

“Esbozo Primario”, el que ha marcado el inicio de una línea de investigación que ha sido tomada como punto de partida por numerosos investigadores del campo. Por otra parte esta idea básica parece coincidir con las investigaciones en neurobiología realizadas hasta el presente.

Un ejemplo (de los muchos existentes) de la grafica que corresponde a estas características –para un modelo dado- se muestra en la siguiente figura

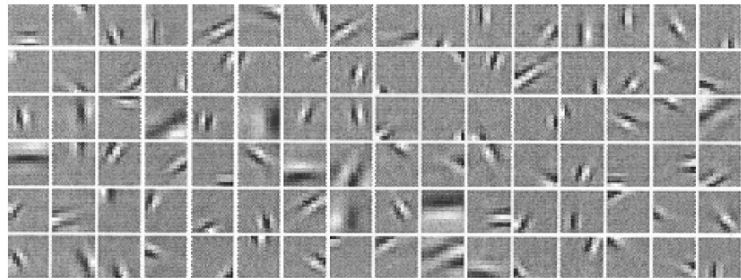


Figura 6.2: Detectores de características empleados en la imagen de la Fig. siguiente.

A su vez la figura siguiente ilustra el resultado de detectar zonas importantes en una imagen a partir de estas características (esbozo primario). Se observa que prácticamente solo aparecen rasgos que pertenecen a objetos de interés y no al fondo –sin información importante- de la imagen.

Esta idea de que el esbozo primario se construye a partir de la detecciones de características, se apoya particularmente en la evidencia de que las neuronas biológicas implementan este tipo de funciones. Lo que ha sido estudiado por diversos neurocientíficos, destacándose entre estos Hubel y Wiesel, quienes comenzaron sus investigaciones a mediados del siglo XX.

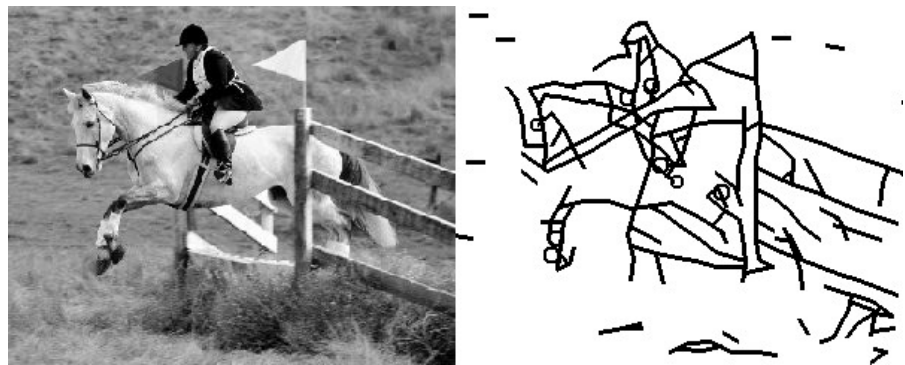


Figura 6.3: Obtención del esbozo primario.

Este concepto de esbozo primario y el modo en que este es detectado, han dado lugar a diversos modelos de visión artificial –en el sentido de reconocimiento visual- que se discuten a continuación.

6.2. Suavizado

El método general de convolución de imágenes digitales, permite entre otras operaciones, una particularmente importante denominada “suavizado”. Como corresponde al método de convolución esto se logra empleando la máscara apropiada. El objetivo de una operación de suavizado es moderar los gradientes de las zonas de alto gradiente. Es decir hacer los bordes menos nítidos, o la imagen más borrosa.

El método principal de suavizado es simplemente la función Gaussiana, es decir los valores de la máscara de convolución corresponden a valores discretos de la campana de gauss.

Veamos un ejemplo. En la figura 4.1 se observa un conjunto de datos expresados como valores discretos. Podrían corresponder, por ejemplo a los niveles de gris de un renglón de una imagen de grises. En la figura 4.1 se incluyen valores negativos, lo que no ocurre realmente en imágenes discretas, pero es válido para comprender el proceso.

En la figura 4.2 vemos una gráfica genérica de la función Gaussiana expresada como función continua y en la 4.3 los valores que corresponden a la función Gaussiana discreta.

Esos valores discretos son usados para formar la máscara de convolución o núcleo, para el caso de que se requiera procesar solo un renglón de la imagen.

El proceso consiste en aplicar la máscara a cada valor original y sus vecinos. Tomemos por ejemplo los valores de los puntos 12, 13, 14, 15 y 16, que son respectivamente;

0.1174 0.1975 0.2349 0.1975 0.1174

y los de la máscara, que son a su vez para los puntos correspondientes

1.0645 0.3893 0.3490 -0.6566 -0.1946

Multipliquemos los valores Gaussianos por los de los datos y sumamos los resultados para obtener el nuevo valor para el punto 14.

