

Capítulo 4

SEGUIMIENTO DE OBJETOS Y ANÁLISIS DEL MOVIMIENTO

En este capítulo se extiende el empleo de los gases neuronales al tratamiento de secuencias de imágenes, en particular para resolver los problemas de seguimiento de objetos y análisis de su movimiento. Aprovechando el carácter dinámico de las redes neuronales auto-organizativas, que permite aprender nuevos patrones de entrada sin tener que reiniciar el aprendizaje, se va a modelar el movimiento de los objetos a partir de los cambios que se producen en la red de interconexión entre los diferentes elementos de procesamiento.

Haciendo uso de este modelo de representación, se plantea el problema del reconocimiento de gestos de la mano. Este se resuelve adaptando la red neuronal a la forma de la mano en cada instante, y readaptándola para cada una de las posturas de un gesto.

El proceso de seguimiento de objetos y análisis del movimiento incorpora operaciones a varios niveles de procesamiento, tal y como se muestra en la Figura 4.1.

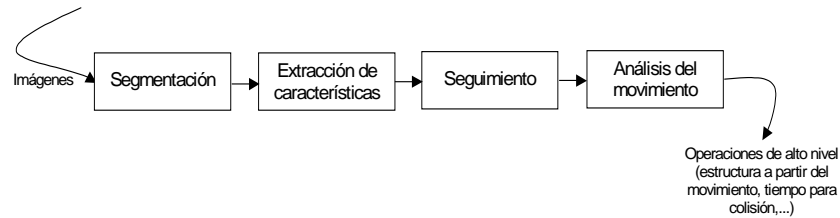


Figura 4.1. Etapas del problema del análisis del movimiento.

4.1 SEGUIMIENTO DE OBJETOS

El seguimiento de un objeto consiste en calcular la posición y forma del mismo en cada una de las imágenes de una secuencia. Como el problema general es muy complicado, se abordan simplificaciones realistas o adecuadas a problemas concretos. Para el movimiento elástico, en particular, se suelen tomar algunas asunciones sobre la dinámica de los objetos (Sonka et al., 1998):

- Velocidad restringida: si se asume que el objeto en movimiento es muestreado a intervalos de tiempo dt , el objeto se encontrará situado dentro de un círculo con centro en su posición anterior y radio $c_{max}dt$, donde c_{max} es la máxima velocidad que puede tomar el objeto.
- Aceleración pequeña: el cambio en la velocidad en un tiempo dt está limitado.
- Movimiento común: todos los puntos del objeto se mueven de forma similar.

Con estas consideraciones, se establece el movimiento de un objeto como las variaciones que realiza un objeto, tanto en posición como en forma, durante un periodo de tiempo:

$$M = [O(t)], t \in [t_0, t_f] \quad (4.1)$$

con sus correspondientes apariencias geométricas y visuales

$$O(t) = [A_G(t), A_V(t)] \quad (4.2)$$

Esto significa que el objeto puede variar su apariencia visual, por ejemplo si se presentan cambios en la iluminación, por lo que la arquitectura representativa debe ser capaz de soportar estas variaciones.

El seguimiento de un objeto puede realizarse mediante la obtención del Grafo Preservador de la Topología del objeto en cada una de las imágenes, y para ello, se deberán tener en cuenta las posibles modificaciones en la apariencia visual del objeto.

Siguiendo el esquema de aprendizaje de la morfología de un objeto propuesta en el capítulo 3, el proceso de seguimiento consta de las siguientes fases:

- cálculo de la función de transformación ψ_T ,
- predicción de la nueva posición del GPT ,
- readaptación del GPT .

4.1.1 CÁLCULO DE LA FUNCIÓN DE TRANSFORMACIÓN

Dado que la apariencia visual del objeto puede variar con el tiempo, es necesaria la modificación de la función de transformación ψ_T en cada instante, para realizar una correcta segmentación del objeto a lo largo de toda la secuencia.

Para ello, se actualiza en cada una de las imágenes los valores de la propiedad T que posee el objeto. Se extraen dichos valores de las posiciones que poseen los nodos del GPT para la imagen anterior:

$$\psi_T(x, y, t) = f_T(I(x, y, t), GPT(t-1)) \quad (4.3)$$

Por ejemplo, si la propiedad visual T es el nivel de gris, y la función que segmenta el objeto f_T es una umbralización, ésta puede variar para cada una de las imágenes; calculando el nuevo umbral a partir de los valores $I(w_i)$, para todos los nodos i del GPT previo.

De modo, que se varía de forma dinámica la función de transformación que calcula la probabilidad de cada uno de los puntos de pertenecer al objeto.

4.1.2 OBTENCIÓN DEL *GPT*

Como resultado del paso anterior, a partir de una imagen $I(t)$, se calcula la correspondiente transformación $\psi_T(t)$. A partir de ésta, se obtiene la representación del objeto, mediante la adaptación del *GPT*.

Para el instante inicial t_i se obtiene el *GPT* correspondiente de forma directa. Sin embargo, dadas las características dinámicas de las redes neuronales auto-organizativas, para las imágenes posteriores, en lugar de partir de cero en el aprendizaje, se toma la red previa como punto de inicio. Esto provoca que con una sola iteración del mapa se consiga su readaptación a la nueva posición del objeto.

