Revista Iberoamericana de Automática e Informática Industrial

CEA comité español de automática

RIAI http://www.revista-riai.org

ISSN: 1697-7912. Vol. 5, Núm. 4, Octubre 2008, pp. 69-76

Modelo Bio-inspirado para el Reconocimiento de Gestos Usando Primitivas de Movimiento en Visión*

Sandra E. Nope* Humberto Loaiza Eduardo Caicedo*****

** Grupo de Percepción y Sistemas Inteligentes (PSI) – E.I.E.E. Universidad del Valle, Ciudad Universitaria Meléndez, Calle 13 nº100-00, Cali, Colombia (e-mail: sandranope@univalle.edu.co)*

***Grupo de Percepción y Sistemas Inteligentes (PSI) – E.I.E.E. Universidad del Valle, Ciudad Universitaria Meléndez, Calle 13 nº100-00, Cali, Colombia (e-mail: hloaiza@univalle.edu.co)*

**** Grupo de Percepción y Sistemas Inteligentes (PSI) – E.I.E.E. Universidad del Valle, Ciudad Universitaria Meléndez, Calle 13 nº100-00, Cali, Colombia (e-mail: ecaicedo@univalle.edu.co)*

Resumen: Se aborda el problema del reconocimiento de gestos usando la información de movimiento con el fin de obtener un modelo bio-inspirado para, en un futuro, utilizarlo en la programación de robots mediante el paradigma del aprendizaje por imitación. En este trabajo se extraen las primitivas de movimiento a partir de imágenes consecutivas, capturadas por una cámara web estándar. Para la programación por imitación de robots se identificó, como primera fase, el reconocimiento de gestos, en el cual es necesario resolver tres aspectos principales: La representación instantánea del movimiento, la integración temporal de dicha información y, la estrategia de clasificación. Estos tres aspectos serán tratados a lo largo de este trabajo y, en contraste con otros, la extracción del movimiento y su codificación está inspirada en el procesamiento del movimiento realizado en el cerebro de macacos. El modelo obtenido fue aplicado al reconocimiento de cuatro tipos de gestos realizados con la mano por diferentes personas. El porcentaje de aciertos varió entre 91.42% y 97.14%, utilizando diferentes estrategias estándar de clasificación.

Palabras Clave: Reconocimiento de gestos, modelo bio-inspirado, primitivas de movimiento, codificación del movimiento, integración temporal, visión artificial.

1. INTRODUCCIÓN

Los sistemas de visión biológica son capaces de extraer muchos tipos de información del ambiente. Algunos pueden detectar color, o ver partes del espectro infrarrojo, o detectar cambios en la polaridad de la luz que pasa a través de la atmósfera; otros, usan varios ojos para determinar la información de profundidad. Pero, definitivamente, el movimiento es un tipo de información que se cree es usada por todos los sistemas biológicos de visión.

El cálculo del movimiento en visión artificial ha sido una de las mayores áreas de investigación, debido a la gran cantidad y diversidad de aplicaciones en las que puede ser empleado. Específicamente, ha sido empleadas en tareas como: (a) codificación y compresión de vídeo; (b) tratamiento de imágenes de satélite; (c) aplicaciones civiles y militares de seguimiento de objetivos y navegación autónoma; (d) evasión de obstáculos en robótica móvil; (e) identificación de anomalías mediante el tratamiento de las imágenes biológicas y médicas; (g) vigilancia y supervisión de lugares; (f) interfaces y realidad virtual; (h) recuperación de la estructura tridimensional; (i) entrenamiento para analizar el desempeño de atletas con respecto a un modelo matemático de desempeño perfecto; (j) monitoreo automático para localización de fallos e identificación de problemas en una línea automatizada; y (k) reconocimiento del habla.
El trabajo que se presenta a continuación utiliza primitivas del movimiento para caracterización y, en el futuro, en la programación de un robot, mediante aprendizaje por demostración; es decir, lograr que los robots adquieran nuevas habilidades a través de la observación y que, de esta forma, aprendan comportamientos complejos e interactúen inteligentemente con el ambiente.

La tarea de construir modelos determinísticos para procesar información visual en el desarrollo de tareas complejas, dentro de ambientes del mundo real, es muy difícil. Los sistemas biológicos han evolucionado hacia una solución simple y robusta, lo que los hace dignos de estudio y de esfuerzos para imitarlos.

Ya que el hombre parece poseer el sistema visual que mejor se ha adaptado a diversas condiciones ambientales, es interesante estudiar el proceso que ocurre en el cerebro durante el procesamiento de movimiento. Sin embargo, por razones obvias, ha sido más estudiado el cerebro de animales, en especial el cerebro de macacos (una especie de monos); debido a la similitud de sus capacidades visuales con las humanas (DeVanois et al., 1974).

Los estudios en neurofisiología sobre el procesamiento de la información visual en el cerebro empiezan siguiendo el recorrido que realiza la información desde los ojos, en donde la retina

transforma los patrones fluctuantes de la luz en patrones de actividad neuronal; pero, esta transformación es sólo el principio de un gran número de transformaciones que se realizan en el sistema nervioso central.

El camino de procesamiento del movimiento en el cerebro del macaco está compuesto por cuatro áreas: área estriada (*V1*), área temporal media *(MT*), área superior media (*MTS*) y *7a* (Bruce y Green, 1990). En (Pomplun et al., 2002) se presenta un modelo inspirado neurológicamente en el procesamiento jerárquico del movimiento primario, en el que se describe un trabajo de simulación que sirvió de inspiración para esta investigación.

Las neuronas en V1 se activan ante una dirección de movimiento particular, y en al menos 3 rangos de velocidad diferentes (Orban et al., 1986). Esta información puede extraerse de los vectores de flujo óptico, una técnica de visión artificial para estimar el movimiento sobre un conjunto de imágenes consecutivas. En este caso, la información de dirección de movimiento se encuentra en el ángulo de los vectores del flujo óptico, mientras que la información de velocidad se encuentra en la magnitud.