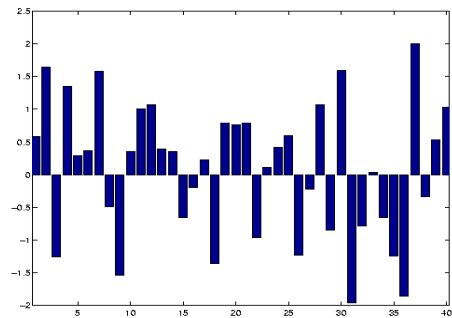


Figura 6.4: Secuencia de valores discretos a los que se aplica un proceso de suavizado.

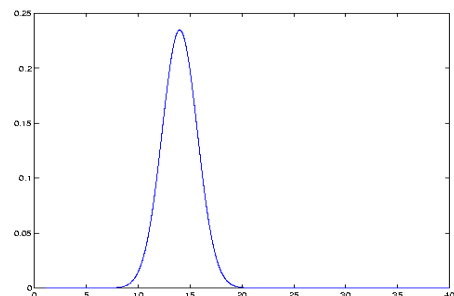


Figura 6.5: Función Gaussiana continua unidimensional.

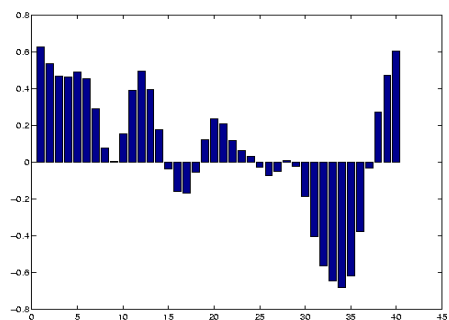


Figura 6.6: Secuencia de valores discretos obtenidos por suavizado a partir de los valores de la Fig.

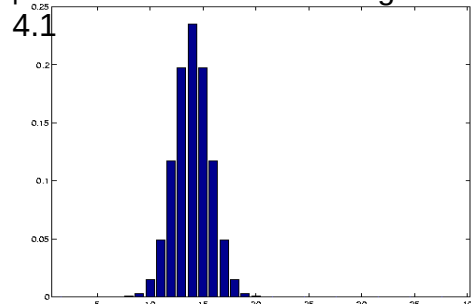


Figura 6.7: Función Gaussiana discreta unidimensional.

... + 0.1174 x 1.0645 + 0.1975 x 0.3893 + 0.2349 x 0.3490 + 0.1975 x -0.6566 + 0.1174 x -0.1946 + ...

Guardamos el Nuevo valor del punto 14, por ejemplo en un nuevo buffer y nos movemos en el original al punto 15 donde repetimos el proceso.

El resultado de aplicar este proceso a cada valor de la Fig. 4.1 se muestra en la Fig. 4.4.

En el caso de las imágenes digitales se aplica una función Gausiana discreta de dos dimensiones a través de una máscara como la mostrada en la Fig. 4.5 (en este caso 5x5).

	1	4	7	4	1
	4	16	26	16	4
$\frac{1}{273}$	7	26	41	26	7
	4	16	26	16	4
	1	4	7	4	1

Figura 6.8: Ejemplo de Mascara de la Función Gausiana discreta bidimensional.

6.3. Pirámide Gausiana

El procedimiento visto de suavizado corresponde a la aplicación a una imagen de un filtro espacial.

El mismo procedimiento puede aplicarse repetidamente sobre la imagen obtenida, (Reduciendo la selectividad de la función Gausiana). De este modo se obtiene una secuencia de imágenes cada vez mas difusa.

Debido a que con cada aplicación de este filtro se pierden detalles de la imagen, es frecuente reducir la dimensión de la imagen obtenida, quitando algunos pixels a la imagen a la que se aplico el suavizado.

Este procedimiento que consiste en lograr un conjunto de imágenes cada vez menos nítidas y eventualmente de menor tamaño se conoce como pirámide Gausiana (en relación a imaginar las imágenes una sobre otra y reduciéndose cada vez mas). La figura muestra el conjunto de imágenes de la misma dimensión pero con distintos niveles de suavizado.

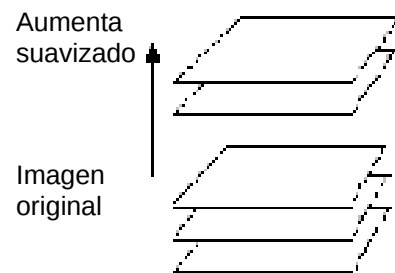


Figura 4.9: La pirámide Gausiana se compone de un conjunto de imágenes derivadas a partir de una imagen inicial para cada una de las uales se aplica un suavizado progresivo.

6.4. Teoría del Espacio Escalar

Haciendo uso de los últimos conceptos vistos, y de una propiedad del suavizado gaussiano, surge una teoría de reconocimiento de objetos llamada Teoría del Espacio Escalar.

La propiedad a la que hicimos referencia podemos explicarla en forma sencilla diciendo que, al aplicar un filtro gaussiano de suavizado a una imagen digital, esta se modifica de un modo tal que pueden desaparecer detalles de la imagen, pero nunca surgir nuevos detalles.

