

Modelo bio-inspirado para el reconocimiento de gestos usando primitivas de movimiento en visión artificial

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 el futuro ser utilizado en la programación de robots, usando el paradigma del aprendizaje por imitación. En este trabajo se utiliza una cámara web para capturar las imágenes utilizadas como base para la extracción de las primitivas del movimiento, que contienen información usada para la percepción del mundo tanto por personas como por animales. Para la programación de robots por imitación se ha identificado como primera fase el reconocimiento de gestos, en el cual es necesario resolver tres aspectos principales: la representación instantánea del movimiento, la integración en el tiempo de dicha información y la estrategia de clasificación. Estos 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 en el reconocimiento de gestos fue de 88.57% usando una estrategia de clasificación con baja carga computacional.

Palabras clave: codificación de movimiento, integración temporal, primitivas de movimiento, reconocimiento de gestos.

I. Introducción

Los sistemas de visión biológica son capaces de extraer muchos tipos de información diferente del ambiente. Algunos pueden detectar color, 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. Sin embargo, hay un tipo de información que se cree es usada por todos los sistemas biológicos de visión: el movimiento.

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 empleada. 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 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 las primitivas del movimiento para su 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, 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 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 se realiza en el cerebro para el procesamiento de movimiento. Sin embargo, por razones obvias ha sido mucho 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 [1], donde se han encontrado áreas equivalentes al cerebro humano y algunas codificadas bajo el mismo nombre.

Las investigaciones neurofisiológicas, 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 a patrones de actividad neuronal; pero esta transformación es solo el principio de un gran número de transformaciones que se realizan en el sistema nervioso central.

El camino de procesamiento de movimiento en el cerebro del macaco está compuesto por cuatro áreas, éstas son: área estriada (*V1*), área temporal media (*MT*), el área superior media (*MTS*) y *7a* [2]. En [3], se presenta un modelo inspirado neurológicamente del procesamiento jerárquico de movimiento primario y en donde describen una implementación para simulación en

computador que sirvió de inspiración para este trabajo.

Las neuronas en *V1* se activan ante una dirección de movimiento particular y en al menos 3 rangos de velocidad diferentes [4]. Esta información puede extraerse de los vectores de flujo óptico, una técnica de visión artificial para estimar el movimiento en 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.

Una alta porción de las neuronas en *MT* se activan en forma similar a *V1*, mientras que otra porción de las neuronas son selectivas a un ángulo particular entre la dirección del movimiento y el gradiente de la velocidad espacial [5].

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 [6,7].

En este trabajo, se presenta un modelo bio-inspirado para la representación y percepción del movimiento dentro de un sistema de visión artificial, que es aplicado 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 II se expone una breve descripción del sistema, en la sección III 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 trabajos futuros.

2. Descripción del sistema

La figura 1 presenta el diagrama de bloques del sistema de reconocimiento de gestos utilizado, 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 gestos.

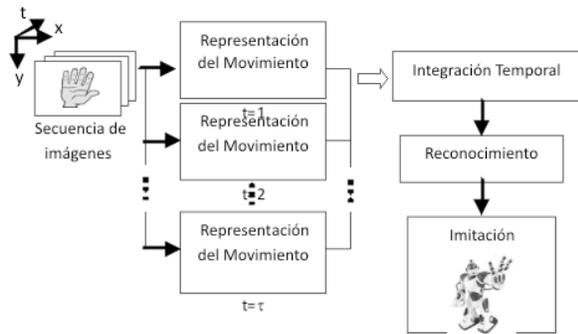


Figura 1. Diagrama de bloques del sistema de reconocimiento de gestos.

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.

como sólo 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 constituye. Así, se utiliza la técnica de detección de color piel por píxel propuesta por [8]. Esta técnica es rápida y simple, pues 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.

En aras de eliminar puntos ruidosos y objetos indeseados del fondo de la imagen que no correspondan a la mano, sólo se escogen aquellos puntos conexos, correspondiéndole a la mano la mayor región dentro de la primera imagen. En imágenes consecutivas, la mano corresponde a la mayor región de color piel dentro de una ventana de búsqueda.