Un alto porcentaje de las neuronas en MT se activan en forma similar a V1, mientras que el resto de las neuronas son selectivas a un ángulo particular entre la dirección del movimiento y el gradiente de la velocidad espacial (Treue y Andersen, 1996).

Por su parte, las neuronas en el área MTS se activan ante patrones de movimiento complejos, como: compresión, expansión y rotaciones con campos receptivos que cubren la mayor parte del campo visual (Graciano et al, 1994; Duffy y Wurtz, 1997).

En este trabajo se presenta un modelo bio-inspirado para la representación y percepción del movimiento, usado para la caracterización de gestos en un sistema de visión artificial, y su posterior aplicación en el reconocimiento de cuatro gestos. Este trabajo constituye la fase inicial en la programación de un brazo robótico mediante aprendizaje por demostración. En la sección 2 se describe el sistema; en la sección 3 se explica la representación del movimiento empleada e inspirada en la biología, y como se realiza la integración temporal de la información del movimiento. En las dos últimas secciones se presentan los resultados obtenidos en el reconocimiento de cuatro gestos, las conclusiones y trabajo futuro.

2. DESCRIPCIÓN DEL SISTEMA

La Figura 1 presenta el diagrama de bloques del sistema de reconocimiento de gestos, el cual utiliza una secuencia de vídeo (conjunto de imágenes) como entrada al bloque de "Representación del Movimiento". La salida de este bloque es un conjunto de respuestas neuronales que codifican el movimiento instantáneo, en donde la variable τ corresponde al número de imágenes del vídeo en el que se realiza un gesto. A continuación está el bloque de "Integración Temporal" que recopila la información de movimiento instantáneo provista por los bloques precedentes; su salida es procesada para reducir la dimensionalidad de los datos y facilitar el reconocimiento de los gestos.

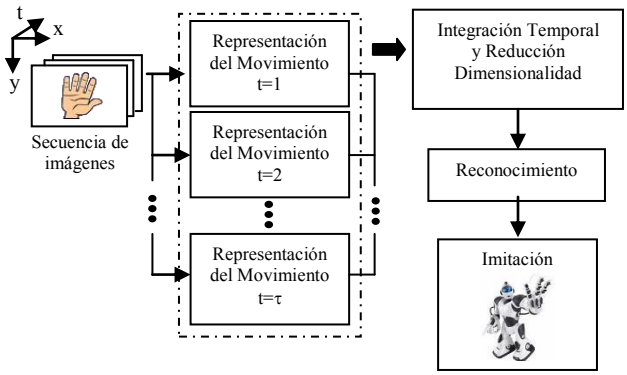


Figura 1. Diagrama de bloques del sistema de reconocimiento de gestos.

Pre-procesamiento

Para la estimación de las derivadas, usadas en la estimación del movimiento, es conveniente pasar las imágenes por un filtro pasa-bajo, en aras de disminuir el efecto del ruido presente en las mismas.

Ya que solo hay interés en analizar el movimiento de la mano que realizan los diferentes gestos grabados en vídeo, se utiliza la información de color presente en las imágenes que lo constituyen. Así, se utiliza la técnica de detección de color piel por píxel propuesta por (Wang y Brandstein, 1999); esta técnica es rápida y simple, ya que únicamente utiliza un umbral en el plano I del espacio de color YIQ, para segmentar el color de la piel. Los resultados reportados en la literatura por los autores fueron de 94.7% de verdaderos positivos y, 30.2% de falsos positivos. En el presente trabajo se adicionó un umbral inferior y superior en el plano Q, con el fin de disminuir los falsos positivos. Los umbrales usados en las pruebas de laboratorio fueron determinados heurísticamente y corresponden a un valor de 13.7 en el plano I, y de -10 y 22 como límites inferior y superior en el plano Q.

La región de la mano queda determinada por los puntos de color piel, conexos mediante la operación morfológica de mayoría en la primera imagen. Esta región permite determinar una ventana que reduce el espacio de búsqueda en imágenes consecutivas, y robustece la segmentación ante la presencia de objetos de color piel que aparezcan repentinamente en la escena.

3. REPRESENTACIÓN DEL MOVIMIENTO

La Figura 2 muestra los bloques constitutivos del bloque de "Representación del Movimiento". Los puntos identificados como pertenecientes a la mano en cada instante de tiempo, son usados por el primer sub-bloque, "Cálculo de flujo óptico", como entrada, así como sus respectivas derivadas espacio-temporales.

3.1 Flujo Óptico Afín

Hay una amplia cantidad de técnicas de estimación de flujo óptico en visión artificial (Nope et al., 2006a). Sin embargo, la técnica del flujo óptico afín fue la que presentó el mejor desempeño para objetos con poca textura, como la mano; también se destacó por su robustez al ruido y a los cambios en la

iluminación. El flujo óptico afín, como lo indica su nombre, combina la ecuación de restricción del flujo óptico con las ecuaciones correspondientes al modelo de formación de imágenes por proyección perspectiva o modelo afín. La aplicación de este algoritmo se basó en el trabajo de (Santos y Sandini, 1996), en el que el flujo óptico afín es determinado por los valores de θ en (1) y descritos por (2).

