

46  
2el.



# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES  
ARAGON

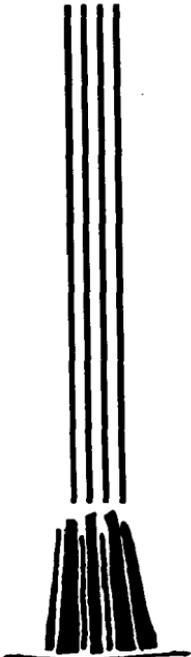
## PROCESAMIENTO DIGITAL DE IMAGENES UTILIZANDO REDES NEURONALES ARTIFICIALES

**T E S I S**  
QUE PARA OBTENER EL TITULO DE:  
**INGENIERO EN COMPUTACION**  
P R E S E N T A N:  
**MARIA ESTHER PALMERIN RUIZ**  
**CELIA PRADO REYES**

ASESOR DE TESIS:  
M. EN I. DAVID GONZALEZ MAXINES

SAN JUAN DE ARAGON, ESTADO DE MEXICO MAYO 1997

TESIS CON  
FALLA DE ORIGEN





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES ARAGON  
JEFATURA DE CARRERA DE INGENIERIA EN COMPUTACION

- M. en I. DAVID JAIME GONZALEZ MAXINEZ
- ING. MANUEL MARTINEZ ORTIZ
- ING. AMILCAR MONTERROSA ESCOBAR
- ING. ARTURO OCAMPO ALVAREZ
- ING. SILVIA VEGA MUYTOY

Informamos a ustedes de la autorización que se le concede a la alumna CELIA PRADO REYES, para desarrollar el trabajo de tesis "PROCESAMIENTO DIGITAL DE IMAGENES UTILIZANDO REDES NEURONALES - ARTIFICIALES", conjuntamente con MARIA ESTHER PALMERIN RUIZ — dirigida por el M. en I. David Jaime González Maxinez, solicitan do a ustedes, sean tan amables de revisar el avance del mismo y hacer las observaciones que consideren pertinentes, o en su caso indicar a las alumnas si dicha revisión se hará a la conclusión del trabajo de tesis.

Sin otro particular, me es grato enviarles un cordial saludo.

A T E N T A M E N T E  
"POR MI RAZA HABLARA EL ESPIRITU"  
San Juan de Aragón, Edo. de Mex., 16 de agosto de 1998  
EL JEFE DE LA CARRERA

ING. SILVIA VEGA MUYTOY





UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
CAMPUS ARAGÓN

UNIDAD ACADÉMICA

Ing. JUAN CASTALDI PÉREZ  
Jefe de la Carrera de Ingeniería en Computación,  
Presente.

En atención a la solicitud de fecha 12 de febrero del año en curso, por la que se comunica que las alumnas MA. ESTHER PALMERÍN RUIZ y CELIA PRADO REYES, de la carrera de Ingeniero en Computación, han concluido su trabajo de investigación intitulado "PROCESAMIENTO DIGITAL DE IMÁGENES UTILIZANDO REDES NEURONALES ARTIFICIALES", y como el mismo ha sido revisado y aprobado por usted, se autoriza su impresión; así como la iniciación de los trámites correspondientes para la celebración del Examen Profesional.

Sin otro particular, le reitero las seguridades de mi atenta consideración.

ATENTAMENTE  
"POR MI RAZA HABLARA EL ESPÍRITU"  
San Juan de Aragón, México, 13 de febrero de 1997  
EL JEFE DE LA UNIDAD

  
Lic. ALBERTO IBARRA ROSAS

c c p Asesor de Tesis.  
c c p Interesado.

AIR'la.

## INDICE

<b>Prólogo.....</b>	<b>1</b>
<b>1 Introducción A Las Redes Neuronales Artificiales .....</b>	<b>3</b>
Introducción.....	3
1.1 Qué es una red neuronal artificial. ....	5
1.2 Funcionamiento de la red.....	6
1.3 Similitudes entre una red neuronal artificial y una neurona biológica. ....	7
1.4 Aplicaciones de redes neuronales artificiales .....	8
<b>2 Componentes Y Modelos De Una Red Neuronal Artificial.....</b>	<b>9</b>
Introducción.....	9
<b>2.1 Componentes de una red neuronal artificial .....</b>	<b>9</b>
2.1.1 Elementos de entrada y elementos de salida .....	9
2.1.2 El Peso en las redes neuronales artificiales .....	10
2.1.3 Vectores de entrada, vectores ocultos y vectores de salida. ....	11
2.1.4 Conexiones .....	15
2.1.5 Elementos de procesamiento.....	17
<b>2.2 Modelos básicos de una red neuronal artificial .....</b>	<b>21</b>
2.2.1 Modelo Adaline .....	22
2.2.2 Modelo Madaline.....	26
<b>3 Introducción Al Procesamiento De Imágenes .....</b>	<b>29</b>
Introducción.....	29
3.1 Serie y transformada de Fourier.....	29
3.2 Qué es una imagen digital .....	35
3.3 Tipos de imágenes .....	35
3.4 Elementos de un sistema de procesamiento digital de imágenes.....	36
3.5 Escala de grís.....	37
3.6 Luminancia.....	38
3.7 Color .....	39
<b>4 Técnicas Utilizadas En El Procesamiento De Imágenes .....</b>	<b>42</b>
Introducción.....	42
<b>4.1 Técnicas de transformación de imágenes.....</b>	<b>42</b>
4.1.1 Transformación unitaria de dos dimensiones.....	42
4.1.2 Transformación unidimensional .....	42

<b>4.2. Técnicas de procesamiento de imágenes empleando transformada.....</b>	<b>43</b>
4.2.1 Transformada de Fourier.....	44
4.2.2 Transformada rápida de Fourier.....	45
4.2.3 Transformada Haar.....	47
4.2.4 Transformada Hadamard.....	48
4.2.5 Transformada Karhunen-Loeve.....	49
4.2.6 Transformada Slant.....	51
4.2.7 Transformada Coseno.....	52
4.2.7.1 Transformada coseno de simetría semejante.....	54
4.2.7.2 Transformada coseno de simetría impar.....	54
4.2.8 Transformada Seno.....	55
<b>5 Las Redes Neuronales y el Reconocimiento de Imágenes.....</b>	<b>57</b>
Introducción.....	57
<b>5.1 Procesamiento.....</b>	<b>60</b>
5.1.1 Técnicas de manipulación de la escala de gris.....	60
5.1.2 Técnicas para el realce de contorno.....	65
5.1.3 Modelos de R.N.A. para percepción de brillo y detección de límite.....	67
5.1.4 Técnicas para eliminar el ruido.....	74
5.1.5 Restauración de imágenes.....	77
<b>5.2 Extracción de características.....</b>	<b>88</b>
5.2.1 Extracción de características usando el momento invariante.....	90
5.2.2 Extracción de características usando transformadas ortogonales.....	94
5.2.3 Extracción de características en el dominio de la transformada de Fourier.....	95
5.2.4 Modelos de R.N.A. para la extracción de características en el dominio de la transformada de Fourier.....	98
<b>6 Aplicaciones de Redes Neuronales Artificiales Para el Reconocimiento de Imágenes.....</b>	<b>110</b>
Introducción.....	110
6.1 Percepción remota.....	110
6.2 Procesamiento de imágenes médicas.....	112
6.3 Procesamiento de huellas digitales.....	114
6.4 Reconocimiento de caracteres.....	117
6.5 Reconocimiento de rostros.....	118
<b>7 Algoritmos De Aprendizaje En Redes Neuronales Artificiales.....</b>	<b>121</b>

---

Introducción.....	121
7.1 Algoritmo de perceptrón.....	121
7.2 Algoritmo de la máquina de Boltzmann.....	123
7.3 Algoritmo de retropropagación.....	129
<b>Conclusiones.....</b>	<b>136</b>
<b>Anexo.....</b>	<b>137</b>
<b>A1 La Neurona Biológica.....</b>	<b>137</b>
<b>Bibliografía.....</b>	<b>144</b>

## *PROLOGO*

La investigación sobre Redes Neuronales Artificiales (R.N.A) en nuestro país es relativamente reciente, por este motivo y considerando que no se cuenta con mucha información sobre el tema, este trabajo de investigación hace referencia a una de las aplicaciones que se le da a estas redes, la cual es " el procesamiento de imágenes empleando redes neuronales artificiales."

Considerando lo anterior, el objetivo de este trabajo es dar a conocer la utilización de R.N.A. para el procesamiento de una imagen, complementando el tema con algunas técnicas de procesamiento de imágenes, como lo es la transformada de Fourier.

Una de las razones por la cual se procesa una imagen es porque cuando ésta se envía por algún medio de comunicación, puede distorsionarse y por ende el receptor emplea técnicas para reconstruirla y de esta forma visualizar la imagen con un mínimo de rasgos perdidos.

Para introducirnos en el tema en el Cap. 1 se da una breve explicación de lo que es una R.N.A. y para que sirve, también se menciona la analogía que tienen estas con el sistema neuronal biológico. En el Cap. 2 se mencionan los elementos con los cuales trabaja la red; como son el peso, los vectores, las conexiones, etc. En este mismo se dan a conocer dos modelos básicos de redes que son el ADALINE y MADALINE.

Siguiendo con otras bases importantes para el desarrollo de este trabajo y con la intención de ir comprendiendo el tema, en el Cap. 3 se da una introducción al procesamiento de imágenes tratando el tema de lo que es una imagen digital, tipos y características de las mismas, y los elementos que se utilizan en un sistema de procesamiento de imágenes. Seguido de esto se desarrollan algunas técnicas que se utilizan en el reconocimiento de imágenes; como lo es la transformada de Fourier y otras más, esto se describe en el Cap. 4.

Evolucionando en el tema y entrando a lo que es el procesamiento de imágenes empleando R.N.A., empezamos con las técnicas para el enfoque de las etapas en una imagen, para que la imagen pueda ser reconstruida necesita pasar por diferentes etapas de tratamiento y para esto se emplean modelos de R.N.A. las cuales se utilizan para percibir el brillo, detectar los límites de los rasgos así como para extraer sus características y eliminar el ruido de las mismas. Todos estos procesos se describen con más detalle en el Cap. 5.

En el Cap. 6 se dan algunos ejemplos de aplicaciones de R.N.A. para el reconocimiento de imágenes.

En el último capítulo se enumeran algunos algoritmos que se aplican en las R.N.A., como lo es el "Algoritmo del Perceptron" por mencionar alguno, se da una explicación teórica y el algoritmo en si.

# Capítulo 1

## INTRODUCCION A LAS REDES NEURONALES ARTIFICIALES

---

### INTRODUCCION

Las Redes Neuronales Artificiales (R.N.A.) son sistemas computacionales inspirados en el sistema biológico nervioso. El principio de operación de éstas redes ha permitido diseñar redes con propósitos específicos como el reconocimiento de patrones, ya sean estos visuales o de otros tipos, éste es radicalmente distinto a la computación hasta ahora tradicional y es muy interesante que las capacidades computacionales de estos sistemas, inspirados biológicamente, sean superiores cuando realizan funciones similares a las que llevan a cabo los sistemas nerviosos, pues es bien conocido que una computadora digital aún la más sencilla supera la velocidad y precisión del cerebro en la relación de operaciones secuenciales, mientras que las operaciones como reconocimiento de patrones, memoria asociativa y en general todas las relacionadas con el comportamiento inteligente parecen imposibles para las concepciones computacionales tradicionales.

Estas similitudes en propiedades entre las Redes Neuronales y el cerebro parecen sugerir que la emulación del comportamiento inteligente o más ambiciosamente la reproducción de "mentes" en "cerebros artificiales", se puede obtener no con algoritmos más perfeccionados sino con cambios más o menos radicales en la concepción básica de las estructuras computacionales (El desarrollo futuro de la Inteligencia Artificial dirá que tan radicales).

Existe una gran cantidad de diseños de Redes Neuronales que son capaces de emular en mayor o menor medida, una serie de funciones llevadas a cabo por el sistema nervioso, como la discriminación de patrones, o propiedades biológicas del cerebro. Pero aún con este sorprendente éxito obtenido por las Redes Neuronales, no está comprobado que esta arquitectura computacional cumpla con el test de **TURIN**<sup>1</sup> y que, por ende, sea la solución al problema de la generación de verdadera Inteligencia Artificial. Para muchos críticos desde el terreno Filosófico, las arquitecturas distribuidas y paralelas, como las Redes Neuronales, no son más que muchos procesadores tradicionales, por ende no poseen ninguna propiedad nueva que evite las objeciones del test de **TURIN**.

---

<sup>1</sup> Demostración matemática de cualquier sistema computacional sin importar su arquitectura y realice una tarea específica similar a la que realiza un sujeto.

Sin embargo una gran cantidad de grupos proporcionan evidencias de que estas redes poseen nuevas propiedades, quizá muy similares a las que tiene el cerebro, aún cuando estas sigan siendo implementadas en sistemas digitales convencionales.

Para demostrar que las Redes Neuronales pueden ser efectivamente la arquitectura capaz de pensar, y en fin, tener mente y ser el sustento de la Inteligencia Artificial, se aplica la prueba de fuego basada en el conocimiento de que no basta ser como un cerebro, en cuanto a la arquitectura se refiere, sino que hay que comportarse como tal para realmente emular sus propiedades básicas, esto significa que la red debe funcionar también en forma analógica como lo hace el sistema nervioso. Es en esta línea que se han dado pasos realmente importantes y que literalmente se han materializado en microchips de sistemas analógicos, con la arquitectura de la misma red tradicional, en este caso inspiradas en las estructuras de dos sistemas biológicos que realizan funciones computacionales, la retina y la coclea de un vertebrado generalizado.

Estos experimentos han sido realizados por el grupo de C. Mead<sup>2</sup> en Caltech.

La retina de Mead consiste en una serie de elementos analógicos neuromórficos que funcionan como fotorreceptores conectados a una red de resistores y condensadores que hacen las veces de la capa neuronal intermedia de la red, finalmente la capa de salida que es análoga a la región de células bipolares en la retina, esta formada por amplificadores que dan una respuesta proporcional a la diferencia entre la señal en la capa fotorreceptora y el promedio en la capa bipolar. De esta forma, las señales importantes en la retina artificial son las zonas de contraste luminoso mientras que las zonas con la luminosidad promedio no pesan tanto. Este es un comportamiento análogo al que ocurre en la retina de los vertebrados. La retina artificial es capaz también de hacer discriminaciones en la dirección del movimiento mediante un principio similar al que operan en la retina verdadera, es decir mediante la inhibición lateral de fotorreceptores llevada a cabo por la red R-C correspondiente a las células horizontales.

Esta implementación en silicón posee propiedades computacionales en principio idénticas a las desplegadas por su análogo biológico, es capaz de aprender y sobre todo, es analógica lo que parece imprimírle una flexibilidad no igualada por las Redes Neuronales implementadas en forma digital. Estas características podrían permitir la futura organización de sistemas inteligentes por módulos independientes con un diseño específico de funcionamiento analógico, que serían más parecidas a un sistema biológico que las computadoras digitales de propósito múltiple, aún cuando éstas pudieran estar construidas por Redes Neuronales quizás entonces quedaría demostrado que la inteligencia no es una propiedad exclusiva del cerebro, sino de cualquier máquina computacional construida con los mismos principios que el

<sup>2</sup> M.A. Mahowald C. Mead, 1991 Dhesilicom Retina, SCI.A.M.

cerebro, es decir, con una conectividad y principio de operación tipo Redes Neuronales y con una implementación analógica.

Las Redes Neuronales Artificiales son motivadas por sistemas biológicos nerviosos. Las computadoras modernas y el cómputo algorítmico son suficientes en tareas bien definidas, los sistemas biológicos nerviosos por otro lado, resuelven de manera sencilla problemas auditivos y de visión bajo amplio rango de condiciones, dichas tareas no pueden ser resueltas adecuadamente por computadoras digitales, esta diferencia ha motivado a las comunidades científicas a estudiar sistemas neuronales biológicos en un esfuerzo por diseñar sistemas computacionales con capacidades semejantes a las del cerebro. Al mismo tiempo, la tecnología de circuitos integrados modernos está ofreciendo su potencialidad para construir redes paralelas masivas en base a elementos de procesamiento simples. Así, la disciplina formal del estudio de redes neuronales, la Neurocomputación sienta las bases necesarias para programar y coordinar el comportamiento de dichos elementos de procesamiento.

Los modelos de redes neuronales están ofreciendo nuevos enfoques para resolver problemas, pueden ser simulados sobre máquinas de propósito específico, aceleradores de hardware con arquitectura neuronal, así como computadoras convencionales. Para una velocidad de procesamiento máximo pueden utilizarse arreglos en base a tecnología de procesamiento óptico o en VLSI de silicio. El aspecto fundamental reside en utilizar el modelo neuronal de procesamiento paralelo en tiempo real con un número masivo de procesadores sencillos.

### ***1.1. Qué es una Red Neuronal Artificial***

Las R.N.A. son sistemas basados en una representación simplificada de estructura y funcionamiento del sistema nervioso ya sea simulado en software o construido en hardware. Se entrenan mediante ejemplos conocidos hasta que son capaces de asociar patrones de entrada con respuestas definidas sin necesidad de una programación explícita para un problema particular. Esta característica permite a las redes resolver problemas para los cuales no es posible diseñar algoritmos de solución, pudiendo entonces enfrentar con éxito aquellos casos en los que no es posible definir una relación entre las causas y los efectos.

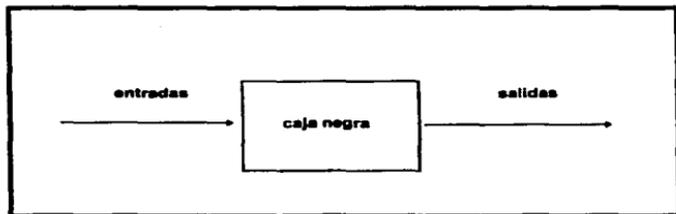
Las redes neuronales son modelos de cómputo paralelo los cuales son utilizados en Ingeniería con el nombre de Redes Neuronales Artificiales, teniendo una alternativa de procedimientos a problemas cuya solución por técnicas tradicionales resulta en la mayoría de los casos de difícil aplicación.

Dando un significado más sencillo, puede decirse que las R.N.A. son como una **caja negra**, aparato que acepta entradas y produce salidas, como se ilustra en la fig. 1.1

Los rasgos más relevantes de las características del medio de aplicación de Redes Neuronales Artificiales son:

**Uso de Redes Neuronales.**

- La aplicación no permite generar reglas explícitas
- El dominio es dinámico
- Se requiere de juicio
- El tiempo de cómputo es importante
- La aplicación es altamente interactiva
- La aplicación tiene restricciones múltiples



**Fig. 1.1 Red neuronal artificial**

Adicionalmente, las Redes Neuronales Artificiales poseen características únicas que las diferencian de otras técnicas, por ejemplo:

- Aprendizaje o entrenamiento a través de ejemplos
- Tolerancia a fallas en la operación de algunas unidades
- Capacidad para sintetizar

**1.2 Funcionamiento de la red**

El principio de funcionamiento de las R.N.A es, en términos generales el de un convertidor vectorial, la red es controlada con la información codificada en forma de vector de entrada y después de un número determinado de intentos la red produce un vector de salida que es la mejor solución que encuentra en la conversión del vector de entrada. Así las respuestas distintas de la red dependen de la arquitectura de ésta, de sus propiedades de conectividad y de las reglas algorítmicas explícitas en el paso

de información de una capa de neuronas a las siguientes. En cuanto a la mejora de las respuestas generadas por la red, depende del entrenamiento de esta mediante el cual cambia, en función de un algoritmo de retroalimentación la forma en que pasa la información entre las neuronas individuales y capas de neuronas.

De esta forma la red realiza funciones de tipo cerebral, en tiempos razonables aunque no en tiempo real. Sin embargo la forma de implementación sigue siendo similar a la de la Inteligencia Artificial clásica, pues todas estas redes se instalan en computadoras algorítmicas y son alimentadas por los entrenadores de la red.

### **1.3 Similitudes entre una Red Neuronal Artificial y una Neurona Biológica**

Para comprender mejor lo que son las Redes Neuronales Artificiales es necesario hacer una breve referencia de la relación y similitudes que tienen estas, con las neuronas biológicas.

El funcionamiento de las Redes Neuronales se basa en el mecanismo que lleva a cabo el sistema nervioso y en especial nos referimos a la neurona, por medio de la cual se conduce la información del cerebro a todo el cuerpo. En general las criaturas vivientes tienen receptores que utilizan para ver, oír y sentir. Lo que interesa principalmente a los biónicos, es el hecho de como estos fenómenos físico - químicos son utilizados para razonar o pensar.

Con algunas respuestas a estas interrogantes, se construyen circuitos electrónicos, que pueden "razonar" y "aprender" tal como lo hacen los seres vivos. Los secretos están basados en la celda nerviosa básica (la neurona). Estos pequeños bloques o estructuras de todos los sistemas vivientes, tanto del cerebro como del sistema nervioso que los científicos ahora conocen, están basados en un sistema de interruptores.

Como es sabido la neurona en el sistema nervioso, se compone de un cuerpo celular y sus prolongaciones (dendritas y axón). Esto es, una neurona tiene muchas entradas (quizá varios cientos) y una salida. Las dendritas en la neurona vienen siendo los receptores de información que viene de otra neurona, dicha información llega al cuerpo celular, de ahí se envía por medio del axón hacia otra neurona. A este proceso de comunicación entre neuronas se le llama sinápsis ( ver apéndice " A1 ").

La transmisión en las neuronas depende de los cierres o abiertos en que se encuentra la neurona en ese momento, esto es, enviar información o no de una neurona a otra. La entrada de una red neuronal biológica esta compuesta por un conjunto de sensores que desempeñan la función de receptores, los cuales proporcionan los estímulos provenientes del exterior al interior de la red; dichos estímulos están representados por impulsos eléctricos que se encargan de transmitir

la información hacia el sistema nervioso central, provocando así, efectos de respuesta humana, la cual puede estar expresada en una gran variedad de acciones. De igual forma que en la neurona, en los circuitos electrónicos existen bajo o alto (0 ó 1), las entradas a los circuitos son acumulados y posteriormente procesados por una función de Activación la cual determina la respuesta de la neurona. Para la función de Activación se utiliza comúnmente una función matemática que varía en forma continua con perfil "sigmoideal". La salida de la neurona o rango dinámico, cambia entre los límites de 0 y 1.

Tanto en la neurona como en los circuitos electrónicos la comunicación se da por medio de un pulso eléctrico. En la célula nerviosa podemos medir el voltaje de la membrana, que es cercanamente igual a 0.1 volt, es decir 100 milivoltios. De igual forma en los alambres conductores también se puede medir el voltaje cuando circula la corriente eléctrica. La transmisión del impulso nervioso se propaga en la célula a una velocidad que oscila entre 27 y 132 metros por segundo. Así también las corrientes en los alambres conductores pueden ser propagados (por efectos de la acción de los campos) a una velocidad cercana a la de la luz (300,000 km./seg.).

#### **1.4 Aplicaciones de Redes Neuronales Artificiales**

Las aplicaciones que se derivan de la computación son abundantes en el campo de procesamiento de señales o reconocimiento de patrones, extracción de características, inspección industrial, pronóstico de negocios, clasificación de crédito, selección de seguridad, diagnóstico médico, procesamiento de voz, entendimiento del lenguaje natural, control de robots y adaptación de procesos de control. Es mucho más individual en el rango de la computación tecnológicamente hablando el filtro adaptivo, o también llamado "**neurona adaptiva**" que cambia las características de la respuesta, combinando los patrones de ruido de la señal de entrada. Casi toda la computación moderna, rápida esta confiada en dicho filtro adaptivo.

El reconocimiento de patrones es el corazón de muchas aplicaciones de redes neuronales. En 1987 una industria bancaria usó un sistema de red neuronal y actualmente toda la industria bancaria usa redes neuronales. Acordado por Banc Tec (Dallas Texas), para leer los números de cheques y cuando es deslizada una tarjeta de crédito. Solamente el año pasado sinápticos de (San José California) anunciaron el primer microchip basado en redes neuronales para leer los números de cheques y los números de las tarjetas de crédito cuando sean deslizadas.

La retina electrónica que es la que se utiliza para leer el número de la tarjeta al ser deslizadas y los números de los cheques, es una implementación análoga en paralelo de la emulación de una red neuronal de un **Ojo óptico**.

# Capítulo 2

## COMPONENTES Y MODELOS DE UNA RED NEURONAL ARTIFICIAL

### INTRODUCCION

Las R.N.A. consisten de elementos de procesamiento y conexión de pesos. La cantidad de estos depende de la implementación y diseño de la red, otros elementos que intervienen dependiendo de la aplicación, son las funciones preestablecidas de los vectores de entrada y los vectores de salida.

### 2.1 COMPONENTES DE UNA RED NEURONAL ARTIFICIAL

#### 2.1.1 Elementos de entrada y elementos de salida

Acercas de los componentes de una red neuronal artificial es conveniente dar a conocer la terminología que se utilizará en la descripción posterior de cada elemento.

Para ilustrar esta terminología, considere la figura 2.1

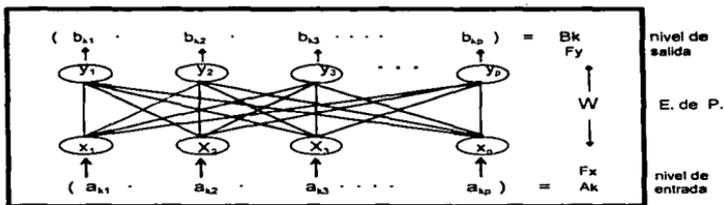


Fig. 2.1. Estructura de una red neuronal artificial de 2 niveles

En la figura 2.1 se observa la estructura de una red neuronal artificial, la cual consta de dos niveles (nivel de entrada y nivel de salida). Dentro de cada nivel se

encuentran los elementos de procesamiento que en conjunto forman vectores de entrada y salida, y cuya función es la de manejar la información proveniente del nivel de entrada y enviar una respuesta hacia la salida.

La nomenclatura para describir a los vectores de entrada y salida se denotará con las primeras letras del alfabeto ( por ejemplo **A, B, C**, etc. ) acompañadas de un subíndice, el cual servirá para indicar la ubicación del vector dentro de la red neuronal. De esta manera, los vectores de entrada se representarán por la siguiente ecuación:

$$A_k = (a_{k1}, a_{k2}, a_{k3}, \dots, a_{kn}) \quad \text{donde } k = 1, 2, 3, \dots, m$$

y los vectores de salida por la ecuación :

$$B_k = (b_{k1}, b_{k2}, b_{k3}, \dots, b_{kp}) \quad \text{donde } k = 1, 2, 3, \dots, m$$

Los Elementos de Procesamiento que se encuentran en un nivel cualquiera, se representan con la misma variable que le fue asignada a dicho nivel. Por su parte, al conjunto de elementos de procesamiento de un nivel se le conoce como **Vector de Elementos de Procesamiento (PE's)** al que se le interpreta con las últimas letras del alfabeto ( por ejemplo **T, U, V**, etc ), aunque en muchos de los casos, el utilizar 3 niveles de elementos de procesamiento, son los suficientes para formar una red neuronal artificial aceptable. De esta manera, los elementos de procesamiento que se encuentren localizados en el nivel de entrada se representarán mediante la siguiente ecuación:

$$F_x = (X_1, X_2, \dots, X_n)$$

De donde cada elemento  $x_i$  recibe información de su correspondiente entrada que se denotará como  $a_{ki}$ . El siguiente nivel de elementos de procesamiento será indicado como  $F_y$ , y así sucesivamente. Volviendo a la figura 2.1, se observa que el segundo nivel de la red es considerado como el nivel de salida, que se representa con la siguiente ecuación:

$$F_y = (y_1, y_2, \dots, y_n)$$

De donde cada elemento  $y_j$  esta correlacionado con el  $j$ -ésimo elemento de  $B_k$ .

### **2.1.2 El Peso en las Redes Neuronales Artificiales**

Los pesos conectados entre los niveles de la red neuronal, sirven para modificar los valores que circulan en las conexiones, son almacenados en un arreglo matricial,

denominado como **Matriz de Peso**, la cual puede estar representada por cualquier letra intermedia del alfabeto ( **T, U, V**, etc ).

Para el caso de la figura 2.1, que muestra una red neuronal de dos niveles, tenemos que su respectiva matriz de pesos es como la que se ilustra en la figura 2.2, que sirve para representar las correspondientes conexiones entre los elementos de procesamiento del nivel de entrada señalado como  $F_x$  y los del nivel  $F_y$  ( nivel de salida ). De acuerdo con la matriz de pesos de la figura 2.2, el valor de  $W_{11}$ , significa la conexión de los pesos de los elementos de procesamiento de  $x_1$  con el de  $y_1$ .

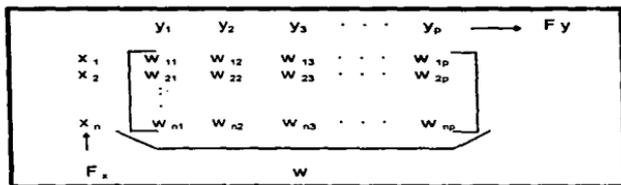


Fig. 2.2.- Matriz de peso de una red neuronal de dos niveles.

### 2.1.3 Vectores de Entrada, Vectores Ocultos y Vectores de Salida

Como se había analizado anteriormente, las redes neuronales artificiales están compuestas por un gran número de unidades interconectadas, de manera que estas a su vez forman Vectores, que dependiendo de su ubicación o nivel en que se encuentren dentro de la red neuronal, se les denominan como **Vectores de Entrada, Vectores Ocultos y Vectores de Salida**.

Dependiendo del número de niveles que conformen a la red neuronal, tendrá esta una cantidad proporcional de niveles ocultos, y a su vez un número determinado de vectores de salida así como los vectores de entrada.

Como se puede observar en la figura 2.3, el proceso de sinápsis se representa mediante un peso que puede ser modificado, el cual es asociado con cada una de las conexiones.

La mayoría de los modelos de redes neuronales artificiales presentan dos desventajas importantes y básicas: por un lado, no tienen la capacidad de proporcionar la geometría más acertada del funcionamiento de las dendritas y los axones, y por el otro, el de asemejar o interpretar las salida eléctrica, como la que se

produce en una neurona biológica, como un simple valor numérico que representa la magnitud de disparo.

De la figura 2.3 se puede observar que cada unidad transforma una actividad proveniente del exterior en un vector de entrada, siendo éste a su vez transmitido a otras unidades; este proceso se lleva a cabo durante dos fases: Primero, el vector de entrada es multiplicado por un factor de peso colocado sobre su conector, dando origen a un valor denominado **Entrada Resultante**, posteriormente se suma simultáneamente con las otras entradas, formándose así, un valor denominado **Entrada Total Resultante**. Segundo, una unidad hace uso de una **Función de Entrada-Salida** con el fin de transformar el valor de entrada total resultante a un valor de salida, llamado **Actividad de Salida**.

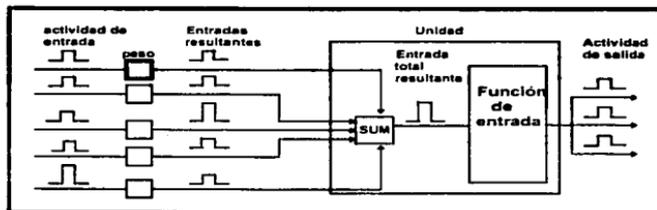


Fig. 2.3.- Proceso de sinapsis en una red neuronal artificial

La función Entrada-Salida y los pesos que se manejan en una red neuronal en perfecta combinación, determinan el desempeño que logre la red. Por su parte, la función Entrada-Salida puede encontrarse en cualquiera de las siguientes tres categorías: **Lineal**, **Umbral** o **Sigmoidal (en función de Sigma)**. Las cuales semejan el comportamiento de una neurona biológica.

En caso de que la función se encuentre en la categoría de unidades lineales, tendríamos que la actividad de salida de la red neuronal será proporcional al total de las entradas. Para el segundo caso, el de las unidades umbrales, la salida únicamente puede tomar uno de dos niveles, el que será determinado dependiendo si el valor de la entrada total resultante sea mayor o menor que alguno de los valores preestablecidos. Por último, para el caso de las unidades de tipo sigmoidal, el valor de salida, constantemente se encuentra en estado de variación, y esto trae como consecuencia ciertos cambios en la entrada; los casos anteriores se muestran en la figura 2.4.

De las tres categorías anteriores, la que se considera más próxima al comportamiento de las neuronas biológicas, es la de las unidades sigmoidales, aunque las tres son consideradas únicamente como aproximaciones al comportamiento real.

La clase de red neuronal artificial que con mayor frecuencia se diseña y se construye es la que se estructura de tres grupos o niveles, que son interconectados de la siguiente manera:

El nivel que contiene a las unidades de entrada se conecta al nivel de las unidades ocultas, siendo éste a su vez conectado con el nivel formado con las unidades de salida. Por su parte, el funcionamiento de las unidades de entrada consiste en suministrar la información del exterior o de realimentación más reciente. En tanto la función de las unidades ocultas se va componiendo de acuerdo a las actividades que realizan las unidades de entrada en combinación con los pesos que han sido colocados sobre las conexiones que se encuentran entre las unidades de entrada y las unidades ocultas.

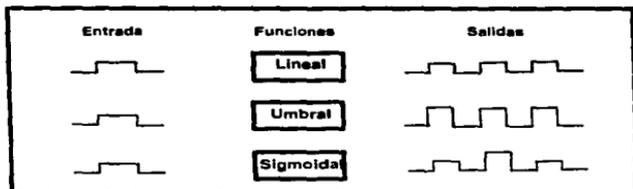


Fig. 2.4.- Función Entrada-Salida en sus tres categorías: Lineal, Umbral y Sigmoidal

De una forma muy similar, el comportamiento de las unidades de salida se deriva tanto de las actividades de las unidades ocultas como de los pesos colocados sobre las conexiones entre ambas unidades. Este tipo de red además de ser uno de los más sencillos, es uno de los más interesantes, debido a que las unidades ocultas tienen la capacidad de construir sus propias interpretaciones acerca de las actividades de entrada, de esta manera, mientras los pesos sufran ciertas modificaciones, darán otras interpretaciones diferentes sobre las actividades de entrada.

Una situación muy interesante, es que no todas las redes neuronales artificiales operan con algunos tipos comunes de vectores, es decir algunos sistemas neuronales requieren para su operación patrones simples (conjunto de valores, con el cual se compara la entrada), a estos sistemas se les conoce con el nombre de Red Neuronal Artificial Autoasociativa; pero también existen los sistemas denominados Redes Neuronales Artificiales Heteroasociativas, debido a que utilizan para su funcionamiento patrones pares (dos conjuntos de valores con los cuales se compara la entrada).

Para las redes del tipo autoasociativa tenemos a un conjunto de patrones que se encuentran previamente almacenadas en la red, los cuales son comparados con el patrón de entrada y si alguno es similar, entonces ambos patrones se asocian, en la figura 2.5 se muestra este tipo de red, en la cual, un cuadrado incompleto, es asociado con el cuadrado que se encuentra en el interior de la red.



Fig. 2.5.- Red neuronal artificial autoasociativa.

En las redes heteroasociativas, los patrones que se encuentran en su interior, están agrupados en pares, y de igual manera que en la red autoasociativa, se compara la entrada con los patrones almacenados, asociándole el par (romboide) que le fue asignado al patrón que se le asemeja (cuadrado). Esto se observa en la figura 2.6.