PREDICCIÓN DE LA NUEVA POSICIÓN DEL *GPT*

Sin embargo, tal y como está presente en la mayor parte de trabajos relacionados con seguimiento de objetos, es necesario predecir la posición del nuevo mapa, previamente a esta adaptación. Ese proceso facilita el aprendizaje ya que se realiza una pseudo-adaptación considerando los movimientos que están desarrollando los diversos elementos de procesamiento.

Por ejemplo, en la Figura 4.2 se muestra el estado de un *GPT* antes de realizar una predicción. Partiendo desde la representación de la instancia previa del objeto, su readaptación no será correcta, ya que es complicado que las neuronas vuelvan a situarse en el dedo correspondiente. Más bien, se intentarán readaptar de modo que, tras varias iteraciones, se obtenga una nueva representación del objeto, sin que tenga relación la posición que ocupa una neurona con la posición que ocupaba previamente.

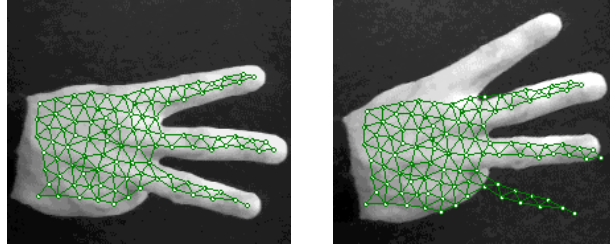


Figura 4.2. Posición del GPT antes de la predicción.

Existen diferentes mecanismos de predicción. Si la frecuencia de muestreo de imágenes es alta en relación con la velocidad del objeto, puede bastar una simple extrapolación, basada en las posiciones previas de cada una de las neuronas

$$w_i^{pred}(t) = 2w_i(t-1) - w_i(t-2) \quad (4.4)$$

Si se desea contar con un histórico de tres GPT se emplea la siguiente expresión:

$$w_i^{pred}(t) = a + b + c \quad (4.5)$$

donde

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0.5 & -1 & 0.5 \\ -1.5 & 2 & -0.5 \\ 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_i(t-1) \\ w_i(t-2) \\ w_i(t-3) \end{bmatrix} \quad (4.6)$$

READAPTACIÓN DEL GPT

A partir del GPT^{pred} , se realiza el proceso de adaptación de la red para la nueva posición y forma del objeto. El resultado final es función de la transformación $\psi_T(t)$ y de $GPT^{pred}(t)$.

La proximidad de la predicción realizada al GPT correspondiente, provoca, en muchos casos, que sea necesaria una sola ejecución del bucle principal del algoritmo de aprendizaje del gas neuronal para completar la readaptación.

De este modo, la segmentación, representación y seguimiento del objeto, se realiza en una única operación de adaptación del grafo.

4.2 ANÁLISIS DEL MOVIMIENTO

El análisis del movimiento comprende el estudio de las variaciones en posición y forma de los objetos.

El análisis del movimiento que realiza un objeto se aborda, generalmente, mediante dos técnicas:

- Basado en el reconocimiento: donde el objeto es reconocido en cada una de las imágenes de la secuencia. De modo, que el movimiento viene determinado por las diversas instancias que toma el objeto a lo largo del tiempo.
- Basado en el movimiento: el análisis viene dado por las trayectorias que toman determinados puntos del objeto. Esta solución es menos costosa ya que se evita el reconocimiento del objeto en cada imagen.

La arquitectura desarrollada basada en la adaptación dinámica de *GPT* permite ambos modos de análisis.

El objeto puede ser reconocido a partir de su *GPT*. En el capítulo 3 se han introducido varias características que permiten identificar las diferentes instancias de un objeto, de modo que se pueda construir una cadena de formas y posiciones del mismo para identificar el movimiento realizado.

Más interesante desde el punto de vista computacional, el modelo aporta una herramienta de análisis basado en el movimiento muy potente, ya que se evita el problema de correspondencia para construir la trayectoria de los puntos de interés. Análogamente a cuando se trabaja con marcadores, la determinación de las trayectorias es inmediata, ya que únicamente hay que realizar un seguimiento a los diferentes nodos del *GPT*.

4.2.1 OBTENCIÓN DE LAS TRAYECTORIAS

El movimiento de un objeto se interpreta como las trayectorias seguidas por cada una de las neuronas del *GPT* :

$$M = [Tray_i], \forall i \in A \quad (4.7)$$

donde la trayectoria viene determinada por la sucesión de posiciones de cada una de las neuronas a lo largo del mapa:

$$Tray_i = \{w_{i_{t_0}}, \dots, w_{i_{t_f}}\} \quad (4.8)$$

El empleo de trayectorias directamente no suele proveer información suficiente para abordar el reconocimiento del movimiento, por lo que se acostumbra a realizar una parametrización de las mismas.