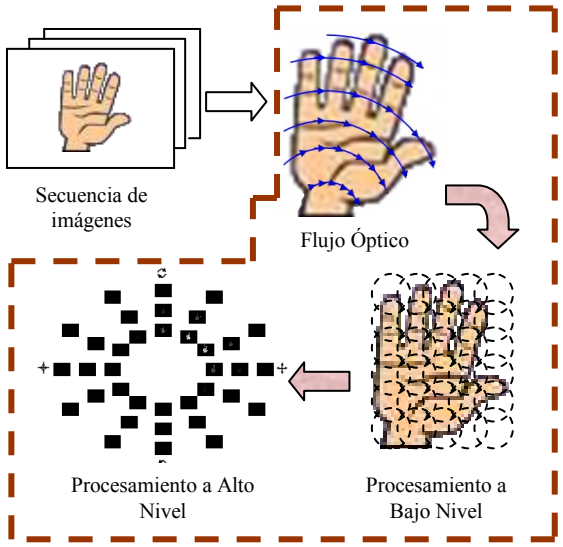


Figura 2. Detalle del bloque "Representación del Movimiento".

$$\begin{bmatrix} I_y & xI_y & yI_y & I_x & xI_x & yI_x \end{bmatrix} \theta = -I_t \qquad (1)$$

donde,

$$\theta = \begin{bmatrix} v_o & v_x & v_y & u_o & u_x & u_y \end{bmatrix} \qquad (2)$$

Donde, I_x, I_y, I_t son las derivadas parciales de la imagen respecto a las coordenadas espaciales (x,y) y el tiempo t.

Con base en estas ecuaciones, el problema de determinar θ, se puede resolver usando seis medidas de las derivadas espacio-temporales de primer orden, aunque usualmente se cuenta con más de seis puntos para su estimación. El procedimiento usado para la estimación de θ, se resume en los siguientes pasos:

a) **Selección Aleatoria**: se escoge un conjunto de puntos (seis o más) de forma aleatoria $[I_x, I_y, I_t]^T$ para obtener una primera estimación de θ.
b) **Cálculo del Error**: se determina el error basado en (1), para estimar θ en el paso anterior.
c) **Repetición del Procedimiento**: se ejecutan los pasos a) y b) hasta alcanzar un error deseado, o hasta un número predeterminado de iteraciones.

El número de puntos elegidos por el algoritmo debe ser lo suficientemente grande como para garantizar que la mayoría de los puntos se ajusten al modelo. Sin embargo, el uso de muchos puntos hace que se requieran mayores tiempos de cómputo para la estimación de θ y, en el caso extremo de usar todos los puntos, se perdería la robustez del método de Ransac y se volvería a una simple solución de mínimos cuadrados, cuya solución es sensible a puntos fuera del modelo real (*outliers*).

3.2 Codificación del movimiento (modelo inspirado biológicamente)

Los vectores de flujo óptico son la representación más simple del movimiento. Sin embargo, para percibir el movimiento que está ocurriendo es necesario procesar la información que contienen dichos vectores. En la codificación del movimiento se aplicaron las ideas principales de (Pomplun et al., 2002) para simular en computador el procesamiento de movimiento en macacos.

La codificación de movimiento realizada aquí se divide en dos partes: Procesamiento a bajo nivel y, codificación del movimiento. La primera reduce la resolución sin pérdida significativa de información, mientras que la segunda permite identificar la velocidad y dirección del movimiento, e identificar entre movimientos complejos.

a) **Codificación a Bajo Nivel:** Para bajar tiempo de procesamiento, sin pérdida relevante de información, se utilizó otra idea de la biología: Los Campos Receptores (*Receptive Fields* - RF). En el cerebro de los macacos, los campos receptores de las neuronas en V1 son circulares y están uniformemente distribuidos a través del campo visual; además, RFs vecinos tienen un solapamiento aproximado del 20%. Los campos receptores se simularon calculando la media de todos los puntos, dentro de círculos fijos solapados, de diámetro D píxeles. La entrada a los RFs corresponde a la matriz de magnitud del flujo óptico, o a la de su ángulo. Matemáticamente, dada la matriz de entrada a los campos receptores $I_{in}(x,y)$, la matriz de salida de los campos receptores $I_{out}(i,j)$ está definida por (3):

$$I_{out}(i,j) = \sum_x \sum_y k(i,j) * I_{in}(x,y) \qquad (3)$$

donde,

$$k(i,j) = \begin{cases} 1/n & if \sqrt{(x-i)^2 + (y-j)^2} \le D/2 \\ 0 & \text{en otro caso} \end{cases} \qquad (4)$$

La Figura 3 está compuesta por cuatro imágenes diferentes, que resultan del movimiento de un octágono que rota en el sentido de las manecillas del reloj; estas imágenes son un ejemplo de ambas situaciones de entrada (la magnitud del flujo óptico o el ángulo del flujo óptico).

La parte superior corresponde a las diferentes entradas de los RFs, mientras que las de la parte inferior corresponden a la respuesta con menor resolución de los RFs. Las imágenes de la izquierda corresponden a la magnitud del flujo óptico, mientras que las de la derecha, al ángulo del flujo óptico. En las imágenes de arriba aparecen círculos blancos que representan los campos receptores; sin embargo, para facilitar la visualización, sólo se muestran algunos de ellos.

b) **Codificación a alto nivel**: La selectividad de las neuronas a una velocidad y dirección particular se simuló mediante filtros Gausianos, sintonizados a una determinada velocidad y dirección de movimiento. Esta respuesta se aproximó mediante la multiplicación de la respuesta de dos filtros Gausianos separados; uno selectivo a una velocidad particular y, otro selectivo a una dirección particular. El

conjunto completo de respuestas neuronales corresponde a las diferentes combinaciones (multiplicaciones) de los diferentes filtros Gausianos, sintonizados a velocidad con los diferentes filtros Gausianos sintonizados a dirección.

c)