Debido al tipo de procesamiento que realice la red neuronal, existen básicamente dos tipos de redes que son: las que únicamente tienen la capacidad de procesar datos en código binario; este tipo de máquinas se pueden ver tanto en el modelo de la **Red Neuronal Artificial de Hopfield**<sup>1</sup> perfeccionado por Amari en 1972, como en el modelo llamado **El cerebro como una Simple Caja de Anderson**<sup>2</sup>.

<sup>1</sup> J.J Hopfield, "Neural Network and Physical Systems with Emergent Collective Computational Abilities", Proc. Natl. Acad. Sci., vol. 79 Apr., 1982

Por el contrario, otros modelos de redes neuronales artificiales, tienen la característica de procesar datos con valores reales, como es el caso del **Método de Realimentación hacia Atrás de Williams**<sup>3</sup>, y del **Vector de Cuantización para Aprendizaje de Kohonen**<sup>4</sup>, cuyo propósito es el de producir un código obtenido de un patrón de entrada de  $n$  dimensiones, el cual es transmitido a través de un canal y utilizado para reconstruir la entrada original con un mínimo de distorsión.

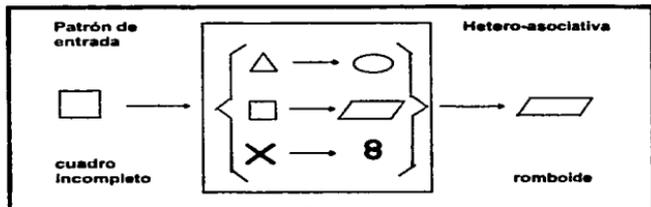


Fig. 2.6.- Red neuronal artificial heteroasociativa.

#### 2.1.4 Conexiones

El Propósito de las conexiones dentro de una gráfica dirigida es la de determinar la dirección del flujo de la información, como se muestra en la figura 2.7. El flujo de la información que se realiza del nivel  $F_x$  a el nivel  $F_y$  se hacen mediante las conexiones representadas con la letra **W**. En los digrafos de las redes neuronales se les asigna un peso a cada trazo (conexión), que representa el valor total de la señal de salida que se trasladará de un nodo (elemento de procesamiento) inferior a un nodo adyacente.

Como se aprecia en la figura 2.7, se utiliza una flecha doble para indicar el sentido del flujo de la información. Una conexión tiene fundamentalmente dos

<sup>3</sup> Anderson, J. J. Silverstein, S ritz, and R. Jones. "Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model". Psych. Rev., vol. 84, 1977.

<sup>4</sup> Jacek M. Zurada "Introduction to Artificial Neural System". West Publishing Company. USA, 1992.

<sup>4</sup> Kohonen, T. "Learning Vector Quantization for Pattern Recognition". Helsinki University of Technology, 1986.

funciones, por un lado la de indicar la dirección del flujo de la información, y por el otro, de modular la cantidad de información que fluye entre los elementos de procesamiento.

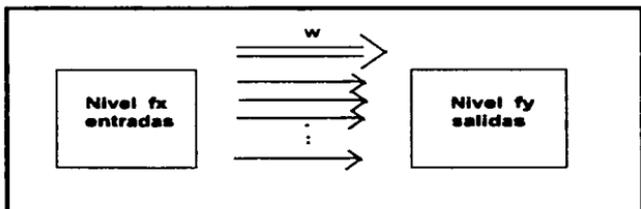


Fig. 2.7.- Conjunto de conexiones de una red neuronal artificial.

Como se aprecia en la figura 2.7, se utiliza una flecha doble para indicar el sentido del flujo de la información. Una conexión tiene fundamentalmente dos funciones, por un lado la de indicar la dirección del flujo de la información, y por el otro, de modular la cantidad de información que fluye entre los elementos de procesamiento.

En tanto, las conexiones denominadas excitatorias son las que contienen pesos con valores positivos, y por el contrario, cuando estos pesos tienen un valor negativo se les denomina como conexiones inhibitorias. Cuando se presenta el caso de que no existe conexión alguna se debe a que el peso de esta conexión presenta un valor de cero o nulo.

Para el diseño de redes neuronales artificiales no es conveniente el uso continuo de conexiones que presenten el valor cero, a excepción de que estas conexiones se encuentren esparcidamente y en números pequeños. Durante el funcionamiento de una red neuronal artificial, una de las condiciones más deseables es la de lograr que los elementos de procesamiento mantengan valores parciales internos (valores preestablecidos).

En la figura 2.8 en donde se observa que el elemento de procesamiento que es señalado como  $y$  está en función de tres conexiones representadas por la siguiente ecuación:  $F_x ( W_1, W_2, W_3 )$  y por un valor preestablecido representado por  $0_j$ .

Como se observa en la figura 2.9, para el mejor desempeño de la red es importante que el valor preestablecido  $\theta_j$  sea utilizado como una conexión extra, marcada como  $w_0$  y proveniente de la expresión  $F_x$  PE y  $x_0$ , considerando que el valor de  $x_0$  siempre tomará el valor de uno.

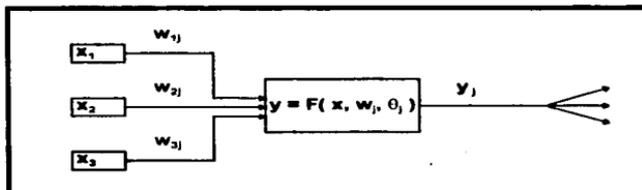


Fig. 2.8.- Elemento de Procesamiento con tres conexiones.

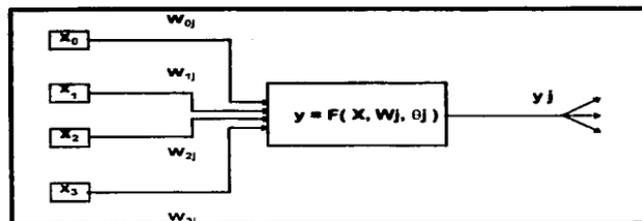


Fig. 2.9.- Elemento de Procesamiento con conexión extra ( $w_0$ ).

### 2.1.5 Elementos de Procesamiento (P.E.)

Los elementos de procesamiento son los elementos más importantes y básicos que componen a una red neuronal artificial. Esta importancia se debe a que en el interior de estos elementos se ponen en funcionamiento los componentes restantes de la red, para realizar las operaciones que nos lleven a los resultados esperados. Uno de los elementos de procesamiento que se usa con mayor frecuencia se muestra en la figura 2.9.

Además, los elementos de procesamiento pueden encontrarse en dos situaciones diferentes: la primera de ellas, se presenta cuando un elemento de procesamiento se encuentra en el nivel de entrada y por consiguiente recibirá un solo patrón de entrada, como se observa en la figura 2.10(a). La segunda ocurre cuando cada elemento de procesamiento que se encuentre en un nivel que no sea el de entrada, tiene varias conexiones de pesos provenientes de otros elementos de procesamiento de diferente nivel, como se ilustra en la figura 2.10(b).

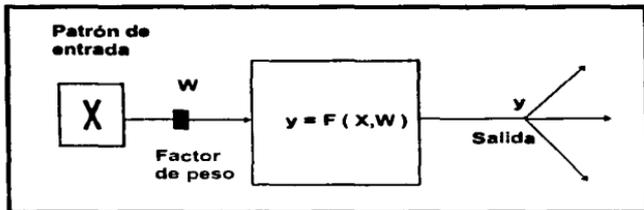


Fig. 2.10(a).- Elemento de Procesamiento con una sola conexión de peso.

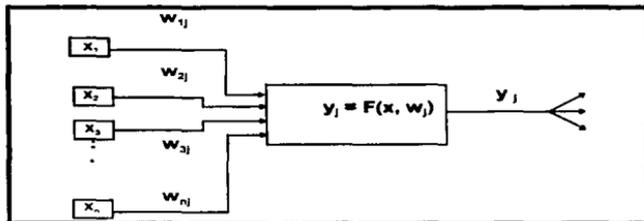


Fig. 2.10(b).- Elemento de Procesamiento con varias conexiones de peso.

Como consecuencia de las situaciones anteriores, se tiene que los elementos de procesamiento tienen dos cualidades importantes:

1. Los elementos de procesamiento necesitan únicamente de información local ubicadas en las entradas de estos, para posteriormente almacenarlas en el interior de dicho elemento y finalmente producir un valor de salida. Una ventaja muy importante de esta característica es que ya no se requiere de una información extra para un mejor desempeño.
2. Durante su funcionamiento, los elementos de procesamiento utilizan los valores de entrada para procesarlos y producir un sólo valor de salida, transmitiéndose mediante las conexiones hacia un elemento de procesamiento receptor o bien puede ser utilizado como valor de salida de la red.

Como se hizo anteriormente con las conexiones, el valor que se le asigna a cada elemento de procesamiento así como a la etiqueta que lo representará, serán denotados con una misma variable, por ejemplo, para representar el  $j$ -ésimo elemento de procesamiento que se localiza en el nivel  $F_y$  se utiliza la expresión  $y_j$ , y por consiguiente el valor que tenga este elemento será también de la forma  $y_j$ .

El valor de salida de un elemento de procesamiento esta en función de las entradas y los pesos que llegan a él, como se puede observar claramente en la figura 2.9. en donde  $y_j$  representa la función de las salidas producidas en un nivel precedente denotada por la ecuación  $F_x = X = (x_1, x_2, \dots, x_n)$  y de los pesos localizados desde el nivel  $F_x$  hasta el elemento  $y_j$ , es decir,  $W = (w_1, w_2, \dots, w_{nj})$ ; de acuerdo a lo anterior, la ecuación que representa al elemento  $y_j$  es:

$$y_j = F(X, W_j) \quad (2.1)$$

Existen básicamente dos tipos de funciones que se utilizan para el diseño de los elementos de procesamiento, de acuerdo a la aplicación que se les vaya a asignar a estos, los cuales son las siguientes:

**Función por Combinación Lineal.-** De los métodos utilizados para el diseño de los elementos de procesamiento el más común es el conocido como **Combinación Lineal** llamado también **Producto Punto**. Este método consiste en realizar el producto punto con los valores de entrada  $X$  y los pesos de conexión  $W_j$ , para ilustrar este método utilizaremos el elemento de procesamiento mostrado en la figura 2.9 en donde la salida  $y$  está expresada por la siguiente ecuación:

$$y_j = f \left[ \sum_{i=0}^n x_i w_{ij} \right] = f(x, w_j) \quad (2.2)$$

De la ecuación 2.2 tenemos que  $W_j = (w_{1j}, w_{2j}, \dots, w_{nj})$ , y  $f$  es una de las funciones no lineal del elemento de procesamiento. Una de las características más importantes del producto punto es que este es muy usado en la computación.

**Función con Conexiones Dobles (Media Varianza).**- En algunas ocasiones, en la representación de un elemento de procesamiento se utilizan dos conexiones interconectadas a dicho elemento como se aprecia en la figura 2.11. Esta doble conexión tiene dos funciones, por un lado la de representar el valor de la clase, y por el otro, la varianza que existe en dicha clase<sup>1</sup>, entendiéndose por varianza, el valor medio de las desviaciones cuadráticas de cada una de las variables.

Para este tipo de conexiones, tanto el valor de las entradas como el valor de las conexiones determinan el valor de la salida del elemento de procesamiento, es decir,  $y_j = F(X, V_j, W_j)$ , de donde las conexiones medias se representan con la expresión  $W_j = (w_{1j}, w_{2j}, \dots, w_{nj})$  y las conexiones que llevan el valor de la varianza se denota con la expresión  $V_j = (v_{1j}, v_{2j}, \dots, v_{nj})$ . El valor de la salida  $y_j$  se determina calculando la diferencia entre el valor de la entrada  $X$  y el valor de peso  $W_j$ , dividirla entre la varianza  $V_j$  y elevar el resultado al cuadrado, esto se hace para transformar el valor resultante mediante la función Gaussiana no lineal del elemento de procesamiento, y obtener finalmente el valor de salida representado como:

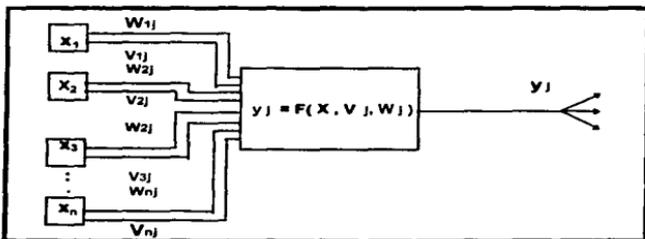


Fig. 2.11.- Elemento de procesamiento con conexiones dobles (medio varianza).

Para este tipo de conexiones, tanto el valor de las entradas como el valor de las conexiones determinan el valor de la salida del elemento de procesamiento, es decir,

<sup>1</sup>Robinson, A. M., Niranjana, and F. Fallside. "Generalizing the nodes of the Error Propagation Network". Cambridge University Engineering Department, 1988.

$y_j = F(X, V_j, W_j)$ , de donde las conexiones medias se representan con la expresión  $W_j = (w_{1j}, w_{2j}, \dots, w_{nj})$  y las conexiones que llevan el valor de la varianza se denota con la expresión  $V_j = (v_{1j}, v_{2j}, \dots, v_{nj})$ . El valor de la salida  $y_j$  se determina calculando la diferencia entre el valor de la entrada  $X$  y el valor de peso  $W_j$ , dividirla entre la varianza  $V_j$  y elevar el resultado al cuadrado, esto se hace para transformar el valor resultante mediante la función Gaussiana no lineal del elemento de procesamiento, y obtener finalmente el valor de salida representado como:

$$y_j = g \left[ \sum_{i=1}^n \frac{(w_{ij} - x_i)^2}{v_{ij}} \right] \quad (2.3)$$

De la ecuación anterior se obtiene que la función no lineal Gaussiana del elemento está dada por:

$$g(x) = \exp. \left( \frac{-x^2}{2} \right) \quad (2.4)$$

## 2.2 MODELOS BASICOS DE UNA RED NEURONAL ARTIFICIAL

Los métodos de procesamientos de señales tuvieron su desarrollo dentro de la rama de la ingeniería con la llegada de una nueva rama en la electrónica, la comunicación. En sus inicios para realizar este tipo de comunicación se requería de filtros analógicos, los cuales consistían de circuitos RLC (resistor-inductor-capacitor) que fueron diseñados con el propósito de eliminar el ruido que existía en las señales de comunicación. Hoy en día, los sistemas de procesamientos de señales junto con sus avances, se han involucrado dentro de una tecnología multifacética, haciendo énfasis principalmente en la evolución que hubo desde la implementación de los primeros circuitos sintonizadores (RLC) hasta la llegada de los procesadores de señales digitales (DSP's).

Las bases para la comunicación residen en el diseño e implementación de filtros, cuya función es la de eliminar el ruido existente en las señales de información generadas y/o recibidas.

Para el desarrollo de los procesadores de señales digitales se pensó en varios modelos de redes neuronales artificiales, de los cuales se tomó de acuerdo a sus características el primer perceptrón, que consistía de dos niveles, por tanto las investigaciones no se detuvieron allí y optaron por el diseño de un perceptrón más

preciso, el cual consistió de múltiples niveles. Dentro de este conjunto de perceptrones, Bernard Widrow y un grupo de investigadores tomaron principalmente dos modelos para su análisis y desarrollo, a los cuales denominaron perceptrones **ADALINE** y **MADALINE**. El Adaline (ADaptive LInear Element "Elemento Lineal Adaptivo") también es conocido como **Unidad Umbral Lineal** y constituye el bloque básico de la estructura de una red Madaline (Many Adaline "Muchas Adaline"). De acuerdo al modelo Adaline, su estructura se constituye de una sola neurona, en tanto el modelo Madaline posee cualquier número mayor de una neurona.

### 2.2.1 Modelo Adaline

El bloque básico de construcción que frecuentemente es usado para el diseño de las redes neuronales artificiales, es el Elemento Lineal Adaptivo mejor conocido como **ADALINE**, que en principio fue llamado Neurona Lineal Adaptiva, como se observa en la figura 2.12

Muchos de los investigadores coinciden en que la estructura del **ADALINE** es semejante a la descripción general de un elemento de procesamiento, para esto, existen dos modificaciones que se requieren para hacer que la estructura de dicho elemento de procesamiento cumpla con las características de una red **ADALINE**, las cuales son:

Adicionar el conjunto de conexiones que llegan al combinador lineal adaptivo, una conexión más, con un peso marcado como  $w$ , el cual siempre tendrá un valor de entrada igual a uno.

Esta consiste, en la colocación de una condición bipolar sobre la salida; dicha condición, funciona de la siguiente forma: si la salida del combinador lineal adaptivo es positivo, la salida de la red tendrá un valor igual a +1, si por el contrario, la salida del combinador lineal adaptivo es negativo, la salida final de la red será -1.

Para comprender mejor el funcionamiento de una red neuronal de tipo **ADALINE**, tomaremos como referencia la red neuronal artificial de la figura 2.12 la cual es capaz de implementar ciertas funciones lógicas.

En ella tenemos 2 posibles patrones de entrada; para esto se pretende que una implementación lógica general sea capaz de clasificar cada patrón, cualquiera que sea su valor, +1 o -1, de acuerdo a la respuesta deseada. De esta manera tenemos 2 posibles funciones lógicas que conectan las  $n$  entradas a una sola salida binaria. Debido a lo anterior un subconjunto de estas funciones, las cuales son conocidas como **Funciones Lógicas Separables Linealmente** o **Funciones Lógicas**

**Prestablecidas<sup>2</sup>**, que pueden ser obtenidas mediante las variaciones posibles sobre los pesos.

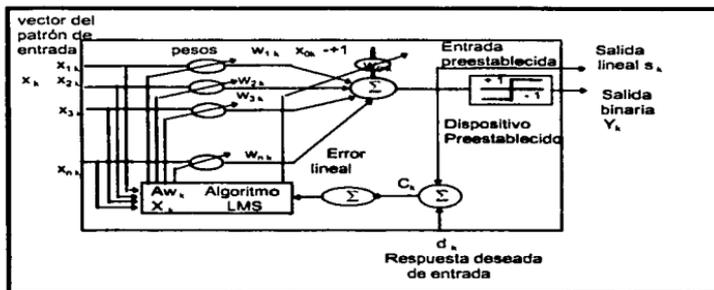


Fig. 2.12 Red neuronal artificial con arquitectura ADALINE

Estas características se aprecian en la figura 2.13

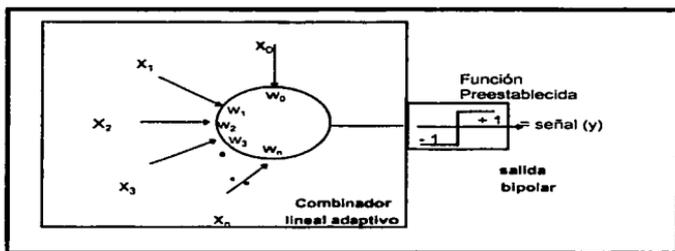


Fig. 2.13 Elemento de procesamiento modificado para funcionar como una estructura ADALINE.

<sup>2</sup>P.M. Lewis II and C. Coates. "Threshold Logic". New York: Wiley, 1967.

Para comprender mejor el funcionamiento de una red neuronal de tipo **ADALINE**, tomaremos como referencia la red neuronal artificial de la figura 2.12 la cual es capaz de implementar ciertas funciones lógicas.

En ella tenemos 2 posibles patrones de entrada; para esto se pretende que una implementación lógica general sea capaz de clasificar cada patrón, cualquiera que sea su valor, +1 o -1, de acuerdo a la respuesta deseada. De esta manera tenemos 2 posibles funciones lógicas que conectan las  $n$  entradas a una sola salida binaria. Debido a lo anterior un subconjunto de estas funciones, las cuales son conocidas como **Funciones Lógicas Separables Linealmente o Funciones Lógicas Preestablecidas**<sup>3</sup>, que pueden ser obtenidas mediante las variaciones posibles sobre los pesos.

Por otro lado, tenemos un caso particular, en el cual el elemento llamado **ADALINE** posee únicamente dos entradas, como se observa en la figura 2.14. Así también, en la figura 2.15 podemos apreciar todas las posibles entradas binarias a dicho elemento, representadas por cuatro puntos extensos, localizados dentro del espacio correspondiente al vector del patrón de entrada. En este espacio, los componentes de dicho vector se encuentran situados a lo largo de los ejes de coordenadas.

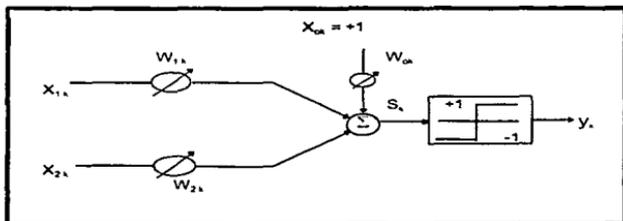


Fig. 2.14 Elemento ADALINE con dos entradas

De acuerdo a las figuras anteriores, durante su funcionamiento, el elemento **ADALINE** lleva a cabo una tarea muy importante, que consiste en separar a los patrones de entrada en dos categorías que dependen de los valores que tengan los pesos. Para esto, tenemos que una condición preestablecida en estado crítico se presenta cuando el valor de la salida lineal es igual a cero, es decir:

<sup>3</sup>P.M. Lewis II and C. Coates. "Threshold Logic". New York: Wiley, 1967.

$$s = x_1 w_1 + x_2 w_2 + w_0 = 0$$

de la ecuación anterior tenemos que:

$$x_2 = - \frac{w_1}{w_2} x_1 - \frac{w_0}{w_2}$$

siendo esta una relación lineal, apreciada en la figura 2.15, la cual contempla tanto una línea inclinada de separación como una intersección, representadas ambas por:

$$\text{pendiente} = - \frac{w_1}{w_2}$$

$$\text{intersección} = - \frac{w_0}{w_2}$$

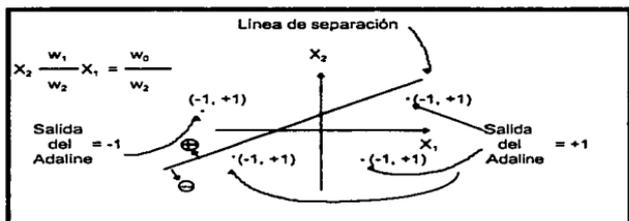


Fig. 2.15 Número posible de entradas binarias para el elemento ADLINE de dos entradas.

Los pesos marcados como  $w_0$ ,  $w_1$  y  $w_2$  son los que determinan la pendiente, la intersección y el lado de la línea de separación que corresponde a la salida positiva, en tanto en el lado opuesto, la línea de separación corresponde a la salida negativa.

De acuerdo a las expresiones anteriores, decimos que en las redes **ADALINE** que contemplan en su estructura cuatro pesos, su límite de separación será siempre un plano, no siendo así para las **ADALINE** que contemplan más de cuatro pesos, ya que en este caso su límite de separación será un hiperplano. Como lo señalamos

anteriormente, cuando el peso adicional marcado como  $w_0$  toma el valor de cero, tenemos que la separación hiperplana será homogénea, es decir, pasará dentro del espacio del vector patrón y por el origen.

De la figura 2.15 obtenemos que los patrones binarios de entrada están clasificados de la siguiente manera:

(+1, +1)	—————>	+1
(+1, -1)	—————>	+1
(-1, -1)	—————>	+1
(-1, +1)	—————>	+1

De acuerdo a los resultados anteriores, observamos que la red **ADALINE** cumple con la función de separabilidad lineal. Por el contrario, si no cumpliera con esta función, estaríamos hablando de que lleva a cabo una función NOR exclusiva con dos entradas, y en la cual los patrones binarios de entrada estarían clasificados de la siguiente forma.:

(+1, +1)	—————>	+1
(+1, -1)	—————>	-1
(-1, -1)	—————>	+1
(-1, +1)	—————>	-1

De esta manera, observamos que una red **ADALINE** de estructura simple sin una fase de preprocesamiento, no tiene la capacidad de implementar una función del tipo NOR exclusiva.

En tanto una red **ADALINE** sencilla pero con dos entradas, es capaz de realizar 14 de las 16 posibles funciones lógicas, que si por el contrario, la red posee un número mayor de entradas, únicamente una pequeña fracción de las funciones lógicas se podrán realizar, pero siempre siendo éstas linealmente separables.

Finalmente tenemos que las combinaciones de los elementos o también llamados redes de elementos, tienen su mayor aplicación en las redes cuyas funciones no son linealmente separables.

### 2.2.2 Modelo Madaline

Como lo analizamos en el tema anterior, uno de los problemas en el desarrollo de las redes, se presenta en las redes **ADALINE** de dos niveles, debido a que estas no pueden procesar funciones como es el caso de la OR exclusiva.

Los primeros modelos de redes neuronales artificiales cuyos niveles se encontraban compuestos por elementos adaptivos múltiples, es decir, una combinación de redes **ADALINE**.

La red **MADALINE** puede ser presentada con un vector de entrada de una dimensión extensa, como por ejemplo, los valores de los píxeles de una imagen rastreada pueden ser tomados como los valores de los patrones de entrada. En este caso, con un entrenamiento adecuado, la red podría enviar como respuesta un 1 binario a cualquiera de los diferentes nodos de salida, en donde cada respuesta corresponde a una de las porciones de la imagen de entrada. Para un patrón de entrada, un nodo deberá tener una salida con valor +1, si en patrón de entrada corresponde al de la clase representada por dicho nodo, de manera que los nodos restantes que forman a la red deberán tener un valor de -1.

Cuando una red neuronal de este tipo ha sido entrenada con una identificación previa de los patrones de entrada, el vector de salida deseado es ya conocido; lo que realmente es desconocido, es el valor de la salida deseada para un nodo conocido el cual se encuentra localizado en uno de los niveles ocultos.

Durante el desarrollo de estas redes, tenemos que al principio de la década de los 60' s se construyó una red con arquitectura **MADALINE** formada por 1000 conexiones de peso; la cual fue utilizada en el área de reconocimiento de patrones, en ella, los pesos que se utilizaron para su elaboración fueron fabricados con memistores, que son resistencias eléctricamente variables.

En la figura 2.16 se observa una red del tipo **MADALINE**, la cual está formada de tres niveles, que a su vez involucran diversos elementos Adaline, los cuales se encuentran conectados en su totalidad, es decir, cada elemento Adaline recibe en sus entradas todas las salidas provenientes del nivel precedente.

Durante su funcionamiento, la respuesta de cada elemento de salida es comparado con su correspondiente respuesta deseada, en tanto, las señales de error asociadas con cada uno de estos elementos son parcialmente procesadas, de esta manera, la adaptación del nivel de salida es directa. Además, uno de los problemas fundamentales de las redes multiniveles con arquitectura **MADALINE**, se encuentra en la obtención de las señales de error para los elementos de las redes Adaline que se encuentran en los niveles ocultos, pero esto no es un factor determinante para decir que una red con alimentación hacia adelante no deba tener una estructura multinivel.

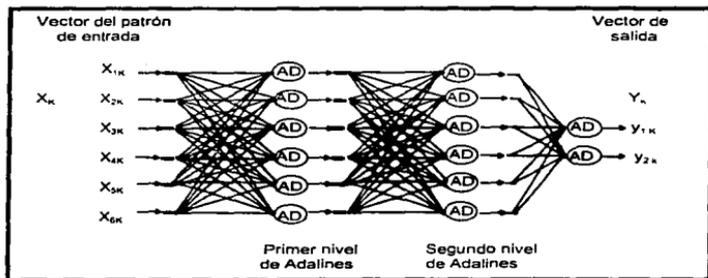


Fig. 2.16 Red Neuronal Artificial del tipo MADALINE, con alimentación hacia adelante y formada con dos niveles ADALINE.

Finalmente, retomando el problema que se presenta frecuentemente en los sistemas de reconocimiento de patrones, tenemos que la invariancia que existe en la salida de la red, ante los cambios presentes tanto en la posición como en el tamaño del patrón de entrada o de la imagen, es mínima. Para esto, existen varias Técnicas que han sido utilizadas para solucionar problemas, entre las que se encuentran; el movimiento de rotación, translación y escala de imágenes, así como la eliminación de los efectos de invariancia en el tiempo. Estas cuatro propiedades de las redes con estructura **MADALINE** son esenciales para cualquier sistema que se desee aplicar dentro del campo de la visión artificial como es el reconocimiento de patrones o los sensores infrarrojos.

# Capítulo 3

## INTRODUCCION AL PROCESAMIENTO DE IMAGENES

---

### INTRODUCCION

La primera etapa en un sistema de comprensión de imagen es la etapa de procesamiento, ya que es de gran importancia y útil en la descripción de áreas y superficies de una imagen. La imagen es adquirida por la grabación reflejando luz de la superficie de los objetos. 4 factores importantes contribuyen a los valores de intensidad de un pixel en una imagen. (1) la geometría, (2) reflejos de superficie visible, (3) iluminación de la escena y (4) el punto de vista.

En un sistema de compresión de imagen, la etapa de procesamiento a menudo ejecuta funciones tales como la manipulación de la escala de gris, detección de etapas, desarrollando la descripción de objetos o formas en la imagen, restauración de imagen, y corrección geométrica. Los cambios de intensidad son importantes indicios en la estructura de una superficie visible. En las etapas de procesamiento tratamos de captar esta información. Hay un número de operaciones en el procesamiento de imágenes que pueden ser ejecutados con conocimientos acerca del contenido de la imagen. Esta operación incluye la manipulación de la escala de gris, filtro, ruido removible, detección de áreas, segmentación, y restauración. Para tratar de realzar la imagen con calidad se usan técnicas convencionales como los modelos de R.N.A.

El procesamiento digital de imágenes es un desarrollo relativamente reciente, además de que se requiere de una computadora para su análisis y manipulación. Una de las causas por las cuales se utiliza la computadora, es por su gran capacidad de procesamiento, rapidez y almacenamiento de datos, pero debido a que las computadoras trabajan con datos que son más bien numéricos que gráficos, se hace necesario que para poder manipular una imagen, ésta sea convertida en su forma numérica.

### ***3.1 Serie y Transformada de Fourier***

Cuando es transmitida una información a través de un sistema de comunicación, la velocidad de transmisión esta relacionada con la rapidez con que las señales cambian en el tiempo. Por lo que cualquier señal que exista en un intervalo de tiempo es posible descomponerla a través de la serie de Fourier.

Existen dos representaciones de la serie de Fourier: la serie trigonométrica y la serie exponencial compleja, la primera se refiere al empleo de senos y cosenos en su serie o bien, al empleo de senos con ángulos de defasamiento y la segunda al empleo de términos exponenciales complejos o  $\exp [ \pm j \eta \omega t ]$ .

La serie exponencial esta representada en la ecuación 3.1

$$f(t) = C_n + \sum_{n=1}^{\infty} \{ C_n \exp [ j \eta \omega t ] + C_{-n} \exp [ - j \eta \omega t ] \} \quad (3.1)$$

$$f(t) = C_n + \sum_{n=-\infty}^{\infty} C_n \exp [ j \eta \omega t ] \quad (3.2)$$

de donde:

$$C_n = \left( \frac{1}{T} \right) \int f(t) \exp [ - j \eta \omega t ] dt, \quad n = 0, \pm 1, \pm 2, \dots \quad (3.3)$$

En las figuras 3.1 y 3.2 se muestran dos ejemplos de funciones discretas en el tiempo (pulso rectangular y pulso triangular); si se obtiene la serie de Fourier para ambas funciones, se representarían como sigue:

$$C_n = A \text{Sa} \left( \frac{n\pi}{2} \right), \quad (3.4)$$

para la figura 3.1

$$C_n = \left( \frac{A}{2} \right) \text{sinc}^2 \left( \frac{n}{2} \right), \quad (3.5)$$

y para la figura 3.2

Los espectros de amplitud para ambas funciones estarían representados como se muestra en las figuras 3.3 y 3.4

Si se observan ambos espectros, se observa que la figura 3.4 tiende más rápido a cero que la figura 3.3.

En la serie de Fourier se analizan funciones periódicas en el tiempo, pero para una aproximación más real es necesario utilizar funciones continuas en el tiempo. Para poder establecer la relación de tiempo-frecuencia de las funciones continuas se introduce el concepto de transformada de Fourier.

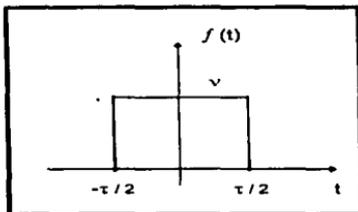


Fig. 3.1 Pulso rectangular

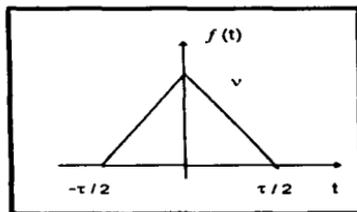
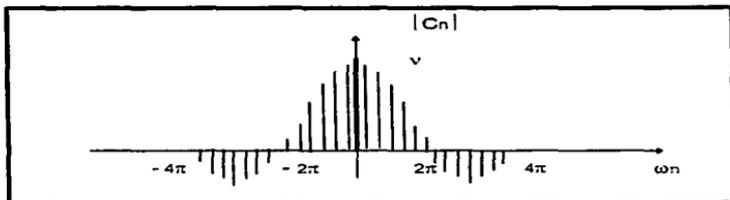
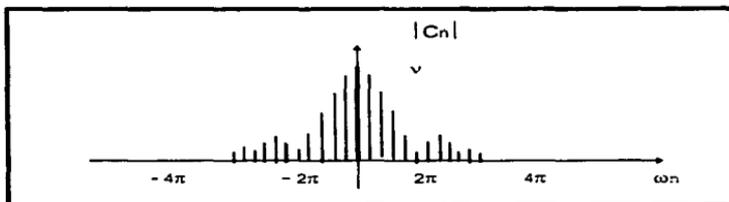


Fig. 3.2 Pulso triangular



3.3 Espectro de amplitud del pulso rectangular



3.4 Espectro de amplitud del pulso triangular

La transformada de Fourier también es conocida como Integral de Fourier, esta definida por la ecuación:

$$F(\omega) = \int_{-\alpha}^{\alpha} f(t) e^{-j\eta\omega t} dt \quad (3.6)$$

Es decir,

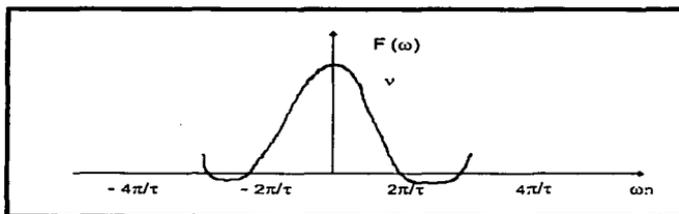
$$F(\omega) = \int_{-\alpha}^{\alpha} f(t) e^{-j\omega t} dt \quad \text{dado que } \eta\omega \rightarrow \omega \quad (3.7)$$

Si se obtiene la transformada de Fourier de la figura 3.1, resulta un espectro en frecuencia como el que se muestra en la figura 3.5, cuyo valor es el siguiente:

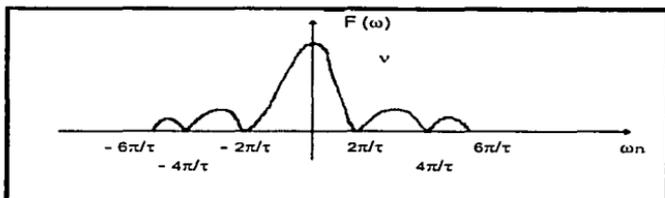
$$F(\omega) = A\tau \text{ Sa} \left( \frac{\omega\tau}{4} \right) \quad (3.8)$$

y en el caso de la figura 3.2, se tiene un espectro como el mostrado en la figura 3.6, con un valor de:

$$F(\omega) = \left( \frac{A\tau}{2} \right) \text{ Sa}^2 \left( \frac{\omega\tau}{4} \right) \quad (3.9)$$



3.5 Espectro del pulso rectangular



**3.6 Espectro del pulso triangular**

Observe como los espectros de la Serie de Fourier son semejantes a los de la Transformada de Fourier lo cual implica la relación entre señales continuas y discretas en el tiempo<sup>1</sup>.

