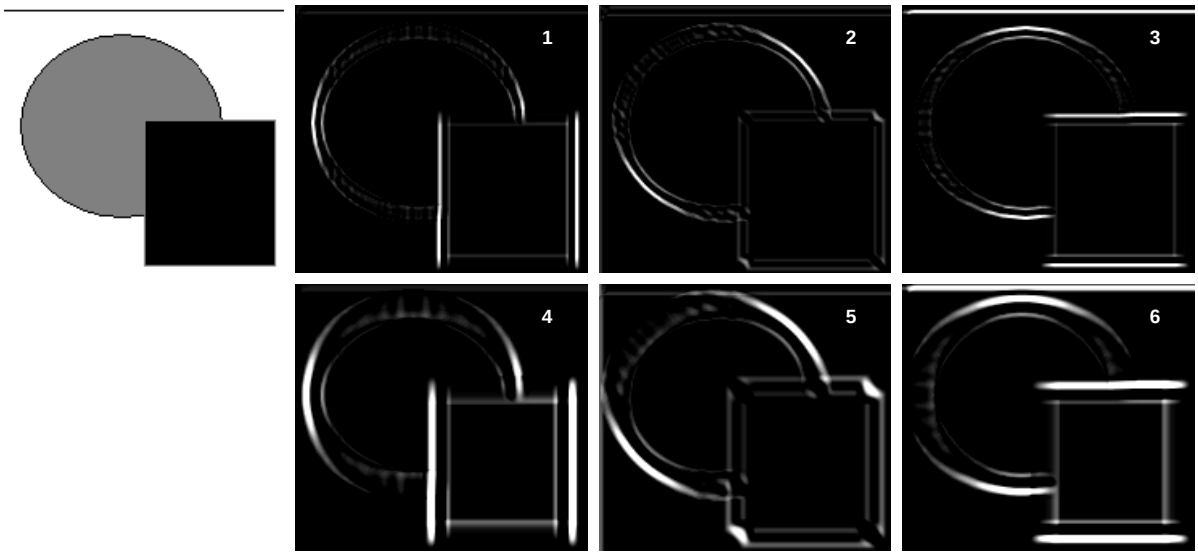


Fig. 2.4 Imagen Izquierda: Imagen original. Derecha: aplicación de banco de filtros Gabor: La escala define el ancho del filtro Gabor. 1: 0º esc. 11. 2: 45º esc. 11. 3: 90º esc. 11. 4: 0º esc. 19. 5: 45º esc. 19. 6: 90º esc. 19.

En el modelo, esta capa se emplea para extraer un conjunto de imágenes a partir de un banco de filtros Gabor, aplicados en distintas orientaciones y en distintas escalas. Las orientaciones utilizadas en el modelo, son cuatro: 0, 45, 90 y 135 grados. Cada una de estas orientaciones emplea 17 escalas (desde 7x7 píxeles hasta 39x39, sólo impares) y se forman 8 grupos de estas imágenes, a cada grupo se le denomina banda. Así, cada banda tiene 8 imágenes. De la banda 1 a la 7, cada banda incluye 2 escalas en sus 4 orientaciones cada escala. En la última banda (banda 8) tiene 12 imágenes, empleando 3 escalas en sus 4 orientaciones. Por lo anterior se tiene un banco de 68 filtros Gabor que se aplican sobre cada imagen que es analizada. Lo anterior se puede resumir en la Tabla 2.1.

Finalmente este banco de filtros es aplicado mediante convolución con cada imagen ya sea de entrenamiento o de prueba.

Tabla 2.1 El banco de filtros Gabor se agrupan en 8 bandas con escalas distintas.

No. Banda	1	2	3	4	5	6	7	8
Escalas	7,9	11,13	15,17	19,21	23,25	27,29	31,33	35,37,39
Orientaciones	0º, 45º, 90º, 135º							

2.1.2 Capa C1

Como se ha comentado, en el enfoque biológico, la capa C1 representa al conjunto de neuronas complejas. Las neuronas complejas, reciben información del conjunto de neuronas simples, que en el modelo han sido representadas en la capa S1. Al inspirarse en el modelo biológico, el resultado de la construcción de la capa C1 fue la aplicación de ciertas operaciones basadas en filtros max que se detallan a continuación.

La capa C1 realiza 3 operaciones fundamentales con cada una de las 8 bandas generadas por imagen en la capa S1. Primeramente, realiza una operación max sobre pares de imágenes de una banda con la misma orientación (pero escala ligeramente distinta) para obtener los máximos del par de imágenes. En la octava banda se aplica el filtro Max sobre las 3 escalas de la imagen. El resultado se muestra en la Fig. 2.5

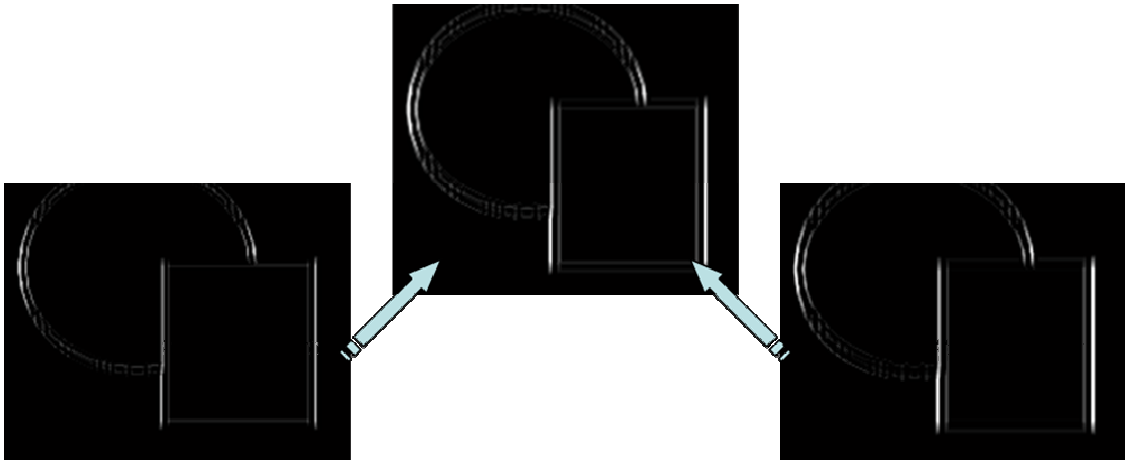


Fig. 2.5 En los extremos dos imágenes en escalas 7 y 9 y orientación 0º de la primera banda. En el centro, el resultado de aplicar la función max.

Posteriormente se extrae información de máximos por vecindad local, se aplican filtros Max en distintos tamaños de ventana en píxeles, es decir, el filtro barre la imagen en ventanas de k $\times$ k píxeles, en proporción a la banda empleada. Este filtro ayuda a presentar cierta invarianza a cambios de escala y en algunos casos de oclusión de objetos por otros. De este modo, si existen pequeñas variaciones en los objetos a reconocer en las imágenes, estas variaciones desaparecen con la aplicación del filtro Max. Los resultados se ilustran en la Fig. 2.6

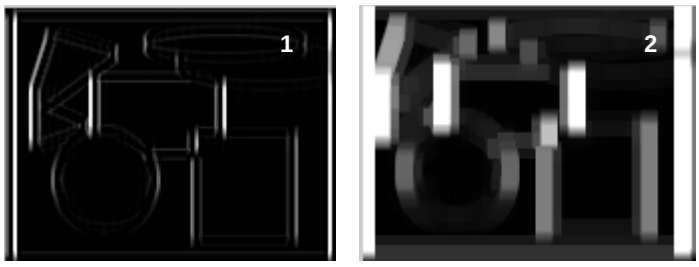


Fig. 2.6 En 1: imagen antes de aplicar el Max local. En 2: imagen después de aplicar el Max local.

En la tercera fase de la capa C1 se aplica un sub-muestreo de manera que se eliminan algunos píxeles. El sub-muestreo toma un píxel de cada 3 píxeles y los demás son eliminados en la primera banda, un píxel cada 5 en la segunda banda y de manera creciente para bandas más altas.

La operación de sub-muestreo quita información redundante reduciendo el tamaño de las imágenes, además al tener imágenes más pequeñas se reduce el costo computacional en las capas superiores. Los resultados se ilustran en la Fig. 2.7

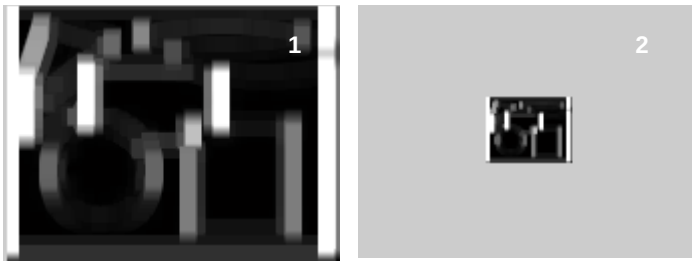


Fig. 2.7 En 1: Imagen antes del submuestreo. En 2: imagen después del submuestreo de C1.

Las bandas y los valores empleados para aplicar el Max en vecindad local así como los valores empleados para el sub-muestreo se resumen en la Tabla 2.2.

Tabla 2.2 Cada banda tiene un distinto tamaño de ventana para aplicar el max local, así como una distancia distinta para el submuestreo. Ambos valores son en píxeles.

#Banda	1	2	3	4	5	6	7	8
Tamaño de Ventana del Filtro Max.	8	10	12	14	16	18	20	22
Sub-muestreo	3	5	7	8	10	12	13	15

2.1.3 Capa S2

En el modelo biológico, la capa S2 representa un conjunto de neuronas simples que emiten un estímulo cuando un objeto es similar a uno observado antes, los estímulos son mayores a mayor similitud del objeto observado; en cambio cuando el objeto no guarda ninguna relación con algo observado antes, estas neuronas simples no se estimulan. Las neuronas de esta capa S2, están conectadas con las neuronas de la capa C1, por tanto, reciben la información generada por C1.

Tanto en el modelo general, como en el modelo simplificado, en la capa S2 se calcula una métrica para identificar la similitud de una región de una imagen del conjunto de entrenamiento habiendo sido procesada por C1 (denominada prototipo) con otra región de una imagen considerando un valor de distancia. Tal valor será mínimo cuando exista una similitud considerable entre el prototipo y una región de la imagen. El prototipo se barre sobre la imagen dando lugar a un conjunto de valores que indican los grados de similitud del prototipo en cada región de una imagen. Este valor se mide mediante:

$$D = IM^2 + PT^2 - 2conv(IM, PT) + C \qquad \textbf{(2.4)}$$

Donde IM^2 es el cuadrado de cada uno de los píxeles de la imagen obtenida de la capa anterior, PT^2 es el cuadrado de cada uno de los píxeles del prototipo y conv representa la convolución entre la imagen y el prototipo. Si una región de la imagen y el prototipo son coincidentes, el resultado D sería mínimo, indicando una similitud exacta. Con la finalidad de evitar el caso de una división entre cero si el valor D se emplea en otros procesos como divisor, C representa una constante con un valor muy cercano a cero para evitar que D llegue a ser cero.

Este conjunto de valores sólo representa a una imagen en una escala y orientación determinada. A este conjunto se le denomina vector de estímulos del prototipo en la imagen. Si aplicamos este prototipo sobre el resto de las escalas y orientaciones obtenemos un conjunto mayor de valores. A dicho conjunto se le denomina matriz de estímulos. En la Fig. 2.8 se esquematiza un prototipo y una imagen.

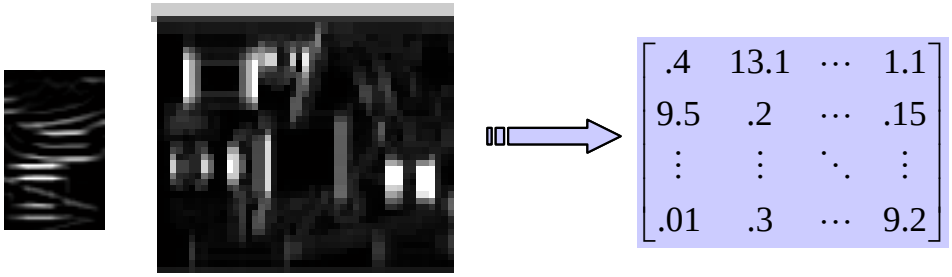


Fig. 2.8 Para el prototipo (izquierda) se obtiene su valor de similitud en la imagen central, que representa una escala y una orientación determinada, aplicado en todas las bandas, se obtiene una matriz de estímulos (derecha).

2.1.4 Creación de prototipos

Los prototipos son creados a partir de las imágenes de prueba positivas, se les aplica las capas S1 y C1, es decir, el banco de filtros Gabor, filtros max y submuestreo, sin embargo, sólo se toma el banco de filtros Gabor en dos escalas, (11x11 y 13x13) y en las 4 orientaciones (0, 45, 90 y 135 grados). Por tanto sólo hay una banda en los prototipos por cada imagen.

El número de prototipos es arbitrario, aunque en el modelo estudiado se manejan 250 prototipos. Emplear más prototipos puede ayudar a un resultado mas preciso en el reconocimiento de objetos, aunque el costo computacional será mayor. Los prototipos son extraídos al azar del conjunto de imágenes y son regiones también aleatorias de estas imágenes. El tamaño de los prototipos es variable, aunque se han sugerido en [13] tamaños de 4x4, 8x8, 12x12 y hasta 16x16 píxeles. Si consideramos que hay 4 orientaciones posibles se tienen 4 prototipos de distintas orientaciones al extraerlos de alguna región de una imagen.

De manera resumida, tenemos que si se considera el número de prototipos sugerido (250) tendríamos 4000 patrones para evaluar. Lo anterior se ilustra en la tabla Tabla 2.3

Tabla 2.3 Cantidad de Prototipos usados para el caso de que se extraigan 250 patrones aleatorios de las imágenes de entrenamiento.

Prototipos por Tamaño:	250 (Extraídos de imágenes aleatorias).
Orientaciones	4 orientaciones (0º, 45º, 90º, 135º)
Tamaños	4x4, 8x8, 12x12, 16x16 (4 tamaños)
Total:	250 x 4 x 4 = 4000 prototipos.

2.1.5 Capa C2

C2 representa a un conjunto de neuronas complejas cuya función es análoga a C1, extraer información invariante a partir de la información recibida por S2

En esta capa se calcula nuevamente la operación max ahora entre dos bandas adyacentes. Como los resultados de similitud de S1 son menores cuando mejor es el estímulo de un prototipo, la operación usada en C2 es un mínimo, análogo a la búsqueda de máximos en C1. Así se buscan mínimos en imágenes de misma orientación y bandas adyacentes. Posteriormente se hace una búsqueda de mínimos locales análogo a la fase de C1. Finalmente se extrae un sub-muestreo equivalente a la última fase de C1. Las operaciones aplicadas en esta capa se resumen en la Tabla 2.4

Tabla 2.4 En C2 se aplica un mínimo entre bandas y un mínimo local. Nuevamente se aplica un sub-muestreo.

No. Banda	1,2	3,4	5,6	7,8
Min local	8	12	16	20
Sub-muestreo	3	7	10	13

En esta capa, por tratarse de una medida de mínimos y un submuestreo, las imágenes resultan ya muy pequeñas y se consideran los valores como elementos de una matriz, en lugar de elementos de una imagen (píxeles).