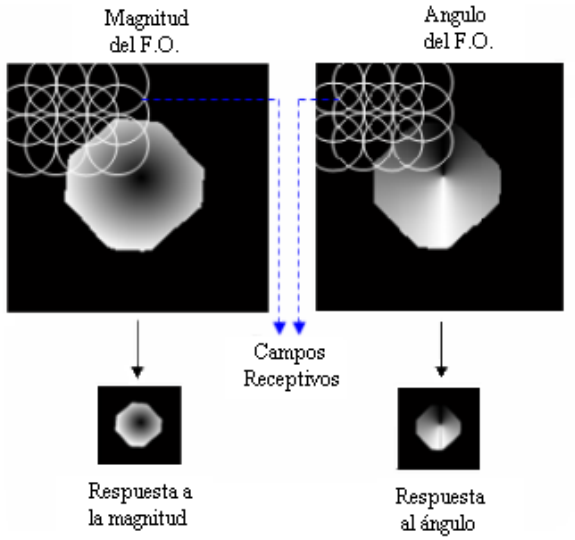


Figura 3. Ejemplo de aplicación de RF a la magnitud y ángulo del flujo óptico ante un movimiento de rotación de un octágono en el sentido de las manecillas del reloj

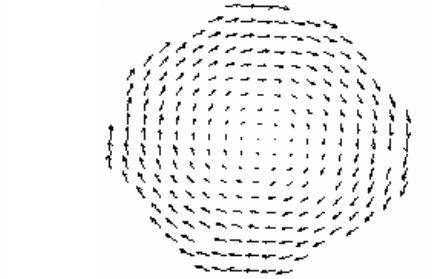
La respuesta de un filtro bidimensional $G(s_k, \theta_p)$, con ajuste de sintonía a la velocidad s_k y a la dirección de movimiento θ_k, está determinada por (5). En donde I_s e I_θ son la respuesta de los RFs ante la magnitud del flujo óptico y ante el ángulo respectivamente y, σ_s y σ_θ, la desviación estándar de los mismos.

$$G(s_k, \theta_p) = e^{\frac{-(I_s - s_k)^2}{\sigma_s}} e^{\frac{-(I_\theta - \theta_p)^2}{\sigma_\theta}} \qquad (5)$$

La Figura 4 muestra en la parte superior los vectores de flujo óptico para un octágono que rota en el sentido de las manecillas del reloj. En la parte inferior aparece el conjunto completo de respuestas neuronales. Los valores de sintonización de los filtros selectivos a velocidad fueron de 0.7, 1.4 y 2.1 píxeles/trama, mientras que la de los filtros selectivos a la dirección del movimiento fueron {0, 30, 60 … 330} grados. Se puede verificar que la respuesta del filtro es, en efecto, más brillante para aquellos vectores de flujo óptico con valores cercanos a los de sintonización del filtro.

Para que la representación del movimiento bio-inspirada obtenida sea robusta ante cambios del punto de vista, se utiliza el ángulo entre los vectores de flujo óptico y el gradiente de la magnitud del flujo óptico (α), en lugar de la dirección de movimiento por sí sola. Así mismo, este ángulo permite identificar movimientos complejos, como rotaciones, expansiones y contracciones.

La Figura 5 presenta, a la izquierda, los vectores de flujo óptico y, a la derecha, los vectores de gradiente de la velocidad para el caso del octágono que gira en el sentido de las manecillas del reloj. Puede apreciarse que el ángulo formado entre estos dos vectores (α), para este caso, es de 90° en todos los puntos sobre el octágono, tomando como referencia los vectores de flujo óptico; así mismo, puede verificarse que los valores de α para una rotación en sentido inverso al de las manecillas del reloj, es de 270°, 0° para un movimiento de expansión, y 180° para un movimiento de compresión. Otros valores angulares significan que el movimiento es el resultado de la combinación de dos de estos movimientos complejos. Por ejemplo, una respuesta fuerte a un ángulo de 60° significa un movimiento producido por una rotación en el sentido de las manecillas del reloj, en mayor medida (↻) y, en menor medida, de un movimiento en expansión (✢).



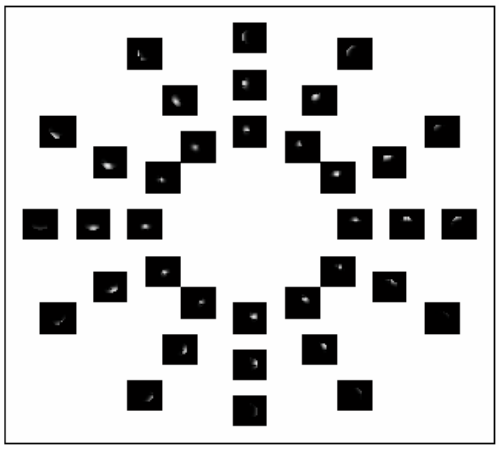


Figura 4. Vectores de flujo óptico en el caso de un octágono que rota en sentido de las manecillas del reloj – arriba. Conjunto completo de respuestas neuronales – abajo.

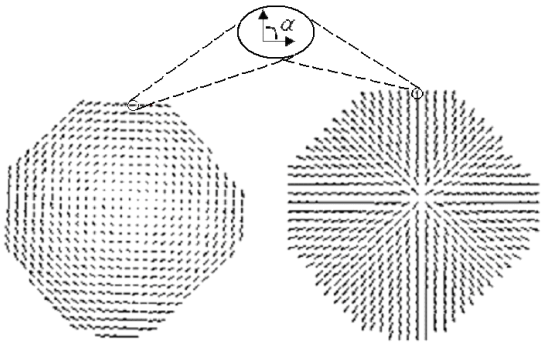


Figura 5. Ejemplo del ángulo formado entre el flujo óptico y el gradiente de la velocidad para un movimiento de rotación en el sentido de las manecillas del reloj.

La Figura 6 presenta el conjunto completo de respuestas neuronales, en el caso del octágono que rota en el sentido de las manecillas del reloj. El anillo interior corresponde a la velocidad más baja e incrementa hacia el anillo exterior. Los tres cuadros, en la parte derecha de la horizontal, corresponden a un ángulo de 0° e incrementa 30° en sentido contrario a las manecillas del reloj. Se aprecia en la figura que, efectivamente, las imágenes más brillantes corresponden **al movimiento de rotación** en el sentido de las manecillas del reloj, en concordancia con las respuestas neuronales más fuertes.