2.1. Representación del Movimiento

La figura 2 muestra los bloques constitutivos del bloque de “Representación del Movimiento” inspirado en el trabajo de [3]. Así, 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.

A. Flujo óptico afín

Hay una gran cantidad de técnicas de estimación de flujo óptico en visión artificial [9], la técnica que mejor se ajusta al reconocimiento de gestos realizado por manos es el Flujo Óptico Afín, ya que produce muy buenas estimaciones con objetos poco texturados. Además, es robusta al ruido y a variaciones en la iluminación. El flujo óptico afín, como lo indica su nombre, combina la ecuación de restricción del flujo óptico y las ecuaciones correspondientes al modelo de formación de imágenes por proyección, perspectiva o modelo afín. La implementación de este algoritmo se basó en el trabajo de [10], en donde el flujo óptico afín es determinado por los valores de θ en la ecuación (1).

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

donde,

$$\theta = [v_o \quad v_x \quad v_y \quad u_o \quad u_x \quad u_y] \quad (2)$$

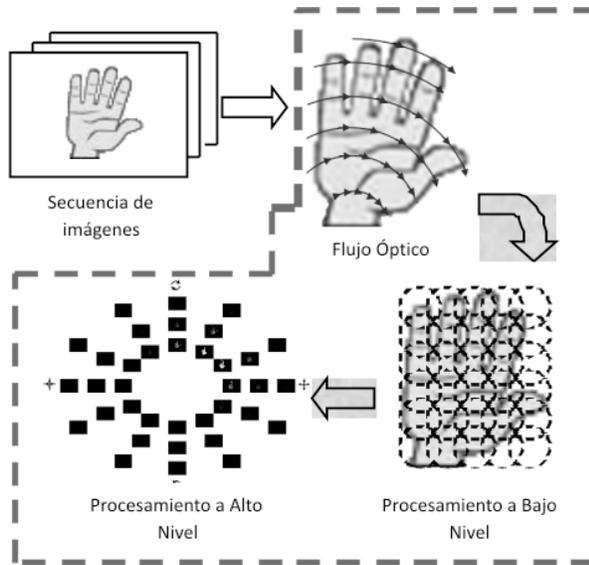


Figura 2. Detalle del bloque de representación del movimiento

De acuerdo con estas ecuaciones, para resolver el problema bastaría con usar seis medidas de las derivadas espacio-temporales de primer orden para determinar θ , aunque usualmente se cuenta con mucho más de seis puntos para su estimación. El procedimiento para la estimación de θ se resume en los siguientes pasos:

- Se escoge al azar un conjunto de puntos $[I_x, I_y, I_t]$ para obtener una primera estimación de θ .
- Se calcula el error de acuerdo con la ecuación (1) para el θ estimado.
- Se repite el procedimiento de a-b hasta que se alcance un error deseado o dicho procedimiento se haya repetido un determinado número de veces.

El número de puntos elegidos por el algoritmo debe ser lo suficientemente grande como para garantizar que la mayoría de los puntos se van a ajustar 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*).

B. 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. Para la codificación del movimiento se usaron las ideas principales utilizadas en [3] 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.

Procesamiento en bajo nivel

Para ahorrar tiempo de cómputo sin pérdida de información relevante se utilizó otra idea de la biología: los Campos Receptores (*Receptive Fields - RF*). En el cerebro de los macacos, los campos receptivos 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 receptivos $I_m(x, y)$, la matriz de salida de los campos receptivos $I_{out}(i, j)$, está definida por la ecuación (3):

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

donde,

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

La figura 3 muestra un ejemplo de aplicación de dichos RFs ante un movimiento de rotación en el sentido de las manecillas del reloj y usando como entrada la magnitud y ángulo del flujo óptico respectivamente. En la parte superior izquierda de la figura aparece la imagen correspondiente a la magnitud del flujo óptico y , sobre ella, unos círculos blancos que corresponden a algunos campos receptores. El resultado de aplicar los campos receptores a dicha magnitud aparece en la parte inferior izquierda. De la misma manera, al lado derecho se presentan los resultados del ejemplo para el caso del ángulo del flujo óptico.