En (Cédras y Shah, 1995) se presentan algunas propuestas de parametrización. Una de las más utilizadas es el cálculo de sus velocidades V_x y V_y , calculadas como:

$$V_{x_t} = \frac{w_{x_t} - w_{x_{t-1}}}{\Delta t} \quad V_{y_t} = \frac{w_{y_t} - w_{y_{t-1}}}{\Delta t} \quad (4.9)$$

Estas velocidades son invariantes a traslación, pero no a rotación o escalado. Por otro lado, la velocidad y la dirección del movimiento se determinan como:

$$V_t = \sqrt{(w_{x_t} - w_{x_{t-1}})^2 + (w_{y_t} - w_{y_{t-1}})^2} \quad (4.10)$$

$$d_t = \arctan\left(\frac{w_{y_t} - w_{y_{t-1}}}{w_{x_t} - w_{x_{t-1}}}\right) \quad (4.11)$$

Estas dos medidas son invariantes a traslación y rotación. Además, la dirección es invariante a escalado. Sin embargo, es muy sensible al ruido, debido al cálculo del arco tangente.

Otra representación usual de las trayectorias es la curvatura espacio-temporal κ , calculada como:

$$\kappa = \frac{\sqrt{A^2 + B^2 + C^2}}{\left((W'_x)^2 + (W'_y)^2 + (t')^2 \right)^{3/2}} \quad (4.12)$$

donde

$$A = \begin{vmatrix} W'_y & t' \\ W''_y & t'' \end{vmatrix} \quad A = \begin{vmatrix} t' & W'_x \\ t'' & W''_x \end{vmatrix} \quad A = \begin{vmatrix} W'_x & W'_y \\ W''_x & W''_y \end{vmatrix}$$

Normalmente, se emplea una discretización para calcular las derivadas, por ejemplo, $W_{x_t} = W_{x_t} - W_{x_{t-1}}$ y $W_{x_t}'' = W_{x_t} - W_{x_{t-1}}$, asumiendo que Δt es constante y, por tanto, $t' = 1$ y $t'' = 0$. Esta representación tiene la ventaja que captura toda la información sobre la trayectoria, sin necesitar dos medidas como en los casos anteriores.

4.2.2 DESCOMPOSICIÓN DEL MOVIMIENTO

El movimiento absoluto del objeto presenta dos componentes (Allmen, 1991):

$$M = M_C + M_R \quad (4.13)$$

definidos como:

- **Movimiento común:** definido como el movimiento que realiza el conjunto del objeto con respecto al observador, de modo que todos sus elementos comparten dicho movimiento. Principalmente determina traslaciones, escalados y rotaciones del objeto.
- **Movimiento relativo:** definido como el movimiento de cada uno de los elementos del objeto con respecto a los otros. Este determina los cambios morfológicos del objeto.

En nuestro caso, el movimiento común va a considerarse como la trayectoria que sigue el centroide del *GPT* :

$$M_C = Tray_{c_m} = \{c_{m_{t_0}}, \dots, c_{m_{t_f}}\} \quad (4.14)$$

Por otro lado, el movimiento relativo va a venir determinado por los cambios de posición de cada una de las neuronas con respecto a dicho centroide:

$$M_R = [Tray_i^{c_m}] \forall i \in A \quad (4.15)$$

donde

$$Tray_i^{c_m} = \{w_{i_{t_0}} - c_{m_{t_0}}, \dots, w_{i_{t_f}} - c_{m_{t_f}}\} \quad (4.16)$$

4.3 EJEMPLO DE APLICACIÓN: RECONOCIMIENTO DE GESTOS

Para mostrar la potencia del modelo desarrollado, se va a tratar el problema del reconocimiento de gestos de la mano. Esta es una de las áreas donde más trabajo se está realizando como aplicación del análisis del movimiento. Este es un problema bastante completo dentro del movimiento no-rígido, siendo además una de las más interesantes de resolver, debido a la complejidad morfológica de la mano. El elevado número de grados de libertad de la mano y los dedos (Figura 4.3) produce multitud de posibles posiciones del objeto (mano) así como de oclusiones, que dificultan considerablemente su seguimiento.

4.3.1 INTRODUCCIÓN

Si bien la comunicación oral puede considerarse como la más importante, los gestos introducen una información adicional que es muy importante para la expresión y comprensión humana. En el caso de ausencia de voz, los gestos adquieren capital importancia. Aunque los gestos son expresados con todo el cuerpo, la mayoría de ellos están relacionados con la mano. Nos comunicamos mediante gestos de la mano.



Figura 4.3. Grados de libertad de la mano (Dorner, 1994).

Los movimientos de la mano tienen diferentes funciones (Cadoz, 1994):

- **Semiótica:** para comunicar información con significado. En este caso los gestos están íntimamente relacionados con el habla
- **Ergótica:** asociada con la noción de trabajo y la capacidad de los humanos de manipular el mundo físico con el objetivo de satisfacer determinadas necesidades. Se incluyen operaciones para empujar, agarrar o soltar objetos
- **Epistémica:** que permite a los humanos aprender del entorno mediante la experiencia o exploración táctil

En este caso se centra la atención en la vertiente semiótica de los gestos de la mano. Dentro del carácter lingüístico de éstos se pueden destacar dos aspectos (McNeill, 1996): gesticulación y lenguaje de signos.