### 3.2 Qué es una imagen digital

El componente individual de los elementos dentro de la imagen digital son referidos como picture elements. Dos abreviaciones comunes para este termino son pixel y pel. El valor digital de cada pixel de la línea 3, muestra 4, en la imagen mostrada en la figura 3.7 es cero. En una imagen en blanco y negro, una intensidad de cero representa negro.

Una imagen es una representación, semejanza o imitación de un objeto o cosa: una descripción de una gráfica o animación, una representación de algo para representar otro algo.

Una imagen digital se considera como un arreglo matricial de números. Un arreglo es mostrado en la fig. 3.7, escena es un cuadrado negro sobre un fondo blanco, en la misma figura se representa una posible representación digital. Cada punto dentro de la representación digital corresponde a una área de un objeto en el espacio; un valor digital es asignado a cada punto dentro de la imagen digital relacionando la intensidad del área del objeto en el espacio.

La figura 3.8 contiene una representación digital diferente de la misma escena mostrada en la figura 3.7; la misma frecuencia de muestreo (que será analizada posteriormente) ha sido usada para generar esta versión digital, pero el número de bits

<sup>1</sup> Para más detalles vease Capítulo 2, Mischa Schwartz, Transmisión de información modulación y ruido.

usado para representar la intensidad digital ha sido incrementada de 1 a 8. Los valores de la intensidad digital pueden ser de un rango de cero a 255, con cero se representa el negro y con 255 se representa el blanco.

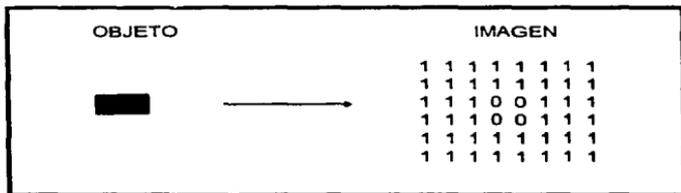


Fig. 3.7 Ejemplo de una imagen digital

Los efectos que ocurren en una representación digital cuando se muestrea una escena continua pueden ser vistos en la figura 3.8. Primero, los valores de intensidad digital representan el cuadro negro. Los cuatro píxeles representan la uniformidad del cuadro negro tienen valores de intensidad digital de 3,2,2, y 4.

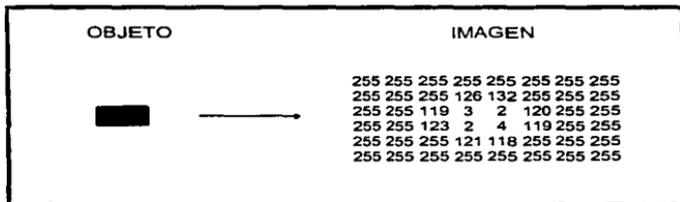


Fig. 3.8 Representación digital de la escena de la figura 3.7, intensidad digital representada como 8 bits por píxel.

La mayoría de los sistemas están provistos de representaciones de muestras digitales de una escena, la cual presenta un grado de ruido no deseado dentro de la imagen muestreada. Si no se presenta ruido durante el proceso de muestreo, la intensidad de los valores digitales de los 4 píxeles sería la misma.

Un segundo efecto que puede ser observado es la falta de una transición aguda de la imagen digital muestreada a los bordes del cuadro negro. En la figura 3.7 el cuadro es claramente definido en los límites, pero en la figura 2.8 se presenta lo que realmente ocurre cuando el límite es muestreado para crear una representación digital discreta de la escena.

### 3.3 Tipos de Imágenes

Las imágenes se presentan en varias formas: visibles, no visibles, abstractas, reales, adecuadas para el análisis por computadora y no adecuadas. Las imágenes se clasifican tomando en cuenta la forma o método con el cual son generadas.

Si consideramos como un conjunto a todos los objetos, las imágenes forman un subconjunto, ver Fig. 3.9.

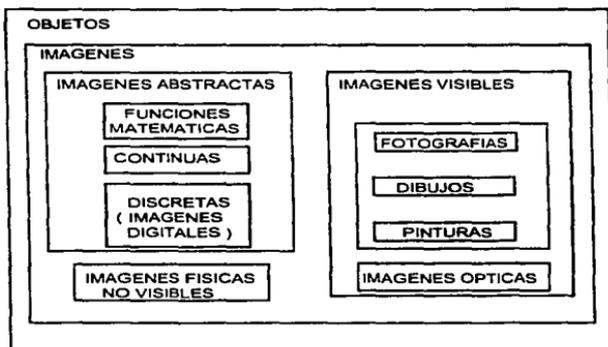


Fig. 3.9 Tipos de imágenes

Dentro del subconjunto de las imágenes, se encuentran varios subconjuntos, uno de ellos es el de las imágenes visibles (o percibidas por el ojo humano), en este se encuentran otros subconjuntos: fotografías, dibujos, pinturas; también está el subconjunto de las imágenes ópticas (generados por lentes y hologramas).

Las imágenes físicas no visibles son distribuciones de propiedades físicas medibles, por ejemplo, temperatura, presión, altitud, densidad de población. El otro subconjunto contiene las imágenes abstractas entre las que se encuentran las funciones matemáticas, las funciones continuas y discretas o imágenes digitales (sólo las imágenes digitales pueden ser procesadas por la computadora).

### 3.4 Elementos de un sistema de procesamiento digital de imágenes

Los elementos básicos de un sistema de procesamiento digital de imágenes son mostrados en la Figura 3.10

#### A) Procesador de imagen.

El procesador es el corazón de un sistema de procesamiento digital de imagen. Un procesador consiste en módulos hardware que ejecutan cuatro funciones básicas: adquisición de imagen, almacenamiento, procesamiento de bajo nivel y despliegue.

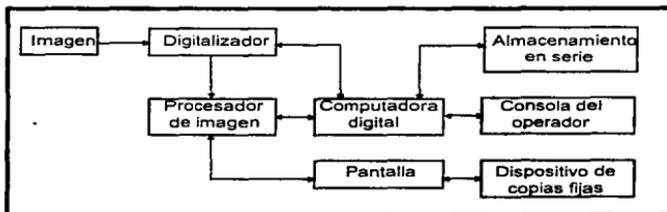


Fig. 3.10 Elementos de un sistema de procesamiento digital de imagen.

El módulo de adquisición de imagen, también llamado frame grabber, generalmente tiene como entrada una señal de video de TV y convierte esta señal a su forma digital. El módulo de almacenamiento, también llamado memoria de refrescamiento (frame buffer), es una memoria capaz de almacenar una imagen digital entera. El módulo de procesamiento ejecuta funciones de bajo nivel como operaciones aritméticas y lógicas, este módulo es también llamado Unidad Aritmética y lógica (ALU). La función del módulo de despliegue es de leer una imagen de memoria, convertir la información digital a una señal de video analógica y direccionar la señal de salida a un monitor u otro dispositivo.

**B) Digitalizador**

Un digitalizador convierte una imagen a una representación numérica propia para una entrada en la computadora digital. Entre los dispositivos más comúnmente usados están los microdensitómetros, lentes luminosos de los scanners, analizadores de imagen, cámaras vidicon y arreglos fotosensitivos de estado sólido. Los dos primeros requieren de la imagen digitalizada, en forma de transparencia o fotografía; los restantes también las aceptan pero tienen la ventaja adicional de digitalizar imágenes naturales que tengan suficiente intensidad de luz para excitar el detector.

**C) Computadora Digital**

Los sistemas de computación usados para procesar una imagen, van desde microprocesadores hasta grandes sistemas de computación capaces de ejecutar funciones computacionalmente intensivas dentro de grandes arreglos de imagen. Los parámetros que influyen en la estructura de una computadora depende de la aplicación y los datos que se requieran para llevarla a cabo. Una imagen digital requiere de gran cantidad de memoria durante su procesamiento.

**D) Dispositivos de almacenamiento**

Una imagen digital consiste en  $512 \times 512$  píxeles, cada uno de los cuales es cuantizado dentro de ocho bits y requiere de 0.25 megabytes de almacenamiento. Los tres dispositivos de almacenamiento usados para este trabajo son discos magnéticos (700 Mb o más de capacidad) almacenan 2,800 imágenes del tamaño antes mencionado; cintas magnéticas (6,400 bytes por cada 2.5 c.) almacenan una imagen por cada cm. y disco óptico (4 Gb de capacidad) almacenan aproximadamente 16,000 imágenes por disco.

**E) Despliegue y dispositivos de grabación**

Los monitores monocromáticos y en color son los principales dispositivos de despliegue usados en los sistemas de procesamiento de imágenes actuales. Los monitores son direccionados por la(s) salida(s) del módulo de despliegue de imagen dentro del procesador de imágenes. Estas señales pueden ser alimento dentro del dispositivo de grabación de imagen cuya función produce una copia fija (pantalla fotográfica, fotografías, y transparencias) de la esencia de la imagen dentro de la pantalla del monitor. Otros medios de despliegue incluye el tubo de rayos catódicos (TRC) y dispositivos de impresión.

**3.5 Escala de gris**

Las imágenes digitalizadas tienen únicamente dos niveles de la escala de gris: blanco y negro. Se pueden tener 8 bits 256 niveles que van desde el blanco hasta el

negro, pasando por una escala de grises con una alta resolución. La resolución de una imagen depende en gran parte de los niveles de la escala de grises.

El ojo humano puede diferenciar aproximadamente 40 diferentes matices de gris, en la práctica el número empleado rebasa la capacidad del ojo humano.

El término imagen se refiere a una función de dos dimensiones luz - intensidad, denotada por  $f(x,y)$ , en donde el valor o amplitud de  $f$  esta dado en coordenadas espaciales  $(x,y)$  de la intensidad (brillo) de la imagen en un punto.  $(x,y)$  es proporcional a la brillantez (nivel de gris) de la imagen en un punto (figura 3.11).

### 3.6 Luminancia

Desde que la luz es una forma de energía,  $f(x,y)$  debe ser finita y no cero, esto es,

$$0 < f(x, y) < \alpha \quad (3.10)$$

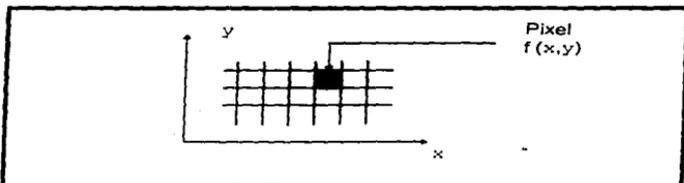


Fig. 3.11 Digitalización de una imagen.

Las imágenes que percibimos en nuestras actividades diarias consisten de luz reflejada de los objetos. La naturaleza básica de  $f(x, y)$  puede ser considerada como su esencia caracterizada por dos componentes. Un componente es la suma del origen de la luz, incidida sobre la escena vista, mientras que el otro es la suma de la luz reflejada por el objeto en la escena. Estos componentes son llamados iluminación y reflejo de los componentes, y son denotados por  $i(x, y)$  y  $r(x, y)$ , respectivamente. Las funciones  $i(x, y)$  y  $r(x, y)$  combinadas como un producto forman  $f(x, y)$ :

$$f(x, y) = i(x, y) r(x, y) \quad (3.11)$$

donde

$$0 < y(x, y) < \alpha \quad (3.12)$$

$$0 < r(x, y) < 1 \quad (3.13)$$

La ecuación 3.13 indica el factor de reflejancia, el cual es limitado por 0 (total absorción) y por 1 (total reflejancia). La naturaleza de  $i(x, y)$  es determinada por el origen de la luz, mientras que  $r(x, y)$  es determinada por las características del objeto en la escena.

La intensidad de una imagen monocromática  $f$  en coordenadas  $(x, y)$  es llamada nivel de gris ( $i$ ) de la imagen en ese punto.

Desde la ecuación 3.11 hasta la 3.13, es evidente que  $i$  es falso en el rango.

$$L_{\min} \leq i \leq L_{\max} \quad (3.14)$$

En teoría, se requiere que  $L_{\min}$  sea positivo, y  $L_{\max}$  sea finito. En la práctica,  $L_{\min} = (i_{\min}) (r_{\min})$  y  $L_{\max} = (i_{\max}) (r_{\max})$ . Usando por encima los valores de iluminación y reflejancia como una norma, uno puede esperar los valores de  $L_{\min}$  es de aprox. 0.005 y para  $L$  es de aprox. 100 dentro de aplicaciones del procesamiento de imágenes.

El intervalo  $[L_{\min}, L_{\max}]$  es llamado escala de gris. En la práctica es común variar estos intervalos numéricos a intervalos  $[0, L]$ , donde  $i = 0$  es considerado como negro y  $i = L$  es considerado como blanco en esta escala. Todos los valores intermedios son considerados matices de gris variando continuamente desde el negro al blanco.

### 3.7 Color

La luz es una forma de energía electromagnética esto es, la imagen esta dada por la distribución de ondas largas. No toda la radiación electromagnética es visible para el ojo humano. De hecho, la porción entera visible de la radiación es sólo dentro del límite de la banda de onda larga de 380 a 780 nm. (Figura 3.12). Esta radiación cuando incide sobre el ojo produce una variedad de sensaciones, por ejemplo, la sensación de brillo producido por el conjunto de varias ondas largas. Sin embargo, la luz de varias ondas largas también producen otra importante información visual llamada color.

Las ondas largas de la luz producen un estímulo visual. Como estas ondas varían, los colores percibidos (o mejor dicho el matiz) de los cambios de luz varían, desde el rojo (corresponde a la onda larga más larga) pasando por el naranja, amarillo, verde y del azul al violeta.

Las mezclas de las ondas largas producen la percepción de matices, algunos de los cuales (por ejemplo, el púrpura) no pueden ser obtenidos desde una sola onda larga.

La luz blanca (o gris) también puede ser obtenida por una mezcla de colores. De hecho, cualquier matiz puede ser obtenido por una mezcla de colores (rojo, verde y azul) en proporción adecuada (fig. 3.13).

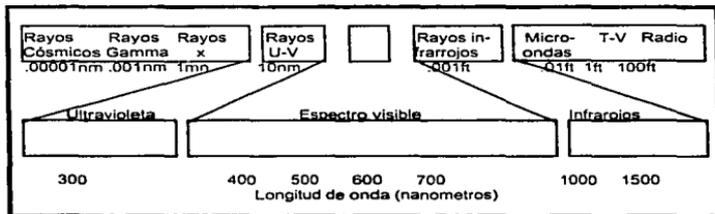


Fig. 3.12 Espectro de energía electromagnética, mostrando la porción del espectro visible.

El color es un importante aspecto de energía radiante visible por el cual un observador puede distinguir entre diferentes composiciones espectrales. Por eso, un estímulo de color, es especificado por energía radiante visible de una intensidad dada y una composición espectral.

El color es generalmente caracterizado por juntar nombres para los diferentes estímulos, por ejemplo, blanco, gris, negro, rojo, verde, azul. Los estímulos de color son generalmente más agradables para el ojo humano que el estímulo "blanco y negro". El colorido de imágenes blanco y negro por transformaciones de intensidad dentro de los colores es llamado pseudo-color. Esta adición de color regularmente se realiza en proporción a la escala de gris, esto es, a cada valor le corresponderá una determinada escala de gris.

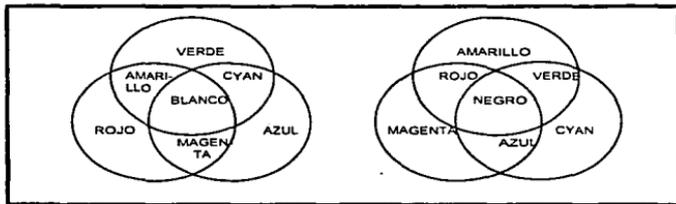


Fig. 3.13 Los tres colores primarios y los matices obtenidos al ser mezclados.

El color es generalmente caracterizado por juntar nombres para los diferentes estímulos, por ejemplo, blanco, gris, negro, rojo, verde, azul. Los estímulos de color son generalmente más agradables para el ojo humano que el estímulo "blanco y negro". El colorido de imágenes blanco y negro por transformaciones de intensidad dentro de los colores es llamado pseudo-color. Esta adición de color regularmente se realiza en proporción a la escala de gris, esto es, a cada valor le corresponderá una determinada escala de gris.

La especificación del color consiste de tres partes: (a) Semejanza del color, (b) Diferencias de color, (c) Apariencia del color (o percepción del color).

El estudio del color es importante en diseños y desarrollo de sistemas de visión del color. El uso del color en el despliegue de imágenes no sólo es más agradable, sino que facilita la recepción de más información visual. El ojo humano puede distinguir sólo cerca de 32 sombras de gris discreto; por el contrario es menos limitado en la discriminación del color y es capaz de distinguir por arriba de varios cientos de sombras de color discreto. Por esta razón, el color es una importante herramienta en el realce de la imagen.

El color varias veces es usado en la interpretación de imágenes que podrían desplegarse en blanco y negro. Cuando esto es así, el color es adherido como un mecanismo de ayuda al ser humano en la interpretación del análisis de datos. Una imagen de 8 bits contiene intensidades desde 0 hasta 255 que pueden ser transformadas dentro de una imagen en color. El color producido en una imagen por translación a una imagen monocromática dentro de la presentación de colores es llamada falso color de imagen. Esto no tiene relación entre el color producido en el falso color de imagen y el verdadero color del objeto que fue imaginado.

El falso color en las imágenes es producido por intensas transformaciones. La intensidad de los rangos de entrada del digitalizador de imagen es transformada dentro de tres rangos de intensidad separados a la salida: uno es para el rojo, otro para el verde y el otro para el azul, estos componentes de un color salen de la composición de la imagen, es decir, de una entrada particular de imagen, se producen tres salidas de imagen.

Los atributos de la percepción son el brillo, el matiz y la saturación. Los brillos representan la luminancia percibida. El matiz de un color se refiere a la "rojura", "verdor" y así sucesivamente. Para el origen de luces monocromáticas, la diferencia en matices son manifestados por la diferencia de ondas largas. La saturación en ese aspecto, es la percepción de más y más luz sumada a una luz monocromática. Estas definiciones son algo imprecisas porque el matiz, la saturación y el brillo cambian cuando la onda larga, la intensidad, el matiz o la suma de la luz blanca en un color es cambiado.

# Capítulo 4

## TECNICAS UTILIZADAS EN EL PROCESAMIENTO DE IMAGENES

---

### INTRODUCCION

La transformación es importante en el procesamiento digital de imágenes, tanto en teoría como en aplicaciones. La transformación de imágenes se ha desarrollado debido a los adelantos tanto en software como la tecnología de circuitos integrados de estado sólido, como en hardware con los algoritmos de convolución rápida.

Los sistemas digitales van adquiriendo mayor flexibilidad en el procesamiento de algoritmos, procesos iterativos, procesos de evaluación de requerimientos y en la toma de decisiones.

### 4.1 TECNICAS DE TRANSFORMACION DE IMAGENES

La transformada de dos dimensiones es usada para el mejoramiento, restauración, codificación y descripción de imágenes.

#### 4.1.1 Transformación unitaria de dos dimensiones.

La transformación unitaria de dos dimensiones ha encontrado tres aplicaciones principales en el procesamiento de imágenes. Primeramente se ha utilizado para extraer las características principales de las imágenes. Por ejemplo, en la transformada de Fourier el valor promedio es proporcional al promedio del brillo de la imagen y en términos de alta frecuencia de una indicación de la amplitud y una orientación del contorno dentro de una imagen. Otra aplicación es la codificación de imagen por transformada en la cual una reducción de ancho de banda es realizada por descartación o una cuantización en conjunto de baja magnitud de coeficientes transformados. La tercera aplicación es la reducción dimensional en computación.

#### 4.1.2 Transformación unidimensional.

Quando una señal unidimensional es representada por una serie ortogonal de funciones base, la imagen puede ampliarse en términos de un juego discreto de arreglos base llamados imágenes base. Estas imágenes base pueden ser generadas por matrices unitarias. Alternativamente, una matriz de imagen de  $N \times N$  puede ser vista como un vector de  $N^2 \times 1$ . Una transformación de imagen provee un juego coordinado o vectores base para el espacio del vector.

Para una función continua, la serie ortogonal de expansión provee series de coeficientes los cuales pueden ser usados para cualquier proceso futuro de las funciones. Para una secuencia unidimensional  $\{u(n), 0 \leq n \leq N-1\}$ , representada como un vector  $u$  de tamaño  $N$ ; una transformación unitaria es descrita como

$$v = A u \quad v(k) = \sum_{n=0}^{N-1} a(k, n) u(n), \quad 0 \leq k \leq N-1 \quad (4.1)$$

donde  $A^{-1} = A^*T$  (unitaria). Esta dado

$$u = A^{-1} v \quad u(n) = \sum_{k=0}^{N-1} v(k) a^*(k, n), \quad 0 \leq n \leq N-1 \quad (4.2)$$

Cabe aclarar que ambas ecuaciones son para el caso discreto. La ecuación 4.2 puede ser vista como una representación en serie de la secuencia  $u(n)$ . Las columnas de  $A^{-1}$ , es decir, los vectores  $a^* = (a^*(k, n), 0 \leq n \leq N-1)^T$ , son llamados los vectores base de  $A$ . La serie de coeficientes  $v(k)$  da una representación de la secuencia original  $u(n)$  y son beneficiosos en el filtrado, compresión de datos y otros análisis.

#### 4.2 TECNICAS DE PROCESAMIENTO DE IMAGENES EMPLEANDO TRANSFORMADA

Entre las transformadas utilizadas para el procesamiento digital de imágenes se encuentra principalmente la transformada de Fourier y sus derivaciones, que tienen un amplio rango de aplicaciones en problemas de procesamiento de imágenes. Con la utilización de otras formas de transformación se han obtenido mejoras al ser aplicadas en el procesamiento de imágenes.

Con la transformada Haar, Slant, Hadamard, Karhunen-Loeve, Coseno Seno, se han obtenido adelantos en su aplicación en casos específicos, porque se han creado algoritmos de rápidos resultados, reducido el número de cálculos y consecuentemente los métodos de aplicación.

#### 4.2.1 Transformada de Fourier

La transformada discreta de Fourier de dos dimensiones de un arreglo de imagen se define por una serie de la siguiente forma:

$$f(u,v) = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k=1}^{N-1} F(j, \kappa) \exp \left[ \frac{-2\pi i}{N} (uj + v\kappa) \right] \quad (4.3)$$

donde  $i = \sqrt{-1}$ . La transformada inversa esta dada por:

$$F(j, \kappa) = \frac{1}{N} \sum_{u=0}^{N-1} \sum_{v=1}^{N-1} f(u,v) \exp \left[ \frac{2\pi i}{N} (uj + v\kappa) \right] \quad (4.4)$$

Los índices  $(u,v)$  son llamados, frecuencias espaciales de la transformación, en analogía con la transformada continua de Fourier. La transformada de dos dimensiones puede ser calculada como una secuencia de renglones y columnas de una transformada unidimensional. Las funciones base de la transformada son de exponenciales complejos que pueden ser descompuestos en componentes seno y coseno. Esto es,

$$\begin{aligned} A(j, \kappa; u, v) &= \exp \left[ \frac{-2\pi i}{N} (uj + v\kappa) \right] \\ &= \cos \left[ \frac{2\pi i}{N} (uj + v\kappa) \right] - i \operatorname{sen} \left[ \frac{2\pi i}{N} (uj + v\kappa) \right] \end{aligned} \quad (4.5)$$

$$\begin{aligned} B(j, \kappa; u, v) &= \exp \left[ \frac{2\pi i}{N} (uj + v\kappa) \right] \\ &= \cos \left[ \frac{2\pi i}{N} (uj + v\kappa) \right] + i \operatorname{sen} \left[ \frac{2\pi i}{N} (uj + v\kappa) \right] \end{aligned} \quad (4.6)$$

La transformada de Fourier de dos dimensiones de una imagen es esencialmente una representación de la serie de Fourier de un área de dos dimensiones. Para que la representación de la serie de Fourier sea valida, el área debe ser periódica.

Si el arreglo de imagen representa una área de luminancia,  $F(j, k)$  sería una función positiva real. Sin embargo, la transformada de Fourier es, en forma general, compleja. La figura 4.1 contiene fotografías de una imagen y varias versiones de la transformada de Fourier.

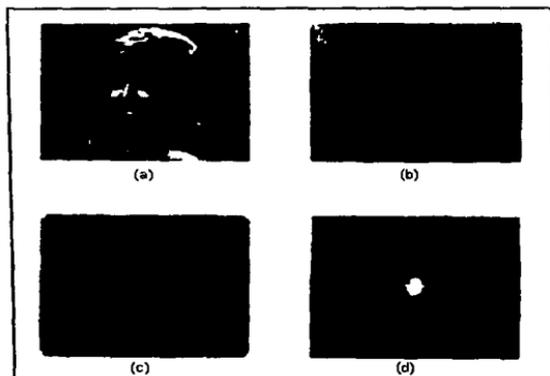


Fig. 4.1 Transformada de Fourier de una imagen, (a) Original, (b) magnitud logarítmica no ordenada, (c) magnitud comprimida no ordenada, (d) magnitud comprimida ordenada.

#### 4.2.2 Transformada rápida de Fourier (FFT)

Considerando un arreglo bidimensional de valores de imagen  $\beta_{m,n}$  obtenidas por muestreo, donde  $m = 0, 1, 2, \dots, M - 1$  y  $n = 0, 1, 2, 3, \dots, N - 1$ . Cada  $\beta_{m,n}$  es proporcional al brillo de la imagen muestreada, su posición esta dada por  $m$  y  $n$ .

La transformada discreta de Fourier de la función discreta  $\beta_{m,n}$  esta dada por:

$$F_{k,l} = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \beta_{m,n} \exp - \left[ j2\pi \left[ \frac{km}{M} + \frac{ln}{N} \right] \right] \quad (4.7)$$

donde  $k = 0, 1, \dots, M - 1$  e  $i = 0, 1, \dots, N - 1$ .

La inversa es :

$$\beta_{m,n} = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=1}^{N-1} F_{k,i} \exp \left[ j2\pi \left[ \frac{km}{M} + \frac{in}{N} \right] \right] \quad (4.8)$$

A fin de representar una matriz de pixeles, con una subposición de tamaño  $L \leq MN$ , de exponenciales complejos de la transformada discreta de Fourier, se escogen  $L$  exponenciales con lo que se alarga la magnitud de los coeficientes, minimizandose el mínimo error cuadrático.

El método usado es dividir la imagen dentro de sub-bloques rectangulares. Cada sub-bloque se expande en una transformada discreta de Fourier y los coeficientes son subejcutados para un umbral adaptivo.

Una cantidad  $L$  es determinada para cada sub-bloque tal que sea proporcional a la parte entera de la desviación estandard  $\sigma$  de los subbloques de los pixeles;

$$\sigma = \left[ \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=1}^{N-1} (\beta_{m,n} - a)^2 \right]^{1/2} \quad (4.9)$$

La frecuencias y amplitudes complejas de los  $L$  coeficientes de Fourier de mayor valor son transmitidos; la cuantización lineal adaptiva de los coeficientes de Fourier son usados como su precedente para su transmisión.

El máximo coeficiente de Fourier al lado de  $F_{0,0}$ , determina el valor de cuantización máximo dentro de cada sub-bloque. Las fases de los coeficientes complejos de Fourier son cuantizados en una escala de 0 a  $2\pi$ ; el número de bits usados para cuantizarlos se determinan por  $\sigma$ .

También se usan 2, 3, 4 o 5 bits para cuantizar todas las fases y amplitudes muestreadas dentro de cada sub-bloque.  $F_{0,0}$  es finalmente cuantizado previo a su transmisión, independientemente del valor de  $\sigma$ . Los coeficientes no transmitidos son puestos cero en el decodificador del receptor.

Son varias las cosas para considerar el tamaño del sub-bloque, entre ellas el valor de  $\sigma$  que se incrementa con el área del sub-bloque, porque la correlación entre los puntos de imagen generalmente decrece conforme la distancia entre los píxeles aumenta. En un sub-bloque de mayor tamaño es muy probable que las áreas de pequeña variación en el brillo sean incluidas con áreas de gran variación. En áreas de poca variación es probable que se den pérdidas, por la presencia del mismo sub-bloque de la áreas de gran variación.

El tamaño del sub-bloque escogido por la implementación hardware es de 16 x 16. En la implementación de este sistema las magnitudes y las muestras de fase dentro de un sub-bloque son linealmente cuantizadas para 2, 3, 4 o 5 bits dependiendo del valor de  $\sigma$ .

#### 4.2.3 Transformada Haar

Haar propuso una transformada simple compuesta de unos, menos unos y ceros. La matriz Haar de 4 x 4 esta dada por:

$$H_4 = \frac{1}{\sqrt{4}} \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{vmatrix} \quad (4.10)$$

La matriz de 8 x 8 queda como sigue:

$$H_8 = \frac{1}{\sqrt{8}} \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{vmatrix} \quad (4.11)$$

Esta transformada, ha sido implementada para procesos de muestreo en el cual los renglones de la matriz transformada muestre la secuencia de datos de entrada con una fina resolución incrementándose en potencia de 2. Sin embargo la calidad de la imagen resultante es mala respecto a otras transformadas por lo que no es muy usada. La fig. 4.2 muestra la Transformada Haar de una imagen.

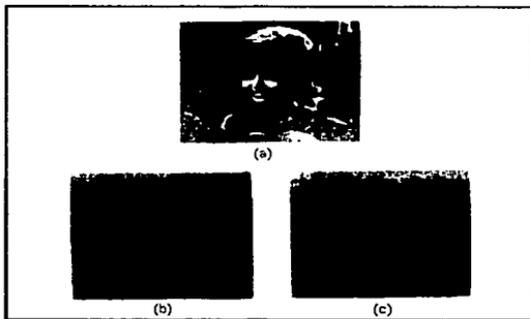


Fig. 4.2 Transformada Haar de una imagen, (a) Original, (b) magnitud logarítmica, (c) magnitud comprimida.

#### 4.2.4 Transformada Hadamard

La transformada Hadamard se basa en la matriz Hadamard la cual es un arreglo de más y menos unos, cuyos renglones y columnas son ortogonales; la matriz Hadamard satisface la relación

$$\mathbf{H} \mathbf{H}^T = \mathbf{I} \quad (4.12)$$

La matriz Hadamard de  $2 \times 2$  esta dada por:

$$H_2 = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \quad (4.13)$$

Esta matriz no se restringe a 2 y es extendible a cualquier valor de  $N = 2^n$  donde  $n$  es un entero, en este caso si  $H_N$  es una matriz de tamaño  $N$  la matriz es

$$H_{2N} = \frac{1}{\sqrt{2}} \begin{vmatrix} H_N & H_N \\ H_N & H_N \end{vmatrix} \quad (4.14)$$

Esta es una matriz de orden  $2^N$ ; H. Harmuth en su artículo "transmission of information by orthogonal functions" sugirió que es posible construir una matriz Hadamard de orden  $N = 2^n$  en la cual el número de signos por renglón se incrementara de 0 hasta  $N - 1$ ; los renglones pueden ser considerados muestras de ondas rectangulares con subperiodo de  $1/N$  unidades; estas son llamadas funciones Walsh. Es por esto, que esta transformada es llamada Walsh o Walsh-Hadamard.

Para matrices simétricas de orden  $N = 2^n$  una transformada Hadamard de 2 dimensiones se puede escribir en series de la siguiente forma:

$$f(u,v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} F(x,y) (-1)^{p(x,y,u,v)} \quad (4.15)$$

donde

$$p(x,y,u,v) = \sum_{i=1}^{N-1} (x_i y_i + u_i v_i) \quad (4.16)$$

Los términos  $x_i$ ,  $y_i$ ,  $u_i$  y  $v_i$  son representaciones binarias de  $u$ ,  $x$ ,  $u$  y  $y$ , respectivamente. La figura 4.3 representa la transformada Hadamard de una imagen.

#### 4.2.5 Transformada Karhunen-Loeve

Las técnicas para la transformación de señales continuas, fue originalmente desarrollada por H. Karhunen y M. Loeve; a H. Hotelling se le acreditan los procesos de conversión de transformación de las señales discretas. Sin embargo, a Karhunen-Loeve se les reconoce ambas transformaciones.

La transformada Karhunen -Loeve es de forma general:

$$F(u,v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) A(x,y;u,v) \quad (4.17)$$

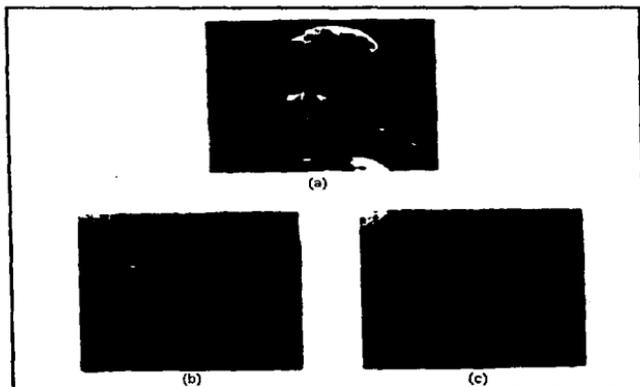


Fig. 4.3 Transformada Hadamard de una imagen, (a) original, (b) magnitud logarítmica, (c) magnitud comprimida.

para la cual  $A(x, y; u, v)$  satisface la ecuación:

$$\lambda(u, v) A(x, y; u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} K_l(x, y; x', y') A(x, y; u, v) \quad (4.18)$$

donde  $K_l(x, y; x', y')$  puede separarse en renglones y columnas resultando:

$$K_l(x, y; x', y') = K_c(x, x') K_r(y, y') \quad (4.19)$$

de (4.18) resulta:

$$A(x, y; u, v) = A_c(x, u) A_r(y, v) \quad (4.20)$$

$$F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) A_C(x, u) A_R(y, v) \quad (4.21)$$

que es la transformada Karhunen-Loeve en forma de serie.