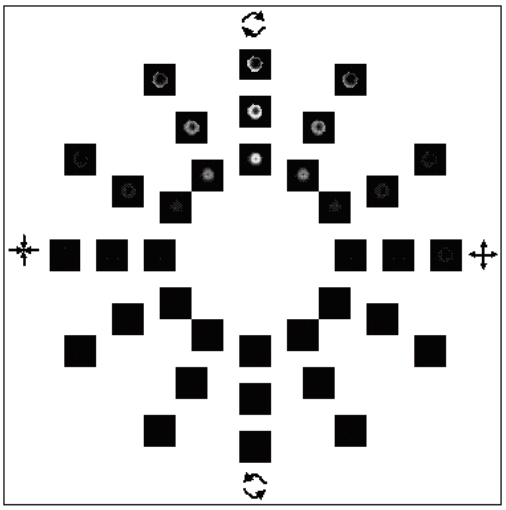


Figura 6. Conjunto completo de respuestas neuronales ante la rotación de un octágono que rota en el sentido de las manecillas del reloj

d) **Integración temporal**: Para analizar la evolución de un gesto se usaron ventanas temporales, que incluyeron toda la ejecución del gesto; de esta forma se logra mayor robustez en el reconocimiento. Los autores (Bobick y Davis, 2001) proponen dos plantillas temporales para el reconocimiento de movimientos: Una Imagen de la Energía del Movimiento (*Motion Energy Image* – MEI) y, una Imagen de la Historia del Movimiento (*Motion History Image* – MHI). Para el reconocimiento de gestos, en este trabajo se propone una Imagen de la Historia del Movimiento, que se inspiró del trabajo de (Bobick y Davis, 2001).

Para construir cualquiera de las plantillas temporales, inicialmente se estima una imagen binaria, que indica las regiones con respuestas neuronales más fuertes en cuanto a velocidad y dirección del movimiento $D_\theta(s_k,\theta_p,t)$ de acuerdo con (6), o con respuestas neuronales más fuertes en cuanto a velocidad y clase de movimiento complejo $D_\alpha(s_k,\alpha_p,t)$, de acuerdo con (7).

$$D_\theta(s_k,\theta_p,t)=\begin{cases}1 & si\ G(s_k,\theta_p,t)>th_\theta\\ 0 & en\ \ otro\ \ caso\end{cases} \quad (6)$$

$$D_\alpha(s_k,\alpha_p,t)=\begin{cases}1 & si\ G(s_k,\alpha_p,t)>th_\alpha\\ 0 & en\ \ otro\ \ caso\end{cases} \quad (7)$$

c.1. Imagen de la historia del movimiento modificada: Sean $H_\theta(s_k,\theta_p,t)$ la imagen de la historia del movimiento modificada para $G(s_k,\theta_p,t)$ y, $H_\alpha(s_k,\alpha_p,t)$ para $G(s_k,\alpha_p,t)$, definidas por (8) y (9) respectivamente. En esta representación propuesta, los puntos con respuesta neuronales más fuertes corresponden a aquellos que se activaron reiterativamente en el tiempo.

$$H_\theta(s_k,\theta_p,t)=\mathrm{H}_\theta(s_k,\theta_p,\mathrm{t}\text{-}1)+1 \quad si\ D_\theta(s_k,\theta_p,t)=1 \quad (8)$$

$$H_\alpha(s_k,\alpha_p,t)=\mathrm{H}_\alpha(s_k,\alpha_p,\mathrm{t}\text{-}1)+1 \quad si\ D_\alpha(s_k,\alpha_p,t)=1 \quad (9)$$

La Figura 7 presenta un ejemplo de la MHI modificada, obtenida a través de (8) y (9) para el caso del gesto 1 (rotar la mano a la derecha y luego a la izquierda -saludar).

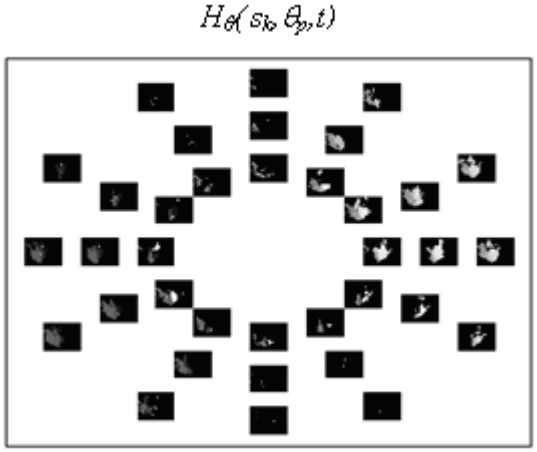


Figura 7. Ejemplo de la Imagen de la Historia del Movimiento para el Gesto 1 (rotar la mano a la derecha y luego a la izquierda).

c.2. Reducción de la dimensionalidad: Estas imágenes contienen información redundante y poseen una alta dimensionalidad, por lo que se hace necesario reducir la dimensión de los datos, con el fin de facilitar el proceso de reconocimiento de los gestos. Esto se realizó mediante el cálculo de histogramas sobre las dos imágenes.

Sean h_θ el histograma para $H_\theta(s_k,\theta_p)$, y h_α el histograma para $H_\alpha(s_k,\alpha_p)$, de acuerdo con (10) y (11) respectivamente.

$$h_\theta(s_{k=i},\theta_{p=j},t)=\frac{1}{M}\sum H_\theta(s_i,\theta_j,t) \quad (10)$$

$$h_\alpha(s_{k=i},\alpha_{p=j},t)=\frac{1}{M}\sum H_\alpha(s_i,\alpha_j,t) \quad (11)$$

Donde M corresponde al tamaño de una respuesta neuronal cualquiera, $H_\theta(s_k,\theta_p)$ o, lo que es lo mismo, el tamaño de una respuesta neuronal $H_\alpha(s_k,\alpha_p)$.