La gesticulación no puede ser considerada como “lenguaje” del cuerpo, sino que complementa la comunicación oral. En el lenguaje de signos cada uno de los gestos corresponde a un concepto, aunque no tenga necesariamente una equivalencia con palabras del lenguaje hablado. Ejemplos de lenguajes de signos son el Lenguaje de Signos Americano (ASL) y el lenguaje de los sordomudos.

Quek (1994) describe diferentes características de los gestos:

1. Los gestos están contenidos en movimientos que comienzan desde una posición de descanso, continúan con una fase de aumento sustancial de velocidad y que finalizan volviendo a una posición de descanso.
2. La mano asume una configuración particular durante el movimiento.
3. Lentos movimientos entre posiciones de descanso no son gestos.
4. Los gestos de la mano deben estar limitados a un cierto área del espacio.
5. Los gestos estáticos de la mano requieren un periodo finito de tiempo para ser reconocidos.
6. Movimientos repetitivos de la mano pueden ser gestos.

Todo esto lleva a considerar (Huang y Pavlovic, 1995) los gestos de la mano como:

- gestos estáticos: caracterizados por la postura que está determinada por una configuración y orientación particular de los dedos y la palma de la mano,
- gestos dinámicos: caracterizados por la configuración de la mano en las posiciones inicial y final de descanso y por su movimiento.

4.3.2 TRABAJOS PREVIOS

Existen diferentes mecanismos para capturar elementos importantes de la mano, pudiendo dividirlos en dos grandes grupos: sistemas basados en visión y dispositivos mecánicos. No hay ningún sistema que aporte todas las características deseadas para la captura de información de la posición, postura y movimiento de la mano, algunos son caros, otros molestos, otros ambiguos en sus lecturas... Los sistemas basados en visión tienen la ventaja de que no requieren que el usuario porte nada sobre su cuerpo, pero tienen el problema de las oclusiones y de los diferentes grados de iluminación. Por otro lado, los dispositivos mecánicos no poseen estos inconvenientes pero son molestos de usar. Los usuarios preferirán un sistema al cual no tengan que prestar atención y que no les limite en su trabajo o movimientos habituales.

Existen multitud de dispositivos mecánicos (Sturman, 1991) para la extracción de la postura de la mano, algunos de los cuales también permiten recuperar la posición y orientación de la misma. Se han diseñado multitud de guantes con transductores que determinan la posición de los dedos.

ANÁLISIS BASADO EN VISIÓN

La generación de trayectorias de movimiento a partir de una secuencia de imágenes implica, normalmente, la detección de elementos en cada imagen, y la correspondencia de cada uno de esos elementos de una imagen en la siguiente. Estos elementos deben ser suficientemente diferenciables para su detección fácil y deben ser estables en el tiempo para que puedan ser correctamente seguidos. Podemos encontrarnos diferentes tipos de elementos: aristas, esquinas, puntos de interés (Figura 4.4), centroides, regiones...

Sin embargo, estos métodos conllevan la solución de varios problemas complejos: segmentación de la mano en un entorno complejo, seguimiento de la posición de la mano, reconocimiento de las posturas de la mano y análisis del movimiento.

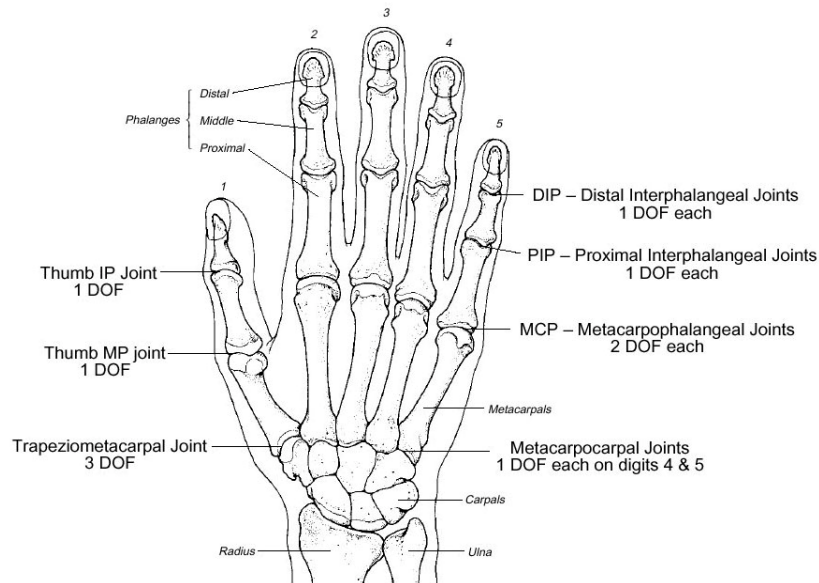


Figura 4.4. Localizaciones típicas de los puntos de interés de la mano (Sturman, 1991).

Para evitar el problema de segmentación, muchos de los sistemas desarrollados emplean marcadores o guantes coloreados para localizar fácilmente los puntos de interés. En otros casos, los gestos son realizados frente un fondo homogéneo para extraer la mano con una simple umbralización.

El resto de problemas son solucionados limitando el tamaño del vocabulario de gestos a reconocer e, incluso, reconociendo únicamente posturas estáticas de la mano.