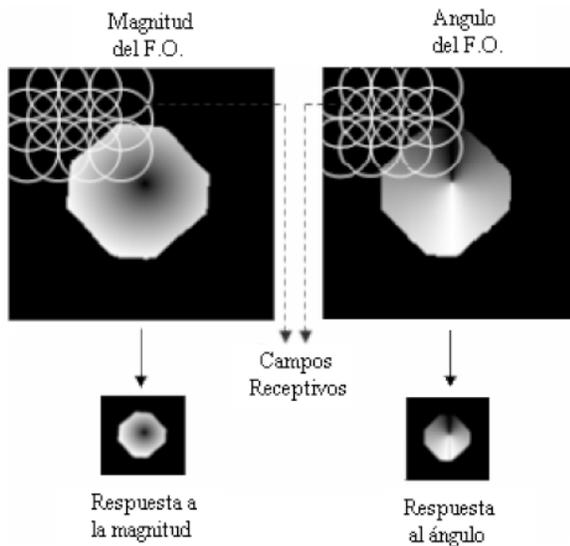


Figura 3. Ejemplo de aplicación de RF a la magnitud y ángulo del flujo óptico ante un movimiento de rotación en el sentido de las manecillas del reloj.

Codificación del movimiento

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 corresponden a las diferentes combinaciones (multiplicaciones) de los diferentes filtros Gausianos sintonizados a velocidad con los diferentes filtros Gausianos sintonizados a dirección.

La respuesta de un filtro bidimensional $G(s_k, \theta_p)$ sintonizado a una velocidad s_k y a una dirección de movimiento θ_p está determinada por la ecuación (5). En donde I_s y 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.

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

La figura 4 muestra los vectores de flujo óptico para el caso de un octágono que rota en el sentido de las manecillas del reloj (derecha), y el conjunto de respuestas neuronales obtenidas en él. Los filtros de velocidad han sido sintonizados a 0.7, 1.4 y 2.1 píxeles por trama, mientras que los sintonizados a la dirección se sintonizan a $\{0, 30, 60 \dots 330\}$ grados. Se observa 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, bioinspirada 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.

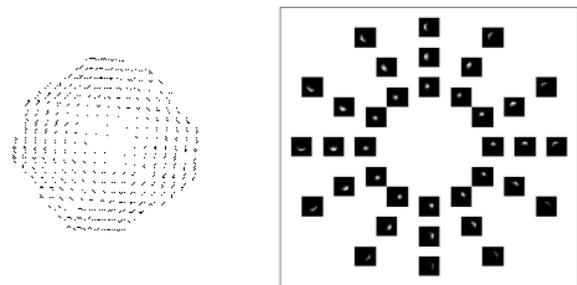


Figura 4. Vectores de flujo óptico para el caso de un octágono que gira en sentido de las manecillas del reloj - izquierda. Conjunto completo de respuestas neuronales - derecha.

La figura 5 presenta los vectores de flujo óptico (flechas azules) y los vectores correspondientes al gradiente de la magnitud del flujo óptico (fucsia) para el caso de: rotación en el sentido de las manecillas del reloj (figura superior izquierda), rotación en el sentido de las manecillas del reloj

(figura superior derecha), movimiento de expansión /figura inferior izquierda) y movimiento de contracción (figura inferior derecha). Se puede apreciar que para el caso de un movimiento de rotación en el sentido de las manecillas del reloj, el ángulo α entre estos vectores corresponde a 90, tomando como referencia el vector de flujo óptico. Así mismo, se puede verificar que para una rotación en el sentido inverso este ángulo es de 270; y para un movimiento de expansión del objeto observado es de 0, mientras que para uno de compresión es de 180. Para una combinación de estos movimientos básicos los valores angulares se encuentran entre estos límites marcados.

En la figura 6 se 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 y va incrementando 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. En este ejemplo, la respuesta más fuerte se presenta a una velocidad de 0.7 píxeles/trama y a un ángulo de 60, esto corresponde a un movimiento, en su mayoría, de rotación en el sentido de las manecillas del reloj (↻) y de expansión en menor grado (+).