2.1.6 Capas Adicionales

Es posible repetir las capas S2 y C2 con la finalidad de tratar de obtener mejores resultados en cuanto a la tarea del reconocimiento de objetos. Si bien el modelo ha sido descrito de manera general, en los trabajos de Serre y Poggio [13] sugieren la elaboración de las capas S3 y C3. De manera breve, la capa S3 representa un modelo de neuronas simples análogo a S2. El objetivo es ir incrementando la selectividad a los estímulos cercanos al objeto presentado en el conjunto de imágenes de entrenamiento. Aquí el modelo general toma un conjunto análogo de prototipos generados en C2 y estos prototipos son medidos con los conjuntos de valores de C2, nuevamente se aplica la métrica de similitud empleada en S2 con ciertas variaciones. En la implementación del modelo general en esta capa se usan mas prototipos (aproximadamente 500, en lugar de 250, como en S2). El impacto computacional de esta operación puede llegar a ser significativo, aunque los patrones de imágenes ya son de menor tamaño. De

manera análoga se forma un vector de estímulos por imagen, y tomando los valores en los diferentes pares de bandas y orientaciones se forma una matriz de estímulos. En lo que respecta a C3, en esta capa se realiza una función análoga a C1 y C2. Mientras en C1 se aplica un máximo en imágenes de dos escalas en una banda, y en C2 se aplica un mínimo en 2 bandas, en C3 se aplica un mínimo para todas las bandas, posteriormente se aplica un mínimo local y no se realiza un sub-muestreo como en las anteriores capas. Los mínimos se consideran ya como elementos numéricos de una matriz de manera análoga a como ocurre en C2. Se extraen mínimos de matrices que contienen valores de similitud, como aparece en la Fig. 2.9

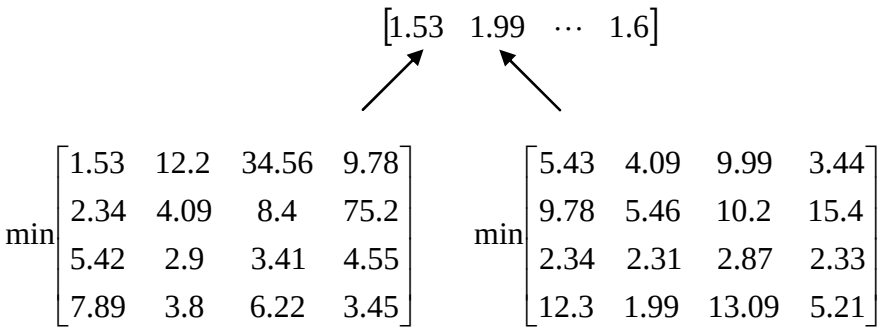


Fig. 2.9 Cada matriz representa los valores de similitud entre los patrones obtenidos de etapas anteriores. Un valor mínimo refleja un grado alto de similitud entre un prototipo y un área de la imagen.

Se describe también la posibilidad de agregar una capa de neuronas simples adicional: S4, que representa a una capa análoga a S3 y S2. Sin embargo, los prototipos son creados previamente. El objetivo es contar con prototipos de control que ayuden en el proceso de reconocimiento, pero que éstos no son tomados del conjunto de entrenamiento. Esta capa es la última del modelo para dar paso a la capa de categorización. Normalmente estos prototipos provienen de imágenes que contienen específicamente el objeto a reconocer, con un fondo en blanco. Esta capa, junto con S3 y C3 pueden omitirse. Un ejemplo de una imagen usada en S4 aparece en la Fig. 2.10

De manera general el proceso desde la Capa S1 hasta las capas adicionales, se ilustra, en su fase de entrenamiento, en la Fig. 2.11

Fig. 2.10 Ejemplo de imagen usada para S4. Esta imagen pasa por todas las capas, para obtener un prototipo del objeto a reconocer. Este prototipo es aplicado en S4 con la finalidad de mejorar la tarea de reconocimiento. En el modelo simplificado esta capa no es usada.

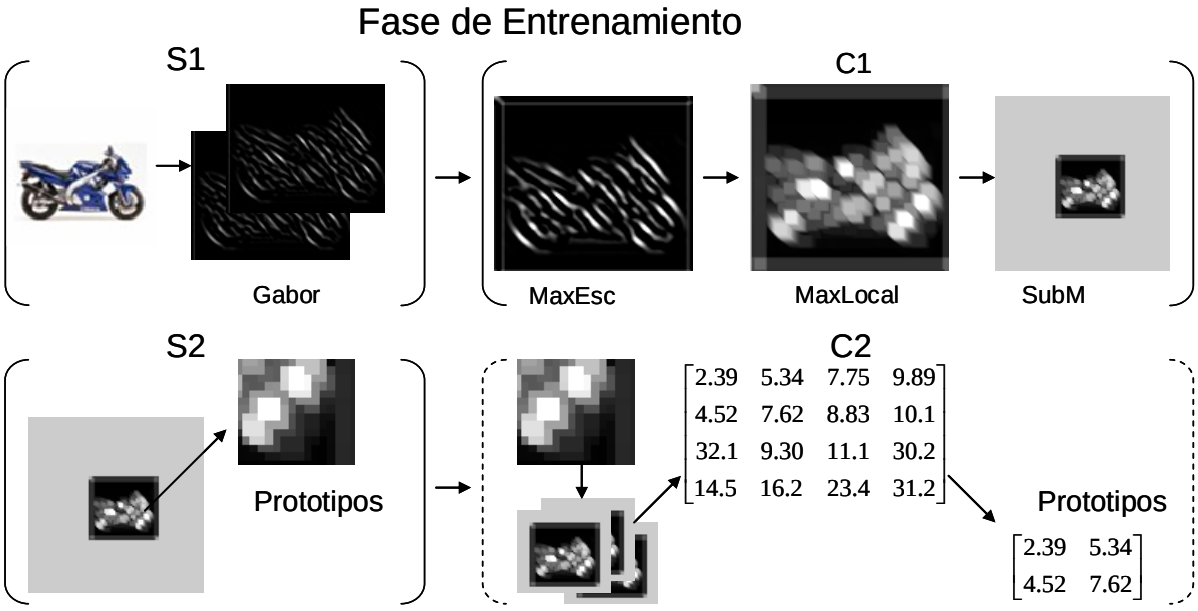


Fig. 2.11 Fase de Entrenamiento del Modelo General. En el recuadro punteado, el entrenamiento para el modelo extendido.

De manera resumida, la Fase de entrenamiento consiste en la aplicación del Filtro Gabor en S1, Aplicación de filtros Max en la capa C1, Extracción de prototipos de manera aleatoria en S2, y para el modelo extendido, extracción de nuevos prototipos a partir de las métricas obtenidas al evaluar los prototipos de S2 en el conjunto de imágenes de entrenamiento positiva. En cuanto a la fase de prueba, se puede observar un diagrama resumido en la Fig. 2.12 .

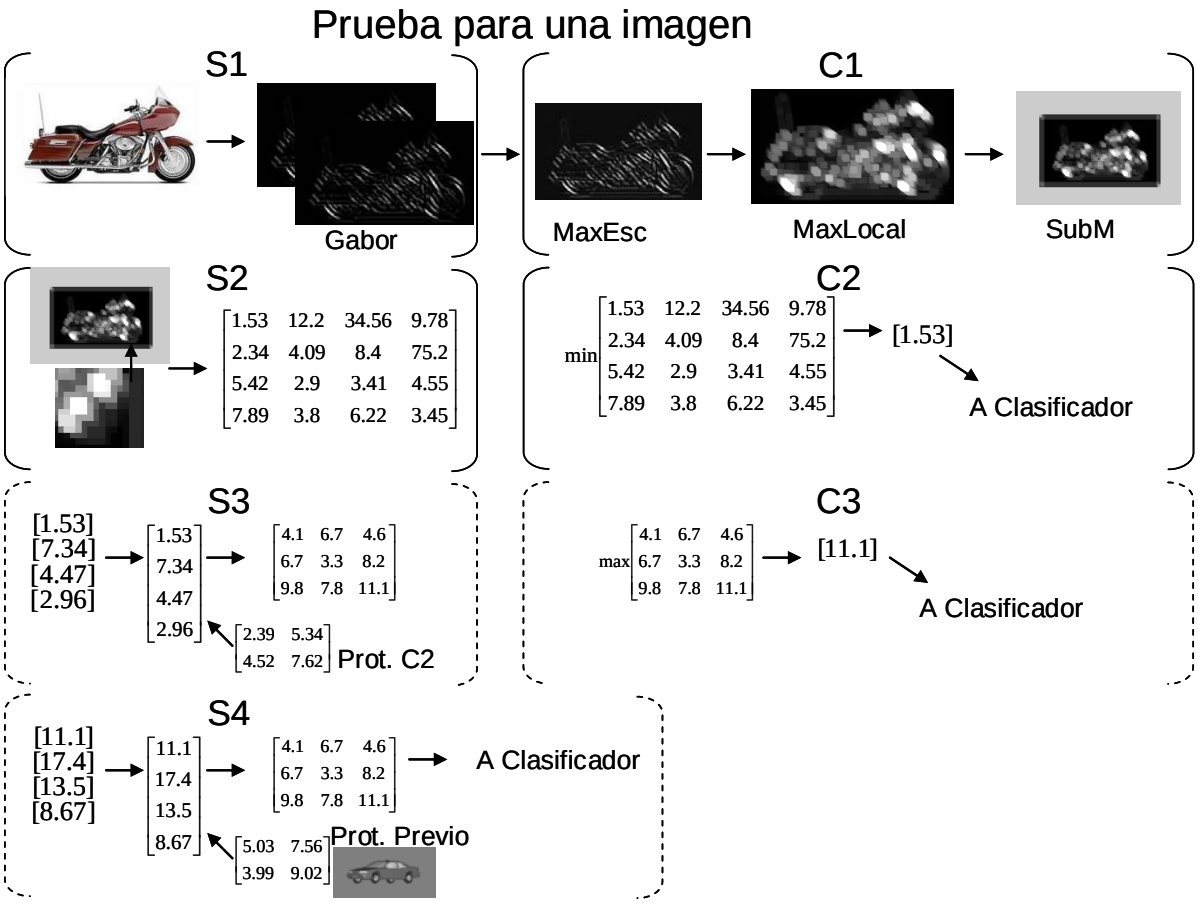


Fig. 2.12 Fase de prueba del Modelo General. En los recuadros punteados, las capas del modelo extendido (S3, C3 y S4)

De manera resumida, en la fase de prueba se aplica en un banco de filtros Gabor en S1, Se aplican los filtros Max y el submuestreo en C1, Se evalúa el grado de similitud de los prototipos en los conjuntos de imágenes a partir de una métrica determinada en S2, Se extraen los valores mínimos generados y, en el *modelo simplificado*, estos datos pasan a un clasificador. Para el modelo extendido en S3 se aplican los prototipos del entrenamiento en C2 aplicando una nueva métrica de similitud. En C3 se extraen valores máximos. En S4 estos valores nuevamente pueden ser evaluados con un conjunto previo de prototipos a partir de imágenes de entrenamiento sin ruido. Finalmente estos valores son enviados a un clasificador determinado, que recibe valores tanto de los conjuntos de entrenamiento como prueba, a fin de entrenarlo y clasificar las imágenes del conjunto de prueba.

2.1.7 Clasificación

Finalmente en esta etapa la información se categoriza y se determina si se ha encontrado un objeto en la imagen acorde con los conjuntos de entrenamiento. Esta categorización se realiza con un clasificador determinado. En el modelo general se sugiere la aplicación del clasificador vecinos más cercanos (nearest neighbor) o máquinas de soporte vectorial (SVM). Estos clasificadores (o cualquier otro, inclusive) recibe la información de los conjuntos de datos provenientes de las imágenes de entrenamiento para la primera fase del clasificador (entrenamiento) y los datos provenientes de las imágenes de prueba para hallar los resultados. Al final se obtiene un rendimiento de los resultados encontrados. La Fig. 2.13 ilustra los datos de entrenamiento y prueba para el clasificador.

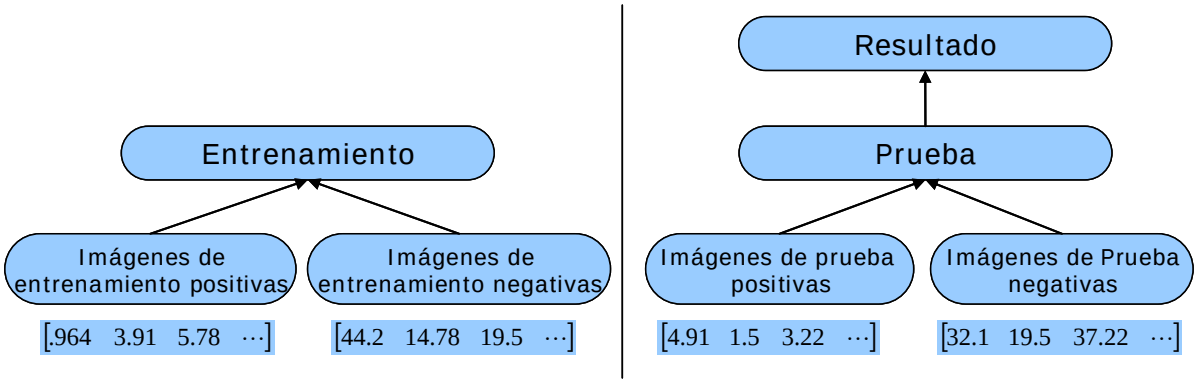


Fig. 2.13 En la fase de entrenamiento los valores generados por la última capa permiten entrenar algún clasificador determinado. En la fase de prueba los valores de la ultima capa son son evaluados para emitir un criterio de clasificación para cada imagen.

2.2 Otros modelos empleados en el reconocimiento de objetos.

Los modelos biológicamente inspirados en el sistema visual, están teniendo resultados interesantes en el campo del reconocimiento de objetos. De manera análoga, existen otros modelos que tienen por objetivo comprender mejor la percepción visual del hombre. Tal es el caso de los trabajos de Orbán y Fiser [11], en donde se argumenta que a partir de un conjunto de formas visuales simples (a manera de piezas de rompecabezas) que arman cierta estructura visual mas compleja, mediante un modelo bayesiano se logran aprender tales estructuras. Este tipo de entrenamiento es no supervisado, inspirándose en la percepción visual del hombre.

Desde un punto de vista más computacional, en la tarea del reconocimiento de objetos se considera que un sistema capaz de reconocer instancias de objetos en una imagen es muy deseable que reúna 3 características:

- **Invariante a la rotación**: Se espera que en una imagen si el objeto rota, el sistema deba ser capaz de seguir reconociendo el objeto. La rotación puede involucrar una rotación en cualquier sentido, tal como en un sistema 3D. Los modelos invariantes a la rotación en la actualidad son los más difíciles de conseguir en comparación con las otras 2 características.
- **Invariante a la escala**: Si la instancia del objeto a reconocer solo cambia de tamaño en la imagen, se espera que un modelo de reconocimiento de objetos sea capaz de reconocerlo si éste ha aumentado o reducido su tamaño.
- **Invariante a la traslación**: Si el objeto se mueve, el modelo debe ser capaz de encontrarlo. El que un modelo sea invariante a la traslación, por lo regular es la característica más fácil de conseguir de las 3 mencionadas.

Conviene indicar que también hay esfuerzos por reconocer objetos en condiciones un poco más adversas, como la oclusión del objeto a reconocer por otro.

En el caso del modelo de Serre y Poggio [13] anteriormente analizado, se indica en sus conclusiones que es un modelo en buena medida invariante a escala y traslación, así como pequeñas oclusiones del objeto, aunque limitado en el aspecto de rotación.

Normalmente la tarea de reconocimiento de objetos se divide en 2 niveles de visión: el procesamiento a bajo nivel (basado en la extracción de características como brillos, matices, formas, bordes, líneas, etc) y el procesamiento a alto nivel (interpretar esas características para darles una descripción simbólica. En Sucar y Gillies [9] se da este enfoque aplicado a la endoscopia en el tubo digestivo. Se emplea una representación de redes bayesianas con la finalidad de obtener un razonamiento probabilista de alto nivel.

Trabajos más enfocados a tareas de reconocimiento los encontramos en Mori y Malik [8], en donde se busca reconocer textos en imágenes aplicando métricas a partir de encontrar formas conocidas. Estos algoritmos se probaron con imágenes de texto

distorsionadas y adicionadas con fondos geométricos (*Captcha software*). En el trabajo de Felzenszwalb y Huttenlocher [12] se desarrollaron estructuras geométricas que buscan patrones de formas conocidas en imágenes. Este tipo de estructuras se enfocaron para el reconocimiento de rostros y del cuerpo humano. Boykov y Huttenlocher [1] y Lieshout [7], proponen de manera independiente un enfoque basado en campos aleatorios de Markov a partir de ciertas métricas aplicadas en instancias de objetos con oclusión parcial en imágenes binarias (blanco y negro).

Como puede observarse, por un lado existe una buena cantidad de trabajos que aplican la tarea de reconocimiento de objetos, aunque no de manera bio-inspirada. Por otro lado, muchos de estos trabajos involucran un enfoque bayesiano para esta tarea. En un esfuerzo tanto por comprender mejor el modelo biológico del sistema visual, como por aplicar un modelo bayesiano flexible, se plantea en esta tesis usar un enfoque bayesiano para el modelo del sistema visual, basado en el modelo de Poggio [13].