Los sistemas de reconocimiento de gestos pueden ser clasificados según la técnica empleada para la extracción y seguimiento de la mano: análisis basados en modelos de la mano (Rehg y Kanade, 1993) (Shimada et al., 1998), análisis usando marcadores (Davis y Shah, 1993) o guantes coloreados (Dorner, 1994), y análisis basados en propiedades de la imagen (Starner y Pentland, 1995) (Starner et al., 1998) (Campbell et al., 1996) (Triesch y von der Malsburg, 1996). Relaciones más amplias de trabajos en este campo se presentan en (Huang y Pavlovic, 1995) (Kohler y Schröter, 1998).

4.3.3 DESCRIPCIÓN GLOBAL DEL SISTEMA PROPUESTO

Se propone un sistema de reconocimiento de gestos que emplea una Growing Neural Gas para caracterizar la postura de la mano. El gesto, determinado por las sucesivas posturas, es extraído a partir de las trayectorias que siguen cada uno de los nodos.

Para reconocer el gesto, se realiza una normalización del *GPT* para hacerlo invariante a traslación y escalado. Por otro lado, se normalizan las trayectorias seguidas por los nodos para hacerlas invariantes a rotaciones.

El esquema del sistema completo se muestra en la Figura 4.5.

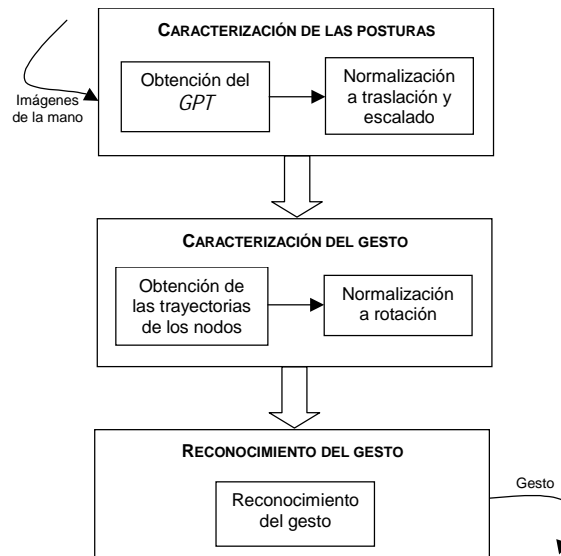


Figura 4.5. Sistema de reconocimiento de gestos.

4.3.4 CARACTERIZACIÓN DE LA POSTURA DE LA MANO CON UNA GNG

OBTENCIÓN DE LA GNG DE LA MANO

Dada una imagen $I \in R^2$ de la mano, se realiza la transformación $\psi_T(x, y) = T(I(x, y))$ que asocia a cada uno de

los puntos de la imagen su probabilidad de pertenecer a la mano, según una propiedad T .

De modo que considerando $\xi = (x, y)$ y $P(\xi) = \psi_T(\xi)$, se aplica el algoritmo de aprendizaje de la GNG a la imagen I , de modo que la red adapta su topología, dando como resultado una esqueletización de la mano.

Diversos resultados de la aplicación del algoritmo de aprendizaje a varias posturas de la mano se presentan en la Figura 4.6.

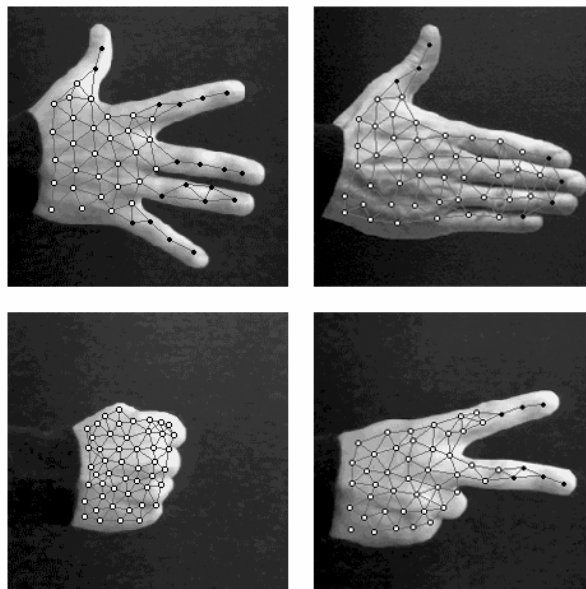


Figura 4.6. Adaptación de la Growing Neural Gas a diferentes posturas de la mano.

NORMALIZACIÓN DE LA GNG

Con el propósito de normalizar la red neuronal obtenida, se realiza una transformación de la GNG para hacerla invariante a traslación y escalado.

Se puede observar que las neuronas con un mayor número de vecinas son aquellas que pertenecen a la palma. Por ello, se

definen los subconjuntos A_p y $A_{\bar{p}}$ como las neuronas que pertenecen o no a la palma:

$$A_p = \{n \in A / \text{vecinas}(n) > \kappa\} \quad (4.17)$$

$$A_{\bar{p}} = A - A_p \quad (4.18)$$

donde κ se establece en función del número de neuronas totales de la GNG.