En el proceso descrito hasta aquí, sólo se tiene una representación instantánea del movimiento. Sin embargo, para lograr el reconocimiento de gestos es necesario integrar dicha información en el tiempo.

2.2. Integración en el Tiempo

Para analizar la evolución de un gesto se usaron ventanas temporales que incluyeron toda la ejecución del gesto para lograr mayor robustez en su reconocimiento. Los autores 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).

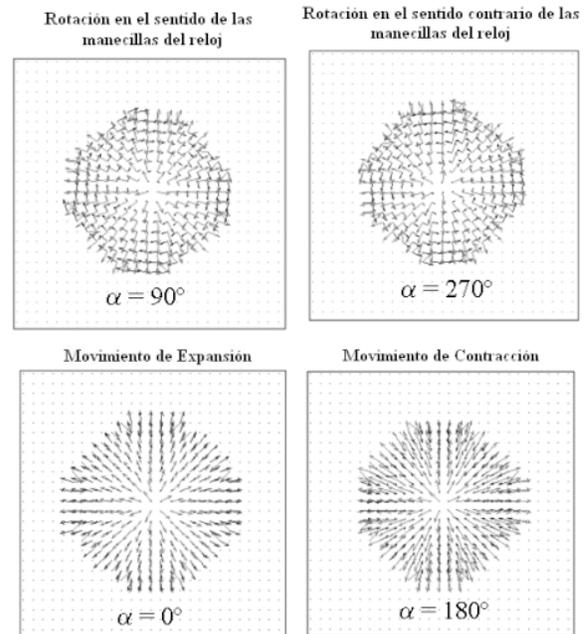


Figura 5. Ejemplo para un octágono que realiza cuatro tipos de movimientos complejos: rotación en sentido de las manecillas del reloj y en sentido inverso – arriba movimientos de expansión y contracción - abajo.

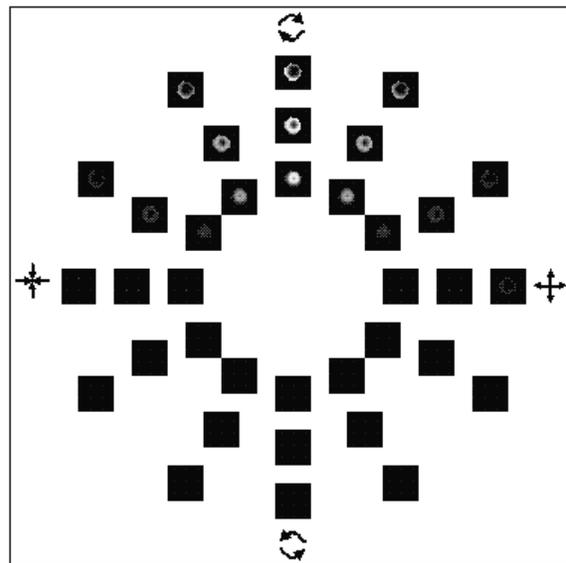


Figura 6. Conjunto completo de respuestas neuronales ante la rotación de un octágono que rota en el sentido de las manecillas del reloj.

Para construir las plantillas, inicialmente se necesita estimar una imagen binaria $D(s_p, \theta_p, t)$. Esta imagen binaria indica las regiones de movimiento de las respuestas neuronales en cuanto a velocidad y dirección del movimiento $D(s_p, \theta_p, t)$ de acuerdo

con la ecuación (6), o para las respuestas neuronales en cuanto a velocidad y clase de movimiento complejo $D_\alpha(s_k, \alpha_p, t)$ de acuerdo con la ecuación (7).

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

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

A. Imagen de la Energía del Movimiento

Las imágenes de energía del movimiento $E_\theta(s_k, \theta_p, t)$ y $E_\alpha(s_k, \alpha_p, t)$ para $G(s_k, \theta_p, t)$ y $G(s_k, \alpha_p, t)$, están definidas por las ecuaciones 8 y 9 respectivamente.

$$E_\theta(s_k, \theta_p, t) = \bigcup_{\tau=0}^t D_\theta(s_k, \theta_p, \tau) \quad (8)$$

$$E_\alpha(s_k, \alpha_p, t) = \bigcup_{\tau=0}^t D_\alpha(s_k, \alpha_p, \tau) \quad (9)$$

La figura 7 presenta un ejemplo de MEI para el caso del gesto 1 (rotar la mano a la derecha y luego a la izquierda), de donde se observa que las MEIs son brillantes (más claras) en los puntos en los que las neuronas se han activado en cualquier instante de tiempo.