2.3 Conclusiones.

En este capítulo se analizó el modelo del sistema visual propuesto por Serre y Poggio. Se comprendió su funcionamiento para, a partir de éste, elaborar un modelo con un enfoque bayesiano que permita simplificar algunas características del modelo original. Es importante comentar que el modelo del que se parte es el modelo simplificado *(Standard Model*) aunque también se analizó brevemente el modelo extendido, que cuenta con mas capas que el modelo simplificado. Al ser este trabajo un primer esfuerzo en el sentido de llevarlo a un enfoque bayesiano, se decidió trabajar sobre el modelo simplificado, el cual presenta una estructura concreta para a partir de ella usar un enfoque bayesiano. Se estudiaron las capas del modelo simplificado y a partir de ello, en el siguiente capítulo se describen aspectos importantes de los modelos bayesianos, a fin de construir un enfoque alternativo que permita, al igual que el propuesto por Poggio, el reconocimiento de objetos en imágenes, previo entrenamiento.

3 Modelos Bayesianos

Los modelos bayesianos parten del Teorema de Bayes. Este enfoque parte de la idea de que es posible obtener información de probabilidad de una hipótesis dada cierta evidencia, conociendo la probabilidad a priori de dicha hipótesis y la probabilidad de la evidencia dada la hipótesis. De manera general se tiene:

$$P(H \mid E) \propto P(H)P(E \mid H) \quad \textbf{(3.1)}$$

La idea principal de este enfoque es construir un modelo que represente relaciones de dependencia entre ciertas variables con la hipótesis que se desea conocer. Estos modelos se pueden representar de manera gráfica mediante nodos las variables y mediante aristas las relaciones de dependencia. En este capítulo primeramente se dará una introducción a la estructura más sencilla: el clasificador bayesiano simple (CBS). Después de esto se estudiarán algunos casos del CBS, con la finalidad de aplicar estas estructuras en el modelo del sistema visual, entre ellas la elaboración de un filtro OR y su extensión en un filtro Max.

3.1 Clasificador Bayesiano simple

Un clasificador bayesiano simple (CBS, o también llamado, naïve bayes) se puede expresar mediante un grafo dirigido acíclico que consta de un número determinado de nodos (atributos) y una cantidad de estados incluida en cada nodo. Un nodo define la ocurrencia de un evento probable. Las probabilidades de ocurrencia de los distintos estados que puede adoptar ese evento se reflejan en una tabla de probabilidad condicional contenida en cada nodo. Esta tabla de probabilidad define las probabilidades de ocurrencia dada una clase (gráficamente descrito como el padre del grafo). Un CBS por definición asume que los atributos son independientes dada la clase. Esta independencia le permite simplificar el cálculo de la probabilidad de la clase, y gráficamente le da una forma de estrella como se muestra en la Fig. 3.1 Conceptualmente, un CBS permite obtener la probabilidad de ocurrencia de la clase, si se sabe de que otros eventos (conocidos como evidencia o atributos) ocurran. Un

ejemplo de un clasificador bayesiano simple y un ejemplo de tablas de probabilidad condicional aparecen descritos en la Fig. 3.2

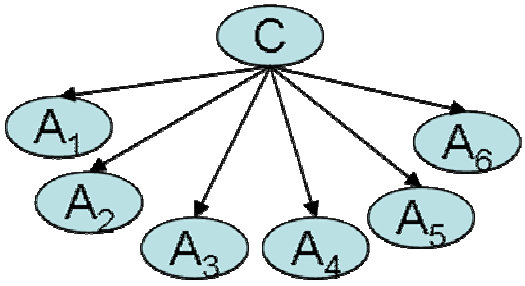


Fig. 3.1 Estructura de estrella de un CBS. C representa la clase y A cada uno de los atributos. Las aristas dirigidas solo emanan de la clase y no existen aristas entre los atributos. Ello representa la independencia de los atributos con respecto a la clase.

A partir de la información anterior se desea conocer la probabilidad de la clase dado ciertos valores de evidencia (para el ejemplo de la Fig. 3.2 , P(OR|A,B)). Por el Teorema de Bayes, tenemos:

$$P(OR \mid A,B) = \frac{P(OR)P(A,B \mid OR)}{P(A,B)} \propto P(OR)P(A,B \mid OR) \qquad \textbf{(3.2)}$$

Como un CBS asume independencia condicional de los atributos con respecto a la clase, el lado derecho de la Ecuación 3.2 es:

$$P(OR)P(A,B \mid OR) \propto P(OR)P(A \mid OR)P(B \mid OR) \qquad \textbf{(3.3)}$$

por lo tanto:

$$P(OR \mid A,B) \propto P(OR)P(A \mid OR)P(B \mid OR) \qquad \textbf{(3.4)}$$

De la Ecuación 3.3 se tiene que cada uno de los elementos del lado derecho se determina por las tablas de probabilidad condicional que aparecen en el ejemplo de CBS de la Fig. 3.2 . La imagen representa un caso particular de un CBS: una implementación (de muchas posibles) del operador OR lógico. Los valores o estados que pueden asumir las variables A y B son {0, 1} únicamente (aparecen como Evidencia 0 y 1 en las tablas). A cada nodo se le dio una tabla de probabilidad condicional: las probabilidades de ocurrencia de un estado dado que la clase asumiera cierto valor determinado. De modo que el total de combinaciones que pueden admitir son las dadas por la Tabla 3.1.

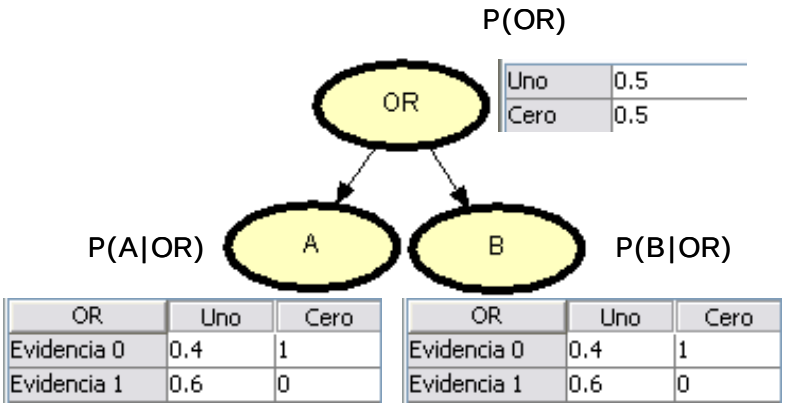


Fig. 3.2 Ejemplo de un CBS la Probabilidad de la clase P(OR) es igual para cada estado (sea cero o uno. Las Tablas de probabilidad condicional reflejan los valores de los estados de los nodos A y B (evidencia 0 ó 1) dada la clase OR.

Tabla 3.1 En la tabla de la izquierda, se muestra OR lógico tradicional. En la tabla de la derecha el OR a partir de un CBS. La información de la tercera columna es un valor de probabilidad de la clase 1, dado por el Teorema de Bayes.

A	B	OR lógico
1	1	1
1	0	1
0	1	1
0	0	0

A	B	OR Bayes
1	1	1.0
1	0	1.0
0	1	1.0
0	0	0.13

De acuerdo con la Tabla 3.1 en el CBS se infiere la probabilidad de que dada la evidencia presentada por A y B el valor de la clase OR sea 1. En los primeros casos, se tiene un 100% de conocimiento que el valor es 1. Tenemos el ejemplo para el caso 1,1:

Para el estado 1 de la clase OR:

$$P(OR \mid 1,1) \propto P(OR)P(1 \mid OR)P(1 \mid OR) = (0.5)(0.6)(0.6)=0.18$$

Para el estado 0 de la clase OR:

$$P(OR \mid 1,1) \propto P(OR)P(1 \mid OR)P(1 \mid OR) = (0.5)(0.0)(0.0)=0.0$$

Normalizando lo anterior, tenemos un valor de probabilidad de 1 para el estado 1. Para los casos (0,1) o (1,0) en el estado 0 de la clase OR, nuevamente habrá un cero entre los factores, produciendo un resultado igual:

Para el estado 1 de la clase OR:

$$P(OR \mid 0,1) \propto P(OR)P(0 \mid OR)P(1 \mid OR) = (0.5)(0.4)(0.6)=0.12$$

Para el estado 0 de la clase OR:

$$P(OR \mid 0,1) \propto P(OR)P(0 \mid OR)P(1 \mid OR) = (0.5)(1.0)(0.0)=0.0$$

Normalizando lo anterior, tenemos un valor de probabilidad 1, para el estado 1 de la clase OR, como se esperaba.

En el caso de que la evidencia presentada sea cero para A y B, el modelo presentado muestra un valor de 13% de probabilidad que sea 1. O dicho de manera complementaria, un 87% de que sea cero:

Para el estado 1 de la clase OR:

$$P(OR \mid 0,0) \propto P(OR)P(0 \mid OR)P(0 \mid OR) = (0.5)(0.4)(0.4)=0.08$$

Para el estado 0 de la clase OR:

$$P(OR \mid 0,0) \propto P(OR)P(0 \mid OR)P(0 \mid OR) = (0.5)(1.0)(1.0)=0.5$$

Normalizando lo anterior:

Uno: (0.08)/(0.08+0.5) = 0.137

Cero: (0.5)/(0.08+0.5) = 0.862

De modo que la probabilidad de que sea uno el resultado, es muy baja.

Este CBS define una operación lógica OR de manera probabilista. Si este concepto se generaliza, se obtiene una equivalencia a un filtro max, es decir un CBS que extraiga el valor máximo de un conjunto determinado. Adicionalmente podemos integrar una mayor cantidad de estados por atributo, es decir, no solamente definir los estados {0,1} como los posibles. Podríamos, definir {0, 1, 2, … k} en donde son k estados por atributo. Asi, generalizando el OR, y extendiendo hacia una mayor cantidad de atributos, podríamos proponer:

$$P(Max \mid A_1, A_2, A_3,...A_n) = \frac{P(Max)P(A_1, A_2,...A_n \mid Max)}{P(A_1, A_2,...A_n)} \propto P(Max)P(A_1, A_2,...A_n \mid Max) \qquad \textbf{(3.5)}$$

En donde análogamente, asumiendo independencia condicional de atributos:

$$P(Max)P(A_1, A_2,...A_n \mid Max) = P(Max)P(A_1 \mid Max)P(A_2 \mid Max)...P(A_n \mid Max) \qquad \textbf{(3.6)}$$

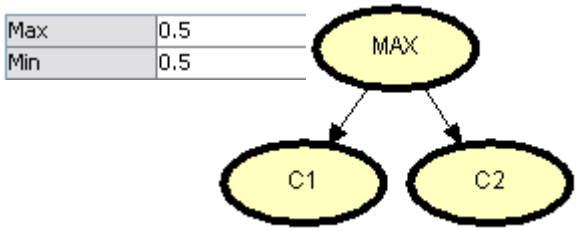
De lo anterior, sólo deben especificarse las tablas de probabilidad condicional para cada atributo dada la clase y la probabilidad a priori de la clase. La construcción de tales tablas de probabilidad, por lo regular no es trivial. Como se verá mas adelante, definir un clasificador bayesiano de manera que tenga cierta utilidad emulando el comportamiento de alguna función, servirá para el diseño de un enfoque probabilista

para el modelo del sistema visual. De esto, se analizará la adaptación de los enfoques bayesianos en el modelo del sistema visual. Adicionalmente, deben observarse algunos aspectos que un enfoque bayesiano involucra, a saber:

- Un CBS trabaja con datos discretos. En el caso de variables numéricas hay que discretizar. Esta operación puede ayudar a establecer un sesgo que puede ser útil al agrupar con mayor selectividad los conjuntos de datos.
- Un CBS resuelve una probabilidad de ocurrencia para cada uno de los estados de la clase, de modo que la información aportada puede resultar más útil al ofrecer un valor de probabilidad de ocurrencia de la clase (comprendido entre cero y uno), y no un valor lógico como uno o cero.
- Un CBS es una estructura que permite encadenarse a su vez a otras estructuras formales (ya sea otros CBS o redes bayesianas) de modo que la información puede propagarse ya sea hacia abajo o hacia arriba.

Si extendemos el modelo del CBS, a una emulación de filtro max, se requerirá también extender el número de estados. Partiendo de la base de las tablas de probabilidad condicional anteriormente expuestas para el OR (Fig. 3.2) se puede construir bajo un esquema equivalente: asignar probabilidades crecientes para la clase uno, (que bien puede denominarse clase max) y probabilidades decrecientes para la clase 0 (que bien puede denominarse clase min). Dado que la suma de probabilidades por columna de los atributos A y B debe ser 1, se puede construir una tabla de probabilidades ascendente y normalizada.

En la Fig. 3.3 aparece un posible CBS emulando una función Max. Adicionalmente, el CBS propuesto puede comportarse como una función Mín, en caso de que se considere la probabilidad del segundo valor de la clase. La tabla presentada en la figura es equivalente para ambos nodos. Es posible, por otra parte, extender la cantidad de atributos de la clase, aunque será necesario hacer ajustes en los valores de las tablas de probabilidad condicional, para evitar probabilidades cercanas a cero.



MAX	Max	Min
Cero	0.025	0.282857
Uno	0.043429	0.228571
Dos	0.08686	0.171429
Tres	0.108574	0.142857
Cuatro	0.162005	0.085714
Cinco	0.17372	0.057143
Seis	0.196434	0.028571
Siete	0.214978	0.002857

Fig. 3.3 CBS emulando una función max. Intrínsecamente tanto es una función max o mín. según la probabilidad de clase que se desee.

Este CBS consta de 8 estados por nodo (los nodos C1 y C2 tienen nombres arbitrarios). Cada nodo pude estar asociado a cualquier fuente de información de evidencia. El nodo superior, consta de dos estados únicamente, estos estados son complementarios, pues la suma de sus probabilidades da uno.

A manera de ejemplo podemos asumir dos casos:

Caso 1. Los píxeles tienen los valores C_1= 0, C_2=3.

Aplicando el Teorema de Bayes en el CBS para la clase Max:

$P(Max \mid C_1, C_2) \propto P(Max)P(C_1 \mid Max)P(C_2 \mid Max)$ = (.5)(.025)(.108574)=1.35x10^{-3}

Para la clase Mín:

$P(Mín \mid C_1, C_2) \propto P(Mín)P(C_1 \mid Mín)P(C_2 \mid Mín)$ =(.5)(.282857)(.142857)=2.02x10^{-2}

Normalizando:

Max: 1.35x10^{-3}/(1.35x10^{-3}+2.02x10^{-2})=.062

Mín: 2.02x10^{-2}/(1.35x10^{-3}+2.02x10^{-2})=.937

En este caso, la probabilidad entregada por Max es muy baja, una probabilidad proporcional al máximo entre los valores iniciales, por el contrario la probabilidad de mín es muy alta, lo cual en efecto nos conduce a que el elemento cero, es un mínimo.

Caso 2. Los píxeles tienen los valores C_1= 6, C_2=7.

Para la clase Max:

$P(Max \mid C_1, C_2) \propto P(Max)P(C_1 \mid Max)P(C_2 \mid Max)$ = (.5)(.1954)(.2149)=.0209

Para la clase Mín:

$P(Mín \mid C_1, C_2) \propto P(Mín)P(C_1 \mid Mín)P(C_2 \mid Mín)$ =(.5)(.0285)(.0028)=3.99x10^{-5}

Normalizando:

Max: .0209/(.0209+3.99x10^{-5})=.9998

Mín: 3.99x10^{-5} /(.0209+3.99x10^{-5})=.1.9x10^{-4}

En este caso, la probabilidad entregada por Max es casi 1, proporcional al valor máximo 7.

3.2 Redes Bayesianas.

Un clasificador bayesiano simple es un caso especial de una red bayesiana. El clasificador bayesiano simple asume que los atributos son condicionalmente independientes, es decir si A y B son atributos y C es la clase, se asume que P(A | B, C) = P(A | C). En un sistema real, por lo regular esta afirmación es falsa, puesto que algunos atributos pueden presentar dependencia condicional de otros. Curiosamente, a pesar de esto, un CBS se comporta bien en las tareas de clasificación. Sin embargo, a partir de este planteamiento, puede ser útil preguntarse como mejorar la calidad de clasificación de un CBS evitando asumir la independencia de de los atributos. Este camino conduce a la generalización de un CBS dando lugar a las redes bayesianas.