Los histogramas se conforman por el promedio de las respuestas de cada una de las 36 neuronas y, pueden verse como un vector que contiene las diferentes probabilidades de que se esté realizando determinado movimiento complejo, o un movimiento en determinada dirección a una velocidad dada.

En el reconocimiento de gestos, la representación del movimiento que finalmente se utiliza es la proyección de los histogramas en espacio producido por el análisis de componentes principales (*Principal Component Analysis* – PCA).

La Figura 8 presenta la distribución de los datos de acuerdo con las tres primeras componentes principales (e1, e2 y e3) para cada uno de los cuatro gestos empleados. Los asteriscos '*' corresponden a histogramas del gesto 1 (rotar la mano a la derecha y luego a la izquierda -saludar); los círculos 'o' al gesto 2 (bajar y subir la mano – abanicar); el signo de suma '+' al gesto 3 (rotar mano en sentido horario - rotar); y, los triángulos 'Δ' al gesto 4 (acercar y alejar la mano de la cámara). De acuerdo con la figura, los gestos 2 y 4 aparecen traslapados, pero podrían ser separados a través de las otras 15 componentes que no aparecen en la gráfica.

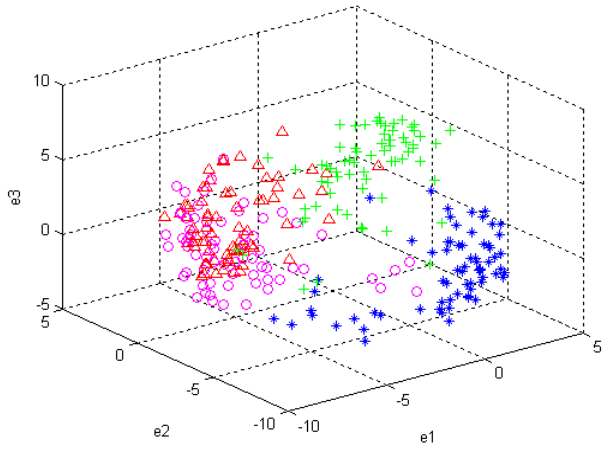


Figura 8. Gráfica de los tres componentes principales para los cuatro gestos de acuerdo con los datos de entrenamiento.

4. APLICACIÓN AL RECONOCIMIENTO DE GESTOS

La codificación de movimiento anteriormente descrita fue aplicada al reconocimiento de 4 gestos: El gesto 1 corresponde a rotar la mano en sentido inverso a las manecillas del reloj y devolverse (saludar). El gesto 2 corresponde a bajar y subir la mano (abanicar). El gesto 3 corresponde a rotar la mano en sentido inverso a las manecillas del reloj. El gesto 4 corresponde a acercar y alejar la mano respecto de la cámara.

La base de datos empleada contiene 70 secuencias de vídeo, grabadas en el laboratorio para cada uno de los cuatro gestos, de las cuales 35 fueron usadas para el entrenamiento y, las 35 restantes para la validación.

Para el reconocimiento de gestos se usaron diferentes clasificadores estándar: Paramétricos (clasificador Bayesiano), No-paramétricos (el vecino más próximo, los *k-vecinos* y distancia al centroide), y Redes neuronales (probabilísticas y perceptron multicapa). La Tabla 1 resume los resultados de clasificación con dichas técnicas de reconocimiento de patrones, si se conserva el 89,23% de la información; esto es, si se usan 18 componentes principales.

En la Tabla 1, se observa que las redes neuronales, junto con la técnica de los *k*-vecinos (k = 5), presentan el mejor desempeño en el reconocimiento de los gestos.

Tabla 1. Resumen de los porcentajes de éxito en el reconocimiento de gestos

Técnica	Porcentaje de éxito (%)
Bayesiano	92.14
Vecino más próximo	95.00
k- vecinos	94.29
Distancia al centroide	91.42
Redes Neuronales Probabilísticas	95.00
Redes Perceptron Multicapa	97.14

Las Figuras de la 9 a la 14 presentan los resultados de reconocimiento, en forma más detallada para cada una de las técnicas. Un reconocimiento perfecto se presenta cuando los valores en la diagonal son del 100%, y cero en el resto.

Los resultados se leen de la siguiente forma: las etiquetas en la parte inferior de las figuras indican con cual gesto fue confundido el gesto identificado por las etiquetas ubicadas a la derecha de las figuras. Por ejemplo, en la Figura 9, el valor más alto por fuera de la diagonal es de 11.43% en dos bloques: uno que surge de la intersección de G3 (inferior) y G1 (derecha) y, el otro, de la intersección de G2 (inferior) y G4 (derecha). Estos valores indican, respectivamente, el porcentaje de veces en que el gesto 1 (G1) fue erróneamente identificado como gesto 3 (G3) y, el porcentaje de veces en que el gesto 4 (G4) fue erróneamente identificado como gesto 2 (G2).

Por otro lado, el mayor error se presenta entre el gesto 2 y 4. Esto se debe a que el ángulo entre la magnitud del flujo óptico y el gradiente de la velocidad, es similar en ambos gestos, y se diferencian más por la dirección del movimiento.