Una vez encontradas las neuronas que forman parte de la palma, se identifica el centro de la mano, como el centroide de estas neuronas:

$$C_M = (x_{C_M}, y_{C_M}) = \left(\frac{m_{1,0}(n)}{m_{0,0}(n)}, \frac{m_{0,1}(n)}{m_{0,0}(n)} \right), n \in A_p \quad (4.19)$$

A continuación, se procede a normalizar la mano, como sigue:

1. Se sitúa el centroide en el centro de la imagen.
2. Se escala la mano, de modo que la palma tenga un tamaño fijo. Para ello, se obtiene la distancia media de las neuronas más alejadas al centroide, estableciendo esta distancia como radio estándar (Figura 4.7). En la imagen, se representan con puntos blancos las neuronas pertenecientes a la palma, y con puntos negros las neuronas de los dedos.

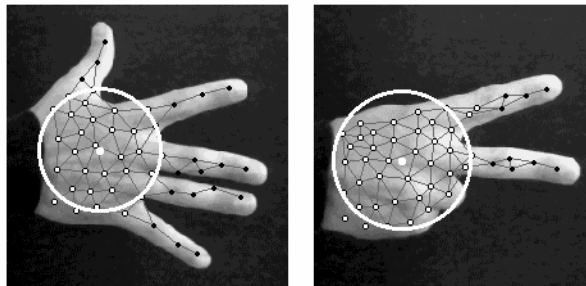


Figura 4.7. Extracción del radio estándar de la palma.

4.3.5 CARACTERIZACIÓN DE LOS GESTOS DE LA MANO CON UNA GNG

OBTENCIÓN DE LA GNG DEL GESTO

Sea $\{M_1, M_2, \dots, M_f\}$ una secuencia de imágenes de la mano que forman un gesto, donde M_1 es la postura inicial y M_f es la postura final del mismo.

Se realiza, como se explicó previamente, el aprendizaje de la GNG para la postura inicial (imagen M_1), normalizando la red obtenida. Posteriormente, se realiza el aprendizaje de la GNG para la siguiente imagen, M_2 , partiendo de la red obtenida para la imagen anterior. Por lo tanto, la adaptación a la nueva postura será muy rápida, dado que únicamente es necesario realizar una iteración al bucle principal de su algoritmo de aprendizaje. De modo similar, se opera con el resto de imágenes, obteniendo para cada una de ellas una GNG que no difiere en demasía de la calculada previamente.

En la Figura 4.8 se muestran las sucesivas adaptaciones de una GNG a un gesto realizado por la mano.

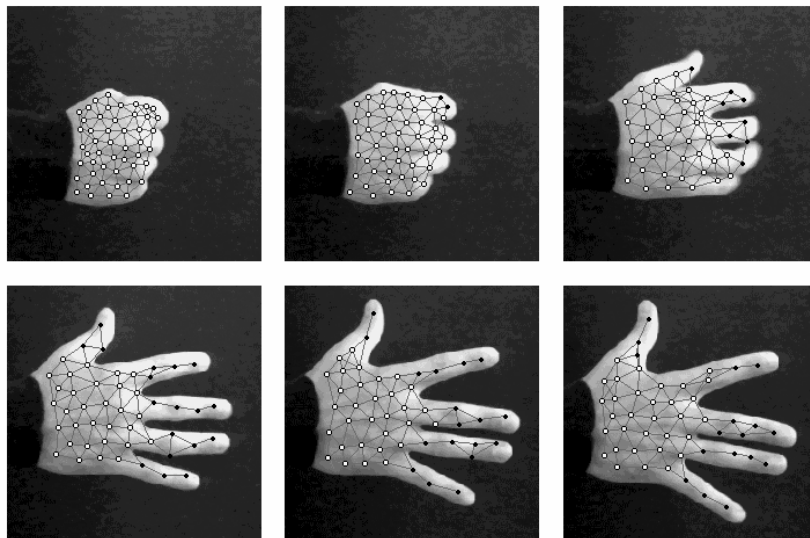


Figura 4.8. Adaptación dinámica de las Growing Neural Gas a un gesto completo.

CARACTERIZACIÓN DEL GESTO

Dado un gesto y su secuencia de GNGs normalizados, se calcula la trayectoria seguida por cada una de las neuronas. Por tanto, para una neurona i se calcula su trayectoria T_i realizada desde la GNG inicial a la final. La trayectoria de cada neurona viene dada por su posición en cada uno de los GNG normalizados:

$$T_i = \{(x_{i_1}, y_{i_1}), (x_{i_2}, y_{i_2}), \dots, (x_{i_n}, y_{i_n})\} \quad (4.20)$$

Sin embargo, a la hora de reconocer un gesto únicamente se van a considerar los puntos inicial y final de cada una de las trayectorias. Esta simplificación ya fue realizada en (Davis y Shah, 1994), rechazando la posible curvatura de las trayectorias, dado que será mínima.

Por tanto, si $G = \{T_1, T_2, \dots, T_n\}$ es un gesto a reconocer, cada trayectoria T_i es determinada por sus puntos inicial y final, $t_{i_0} = (x_{i_0}, y_{i_0})$ y $t_{i_f} = (x_{i_f}, y_{i_f})$.