Una red bayesiana es una estructura en la que se capturan las relaciones de dependencia existentes entre los atributos de los datos observados. Las redes bayesianas describen la distribución de probabilidad concernientes a un conjunto de variables especificando suposiciones de independencia condicional junto con probabilidades condicionales. Así, las redes permiten especificar relaciones de independencia entre conjuntos de variables, de una manera más general que en el caso de un CBS.

Aunque en el enfoque bayesiano que se realiza en esta tesis, no se emplean redes bayesianas, conviene especificar su estructura y funcionamiento, al ser una generalización de un CBS.

Una red bayesiana es un grafo acíclico dirigido que describe la distribución de probabilidad conjunta que gobierna un conjunto de variables aleatorias.

Sea $U = \{X_1, X_2, ..., X_n\}$ un conjunto de variables aleatorias.

Formalmente, una red Bayesiana para U es un par $B = <G, T>$ en el que:

- G es un grafo acíclico dirigido en el que cada nodo representa una de las variables $X_1, X_2, ..., X_n$, y cada arco representa relaciones de dependencia directas entre las variables. La dirección de los arcos indica que la variable "apuntada" por el arco depende de la variable situada en su origen.
- T es un conjunto de parámetros que cuantifica la red. Contiene las probabilidades $P_B(x_i \mid \pi_{xi})$ para cada posible valor o estado x_i de cada variable X_i y cada posible valor π_{xi} de F(Xi), donde F denota al conjunto de padres de X_i en G.

La inclusión de las relaciones de independencia en la propia estructura del grafo hace de las redes bayesianas una buena herramienta para representar conocimiento de forma compacta, reduciendo el número de parámetros necesarios). Además, proporcionan métodos flexibles de razonamiento basados en la propagación de las probabilidades a lo largo de la red de acuerdo con las leyes de la teoría de la probabilidad. Este tipo de ventajas también se reflejan claramente en un CBS, aunque, como es de esperarse, el costo computacional de las redes bayesianas es mayor que en el caso de un CBS.

Existen varios trabajos relacionados con el empleo de redes bayesianas aplicadas para la representación de conocimiento en visión computacional (tanto imágenes naturales como sintéticas) como los de Sucar y Gillies [9] y Lieshout [7] los cuales se vieron anteriormente en el capítulo 2.

3.3 Resumen

En este capítulo se estudió el enfoque bayesiano con el clasificador bayesiano simple, la estructura mas sencilla basada en el teorema de Bayes, la cual asume independencia de atributos dada la clase. Se analizó un enfoque simple para encontrar la probabilidad de ocurrencia de una clase (clase OR) dada cierta información de evidencia por dos atributos. El objetivo fue modelar un OR lógico empleando un enfoque bayesiano. La generalización de esta estructura, con un mayor número de

atributos y estados, se puede interpretar como un filtro max. La elaboración de tablas de probabilidad es una tarea importante para lograr resultados deseados. Todo lo analizado en el presente capítulo, servirá de base para el desarrollo del enfoque bayesiano del sistema visual.

4 Modelo Probabilista del Sistema Visual

En este capítulo se plantea un modelo probabilista del modelo biológico del sistema visual.

De manera general, en el presente trabajo los modelos probabilistas están fundamentados en un clasificador bayesiano simple que, como se ha visto, es una estructura común, sencilla y flexible. Asimismo se tiene el cuidado de construir las tablas de probabilidad condicional de manera que presenten resultados análogos a los vistos en el modelo original y en el modelo simplificado. También en la construcción de este modelo se mantuvo un enfoque a modo de pirámide entre las estructuras bayesianas, esto con el objeto de que exista la posibilidad de fusionar las estructuras para crear una red bayesiana “integral”, es decir, que represente todo el modelo y en la cual en el futuro se pueda realizar inferencia de arriba hacia abajo.

4.1 Esquema General

El modelo probabilista propuesto está basado en el modelo simplificado que aparece en Serre y Poggio [13]. El modelo cuenta con las capas S1, C1, S2, C2 y una etapa de clasificación. Al ser un primer acercamiento probabilista no se considera el modelo extendido que itera más capas simples y complejas (S3, C3, S4). Un modelo con más capas presenta una estructura análoga, quizás con resultados más precisos aunque con un mayor costo computacional. El modelo probabilista propuesto emplea clasificadores bayesianos para hacer operaciones análogas al modelo simplificado, y sustentar las propiedades del modelo original. El modelo basado en clasificadores bayesianos presenta una estructura jerárquica con la posibilidad de conformar una red de mayor tamaño. En el enfoque presentado en esta tesis cada capa del modelo es flexible y puede modificarse. El esquema general se muestra en la Fig. 4.1 Este esquema general presenta las etapas del modelo desde la lectura de la imagen y su procesamiento por un banco de filtros Gabor en S1, la aplicación de filtros Max en C1, la extracción y evaluación de prototipos o patrones en S2 en la fase de entrenamiento y prueba, y nuevamente otra aplicación de filtros Max en C2. Los valores de esta capa

sirven de entrada para la categorización de las imágenes. La clasificación aunque puede realizarse con algún clasificador determinado, en este modelo por tratarse de emplear un enfoque bayesiano, también se usó un CBS.

Por lo anterior cada capa se reconstruye empleando el enfoque probabilista, lo cual se desarrolla en las siguientes secciones.

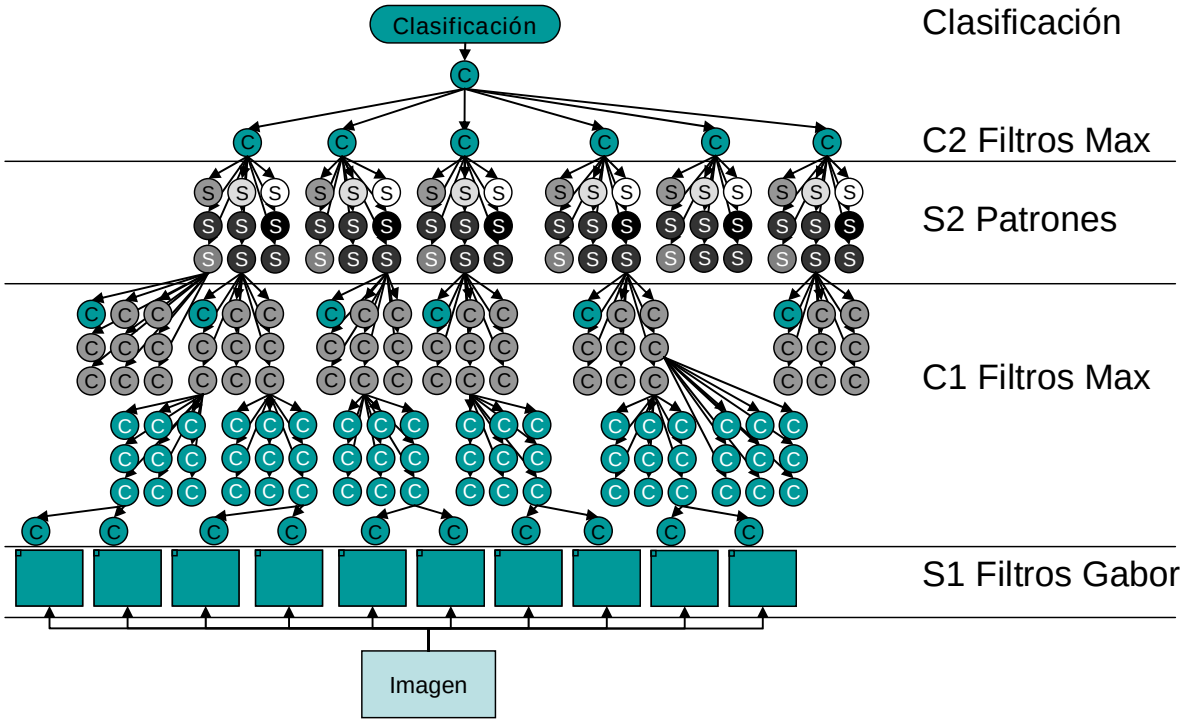


Fig. 4.1 Esquema general del modelo probabilista propuesto. Tiene 5 capas. La capa C1 tiene 3 fases. La estructura jerárquica del modelo permite poder construir un esquema bayesiano análogo a manera de una red. No se presentan todas las estructuras gráficas en la imagen por razones de espacio.

4.2 Modelado por capas

4.2.1 Capa S1: Banco de Filtros Gabor

La capa S1 es equivalente al banco de filtros Gabor. Esta capa se considera importante para extraer información de alguna orientación en las imágenes. El filtro Gabor realiza una extracción de bordes en una imagen dado cierto ángulo y cierta intensidad del borde, lo cual se refleja en información en cierta rotación y escala. Esta capa al considerarse más bien una etapa de pre-procesamiento de las imágenes a nivel computacional, no se consideró alguna alternativa probabilista. Construir un modelo

probabilista desde esta parte, llevaría a alejarse de las ideas propuestas por Hubel y Wiesel [2] en donde se explica la sensibilidad a la orientación de las neuronas simples que de acuerdo a los trabajos de Fukushima [4] denominan S1. En los modelos inspirados en el sistema visual estudiados, éstos parten de la idea del filtro Gabor, para obtener un conjunto de imágenes en cierta rotación y escala a partir de la original. Por lo anterior, el modelo probabilista se empieza a construir a partir de C1. El banco de filtros Gabor se aplica también en la misma cantidad de rotaciones y escalas que el modelo original. Sin embargo, una modificación importante en esta capa es que mientras en el modelo simplificado las imágenes generadas por el banco del filtro Gabor pasan tal cual a la siguiente capa, en este modelo se realiza una cuantización posterior al filtro Gabor. Esto se realiza puesto que la siguiente capa trabaja con unidades de valor discreto, por lo que un número en escala de grises que oscila entre 0 y 255 pudiera no ser conveniente. La capa de cuantización reduce los niveles de gris a 8 únicamente. El efecto de la cuantización se muestra en la Fig. 4.2

Fig. 4.2 A la imagen original (1), después de aplicar el filtro Gabor (2) se aplica una cuantización (3) para reducir los niveles de gris, de 256 a 8 niveles.

Aunque hay cierta pérdida de información en este proceso, el histograma de valores de gris en las imágenes generadas por el filtro gabor refleja pocos valores de gris distribuidos en la imagen, por el contrario, en su mayoría los valores generados por el filtro gabor se definen en píxeles blancos o negros. Esto refleja que una discretización no resulta en una pérdida significativa de información. Un ejemplo de un histograma puede observarse en la Fig. 4.3

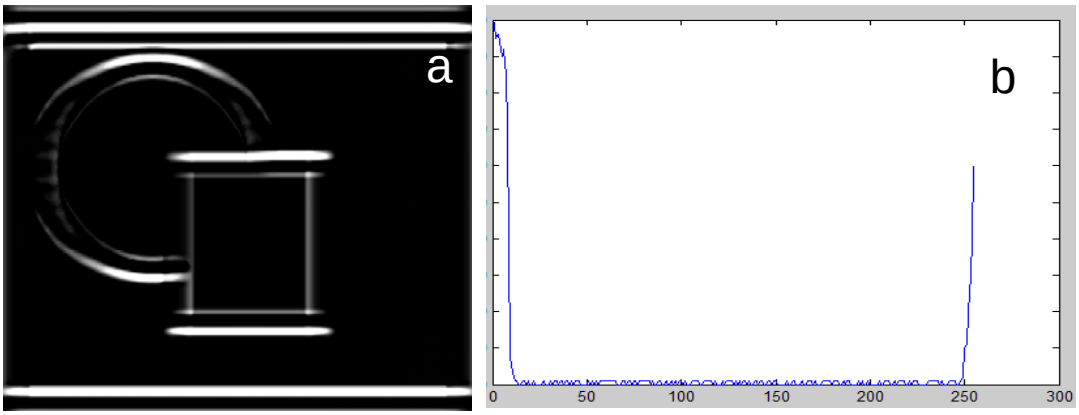


Fig. 4.3 El histograma (b) generado de una imagen (a) después de aplicar el filtro Gabor indica que una discretización no representa una pérdida significativa de información.

Debe tenerse en cuenta que la generación de filtros Gabor es congruente con la Tabla 2.1 del capítulo 2, es decir, se crean 8 bandas conteniendo cada banda elementos en diversas orientaciones y 2 escalas. Estas 2 escalas se fusionan en la próxima capa mediante un filtro Max.

4.2.2 Capa C1

En esta capa se aplican las operaciones del modelo simplificado de la capa C1. Primeramente se emplea un CBS para semejar la función max sobre el par de imágenes de igual orientación pero escala distinta, para cada banda.

Los píxeles de la imagen se asocian de manera proporcional a un estado de cada atributo en función de su intensidad de color. El estado cero es para píxeles oscuros y el estado siete es para píxeles claros. Debe tenerse en cuenta que esta asociación es directa dado que al final del filtro Gabor se cuantizó la imagen en 8 niveles de gris, que son correspondientes con los 8 estados de cada atributo en el CBS propuesto.

En términos de probabilidad, como se vio en el capítulo 3, se desea hallar el valor de probabilidad de que se haya encontrado un píxel máximo, dado los valores de los píxeles en C_1 y C_2 (uno de cada imagen en escala distinta), de manera análoga el CBS también se puede indicar la probabilidad de que se haya encontrado un píxel mínimo. En el caso de la probabilidad de un máximo, para obtener ese resultado por Bayes requerimos: $P(C_1, C_2 \mid Max) = P(C_1 \mid Max)\ P(C_2 \mid Max)$, dado que se considera en un

CBS que los atributos C_1 y C_2 son independientes. La probabilidad de la clase Max obtenida se asocia de manera proporcional a algún nivel de los 8 posibles en escala de grises, así probabilidades bajas se corresponden con colores oscuros (mínimos) y probabilidades altas con colores claros (máximos).

Los resultados de aplicar lo anterior se ilustran en la Fig. 4.4 En la parte izquierda se aplicó el máximo con una función Max normal, y en la parte derecha se empleó el clasificador bayesiano simple. La cantidad de información que se extrae es menor, aunque probablemente más significativa de la orientación de los bordes en la imagen. Por otra parte, conviene señalar que es posible modificar las tablas de probabilidad condicional, con lo cual se gana una mayor flexibilidad del modelo.

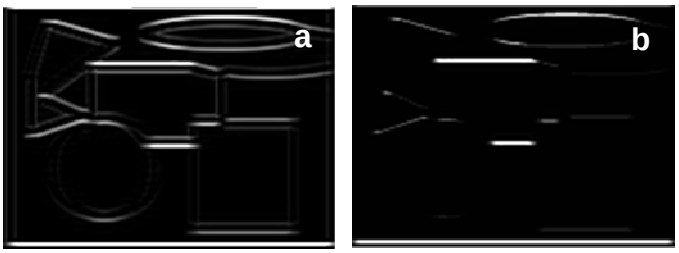


Fig. 4.4 La aplicación de un max normal en la imagen (a) se compara con un CBS en la imagen (b).

Tomando como referencia la Fig. 3.3 es posible observar de manera numérica el porqué se reduce la cantidad de información obtenida a través de un filtro Max bayesiano, mientras en un filtro Max normal se obtendría de Max(7,0) = 7, en un filtro bayesiano, la operación está dada por:

$$P(Max \mid 7,0) \propto P(Max)P(7 \mid Max)P(0 \mid Max) = (.5)(.214)(.025) = 2.675\text{x}10^{-3}$$

$$P(Min \mid 7,0) \propto P(Min)P(7 \mid Min)P(0 \mid Min) = (.5)(.002)(.282) = 2.82\text{x}10^{-4}$$

Normalizando: $2.675\text{x}10^{-3}/(2.675\text{x}10^{-3}+2.82\text{x}10^{-4}) = .904$

Si convertimos ese .904 a valores discretos entre 0 y 7, correspondería a 6. Por tanto tenemos un modelo más restrictivo a la operación Max, aunque más flexible, puesto que podríamos obtener resultados diferentes variando las tablas de probabilidad condicional.

El clasificador empleado es gráficamente semejante al mostrado en el capítulo 3. Se muestra en la Fig. 4.5 el resultado de aplicar la información de evidencia al clasificador es un valor de probabilidad el cual se toma proporcional al valor de píxel,

probabilidades más bajas para valores de píxel más oscuros, y probabilidades más altas para valores claros.