B. Imagen de la Historia del Movimiento Modificada

Sea $H_\theta(s_k, \theta_p, t)$ la imagen de la historia del movimiento modificada para $G(s_k, \theta_p, t)$ y $H_\alpha(s_k, \alpha_p, t)$ para $G(s_k, \alpha_p, t)$, definidos por las ecuaciones 10 y 11 respectivamente. En esta representación los puntos más brillantes corresponden a aquellos puntos que se activaron reiterativamente en cualquiera de los instantes de tiempo.

$$H_\theta(s_k, \theta_p, t) = H_\theta(s_k, \theta_p, t-1) + 1 \quad \text{si } D_\theta(s_k, \theta_p, t) = 1 \quad (10)$$

$$H_\alpha(s_k, \alpha_p, t) = H_\alpha(s_k, \alpha_p, t-1) + 1 \quad \text{si } D_\alpha(s_k, \alpha_p, t) = 1 \quad (11)$$

través de las ecuaciones 10 y 11 para el caso del gesto 1 (rotar la mano a la derecha y luego a la izquierda).

Estas nuevas 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, trabajo que será abordado posteriormente.

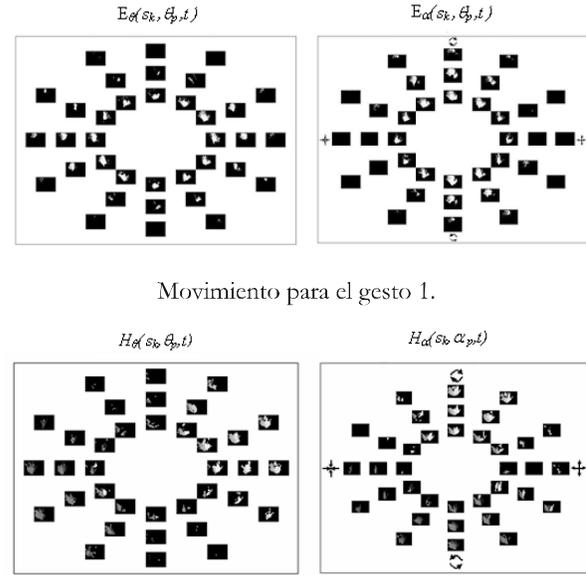


Figura 8. Imágenes de la Historia del Movimiento usando las ecuaciones 10 y 11 para el Gesto 1.

I. 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 a la cámara.

Los resultados del reconocimiento se obtuvieron utilizando la estrategia de clasificación de distancia al centroide y se muestran tabulados en la figura 9. Se utilizaron 70 secuencias de vídeo para cada uno de los cuatro gestos, de las cuales, 35 fueron usadas para el entrenamiento y las 35 restantes para la validación. Un clasificador sin errores generaría una gráfica en donde los valores de los porcentajes en la diagonal sean 100% y el resto cero. El valor

más alto por fuera de la diagonal corresponde a un porcentaje de 22.86% e indica el número de casos dentro de la clase en el que el gesto 4 fue confundido con el gesto 2. De la gráfica se observa que el sistema tiende a confundir el gesto 2 con el 4 y viceversa; esto se debe a que el ángulo entre los vectores de flujo óptico y el gradiente de la magnitud es similar y se diferencian más por la dirección del movimiento, la cual no siempre es suficiente para discriminar entre dichas clases. Con el clasificador por distancia al centroide se obtuvo un porcentaje de aciertos del 88.57%.