#### 4.2.6 Transformada Slant.

La transformada Slant fue diseñada específicamente para codificación de imagen; utiliza un algoritmo computacional rápido, lo que provoca que se obtenga una mayor reproducción de calidad de imagen. Por sus características de funcionamiento es un poco menos óptimo que la transformada Karhunen-Loeve, pero más compleja que la transformada rápida de Fourier. Para un vector de longitud  $N = 2$  la transformada Slant es idéntica a la transformada Hadamard de orden 2, es decir:

$$S_2 = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} \quad (4.22)$$

La Transformada Slant de orden 4 es obtenida por la operación:

$$S_2 = \frac{1}{2^{1/2}} \begin{vmatrix} 1 & 0 & 1 & 0 \\ a_1 & b_1 & -a_1 & -b_1 \\ 0 & 1 & 0 & -1 \\ -b_1 & a_1 & b_1 & a_1 \end{vmatrix} \begin{vmatrix} S_2 & 0 \\ 0 & S_2 \end{vmatrix} \quad (4.23)$$

ó bien,

$$S_4 = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 1 & 1 & 1 \\ a_1 + b_1 & a_1 - b_1 & -a_1 + b_1 & -a_1 - b_1 \\ 1 & -1 & 1 & -1 \\ a_1 - b_1 & -a_1 - b_1 & a_1 + b_1 & -a_1 - b_1 \end{vmatrix} \quad (4.23a)$$

donde  $a_1$  y  $b_1$  son constantes de escala real cuyos valores se determinan de acuerdo con las condiciones que hacen a  $S_4$  ortogonal, se ha encontrado que este valor es  $a_1 = 2b_1$  y condición de ortogonalidad dice que  $S_4 S_4^T = I$  lleva a encontrar  $b_1 = 1/\sqrt{5}$ .

La matriz general para N términos, esta dada por:

$$S_N = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ -aN & bN & & -aN & bN & \\ 0 & & I(N/2-2) & 0 & & I(N/2-2) \\ 0 & 1 & 0 & 0 & -1 & 0 \\ -bN & aN & & bN & aN & \\ 0 & & I(N/2-2) & 0 & & -I(N/2-2) \end{vmatrix} \begin{vmatrix} SN/2 & 0 \\ & \\ & \\ & \\ 0 & SN/2 \end{vmatrix} \quad (4.24)$$

donde  $I(N/2-2)$  es la matriz identidad de dimensión  $(N/2)-2$ . Las constantes  $aN$  y  $bN$  pueden ser encontradas por las siguientes ecuaciones:

$$aN = 1 \quad (4.25)$$

$$bN = [1 + 4(aN/2)^2]^{-1/2} \quad (4.26)$$

$$aN = 2bN \quad (4.27)$$

$$aN = [(3N^2) / (4N^2 - 1)]^{1/2} \quad (4.28)$$

$$2N = [(N^2 - 1) / (4N^2 - 1)]^{1/2} \quad (4.29)$$

La transformada Slant requiere de un total de  $N \log_2 N + (N/2)-2$  adiciones y subtracciones más  $2N - 4$  multiplicaciones para un vector con una dimensión de datos  $N$ . Fotografías de la transformada Slant de una imagen se muestran en la figura (4.4)

#### 4.2.7 Transformada Coseno

Es conocido que en las series de Fourier para cualquier función real continua y simétrica, contienen únicamente coeficientes reales que corresponden a los términos coseno de dicha series.

Esto puede extenderse a la transformada de Fourier bajo su propia interpretación. Existen dos cambios en los cuales una imagen puede hacerse simétrica, como se muestra en la fig. 4.5 Por la primer técnica las imágenes son dobladas hacia un contorno y en el segundo método las imágenes son dobladas y superpuestas por un pixel.

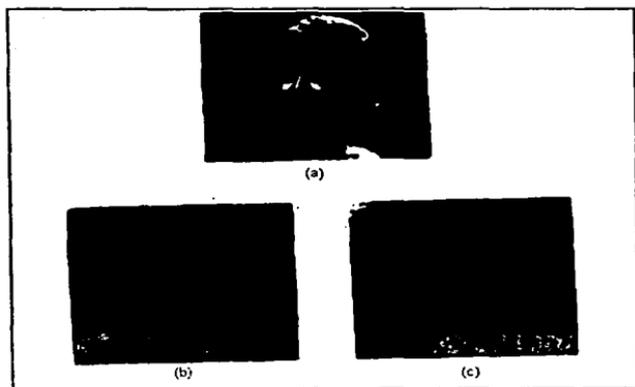


Fig. 4.4 Transformada Slant de una imagen, (a) original, (b) magnitud logarítmica, (c) magnitud comprimida.

De este modo, para una imagen de  $N \times N$  píxeles, el primer método es llamado Transformada Coseno de Simetría Semejante, resultando un arreglo de  $2N \times 2N$  píxeles; el segundo método de Transformada Coseno de Simetría Impar produce un arreglo de  $2N - 1 \times 2N - 1$  píxeles.

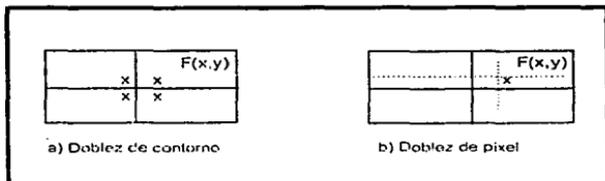


Fig. 4.5 Simetría para Transformada coseno.

**4.2.7.1 Transformada coseno de simetría semejante.**

Si se tiene un arreglo de imagen simétrica será formada por reflexión del arreglo de imagen hacia los contornos de acuerdo a la relación:

$$F_s(x,y) = \begin{cases} F(x,y) & x \geq 0, y \geq 0 \\ F(-1-x, y) & x < 0, y > 0 \\ F(x, -1-y) & x \geq 0, y < 0 \\ F(-1-j, -1-k) & x < 0, y < 0 \end{cases} \quad (4.30)$$

Para esta construcción  $F_s(x,y)$  es simétrica hacia el punto

$x = -1/2, y = -1/2$  y tomando la transformada de Fourier hacia el punto de simetría resulta:

$$f_s(u,v) = -\frac{1}{2N} \sum_{x=-N}^{N-1} \sum_{y=-N}^{N-1} F_s(x,y) \exp \left\{ \frac{-2\pi i}{2N} \left[ \left(x + \frac{1}{2}\right) + v \left(y + \frac{1}{2}\right) \right] \right\} \quad (4.31)$$

para  $u,v = -N, \dots, -1, 0, 1, \dots, N-1$ ; debido a que  $F_s(x,y)$  es real y simétrica, la ecuación (4.29) se reduce a:

$$f_s(u,v) = \frac{2}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} F(x,y) \cos \left[ -\frac{\pi}{N} u \left(x + \frac{2}{2}\right) \right] \cos \left[ \frac{\pi}{N} u \left(y + \frac{2}{2}\right) \right] \quad (4.32)$$

De manera alternativa, los términos de las ecuaciones (4.29) y (4.30) pueden ser obtenidas por una transformada de Fourier de  $F(x,y)$  sobre  $2N$  puntos obteniéndose:

$$f_s(u,v) = \frac{2}{N} R_o \left\{ \exp \left[ \frac{i\pi u}{2N} \right] \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} F(x,y) \exp \left[ \frac{-2\pi i}{2N} (u x + v y) \right] \right\} \quad (4.33)$$

**4.2.7.2 Transformada coseno de simetría impar**

Para la transformada coseno de simetría impar, el arreglo de la imagen queda definida de la siguiente forma:

$$\begin{aligned}
 F_s(x, y) &= F(x, y) & x, y \geq 0 \\
 F_s(x, y) &= F(-x, y) & x < 0, y \geq 0 \\
 F_s(x, y) &= F(x, -y) & x \geq 0, y < 0 \\
 F_s(x, y) &= F(-x, -y) & x, y < 0
 \end{aligned}
 \tag{4.34}$$

Tomando en cuenta la transformada de Fourier se genera el arreglo:

$$f_s(u, v) = \frac{1}{2N-1} \sum_{x=-N+1}^{N-1} \sum_{y=-N+1}^{N-1} F_s(x, y) \exp \left[ \frac{-2\pi i}{2N-1} (x u, y v) \right]$$

para  $u, v = N+1, \dots, -1, \dots, N-1$ .

(4.35)

#### 4.2.8 Transformada seno

La transformada seno, fue introducida como un algoritmo rápido para substituir a la transformada Karhuen-Loeve.

Esta transformada es definida como:

$$f_s(u, v) = \frac{1}{N+1} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} F_s(x, y) \operatorname{sen} \left[ \frac{(x+1)(u+1)}{N+1} \pi \right] \operatorname{sen} \left[ \frac{(y+1)(v+1)}{N+1} \pi \right]$$

(4.36)

su inverso es de forma similar. La transformada seno puede también ser calculada con un algoritmo de la transformada de Fourier. La figura 4.6 muestra las fotografías de la transformada seno de una imagen.

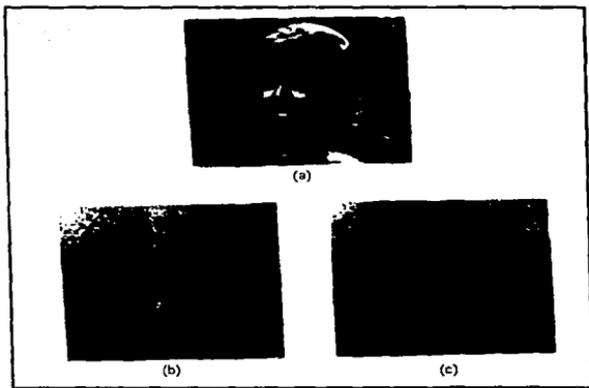


Fig. 4.6 Transformada Seno de una imagen, (a) original, (b) magnitud logarítmica, (c) magnitud comprimida.

# Capítulo 5

## LAS REDES NEURONALES Y EL RECONOCIMIENTO DE IMAGENES

### INTRODUCCION

El reconocimiento de un objeto en una imagen es una tarea compleja que envuelve un ancho rango de técnicas. El sistema de comprensión de imagen como se muestra en la Fig. 5.1 consiste de seis etapas: adquisición de imagen, procesamiento, extracción de características, almacenamiento asociativo, base de conocimientos y reconocimiento. Estas etapas esencialmente a un nivel bajo, medio y alto de procesamiento.

La primera etapa en el proceso es la adquisición de la imagen: adquiriendo una imagen digital. El término imagen se refiere a la función de dos dimensiones  $f(x,y)$  de intensidad de luz, donde las coordenadas espaciales  $x$  y  $y$  y los valores de  $f$  en un punto  $(x,y)$  es proporcional al brillo y al nivel de gris de la imagen en este punto.

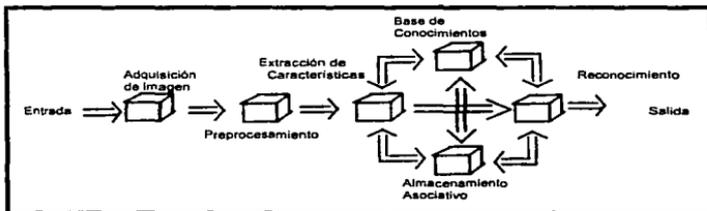


Fig. 5.1 Modelo de un sistema de reconocimiento de imagen

Una imagen digitalizada es una imagen  $(x,y)$  en la cual ambas coordenadas espaciales y el brillo han sido digitalizados. Un digitalizador de imagen puede ser considerado como una matriz de elementos que representa el nivel de gris en este punto. Los elementos de tal arreglo digitalizado es llamado cuadro de elementos o

pixeles. La etapa de adquisición de imagen esta compuesta de sensores que capturan la imagen. Los sensores pueden ser una camara o un scanner. La naturaleza de los sensores y la imagen que produce son determinados por la aplicación.

Después de que una imagen ha sido digitalizada, el siguiente paso es procesarla, el cual corresponde a una primera visión de procesamiento o bajo nivel de proceso. Una cantidad de técnicas de procesamiento son aplicables en la práctica. Estas incluyen la manipulación de la escala de gris, filtro de ruido, aislamiento de regiones, corrección geométrica, restauración, reconstrucción y segmentación. Las técnicas de realce de imagen pueden ser clasificadas en dos grandes categorías: métodos de dominio espacial y métodos de dominio en frecuencia. Los métodos de dominio espacial están basados en la manipulación directa de los valores de gris del pixel en la imagen. Las técnicas de proceso de dominio en frecuencia están basadas en la modificación de la transformada de Fourier de una imagen. En la manipulación de la escala de gris el realce de algún punto en la imagen puede depender de los valores de gris en ese punto, o puede depender de los valores de gris de los puntos cercanos a el. El realce en algunos pixeles depende de los valores de gris del pixel y hace uso de otros que generan sombras para definir el lineamiento del pixel. Algunas implementaciones de hardware y software son aplicables en la práctica para implementar la técnica de realce.

Los modelos de Redes Neuronal Artificiales (R.N.A.) proveen una alternativa de enfoque a la implementación de técnicas de realce. Investigadores tales como Grossberg ( 1988)<sup>1</sup> han considerado propiedades del sistema de visión humano y han propuesto arquitecturas de redes para percepción de brillo bajo una constante y variable de condiciones de iluminación. El sistema visual humano toma en cuenta variablemente la iluminación. El estado de procesamiento en un sistema de reconocimiento puede tratar con percepción de brillo, también con problemas de restauración y reconstrucción de imagen. El sistema de adquisición de imagen en la práctica puede no ser perfecto. Los métodos de restauración de imagen trata con estimados de una imagen original degradada.

El siguiente nivel de procesamiento es el nivel intermedio. En este nivel se trata de tener una coalición de muestras que son obtenidas en el bajo nivel de procesamiento y extraer entidades significantes. Un bien conocido nivel de procesamiento es la extracción de características, el cual consiste en mapear un vector de observación en un espacio de características. La principal tarea de la extracción de características es reducir datos midiendo ciertas características que distinguen la entrada de patrones. Al extraer características, se puede seleccionar un vector de entrada de observación, o uno puede transformar el vector de observación de entrada a un vector de características usando algunas funciones básicas ortogonales. En algunas aplicaciones, el vector de observación es obtenido por una

<sup>1</sup> Grossberg, S. (1988). *Neural Networks and Natural Intelligence*. Bradford Books, MIT Press, Cambridge, MA.

simple entrada de imagen que representa alta correlación de datos. Se reduce la dimensionalidad mientras se retiene la información, el vector de observación es mapeado sobre un espacio de dominio de características. Los datos en el dominio transformado pueden ser ordenados de acuerdo al grado de significancia de la información contenida y la calidad de los patrones portados.

En los pasados 30 años algunas técnicas fueron desarrolladas para extracción de características. Esas pueden ser la transformada de Fourier, momento invariante, la distribución Wigner, la transformada Hough, polinomios ortogonales, función Gabor, etc. En la actualidad algunos modelos de R.N.A. se han desarrollado para extraer características. El problema de reconocimiento de objetos invariantes es a menudo tratado en una etapa de extracción de características. Esto es porque para considerar translación, rotación y diferentes escalas en la imagen, el sistema de reconocimiento necesita tener un gran número de entrenamientos simples, para obtener características invariantes, para esto se usan las propiedades de la serie de la transformada de Fourier. El sistema visual humano es sensitivo a la variación de texturas en las superficies de los objetos. Las características de textura se usan para reconocer objetos. La textura generalmente es reconocida como la percepción fundamental. Algunos métodos estáticos y estructurales como los modelos de R.N.A. para análisis de textura son utilizados en la práctica. Los métodos estáticos de textura se basan en los valores de gris de un pixel en la imagen. Recientemente, modelos de R.N.A. basados en las funciones Gabor (Dugman, 1988) <sup>2</sup> se han usados para la segmentación de textura. El uso de la función Gabor se ha desarrollado porque los campos receptores de una neurona en un corte visual son conocidos para tener una forma que se aproxime a la función Gabor de dos dimensiones. La etapa de extracción de características trata con la extracción de características de textura.

Las últimas tres etapas, almacenamiento asociativo, base de conocimientos, y la etapa de reconocimiento, corresponde a un procesamiento de alto nivel. La memoria humana puede a menudo recordar información completa o parcial. La memoria asociativa contiene direcciones de memoria. La habilidad para pasar de una representación a otra o inferior a una porción de una compleja, forma las bases de una memoria asociativa. La función básica de una memoria de este tipo es asociar pares de patrones a través de una organización de procesos y produce patrones de respuesta apropiada en la presentación del correspondiente patrón de estímulo. La memoria asociativa es útil para el reconocimiento de objetos. En años pasados se han renovado modelos de memoria asociativa, de gran interés. Libros de texto de Kohonen (1988) <sup>3</sup>, Hinton y Anderson (1989) <sup>4</sup> describen memorias asociativas. La etapa de reconocimiento trata con clasificación, asigna una etiqueta a un objeto basado en la información provista por descriptores. Las técnicas de clasificación convencional son agrupadas en dos categorías: técnicas supervisadas y técnicas no

<sup>2</sup> Daugman, J.G. (1988). Complete discrete 2-D Gabor Transforms by neural networks for image analysis and compression.

<sup>3</sup> Kohonen, T. (1988). Self Organization and Associative Memory (de.2). Springer-Verlag, Berlin.

<sup>4</sup> Hinton, G., and Anderson, J. (1989). Parallel Models of Associative Memory. Lawrence Erlbaum Associates, Hillsdale, NJ.

supervisadas. En un modo supervisado, se clasifica el aprendizaje sin entrenamiento. Las redes neuronales representan un poder y alternativa razonable para una clasificación convencional. Modelos de R.N.A. con algoritmos de aprendizaje como el back-propagation se han usado como un clasificador supervisado y redes neuronales de auto organización con algoritmos de aprendizaje tales como un aprendizaje competitivo o teoría de resonancia adaptiva (ART), se usan como un clasificador supervisado. Métodos estáticos y clasificadores de redes neuronales se usan acertadamente en varios problemas de reconocimiento. Sin embargo, hay algunos problemas donde los métodos estadísticos son inapropiados y los métodos descriptivos son más apropiados. Los métodos descriptivos se basan en las reglas de clasificación que en éste caso pueden ser almacenadas en una base de conocimientos. La interacción entre la base de conocimientos y otros módulos en un sistema de reconocimiento se muestra en la figura 5.1. La base de conocimientos interactúa no sólo con la extracción de características y estados de reconocimiento sino también con el almacenamiento asociativo. A menudo conocimientos prioritarios acerca de un objeto pueden ser ubicados en la base de conocimientos. Los conocimientos pueden ser tan simples como una región detallada de una imagen, la información de interés es conocida para ser localizada, así limitando la búsqueda que tiene que ser conducida a encontrar la información. La base de conocimientos puede ser compleja. El diseño de una máquina de sistema de reconocimiento necesita abarcar todas las etapas de procesamiento.

### **5.1 Procesamiento**

La etapa de procesamiento, como ya se mencionó en el *Cáp. 3*, es el primer paso a seguir en el sistema de comprensión de imágenes ya que es en éste donde se describe la imagen.

#### **5.1.1 Técnicas de manipulación de la escala de gris**

Para hacer el máximo uso de la información o extraer características de la imagen, es necesario emplear alguna clase de técnica de realce. Algunos libros de texto y artículos de revistas como esos por Andrews (1970)<sup>5</sup>, Rosenfeld and Kak (1982)<sup>6</sup>, trataron con mapeo de punto a punto. Considerando píxeles colindantes y calculando el valor de los píxeles en la imagen de salida. Tales operaciones como la manipulación de la escala de gris y filtro son ejecutadas por el sistema de visión primaria.

La manipulación de la escala de gris es una técnica de mapear la intensidad en el cual cada pixel es asignado a un nuevo valor de gris mejorando el contraste en la

<sup>5</sup> Andrews, H.C. (1970). *Computer Techniques in Image Processing*. Academic Press, New York.

<sup>6</sup> Rosenfeld, A., and Kak, A. (1982). *Digital Image Processing, Vols. I and II*. Academic Press, Orlando, FL.

imagen. En la manipulación de la escala de gris a menudo el nivel de gris se diferencia entre el objeto y el fondo. Las técnicas de la escala de gris pueden ser definidas como:

$$g'(i, j) = T [ g (i, j) ] \quad (5.1)$$

donde  $g'$  es la imagen realizada,  $g$  es la imagen original,  $T(g)$  es la función transformada e  $(i, j)$  representa el número de columnas y renglones de los píxeles. Comúnmente se usa el traslado de funciones como se muestra en la Fig. 5.2. Los píxeles son mapeados de la imagen de entrada a la imagen de salida usando tablas. Las imágenes original y realizada usan una función de transferencia lineal, como se muestra en la Fig. 5.3 y 5.4 respectivamente.

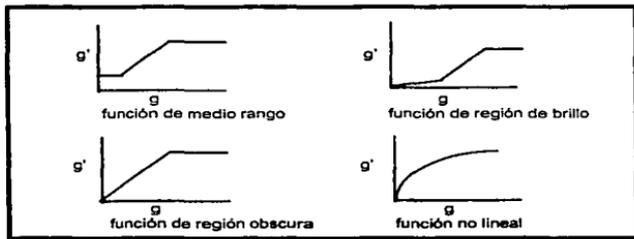


Figura 5.2 Funciones de transformación de la escala de gris

El histograma de igualación es uno de los más comúnmente usados en las técnicas de manipulación de la escala de gris. El histograma típico de una imagen de entrada y una de salida se muestra en la Fig. 5.5. Se puede ver en esa figura que los valores de gris en la imagen de salida están uniformemente distribuidos.

La función distribuida acumulativa (CDF) de una imagen típica se muestra en la Fig. 5.6. La línea deseada CDF corresponde al histograma igualado y se muestra en la Fig. 5.6.



Fig. 5.3 Imagen original (dato de satélite)



Fig. 5.4 Imagen reconstruida

El Procedimiento de linealización puede ser formulado como:

$$g'(i, j) = (g_{\max} - g_{\min}) P [g(i, j)] + g_{\min} \quad (5.2)$$

donde  $g'$  representa la imagen realzada,  $g$  representa la imagen de entrada,  $p [g(i, j)]$  es el valore de CDF en un nivel de gris  $g(i, j)$ ,  $g_{\min}$  y  $g_{\max}$  son el mínimo y máximo nivel de gris respectivamente. La técnica de filtrado se usa para eliminar el ruido o realzar el corte y pequeños detalles en una imagen. El filtro puede ser transportado fuera del dominio de la frecuencia espacial o en el dominio espacial.

Una imagen dada puede ser descompuesta en series de Fourier. Los componentes de baja frecuencia corresponden a objetos homogéneos y fondo, considerando que los componentes de baja frecuencia corresponden a cortes pequeños y detalles en la imagen. Los filtros paso bajas y paso altas son usados para

refinar la imagen o realzar los cortes y pequeños detalles. Duda y Hart (1973)<sup>7</sup> consideran el filtro paso bajas con la función de transferencia:

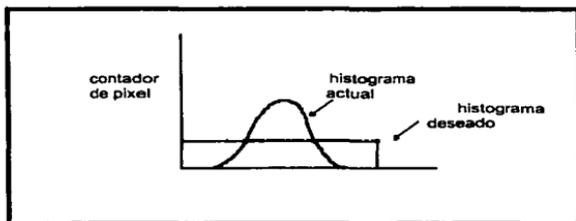


Figura 5.5 Igualación de histogramas

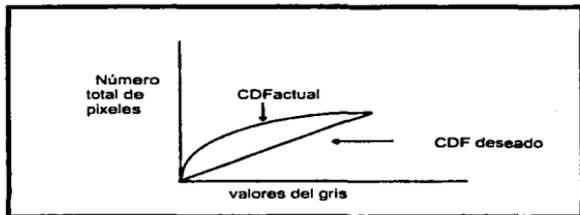


Figura 5.6 Función de acumulación distribuida de una imagen

$$H(u, v) = [(\cos \pi u)(\cos \pi v)]^\alpha \quad (5.3)$$

donde  $\alpha \geq 1$ ; y el filtro paso altas con la función de transferencia

<sup>7</sup> Duda, R.O., and Hart, P.E. (1973). *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York.

$$H(u, v) = 1 - [(\cos \pi u)(\cos \pi v)]^\alpha \quad (5.4)$$

donde  $\alpha \geq 1$ . En la ecuación 5.3 y 5.4,  $u$  y  $v$  representan frecuencia espacial en las direcciones  $x$  y  $y$  respectivamente.

El filtrado en un dominio espacial se puede representar como

$$g'(i, j) = C_1 [g_L(i, j)] + C_2 [g(i, j) - g_L(i, j)] \quad (5.5)$$

donde  $g'$  es la imagen realzada,  $g$  es la imagen de entrada original,  $C_1$  y  $C_2$  son constantes tales que  $C_1 \leq 1$  y  $C_2 \geq 1$ ,  $(i, j)$  corresponden al número de renglones y columnas respectivamente, de los píxeles, y  $g_L$  es el nivel de gris local. Y con  $C_1 = 1$  y  $C_2 = 0$ , la operación es de un simple refinado. Con  $C_1 \leq 1$  y  $C_2 \geq 1$ , los cortes y pequeños detalles son realzados. La imagen original y la versión paso altas obtenida por el filtro espacial se muestra en la Fig. 5.7a y b, respectivamente.

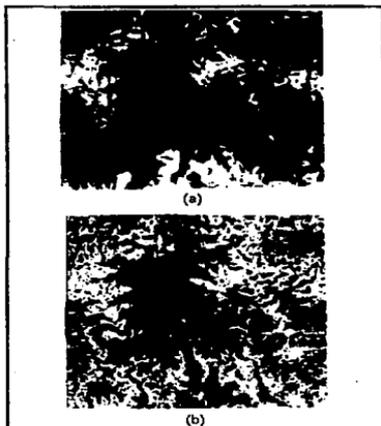


Fig. 5.7 Filtro paso altas

### 5.1.2 Técnicas para el realce de contorno

En el realce de contorno, intentamos hacer más visibles los contornos sin la imagen. La técnica de realce de contornos esencialmente incrementa la diferencia del nivel de gris entre dos regiones. Una señal de contorno y su derivada se muestra en la Fig. 5.8. La Línea de contorno y su derivada se muestran en la Fig. 5.9. Es obvio que para una señal de contorno, la primera derivada muestra un máximo en el punto del corte, no obstante para la línea de contorno la primera derivada es cero en el punto de corte. La segunda derivada en caso de una huella es cero. Mientras tanto, para la línea de contorno la segunda derivada es máxima en el punto de corte. Por lo tanto, los contornos en la imagen pueden ser detectados por los límites reparados en los valores de la primera y segunda derivada. Si la imagen es de dos dimensiones, las derivadas se consideran en ambas direcciones. La variación del nivel de gris es diferente para distintos tipos de contornos, y la presencia de ruido en la imagen puede dificultar detectar los cortes.

Para detectar los cortes en la imagen se pueden aplicar operadores como Laplace, gradiente. Esto se puede describir como:

$$g'(i, j) = \text{abs} [4 g(i, j) - g(i+1, j) - g(i, j+1) - g(i-1, j) - g(i, j-1)] \quad (5.6)$$

$$g'(i, j) = \left[ [g(i, j) - g(i, j-1)]^2 + [g(i, j) - g(i-1, j)]^2 \right]^{1/2} \quad (5.7)$$

$$g'(i, j) = \text{abs} [g(i, j) - g(i+1, j+1)] \quad (5.8)$$

En las ecuaciones 5.6, 5.7 y 5.8  $g(i, j)$  corresponden a los valores de gris de entrada a la imagen en  $(i, j)$ ;  $g'(i, j)$  corresponden a los valores de gris de la imagen de salida en  $(i, j)$ . En todo caso para determinar un pixel en  $(i, j)$  se emplean algunos límites de T:

si  $g'(i, j) \geq T$ , es un corte de pixel, sin embargo

si  $g'(i, j) < T$ , no es un corte de pixel.

Los operadores de corte pueden ser definidos en términos de máscaras. Considerando una región de tamaño de  $3 \times 3$  como se muestra en la Fig. 5.10, con  $g_0$  como el pixel central, los valores de gris en la región pueden ser representados por un vector  $g [g_0, g_1, \dots, g_8]$ . El peso del vector de coeficientes puede ser representado como  $w = [w_0, w_1, \dots, w_8]$ .

Los cortes pueden ser detectados reparando los límites en los valores de  $s$ , el cual esta dado por

$$s = \mathbf{g} \cdot \mathbf{w} = \sum_{i=1}^9 g_i w_i \quad (5.9)$$

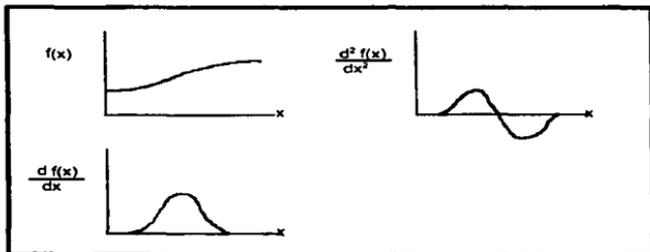


Figura 5.8 Etapas de un corte y sus derivadas

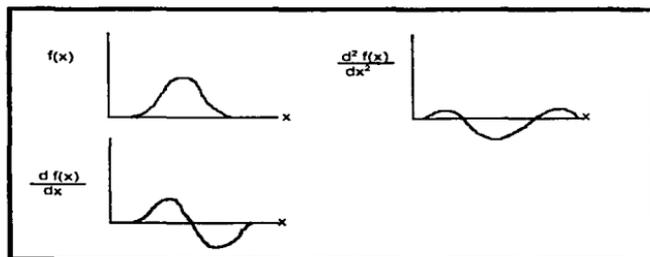


Figura 5.9 Una línea y sus derivadas

Diferentes mascarar pueden ser usadas para detectar cortes en diferentes direcciones. La mascarar para los operadores Sobel y Khrish se muestran en la Fig. 5.11 y 5.12, respectivamente. Varias mascarar se han propuesto para detectar cortes

y líneas en una superficie de una imagen en el lineamiento de cada pixel. Hueckel (1973)<sup>9</sup> trató la superficie usando una versión polar de la base ortogonal de Fourier. Heralick (1980)<sup>9</sup>, Morgenthaler y Rosenfeld (1981)<sup>10</sup> usaron polinomio ortogonal multidimensional como función básica.

$$\begin{bmatrix} g_1 & g_2 & g_3 \\ g_4 & g_5 & g_6 \\ g_7 & g_8 & g_9 \end{bmatrix}$$

Fig. 5.10 Matriz de una imagen de 3 X 3

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Fig. 5.11 Operador Sobel

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix}$$

Fig. 5.12 Operador Krisch

### 5.1.3 Modelos de R.N.A para percepción de brillo y detección de límite

<sup>9</sup> Hueckel, M. (1973). A local operator which recognizes edges and lines. Journal of Association of Computer Machinery 20:634-647.

<sup>9</sup> Heralick, R. M. (1980). Edge and region analysis for digital image data. Computer Vision and Graphics and Image Processing 12:60-73.

<sup>10</sup> Morgenthaler, D.G., and Rosenfeld, A. (1981). Multidimensional edge detection by hyper surface fitting. IEEE Transaction on Pattern Analysis and Machine Intelligence 3:482-486.

Grosberg y Todorovic (1988)<sup>11</sup> han considerado propiedades de un sistema visual humano y han propuesto una arquitectura para percepción de brillo bajo constantes y variables de condiciones de iluminación. La arquitectura consiste de un sistema contrario de características (FCS) y un sistema contrario de límite (BCS). Una respuesta fundamental en una percepción visual y el diseño de un sistema de visión real, es como el sistema visual transforma la distribución de entrada de iluminación a la distribución de brillo percibido.

La transformación de iluminación en brillo es un proceso complejo. Para limitar un sistema visual humano, necesitamos considerar dos fenómenos: constancia de brillo y contraste de brillo. Los fenómenos de constancia de brillo puede ser explicado por el factor de que dos piezas de papel idénticas se ven con brillo iguales aún cuando una esta bien iluminada y la otra sombreada.

El sistema visual toma importancia en la variabilidad de iluminación. El fenómeno de contraste de brillo es por el factor de que una misma pieza de papel gris se ve con más brillo si esta sobre un fondo negro que sobre un fondo blanco. El sistema de límite contrario es una segmentación y el proceso de complejión de límite basado en diferente iluminación a través de la imagen. Como lo propuesto por Grosberg y Mingolla (1987)<sup>12</sup>, BCS contiene dos subsistemas sucesivos: un filtro de contraste orientado (OC) y un lazo o vuelta de cooperativa competitiva (CC). El filtro OC obtiene la amplitud, dirección y localización de un gradiente de iluminación en una imagen. El subsistema CC es una red neuronal de 4 capas que permite la interacción entre el campo de datos receptor orientado. Van Allen y Kolodzy (1988)<sup>13</sup> han desarrollado software para simular el BCS en una computadora de objetivo general. Ellos han considerado dos imágenes: el cuadro Kanizsa y la figura Ehrenstein. La imágenes antes y después de procesadas se muestran en la Fig. 5.13 y 5.14. El contraste visible producido por el BCS en donde no se localiza contraste de iluminación existe en la imagen original.

Recientemente, Carpenter y otros (1989)<sup>14</sup> desarrollaron una arquitectura de red neuronal llamada el filtro CORT- X para segmentación de límite en la presencia de ruido. El elemento básico de sus modelos es una célula sensitiva que detecta la dirección y cantidad de contraste entre dos niveles de gris. La salida de una célula simple esta dada por:

$$\text{Salida} = \max [ L_s(x, \kappa) - \alpha_s R_s(x, \kappa) - \beta_s, 0 ] \quad (5.10)$$

<sup>11</sup> Grosberg, S. y Todorovic, D. (1988). A neural network architecture for brightness perception under constant and variable illumination conditions.

<sup>12</sup> Grosberg, S., and Mingolla, E. (1987). Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading. *International Journal of Computer Vision, Graphics and Image Processing* 37:116-165.

<sup>13</sup> Van Allen, E. J., and Kolodzy, P. J. (1988). Application of a boundary counter neural network to illusion and infrared sensor imagery. *Proceedings of International Joint Conference on Neural Network, San Diego*, pp. 193-202.

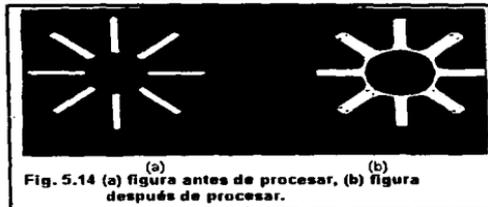
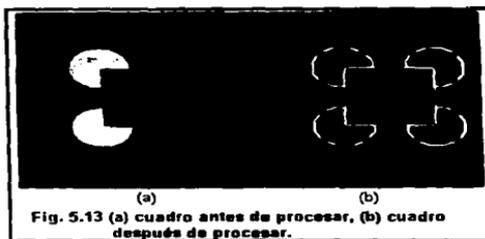
<sup>14</sup> Carpenter, G. A., et al. (1989). Invariant recognition of cluttered scenes by a self organizing ART architecture: CORT-X boundary segmentation. *Neural Networks* 2:169-181.

donde

$$L_s(x, \kappa) = \int_{\text{media izquierda}}^{\text{media}} I(y) dy / \int_{\text{media izquierda}}^{\text{media}} dy$$

$$R_s(x, \kappa) = \int_{\text{media izquierda}}^{\text{media}} I(y) dy / \int_{\text{media izquierda}}^{\text{media}} dy$$

donde  $x$  representa la posición del campo central receptor,  $\kappa$  es la orientación del campo receptor,  $\alpha_s$  es un parámetro de contraste ( $1 < \alpha_s$ ),  $\beta_s$  es un parámetro límite tal que  $0 < \beta_s < 1$ ,  $L_s(x, \kappa)$  es la activación total de la media izquierda del campo receptor,  $R_s(x, \kappa)$  es la activación total de la media derecha del campo receptor, e  $I(y)$  representa la iluminación en  $y$ . La célula simple descrita responde a pasos de luminancia, diferencia de densidades, cambio gradual en luminancia y otros cambios de contraste.