En cuanto a la teoría del Espacio Escalar su hipótesis es que los sistemas de visión biológicos reconocen los objetos en diferentes escalas, y que esto es necesario en visión artificial.

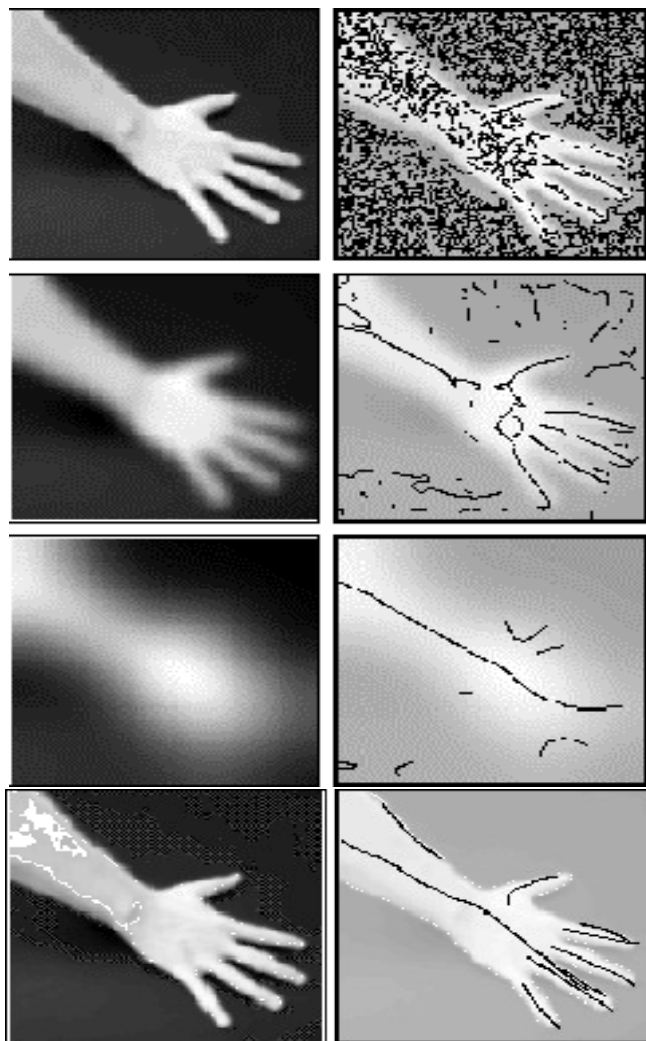


Figura 6.11: Para distintos niveles de suavizado es posible obtener distintas características que corresponden a un mismo objeto en una imagen. (Tales como bordes, esquinas, esqueletos, regiones, etc)

De este modo un sistema puede reconocer “características” (círculos, esquinas, segmentos, etc) que definen un objeto en diferentes escalas y componerlas de modo de reconocer el objeto, en una escala dada. En las figuras se muestra la detección de distintas características de un brazo a diferentes escalas, y su resultante.

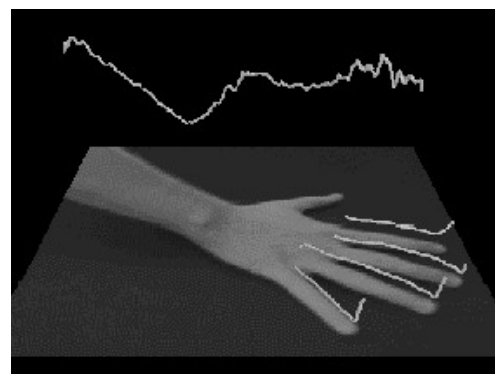


Figura 6.10: El objeto en una imagen, (en este ejemplo la mano y parte del antebrazo) puede ser caracterizado por aquellas características (bordes y otros) que muestran un comportamiento mas estable, es decir que son detectadas mas frecuentemente. Es deseable además que tales características sean invariantes a la rotación, escalación y otras transformaciones. Esto último permite lograr una mayor robustez en el reconocimiento.

6.5. Reconocimiento por Invariantes

Para detectar objetos en imágenes, algunos métodos se basan en extraer las características (en lo posible invariantes a la rotación y otras transformaciones) de un objeto, y almacenarlas asociadas por un índice o etiqueta a ese objeto, conteniendo información acerca de cada característica. (Con la palabra característica nos referimos aquí a un borde, una esquina, un círculo, etc.)

Cuando se examina una imagen, se comparan las características detectadas con las que corresponden a las que describen los objetos almacenados. Se prueba luego la correspondencia de las que corresponden a los objetos más prometedores, y eventualmente se reconocen si aparecen en la imagen.

Este procedimiento requiere realizar extensas búsquedas para encontrar conjuntos de características que correspondan a un objeto. Para esto se emplea un método especial llamado árbol K-D. A menudo se prefiere evitar el empleo de tablas de Hashing para realizar la búsqueda porque por lo general no son tan eficientes.