A partir de esta definición del gesto, se establece el movimiento común de la mano como el movimiento que realizan las neuronas que pertenecen a la palma y, en particular, a la trayectoria que sigue el centroide C_M .

Por otro lado, se define el movimiento relativo de la mano como los cambios morfológicos que sufre a lo largo del gesto, determinados por los movimientos de cada una de las neuronas. Sin embargo, dado que los cambios en la postura de la mano lo determinan los dedos, únicamente van a ser consideradas las trayectorias de aquellas neuronas que, en alguna ocasión durante el gesto, no han pertenecido a la palma; es decir, pertenecieron a los dedos.

En resumen, el movimiento común de la mano viene determinado por los nodos pertenecientes a la palma, mientras que el movimiento relativo viene determinado por las neuronas de los dedos.

NORMALIZACIÓN DE LOS GESTOS

Con el propósito de normalizar el movimiento relativo de la mano para hacerlo invariante a rotación, se extrae la dirección media de las trayectorias seguidas por las neuronas de los dedos, rotándolas para situar ésta en la vertical (Figura 4.9).

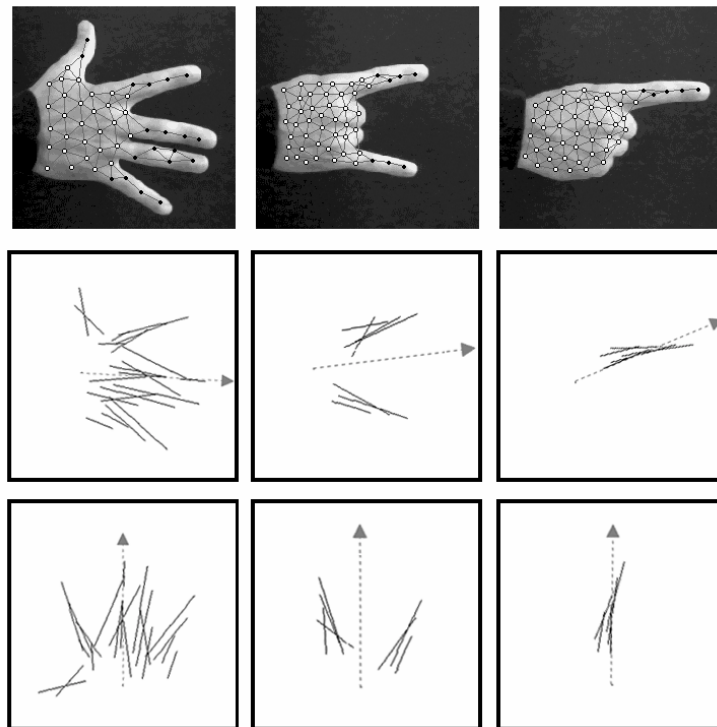


Figura 4.9. Normalización de las trayectorias del gesto.

De este modo, junto con la normalización realizada a cada una de las posturas de la mano, se obtiene una representación del gesto invariante a traslación, escalado y rotación.

4.3.6 RECONOCIMIENTO

El reconocimiento del gesto realizado, se efectúa comparando las trayectorias del mismo con las de otros gestos que se encuentran

en una base de datos, de modo que el patrón $P = \{p_1, p_2, \dots, p_{n'}\}$ más similar al gesto G es aquel que minimiza la expresión:

$$\frac{\sum_{j=1}^n \left\{ \min_{j=1}^{n'} \left(\|t_{i_o} - p_{j_o}\| + \|t_{i_f} - p_{j_f}\| \right) \right\} + \sum_{j=1}^{n'} \left\{ \min_{j=1}^n \left(\|t_{i_o} - p_{j_o}\| + \|t_{i_f} - p_{j_f}\| \right) \right\}}{n+n'} \quad (4.21)$$

Esto es, para cada una de las trayectorias del gesto G se calcula su distancia a la trayectoria más similar del patrón P , obtenida como la suma de las distancias entre los inicios y fines de ambas trayectorias. Este proceso se repite para todas las trayectorias de G . De igual modo, se realiza la operación en sentido contrario, comparando cada una de la trayectorias de P con su correspondiente en G . Por último, se calcula la distancia media dividiendo entre el número total de trayectorias tomadas en consideración.

4.3.7 EXPERIMENTACIÓN Y RESULTADOS

Para realizar los experimentos se ha diseñado un vocabulario de 12 gestos, algunos de los cuales tienen características similares. En la Figura 4.10 se muestran los estados iniciales y finales de cada uno de los gestos. Los gestos A a J comienzan con el puño cerrado, mientras que los dos últimos gestos comienzan con la palma extendida.

Los gestos han sido tomados con una resolución de 320 x 240 puntos y con 256 tonos de grises. Para evitar el problema de segmentación, se han realizado los gestos frente un fondo oscuro, de modo que el filtro ψ_T es una función umbral:

$$\psi_T(x, y) = \begin{cases} 0, & \text{si } I(x, y) < \text{umbral} \\ 1, & \text{si } I(x, y) > \text{umbral} \end{cases} \quad (4.22)$$

La aplicación de este sistema a entornos más complejos, donde se emplean cámaras en color o los fondos sean heterogéneos, se reduce a encontrar la transformación adecuada para realizar una correcta segmentación de la mano.