El siguiente paso de procesamiento es el detector que es sensitivo a la orientación, cantidad y escala espacial de contraste en una imagen dada pero no a la dirección de contraste. Este tipo de detector puede ser comparado a una célula compleja en la corteza visual. La respuesta  $C_c(x, \kappa)$  de una célula compleja centrada en la posición  $x$  y orientación  $\kappa$  puede ser obtenida usando un par de células simples con la misma dirección pero de contraste opuesto.

$$C_c(x, \kappa) = \max [ L_s(x, \kappa) - \alpha_s R_s(x, \kappa) - \beta_s, 0 ] + \max [ R_s(x, \kappa) - \alpha_s L_s(x, \kappa) - \beta_s, 0 ] \quad (5.11)$$

Un orden para detectar áreas en la presencia de ruido, se pueden usar filtros de diferentes tamaños. Una célula simple y compleja se muestra en la Fig. 5.15. Carpenter y otros (1989) ha demostrado que los filtros más pequeños hacen un mejor trabajo de localización de límite que un filtro más grande, mientras que el segundo hace un mejor trabajo de supresión de ruido. Han inventado un método para combinar respuestas de ambos filtros, en orden para lograr las propiedades deseables de ambos. La interacción es definida por:

$$B_{12}(x) = D_1(x) \sum D_2(y) U(y, x) \quad (5.12)$$

En la Eq. 5.12 el núcleo excitatorio  $U(y, x)$  es el próximo núcleo tal que  $\sum U(y, x) = 1$ . El factor  $D_1(x)$  y  $D_2(y)$  representan la supercompleja activación de patrones correspondiente al filtro más pequeño y al filtro más grande respectivamente. El factor  $D_1(x)$  es generalmente positivo en el límite; así exactamente localiza segmentos de límite y suprime el ruido cerca de los límites.

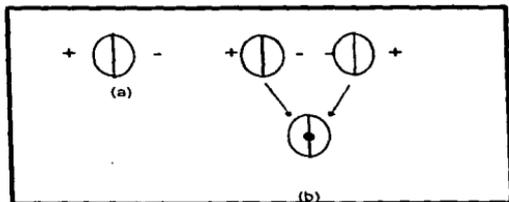


Fig. 5.15 Células simple y compleja

El factor  $D_2(x)$  suprime el ruido lejos de los límites. El software para simular el filtro **CORT-X** ha sido desarrollado y probado con una imagen. El resultado se muestra en la Fig. 5.16a, esta fig. representa la imagen original; la Fig. 5.16b 5.16c muestra la salida obtenida por filtros de 3X3 y 5X5, respectivamente. La respuesta combinada de ambos filtros se muestra en la Fig. 5.16d.

Dhawan y Dufresne (1990)<sup>15</sup> han sugerido una misma organización de redes neuronales que realzan y restauran el nivel de gris de la imagen para aplicación en procesamiento de imágenes en bajo nivel. El modelo consiste de una red neuronal multicapa, llamada "neurona piramide" para la representación y reconocimiento de la imagen con un origen específico tal como un pixel, una etapa o una región. El resultado de la primera capa en el modelo, se presenta aquí. Cada pixel es representado por una neurona con un estado  $o(j, \kappa) = 1, 2, \dots, N$ ; el cual es igual al valor del nivel de gris del pixel para  $j = 1, 2, \dots, J$ ;  $\kappa = 1, 2, \dots, K$ , donde  $J \times K$  es el tamaño de la imagen. Cada unidad es conectada con otra unidad cercana con el conector de orden  $m$ -th. La fuerza de esos conectores es proporcional al contraste entre pixeles.

Las capas de la red analizan el estado de cada unidad con respecto a la fuerza de las conexiones bajo un ambiente local estrecho homogéneo. Algunos cambios en los valores del estado unitario establece un foco de atención de análisis y pruebas para propagar el cambio a la unidad cercana. Los estados de unión son adaptados, y el proceso es iterado hasta que converja el criterio puesto por la función de evaluación de la imagen. Las conexiones entre dos unidades son definidas como valores de contraste, los cuales son calculados como:

$$W_m(i, j, \kappa, l) = [f(i, j) - f(\kappa, l)]/N \quad (5.13)$$

donde  $m=1$  corresponde al primer orden de conexión y  $m=2$  corresponde al segundo orden de conexión,  $N$  es el número del nivel de gris en la imagen, y  $f(i, j)$  corresponde al valor de gris del pixel  $(i, j)$ .

Dhawan y Dufresne (1990) han considerado el primer y segundo orden de conexión de contraste. La primera conexión de contraste es definida como la más cercana, mientras que la segunda es definida como los dos pixeles sobre el pixel central. El estado inicial de unidad es puesta igual a los valores del nivel de gris de los pixeles respectivamente. Para empezar con una unidad, esta es seleccionada al azar. El estado de la unidad es analizado en base a las conexiones de peso de más intensidad. Las conexiones de peso de más intensidad se definen como:

<sup>15</sup> Dhawan, A. P., and Dufresne, T. (1990). Low level image processing and edge enhancement using a self organizing neural network. Proceedings of International Joint Conference on Neural Networks. San Diego, Vol. Y, pp. 503-510.

$$w_i(i, j) = \alpha \sum_{\kappa, l \in R_1} w_{1,i}(i, j, \kappa, l) + \beta \sum_{\kappa, l \in R_2} w_{2,i}(i, j, \kappa, l) \quad (5.14)$$

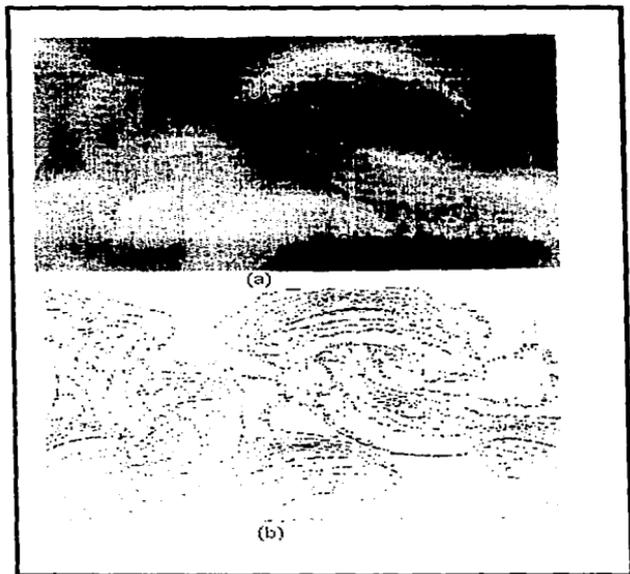


Fig. 5.16 Procesamiento de una imagen por un filtro CORTE-X.

- (a) imagen original
- (b) salida de un filtro CORTE-X de 3 X 3
- (c) salida de un filtro CORTE-X de 5 X 5
- (d) respuesta combinada.

donde  $R_1$  y  $R_2$  representan el primer y segundo orden de regiones, respectivamente, y  $\alpha$  y  $\beta$  son las constantes. La suma de las conexiones de peso de más fuerza  $w_i(i, j)$  es comparada con un límite  $T$  para poner la unidad de estado.  $w_i(i, j)$  es comparada con un límite  $T$  para poner a tiempo la unidad de estado. Si  $w_i(i, j)$  es mayor que el límite  $T$ , entonces la unidad de estado  $(i, j)$  es cambiada sumando o restando 1 al valor de gris.

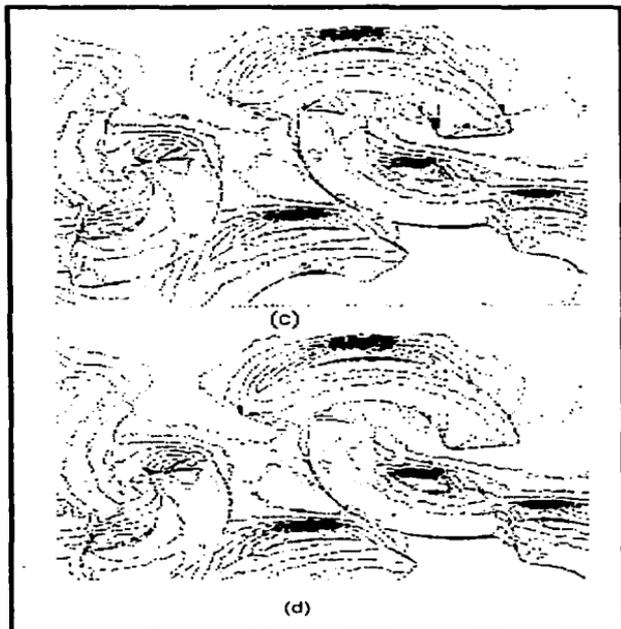
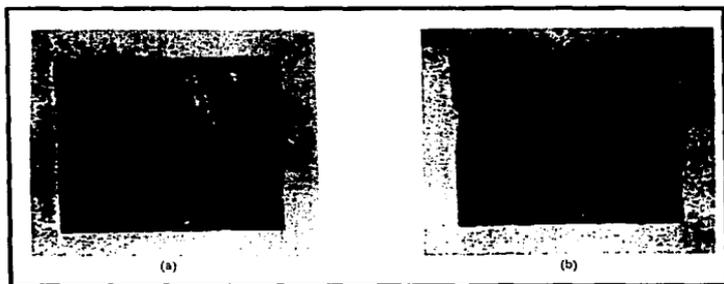


Fig. 5.16 (continuación)

Se puede notar que los valores altos de  $w_{(i, j)}$  corresponden a las regiones no homogéneas o cortes en la imagen. Sumando o restando incrementos de 1, el contraste en tal región es no homogéneo. Este procedimiento es esencialmente similar al filtro paso-altas en el dominio del espacio.

Dhawan y Dufresne (1990) asignan un estado inhibitorio, una posibilidad de indicar la probabilidad de poner a tiempo la unidad de estado. En las siguientes todas las unidades se activan, como el estado inhibitorio de todas las unidades son puestas en 0. Después de unas cuantas iteraciones el estado inhibitorio de algunas unidades llega a ser 1. El estado inhibitorio de cada unidad es cambiado tal que el foco de atención es extendido de una área activa a una área inactiva.

Para detener el proceso iterativo, se ha usado un criterio selectivo influenciado por la homogeneidad local. Ellos seleccionaron tres regiones especialmente distribuidas con mínima variación de contraste que están bien separadas. El cambio en variación de esas regiones seleccionadas es calculada y normalizada. Si las variaciones combinadas exceden tres tiempos en la medida de variación inicial, las iteraciones son detenidas. El resultado de la última iteración es utilizado como una solución final, ellos probaron el algoritmo con una imagen. Las imágenes original y realzada se muestran en las Fig. 5.17a y 5.17b respectivamente.



5.17 Realce de una imagen por un modelo de R.N.A. (a) imagen original, (b) imagen realzada por un modelo de R.N.A.

#### 5.1.4 Técnicas para eliminar el ruido

Si una imagen contiene ruido, se usan técnicas de afinación para limpiar dicho ruido. Sin embargo algunas técnicas de afinación empañan la imagen. Así las etapas de realce pueden ser necesarias después. La simple técnica de afinación es igual a un peso promediando las cercanías de un pixel. Puede ser expresado como

$$g'(i, j) = \sum_{p=-m}^m \sum_{q=-n}^n w(p, q)g(i-p, j-q) \quad (5.15)$$

donde los coeficientes de peso  $w(p,q)$  están dados por

$$w(p,q) = 1/(2m+1)(2n+1) \quad (5.16)$$

La Eq. 5.15 reemplaza el valor de gris en  $(i, j)$  por un valor de gris promediado sobre una cercanía rectangular  $(2m+1) \times (2n+1)$  circundante  $(i, j)$ , y  $g'(i, j)$  representa la imagen de salida. Esta simple técnica de promedio elimina el ruido. Sin embargo, empaña la imagen. Para reducir los efectos de empañamiento, varias y distintas técnicas de pulido son sugeridas. Estas son

$$w = \begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 1.0 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} \quad (5.17)$$

$$w = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (5.18)$$

La desigualdad ponderada desvanece el ruido de la señal en la imagen de salida y causa menos empañamiento que la simple técnica promedio.

Hay varios métodos que hacen uso de propiedades estáticas del pixel. Yasuoka y Haralick (1983)<sup>16</sup> han propuesto un esquema usando un modelo de faceta inclinada, con un ensayo t para limpiar el ruido de sal y pimienta. En su modelo lineal estocástico, el nivel de gris de un pixel se puede expresar como:

$$g(i, j) = \alpha y + \beta z + \gamma + \rho \epsilon(i, j) \quad (5.19)$$

<sup>16</sup> Yasuoka, Y., and Haralick, R.M. (1983). Peak noise removal by a facet model. *Pattern Recognition* 16:23-29.

donde  $i$  representa el número de renglones,  $j$  es el número de columnas,  $\varepsilon$  representa el ruido independientemente distribuido idénticamente con desviación estándar  $\rho$ , y  $\alpha$ ,  $\beta$ ,  $\gamma$  y  $\rho$  son los parámetros del modelo.

Cada pixel para ruido considera un acercamiento de 3X3. El modelo es apropiado para un modelo de 3X3. Los factores  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{\gamma}$  y  $\hat{\rho}$  son encontrados por la función de criterio  $J$ , el cual en este caso es el total de error cuadrático para el bloque  $\Omega$ .

$$J = \sum_{i,j \in \Omega} [g(i, j) - \alpha i - \beta j - \gamma]^2 \quad (5.20)$$

y minimizando  $J$  con referencia a  $\alpha$ ,  $\beta$  y  $\gamma$  se obtiene

$$\hat{\alpha} = \frac{\sum_{i,j} i g(i, j)}{\sum_{i,j} i^2} \quad (5.21)$$

$$\hat{\beta} = \frac{\sum_{i,j} j g(i, j)}{\sum_{i,j} j^2} \quad (5.22)$$

$$\hat{\gamma} = \frac{\sum_{i,j} g(i, j)}{\sum_{i,j} 1} \quad (5.23)$$

En todas las sumatorias,  $i$  y  $j$  varían de  $-1$  a  $+1$ , de esos estimados  $\rho$  es encontrado como:

$$\hat{\rho}^2 = \left[ \sum_{i,j} g(i, j) - \hat{\alpha} i - \hat{\beta} j - \hat{\gamma} \right]^2 / N \quad (5.24)$$

donde  $N$  es el número de elementos en el block (en este caso  $N = 9$ ). El valor de gris estimado del pixel puede ser expresado como:

$$g(i, j) = \hat{\alpha} i + \hat{\beta} j - \hat{\gamma} \hat{\rho} \varepsilon(i, j) \quad (5.25)$$

El ensayo  $t$  es usado para probar la hipótesis  $H_0: g(i, j) = \hat{g}(i, j)$ ; i.e., el valor estimado y el valor actual son iguales, y por lo tanto  $g(i, j)$  no es un pixel con ruido. Aquí  $t'$  es definido como:

$$t' = [g(i, j) - \hat{g}(i, j)] / \hat{\rho} / N$$

Para  $N = 9$  y  $\rho = \hat{\rho}$ , el valor límite de  $t'$  es tomado como  $t(N - 1, 0.05)$  usando un nivel de confianza del 95 %, y el valor de  $t$  puede leerse de tablas. Si  $t' < t(N - 1, 0.05)$ , aceptamos la hipótesis  $H_0$ ; i.e.,  $g(i, j)$  no es elemento de ruido y no es reemplazable. Si  $t' \geq t(N - 1, 0.05)$ , rechazamos la hipótesis  $H_0$ ; i.e.,  $g(i, j)$  es un elemento de ruido, y lo reemplazamos por  $\hat{g}(i, j)$ . El esquema es dócil para iteraciones. Una simulación de software de dicho algoritmo se ha desarrollado e implementado, el algoritmo se aplica a una imagen de prueba. La imagen original con ruido se muestra en la Fig. 5.18a, y la imagen después del filtro se muestra en la Fig. 5.18b (Kulkarni, 1986).

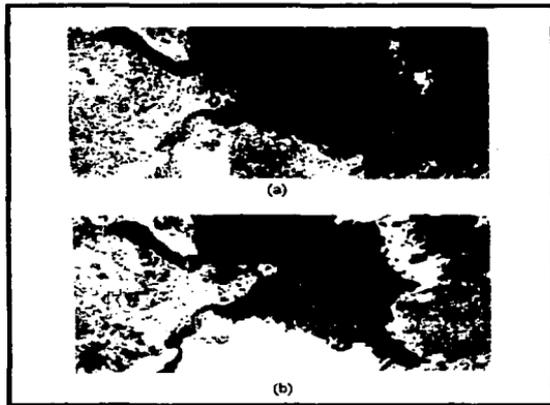


Fig. 5.18 Eliminación de ruido. (a) imagen del planeta tierra, con ruido, (b) imagen del planeta tierra, sin ruido.

### 5.1.5 Restauración de Imágenes

Imagine que un sistema en la práctica es perfecto y las imágenes obtenidas se degradan. Las técnicas de restauración se aplican para compensar la degradación de sistemas tal como lo es el empañamiento, turbulencia atmosférica y difracción óptica. Algunos métodos para la restauración de imágenes se han sugerido en la práctica.

Esos métodos incluyen el significado de error de filtros, técnica de descomposición de valores y otros. En todos esos métodos los sistemas son modelados por un sistema lineal, y el sistema es caracterizado por su amplia función de puntos. Las técnicas de restauración básicamente tratan con inversión de los procesos degradados. La implementación del software de métodos para restauración convencional a menudo consumen tiempo y no son útiles para aplicaciones de tiempo real. Recientemente, modelos de R.N.A. para restauración de imágenes se han sugerido (Zhou y otros, 1988<sup>17</sup>; Kulkarni, 1991<sup>18</sup>; Farehat and Bai 1989<sup>19</sup>). El modelo sugerido por Zhou y otros (1988) usa la red de Hopfield, mientras que el modelo sugerido por Kulkarni (1990 a, 1991) es una red de seis capas y de alimentación hacia adelante (feed-forward), y esta basada en la singular técnica de descomposición de valores.

Algunos sistemas de imágenes pueden ser modelados en la práctica por una ecuación de integrales (Huang y otros, 1971):

$$g(x,y) + \varepsilon(x,y) = \int_a^b \int_a^b f(x',y') h(x,x'; y,y') dx' dy' \quad (5.26)$$

para  $a \leq x$ ,  $y \leq b$  donde  $f(x',y')$  es un objeto (imagen ideal), y  $h(x,x'; y,y')$  es la función de respuesta de impulso del sistema registrado. La imagen degradada esta dada por  $g(x,y)$ , y  $\varepsilon(x,y)$  representa el ruido adaptivo en la medición de  $g(x,y)$ .  $\varepsilon(x,y)$  puede representar el detector de ruido y/o redondear el error. Es posible reducir la Eq. 5.26 usando la búsqueda o la operación de la ecuación de integrales del primer tipo:

$$g(y) + \varepsilon(y) = \int_a^b f(x) h(x,y) dx \quad (5.27)$$

para  $a \leq y \leq b$ .

El problema de restauración se reduce para solventar la ecuación 5.27 para  $f(x)$ , donde  $g(y)$  y  $h(x,y)$  son conocidas. Para resolver la Eq. 5.27 numéricamente o para un modelo de R.N.A. es necesario hacer la variable discreta y remplazar la

<sup>17</sup> Zhou, Y.T., (1988). A novel approach to image restoration based on a neural network. Proceedings of International Joint Conference on Neural Networks, San Diego, Vol. 4, pp. 269-276.

<sup>18</sup> Kulkarni, A. D. (1991). Solving ill-posed problems with artificial neural networks. *Neural Networks* 4:477-484.

<sup>19</sup> Farehat, N. H., and Bai, B (1989). Echo inversion and target shape estimation by neuromorphic processing. *Neural Networks* 2:117-125.

ecuación de integrales por una ecuación integral finita. Las variables continuas  $x$  y  $y$  pueden ser reemplazadas por una maya finita de puntos.

$$a \leq y_1 \leq y_2 \dots \leq y_{n_1-1} \leq y_{n_1} \leq b$$

$$a \leq x_1 \leq x_2 \dots \leq x_{n_1-1} \leq x_{n_1} \leq b$$

donde  $n_1 \geq n_2$ . La ecuación de integrales entonces puede ser reemplazada por la ecuación

$$g(y_i) + \varepsilon(y_i) = \int_a^b h(y_i, x) f(x) dx = \sum_{j=1}^{n_2} d_j h(y_i, x_j) f(x_j) \quad (5.28)$$

donde  $i = 1, 2, \dots, n_1$ ,  $j = 1, 2, \dots, n_2$ , y  $d_1, d_2, \dots, d_{n_2}$  son los coeficientes de peso para la fórmula usada. Para escribir en orden la ecuación 5.28 en forma de matriz, definimos vectores  $f$ ,  $g$ ,  $\varepsilon$  y una matriz  $H$  como  $f = f(x_j)$ ,  $\varepsilon = \varepsilon(y_i)$ ,  $g = g(y_i)$ , y  $H = d_j h(y_i, x_j)$ . Con la notación mencionada, la ecuación 5.28 se puede escribir como:

$$Hf = g + \varepsilon \quad (5.29)$$

La solución del sistema de ecuaciones en la ecuación 5.29 por inversión directa es dada por:

$$f = H^* (g + \varepsilon) \quad (5.30)$$

donde  $H^*$  representa la inversa  $-g$  de  $H$ .

Kulkarni (1991) desarrolló un modelo de R.N.A. basado en el enfoque Descomposición de Valor Singular (SVD) en el cual el peso en la red es inicializada usando los componentes SVD. El "peso crítico" en la red es adaptado iterativamente para obtener una fina solución. La principal ventaja de la implementación de una red neuronal es que el modelo de la red es paralelo y puede ser implementado electrónicamente o electrópticamente. El modelo de la R.N.A. puede ser usado con un sistema de registro para obtener la imagen restaurada.

El modelo sugerido por Kulkarni (1991) es una red lineal de alimentación hacia adelante como se muestra en la fig. 5.1, esta red lineal es una base para comprender el funcionamiento de las redes no lineales. El modelo hace uso de la unidad sigma así como de la unidad pi. Si la entrada a la red para una unidad es la suma de todas las entradas de peso, entonces la unidad es llamada unidad sigma. Otro tipo de unidad es la unidad pi. En la unidad pi la entrada de la red de varias unidades es un

producto de todas las entradas de peso. En algunas situaciones, tal como calculando coeficientes de autocorrelación, la unidad pi es muy usada.

Considerando la ecuación 5.30; usando la descomposición singular de valor (SVD) de la matriz  $H$ , tenemos:

$$f = U [ D, O ] V^T (g + \varepsilon) \quad (5.31)$$

donde las columnas de  $U$  son vectores de  $H^T H$  los renglones de  $V$  son los vectores de  $H H^T$ ,  $D = \text{diag} [ 1/\mu_1, 1/\mu_2, \dots, 1/\mu_n ]$ , donde  $\mu_i$  es el valor singular de  $H$ . en la fig. 5.19, las capas de conexión de pesos  $L_1, L_2, L_3, L_4, L_5, L_6$  son inicializadas usando los elementos de matriz  $V^T, [ D, O ]$  y  $U$ , respectivamente. La entrada a la red es el vector de observación  $g$ . El vector de salida  $f$  obtenido en la capa  $L_4$  esta dado por:

$$f_i = \sum_{j=1}^{n_2} U_{ij} \beta_j w_j \quad (5.32)$$

donde  $\beta_j$  es elemento del vector  $\beta$ , el cual es dado por  $\beta = V^T (g + \varepsilon)$ , y  $w_j = 1/\mu_j$

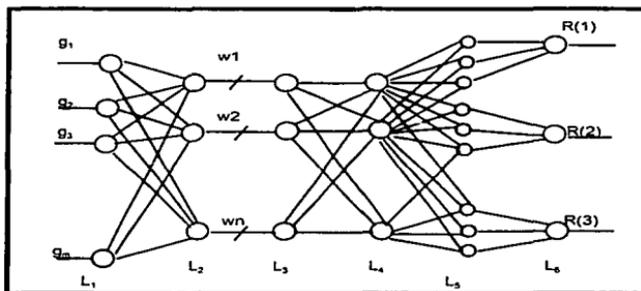


Fig. 5.19 Modelo de Red Neuronal Artificial para restauración de imágenes

El componente en la ecuación 5.32 difiere de sus valores verdaderos porque la presencia de ruido  $\epsilon$  en la medida de  $g$  se puede ver en la ecuación 5.32 que la solución de  $f$  es sensitiva al error en valores  $\beta$  que corresponde a valores pequeños; mientras que la solución es sensitiva a errores en componentes  $\beta$  que corresponden a grandes valores singulares. La distinción entre un valor singular grande y uno pequeño se puede hacer como sigue. Resolviendo la ecuación 5.30, el radio  $\phi_1$  de la señal de ruido en la solución  $f$ , y el radio  $\phi_2$  de la señal de ruido en la medida de  $g$ , es relacionado por:

$$\phi_1 \leq (1/C) \phi_2 \quad (5.33)$$

donde la condición número  $C$  es el radio del valor singular de  $H$  del más grande al más pequeño. En una situación física, dependiendo del límite más bajo aceptado en  $\phi_1$  ( $\phi_{1min}$ ) y precisión en medida de  $g$  (i.e.,  $\phi_2 = \|g\| / \|e_2\|$ , como detector de ruido, truncación y redondeo de errores, etc.), el valor singular  $\mu_1$  para el cual la desigualdad  $\phi_{1min} \leq (\mu_{max} / \mu_1) \phi_2$  se cumple considerando un valor singular pequeño, y el correspondiente peso  $w_1 = 1/\mu_1$  es considerado peso crítico.

La autocorrelación estimada o la capa de salida  $L_e$  esta dada por:

$$R(k) = \sum_{i=1}^{n_2-k} f_i f_{i+k} \quad (5.34)$$

donde  $k = 1, 2, \dots, K$ . Si comparamos la función estimada de autocorrelación  $R(i)$ , obtenemos un sistema de ecuaciones dado por:

$$R(i) - \hat{R}(i) = 0 \quad (5.35)$$

para  $i = 1, 2, \dots, K$ . Usando la ecuación 5.32 y 5.34 podemos escribir la ecuación 5.35 como:

$$F_i (w_{n_2 - m + 1}, \dots, w_{n_2 - 1}, w_{n_2}) = 0 \quad \text{para } i = 1, 2, \dots, K \quad (5.36)$$

donde  $F_i$  representa una ecuación cuadrática, peso crítico  $m$ . El peso crítico ha sido considerado por ser desconocido.

Para resolver el sistema en la ecuación 5.36, considere la función:

$$S ( w_{n_2}, \dots, w_{n_2-1}, w_{n_2} ) = \sum_{i=1}^K [ F_i ( w_{n_2}, w_{n_2-1}, \dots, w_{n_2-1}, w_{n_2} ) ] \quad (5.37)$$

La solución del sistema de ecuaciones en la ecuación 5.37 reduce la función S a cero; el valor de peso para el cual el error es minimizado son las raíces del sistema de ecuaciones en la ecuación 5.36. La función S representa el error cuadrático medio en la función de autocorrelación. El peso crítico en el modelo es ajustado tal que el error cuadrático medio en la función de autocorrelación es minimizado. El algoritmo para actualizar el peso se describe en los siguientes pasos ( Kulkarni, 1991):

1. Inicializar los pesos usando valores y vectores de la matriz **H**.
2. Propagar el vector de entrada **g** para obtener la solución **f** y autocorrelacionar coeficientes **R(i)**.
3. Obtener el error cuadrático medio como:

$$\Delta E = \sum_{i=1}^K [ R(i) - \hat{R}(i) ]^2 \quad (5.38)$$

4. El cambio en el peso crítico esta dado por:

$$\Delta w_i \begin{cases} \alpha \Delta E w_i & \text{si } | \Delta w_i | \leq | w_i | \\ 0 & \text{de otra forma} \end{cases} \quad (5.39)$$

donde  $\alpha$  es una constante

5. Actualizando el peso:

$$w(t+1) = w(t) - | \Delta w | \quad \text{para toda } i \quad (5.40)$$

6. Repetir el paso 2 al 5 hasta que el error  $\Delta E$  llegue a ser más pequeño que la constante  $E_{min}$  o el cambio en el peso  $\Delta w$  llegue a ser cero.

La constante  $\alpha$  en la ecuación 5.39 es representada tanto que el peso representa el valor singular más pequeño a el cual la solución sensitiva es reducida; el peso crítico que contribuye más a el error es reducido a cada iteración. Esto causa la solución para ser más fina con cada iteración. El algoritmo termina cuando el error

cuadrático medio en la autocorrelación de coeficientes reduce a  $E_{\min}$ . En caso extremo el peso crítico llega a ser cero, y la red produce la solución más exactas. Así el algoritmo es garantizado para soluciones exactas.

Kulkarni (1991) desarrolló una simulación de software del modelo y ha considerado dos imágenes: la imagen representando círculos y la otra obtenida de datos del satélite. El censor usado para obtener la imagen del satélite es llamado el mapeador temático, el cual provee datos en siete bandas espectrales. Las imágenes fueron empañadas usando el núcleo dado por:

$$K(z) = \begin{cases} 1 + \cos(2\pi z/3) & \text{para } z \leq 3 \\ 0 & \text{de otra forma} \end{cases} \quad (5.41)$$

La gráfica para la ecuación 5.41 se muestra en la fig. 5.20. Se asume que la distorsión esta sólo en la dirección  $x$ . Esto puede ser logrado orientado adecuadamente el sistema de coordenadas. La imagen original se muestra en las fig. 5.21a y 5.22a. La imagen degradada se muestra en las fig. 5.21b y 5.22b. Las imágenes obtenidas por inversión directa se muestran en las fig. 5.21c y 5.22c. Las imágenes restauradas por el modelo de R.N.A. se muestran en las fig. 5.21d y 5.22d. Las imágenes fueron restauradas línea por línea. Medidas del núcleo o el punto de medida para la degradación de la función tal como la degradación del movimiento, turbulencia atmosférica, o aberraciones del lente se presentan a menudo en la práctica. Además medidas de función de autocorrelación para varias clases de imágenes se presentan en la practica. La exactitud de la solución depende del ruido de la señal de radio del vector  $\mathbf{g}$ , la exactitud de la correlación de coeficientes verdaderos, y el número de condición de la degradación del núcleo.

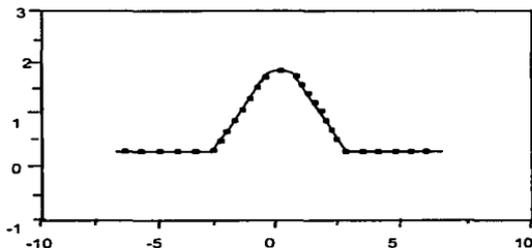


Fig. 5.20 Función núcleo

El ruido en  $g$  incluye el registro de ruido medio, redondeo y truncación de errores. Además desde que la solución  $f$  es de tamaño finito, los coeficientes de correlación son conocidos con una exactitud finita. La arquitectura sugerida puede ser implementada electrónicamente o electropicamente, y de gran importancia es que puede ser usada para aplicaciones de tiempo real.

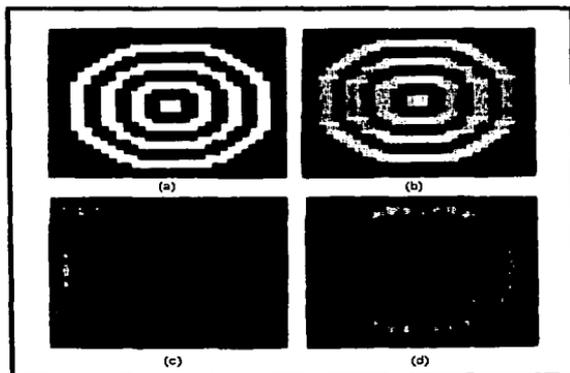


Fig. 5.21 Restauración de Imagen

- a) Imagen original  
 b) Imagen degradada  
 c) Imagen restaurada por inversión directa  
 d) Imagen restaurada por un modelo de R.N.A.

donde:

- $C_i$  = capacitancia de unidad  $i$   
 $\mu_i$  = potencial de entrada  
 $T_{ij}$  = fuerza sináptica entre  $i$  e  $j$   
 $v_i$  =  $g(\mu_i)$   
 $g(\cdot)$  = función sigmoial, describe el cambio del voltaje en la salida  
 $R_i$  = Resistencia interna  
 $I_i$  = Corriente externa

Una alternativa de enfoque para restauración de imágenes es usar el criterio de error cuadrático medio. El método más directo para resolver la ecuación 5.30 es encontrar la solución al cuadrático medio que minimiza la función de error  $E$  dado por:

$$E = \frac{1}{2} \|g - Hf\|^2 \quad (5.42)$$

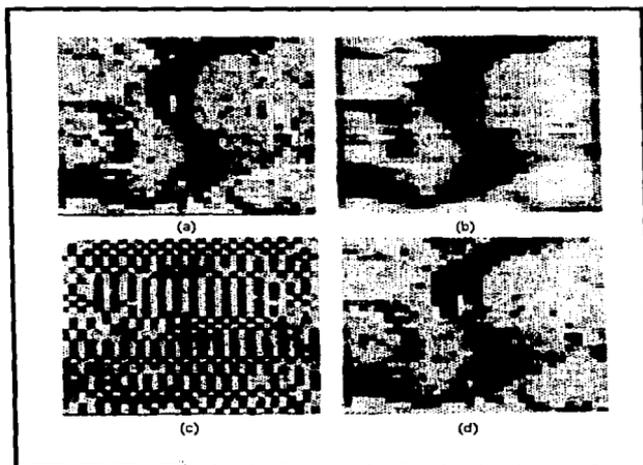


Fig. 5.22 Restauración de imagen

- a) imagen original (dato de satélite)
- b) imagen degradada
- c) imagen restaurada por inversión directa
- d) imagen restaurada por un modelo de R.N.A.