Un método “popular” para codificar invariantes y luego reconocer objetos se denomina SIFT keys, (descriptores invariantes a la escala y la rotación, -Scale Invariant Feature Transform-). En este método se almacenan los gradientes de gris en el entorno del punto que se desea caracterizar (Fig)

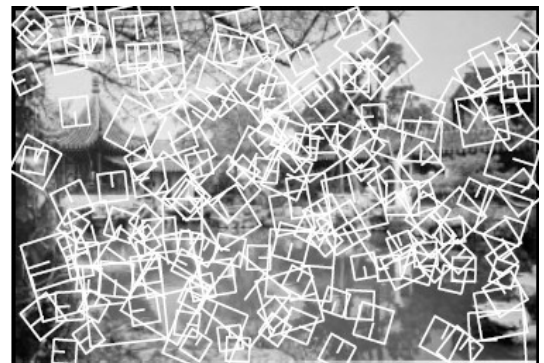
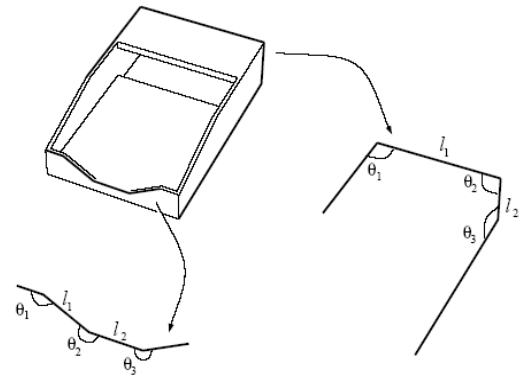


Figura 6.12: En la figura superior se representan esquemáticamente algunas características que corresponden a un objeto. En la figura central se observa un conjunto de características que son detectadas en una imagen al someterla a un proceso de detección de características. La figura inferior muestra las características invariantes que corresponden a ciertos objetos.

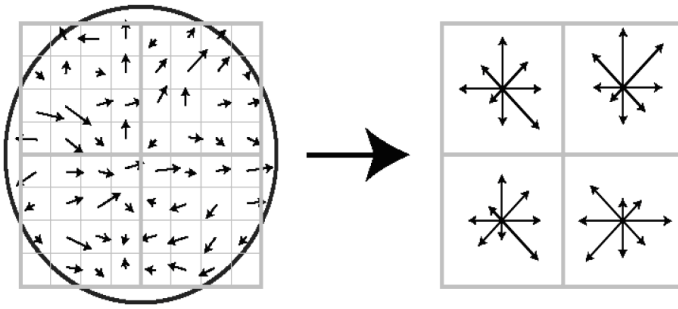


Figura 6.13: Dado un pixel situado en el centro de la imagen izquierda, se evalúan los gradientes de gris alrededor de este punto. Posteriormente se cuantifican y suma obteniéndose los gradientes de la imagen derecha. Estos valores y su orientación son almacenados constituyendo el descriptor de un punto. –Histograma de Gradientes-. (Esto se realiza considerando el espacio de escalas en el que se opera)

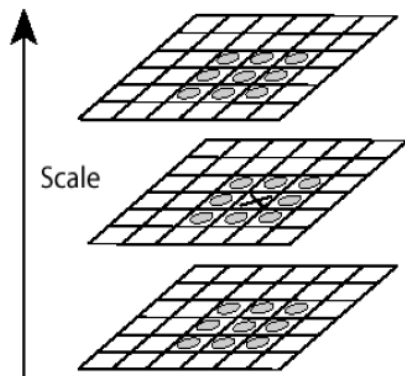


Figura 6.15: Un aspecto adicional en la construcción de las claves SIFT es la necesidad de buscar máximos, no solo en una imagen, sino también entre las adyacentes.

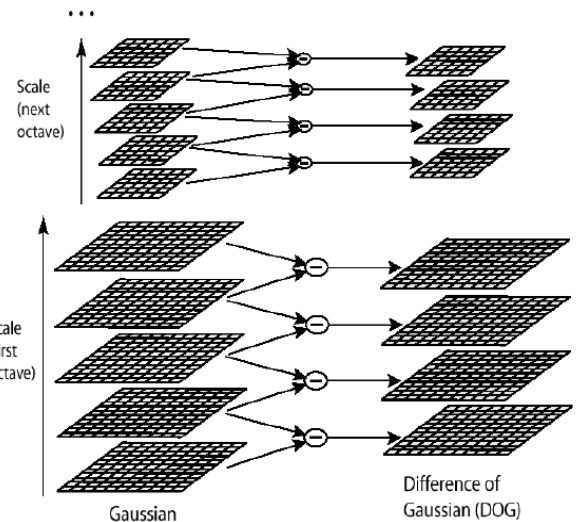


Figura 6.14: El método SIFT emplea una pirámide de escalas múltiple, esto significa que se generan varias imágenes del mismo tamaño para grupos de escalas.