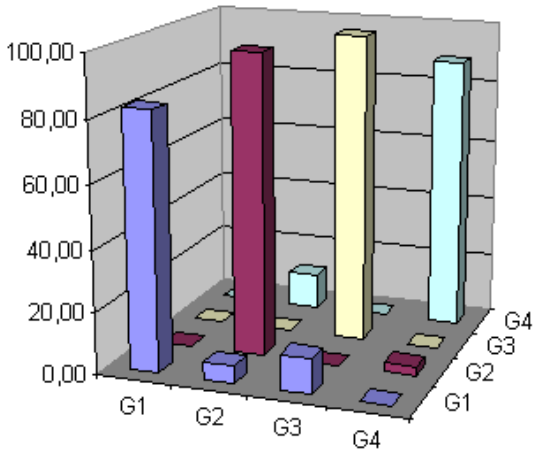


Figura 9. Gráfica de confusión usando la técnica Bayesiana.

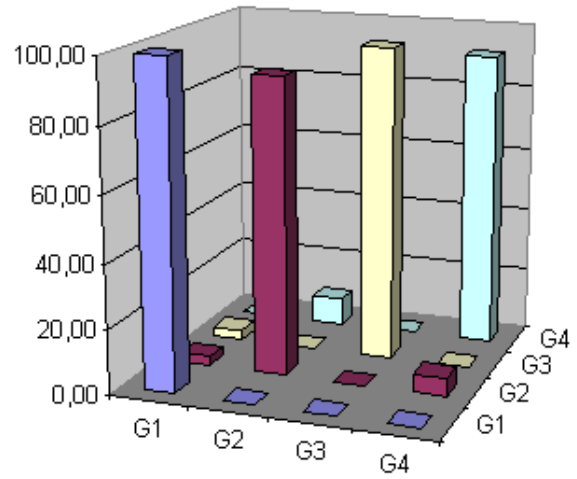


Figura 10. Gráfica de confusión usando la técnica del vecino más próximo.

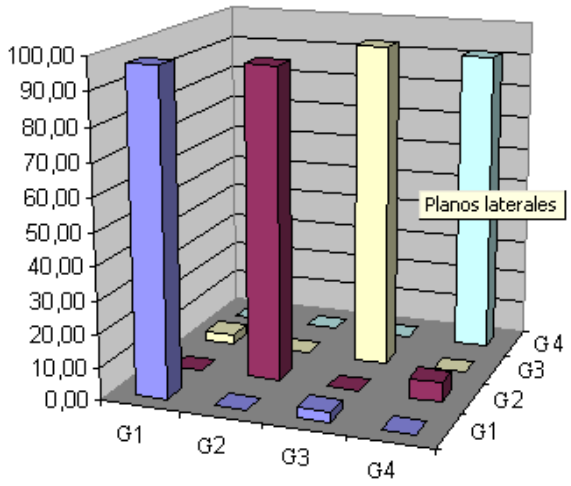


Figura 11. Gráfica de confusión usando la técnica de k-vecinos.

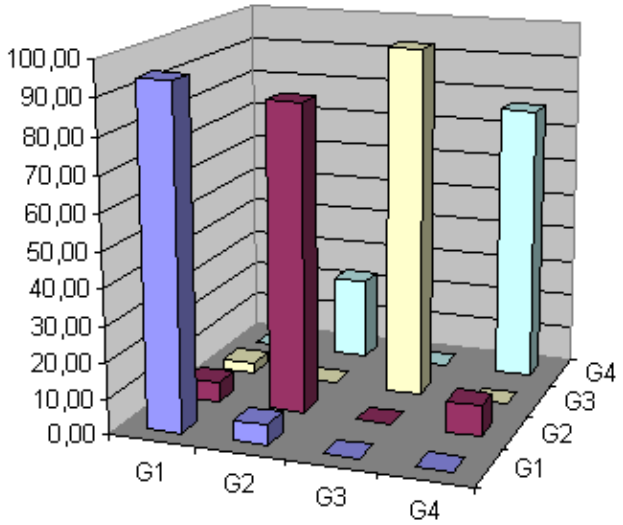


Figura 12. Gráfica de confusión usando la técnica de distancia al centroide.

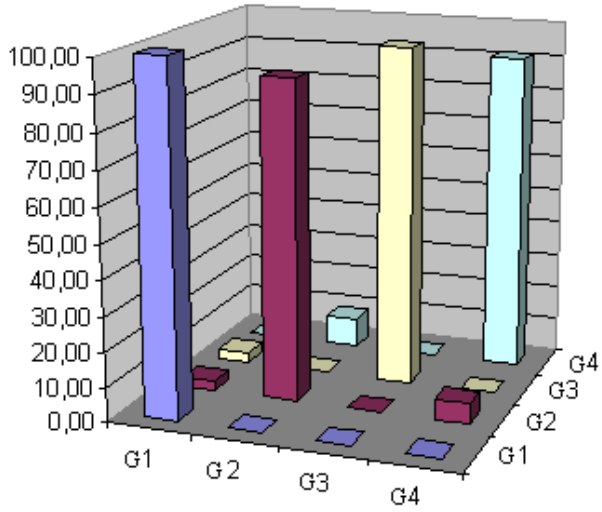


Figura 13. Gráfica de confusión usando redes neuronales probabilísticas.

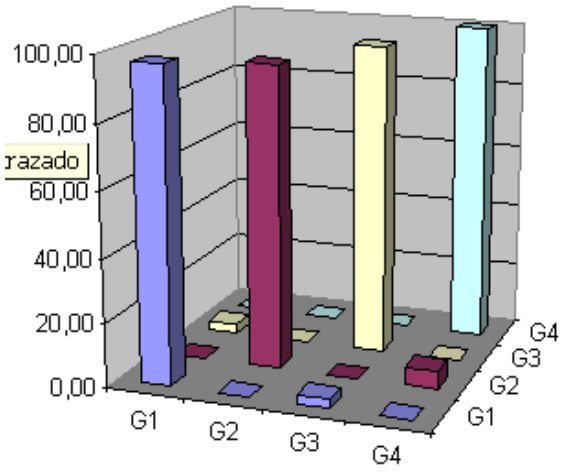


Figura 14. Gráfica de confusión usando redes perceptron multicapa.