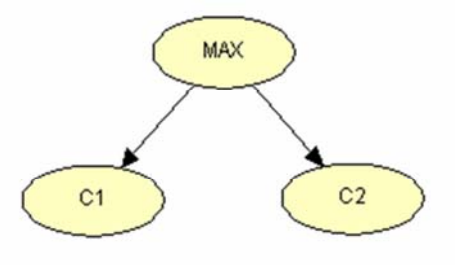


Fig. 4.5 El modelo de un clasificador bayesiano simple empleado en la primera etapa de la capa C1.

En la segunda fase de la capa C1 se aplica el filtro Max pero de manera local en la imagen. Nuevamente se modela un CBS, con más nodos. Por ejemplo se usaron 9 nodos para representar una ventana 3x3 (cada atributo es asociado a un valor de píxel). Para representar un Max local que abarque más área se aplica un CBS con una mayor cantidad de nodos (igualmente, uno para cada píxel). En la Fig. 4.6 se muestra un caso para nueve nodos, es decir, nueve píxeles. De manera análoga que en el CBS anterior, la probabilidad de la hipótesis que sea máximo (Max) dado la evidencia de cada nodo puede calcularse por el Teorema de Bayes:

$$P(C_1, C_2, C_3 ... C_n \mid Max) \propto P(C_1 \mid Max) P(C_2 \mid Max) P(C_3 \mid Max) ... P(C_n \mid Max) \qquad \textbf{(4.1)}$$

Sólo se requiere una multiplicación de probabilidades considerando que los atributos son independientes dada la clase.

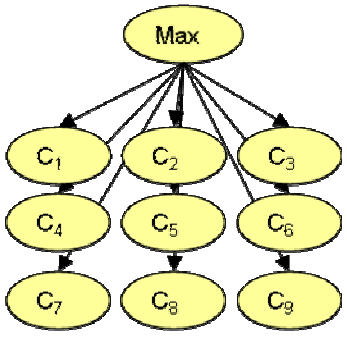


Fig. 4.6 CBS para vecindades de 9 nodos.

La tabla de estados para cada nodo es equivalente (8 estados) a la fase anterior (uno por cada nivel de gris). Debe observarse que se encontraron buenos resultados

cuando las tablas de probabilidad condicional para $P(C_k|Max)$ presentaban una estructura creciente exponencial de la forma:

$$y_j^+ = \frac{\exp(x_j)}{\sum_{\forall j \in E} \exp(x_j)}, \quad E = \{0,1,...n\} \tag{4.2}$$

Donde *n* representa el número de estados dados por la tabla de probabilidad condicional y E es el conjunto de estados para cada C_k. x_j se define:

$$x_j = \frac{j}{\sum_{\forall j \in E} j}, \; j \in E \tag{4.3}$$

Considerando los 8 estados (n=8), se representa una gráfica de los valores con la formula antes mencionada en la Fig. 4.7 De manera análoga para las probabilidades de la tabla dado el valor de la clase min, es decir, $P(C_k|Min)$, se sigue una curva exponencial decreciente, en este caso la fórmula difiere ligeramente quedando:

$$y_j^- = \frac{\max(X_j) - \exp(x_j)}{\sum_{\forall j \in E} \exp(x_j)}, \quad \mathrm{X_j} = \{x_j | \exp(x_j), j \in E\}, \quad E = \{0,1,...n\} \tag{4.4}$$

La gráfica decreciente de la ecuación anterior aparece en la Fig. 4.8

Los valores empleados a partir de las ecuaciones exponenciales se muestran en la Tabla 4.1. Los valores de dicha tabla están normalizados. Un ejemplo al aplicar el CBS de 9 nodos (para un filtro max 3x3) se muestra en la Fig. 4.9 .

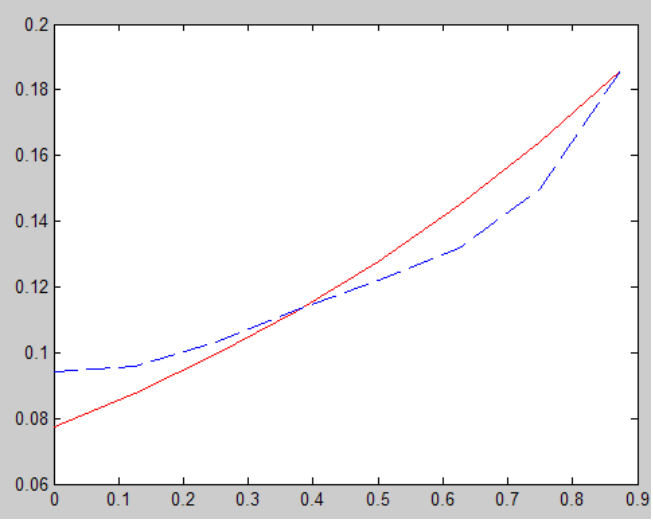


Fig. 4.7 grafica exponencial para la tabla del CBS en cada atributo dada la clase max. Ambas son exponenciales crecientes. La línea continua es un modelo de exponencial. La linea punteada es el modelo usado en la tabla, que es una aproximación de la exponencial.

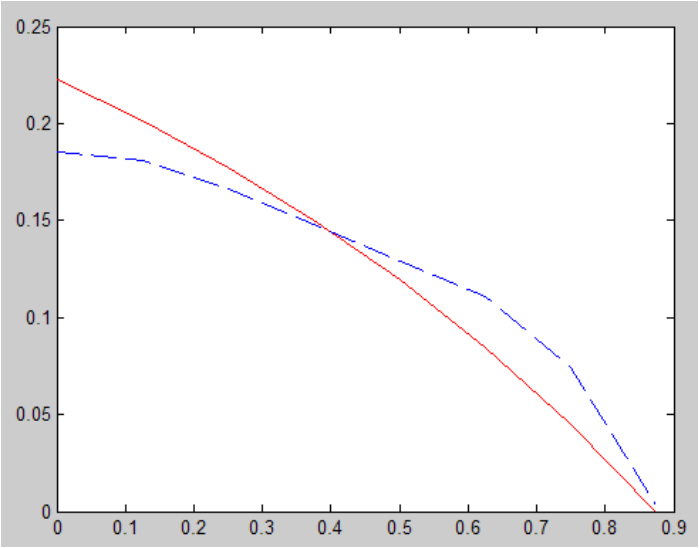


Fig. 4.8 Grafica exponencial decreciente correspondiente a la clase min. La curva continua es un modelo exponencial, la curva punteada es el modelo usado en la tabla, una aproximación de la exponencial decreciente.

Tabla 4.1 Tabla de Valores de probabilidad aplicados en cada nodo.

MAX	max, (y_j^+)	min (y_j^-)
Cero	0.094	0.185
Uno	0.096	0.181
Dos	0.103	0.166
Tres	0.113	0.148
Cuatro	0.122	0.129
Cinco	0.132	0.111
Seis	0.150	0.074
Siete	0.186	0.003

Los valores de la Tabla 4.1 han sido definidos de acuerdo a ensayos que permitan acercar los resultados a los de un filtro max local, por ello difieren ligeramente de una estricta curva exponencial. Si se desea obtener un resultado distinto a un max local, basta con cambiar la tabla de valores de probabilidad semejando alguna otra curva para obtener otros resultados.

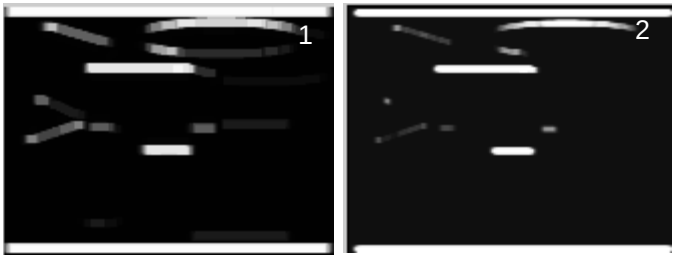


Fig. 4.9 En la imagen (1) el resultado de aplicar el filtro max (3x3) del modelo original. En la imagen (2) se usó el CBS como filtro max.

En la tercera fase, el sub-muestreo es aplicado empleando un CBS, dado que en un submuestreo solo interesa un valor de píxel y el resto es ignorado, al crear las tablas de probabilidad solo un atributo del CBS contiene una tabla que sea proporcional al valor de cada estado. El resto de los atributos contienen probabilidades iguales para cada estado (de 0.125) para evitar que tales nodos aporten información significativa. Gráficamente se representa como el diagrama de la Fig. 4.10 El objetivo es crear una estructura análoga al sub-muestreo, pero con la posibilidad de poder modificar las tablas y con ello decidir el elemento de la imagen a muestrear, así como inclusive que en el sub-muestreo dos píxeles o más puedan aportar información en vez de uno sólo. Las tablas del elemento a considerar y del resto de los nodos se muestran en la Tabla 4.2. Se debe resaltar que este CBS, al igual que los CBS de las etapas anteriores da un valor de probabilidad de la clase, que en este caso es proporcional al valor del píxel de cierto atributo. Las operaciones realizadas en esta capa pueden resumirse con el diagrama ilustrado en la Fig. 4.11 .

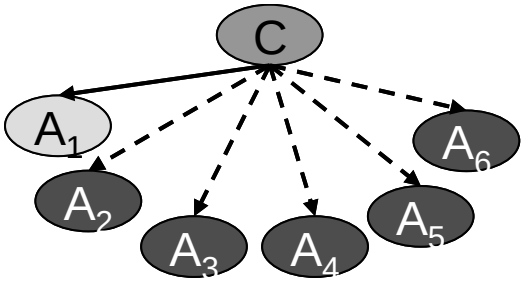


Fig. 4.10 CBS para la fase del submuestreo. Los atributos con líneas punteadas no aportan información significativa, puesto que sus tablas de probabilidad condicional contienen valores iguales.

La creación de éstas tablas permite una mayor flexibilidad en el manejo de estas capas, puesto que modificando las probabilidades se pueden tener resultados distintos

que, acordes con el modelo original, podrán aportar mayor o menor información al resto del modelo.

Tabla 4.2 La tabla izquierda representa el píxel a tomar en cuenta en el sub-muestreo. La tabla derecha se aplica al resto de los nodos.

SUB	ValorP	ValorN
Cero	0.002	0.228
Uno	0.034	0.202
Dos	0.068	0.173
Tres	0.102	0.144
Cuatro	0.136	0.115
Cinco	0.170	0.086
Seis	0.217	0.046
Siete	0.269	0.002

SUB	ValorP	ValorN
Cero	0.125	0.125
Uno	0.125	0.125
Dos	0.125	0.125
Tres	0.125	0.125
Cuatro	0.125	0.125
Cinco	0.125	0.125
Seis	0.125	0.125
Siete	0.125	0.125

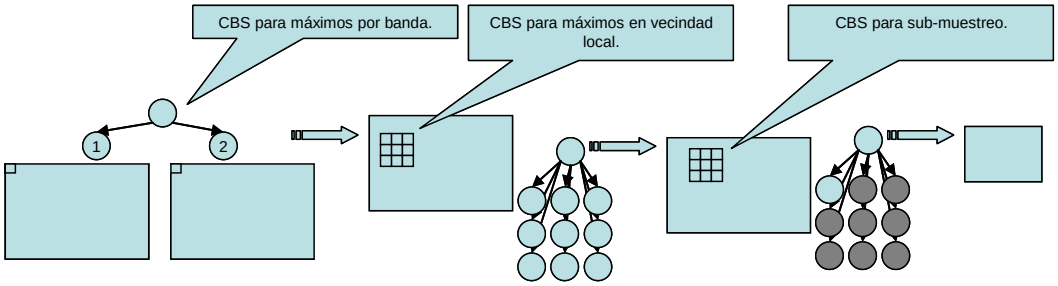


Fig. 4.11 Esquema de los filtros aplicados en la capa C1. Se extrae el máximo por píxeles de 2 imágenes de escalas distintas. Posteriormente se extrae el máximo local de una vecindad determinada. Finalmente se realiza un sub-muestreo que solo considera un píxel de una vecindad.

4.2.3 Capa S2

En esta capa se entrena un CBS con un prototipo extraído del conjunto de imágenes de entrenamiento, donde cada píxel es entrenado en el clasificador. Un prototipo es una región de una imagen de entrenamiento positiva tomada aleatoriamente, de modo que puede contener un patrón útil de un objeto a reconocer. El tamaño del prototipo puede ser variable. En el modelo se han trabajado prototipos de diversos tamaños como 9x9 píxeles y 11x11 píxeles. La manera de entrenar este clasificador es asignar una probabilidad alta al estado de un atributo relacionado proporcionalmente con el valor de píxel del prototipo procesado por las capas anteriores. A manera de ejemplo si el prototipo en uno de sus píxeles tiene el valor 5, la tabla del nodo correspondiente tendrá una mayor probabilidad en ese estado (el estado

5, de los 8 estados posibles). Este ejemplo se ilustra en la Tabla 4.3. Esta asignación de probabilidad busca una distribución normal con un valor pequeño para la desviación estándar (en las pruebas, fue de .3) de modo que la variación de probabilidad sea pequeña entre dos valores adyacentes. La asignación de probabilidades esta dada entonces por:

$$x_n = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_n-\mu)^2}{2\sigma^2}} \qquad \textbf{(4.5)}$$

Donde x_n representa cada uno de los estados, (8 estados), μ es el valor del estado que se desea reciba la mayor probabilidad (que se corresponde con el valor del píxel del prototipo de entrenamiento), σ es la desviación estándar.

Al final dichos valores obtenidos se multiplican por un factor de ajuste y son normalizados, para ser agregados a las tablas del clasificador bayesiano. El factor de ajuste se utiliza para tener una campana de Gauss congruente con los valores de probabilidad comprendidos entre cero y uno. Este factor es aproximado a .7 y además evita que aparezcan probabilidades iguales a uno en las tablas de probabilidad condicional. Empíricamente se ha comprobado que una desviación numéricamente muy pequeña no ayuda a encontrar probabilidades proporcionales para generar una confiable métrica de similitud con el prototipo a detectar, por tanto, los valores de probabilidad en la fase de entrenamiento del CBS son asignados de acuerdo a una distribución normal con una media asociada al píxel del patrón y con desviación estándar alrededor de 0.3. Esta operación es realizada con cada uno de los píxeles del prototipo seleccionado, de este modo el CBS tiene tantos nodos como píxeles tenga el prototipo. Una gráfica de la curva gausiana se muestra en la Fig. 4.12 Si bien con valores asignados a partir de curvas gausianas se han encontrado resultados aceptables en las pruebas realizadas, queda abierta la posibilidad a continuar haciendo modificaciones en función de las necesidades de los patrones a reconocer. Una tabla de valores derivada de la gráfica aparece en la Tabla 4.3. Finalmente este clasificador es barrido sobre las imágenes de prueba para encontrar regiones en las imágenes de prueba que sean semejantes al prototipo con que se entrenó el clasificador. El resultado es un conjunto de probabilidades de semejanza con el prototipo. Las operaciones en esta capa se resumen en la Fig. 4.13 .

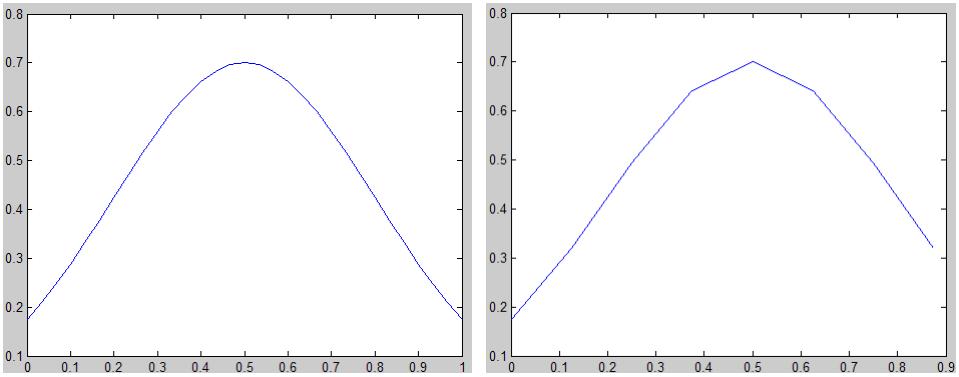


Fig. 4.12 La curva gausiana asigna mayor probabilidad al elemento que se asigne como media. En la gráfica la media es .5. Curva gausiana a la izquierda y curva empleada en las tablas de probabilidad condicional a la derecha, que es aproximación de la gausiana.