6.7. El Neocognitron

El neocognitron es una red jerárquica de capas múltiples propuesta por Fukushima. Las versiones básicas difieren en su principio de entrenamiento supervisado o no supervisado. Aquí se describe el primer caso. Entre las características favorables de esta red esta la de ser insensible a la posición espacial, escala, ruido y oclusión ante patrones muy sencillos.

La idea básica es una extracción jerárquica de características. Esta consiste en llevar a cabo la extracción en varias etapas. Las características mas simples (usualmente líneas rotadas) se extraen en la primera etapa. Las características mas complejas son extraídas en las etapas siguientes. Es importante considerar que solo la información obtenida de la primera etapa es empleada para extracción de características en las etapas siguientes. En la figura 1.1 se muestra simbólicamente una jerarquía de características correspondiente al reconocimiento del dígito cero.

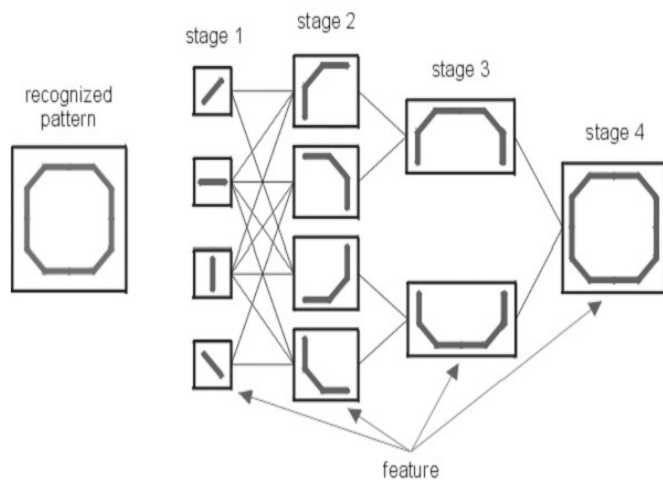


Fig. 1.1 - Extracción jerárquica de características

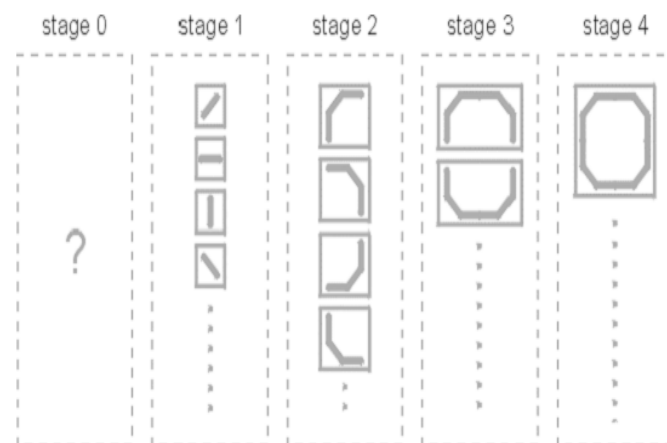


Fig. 1.2: Layers

6.7.1. Etapas

La red esta constituida por un conjunto de etapas. En cada etapa se extrae un nivel de la jerarquía de características. Existe una etapa inicial (Stage 0), que no es empleada en extracción de características. Las etapas usuales se denominan 0 a 4 y se muestran en la figura 1.2 en correspondencia con el ejemplo anterior.

El número de etapas depende de la complejidad de los patrones a reconocer, a mayor complejidad mayor cantidad de etapas. Cada capa del neocognitron consiste en un cierto numero de layers de

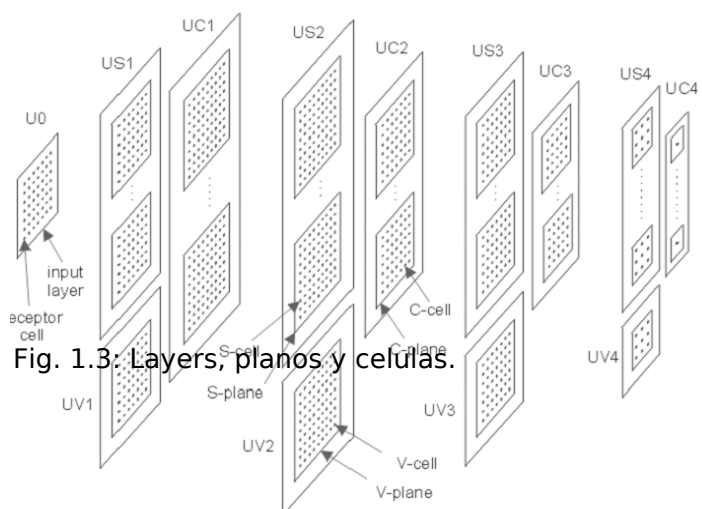


Fig. 1.3: Layers, planos y células.

distinto tipo. En la figura se observan cuatro tipos de layers diferentes. La etapa cero siempre tiene un solo layer. Las restantes consisten siempre de un layer S, un layer V y un layer C.

6.7.2. Planos de células

Cada layer consiste de cierto número de planos de células del mismo tipo. Excepto el layer de entrada para el que no están definidos planos de células.