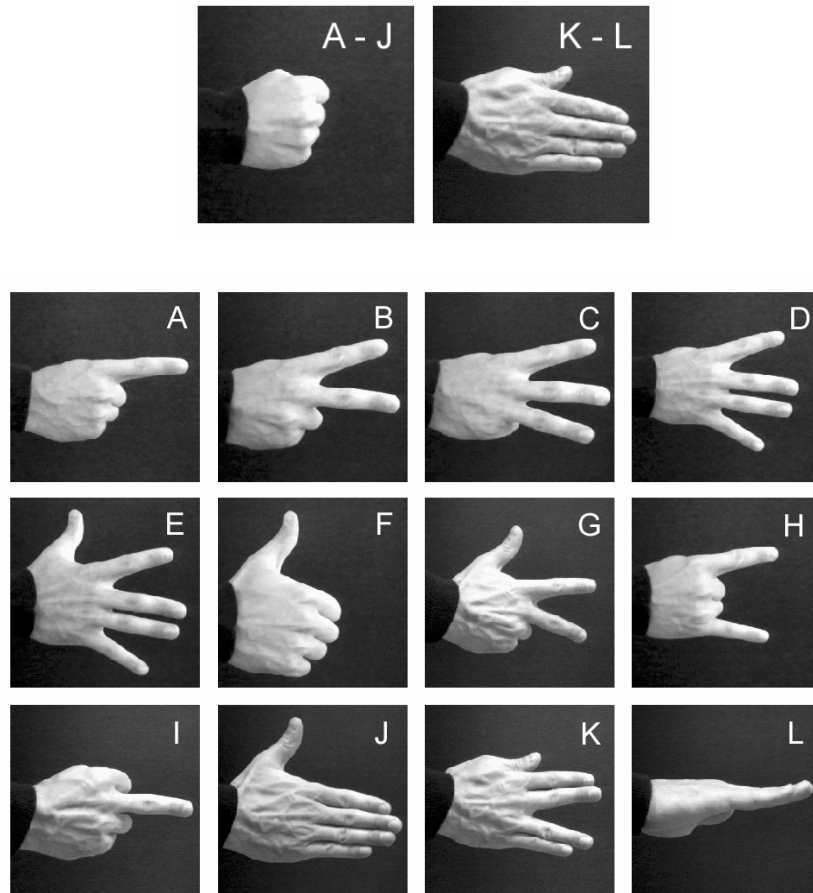


Figura 4.10. Posturas de la mano inicial y final de cada uno de los gestos.

Para entrenar al sistema, se ha realizado cada uno de los gestos 5 veces. Posteriormente, se ha procedido a comprobar su funcionamiento, realizando cada uno de los gestos en 20 ocasiones. En la tabla se presenta la tasa de acierto para cada uno de ellos.

Como se observa, los gestos similares introducen cierta confusión en el sistema, errando algunos de los reconocimientos. Dentro de estos casos, se encuentran los gestos B y H; A, F e I. Asimismo, el gesto D puede ser interpretado como C, si los dedos corazón y anular no están muy separados.

Gesto	Aciertos	Tasa de aciertos
A	20/20	100.00 %
B	18/20	90.00 %
C	19/20	95.00 %
D	19/20	95.00 %
E	20/20	100.00 %
F	20/20	100.00 %
G	20/20	100.00 %
H	19/20	95.00 %
I	19/20	95.00 %
J	20/20	100.00 %
K	20/20	100.00 %
L	19/20	95.00 %
Tasa media de acierto	233/240	97.08 %

Tabla 4.1. Tasa de acierto en el reconocimiento.

4.4 CONCLUSIONES

En este capítulo se ha desarrollado un modelo de representación del movimiento, que aporta diversas ventajas con respecto a la mayor parte de las técnicas existentes.

En primer lugar, las tres primeras etapas del problema del análisis del movimiento (segmentación, extracción de características y seguimiento de objetos) se realizan en un único proceso: la obtención del Grafo Preservador de la Topología del objeto para cada una de las imágenes.

Por otro lado, el modelo es capaz de adaptar su topología a las diferentes evoluciones que puede tomar un objeto (ver punto 1.1.2):

- cambios morfológicos: mediante la obtención del movimiento relativo de cada una de las neuronas,
- cambios fotométricos: mediante el cálculo de la función de transformación, dinámicamente para cada una de las imágenes,
- desplazamientos y giros: los desplazamientos del objeto vienen determinados por el movimiento común del Grafo Preservador de la Topología; mientras que los giros los determina la orientación de las neuronas con respecto al centroide, que queda recogido en su movimiento relativo.

El análisis del movimiento se simplifica debido a que no es necesario realizar un seguimiento de diferentes partes del objeto, localizándolas en cada una de las imágenes (problema de correspondencia). Este viene determinado por los diferentes vectores de características de cada una de las neuronas.

Finalmente, el sistema de análisis de movimiento basado en la adaptación de la Growing Neural Gas a las sucesivas posturas de la mano, ha demostrado ser un buen método para caracterizar los gestos. De hecho, la tasa de acierto es elevada incluso para gestos que pueden ser confundidos con otros muy similares.