Zout y otros (1988) han usado una red Hopfield para obtener la solución por minimización de la función  $E$ . Para hacer esto en orden, es necesario mapear la función  $E$  a la función de la red Hopfield. Una red Hopfield consiste de una sola capa, como se muestra en la fig. 5.23. Cada unidad tiene una simple función de activación

que suma las entradas de otras unidades. Una red Hopfield puede ser descrita por un sistema de ecuaciones diferenciales no lineales ( Hopfield y Tank, 1985 ).

$$C_i (d\mu_i / dt) = \sum_{j=1}^N T_{ij} v_j - \mu_i / R_i + I_i \quad \text{para } i = 1, 2, \dots, N \quad (5.43)$$

Una importante propiedad de la red Hopfield es que dada una matriz de conectividad simétrica ( i.e.,  $T_{ij} = T_{ji}$ ,  $T_{ii} = 0$  ), el vector de salida  $v = v_1, v_2, \dots, v_n$  converge a un estado estable. La función de energía para la red Hopfield es dada por:

$$E = - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N T_{ij} v_i v_j - \sum_{i=1}^N I_i v_i \quad (5.44)$$

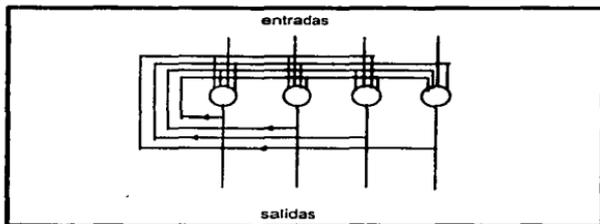


Fig. 5.23 Red Hopfield

Para resolver en orden el problema de restauración, necesitamos mapear la función de error en la ecuación 5.42 a la función de energía en la ecuación 5.44 de la red de Hopfield. Los parámetros de la red tales como la fuerza de interconexión y entradas predispuestas son determinadas en términos de función de energía de la red. Zhou y otros (1988) ha considerado este mapeo. Su modelo consiste de  $L^2 \times M$  de unidades conectadas mutuamente, donde  $L$  es el tamaño de la imagen y  $M$  es el valor máximo de la función del nivel de gris. Su modelo se explica como sigue:

Permite que  $v = \{ v_{i,k}, 1 \leq i \leq L^2, 1 \leq k \leq M \}$  sea un estado binario, pone la red neuronal con  $v_{i,k}$  denotando el estado de la unidad  $(i,k)$ . Permite que  $T_{i,k;j,l}$  denote la fuerza de interconexión entre la unidad  $(i,k)$  y la unidad  $(j,l)$ .

$$T_{i,k;j} = T_{j,i;i,k} \quad (5.45)$$

para  $1 \leq i, j \leq L^2$  y  $1 \leq i, k \leq M$ . Zhou y otros (1988) ha asumido que la unidad tiene realimentación. La imagen de entrada es representada por una función finita de nivel de gris  $\{x(i, j); 1 \leq i, j \leq L\}$ , y la función del nivel de gris es dada por una simple suma de un estado variable de neuronas.

$$x(i, j) = \sum_{k=1}^M v_{m,k} \quad (5.46)$$

donde  $m = i \times L + j$

si rastreamos la imagen a 2 dimensiones por renglones y los juntamos como un vector, entonces el vector degradado puede ser representado como la ecuación 5.47.

$$y = Hx + c \quad (5.47)$$

donde  $H$  es la matriz  $L^2 \times L^2$  representando la función del punto de amplitud,  $x$  y  $c$  son el original vector degradado y de ruido, respectivamente. La función de energía de la red neuronal puede escribirse como:

$$E = \frac{1}{2} \sum_{i=1}^{L^2} \sum_{j=1}^{L^2} \sum_{k=1}^M \sum_{L=1}^M T_{i,k;j,L} v_{i,k} v_{j,L} - \sum_{i=1}^{L^2} \sum_{k=1}^M I_{i,k} v_{i,k} \quad (5.48)$$

comparando términos en la expansión de la ecuación 5.42 con los términos correspondientes en la ecuación 5.48, podemos determinar la fuerza de interconexión y las entradas predispuetas.

$$T_{i,k;j,L} = \sum_{p=1}^{L^2} h_{p,i} h_{p,j}$$

$y$

$$I_{i,k} = \sum_{p=1}^{L^2} y_p h_{p,i} \quad (5.49)$$

De la ecuación 5.49, se puede ver que la fuerza de interconexión se determina por la función de degradación. De aquí  $T_{i,k,j,l}$  puede ser calculado sin ningún error. Si la imagen es degradada por una función de transferencia, entonces  $I_{i,k}$  puede estimarse perfectamente. Una vez que los parámetros  $T_{i,k,j,l}$  e  $I_{i,k}$  se obtienen usando la ecuación 5.49, cada neurona puede libremente y asincrónicamente evaluar su estado y reajustarse usando la ecuación 5.43. Cuando uno de los puntos de energía es buscado, la imagen es reconstruida usando la ecuación 5.43. El algoritmo descrito es difícil para simular una computadora convencional por su alta complejidad. Zhou y otros (1988) han simplificado el algoritmo y han sugerido un método práctico. Los resultados de su experimento se muestran en las figuras 5.24a hasta la 5.24d.

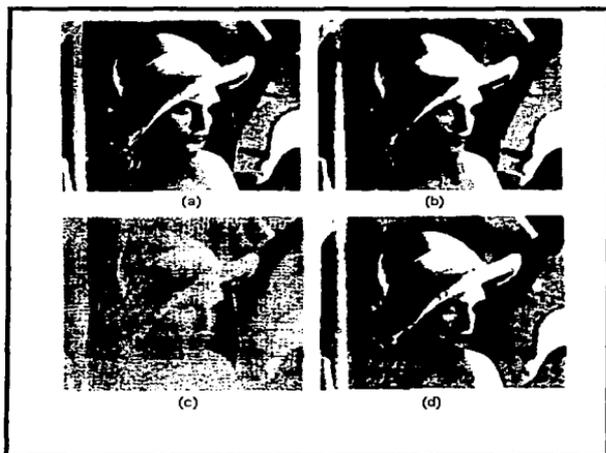


Fig. 5.24 Restauración de imagen usando una Red Hopfield. a) imagen original, b) imagen degradada, c) imagen restaurada por inversión directa, d) imagen restaurada por una R.N.A.

## 5.2 EXTRACCIÓN DE CARACTERÍSTICAS

Los sistemas de reconocimiento de patrones usualmente consideran un espacio de características en el cual el vector de observación es primero mapeado. El vector de características es usado entonces para decidir la clase a la cual pertenece el vector de observación. El propósito de extraer las características es reducir datos mediante ciertas "características" o propiedades que distinguen los patrones de entrada en la extracción de características transformamos un vector de observación de entrada a un vector de características usando algunas funciones básicas ortogonales o no ortogonales tal que los datos en el espacio de características no están correlacionados.

Para la extracción de características se han desarrollado una variedad de enfoques. Las técnicas de características de espacio usadas comúnmente incluyen la Transformada de Fourier (FT), momento del espacio de características, gráfica de distribución de espacio de características, la Transformada Hough, espacio de distribución de características Wigner, polinomios ortogonales, y funciones Gabor. En la comprensión de datos son también usadas las transformadas ortogonales como la Walsh, Walsh Hadaward, o coseno discreto (Ahmed y Rao, 1975)<sup>20</sup> para extracción de características.

Una vez obtenido el ajuste de las características que distinguen patrones de entrada, la tarea de reconocimiento se reduce a particionar el espacio de características. En la práctica se han implementado muchas técnicas de extracción de características con ayuda de sistemas digitales y/o ópticos. La principal ventaja de las técnicas digitales es la flexibilidad para operaciones no lineales. Las técnicas ópticas procesan píxeles en paralelo y la velocidad de proceso esta limitada por la velocidad de la luz. Los modelos de redes neuronales artificiales (ANN) son robustos y tolerantes a fallas.; pueden ser implementados electrónicamente o electroópticamente, y trabajan en paralelo.

Mediante el uso de modelos de ANN para la extracción de características es posible convinar las ventajas tanto del proceso óptico como digital.

Consideremos el problema de reconocimiento de objetos invariantes, y consideremos la invarianza con respecto a diferencias de translación rotación y escalamiento en imágenes de entrada. El problema de la invarianza es difícil debido a que el número de muestras de entrenamiento para el cual el clasificador necesita ser entrenado. Para aliviar esto, el problema de invarianza es a menudo mencionado en el estado de extracción de características. Hay un número de técnicas que pueden ser usadas para la extracción de características invariantes. El método de extracción de características que usa momentos invariantes produce las características que son

---

<sup>20</sup>Ahmed, N., and Rao, K.R. (1975). Orthogonal Transforms for Digital Signal Processing. Springer-Verlag, Berlin.

invariantes con respecto a las diferencias de translación, rotación y escalamiento (Hu, 1992)<sup>21</sup>.

Podemos además extraer características invariantes usando el espacio de características FT. La distribución FT bidimensional tiene algunas propiedades interesantes: (1) La magnitud de FT es invariante al desplazamiento, (2) Las frecuencias espaciales más altas de entrada corresponden a valores de amplitud más altos más aya del origen del plano FT, y (3) si la imagen de entrada se rota, la distribución en el plano FT también rota (Goodman, 1968)<sup>22</sup>. Estas propiedades de la FT y otras propiedades similares de otras transformadas se pueden usar para la extracción de características de invarianza. Además, durante la extracción de características de invarianza, es posible la reducción de datos significantes.

### 5.2.1 Extracción de características usando momento invariante

Uno de los métodos bien conocidos para extracción de características invariantes es el uso de momento invariante. Hu (1962) ha definido una serie de siete momentos invariantes que son invariantes a la translación, escala y diferencia rotacional en patrones de entrada. La principal desventaja de la técnica de momento invariante es que no hay garantía de los momento invariantes, los que en número son exactamente siete, de que formen un grupo completo de descriptores. Sin embargo, para mayores aplicaciones prácticas el grupo de siete invariantes es adecuado para distinguir entre patrones de entrada. Los descriptores Fourier forman un grupo completo de características; sin embargo, extracción de características son usualmente grandes en número y son inmanejables para usar en clasificaciones a menos que estén agrupadas por algún criterio de agrupación. Los siete momentos invariantes son descritos a continuación.

Digamos  $g(x,y)$  es una imagen de tamaño  $n \times n$ , el  $(p+q)$ ésimo momento geométrico para una imagen de tamaño  $(2n+1) \times (2n+1)$  esta dada por:

$$m_{pq} = \sum_{x=-n}^n \sum_{y=-n}^n x^p y^q g(x,y) \quad (5.50)$$

para  $p, q = 0, 1, 2, 3, \dots$

Para hacer los momentos invariantes para translación, podemos definir un momento central como:

<sup>21</sup>Hu, M.K. (1962). Visual Pattern recognition by moment invariants. IRE Transactions on Information Theory IT-8:28-32

<sup>22</sup>Goodman, J.W. (1968). Introduction to Fourier Optics. McGraw Hill. New York

$$m_{pq} = \sum_{x=0}^n \sum_{y=0}^n (x-x')^p (y-y')^q g(x,y) \quad (5.51)$$

Donde  $X' = m_{10} / m_{00}$  y  $Y' = m_{01} / m_{00}$ . Los momentos centrales en la ec. 5.51 pueden ser normalizados por escala de invariancia como se muestra abajo.

$$\mu_{pq} = m_{pq}^{\gamma} / m_{00} \quad (5.52)$$

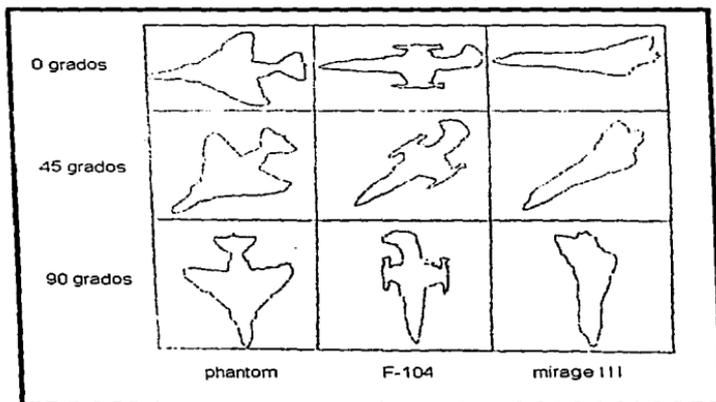
Donde:  $\gamma = (p+q) / 2 + 1$

Hu (1962) desarrolló las siguientes siete funciones de momentos centrales que son invariantes para rotacional y diferencias de escala.

$$\begin{aligned} \phi_1 &= (\mu_{20} + \mu_{02}) \\ \phi_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\ \phi_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\ \phi_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\ \phi_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12}) \cdot [(\mu_{30} + \mu_{12})^2 \cdot 3(\mu_{21} + \mu_{03})^2] \\ &\quad + (3\mu_{21} + \mu_{03})(\mu_{21} + \mu_{03}) [3(\mu_{30} + \mu_{12})^2 - (\mu_{12} + \mu_{03})^2] \\ \phi_6 &= (\mu_{20} - \mu_{02}) [(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\ \phi_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12}) [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &\quad - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03}) [(3\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \quad (5.53)$$

Kulkani y otros (1990) han usado estos momentos invariantes para identificación de aeromágenes. Ellos han considerado tres tipos de aviones El Phantom, F104, y Mirage III. Las imágenes originales e imágenes con rotación se muestran en la fig. 5.25. Las imágenes originales escaladas por el avión se muestran en la fig. 5.26. Su sistema de reconocimiento consiste de dos periodos: el periodo de extracción de características y el periodo de reconocimiento. En el periodo de extracción de características ellos obtuvieron los momentos invariantes que fueron usados como entrada para el periodo de reconocimiento. En el periodo de reconocimiento ellos usaron una red de tres-capas de alimentación-directa con un algoritmo de aprendizaje de retro-propagación. Los momentos invariantes son mostrados en la tabla 5.1

Durante el proceso de aprendizaje las imágenes con rotación de 0° fueron usadas para que la red aprenda. Para estabilizar la red, el proceso de aprendizaje fue 300 veces iterado. Durante la fase de haciendo decisión las imágenes rotadas y escaladas del avión fueron usadas como imágenes de entrada. El modelo bien trabajado para imágenes con diferencias rotacionales y traslacionales, sin embargo el modelo no funciona bien en el reconocimiento de objetos con escalas diferentes. Hu(1962) había mostrado analíticamente esos momentos invariantes descritos por la ec.5.53 son insensitivos a traslacionales, rotacionales, y diferencias en escalas de imágenes de entrada. Aunque, los valores de estos invariantes son pequeños, y ellos son muy sensitivos a errores redondeados y/o ruido, los que hacen difícil el reconocimiento. Una de las soluciones para el problema de sensibilidad es considerar los valores logarítmicos regulados de los momentos invariantes. En la práctica se proporciona uno de los métodos más comúnmente usados de momentos invariantes.



#### 5.25 Imágenes de avión en diferentes rotaciones

Khotanzad y Lu (1987)<sup>23</sup> habían usado momentos invariantes como características para el reconocimiento de caracteres.

<sup>23</sup>Khotanzad, A., and Lu, J.H. (1987) Distorted Invariant character recognizing by multi-layer perceptron and backpropagation learning

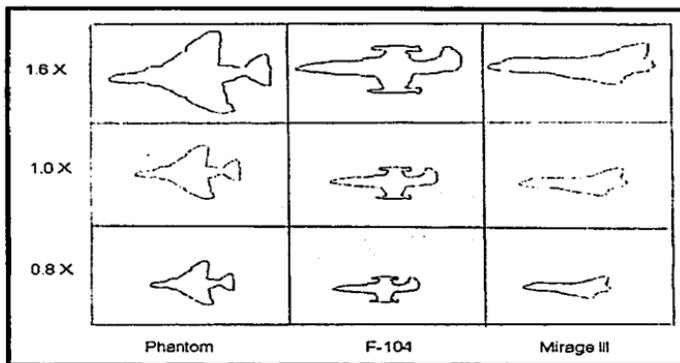


Fig. 5.26

TABLA 5.1 CARACTERÍSTICAS POR MOMENTO INVARIANTE

Transformación	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$	$\phi_5$	$\phi_6$
			F-1404			
0°	0.404	0.114	0.402	0.149	0.0361	0.043
45°	0.435	0.165	0.483	0.219	0.067	0.794
90°	0.404	0.114	0.420	0.149	0.361	0.043
Scale 2 x	0.4831	0.184	2.265	1.318	2.473	0.516
			Phantom			
0°	0.339	0.073	0.074	0.011	0.001	0.002
45°	0.429	0.121	0.192	0.047	0.004	0.016
90°	0.339	0.073	0.074	0.011	0.001	0.002
Scale 2 x	0.332	0.066	0.506	0.125	0.031	0.0325
			Mirage III			
0°	0.697	0.445	2.596	1.308	3.913	1.164
45°	0.362	0.683	3.400	2.351	6.632	1.372
90°	0.697	0.445	2.566	1.308	3.913	1.164
Scale 2 x	0.636	0.366	9.581	6.997	58.10	4.093

### 5.2.2 Extracción de características usando transformadas ortogonales

En muchos problemas prácticos el vector observación es obtenido muestreando un proceso aleatorio, y este usualmente consiste de un grupo de variables aleatorias correlacionadas. Además vectores observación son a veces grandes en dimensión, y estos contienen información redundante. A fin de reducir la dimensionalidad mientras la mayor información, el vector observación puede ser mapeado sobre un espacio característico usando funciones básicas ortogonales tal como la información transformada no es correlacionada.

Los datos no correlacionados en el dominio transformado pueden ser clasificados de acuerdo al grado de significancia de su contribución al contenido de la información. Muchas transformaciones ortogonales están siendo usadas en la práctica; la mejor incluyó la Transformada de Fourier (FT), la transformada Walsh Hadamard (WHT), Karhunen Loeve(KL), polinomios ortogonales y la transformada coseno discreto (DCT). Libros de texto de Ahmed y Rao (1975)<sup>30</sup>, Rosenfeld y Kak (1982), y González y Wintz (1987) describen estas transformadas.

Digamos que  $\mathbf{H}$  es la matriz transformada dada por:

$$\mathbf{H}^T = [\phi_1, \phi_2, \dots, \phi_n] \quad (5.54)$$

Donde  $\phi_i$  es realmente valuada, vectores bases ortonormales; i.e.,

$$\phi_i^T \phi_j = \begin{cases} 1 & \text{para } i = j \\ 0 & \text{para } i \neq j \end{cases} \quad (5.55)$$

Digamos que  $\mathbf{x}$  es un vector observación, entonces el vector transformada y es dado por:

$$\mathbf{y} = \mathbf{H}\mathbf{x} \quad (5.56)$$

donde  $\mathbf{x}^T = (x_1, x_2, \dots, x_n)$  y  $\mathbf{y}^T = (y_1, y_2, \dots, y_n)$ .

De la ecuación 5.54 y 5.55 tenemos

$$\mathbf{H}^T \mathbf{H} = \mathbf{I} \quad (5.57)$$

Por lo tanto, obtenemos

$$\mathbf{x} = \mathbf{H}^T \mathbf{y}$$

$$x_i = \sum_{j=1}^n y_j \phi_{ij} \quad (5.58)$$

A veces, a fin de reducir la dimensión, podemos mantener un subgrupo  $\{y_1, y_2, \dots, y_m\}$  de componentes de  $\mathbf{y}$ , y todavía estimar  $\mathbf{x}$ . Este resulta reemplazando los  $n$ -m componentes de  $\mathbf{y}$  por el preseleccionamiento  $b_j$  para obtener

$$x_i = \sum_{j=1}^m y_j \phi_{ij} + \sum_{j=m+1}^n b_j \phi_{ij} \quad (5.59)$$

El error estimado de  $\mathbf{x}$  es dado por

$$\Delta x_i = \sum_{j=m+1}^n (y_j - b_j) \phi_{ij} \quad (5.60)$$

El error cuadrático medio está dado por:

$$\begin{aligned} \Delta E &= E \{ (\Delta \mathbf{x})^T (\Delta \mathbf{x}) \} \\ &= \sum_{j=m+1}^n E \{ (y_j - b_j)^2 \} \end{aligned} \quad (5.61)$$

El problema de minimizar  $\Delta E$  es generalmente llamado factor de análisis o análisis componente principal, y la correspondiente transformada es llamada la transformada Karhunen-Loeve (KL). Digamos que  $\phi_i$  representa eigenvectores de la matriz covarianza  $\Sigma_x$  de una entrada  $\mathbf{x}$  de observación, y  $\beta_i$  son los correspondientes eigenvalores. El error cuadrático medio mínimo esta dado por

$$\Delta E_{m,n} = \sum_{i=m+1}^n \beta_i \quad (5.62)$$

La transformada KL es la óptima con respecto al criterio de error cuadrático medio; sin embargo esto envuelve cálculos (Ahmed y Rao, 1975; Oja, 1989).

### 5.2.3 Extracción de características en el dominio de la transformada de Fourier

La transformada de Fourier (FT) es una técnica bien conocida de transformada usada en compresión de datos. Para procesos digitales de imágenes necesitamos considerar la transformada discreta de Fourier (DFT). Si  $\{x(n)\}$  denota una secuencia  $x(n)$ ,  $n = 0, 1, 2, \dots, N-1$  de  $N$  finitos valores reales o números complejos, entonces la transformada de Fourier es definida como (Rosenfeld y Kak, 1982)<sup>24</sup>

$$c(u) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-2\pi j u n / N} \quad (5.63)$$

Para  $u = 0, 1, 2, \dots, N-1$  y  $j = \sqrt{-1}$ . Las funciones exponenciales en la eq. 5.63 son ortogonales. La Transformada inversa discreta de Fourier (IDFT) es definida como.

$$x(n) = \sum_{u=0}^{N-1} c(u) e^{2\pi j u n / N} \quad (5.64)$$

Las ecuaciones 5.63 y 5.64 representan la DFT y la IDFT para una secuencia de datos unidimensionales. Las imágenes son inherentemente bidimensionales en naturaleza, y necesitamos considerar la bidimensional transformada de Fourier. La bidimensional transformada de Fourier discreta esta dada por.

$$F(u,v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) e^{-2\pi j (mu / M + nv / N)} \quad (5.65)$$

para  $u = 0, 1, 2, \dots, M-1$  y  $v = 0, 1, 2, \dots, N-1$ . La Transformada inversa discreta de Fourier esta dada por.

$$f(m,n) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{2\pi j (mu / M + nv / N)} \quad (5.66)$$

para  $m = 0, 1, 2, \dots, M-1$  y  $n = 0, 1, 2, \dots, N-1$ .

La doble sumatoria es la ec.5.65 puede ser escrita en forma natural como

$$F_i = PFQ \quad (5.67)$$

---

<sup>24</sup>Rosenfeld, A., and Kak, A. (1982). Digital Image Processing, Vols. 1 and 2. Academic Press, Orlando, FL.

donde **Q** y **P** son matrices no singulares ortogonales de tamaño  $M \times M$  y  $N \times N$  respectivamente; **F** representa la matriz de imágenes de entrada y los elementos de matrices **P** y **Q** están dadas por la ec. 5.68 y 5.69, respectivamente

$$q_{vn} = \frac{1}{N} e^{-2\pi j n v / N} \quad (5.68)$$

donde  $v = 0, 1, 2, \dots, N-1$  y  $n = 0, 1, \dots, N-1$

$$p_{um} = \frac{1}{M} e^{-2\pi j m u / M} \quad (5.69)$$

donde  $u = 0, 1, 2, \dots, M-1$  y  $m = 0, 1, 2, \dots, M-1$ . La Transformada inversa de Fourier esta dada por.

$$F = P^{-1} F_1 Q^{-1} \quad (5.70)$$

Los algoritmos de extracción de características usando el plano TF han sido implementados ópticamente. Algunas de las propiedades de la TF (estas magnitudes de los coeficientes de la TF son invariantes a altas frecuencias en imágenes de entrada, corresponden a grandes valores más aya de los originales, y con una rotación de la imagen de entrada la distribución de la TF también es rotada) son usadas para extracción de características invariantes. A fin de obtener características invariantes, el plano TF es muestreado con depósitos angular y radial. La extracción de características también resulta en significativa reducción de datos. Muestreando el plano TF con elementos de información en rotación puede ser capturada en la extracción de características. Similarmenete muestreando el plano con ángulos o elementos en forma de anillo, la información en escala puede ser capturada en la extracción de características. Entonces los coeficientes de la TF son simétricos en el plano TF, es posible usar la mitad del plano TF para ángulos y la mitad del plano para depósitos radiales, como se muestra en la fig. 5.27 . Los radiales están dados por

$$F_{r_1, r_2} = \iint F^2(u, v) du dv \quad (5.71)$$

donde los limites de integración están definidos por

$$r_1^2 \leq u^2 + v^2 \leq r_2^2 \quad y \quad 0 \leq u, v \leq n-1 \quad (5.72)$$

Características radiales son insensitivas para diferencias rotacionales de imágenes de entrada, y pueden ser usadas para extracción de reconocimiento de imagen rotación-invariante. Los angulares están dados por.

$$\forall \phi_1, \phi_2 = \iint F^2(u,v) du dv \quad (5.73)$$

Donde los límites de interacción están definidos por  $\phi_1 \leq \tan^{-1}(u/v) \leq \phi_2$  y  $0 \leq u, v \leq n-1$ . Características angulares son insensitivas para rugosidades indefinidas de imágenes de entrada, y estas pueden ser usadas para extracción de componentes de escala invariante.

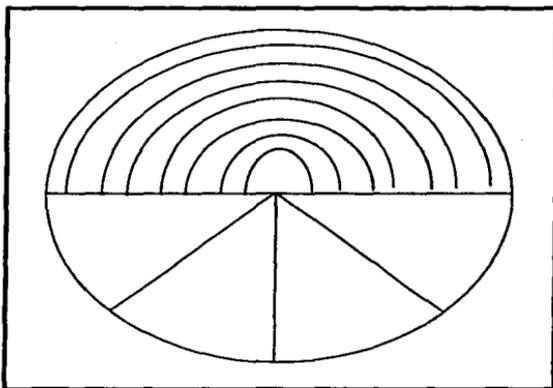


Fig. 5.27 Radiales y angulares

#### 5.2.4 Modelos de R.N.A. para la extracción de características en el dominio de la transformada de Fourier

Los modelos R.N.A. consisten de un número de capas, y cada una de esas capas consisten de varias unidades de procesamiento que pueden tener funciones de

activación lineales o no lineales. Frecuentemente en los modelos R.N.A. que tienen unidades con función de activación lineal, la propagación de información puede describirse matemáticamente como una serie de multiplicaciones de vector-matriz. Cada multiplicación vector-matriz puede representarse por:

$$y = Wx \quad (5.74)$$

donde  $x$  representa el vector de entrada,  $y$  representa el vector de salida, y  $W$  es una matriz de peso. En las redes neuronales biológicas, esta operación es realizada por un gran número de neuronas operando en paralelo.

En la figura 5.28 se muestra un sistema electroóptico. El sistema usa un lente cilíndrico con fotodetectores para los vectores de entrada y salida y una película fotográfica en la cual la transmisión de cada cuadro es proporcional al peso correspondiente. Del lado de la salida unos lentes enfocan una máscara sobre el correspondiente fotodetector. Los modelos R.N.A. requieren dos tipos de operaciones: cálculo y comunicaciones. Las funciones de cálculo son realizadas más bien por los sistemas electrónicos, aunque una tarea de comunicaciones para un gran número de unidades se realiza mejor mediante sistemas ópticos. Un sistema como el que se muestra en la figura 5.28 es el adecuado para las operaciones en paralelo.

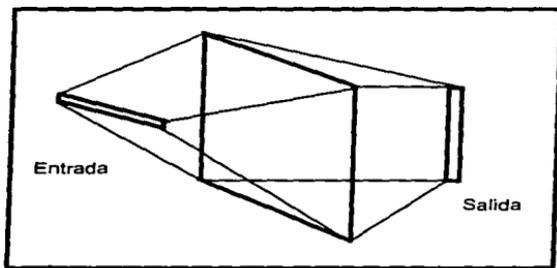


Fig. 5.28 Matriz de vector multiplicador

Kulkarni y Byars (1991a,b)<sup>25</sup> han desarrollado un modelo R.N.A. para la extracción de características usando el plano FT (figura 5.29) que consiste de siete

<sup>25</sup>Kulkarni, A.D., and Byars, P. (1991a). Artificial Neural Network models for imagen understanding. In: Proceedings of SPIE Conference on Image Processing Algorithms and Techniques II, San Jose, CA, 1452:512-522

capas. La capa L1 es la capa de entrada y representa la distribución de la imagen de entrada. La capa L6 representa la amplitud de los coeficientes de Fourier en el dominio de la transformada. La capa L7 representa las características de salida. Las capas de la L2 a L5 son usadas para obtener los coeficientes de Fourier. Los pesos que conectan esas capas pueden obtenerse como sigue.

La ecuación .5.65 puede reescribirse como:

$$F(u,v) = \frac{1}{MN} \sum_{m=0}^{M-1} e^{-2j\pi mu / M} \sum_{n=0}^{N-1} f(n,m) e^{-2j\pi nv / N} \quad (5.75)$$

para  $u = 0, 1, \dots, M-1$  y  $v = 0, 1, \dots, N-1$ . Substituyendo

$$e^{-2j\pi mu / M} = \cos(2\pi mu / M) - j \sin(2\pi mu / M)$$

y

$$e^{-2j\pi nv / N} = \cos(2\pi nv / N) - j \sin(2\pi nv / N)$$

se obtiene

$$\begin{aligned} F(u,v) = & \frac{1}{MN} \left[ \sum_{m=0}^{M-1} \cos(2\pi mu / M) \sum_{n=0}^{N-1} f(n,m) \cos(2\pi nv / N) \right. \\ & - \sum_{m=0}^{M-1} \sin(2\pi mu / M) \sum_{n=0}^{N-1} f(n,m) \sin(2\pi nv / N) \left. \right] \\ & - j \left[ \sum_{m=0}^{M-1} \cos(2\pi mu / M) \sum_{n=0}^{N-1} f(n,m) \sin(2\pi nv / N) \right. \\ & \left. + \sum_{m=0}^{M-1} \sin(2\pi mu / M) \sum_{n=0}^{N-1} f(n,m) \cos(2\pi nv / N) \right] \quad (5.76) \end{aligned}$$

$$= 1 / MN [(P_C f Q_C - P_S f Q_S) - j (P_C f Q_S + P_S f Q_C)] \quad (5.77)$$

donde los elementos de las matrices  $P_C$ ,  $P_S$ ,  $Q_C$ , y  $Q_S$  están dados por:

$$P_C(u, m) = \cos(2\pi mu / M)$$

$$P_s(u, m) = \sin(2\pi mu / M)$$

para  $u = 1, 2, \dots, M - 1$  y  $m = 1, 2, \dots, N - 1$

$$Q_c(v, n) = \cos(2\pi nv / N)$$

$$Q_s(v, n) = \sin(2\pi nv / N)$$

para  $v = 1, 2, \dots, N - 1$  y  $n = 1, 2, \dots, N - 1$

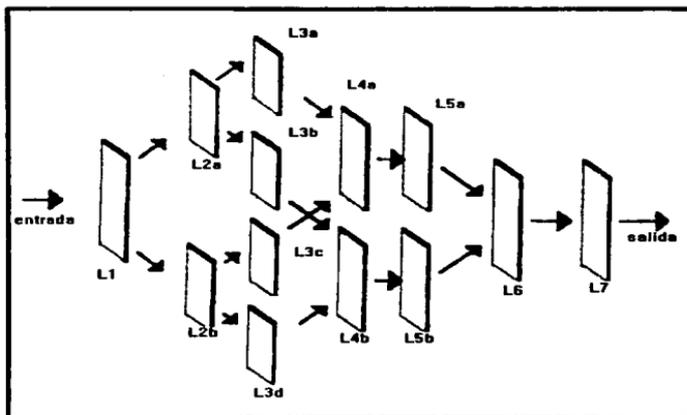


Fig. 5.29 Modelo de R.N.A. en el dominio de la Transformada de Fourier para la extracción de características.

En la figura 5.29 los pesos entre las capas  $L_1L_{2a}$ , y  $L_1L_{2b}$  son los elementos de las matrices  $Q_c$  y  $Q_s$  respectivamente. Los pesos entre las capas  $L_{2a}L_{3a}$ ,  $L_{2a}L_{3b}$ ,  $L_{2b}L_{3a}$ , y  $L_{2b}L_{3b}$  son los elementos de las matrices  $Q_c$ ,  $Q_s$ ,  $Q_s$  y  $Q_c$  respectivamente. Las unidades en las capas  $L_{4a}$  y  $L_{4b}$  corresponden a la parte real e imaginaria de los coeficientes de Fourier. Las unidades en las capas  $L_{5a}$  y  $L_{5b}$  representan el cuadrado de los valores real e imaginario de los coeficientes de Fourier. La capa  $L_6$  representa

las amplitudes de los coeficientes de Fourier. Las unidades en la capa  $L_7$  representan las características de salida. Los pesos que conectan a las capas  $L_6$  y  $L_7$  corresponden a los elementos de los trazos angulares y radiales

Kulkarni y Bryans (1991a,b)<sup>35</sup> han desarrollado un simulador con software para un modelo de R.N.A usando la técnica de extracción de características en el plano TF. Ellos consideraron dos ejemplos. El primero usando tres patrones de prueba representando barras verticales generadas al usar funciones sinusoidales, los patrones con rotación de  $0^\circ$ ,  $45^\circ$  y  $90^\circ$  y los planos de distribución TF son mostrados en las figuras 5.30 y 5.31 respectivamente. Para la extracción de características rotación-invariante, usaron un trazo con seis. Las características obtenidas de un trazo radial son mostradas en la tabla 5.2. En el segundo ejemplo usaron imágenes de tres tipos de aviones, Phantom, Mirage III, F104. Las imágenes con  $0^\circ$ ,  $45^\circ$  y  $90^\circ$  de rotación y sus planos de distribución TF son mostrados en las figuras 5.25 y 5.32 respectivamente. A fin de obtener las características rotación-invariante, el plano TF fue muestreado con anillos detectores. Como en el ejemplo previo, usaron un trazo con seis anillos. Las imágenes de avión con diferentes escalas y sus planos de distribución TF son mostrados en las figuras 5.26 y 5.33 respectivamente. Para obtener características de escala invariante, usaron un trazo radial con ocho separadores. Las características de salida son mostradas en las tablas 5.3 y 5.4 respectivamente. En ambos ejemplos los vectores característicos fueron presentados para un período de reconocimiento que consistió de una red tri-capa de alimentación directa con algoritmos de aprendizaje retro-propagación.

TALA 5.2 CARACTERÍSTICAS PARA TRAZOS RADIALES (PATRON)

Rotación (grados)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$
0	0.74	0.07	0.04	0.02	0.01	0.01
45	0.43	0.70	0.37	0.21	0.11	0.08
90	0.80	0.09	0.05	0.03	0.02	0.02
180	0.91	0.17	0.11	0.07	0.51	0.04

La arquitectura RNA mostrada en la figura 5.29 puede ser aun más simplificada usando la transformada Radon. La transformada Radon es definida como (Gindi y Gmitro, 1984)

$$\beta(x', \phi) = \int f(r) \delta(x' - r, \bar{n}) d^2 r \quad (5.78)$$

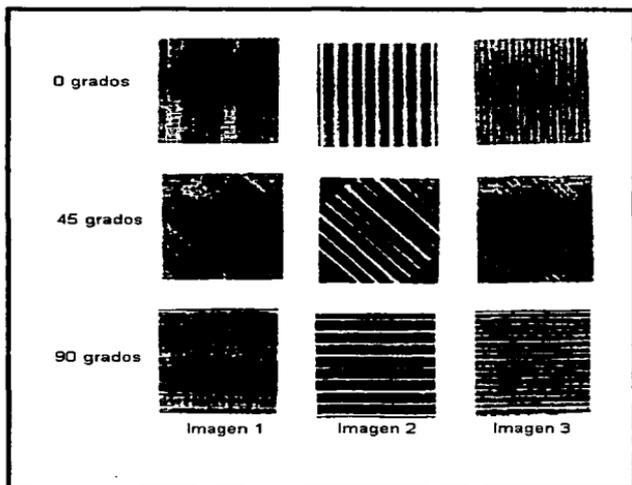


Fig. 5.30 Patrones de prueba

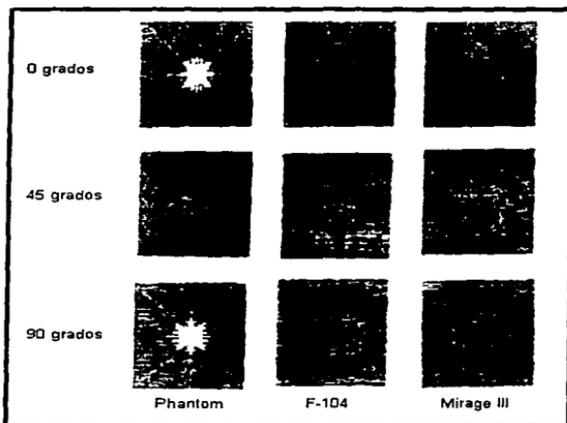


Fig. 5.31 TF para la distribución de patrones de la fig. 5.30

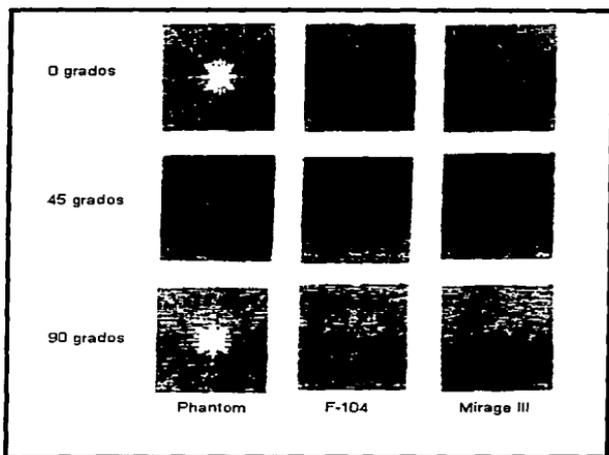


Fig. 5.32 Distribución en la TF para imágenes de aviones

donde  $r(x,y)$  es un vector de posición bidimensional en el espacio,  $f(r)$  define el objeto, y  $\vec{n}$  es el vector unitario que forma un ángulo  $\phi$  con respecto al eje  $x$ . Es conveniente pensar en la transformada Radon como una proyección de la función  $f(x,y)$  sobre el eje  $x'$ , donde  $x'$  es el eje rotado a un ángulo  $\phi$  del eje  $x$ .