El número de planos en cada layer depende del número de características extraídas en la correspondiente etapa de la red. Cada layer tipo V siempre consiste de un solo plano. En la figura se representan los planos celulares. Estos pueden ser tipos S V y C.

Cada plano S, V, C y el layer de entrada consiste de células de cierto tipo.

El tamaño de las matrices de células es igual para todos los planos en un layer y decrece al avanzar en las etapas. Cada plano C en la etapa mayor contiene solo una célula. Su salida representa la medida de pertenencia del patrón presentado a la clase que corresponde a esa célula. La dimensión del arreglo de células en cada plano V es la misma que en los planos S de la misma etapa. Existen cuatro tipos de células receptoras, S, V y C.

Cada célula V evalúa la salida de las células C (o las receptoras en el caso de U0) desde ciertas áreas conectadas del layer anterior. La cantidad de células en cada una de estas áreas conectadas es la misma para todas las células V y S en una etapa de la red y son determinadas al construir la red. (Fig 1.5). salida de la célula V representa el promedio de actividad de las células desde las áreas conectadas y son usadas para inhibir la actividad de las correspondientes células S. La especificación exacta de la función es descrita mediante una función matemática.

Cada unidad S evalúa la salida de unidades C (o receptoras) desde ciertas áreas conectadas desde el previo layer C (o layer de entrada). La medida es la misma en todo un layer (Fig. 1.6).

La función de cada unidad S es extraer cierta característica en cierta posición en el layer de entrada (campo receptivo). Para extraer la característica una célula S emplea solo la información obtenida sobre la actividad desde su área de conexión desde la respectiva célula V. Todas las unidades S en un plano S extraen la misma característica (tienen los mismos pesos). El significado de los pesos es mas claro en el primer layer US1. Aquí cada célula S tiene solo un área de conexión y esta es al mismo tiempo el campo receptivo de la célula S. Así los pesos contienen directamente la representación de una cierta característica. En los layers superiores la correspondencia entre la característica extraída y los pesos no es tan obvia. (fig 9.2). El valor de salida de las células S es determinado por la ecuación respectiva. Una ecuación simplificada es:

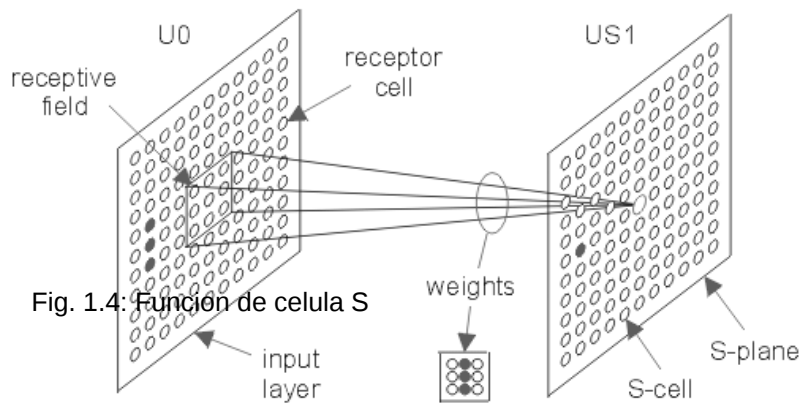


Fig. 1.4: Función de célula S

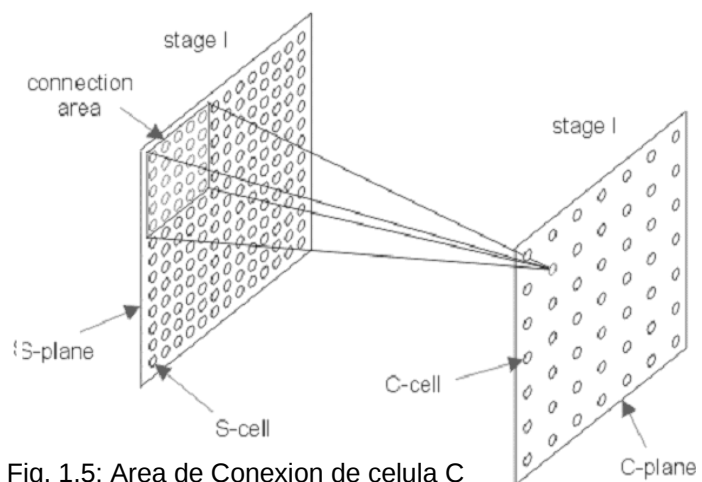


Fig. 1.5: Area de Conexión de célula C

$$u_s = \varphi * [E (a, u_c) / I (r, b, u_v)]$$

Donde:

u_s salida de célula S
 φ^* función no lineal
 E parte excitatoria
 a a-pesos
 u_c salida de célula C desde las áreas de conexión
 I parte inhibitoria
 R selectividad
 B b-pesos
 u_v salida célula V

Las células S no solo pueden extraer las características aprehendidas sino también representaciones deformadas de estas bajo la influencia del parámetro selectividad.