Tabla 4.3 Tabla de probabilidad para un nodo cuando su valor de píxel es 5. Esto produce probabilidades altas para patrones semejantes al prototipo con el que este clasificador haya sido entrenado.

MAX	afinidad al patrón	patrón diferente
Cero	0.108	0.134
Uno	0.116	0.130
Dos	0.118	0.126
Tres	0.133	0.116
Cuatro	0.134	0.113
Cinco	0.146	0.110
Seis	0.122	0.128
Siete	0.121	0.139

4.2.4 Capa C2

En esta capa se buscan probabilidades máximas (equivalente a la extracción de estímulos mínimos en el modelo simplificado). La búsqueda de estas probabilidades emplea un clasificador bayesiano entrenado previamente, que tiene tantos nodos como valores de probabilidad se consideren. Este CBS es similar al empleado en la capa C1 para la extracción de valores máximos locales. Las variaciones estriban en que la tabla de valores construida para cada nodo pone más énfasis en las probabilidades más altas, las cuales reflejan una mayor similitud de los prototipos empleados dentro de las imágenes. El diagrama de esta capa se muestra en la Fig. 4.14 .

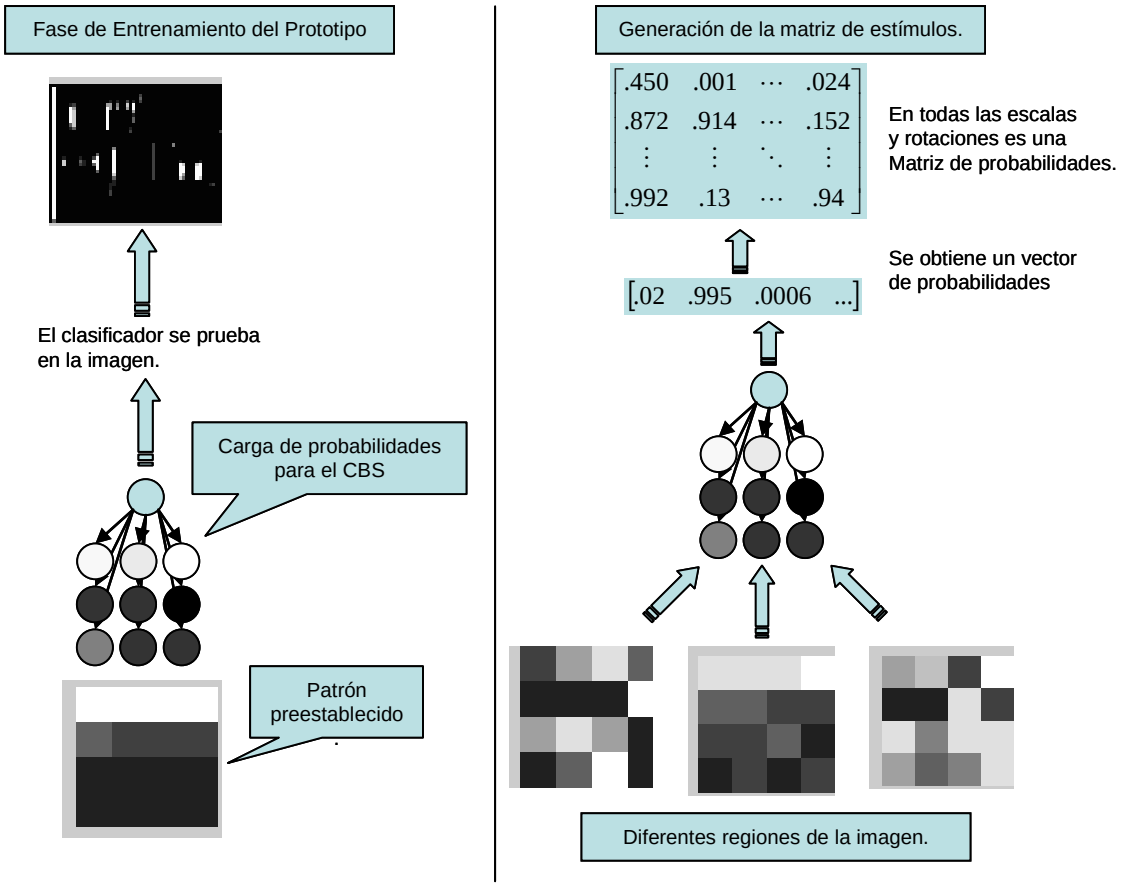


Fig. 4.13 El patrón preestablecido o prototipo se aprende por un CBS el cual cada tabla de sus nodos tiene probabilidades coincidentes con los píxeles del prototipo. En la fase de prueba se genera la matriz de estímulos, ahora matriz de probabilidades, al barrer el CBS sobre la imagen para buscar un patrón semejante al aprendido.

4.2.5 Clasificación

En esta fase la información proveniente de la capa C2 se pasa a un clasificador. Si bien en el modelo original y en el simplificado el clasificador empleado es vecinos más cercanos (*nearest neighbor*), en el modelo propuesto se emplea un clasificador bayesiano, con su fase de entrenamiento y prueba. La fase de entrenamiento emplea el conjunto de imágenes positivas y negativas para entrenamiento y la fase de prueba emplea por tanto el segundo conjunto (imágenes de prueba). Finalmente se extrae una estimación de éxito medida por el clasificador.

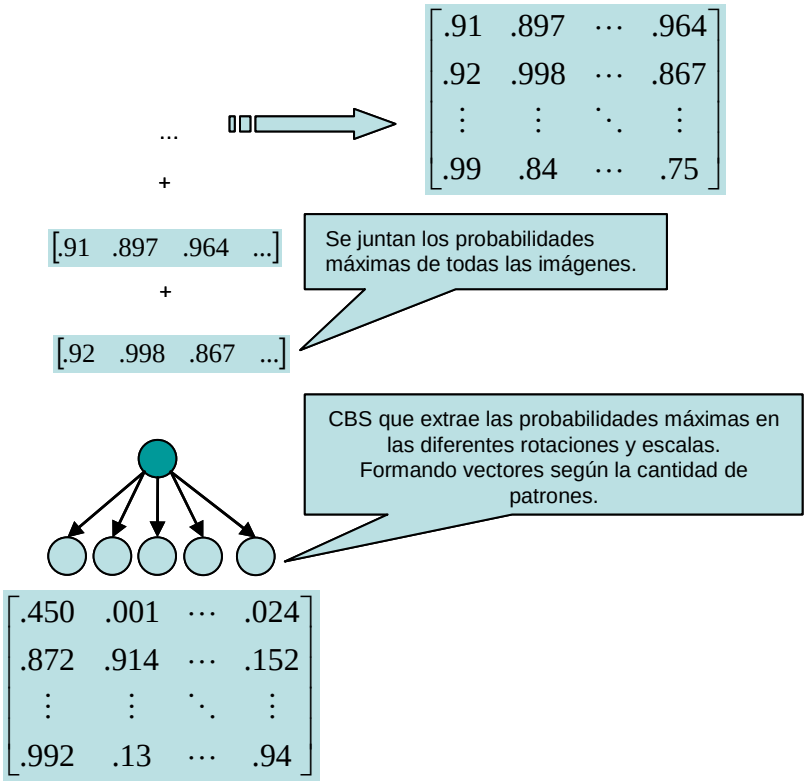


Fig. 4.14 Diagrama de C2. el CBS extrae el máximo del estímulo (valor de probabilidad) encontrado de cada prototipo en todas las rotaciones y escalas. Posteriormente se juntan los vectores máximos de todas las imágenes formando la matriz de probabilidades máximas.

Dado que las probabilidades obtenidas suelen estar numéricamente en valores muy cercanos a cero o uno, se cambia su escala por medio de una función logarítmica. En algunas observaciones realizadas los valores antes de aplicar la función logarítmica están en un intervalo de [$1.0x10^{-100}$, $1.0x10^{-9}$] Una vez aplicada la función logarítmica los valores obtenidos se discretizan a un numero de estados determinado para que puedan entrenarse en un CBS. Con la finalidad de obtener una mayor precisión en la discretización, el número de estados se puede incrementar en lugar de seguir trabajando con 8 estados como se hizo en los anteriores CBS. De manera empírica se probaron 16 y 32 estados obteniendo un incremento en la precisión de la clasificación. En la fase de entrenamiento son llenadas las tablas de probabilidad a partir de los valores obtenidos de los conjuntos de imágenes de prueba positivas y negativas. En la fase de prueba cada patrón emite un criterio de probabilidad por imagen para indicar si se ha encontrado un objeto o no. El valor umbral es movible aunque inicialmente se ha colocado para la clasificación en 0.5. De esta manera en la fase de prueba la

probabilidad de presencia (+) (o ausencia (-)) de un objeto en la imagen viene dado por las ecuaciones:

$$P(+\mid \mathrm{R}) \propto P(+)P(R_1 \mid +)P(R_2 \mid +)...P(R_n \mid +) \quad \textbf{(4.6)}$$

$$P(-\mid \mathrm{R}) \propto P(-)P(R_1 \mid -)P(R_2 \mid -)...P(R_n \mid -) \quad \textbf{(4.7)}$$

Donde R_k representa un prototipo determinado. $P(R_k|+)$ define la probabilidad dada por un patrón con respecto a la clase de la presencia del objeto entrenado, un bosquejo de esta etapa aparece en la Fig. 4.15

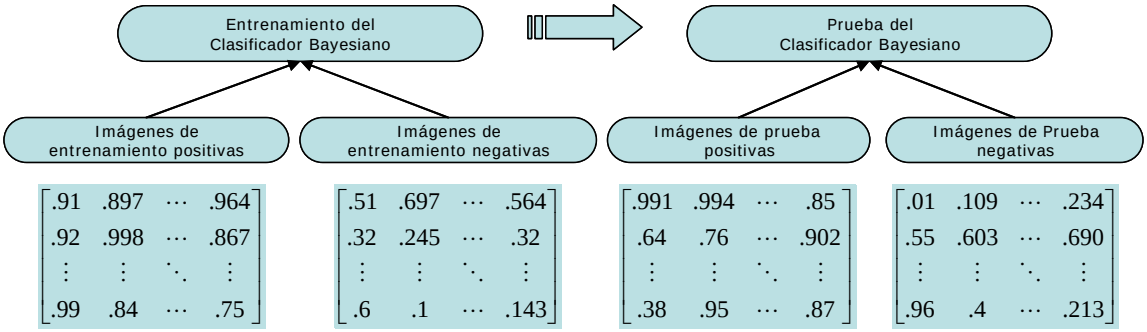


Fig. 4.15 Entrenamiento y prueba en la etapa del Clasificador. Se emplean las matrices obtenidas de C2. Se observa que el clasificador puede tener más estados por atributo, a fin de contar con una discretización más fina de los valores dados por las matrices de C2.

4.3 Resumen

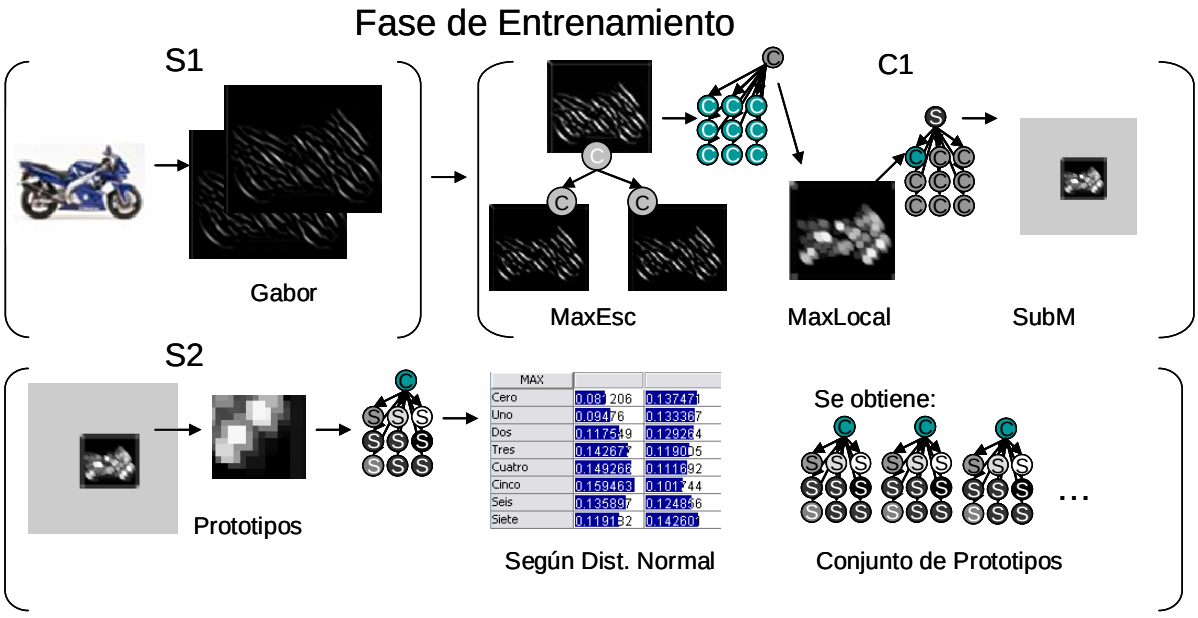


Fig. 4.16 Esquema de resumen general de la fase de entrenamiento del modelo bayesiano.

En este capítulo se desarrolló el enfoque bayesiano para el modelo del sistema visual. Está basado en el modelo simplificado del sistema visual, el cual consta de 5 capas, con una fase de entrenamiento y una fase de prueba. La fase de entrenamiento se ilustra en la Fig. 4.16 Se realizó un modelado por capas en donde la primera capa se manejó como una etapa de pre-procesamiento de la imagen, dejando el banco de filtros Gabor tal cual del modelo original. En las siguientes capas, cada una de ellas se modeló empleando clasificadores bayesianos. Se realizó una discretización posterior al filtro Gabor con la finalidad de poder trabajar con el CBS. La capa C1 constó de la aplicación de 2 fases de filtros Max y una fase de submuestreo, los resultados de esta capa se utilizan para la creación de prototipos de imágenes tomadas del conjunto de entrenamiento que en este enfoque bayesiano permiten entrenar un conjunto de CBS, de acuerdo a la cantidad de prototipos generados (recordar que cada píxel es un atributo, y como tal tiene su propia tabla de probabilidad condicional, donde los estados son los valores posibles que puede tomar el píxel, de 8 posibles). La cantidad de CBS entrenados es equivalente a la cantidad de prototipos deseados.

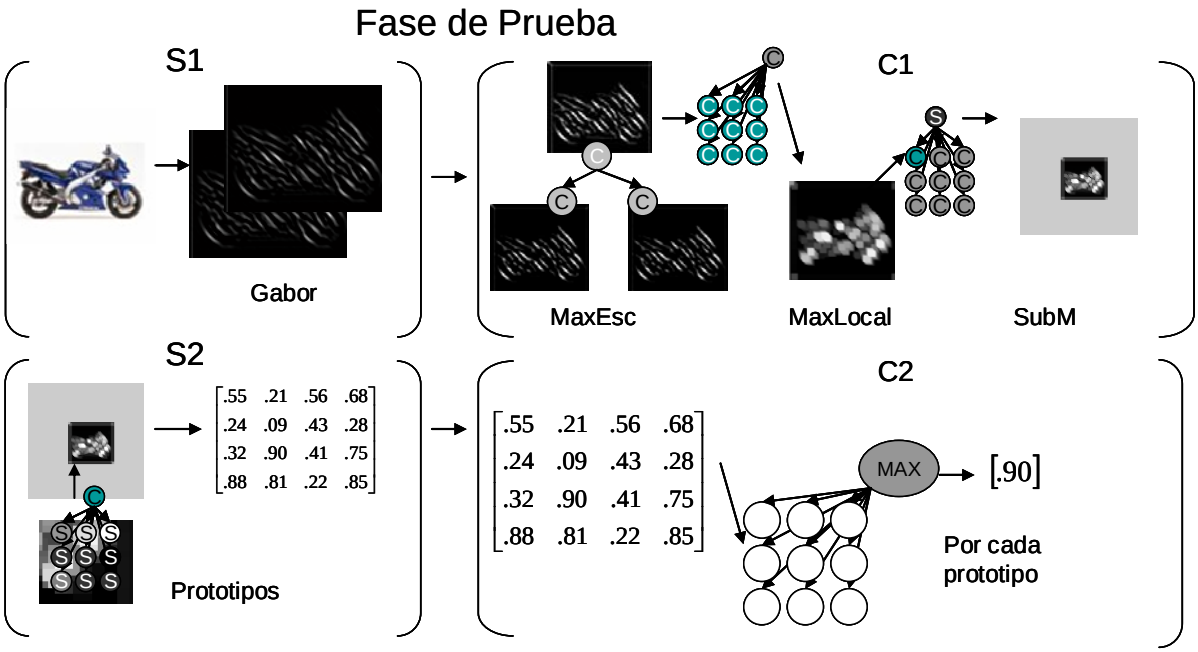


Fig. 4.17 Esquema de resumen general de la fase de prueba del modelo bayesiano.