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.79)$$

Para un ángulo fijo  $\phi$ ,  $\beta(x', \phi)$  es una función unidimensional de  $x' = \beta(x', \phi)$  es esencialmente una proyección para un ángulo  $\phi$ , como se muestra en la figura 5.34. La transformada Inversa Radon es frecuentemente usada para estimar un objeto bidimensional (2-D) de estas proyecciones (1-D) unidimensionales.

La transformada inversa es bien conocida por sus exitosas aplicaciones en tomografía médica computarizada. El teorema central de Slice (TCS) relaciona. La transformada Random de  $f(x,y)$  para su transformada de Fourier. La TCS establece que la FT 1-D de una proyección para un ángulo  $\phi$  es igual a la FT 2-D de  $f(x,y)$  evaluada a lo largo de una línea en el espacio 2-D colineal al vector unitario  $\hat{n}$  que especifica la dirección de la proyección. Así, la transformada de Random junto con el teorema central slice es usado para evaluar los coeficientes de la FT 2-D de proyecciones 1-D de la función  $f(x,y)$ .

El modelo de R.N.A para extracción de características usando la transformada Random (figura 5.35) consiste de cinco capas. La capa  $L_1$  es una capa de entrada representando la distribución de imagen de entrada. La capa  $L_2$  es la transformada Random espacio. Las capas  $L_{3a}$  y  $L_{3b}$  representan las partes real e imaginaria  $P_C$  y  $P_S$ , y son evaluados en el siguiente material (Kulkarni y Byars, 1991c).

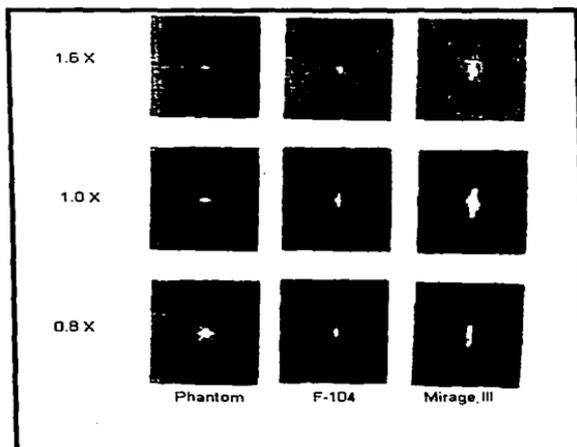
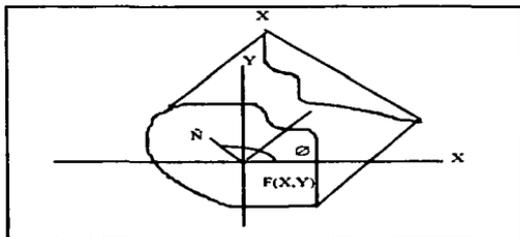


Fig. 5.33 Distribución de la TF para imágenes de aviones de la fig. 5.26

**TABLA 5.3 CARACTERÍSTICAS PARA UN TRAZO RADIAL (IMAGENES DE AVION)**

Rotación en grados	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$
0	0.39	0.23	0.14	0.14	0.82	0.04
45	0.28	0.19	0.18	0.14	0.84	0.04
90	0.39	0.23	0.14	0.14	0.08	0.04
180	0.39	0.23	0.14	0.14	0.08	0.04
0	0.28	0.24	0.13	0.12	0.07	0.04
45	0.27	0.14	0.17	0.12	0.07	0.04
90	0.28	0.24	0.13	0.12	0.07	0.04
180	0.28	0.24	0.13	0.12	0.07	0.04
0	0.39	0.20	0.18	0.18	0.10	0.06
45	0.53	0.25	0.18	0.16	0.10	0.06
90	0.39	0.20	0.18	0.18	0.10	0.06
180	0.39	0.20	0.18	0.18	0.10	0.06



**Fig. 5.34 Geometría para la transformada Random**

Consideremos la transformada de Fourier discreta de 1-D (DFT) de la función  $f(n)$ .

$$F(u) = \frac{1}{N} \sum_{n=0}^{N-1} f(n) e^{-2\pi jnu/N} \quad (5.80)$$

para  $u = 0, 1, 2, \dots, N-1$ . La ecuación 5.80 puede ser representada en forma de matriz como

$$\mathbf{f}_r = \mathbf{P}_c \mathbf{f} - j\mathbf{Q}_c \mathbf{f} \quad (5.81)$$

**TABLA 5.4 CARACTERÍSTICAS PARA TRAZOS ANGULARES (IMAGENES DE AVION)**

Escala	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$
1.6	0.13	0.10	0.13	0.27	0.12	0.08	0.68	0.99
1.2	0.13	0.10	0.11	0.27	0.12	0.08	0.07	0.10
1.0	0.14	0.10	0.11	0.26	0.13	0.08	0.07	0.11
0.8	0.13	0.10	0.12	0.25	0.13	0.08	0.07	0.11
1.6	0.15	0.07	0.01	0.36	0.10	0.06	0.07	0.09
1.2	0.15	0.07	0.10	0.34	0.10	0.05	0.07	0.11
1.0	0.14	0.07	0.09	0.35	0.11	0.05	0.07	0.10
0.8	0.15	0.07	0.08	0.37	0.11	0.05	0.06	0.10
1.6	0.10	0.11	0.18	0.21	0.12	0.12	0.08	0.09
1.2	0.10	0.12	0.16	0.22	0.12	0.12	0.09	0.08
1.0	0.10	0.12	0.16	0.21	0.12	0.12	0.09	0.09
0.8	0.10	0.12	0.15	0.20	0.12	0.17	0.09	0.09

donde  $\mathbf{f}$  representa el vector entrada,  $\mathbf{f}_r$  representa los coeficientes de Fourier, y los elementos de las matrices  $\mathbf{P}_c$  y  $\mathbf{P}_s$  están dados por  $P_c(u, n) = 1/N \cos(2\pi nu/N)$  y  $P_s(u, n) = 1/N \sin(2\pi nu/N)$  para  $u=1, 2, \dots, N-1$ . Las unidades en la capa  $L_4$  representan amplitudes de los coeficientes de la TF obtenidos de la proyección de  $f(x, y)$  para el ángulo  $\phi$ . Así sumando las salidas de las unidades en cada línea obtenemos características radiales. Similarmente, sumando las salidas de las unidades en cada columna obtenemos características angulares. Estas características son usadas como entradas para el periodo de reconocimiento. Kulkarni y Byars (1991b)<sup>26</sup> han desarrollado un simulador de software para el modelo R.N.A. mostrado en la figura 5.35. Ellos han usado exitosamente el modelo para reconocimiento de textura y problemas de reconocimiento de caracteres.

<sup>26</sup>Kulkarni, A.D., and Byars, P. (1991b). Neural nets for invariant object recognition. In: Proceedings of the Symposium on Applied Computing, Kansas City, MO, pp. 336-344.

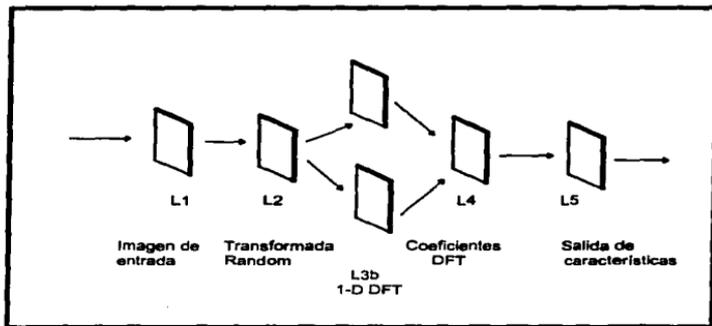


Fig. 5.35 Modelo de R.N.A. para extracción de características usando la transformada Random

# Capítulo 6

## APLICACIONES DE REDES NEURONALES ARTIFICIALES PARA EL RECONOCIMIENTO DE IMÁGENES

---

### INTRODUCCION

Las redes neuronales representan una poderosa y razonable alternativa de aproximaciones tradicionales para comprensión de imágenes. Algunas de las aplicaciones de las redes neuronales para comprensión de imágenes son las siguientes: percepción remota, imágenes médicas, robótica, cartografía, reconocimiento de caracteres, reconocimiento de rostro, etc.

#### 6.1 *Percepción remota*

La percepción remota es la ciencia de la información que se deriva de un objeto hecho de medidas a cierta distancia de este. La señal remotamente medida se expresa como una función de longitud de onda frecuentemente referida como "señal espectral" de la tarjeta objeto que tiene medidas dadas. En principio una señal espectral es única, diferentes objetos tienen diferentes señales espectrales. Por tanto es posible identificar un objeto por su señal espectral. Esto es el principio de un sensor remoto multiespectral, cuya potente técnica de monitores de recurso natural y medio ambiente. Remotamente el sentido espectral de medición puede ser una fuente de información de muchas aplicaciones, incluyendo agricultura, silvicultura, hidrología, recursos minerales, recursos de agua, geografía, cartografía, meteorología y militar.

El censor remoto del satélite comienza con una señal, el lanzamiento del primer satélite de la tierra en 1972 por la NASA en USA., fue seguido por Landsat 2 y landsat 3 en 1975 y 1978, respectivamente. Landsat 2 y Landsat 3 portaron un rastreador multiespectral (MSS). Ese satélite tiene un censor adicional llamado mapa temático (MT). El MT tiene siete bandas espectrales de las cuales una es banda térmica.

Las imágenes multiespectrales son analizadas usando técnicas estadísticas convencionales. Redes Neuronales Artificiales representan una alternativa de clasificación tradicional. Kulkarni y Byars (1991) usaron modelos de redes neuronales para análisis de imágenes multiespectrales. Tienen que usar la retro-propagación

(BP), el algoritmo de aprendizaje competitivo clasificarlo como supervisado y al análisis de imágenes MT como no supervisado.

Un MT científico es usualmente de tamaño 5300 scanners x 4000 píxeles. Se usan métodos convencionales para clasificación multiespectral incluyendo un clasificador de probabilidad - máxima, clasificador distancia - mínima y varias técnicas.

Como una ilustración Kulkarni y Byars (1991) consideraron una ciencia a TM (mapa temático) (# y 4018116055, enero 1983). La ciencia analizó usando una red de tres capas con alimentación hacia adelante, usó un algoritmo de aprendizaje de retro-propagación. La misma ciencia tuvo que analizar usando una red de alimentación hacia adelante de dos capas aplicando un algoritmo de aprendizaje competitivo.

En el caso del aprendizaje competitivo, bandas espectrales 3,4,5 y 7 fueron escogidas porque estas bandas muestran la variación máxima y contienen máxima información. La red puede reconocer únicamente dos clases: agua y aceite.

En las técnicas de clasificación convencional de tiempo - consumo cada píxel es probado por todas las clases posibles y el píxel es asignado a la clase con una máxima y una probabilidad posterior. Con redes neuronales artificiales cada modelo es entrenado directamente, la red mapea el vector observación de entrada a la categoría de salida. Eberlein y Yates (1991)<sup>1</sup> usaron modelos de redes para análisis de datos para un algoritmo de aprendizaje de retropropagación en un sistema de clasificación geológica.

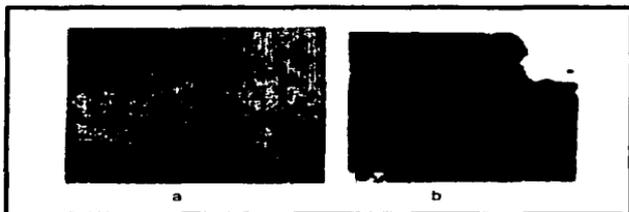


Fig. 6.1 Clasificación meteorológica. a) Imagen original; b) salida clasificada

<sup>1</sup>Eberlein, S., and Yates, G. (1991). Neural Network based system for autonomous data analysis and control. In: Progress in Neural Networks, Vol. 1, O. Omidvar, (de.), pp. 25-55. Ablex, New York.

La implementación del sistema de visión de bajo nivel con un sistema basado en un número operando con teoría biológicamente plausible a través de todos los canales de entrada. La representación del procesamiento es seguido por la representación clasificada. El sistema automatizado genera productos meteorológicos en tiempo real. El vector característico es construido para añadir datos actuales con productos derivados.

Smotroff (1990) usó funciones Gabor 2-d para extraer características de textura empleando una red retropropagación como un clasificador supervisado. Las salidas de estas imágenes meteorológicas provenientes de satélites se muestran en las figuras 6.1 y 6.2.



Fig. 6.2 Clasificación meteorológica. a) imagen original; b) salida clasificada

### 6.2 Procesamiento de imágenes médicas

Las técnicas estadísticas convencionales han sido usadas en la práctica del análisis de rayos -X, microscopio y resonancia magnética (MR) de imágenes. Las redes neuronales representan una alternativa de activación a técnicas convencionales. Kaufman et al (1990) uso modelos de redes neuronales artificiales para valorar la curación de fracturación de huesos. Las imágenes de resonancia magnética se usaron recientemente como un diagnostico estándar de herramientas. La interpretación reproducible y acertada de una imagen de resonancia magnética (MR) es ejecutada por un entrenamiento físico remanente y extremadamente de tiempo - consumido y tarea costosa. Imágenes diferentes a rayos - X, MRI permiten medir los tres parámetros de giro específico; el giro de relajación de tejido (T2), el

giro-enrejado de relajación de tejido (T1) y la densidad de protón (PD). Aquí cada pixel puede ser representado como un vector de espacio tridimensional del cual las componentes T1-peso, T2-peso y valor de peso PD. El T1-peso, T2-peso y valor de peso PD similarmente pueden ser trasladados a una banda espectral de imágenes en un análisis multispectral de imágenes pudiendo ser usados por imágenes MR.



Fig. 10.3 Imágenes segmentadas (caso patológico). a) Red Neuronal; b) cubierta máxima.



Fig. 10.4 Imágenes segmentadas (voluntario normal). a) Red Neuronal; b) cubierta máxima.

Usando tres unidades en la capa de entrada y cuatro unidades en la capa de salida. Las tres unidades de la capa de entrada representan el T1-peso, T2-peso y valores PD para pixel. Las cuatro unidades de la capa de salida representan los cuatro tipos de tejido.

La imagen segmentada se obtiene usando un clasificador de red neuronal y un clasificador convencional se muestra en la figura 6.3 y 6.4. El modelo de redes neuronales esta basado en el límite de un sistema contador (BCS) y la característica de un sistema contador (FCS) pueden usarse por una imagen médica. Lehar et al (1990) uso BCS y FCS modelos para procesamiento de imágenes MR. Ejemplo de esto se muestra en la figura 6.5.

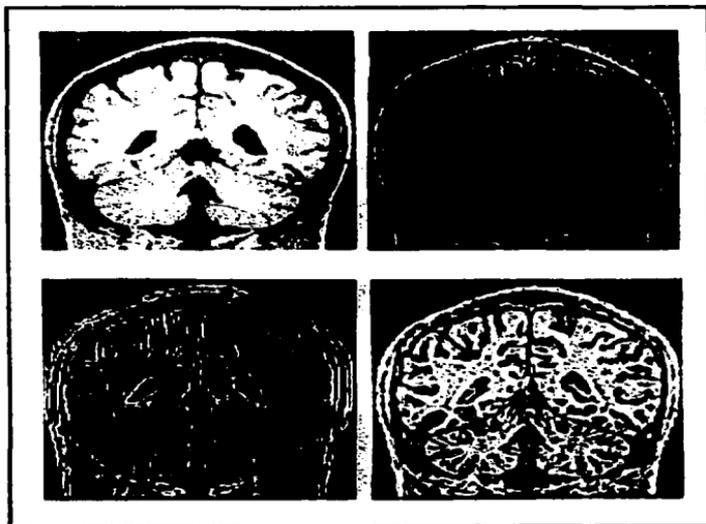


Fig. 6.5 modelo para procesamiento de imágenes MR.

### 6.3 Procesamiento de huellas digitales

Es bien sabido que el cerebro humano calcula la respuesta cuando se presenta un problema tal como extracción de características y reconocimiento de patrones. Un

problema es la extracción de pequeños detalles a gran escala de huellas digitales. Las huellas digitales son usadas para identificaciones individuales porque cada persona tiene su propia huella digital. La clasificación usualmente se realiza por ciertas características de pequeños detalles. Estos están colocados al final de la cumbre o cuando es dividida en dos. Las aproximaciones convencionales para el análisis de las huellas digitales incluyendo la transformada de fourier. Leung y otros (1990) diseñó un sistema de extracción de pequeños detalles a gran escala de imágenes usando filtros Gabor. Una pequeña región de una huella con rasgos verticales, se muestra en la fig. 6.6. Los filtros Gabor tienen dirección sensitiva; del lado del borde en dirección vertical, todas las partes de esta figura, excepto la parte que envuelve los pequeños detalles se tratan de igualar a la señal de amplitud que se tiene en el filtro. La localización de los pequeños detalles es caracterizada por regiones en donde hay significativamente cambios de presentación. La magnitud y fase de un filtrado Gabor de una imagen de huella digital se muestra en las figuras 6.7 y 6.8, respectivamente. El barrido final de la red neuronal de las huellas digitales es mostrado en la figura 6.9. El resultado de la magnitud y la fase de la imagen filtrada usa una red de retropropagación para localizar e identificar pequeños detalles.

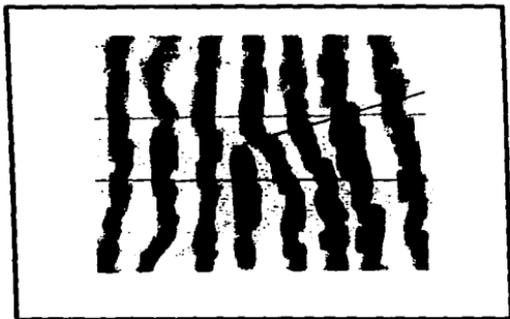


Fig. 6.6 Pequeña región de una huella con rasgos verticales

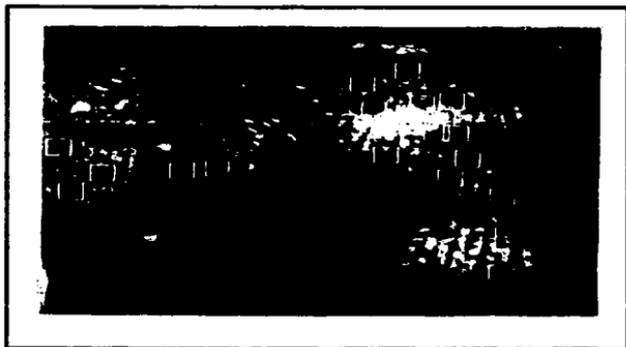


Fig. 6.7 Magnitud de un filtro Gabor de 0°.

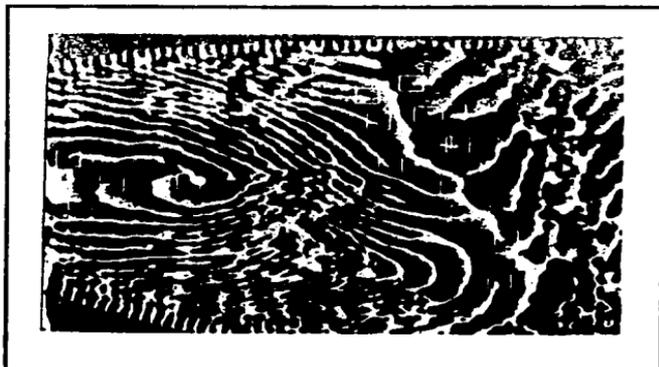


Fig. 6.8 Fase de un filtrado Gabor de una imagen de una huella

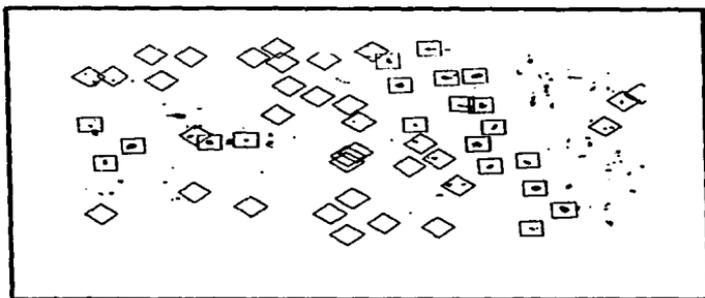


Fig. 6.9 Salida final de una Red Neuronal indicando con un cuadro la localización detallada.

#### 6.4 Reconocimiento de caracteres

En años recientes las Redes Neuronales tuvieron que ser usadas en tiempo real para reconocimiento de caracteres. Los algoritmos paralelos de redes neuronales muestran grandes promesas que proveen una alta exactitud y resistencia al ruido en la aplicación al reconocimiento de caracteres (Omidvar y Wilson, 1991; Jackel et al, 1998). El rango de aplicación de reconocimiento de caracteres incluye lectura de código postal, entrada de datos automática, reconocimiento de escritura, cartografía automatizada, banca y lectura de servicios para ciegos, conversión de imágenes de manuscritos e impresión de caracteres de máquina a representación de computadora fueron estudiados a detalle. Ambos propósitos sistemas de software y hardware para reconocimiento de caracteres han sido desarrollados.

Las Redes Neuronales presentan una alternativa para problemas de reconocimiento de caracteres. Una salida crítica en reconocimiento de caracteres es selección de características como la precisión para cada decisión depende de las características seleccionadas. Cada carácter tiene características que lo distinguen de otro carácter. Algunas características prominentes usadas para el reconocimiento de caracteres son ciclos, agujeros, golpes, líneas verticales, cúspides, etc., más de los algoritmos existentes envuelven procesos extensivos sobre la imagen.

El problema del reconocimiento por computadora para un documento usualmente es dividido en tres operaciones. Primero las áreas relevantes contienen

un texto localizado. La imagen global contiene imágenes de uno o más caracteres divididos dentro de la imagen en caracteres individuales. Este proceso es referido usualmente a segmentación. Esta separación de caracteres forma la entrada para el sistema de reconocimiento de caracteres. Rajavelu et. Al. (1989) uso la función Walsh para la extracción de características para el reconocimiento de caracteres. La función Walsh fue comúnmente usada en procesamiento de señales para representar datos como una función ortogonal en serie. El número de la función de Walsh requerido para que los datos de la imagen concuerden en una serie ortogonal de expansión depende exactamente de la aplicación necesitada. Rajavelu y otros (1989) uso proyecciones en direcciones horizontal y vertical y uso la transformada Walsh-Hadamard para extraer las características del reconocimiento de caracteres. El proceso de extracción de características en 20 características (coeficiente-expansión). Entonces es usada una red de tres capas con alimentación hacia adelante en un algoritmo de aprendizaje retropropagación como un clasificador supervisado. La red retropropagación tiene 20 unidades en la capa de entrada las cuales corresponden a las 20 características obtenidas en la fase de extracción de características. El número de unidades de salida en la representación del sistema es un estándar que se da y el sistema fue probado con 20,30,40 y 50 unidades en la capa media.

El sistema de reconocimiento uso mayúsculas del alfabeto para un tiempo fuente. El número de pruebas que necesita la red para converger es de 1215.

Weideman y otros (1989) usaron una red retropropagación para reconocimiento de caracteres, comparó la ejecución de una red retropropagación con un clasificador. Usando caracteres determinados en donde cada carácter es representado por una matriz de 32 scanner x 24 píxeles. Usando características que incluyen coeficientes 2-D FT, de momentos geométricos y características topológicas. Las características topológicas incluyen distancias que de las 4 esquinas a el primer pixel negro encontrado. Usando una red retropropagación de tres capas como clasificador, la red retropropagación contiene 100,90 y 10 unidades en la capa de entrada, capa media y capa de salida respectivamente. Usando datos de números generados con 3000 letras impresas de una agencia gubernamental de EUA. La base de datos tiene aproximadamente 86, 000 caracteres. Un aprendizaje determinado de 20,000 caracteres se usan durante el aprendizaje de la red, estos se clasifican con el clasificador Nearest-Neighbor.

### **6.5 Reconocimiento de rostros**

El reconocimiento de rostros es una técnica extremadamente importante además es diferente a cualquier otro método, se usan distancias entre marcas específicas de características de cada cara Fleming y Cottrell (1990) sugirieron para una aproximación de reconocimiento de rostros usando una red neuronal. Este modelo consiste en dos representaciones: La representación de extracción de

características y la representación de reconocimiento, se usa una red de tres capas con un algoritmo de aprendizaje supervisado de retropropagación. La red también puede usarse para reconocimiento de imágenes. La entrada y salida son relativamente largas a comparación de la capa media, este tamaño diferencial necesita una representación más compacta la información de imagen de unidad nivel de media para la manipulación de la distribución del tamaño de la capa entrada/salida a tamaño capa media, es necesaria una actividad de comprensión de datos. En la segunda representación para el reconocimiento del sistema se usa una red retropropagación como clasificador supervisado. La primera etapa trabaja como una larga red autoasociativa. Como ilustración se usa un aprendizaje determinado que contiene 64 imágenes de rostros, algunas de estas imágenes se muestran en la figura 6.10 El modelo fue efectuado con una alta exactitud sobre una gran variedad de estímulos. La salida del clasificador indica identidad, sexo y rasgos faciales de cada imagen de entrada.



Fig. 6.10 Imagen de rostros

Estas son mostradas con el incremento en el número de unidades de la capa media la señal de ruido tiene incrementos proporcionales, pero el SNR no fue tan bueno como en los esquemas convencionales. Los efectos de aprendizaje con coeficientes DCT como pesos iniciales son investigados y se encontró que tienen casi las mismas características SNR obtenidas por pesos aleatorios.

# Capítulo 7

## ALGORITMOS DE APRENDIZAJE DE REDES NEURONALES ARTIFICIALES

---

### INTRODUCCION

Las primeras investigaciones en el campo de las reglas del aprendizaje o entrenamiento de las redes neuronales se deben a McCulloch - Pitts<sup>1</sup>.

El proceso de aprendizaje consiste en que, una vez que se obtiene una salida o respuesta del EP (Elemento Procesador ( célula o neurona ) ), esta se puede " retroalimentar " mediante un ciclo de control hacia el mismo EP; si el proceso de control logra hacer variar el valor de los pesos que llegan a el EP, entonces se podrá intentar modificar la salida de la red con respecto a la que proporcionó antes, y por el contrario, si sólo se retroalimenta la salida (sin variar los pesos), entonces las salidas logradas en posteriores pasos no variarán; en otras palabras, se puede considerar que si un EP "desinhibida " bajo ciertas condiciones, se puede lograr que en el siguiente paso se " inhiba ";, si se hacen los cambios apropiados a los pesos que permiten calcular el ingreso neto de la salida !!

### 7.1 Algoritmo del perceptrón

Una salida al problema de la red neuronal a sido tradicionalmente el problema de separación de patrones en dos categorías. Un primer tipo de neurodo es aplicado para este propósito por Warren S. McCulloch u Walter Pitts en 1933 ellos sugirieron un simple dispositivo como el que se muestra en la figura 7.1.

El neurodo de McCulloch - Pitts es la base para las redes neuronales. La idea es extremadamente simple. El calculo de la suma de los pesos del neurodo para la señal de entrada y comparar este peso de entrada de la red con un valor de umbral T. Si el peso de entrada es mayor o igual al de umbral la salida del neurodo es +1, si no la salida es -1. Este neurodo McCulloch - Pitts es virtualmente indistinguible para los neurodos usados en el perceptrón y el adaline y es el claro antecesor de casi todos los neurodos en la corriente de las redes neuronales. Así es encontrado ahora, este neurodo uso la función de transferencia siguiente:

---

<sup>1</sup>McCulloch-Pitts Network, G. Palm and FT Sommer Cu O Vogt Institut Für Hirnforschung, Universität 44sstrasse 1, D-400 Disseldorf, Federal Republic of Germany.

$$I = E = \sum_{i=1}^n W_i X_i$$

$$Y = +1, \text{ si } I > T$$

$$-1, \text{ si } I < T$$

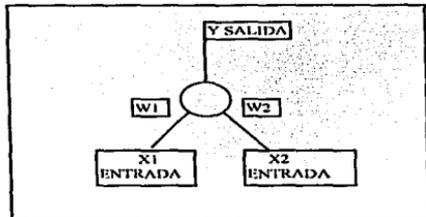


Fig. 7.1 Perceptrón

Aquí el peso de entrada del neurodo  $W_i$  y  $X_i$  son componentes del vector peso y el vector entrada respectivamente, y  $Y$  es la salida del perceptrón. El valor de umbral  $T$  es la mínima actividad requerida para que el neurodo tenga una salida positiva. Para propósitos de discusión  $T = 0$ .

Si McCulloch y Pitts definieron un modelo simple de neurodo que se tomo como estandar. Si Frank Rosenblat transformo este neurodo en la primera red neuronal entrenable. El algoritmo de aprendizaje del perceptrón, introducido en 1958, proveniente del primer procedimiento pudo ser usado para permitir que una red aprendiera una tarea. Y la tarea de aprendizaje de perceptrón es una de las clásicas. La separación de patrones en dos categorías.

Rosenblatt's noto que la ley es totalmente simple:

$$W_{new} = W_{old} + B_{yx}$$

$B = +1$ , si la respuesta del perceptrón es correcta y  
 $-1$ , si la respuesta del perceptrón es falsa.

Y = respuesta del perceptrón

El vector peso en esta expresión es representado por W y el vector patrón de entrada por X.

El algoritmo de aprendizaje del perceptrón es en un contexto de manera gráfica descrito como un procedimiento de trayectoria de un vector. En el seguimiento del siguiente listado también se da una descripción del pseudocódigo del algoritmo.

### ALGORITMO DE APRENDIZAJE DEL PERCEPTRÓN

Para cada patrón en el aprendizaje se determina:

```
{
  aplicación del siguiente patrón para el perceptrón
  registro de respuesta del perceptrón (ambo +1 o -1)
  si la respuesta del perceptrón es correcta
  {
    y la respuesta fue +1, entonces
    el nuevo vector peso = el viejo vector peso + la entrada del vector
      patrón
    y la respuesta fue -1, entonces
    el nuevo vector peso = el viejo vector peso - la entrada del vector
      patrón
  }
  si la respuesta del perceptrón es incorrecta
  {
    y la respuesta del perceptrón fue +1, entonces
    el nuevo vector peso = el viejo vector peso - la entrada del vector
      patrón
    y la respuesta del perceptrón fue -1, entonces
    el nuevo vector peso = el viejo vector peso + la entrada del vector
      patrón
  }
} /* fin de cada patrón de el aprendizaje determinado */.
```

#### 7.2 Algoritmo de la máquina de Boltzmann

La Red de Tracción tiene una energía asociada con su estado actual, que incluye ambos estados para los pesos sobre las conexiones y los niveles de actividad actuales para las neuronas en la Red. La energía surge de un modelo insignificante usado dentro de la operación de la Red de tracción. Cuando aplicamos para una Red

de tracción de capa simple el resultado nos lleva a un nuevo modelo de Red *La máquina de Boltzmann*.

La capa simple de tracción para la Red Hopfield (figura 7.2) binaria acepta entradas, y las neuronas de esta Red calculan estas entradas con la fórmula:

$$Y = E = \sum_{i=1}^n W_i X_i$$

$$Y = +1, \text{ si } Y > 0;$$

$$-1, \text{ si } Y < 0;$$

La Red de tracción tiene una función de energía asociada con las análogas para una función de energía encontrada en un sistema físico. Si se asume que el umbral para una neurona es cero, la energía global para una neurona es cero, la energía global para la Red de tracción es calculada por.

$$E = - \sum_{i=1}^n \sum_{j=1}^n W_{ij} Y_i Y_j$$

La energía es calculada usando el nivel de actividad actual de la salida de cada neurona; de esta manera los términos usados en la ecuación para  $y_s$ ,  $x_s$ .

$$E = -1/2 \sum_{i=1}^n \sum_{j=1}^n W_{ij} Y_i Y_j - 1/2 \sum_{i=1}^n \sum_{j=1}^n W_{ij} Y_i Y_j$$

Esta formula reduciéndola para simplificar el modelo de Hopfield; un peso simétrico para una matriz con diagonal cero (En otras palabras, cuando las neuronas no permiten alimentar las salidas directamente respaldadas por ellas mismas).

Como la actividad fluye directa de la red, el estado del cambio en las neuronas de +1 a -1 y -1 a +1. Obviamente, en cada tiempo sucede un cambio en la energía global. Hay una simple relación que expresa como la energía cambia de modo semejante en cada transición

$$DE = \langle \text{Energía si } Y_j \text{ es } -1 \rangle - \langle \text{Energía si } Y_j \text{ es } +1 \rangle$$

$$DE = \sum_{j=1}^n W_{ij} Y_j$$

Uno de los caminos considerados para la función de transferencia por la red de Hopfield es calculada por la diferencia relativa en la energía global de la red basada en si cada neurona de la red se mueve en un estado de +1 a -1. Porque la salida de cada neurona es determinada por este calculo, cada cambio de tiempo una neurona va del estado +1 a -1 o de -1 a +1, se tiene el efecto de disminución de la energía de la red. Sobre un nivel simple, la función energía puede ser semejante cuando se tiene una pelota que rueda a la parte más baja de la curva. Si la consistencia de la curva con simple descenso es semejante al mostrado en la figura 1, la pelota hace un recorrido hacia lo más bajo de la curva, en el caso de que en el intermedio hubiera una protuberancia y un hueco entre la posición inicial de la pelota y la parte mas baja de la curva la pelota no podrá hacer el recorrido una vez. La figura 7.2 muestra esta situación. En la vida real frecuentemente las pelotas siempre llegan a la parte más baja de las curvas. Otras redes usan directamente el concepto de momento; la red Hopfield no tiene una analogía directa, sin embargo hay otro camino para asegurar el problema teniendo una estructura semejante de "mínimo local", sin embargo la actividad de las neuronas son un proceso estadístico. Los físicos desarrollan fórmulas matemáticas para describir acciones de números largos.