6.7.3. Selectividad

El proceso de extracción de características es influenciado por la selectividad en un grado importante. Para cada layer S es posible asignar diferentes cantidades de selectividad. Modificando la selectividad se modifica el efecto de la parte inhibitoria en el valor de salida de la célula S.

Reducir la selectividad causa la disminución del efecto de la parte inhibitoria. La consecuencia es la disminución de la habilidad de la célula S para distinguir la característica aprehendida correctamente. Es decir que la célula S considerara también características aun mas deformadas como correctas. Las células S se tornan inactivas si la parte inhibitoria es mayor o igual a la excitatoria. (La parte inhibitoria afectara mas a las respuestas mas débiles).

6.7.4 Unidades C

Cada célula C evalúa las salidas de las células S desde un área de conexión desde los planos S del previo layer S. La medida de todas las áreas de conexión en un layer es la misma. (fig 1.8)

El valor de salida de las células C depende de la actividad de las células S del área de conexión. A mayor número de células S activas o mayor actividad tanto mayor la salida de la célula C. Para activar una célula C es suficiente que al menos una S activa este presente en el área de conexión.

Con respecto al solapamiento de áreas de conexión de células S que corresponden a células C vecinas, la actividad de una célula S afecta la actividad de varias

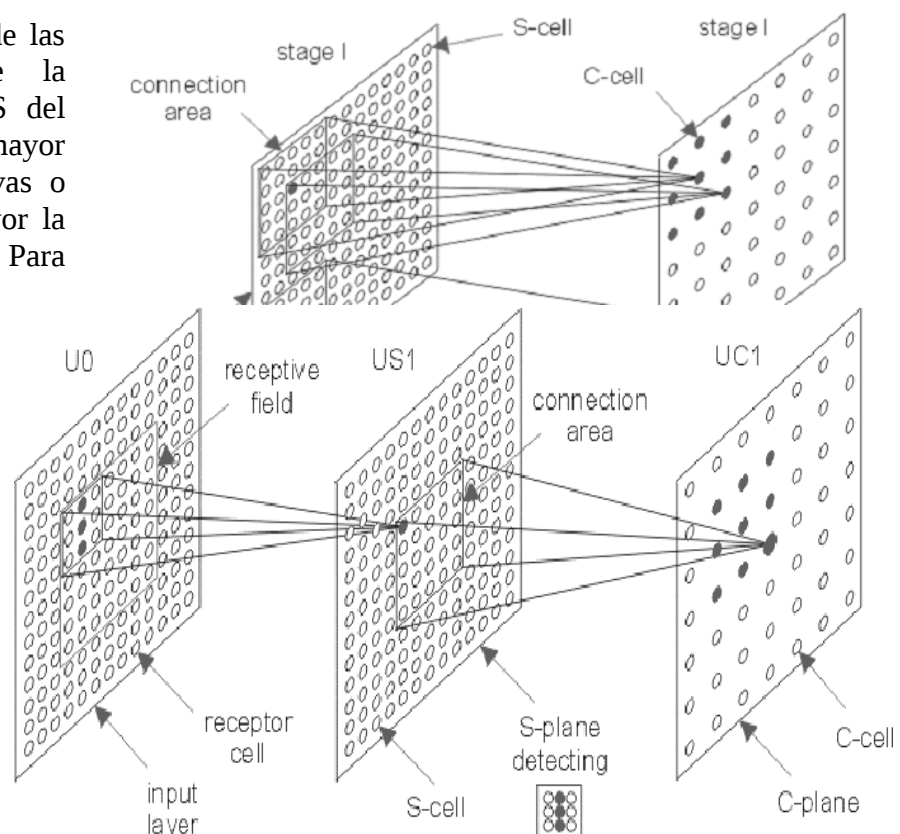


Fig. 1.7: Tolerancia a desplazamientos

células C. Como consecuencia de esto, el plano C contiene una representación borrosa del contenido del plano S (figura 1.10). Un importante aporte de la célula C es la capacidad de esta para comprimir en algún modo el contenido del área de conexión. Otra importante consecuencia de esto es que decrece la densidad de las células en el layer C a la mitad de la densidad de células en el layer previo S (En ciertos casos).

6.7.5. Tolerancia a desplazamientos

Esta propiedad es debida a las células C. El área de conexión de una célula C es mostrada en la figura. La célula C se vuelve activa solo si tiene una conexión con una célula S. Se corresponde con la presencia de una característica en el layer de entrada. Si esta característica se desplaza a otra célula S. Si esta nueva posición corresponde a la célula C permanecerá activa.

El campo receptivo de la célula C consideramos que resulta activada solo si la célula S detecta correctamente la característica.