Para la fase de prueba (Fig. 4.17), estos CBS evalúan las imágenes para obtener una métrica de similitud, habiendo barrido todas las imágenes cada uno de los CBS, teniendo una matriz de resultados por cada imagen. A diferencia del modelo

simplificado, en la siguiente capa C2, emplean un filtro Max, en vez de un filtro Mín análogo. Esto es así porque los CBS entregan una probabilidad cercana a 1 cuando alcanzan un grado alto de similitud y una probabilidad cercana a cero en caso contrario. Como los filtros Max empleando un enfoque bayesiano fueron usados en la capa C1, se crean CBS análogos en la capa C2.

Finalmente en la última capa (Fig. 4.18), se clasifican los datos previo ajuste con una función logarítmica así como una discretización. Esto con la finalidad de mejorar la distribución de los valores de probabilidad que entrenan las tablas de probabilidad condicional. En la prueba cada prototipo emite un valor que es evaluado en el CBS, entregando un valor indicativo si la imagen pertenece a la clase de objeto que se desea clasificar. Este modelo se probará en el siguiente capítulo, comparándolo con el modelo original y observando su comportamiento frente a variaciones de escala y rotación.

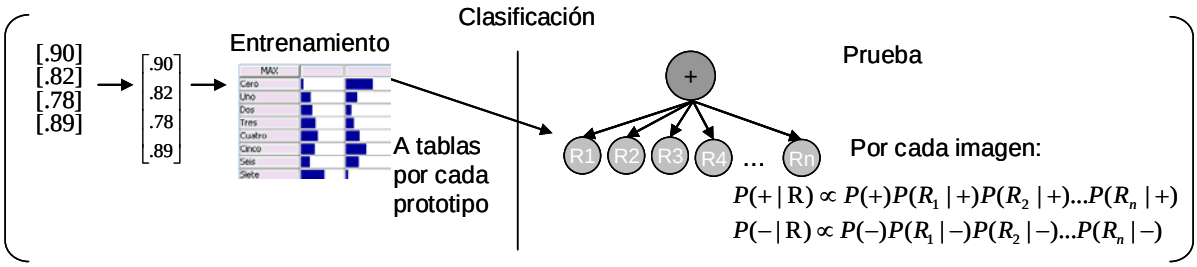


Fig. 4.18 Esquema de la fase de clasificación del modelo bayesiano.

5 Experimentos y Resultados

En este capítulo se presentan las pruebas realizadas al enfoque bayesiano del sistema visual, comparándolo con el modelo simplificado del sistema visual. En términos generales se encontraron resultados similares para ambos modelos, aunque el enfoque bayesiano resultó algo más eficiente en tiempos computacionales. Se encontró que existe una significativa invarianza a escala, aunque una limitada invarianza a rotación.

5.1 Metodología de Pruebas.

El enfoque bayesiano propuesto se comparó con el modelo simplificado, el código fuente en matlab del *Standard Model* puede ser descargado desde [16].

Para evaluar el modelo se requieren 4 conjuntos de imágenes, cada uno con una cantidad variable de imágenes, los cuales se dividen en 2 conjuntos para las imágenes de entrenamiento y otros 2 conjuntos para las imágenes de prueba. En cada caso, uno de ellos es para las imágenes positivas, que siempre contienen el objeto que se desea categorizar, y el otro es para las imágenes negativas, donde este objeto no aparece en las imágenes.

Al finalizar la prueba cada modelo emite un vector de resultados indicando si en cada una de las imágenes de prueba está presente la categoría del objeto a reconocer. Imágenes positivas clasificadas correctamente se denominan verdaderos positivos, y las imágenes positivas clasificadas incorrectamente se denominan falsos negativos. Análogamente, imágenes negativas clasificadas correctamente son verdaderos negativos, mientras que las imágenes negativas clasificadas de manera incorrecta, son falsos positivos.

Se realizaron diversas pruebas con repositorios de imágenes como la biblioteca de 101 categorías de imágenes Caltech [17], la cual está basada en el motor de búsqueda de imágenes de Google. El repositorio de imágenes de “background” se usó para emplearlos como conjuntos de entrenamiento y prueba negativos. Imágenes de ejemplo para cada repositorio se muestran en las figuras Fig. 5.1 y Fig. 5.2

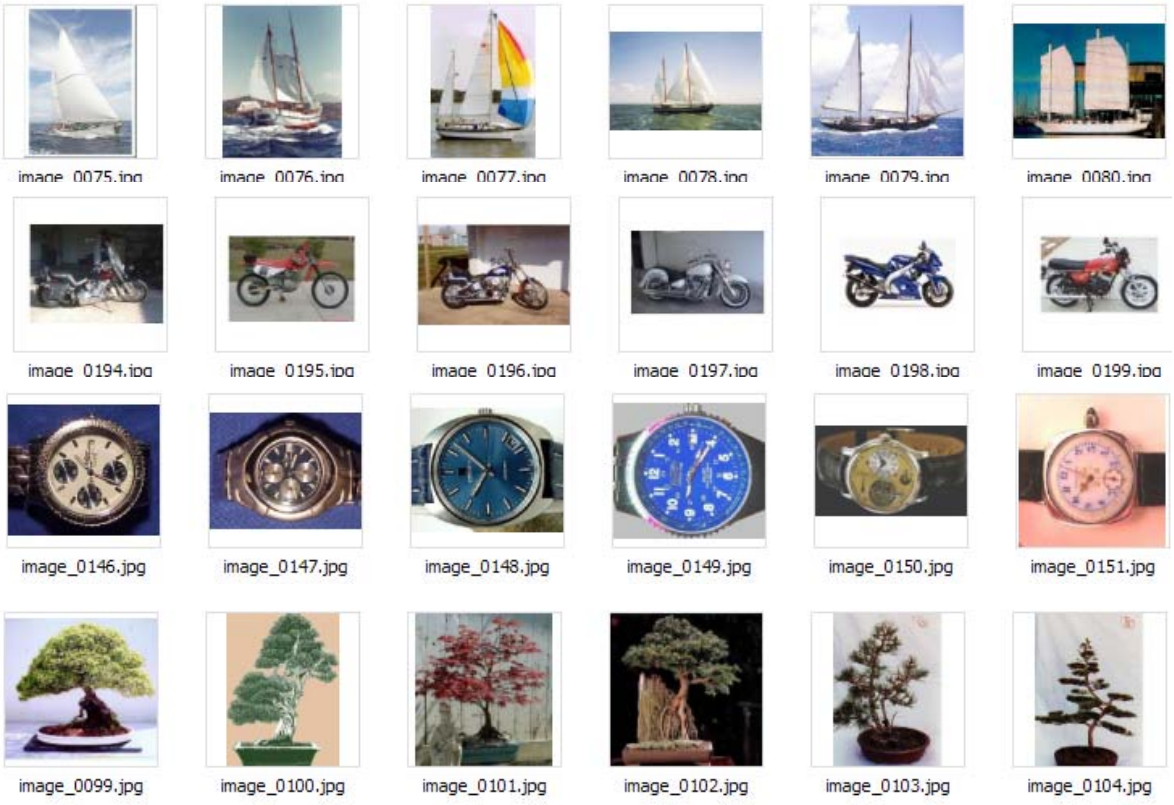


Fig. 5.1 La biblioteca Caltech contiene 101 categorías de objetos. En la imagen aparecen algunos de ejemplos de categorías como veleros, motocicletas, relojes y árboles.

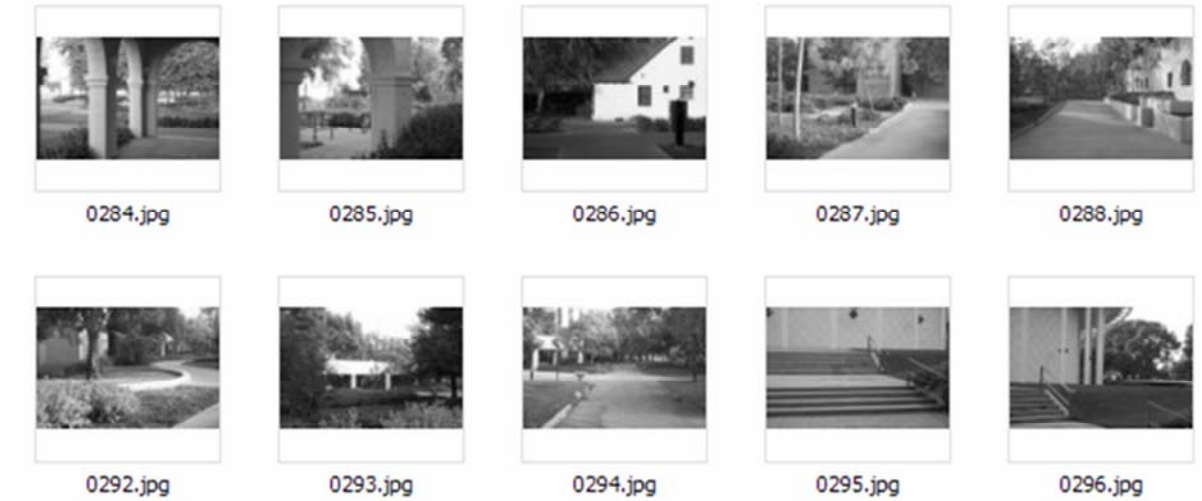


Fig. 5.2 La biblioteca background contiene imágenes naturales de diversas áreas de ciudad o campo sin contener algun objeto en particular.

5.2 Experimentos Realizados

Los experimentos realizados involucraron evaluar diversos conjuntos de imágenes de entrenamiento y prueba. Se emplearon categorías diversas como motocicletas, rostros, carros, casas e imágenes sintéticas, entre otras imágenes. La cantidad de imágenes empleada para entrenar osciló entre 50 y 120 imágenes para imágenes positivas, es decir conteniendo el objeto a reconocer, y un igual número para las imágenes negativas, que contengan otro tipo de información pero no el objeto. Esto aplicado tanto en los conjuntos de entrenamiento como de prueba. Se encontraron mejores resultados cuando el número de imágenes era relativamente grande (del orden de 100 imágenes). Con 100 imágenes en cada conjunto se trabajó en total con 400 imágenes, 200 en entrenamiento positivas (+) y negativas (-) y un igual número para las imágenes de prueba. Se trabajó en su mayoría con imágenes naturales, obteniendo un mejor resultado que al trabajar en imágenes sintéticas. Las imágenes sintéticas consistieron en letras y formas geométricas con colores sólidos en escala de grises, incluyendo cierto ruido como otras formas y letras. Este modelo fue implementado en código de Matlab, corriendo pruebas tanto en ambiente Windows XP Professional como Linux Ubuntu 7.04.

5.3 Resultados

5.3.1 Comparación con el modelo original

Al comparar los modelos con parámetros equivalentes de manera resumida se han obtenido los resultados que se muestran en la Tabla 5.1, donde ambos modelos presentan resultados semejantes.

La información entre paréntesis de la Tabla 5.1 refleja el porcentaje de aciertos en los conjuntos de prueba positivos y negativos. Se emplearon aproximadamente 400 imágenes en cada prueba (200 para entrenar y 200 para evaluar). Conviene señalar que en algunas pruebas tuvieron que ajustarse ciertos parámetros: (i) La cantidad de

orientaciones y escalas del filtro Gabor, (ii) el tamaño y cantidad de los patrones extraídos del conjunto de entrenamiento, (iii) la cantidad y calidad de imágenes del conjunto de entrenamiento. El ajuste que consistió en reducir la cantidad de escalas del filtro Gabor, logró una mejora de los tiempos computacionales y tuvo una pérdida de precisión en la clasificación prácticamente nula.

Tabla 5.1 Porcentajes de reconocimiento para entre el modelo Bayesiano y el Modelo de Poggio para ciertos conjuntos de imágenes. En paréntesis se muestra los porcentajes de clasificación para los conjuntos de prueba positivos y negativos.

	Modelo Bayesiano	Modelo Original
Motocicletas	90% (100%+,81%-)	89% (94%+, 85%-)
Casas	74% (86%+, 62%-)	73% (88%+, 58%-)
Rostros	85% (100%+, 71%-)	83% (100%+, 66%-)
Carros	81% (93%+, 69%-)	85% (100%+, 70%-)
Sintéticas	70% (74%+, 66%-)	68% (80%+, 56%-)
Rostros Similares	92% (100%+, 84%-)	99% (100%+, 98%-)
Promedio	82% (92%+, 72%-)	83% (93%+, 72%-)

En el caso del tamaño y cantidad de los prototipos extraídos, se probaron diversos tamaños que iban desde 5x5 píxeles hasta algunos de 16x16, teniendo mejores resultados con tamaños de 9x9, 10x10 y 11x11 píxeles. En cuanto a la cantidad de prototipos empleados, se probaron casos desde 10 prototipos hasta 250, teniendo resultados pobres con pocos prototipos. La clasificación presentada en las tablas corresponde con el empleo de 100 prototipos. La clasificación sólo mejora levemente para casos superiores a 100 prototipos. Por el lado de las imágenes las bases de datos empleadas contenían imágenes en dimensiones variables, que iban desde 100x100 píxeles hasta 300x300 píxeles. Este factor es el que más afecta el costo computacional del modelo. Cuando la cantidad de imágenes crece y también las dimensiones de las mismas, el rendimiento en ambos modelos se degrada de manera proporcional, es decir imágenes de 200x200 píxeles emplearán 4 veces más tiempo computacional que imágenes de 100x100 píxeles.

Tanto en el modelo original como en el enfoque bayesiano, se encontró una tendencia a la creación de falsos positivos, puesto que los porcentajes de clasificación correcta entre las imágenes negativas es marcadamente menor que en el caso de las

imágenes positivas, que sí contienen el objeto de la categoría para la cual se haya entrenado el modelo.

Tanto el modelo original como el modelo bayesiano propuesto presentó mejores resultados para la detección de rostros. Los casos de clasificación peores se encontraron en el conjunto de imágenes sintéticas (formas geométricas). Esta variación puede deberse a que los componentes de las imágenes sintéticas presentan una menor distorsión que en el caso de imágenes naturales, es decir, las imágenes sintéticas contienen más líneas rectas, mientras que las imágenes naturales presentan variaciones en la constitución del objeto. Dado que la extracción de prototipos es aleatoria y sólo corresponde a una parte de la imagen, es posible que la tarea de categorización reconozca unidades más elementales de los objetos geométricos.

En cada tarea de categorización, se probó variar el clasificador de la última capa a fin de conocer el mejor umbral de clasificación. En el promedio de los casos se obtuvieron curvas ROC del tipo mostradas en la Fig. 5.3 Una curva ROC (receiver operating characteristic) define una relación entre la razón de verdaderos positivos como ordenada (RVP) y la razón de falsos positivos como abscisa (RFP). También se le denomina una relación entre la sensitividad (RVP) contra (1 – especificidad) (RFP). Los RVP y RFP se pueden tomar de las matrices de confusión generadas al ir variando el umbral de clasificación desde 0 hasta 1 (el valor del umbral por defecto es 0.5). Las curvas ROC describen el comportamiento del clasificador; este comportamiento será mejor cuando la curva se pega a la esquina superior izquierda, y muestra un comportamiento opuesto cuando lo hace pegado a la esquina inferior derecha. En el caso de que la curva esté cerca de la diagonal principal indica que el clasificador no es efectivo pues está clasificando como un proceso aleatorio. En la Fig. 5.3 se muestra un caso para la detección de carros en imágenes. En general, el umbral de clasificación predeterminado es 0.5 (en escala de 0 a 1) y fue el que presentó mejores resultados.