La Tabla 2 presenta la matriz de confusión promedio de los resultados, obtenidos por las diferentes técnicas de clasificación estándar. El mayor valor fuera de la diagonal es de 10.48%, y corresponde al porcentaje de veces en que el gesto 4 fue confundido con el gesto 2. Por otro lado, el porcentaje más alto de reconocimiento exitoso ocurre con el gesto 1, con un 97,62%

Tabla 2. Matriz de confusión promedio

	Gesto 1	Gesto 2	Gesto 3	Gesto 4
Gesto 1	96,19	1,43	2,38	0,00
Gesto 2	2,38	93,33	0,00	4,29
Gesto 3	2,38	0,00	97,62	0,00
Gesto 4	0,00	10,48	0,00	89,52

5. CONCLUSIONES

Se presentó un modelo bio-inspirado de la codificación de movimiento, que se utilizó en el reconocimiento visual de cuatro gestos.

Los porcentajes de reconocimiento obtenidos oscilan entre 91.42% usando la estrategia de reconocimiento de distancia al centroide, hasta 97.14% usando redes perceptron multicapa. Estos resultados indican que las características usadas en el reconocimiento son bastante representativas de los gestos, permitiendo así una buena discriminación entre ellos. Por otro lado, dado que los resultados en el reconocimiento son bastante buenos (superiores a 90%) y difieren, en el peor de los casos, en un 6% aproximadamente, la selección final de la técnica de reconocimiento recae en la complejidad computacional y en la facilidad de reentrenamiento ante nuevos datos, nuevos gestos o variaciones de la dimensionalidad que brinde cada técnica.

La Imagen de la Historia del Movimiento, propuesta en este trabajo, genera mayores porcentajes de reconocimiento de gestos que las imágenes propuestas en el trabajo de (Bobick y Davis, 2001), al aplicarlas sobre las respuestas neuronales $G(s_k,\theta_p,t)$ y $G(s_k,\alpha_p,t)$ (Nope et al., 2006b).

El trabajo futuro incluye investigar en la composición de gestos, basándose para ello en gestos básicos aprendidos, y lograr la imitación de los gestos aprendidos por el sistema. Finalmente, este sistema de reconocimiento de gestos se acoplará a un brazo robótico, en una aplicación real.

AGRADECIMIENTOS

Agradecemos al Programa de Apoyo a Doctorados de Conciencias (Colombia), a la Universidad del Valle, y al Instituto Técnico Superior (IST) – Portugal, por el soporte a este trabajo.

REFERENCIAS

Bobick A.F. y J.W. Davis (2001). The Recognition of Human Movement using Temporal Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **Vol. 23**, 257-267.

Bruce V. y P.R. Green (1990). *Visual Perception: Physiology, Psychology and Ecology*. Nottingham: Lawrence Erlbaum Associates.

DeVanois R.L., M.C. Morgan y D.M. Snodderly (1974). Psychophysical studies of monkey vision. III. Spatial luminance contrast sensitivity test of macaque and human observers. *Vision Research* **Vol. 14**, 1974.

Duffy C. y R. Wurtz (1997). MTS neurons respond to speed patterns in optical flor. *Journal of Neuroscience* **Vol. 17(8)**, pp. 2839-2851.

Graciano M., R. Andersen y R. Snowden (1994). Tuning of MTS neurons to spiral motions. *Journal of Neuroscience* **Vol. 14(1)**, pp. 54-67.

Nope S., H. Loaiza y E. Caicedo (2006a). Review of Techniques for Motion Estimation in Artificial Vision. *Revista Colombiana de Tecnologías de Avanzada* **Vol. 2**, pp. 102-108.

Nope S., H. Loaiza y E. Caicedo (2006b). Aplicaciones del Movimiento y su Representación Biológica en el Reconocimiento de Gestos. *Ingeniería y competitividad* **Vol. 8(2)**, pp. 55-63

Orban G.A., H. Kennedy y J. Bullier (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *Journal of Neurophysiology* **Vol. 56(2)**, pp. 462-480.

Pomplun M., J. Martinez-Trujillo, E. Simine, Y. Liu, S. Treue y J.K. Tsotsos (2002). A Neurally-Inspired Model for Detecting and Localizing Simple Motion Patterns in Image Sequences. *Workshop on Dynamic Perception*. Bochum, Alemania.

Santos-Victor S. y G. Sandini (1996). Uncelebrated obstacle detection using normal flor. *Matching Vision and Applications* **Vol. 9**, pp. 130-137.

Treue S. y R.A. Andersen (1996). Neural responses to velocity gradients in macaque cortical area MT. *Visual Neuroscience* **Vol. 13**, pp. 797-804.

Wang C. y M. Brandstein (1999). Multi-source face tracking with audio and visual data. *IEEE MMSP*, pp. 169-174.

.

Este trabajo hace parte de la tesis doctoral de Sandra Esperanza Nope Rodríguez, sobre una arquitectura de control basada en el aprendizaje por imitación de gestos, aplicada en robótica, becaria Colciencias en el programa de apoyo a doctorados nacionales.

Sandra E. Nope R. Especialista en Gerencia de Proyectos de la Universidad del Cauca. Decana de la Facultad de Ingenierías de la Corporación Universitaria Autónoma del Cauca. Candidata a doctorado en Ingeniería, becaria Colciencias, sandrano@univalle.edu.co

Eduardo Caicedo Bravo. Doctor en Informática Industrial de la Universidad Politécnica de Madrid. Profesor de la Universidad del Valle, ecaicedo@univalle.edu.co

Humberto Loaiza Correa. Doctor en Robótica de la Université d'Evry, Francia. Profesor de la Universidad del Valle, hloaiza@univalle.edu.co