La maquina de Boltzmann es una Red Hopfield la cual opera en un camino único. La actividad neuronal no es determinada por un signo de entrada, pero si por una función probabilística. Cada neurodo determina su actividad por un calculo inicial de una función probabilística. De el ith neurodo.

$$P_i = y_i / 1 + e^{-DE_i/T}$$

Donde:

$$DE_i = \sum_{j=1}^n W_{ij} Y_j$$

En la ecuación de arriba T es temperatura Boltzmann, en otras palabras T es un parámetro que puede ser usado para ajustar la acción de la Red. Una vez que es calculada esta probabilidad, cada neurona determina su salida a +1 con probabilidad igual a r. Por ejemplo en una simulación este puede ser acompañado por un número aleatorio generado con el regreso a valor aleatorio entre 0.0 a 1.0.

Si el valor aleatorio generado es menor o igual a r, determina su salida a +1; en otro caso determina su salida a -1. Usando este procedimiento algunas características de la maquina de Boltzmann son evidentes. Por ejemplo cuando los parámetros de la temperatura son relativamente altos a el cambio de energía (el peso de entrada de la red), la aproximación del determinador de limite 2, y la probabilidad r aproximadamente 0.5. En otras palabras las neuronas son solamente como aproximación para generar una salida de +1 o una salida de -1. Como los parámetros de temperatura decrecen con valores muy pequeños relativos al peso de entrada de la neurona (o de la diferencia de los estados de la energía entre +1 y -1), cada



La red artificial considerada (de esta manera) lejos de tener pocas noticias de esta falta de coordinación. Un tratamiento teórico a más detalle, es la demostración del BAM de sí los neurodos en una capa dada del BAM son actualizados ellos mismos sincrónicamente y los otros neurodos de cualquier otra capa sincrónicamente.

La máquina de Boltzmann su operación esta basada en el azar, en la acción probabilística de un sistema estadístico, de esta manera el orden con que los neurodos cambian de estado es por medio del carácter de aleatorio probabilístico. En particular la maquina de Boltzmann escoge un neurodo aleatorio para actualizarlo permitiendo que ese neurodo cambie de estado con una continuidad menor a un 50%, de esta manera bajara la temperatura de la red a el siguiente nivel. Algunas veces debido al tiempo que existe en el cambio de un estado a otro algunos neurodos no se actualizan y algunos actualizan a tiempo muy severo pero estadísticamente la red esta en constante movimiento aun en un estado bajo de energía.

El diagrama de niveles simples de energía de la figura 7.1 y 7.2. están un poco desviados la "gráfica horizontal" de energía de una red con más de uno o dos neurodos es altamente compleja, la curva multidimensional que contiene estructuras complicadas de curvas tales como puntos de "silla de montar" y picos, Aunque esta complejidad causa cierta dificultad en la visualización o trazado de la gráfica horizontal, esto no afecta la comprensión básica de como trabaja la superficie de energía. Esto significa aproximadamente que el pozo simple de los diagramas bidimensionales podrían tener "extremos de salida" que lo harían aun más sencillo para que la red escape a un estado de energía todavía más bajo.

Para complicar este ejercicio se necesita usar una tabla de números aleatorios o tener algún otro método de generación de números aleatorios.

Cuando se presenta un patrón a la red, se permite actualizar los pesos hasta que cada neurodo tiene un 50% de probabilidad de haber actualizado su estado si se elige aleatoriamente el neurodo que actualiza de entre 6 neurodos en la red, entonces cada vez que cualquier neurodo dado tiene una oportunidad de 6 de ser seleccionado. La probabilidad de que un neurodo específico no sea seleccionado ni la 1a vez ni la 2a vez es el 69%, la probabilidad de que este mismo neurodo no este seleccionado después de una tercera actualización es un 58% y la probabilidad de que el mismo neurodo no se a seleccionado en 4 tiempos es 48%. Este medio de el 4 consecutivo y la actualización aleatoria independientemente tiene un 48% de oportunidad de que ese neurodo no sea seleccionado, implicando una oportunidad de 52% que durante el proceso algunas veces sea seleccionado . Esta es la condición necesaria para determinar cuando el ciclo termina.

#### 1.- Llamar a los patrones guardados dentro de la red de la figura 7.3

$$A1 = 1 \ -1 \ -1 \ 1 \ -1 \ -1$$

$$A2 = -1 \ 1 \ -1 \ -1 \ 1 \ -1$$

$$A3 = -1 \ -1 \ 1 \ -1 \ -1 \ 1$$

Asumimos que el patrón A2 es presentado a la red con elementos de entrada y todos esos neurodos comienzan con un nivel de actividad cero, la entrada inicial a los neurodos es el peso neto de entrada de las conexiones. En este caso, este es calculado por la multiplicación descendente de los pesos. El peso neto de entrada para el primer neurodo es la suma de los productos de -1 (el primer elemento) y cada introducción vertical baja los pesos empezando por el que esta a la izquierda (0, -1, -1, 3, -1 y -1). Esta suma de pesos es 1. La suma que queda de los pesos de los 6 neurodos es la construcción de la diferencia en las redes, el estado de energía en grupo si cada neurodo individual es determinado a+1 estado instancia del estado de salida a-1.

$$DE \text{ para cada neurodo: } 1 \ -1 \ 1 \ 1 \ -1 \ 1$$

2.- Aleatoriamente se escoge un neurodo por selección entre 6 y 1. Supongamos que el neurodo escogido es el neurodo 2, este neurodo tiene un valor de entrada con peso neto de -1. La probabilidad de que este neurodo llegue a un estado final arriba de a+1 es calculado por:

$$P_i = \frac{1}{1 + e^{- (DE_i / T)}}$$

En este caso el exponente de e es  $-(-1/10)$ , 0 +0.10 calculando esta probabilidad dada r es segunda salida del neurodo n+1 a 0.475 o 47.5%.

3.- Para completar este paso a través del ciclo de lista, se genera un número aleatorio entre 1 y 100, si el número aleatorio generado es 47 o menor, el neurodo de salida 2 es cambiado a+1; y su salida queda -1. Suponiendo que el número generado es 4; la salida del neurodo 2 es ahora +1.

### ALGORITMO DE LA MAQUINA DE BOLTZMANN

Incrementar temperatura = 1 / \* bajar la temperatura a 1 "grado" cada vez \*/  
Para todos los patrones de entrada

```

Determinar T = inicio de temperatura /* inicia con una temperatura de 10
"grados" */
mientras T > hacer
{
  hacer
  {
    Seleccionar un neurodo aleatorio
    Calcular el peso neto de entrada para ese neurodo
    (también igual a DE )
    
$$PI = \frac{1}{1 + e^{-(DEI / T)}}$$

    Generar un número aleatorio , N, entre 0.0 y 1.0
    Si N < o = r, se determina la actividad del neurodo a+1
    Si no se determina la actividad del neurodo a-1
  }
  Hasta que todos los neurodos tengan un 50% de oportunidad de
  modificar ese estado
  T = T - incremento de temperatura
}

```

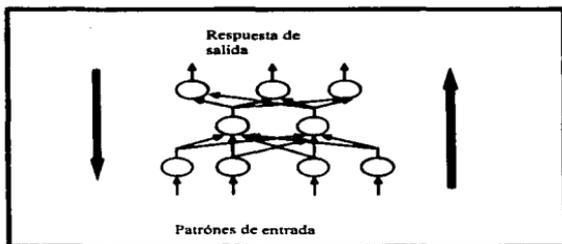
### 7.3 algoritmo de retropropagacion

Las redes de retropropagación dependen de la variación de aprendizaje de error mínimo del Adaline. El Adaline usa una ley de aprendizaje llamada LMS o regla Delta. Una red de retropropagación modifica la regla delta para adecuarla a una red multicapa; el resultado es llamado regla delta generalizada.

Una red de retropropagación opera en una secuencia de dos pasos durante su entrenamiento. Primero, se presenta un patrón de entrada a la capa de entrada de la red. (Figura 7.3.). La actividad resultante fluye a través de la red de una capa a otra hasta que la respuesta de la red es generada en la capa de salida. En un segundo paso, la salida de la red se compara con la salida deseada para ese patrón de entrada en particular. Si no es correcto, se genera un error el cual es repropagado a través de la red de la capa de salida a la capa de entrada, con la modificación de los pesos en las conexiones intracapa de acuerdo al error retropropagado.

Una red de retropropagación, como cualquier otra red neuronal es determinada por la conexiones entre sus neurodos, la función de transferencia usada por los neurodos y la ley de cambio de pesos que controla el entrenamiento de la red. Típicamente una red de retropropagación consiste de tres o más capas completamente conectadas. En otras palabras cada neurodo en cada capa tiene una conexión de salida para cada neurodo en la siguiente capa. Generalmente no hay

conexiones entre neuronos de la misma capa (como en la red kohonen), y usualmente los neuronos se conectan solamente con la capa contigua. Así los neuronos de la capa de entrada normalmente se conectan a los neuronos de la capa media, los cuales a su vez solamente se conectan a los neuronos de la capa de salida.



**Figura 7.3. Una Red Retropropagación aprende con dos procedimientos. La actividad hacia los patrones de entrada fluyen hacia adelante a través de la red, y la señal de error fluye hacia atrás para ajustar**

Los neuronos en la red de retropropagación usan una función de transferencia sigmoidea, o con forma de S. Cada neurona determina su salida en la siguiente etapa. Primero se calcula su peso neto de entrada

$$I_i = \sum_{j=1}^n w_{ij} x_j$$

A continuación, este valor de entrada pasa a través de una función de activación, llamada algunas veces función "squashing". Esto puede ser generalmente cualquier función de forma S, pero comúnmente es la función:

$$f(I) = \frac{1}{1 + e^{-I}}$$

Esta función tiene la propiedad de que su derivación es extremadamente fácil de calcular siendo simplemente:

$$\begin{aligned} \frac{df(I)}{dt} &= \frac{e^{-I}}{(1+e^{-I})^2} = \frac{[(1+e^{-I})^{-1}]}{(1+e^{-I})^2} \\ &= \frac{1^1}{(1+e^{-I})} = 1 - \frac{1}{(1+e^{-I})} = f(I) [1-f(I)] \end{aligned}$$

La figura 7.4 es una gráfica de  $f(I)$ , ilustrando la forma de la curva, un punto interesante es que  $f(I)$  tiene un valor mínimo de 0 y un valor máximo de +1 es totalmente diferenciable con una pendiente positiva. La salida de cada neurodo en la red es generalmente más lejana éste valor de activación o el valor de activación menos un umbral.

El único elemento que permanece definido en la red es la ley cambio de peso, o ley de aprendizaje. La regla delta generalizada especifica el cambio en una conexión dada de peso :

$$\Delta W_{ij} = \beta E f(I)$$

$E$  es el error para este neurodo,  $\beta$  es la constante de aprendizaje, un parámetro entre uno y cero, y  $f(I)$ , es la entrada al neurodo. Esta debe verse similar a la regla delta de la adaline; la principal diferencia es que la entrada de la red es modificada por la función sigmoideal. Hay otras diferencias entre las dos, y esas están en el calculo del valor error  $E$ .

Recordemos que el adaline generó este valor de error al substraer la salida actual del neurodo de la salida deseada. Para la red retropropagación, este mismo procedimiento puede ser usado por la capa de salida, entonces la salida deseada es conocida por esa capa. Pero en la mitad de la capa, un más complejo procedimiento es seguido.

El error en la capa de salida es regresado a los neurodos capa-media y evaluado por la misma corrección de pesos que modificó la activación que fluye hacia adelante. El error de la red en cada neurodo capa-media es de esta manera la suma evaluada de la contribución de error de cada uno de los neurodos capa-salida. Por razones de estabilidad matemática, esta red error evaluado es multiplicada por la derivada de la función activación de los neurodos capa-media. La forma final del cálculo del error es por tanto:

$$E_j^{\text{salida}} = Y_j^{\text{deseada}} - y_j^{\text{actual}}$$

$$E_i^{\text{media}} = \frac{df(j_i^{\text{media}})}{dt} \sum_{j=1}^n (W_{ij} E_j^{\text{salida}})$$

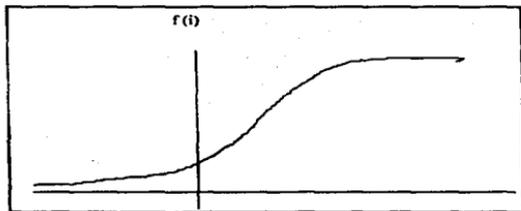


Figura 7.4 La función sigmoideal  $f(i)$ . El rango de la función es limitado por los valores entre 0 y 1.1, y la pendiente (derivada) de la curva es positiva.

En estas expresiones, las indicaciones "salida" y "media" designa la capa del neurodo que está modificando los pesos sobre las conexiones de entrada. El neurodo  $j$  es un neurodo arbitrario en la capa de salida; el neurodo  $i$  es un neurodo arbitrario en la capa media. Nótese la convención del subíndice  $W_{ij}$  es que ella designa el peso de la conexión que va del neurodo  $j$  (en la capa media) al neurodo  $i$  (en la capa de salida).

Esta parece retroceder, pero necesariamente si entonces el uso normal de  $W_{ij}$  es cuando procesaran un flujo de actividad hacia adelante sobre un neuróto de la capa de salida  $j$  colecta entradas sobre todas estas señales entrantes de los neurodos de la capa media  $i$ . Por direcciones opuestas durante el flujo de error de retroceso, subíndices y notaciones son también necesariamente opuestas. La derivación es tomada sobre la actividad del neurodo en la capa-media durante el pase hacia adelante. Para donificar la operación de la red retropropagación, el listado 7.3.1 da un perfil de la operación de la red.

La regla delta generalizada de retropropagación es un sistema de gradiente descendente. La regla delta generalizada mueve el vector peso en tal manera como hacer que esta proyección se mueva bajo una pendiente descendente del tazón error (fig. 7.5). La principal diferencia, sin embargo, es que el error ya no es una simple función paraboloides, porque la retropropagación reparte en entradas complejas multidimensionales y espacios de salida, la simple paraboloides es reemplazado por un función extremadamente compleja en un espacio multidimensional. Donde quizás

existan múltiples mínimas locales y formas complejas de la parte más baja del tazón esto hace más fácil y más difícil tener al error mínimo global. Esto se dificulta más porque hay muchos mínimos locales que pueden estorbar; esto es más fácil porque estos mínimos locales pueden a menudo ser evitados moviéndose lateralmente en otra dirección. Pocas mínimas locales existen en todas las dimensiones del hiperespacio, así hay usualmente una salida de la trampa.

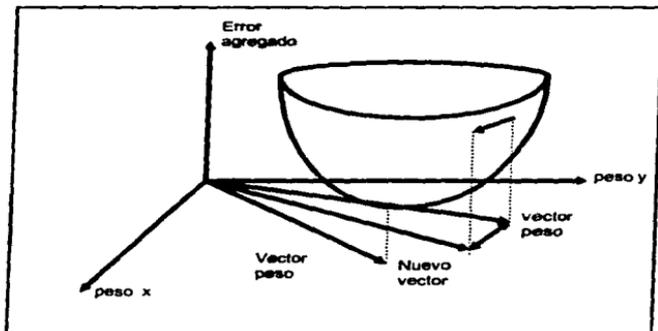


Figura 7.5 La regla delta Generalizada es un sistema de gradiente descendente

No obstante, es totalmente posible que la red retropropagación puede estar sujeta a una de estas trampas. Es fácil notar cuando esto sucede porque el error total detiene su decrecimiento. Muy frecuentemente la red puede ser reinicializada para nuevos pesos (En otras palabras, cambiando la posición de arranque sobre la superficie error), y cuando ha sido convertida, la red usualmente trabaja correctamente

#### ALGORITMO RETROPROPAGACION

##### Conjunto máximo de errores aceptables

/\* valor usuario específico, frecuentemente este elemento caso-peor en cualquier patrón esta dentro del 10 % del valor deseado. \*/

repita

{  
error total = 0;

```

Para cada patrón en el conjunto de instrucccionamiento hacer
{ /* actividad fluye hacia adelante */
continuar al siguiente patrón
para cada neurodo en la capa media hacer
{
calcular la red de entrada = suma calculada de elementos de patrones de
entrada
aplicar la función señal  $f(i) = 1/(1+e^{-x})$ 
/* más comunmente usada la función señal; otras pueden ser sustituidas */
salvar la red de entrada /* necesario para el pase de error hacia atrás y el
cálculo derivativo */
} /* fin para cada neuroda capa-media */
para cada neurodo en la capa de salida hacer
{
calcular la red de entrada = suma pesada de los elementos de salida capa-
media
aplicar la función señal  $f(i) = 1/(1+e^{-x})$ 
representación de salida
} /* fin para cada neurodo capa-salida */
/* paso error retroceso; omitir este paso si no instrucciona activamente la red */
/* note que este no es el más eficiente procedimiento, pero este clarifica el
orden de cada paso */
para cada neurodo en la capa de salida hacer
{
/* calcular el error para cada neurodo capa salida */
calcular el error = salida deseada - salida actual
error total = error total + error
/* note que este puede ser calculado como un error RMS se desea */
}
/* fin para cada neurodo capa salida */

para cada neurodo en la capa media hacer
{
/* retropropagar el neurodo capa de salida a la capa media */
calcular el error de entrada = suma pesada de la capa salida
calcular el error final = error de llegada * (red entrada) * (1 - red entrada)
/* porque de escoger la función simoidal,  $df/dl = f(1) * [1-f(1)]$ 
red entrada, 1, en esta formula son estos neurodos de las redes de entrada
son las que calculan el flujo de activación hacia adelante */
} /* fin para cada neurodo capa media */

para cada neurodo en la capa de salida hacer
{
/* ajustar los pasos entre la capa media y la capa de salida */
para cada peso desde un neurodo capa media hacer

```

```
{
/*estos son los pasos de llegada a la capa de salida */
calcular el cambio de peso =  $\beta + E * 1$ 
/*1 es la actividad de entrada a lo largo de esta conexión; e es el error de
este neurodo  $\beta$  es la constante de aprendizaje. */
peso actualizado
}

/*fin para cada peso*/
}/*fin para cada neurodo capa de salida*/
para cada neurodo en la capa media hacer
{ /* ajustar los pesos entre la capa de entrada y la capa media */
para cada peso desde un neurodo capa - entrada hacer
{ /* estos son los pesos que llegan a la capa media */
calcular el cambio de peso =  $\beta + E * 1$ 
/* 1 es la actividad de entrada a lo largo de esta corrección; E es el error de
este neurodo;  $\beta$  es la constante de aprendizaje */
peso actualizado
}/* fin para cada peso */
}/* fin para cada neurodo capa - media */
} /* fin para cada patrón en la selección del aprendizaje */
} /* fin repetir ... hasta */
hasta ( error total < máximo error aceptable ).
```

## CONCLUSIONES

La utilización de técnicas neurocomputacionales en diferentes áreas, ha permitido realizar tareas que, de otra manera, no serían factibles de ser implementadas. Esto, en parte es debido a lo complejo de la mayoría de los procesos relevantes como es nuestro caso es particular de reconstrucción y corrección de imágenes distorsionadas cuando estas son enviadas a través de un medio de comunicación pudiendo ser este un satélite (*Aunque también se pretende dar una visión general lo más amplia y clara posible de lo que es una red neuronal artificial, de sus similitudes con las redes neuronales biológicas, sus antecedentes (Inteligencia Artificial), sus principios de operación, algoritmos de aprendizaje (perceptrón, retropropagación) aplicaciones y tendencias*), o en ocasiones a las características intrínsecas de las redes neuronales artificiales y que no poseen las otras opciones de procesamiento conocidas. En Particular, la habilidad de las redes neuronales artificiales para procesamiento de imágenes se ve como una de las características más atractivas de la neurocomputación.

Creemos que debido a la gran cantidad de información existente del tema tratado, aunque en nuestro país aun no se tiene el auge como en los países del primer mundo, no se concluyeron algunos puntos, se deja la puerta abierta a la investigación.

# Anexo A1

## LA NEURONA BIOLÓGICA

---

Un corte que atravesara el cerebro permitiría reconocer una capa gris superficial y otra blanca, más profunda. Por ello se habla de sustancia blanca y sustancia gris. El color grisáceo de la corteza cerebral se debe a las células nerviosas, mientras que la sustancia blanca debe su aspecto a las vainas que rodean y envuelven las fibras nerviosas.

En el sistema nervioso central se distinguen dos clases de células, las ganglionares y las de guía. Las ganglionares de forma irregular contienen un núcleo grande y su cuerpo se extiende en numerosas prolongaciones de dos clases : unas más cortas, las dendritas; otras muy largas, las neuritas.

Cada célula nerviosa y sus prolongaciones constituyen una unidad funcional la **neurona**, como estableció el investigador Ramón y Cajal. Las dendritas captan las excitaciones y las transmiten al cuerpo celular, mientras que las neuritas o cilindros ejes emiten los impulsos nerviosos hacia la periferia. En la corteza cerebral existen unos 10.000 millones de neuronas, cuyas neuritas bajan a la base cerebral, tronco, bulbo raquídeo y médula espinal. Cada cilindro eje esta rodeado de una vaina de mielina a la que se debe el color de la sustancia blanca.

La **Neurona** es la célula nerviosa formada por un cuerpo celular y sus prolongaciones (dendritas y cilindro eje o axón). Es la unidad funcional del tejido nervioso.

La neurona es de origen ectodérmico; primitivamente recibe el nombre de neuroblasto, y tiene únicamente una prolongación axónica.

La neurona adulta posee únicamente dos tipos de prolongaciones, y no presenta fenómenos de división por carecer de centrosoma. El cuerpo celular soma, de las neuronas es de tamaño muy variable: las más pequeñas miden 4 micras de diámetro, y las más grandes puede llegar a medir 130 micras ( $10^{-6}$ ). Su forma puede ser piramidal, globulosa, piriforme, etc. y su núcleo guarda respecto al soma diferentes proporciones, según sea el tipo de célula.

En la neurona se puede diferenciar dos procesos, aunque en realidad sean dos modalidades de un mismo fenómeno: el proceso de excitación local y el de excitación propagada. En condiciones fisiológicas, la neurona es excitada localmente por un receptor, otra neurona o por cambios fisicoquímicos, y puede serlo en cualquier punto de su trayecto.

La excitación propagada o impulso nervioso va acompañado de fenómenos metabólicos observándose modificaciones del potencial eléctrico celular. La transmisión del impulso nervioso se propaga en la célula a una velocidad que oscila entre 27 y 132 m/s, siendo mayor la velocidad en las neuronas de los nervios centrípetos que en los elementos centrípetos.

### ***El Potencial de Membrana de Nervio***

Una gran variedad de procesos físicos y químicos ocurren en la celda envlosa, lo mismo ocurre cuando esta conduciendo información, que cuando está en reposo. El oxígeno es consumido, el dióxido de carbono y el calor están saliendo y las sustancias energéticas son metabolizadas formando varias sustancias bio-químicas o alterandolas constantemente.

Estos procesos ocurren en la celda nerviosa o en otro tipo de células. Puesto que la neurona es una celda especializada diseñada para conducir información, un número determinado de fenómenos químico-físicos parecen estar asociados con esta actividad especializada.

Una neurona tiene muchas entradas (quizá varios cientos) y una salida.

Otras tienden a mantener estas en un estado de alerta, ya sea que dispara o no, depende del balance de "cierres" o "abiertos" que existan en las entradas, en cualquier momento dado.

### ***Anatomía de las Células Nerviosas***

Una motoneurona típica, como la que se muestra en el dibujo, tiene un largo axón filamentososo que se origina en una región un tanto engrosada del cuerpo celular, el cono axial. A corta distancia de su origen, el axón o cilindro eje adquiere una vaina de mielina, constituida por un complejo lipoproteico.

La vaina de mielina envuelve al axón, excepto en su terminación y a trechos de aproximadamente 1 mm., formando constricciones periódicas llamadas nodos de ranvier. La vaina de mielina de algunos mamíferos y de la mayoría de los invertebrados es muy delgada o no existe. Del cuerpo celular parten prolongaciones llamadas dendritas a considerables distancias, ramificandose extensamente. La zona dendrítica de la neurona es la membrana receptora de la misma.

El axón es la única prolongación citoplasmática alargada de la neurona, cuya función especializada es la de conducir desde la zona dendrítica.

El axón termina en cierto número de botones terminales o telodendria axónica. El cuerpo celular a menudo está situado en el extremo de la zona dendrítica del axón, pero puede estar intercalado en el axón como en las neuronas auditivas, o unido a él lateralmente como las neuronas bipolares. Su situación no importa por lo que se refiere a la función receptora de la zona dendrítica y a la función transmisora del axón.

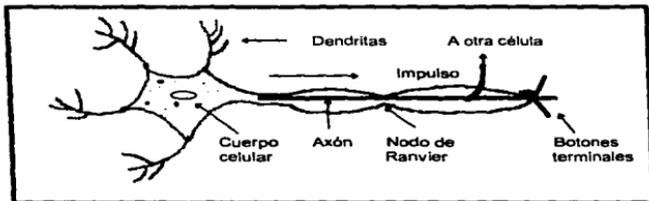


Fig. 1 Llave de "conmutación" en las criaturas vivientes: la neurona. Cuando llegan los impulsos a las dendritas, el cuerpo celular dispara. El pulso de salida deja al axón pasando a la siguiente célula, por medio de la sinápsis.

El tamaño de las neuronas y la longitud de sus prolongaciones varía considerablemente en diferentes partes del sistema nervioso.

La célula nerviosa tiene un umbral bajo de excitación; los excitantes pueden ser: eléctricos, químicos o mecánicos. El trastorno físico-químico creado por estos estímulos, el impulso, es normalmente transmitido (o conducido) desde el cuerpo celular a lo largo del cilindro eje, hasta su terminación. (fig. 1)

### **Transmisión en la Sinápsis**

Los impulsos son transmitidos de una neurona a otra en las sinápsis, es decir, las uniones donde el axón de una célula (célula presináptica) termina en el soma o en las dendritas de otra célula (célula postsináptica). Los impulsos de las fibras presinápticas contribuyen a la iniciación de las respuestas conducidas por la neurona postsináptica, pero la transmisión no es un mero salto del potencial de acción presináptico a la célula postsináptica, sino un proceso complicado que permite la graduación y la modulación de la actividad nerviosa para la función normal. La célula nerviosa activa, genera grandes voltajes que son fácilmente medibles. hay también procesos metabólicos relativos a la actividad nerviosa, pero estos son considerablemente más difíciles de medir directamente. Los eventos eléctricos de la actividad nerviosa constituyen tal vez el mecanismo más fundamental que determina la transmisión de información a lo largo de la fibra nerviosa. Cuando un nervio está en

reposo y no conduce información, existe un estado denominado: **potencial de membrana en reposo**.

Si pudiéramos medir el voltaje de la membrana de un nervio en reposo (la estructura que forma el límite a través de la membrana de la célula nerviosa) podríamos hallar que es cercanamente igual a 0.1 de volt, es decir 100 milivoltios.

Las mediciones más recientes del potencial en reposo fueron obtenidas simplemente aplastando un extremo de un nervio largo (un mazo de fibras nerviosas) o un músculo y grabar "el potencial de lesión", es decir, la diferencia de potencial entre las porciones intactas y dañadas de la muestra. Estas mediciones fueron inexactas, pero demostraron la existencia de una diferencia de potencial.

En atención al potencial en reposo, las celdas nerviosas también exhiben al llamado **potencial de acción propagado**. Cuando una celda nerviosa es activada, un cambio en el voltaje se origina y **viaja** a través del axón.

Este potencial de acción no es simplemente una corriente eléctrica que fluye a través del axón.

Las corrientes en los alambres conductores pueden ser propagados (por efecto de la acción de los campos) a una velocidad cercana a la de la luz (300 000 km/seg). Las corrientes eléctricas son por supuesto, originadas por la acción del potencial, pero la secuencia actual de eventos es mucho más complicada, ya que involucra movimientos de varios iones a través de la membrana.

El mecanismo por virtud del cual las terminaciones presinápticas excitan la neurona postsináptica, se cree que es la secreción de una sustancia excitadora denominada **transmisor excitador**.

El botón terminal esta separado del soma neuronal por una hendidura sináptica con amplitud media de 200 angstroms (unidad de longitud que se utiliza para medir longitudes de onda y dimensiones atómicas. ( $10^{-10}$  m) ).

El botón terminal tiene dos estructuras internas importantes para la transmisión de los impulsos a través de la sinápsis: las vesículas sinápticas contienen al transmisor excitatorio, que cuando es liberado en la hendidura sináptica, excita a la membrana del soma neuronal. El transmisor a su vez, probablemente sea sintetizado por las mitocondrias.

Cuando un potencial se acciona difunde una terminal pesináptica, la despolarización de la membrana origina vaciamiento automático de un pequeño vasículo en la hendidura; el transmisor liberado a su vez origina un aumento inmediato de permeabilidad en la membrana somática subsináptica, que permita que los iones de sodio penetren libremente desde los líquidos extracelulares al interior de

la neurona, provocando un aumento de electronegatividad por fuera y electropositividad por dentro. Si penetra un número suficiente de iones de sodio hacia el interior, la neurona se descarga.

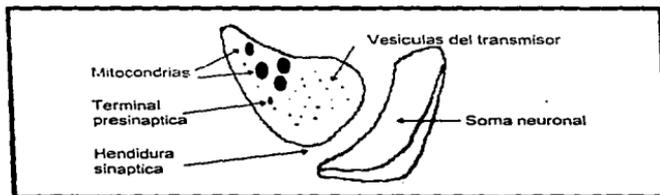
Por desgracia, solo podemos especular acerca del mecanismo por virtud del cual, cuando un potencial de acción alcanza el botón terminal, hace que las vesículas liberen sustancias transmisoras hacia la hendidura sináptica. Sin embargo, el número de vesículas caciadas con cada potencial de acción, disminuye considerablemente:

- cuando la cantidad de iones de sodio en el líquido extracelular esta disminuido.
- cuando la cantidad de iones de magnesio en el líquido extracelular esta aumentada.
- cuando la membrana de la terminal presináptica ya ha sido parcialmente despolarizada antes de transmitirse al potencial de acción, de manera que el potencial de acción es más débil que de ordinario.

Fundandose en estas características, se ha sugerido que la difusión del potencial de acción por la membrana del botón terminal, provoca atracción eléctrica de las vesículas del transmisor por la membrana, y simultaneamente aumenta mucho la porosidad de la membrana para sustancias de la vesícula, permitiendo que estas se difundan por la hendidura sináptica, fig.2.

La penetración de iones de calcio a través de la membrana del botón terminal al tiempo del potencial de acción, quizá sea el estímulo que rompe las vesículas.

Entre las sustancias transmisoras según se cree, se encuentra la acetilcolina, la noradrenalina, la 5-hidroxitriptamina y la sustancia p.



**Fig. 2 Sinápsis de una neurona.  
Propiedades de la sinápsis**

1.- **Conducción unidireccional.**- en contraste con lo que sucede con los potenciales de acción en una fibra, entendiendo por fibra un conjunto (mazo) de axones; que son conducidos en ambas direcciones en la cadena neuronal. La sinápsis es una "válvula de un sólo paso" que determina la dirección de la transmisión.

2.- **Descarga repetitiva.**- una fibra nerviosa, generalmente responde una vez a un estímulo único y breve. Una descarga sincrónica de impulsos que se aplica sobre una vía presináptica de una neurona, con frecuencia, evoca una descarga o tren de impulsos en la descarga postsináptica. La frecuencia de la descarga postsináptica varía por lo general, pero puede alcanzar una cifra de 500 a 1000 impulsos por segundo para períodos cortos. La descarga repetitiva es una forma de amplificación de la actividad nerviosa que ejerce la sinápsis.

3.- **Imposibilidad para transmitir exactamente las frecuencias de las descargas presinápticas.**- cuando se estimula una forma repetitiva de una fibra nerviosa, cada estímulo genera un potencial de acción, siempre que el intervalo entre dos estímulos sucesivos sea mayor que el período refractario. De esta manera la refractariedad es la única limitación que se impone a la transmisión fiel de las frecuencias de estimulación. Con todo, en una cadena de neuronas, la neurona postsináptica puede responder solamente a la primera serie que la alcanza. En general, cuanto más larga sea la cadena neuronal, es decir, cuanto mayor sea el número de sinápsis, menos será su capacidad para seguir fielmente las frecuencias impuestas. El significado especial de esta propiedad de las sinápsis es que el patrón temporal de la descarga que se inicia en la vía presináptica, se altera en forma significativa al atravesar las sucesivas sinápsis de una cadena.

4.- **Suceptibilidad a la asfisia y a las drogas sepresivas.**- la sinápsis es una región de bajo factor de seguridad donde la transmisión es fácilmente bloqueada. Una fibra nerviosa continuara conduciendo muchos minutos después de un paro cardiaco, si bien la transmisión sináptica sucumbe mucho antes.

En general, las cadenas largas con sinápsis múltiples se bloquean mucho más fácilmente que las cortas, más simples. La eficacia de los enestésicos generales, dependen en su mayor parte de la capacidad para bloquear la transmisión sináptica.

Los movimientos, reflejos, las sensaciones y la conciencia se pierden, pero la excitabilidad de los troncos nerviosos se afecta escasamente, como se demuestra por el hecho de que la excitación directa del nervio produce siempre una viva contracción muscular.

5.- **Retraso sináptico.**- la conducción a través de los axones se efectúa en forma continua e ininterrumpida y el ritmo de esta conducción está determinado por el diámetro del axon. El proceso sináptico de la transmisión consume un intervalo medible de tiempo.

Por lo tanto el tiempo de conducción a través de una cadena neuronal es mayor que la suma de los tiempos de conexión axónica, discrepancia esta que aumenta con el número de sinápsis de la cadena.

**BIBLIOGRAFÍA**

II Ciclo de conferencias sobre neurocomputación

**"Aplicaciones y Tendencias"**

Instituto Politécnico Nacional - Centro Nacional de Cálculo

A. cichocki - R. Unbejauen

**Neural Networks for optimization and signal processing**

Arun D. Kulkarni

**Artificial Neural Networks for image understanding.**

Carlos Flores Beceril

Eugenio Villalobos Sánchez

**Elementos de una red neuronal artificial**

C J Pérez Vicente, J Caarrabina and E. Valderrama

**Learning algorithm for neural nets with discrete synapses.**

Dr. Dobb's Journal, April 1991

Casimir C. "Casey" Klimasauskas

**A technique for explaining a neural network's decision-making process.**

Dr. Dobb's Journal, April 1993

**What is cognitive computing?**

**Intelligence from nature conquers tough programming problems.**

Grossberg (1967-1968)

**Yntroduced the additive model of a neuron, involving nonlinear difference-differential equation and explored the use of the model as a basis for short-term memory.**

Lilia Enciso García

**Procesamiento Digital de Imágenes**

Maureen Caudill and Charles Butler

**Understanding Neural Networks**

Volume I y II

Pankaj and Benjamin W. Wah

**Artificial Neural Networks: Concepts and Theory**