5.3.2 Invarianza a rotación

El modelo presenta una pequeña invarianza a la rotación. En las pruebas realizadas el modelo aún presenta una buena clasificación al variar como máximo 15º la inclinación de las imágenes del conjunto de prueba. A orientaciones mayores la

clasificación decrece notablemente. Una inclinación de 90 grados con respecto al conjunto original hace que el reconocimiento del objeto resulte muy difícil. En varios casos, el modelo ya no encuentra el objeto, presentando porcentajes de clasificación inferiores al 50%, indicando que los conjuntos de prueba ya no corresponden con los conjuntos de entrenamiento. Si bien el modelo puede alcanzar a detectar pequeñas variaciones en las rotaciones entre 5 y 10 grados, para rotaciones superiores a 45º los prototipos tendrían que rotarse en el mismo ángulo que los objetos a reconocer, algo que en primera instancia no se conoce. De intentarse probar con rotaciones diferentes para los patrones, es probable también encontrar un incremento de falsos positivos, aunque, se estaría dejando de lado seguir el modelo biológicamente inspirado, lo cual tampoco sería el objetivo planteado a seguir. En la Tabla 5.2 se presentan algunos resultados con diversas orientaciones.

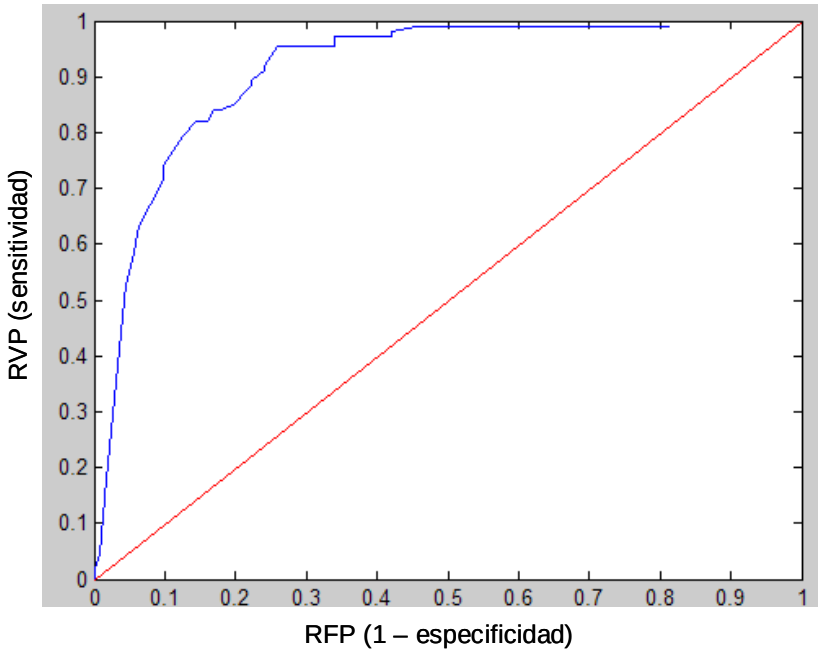


Fig. 5.3 Curva Roc para un conjunto de imágenes de carros.

5.3.3 Invarianza a escala.

La varianza en escala presenta mejores resultados si en el conjunto de prueba las imágenes han aumentado ligeramente su tamaño. Al disminuirlo la clasificación decrece

notablemente. Disminuciones de escala mayores al 50% de los tamaños del conjunto de prueba dificultan la detección de objetos. Los resultados de estas pruebas son análogos a la rotación. Como es de esperarse sólo decrecen los porcentajes de clasificación en los conjuntos de prueba positivos.

Tabla 5.2 Resultados de la clasificación con un conjunto de imágenes en diversas orientaciones.

	Modelo Bayesiano	Modelo Original
Motos a 0º	90% (100%+, 81%-)	89% (94%+, 85%-)
Motos a 5º	88% (96%+, 81%-)	89% (93%+, 85%-)
Motos a 10º	86% (92%+, 81%-)	87% (89%+, 85%-)
Motos a 15º	84% (87%+, 81%-)	83% (81%+, 85%-)
Motos a 90º	59% (37%+, 81%-)	58% (30%+, 85%-)

Tabla 5.3 Resultados de la clasificación con un conjunto de imágenes de rostros similares en diversas escalas

	Modelo Bayesiano	Modelo Original
Rostros Similares	93% (100%+, 86%-)	99% (100%+, 98%-)
Rostros a 110%	93% (100%+, 86%-)	98% (98%+, 98%-)
Rostros a 120%	92% (98%+, 86%-)	97% (96%+, 98%-)
Rostros a 150%	90% (94%+, 86%-)	94% (89%+, 98%-)
Rostros a 90%	91% (96%+, 86%-)	97% (97%+, 98%-)
Rostros a 80%	85% (84%+, 86%-)	92% (86%+, 98%-)
Rostros a 70%	82% (78%+, 86%-)	87% (75%+, 98%-)
Rostros a 50%	63% (40%+, 86%-)	67% (36%+, 98%-)

Cuando el objeto aparece de un tamaño mayor en la imagen, no hay pérdida de información significativa, por tanto se comprende que el modelo siga teniendo buenos resultados al ampliar su tamaño. Por el contrario al reducir la imagen a 50%, la pérdida de información contenida en ella es de aproximadamente 75%. Se encontró que al reducir la escala, se tenían mejores resultados si también se enfatizaba en usar escalas más pequeñas en el filtro Gabor. Por el contrario para el aumento de las dimensiones de los objetos, escalas altas también funcionaban bien.

5.4 Tiempos de Entrenamiento y prueba

Ambos modelos requieren un tiempo computacional considerable para la ejecución de la tarea. Para evitar tener tiempos muy largos de espera, se hicieron algunos experimentos con una reducción en la cantidad de escalas, así como en el número de prototipos. En un equipo Intel Pentium IV HyperThreading con una velocidad de 3.33 Ghz, con memoria RAM de 1.5 Ghz, cada tarea del modelo original, (como las que son presentadas en las tablas) empleó un tiempo promedio de 1 hora 15 minutos. El modelo requirió para el entrenamiento un poco menos de la mitad del tiempo (35 minutos aproximadamente). Conviene indicar que en estas pruebas se emplearon 100 prototipos de tamaño 9x9 y 11x11, 400 imágenes en total, imágenes de dimensión variable, entre 115x115 píxeles hasta 280x280 píxeles con un promedio de tamaños de 160x140 píxeles.

Por parte del enfoque bayesiano con características semejantes, el modelo tardó en promedio 35 minutos, empleando también poco menos de la mitad del tiempo (15 minutos) para la fase de entrenamiento. Cuando se experimentó con una mayor cantidad de prototipos, el modelo original podía llegar a emplear 6 horas su procesamiento, aunque según se observó los resultados mejoraban aunque ya no de manera significativa. En el caso del enfoque bayesiano, el modelo llegaba a tomar el tiempo de 2 horas y media. También se hicieron experimentos con una mínima cantidad de prototipos y una mínima cantidad de imágenes (10 imágenes en cada conjunto) aunque los resultados de clasificación son malos, a manera de referencia su procesamiento ocupó menos de 3 minutos en el modelo simplificado y menos de minuto y medio en el modelo bayesiano. En resumen el costo computacional del enfoque bayesiano es alrededor de 2 veces más rápido computacionalmente en promedio que el modelo original con un rendimiento (precisión en la clasificación) comparable.

5.5 Análisis

Al estudiar los resultados obtenidos de ambos modelos se puede concluir que ambos modelos presentan resultados similares mientras los parámetros resulten equivalentes, aunque se encontró que al restringir la cantidad de escalas, ambos modelos seguían presentando buenos resultados, pero el modelo bayesiano aún mostró

buenos resultados incluso al trabajar con solamente 4 escalas (por tanto, únicamente 2 bandas). Tanto el modelo bayesiano como el simplificado presentaron resultados buenos con restricciones de orientación (usar 1 ó 2 orientaciones en lugar de las 4 propuestas). Por el lado del número de prototipos, se encontró que sí es requerida una buena cantidad de ellos para obtener buenos resultados. Aunque a partir de una cantidad de prototipos superior a 100 ambos modelos presentan una mejora levemente creciente. Se pudo observar también que ambos modelos presentan una ligera tendencia a tener falsos positivos, ya que clasifican mejor los conjuntos de imágenes positivas que negativas, y esta tendencia se notó más marcada en el modelo bayesiano. Por otra parte en cuanto a los tiempos computacionales el costo del modelo bayesiano es significativamente menor, aunque por un factor constante. En las pruebas realizadas se concluye que al modelo bayesiano le toma aproximadamente la mitad del tiempo computacional con respecto del modelo original. En cuanto a la parte de rotación y escala, ambos modelos también presentaron resultados semejantes, siendo mayormente invariantes a escala que a rotación. La escala es bien soportada en casos de aumento, aunque no tanto en reducción. La rotación sólo es bien soportada para variaciones menores a 15 grados con respecto al conjunto de imágenes de entrenamiento. Finalmente ambos modelos resultan comparables en cuanto a los resultados mostrados, sin embargo, el modelo bayesiano planteado en esta tesis, resulta más flexible en cada de una de sus capas, como el que cada uno de los filtros Max son modificables en sus tablas de probabilidad, los prototipos extraídos en la capa S1 pueden modificarse al variar las tablas de probabilidad condicional de cada atributo-píxel, y la métrica de similitud de prototipos también es más flexible a ser modificada. También, considerando que el modelo bayesiano se diseñó con una estructura piramidal, es factible de diseñar un esquema de una red bayesiana global.

6 Conclusiones y Trabajo Futuro

6.1 Resumen

En esta tesis se desarrolló un enfoque bayesiano para un modelo del sistema visual. Se buscó comprender mejor el funcionamiento del sistema visual, desde un punto de vista computacional. En este sentido primeramente se estudió el modelo recientemente propuesto por Serre y Poggio [13], así como también algunos trabajos relacionados [4], [10]. A partir de clasificadores bayesianos simples, se realizó un modelo alternativo partiendo de semejar cada una de las capas del modelo simplificado propuesto por Poggio, en un modelo probabilista. Se probaron de manera comparativa ambos modelos con la finalidad de observar su comportamiento con algunas bases de datos de categorías de imágenes y con ciertos ajustes en varios parámetros de manera equivalente en ambos modelos, y se encontraron resultados similares. Se analizó su invarianza a rotación y escala y análogamente se encontraron resultados similares, presentando mejor invarianza a escala para casos de aumento de escala, mientras en rotación los resultados solo fueron buenos cuando la variación en grados no era muy grande (inferior a 15 grados). En el caso del costo computacional, ambos algoritmos requieren un tiempo superior a 30 minutos cuando se trata de grandes conjuntos de imágenes, aunque, en promedio, el costo computacional del modelo bayesiano es aproximadamente la mitad del costo computacional del modelo simplificado.

6.2 Conclusiones del modelo probabilista del sistema visual

Se ha encontrado que el modelo probabilista presenta buenos resultados para imágenes claras y definidas donde el objeto a reconocer aparece en primer plano. Al igual que en el modelo original, la invarianza a la rotación y escala es limitada, aunque esta podría mejorar si se agregan en los conjuntos de entrenamiento las imágenes en diferentes rotaciones y escalas. La cantidad de parámetros que pueden modificarse lo hace un modelo muy flexible que le permite crecer a fin de mejorar el reconocimiento de objetos. El modelo propuesto cumple con los objetivos planteados al inicio del desarrollo de esta tesis. Se logró realizar un enfoque bayesiano, más comprensible, fundamentado

en el sistema visual. Este enfoque interpreta de manera probabilista la información de cada capa, que a pesar de existir cierta pérdida de información en el proceso, presentó resultados similares que el modelo original, con una significativa mejora en cuanto al tiempo computacional para llevar a cabo la tarea. Adicionalmente se abre la posibilidad de continuar el desarrollo de enfoques probabilistas en el desarrollo de diversos modelos no solamente del sistema visual, sino también en otras áreas del análisis de imágenes.

6.3 Trabajo Futuro

Existen varias ramas por las cuales el presente trabajo se puede extender a fin de profundizar aún más en el empleo de modelos bioinspirados para el reconocimiento de objetos en imágenes.

- Construir un esquema global, empleando una red bayesiana que represente el modelo propuesto. La estructura por capas presentada en esta tesis, permite la construcción de un modelo global, pues en su diseño posee una estructura piramidal, que le permite a partir de ella, diseñar un modelo global. Con esto se podría lograr una inferencia en ambos sentidos así como incluir conocimiento previo.
- Trabajar en el desarrollo del modelo extendido a partir de este modelo simplificado (Standard Model) diseñando nuevas capas simples y complejas con el objetivo de analizar su funcionamiento para la tarea del reconocimiento de objetos.
- Mejorar el costo computacional del algoritmo a fin de poder procesar imágenes de mayor tamaño con mayor velocidad. El modelo reduce notablemente su rendimiento computacional cuando se emplean imágenes grandes (superiores a 640x480), es por ello que pudiera ser interesante diseñar algoritmos que estén pensados para computadores de procesamiento paralelo, como FPGA’s, a fin de mejorar el rendimiento del algoritmo. En el caso de una estructura bayesiana como los CBS, estos admiten realizar en paralelo una gran cantidad de operaciones.

- Aprender a mejorar los parámetros que emplea el modelo en sus capas a partir de ejemplos. Este enfoque buscaría establecer un esquema de aprendizaje para mejorar algunos parámetros que fueron dados en algunas capas del modelo.

REFERENCIAS

[1] Y. Boykov, D. P. Huttenlocher. A New Bayesian Approach to Object Recognition, Proceedings of IEEE CVPR, pp. 517-523, 1999

[2] D. H. Hubel, T. N. Wiesel (1968) Receptive Fields and functional architecture of monkey striate cortex. J. Phisiol., 195, pp. 215-243

[3] L. Fei-Fei , R Fergus and P. Perona. (2004) Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories.. In CVPR, Workshop on generative-Model Based vision.

[4] K. Fukushima (1980) Neocognitron: A self-organizing neural netowork model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics., Vol 36, No. 4., pp. 193-202.

[5] J. Jones, L. Palmer (1987) An evaluation of two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiology, Vol. 58, pp. 1223-1258

[6] Y. LeCun, F.J. Huang, L. Bottou (2004) Learning methods for generic object recognition with invariance to pose and lighting. In Proc. of CVPR 2004. IEEE Press.

[7] M. N. M. van Lieshout. A Bayesian approach to object recognition. (1991). Geometrical problems of image processing. Akademie Verlag. jan 1991. p.p. 185-190

[8] G. Mori, and J. Malik. Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA CVPR 2003

[9] L. Enrique Sucar, Duncan F Gillies. Probabilistic reasoning in high-level vision. Image and Vision Computing. Volume 12 number 1 January 1994 p.p. 42-60

[10] M. Riesenhuber and T. Poggio. (1999). Hierarchical models of object recognition in cortex. Nature Neuroscience., Vol 2, nov 1999 pp. 1019–1025.

[11] G. Orbán, J. Fiser, R. Aslin, M. Lengyel Bayesian model learning in human visual perception. Advances in Neural Information Processing Systems 18. 2000

[12] P. Felzenszwalb and D. P. Huttenlocher. Pictorial Structures for Object Recognition. Intl. Journal of Computer Vision, 61(1), pp. 55-79, January 2005

[13] Thomas Serre, Minjouon Kouh, Charles Cadieu, Ulf Knoblich, Gabriel Kreiman and Tomasso Poggio. A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. Center for biological Computational Learning. December 19, 2005

[14] D. Y. Tsao, W. A. Freiwald. Faces and objects in macaque cerebral cortex. Nature Neuroscience. Volume 6 number 9. September 2003

[15] H. Wersing and E. Corner (2003) Learning Optimizad features for hierarchical models of invariant recognition. Neural Comp., Vol 14., pp. 1559-1588.

[16] A new biologically motivated object recognition system. T. Serre, L. Wolf, T. Poggio. http://cbcl.mit.edu/software-datasets/standardmodel/index.html

[17] Caltech 101. L. Fei-Fei, M. Andreetto, M. A. Ranzato. http://vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html