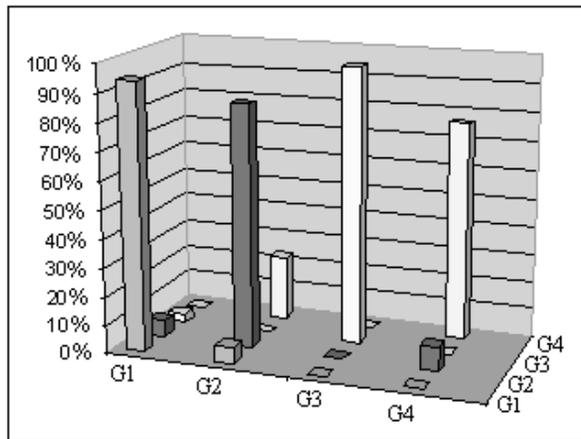


Figura 9. Matriz de confusión (G_i = Gestos)

II. Conclusiones y trabajo futuro

Se presentó un modelo bio-inspirado de la codificación de movimiento para ser utilizado en el reconocimiento visual de cuatro gestos. Los resultados de las pruebas indican una mejora significativa en el reconocimiento de gestos respecto al trabajo de [12], quienes reportaron un porcentaje de correcta imitación del 81% en 20 pruebas. Sin embargo, los autores sólo consideraron 3 de los cuatro gestos usados en este trabajo.

Por otra parte, dada las aproximaciones que hace implícitamente la estrategia de clasificación usada (distancia al centroide), estos resultados no sólo indican que las características usadas en el reconocimiento son bastante representativas de los gestos, permitiendo una buena discriminación entre ellos, sino que pueden ser mejorados a través de técnicas más sofisticadas de clasificación.

El trabajo futuro incluye probar con diferentes clasificadores, investigar en la composición de gestos basándose 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 Colciencias, a la Universidad del Valle y al Instituto Técnico Superior (ISTI) – Portugal, por el soporte a este trabajo.

Bibliografía

- [1] R.L. DeVanois, M.C. Morgan, and D.M. Snodderly, "Psychophysical studies of monkey vision. III. Spatial luminance contrast sensitivity test of macaque and human observers," *Vision Research*, vol. 14, pp. 53-67, 1974.
- [2] V. Bruce and P.R. Green, *Visual Perception Physiology, Psychology and Ecology*. Nottingham: Lawrence Erlbaum Associates, 1990.
- [3] M. Pomplun, J. Martínez-Trujillo, E. Simine, Y. Liu, S. Treue, and J. K. Tsotsos, "A Neurally-Inspired Model for Detecting and Localizing Simple Motion Patterns in Image Sequences," presented at Workshop on Dynamic Perception, Bochum, Germany, 2002.
- [4] G.A. Orban, H. Kennedy, and J. Bullier, "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, 1986.
- [5] S. Treue and R.A. Andersen, "Neural responses to velocity gradients in macaque cortical area MT," in *Visual Neuroscience*, vol. 13, 1996, pp. 797-804.
- [6] M. Graziano, R. Andersen, and R. Snowden, "Tuning of MTS neurons to spiral motions," *Journal of Neuroscience*, vol. 14(1), pp. 54-67, 1994.
- [7] C. Duffy and R. Wurtz, "MTS neurons respond to speed patterns in optical flow," *Journal of Neuroscience*, vol. 17(8), pp. 2839-2851, 1997.
- [8] C. Wang and M. Brandstein, "Multi-source face tracking with audio and visual data," *IEEE MMSP*, pp. 169-174, 1999.

[9] S. Nope, H. Loaiza, and E. Caicedo, "Review of Techniques for Motion Estimation in Artificial Vision," in *Revista Colombiana de Tecnologías de Avanzada*, vol. 2, 2006, pp. 102-108.

[10] J. Santos-Victor and G. Sandini, "Uncelebrated obstacle detection using normal flow," *Matching Vision and Applications*, vol. 9, pp. 130-137, 1996.

[11] A. F. Bobick and J. W. Davis, "The Recognition of Human Movement using Temporal Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 257-267, 2001.

[12] Y. Kuniyoshi, Y. Yorozu, M. Inaba, and H. Inoue, "From Visuo-Motor Self Learning to Early Imitation - A Neural Architecture for Human Learning," presented at International Conference on Robotics and Automation, Taipei, Taiwan, 2003.

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.