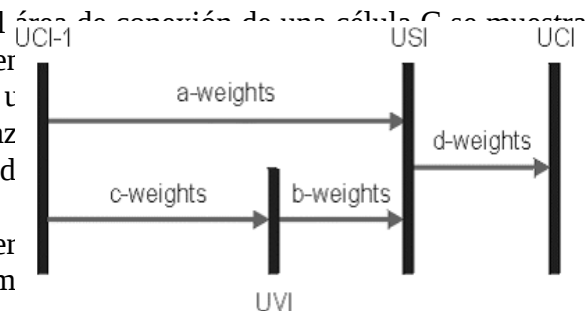


Fig. 1.8: Pesos en el neocognitron

6.7.6. Pesos y conexiones

El neocognitron se caracteriza no solo por el gran número de células sino también por el número de conexiones.

Existen cuatro tipos de pesos en el neocognitron. Fig 13.1.

Pesos Compartidos es un mecanismo del que resulta que todas las células en un mismo plano usan los mismos pesos en sus áreas de conexión. Esto garantiza que todas las células de un plano siempre detectan la misma característica. Podemos clasificarlos según la forma de obtenerlos:

- Pesos modificados por aprendizaje
 - o Pesos a
 - o Pesos b
- Pesos fijos
 - o Pesos c
 - o Pesos d

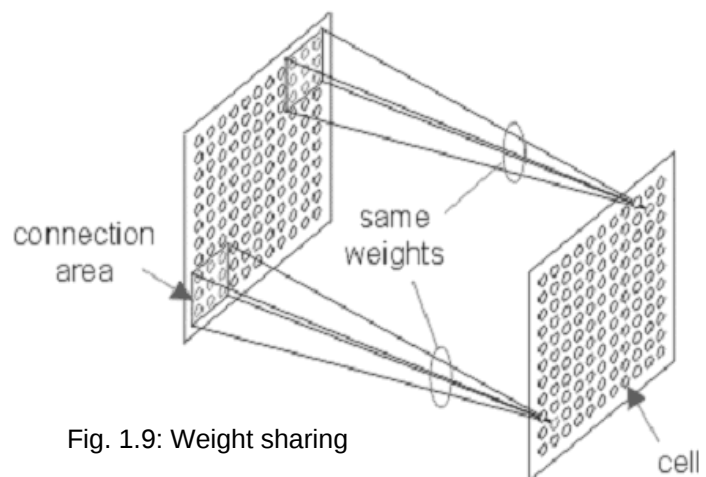


Fig. 1.9: Weight sharing

Los pesos a son el primer tipo de pesos modificados por aprendizaje. Estos se emplean en las conexiones entre células C y S las que corresponden a las áreas de conexión. Las características extraídas por las células S están codificadas en estos pesos a.

Los pesos b son el Segundo tipo de pesos modificados por aprendizaje. Estos pesos son empleados para conexiones entre células V y células S.

Los pesos c son usados en conexiones entre células C y V que pertenecen a las áreas de conexión. Determinados al construir la red. Reducen la transferencia de información de la periferia respecto al centro del área de conexión (filtro gaussiano).

Los pesos d son usados entre células S y C . Correspondiendo a las áreas de conexión. Determinados al construir la red. También ejecutan la función de atenuación periferia-centro.

6.7.7. Aprendizaje

En el aprendizaje el neocognitron procede etapa por etapa adaptando los pesos modificables (a y b) de acuerdo a la respuesta de las partes ya aprehendidas de los patrones presentados. Para cada plano S se emplea usualmente un patrón de entrenamiento y generalmente se presenta solo una vez. Al comienzo se establecen los pesos a y b a cero. Luego se selecciona un plano S del layer $US1$ y en este se selecciona una célula (semilla). El siguiente paso es presentar el patrón de entrenamiento dado para este plano S en el layer de entrada $U0$.

Finalmente se ajustan los pesos de la semilla de acuerdo a las ecuaciones ya citadas.

Debido al mecanismo de pesos compartidos usado el ajuste de los pesos de todas las otras células en ese plano S ocurren simultáneamente.

Si existen mas patrones de entrenamiento para el plano S elegido se presentan sucesivamente repitiendo el proceso, en caso contrario se continua con el siguiente plano S .

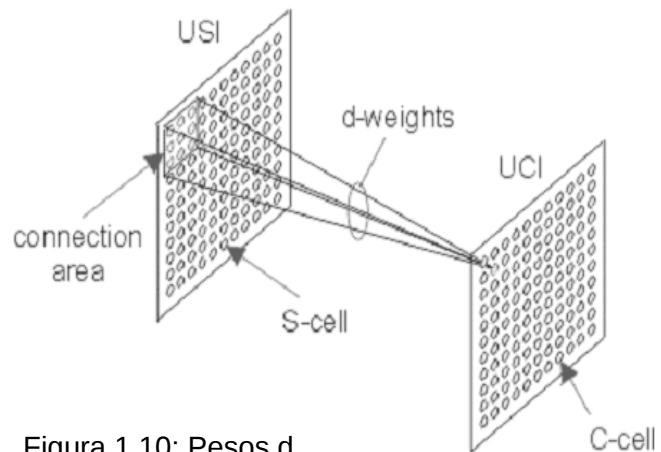


Figura 1.10: Pesos d

