

Aplicaciones del movimiento y su representación biológica en el reconocimiento de gestos

Sandra E. Nope*[§], Eduardo Caicedo*, Humberto Loaiza*

*Grupo PSI, Escuela de Ingeniería Eléctrica y Electrónica, Universidad del Valle, Cali, Colombia
§ e-mail: sandrano@univalle.edu.co

(Recibido: Octubre 4 de 2005 - Aceptado: Octubre 27 de 2006)

Resumen

En este artículo se aborda el problema de reconocimiento de gestos usando la información de movimiento. El movimiento contiene información importante que es usada para la percepción del mundo tanto por personas como por animales. En este trabajo se utiliza una cámara *web* para capturar las imágenes utilizadas como base para la extracción de la información de movimiento. El modelo presentado será utilizado en el futuro en un sistema de aprendizaje por demostración aplicado en robótica. Para abordar dicho problema, se ha identificado como primera fase al reconocimiento de gestos y para ello, la arquitectura requiere que se resuelvan tres aspectos principales: la representación instantánea del movimiento, la integración en el tiempo de dicha información y la estrategia de clasificación. En este trabajo se aborda sólo el primero de ellos. En contraste con otros trabajos, en éste la extracción del movimiento y su codificación está inspirada en el procesamiento del movimiento realizado en el cerebro de macacos. El modelo mostrado ha sido aplicado al reconocimiento de cuatro gestos realizados por diferentes personas, cuyo porcentaje de reconocimiento de gestos fue de 89.6% y con una robustez ante diferentes puntos de vista de 96.9%.

Palabras clave: Codificación de movimiento, Flujo óptico, Reconocimiento de gestos.

Applications of motion and its biological representation in recognition of gestures

Abstract

In this article, the problem of gesture recognition is approached using motion information. Motion contains important information that is used by both people and animals for perception of their surroundings. In this work, a web camera is used to capture the images from which motion information is extracted. The model obtained will be used in the future in a demonstration learning system applied to robotics. To approach this problem, gesture recognition has been identified as the first step, and for this purpose, the architecture requires that three main aspects be solved: the instantaneous representation of motion, the integration of this information into time, and the strategy of classification. In this work, only the first of them is considered. In contrast to previous works, in this paper, the extraction of motion and its codification are inspired in the motion processing made by the brain of macaques. The model shown has been applied to the recognition of four gestures made by different people for which the percentage of gesture recognition was 89.6% with a robustness of 96.9% from different points of view.

Keywords: Motion coding, Optical flow, Gesture recognition.

1. Introducción

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

Hay un gran número de aplicaciones en donde la información de movimiento es usada, estas son:

- Codificación de vídeo para mejorar la resolución de imágenes multi-trama (Nestares & Heeger, 2000) y compresión de vídeo (Hang et al., 1997; Dufaux & Konrad, 2000; Keller & Averbuch, 2003).
- Inferir las propiedades tridimensionales de objetos rígidos (Nicolescu & Medioni, 2003)
- En segmentación (Nicolescu & Medioni, 2003) o como etapa de pre-procesamiento.
- En robótica es fundamental en tareas tales como: navegación (Zucchelli et al., 2002), detección de obstáculos (Santos-Victor & Sandini, 1996), coordinación de miembros y en general, para interactuar con ambientes complejos y cambiantes.
- En sistemas de seguridad para eliminar falsas alarmas generadas por animales o por el viento o para el reconocimiento de actividades (Davis & Taylor, 2002).
- Como forma de comunicación ya sea mediante lenguaje gestual o lectura de labios, aspectos importantes en la comunicación e interacción de robots con humanos (Cohn et al., 1998).

El trabajo que se presenta a continuación utiliza el movimiento como forma de comunicación con el objetivo de programar 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, el cerebro de animales ha sido mucho más estudiado. En especial, el cerebro de macacos (una especie de monos), debido a la similitud de sus capacidades visuales con las humanas (DeValois et al., 1974). Es de resaltar que se han encontrado áreas equivalentes en el cerebro humano, 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 sólo 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, que son: área estriada (V1), área temporal media (MT), área superior media (MTS) y 7a (Bruce & Green, 1990). Pomplun et al. (2002) presentaron un modelo, biológicamente inspirado en el cerebro del macaco, acerca del procesamiento jerárquico del movimiento primario, y describen una implementación para simulación en computador.

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). 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 (Treue & 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 (Graziano et al., 1994; Duffy & Wurtz, 1997).

En este trabajo, se presenta un modelo para la representación y percepción del movimiento dentro de un sistema de visión artificial, y que es aplicado en el reconocimiento de cuatro gestos. Este trabajo es considerado como la fase inicial en la programación de un brazo robótico mediante aprendizaje por demostración. En la sección 2, se hace una breve descripción de las componentes del sistema ilustrado en un diagrama en bloques. En la sección 2.1, se introduce el concepto de flujo óptico y se explica el procedimiento de estimación del flujo óptico afín. En la sección 2.2, se aborda el procesamiento en bajo nivel y la codificación del movimiento inspirada biológicamente usando la información contenida en los vectores de flujo óptico. En la sección 3, se presentan los resultados obtenidos en el reconocimiento de cuatro gestos, incluyendo pruebas con cambio en el punto de vista. Las conclusiones y trabajo futuro se presentan en la sección 4.

2. Descripción del sistema

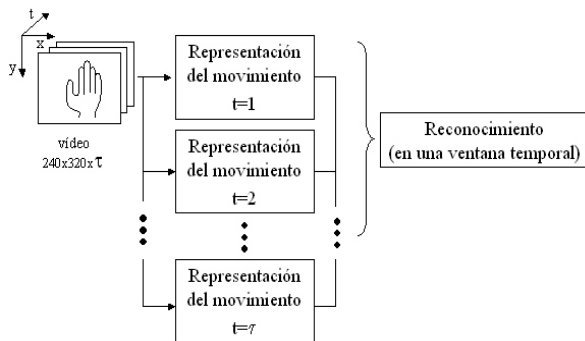


Figura 1. Diagrama en bloques del sistema.

El sistema para el reconocimiento de gestos utiliza una secuencia de vídeo como entrada y está compuesto a grandes rasgos por un bloque que realiza la representación del movimiento y el bloque que realiza el reconocimiento de gestos en sí (ver Figura 1), en donde la variable τ corresponde al número de imágenes del vídeo en el que se realiza un gesto. En este artículo, se hará énfasis en el bloque correspondiente a la representación del movimiento (ver Figura 2).

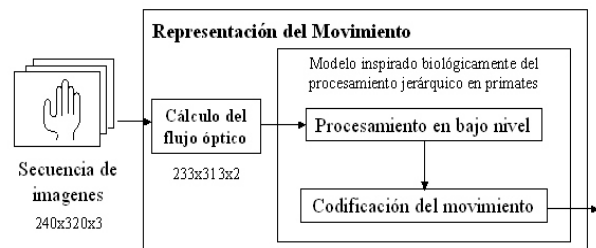


Figura 2. Detalle del bloque de representación del movimiento.

El primer paso dentro del bloque de representación del movimiento corresponde al cálculo de flujo óptico en cada instante de tiempo. Más detalles de este procedimiento se estudian en la sección 2.1. Una vez se tiene la información del flujo óptico, se continúa con su codificación con el fin de alcanzar la percepción del movimiento que se está realizando y al mismo tiempo, obtener robustez al punto de vista. Este procesamiento es explicado en la sección 2.2.

2.1 Flujo óptico

Las imágenes obtenidas por las cámaras corresponden a una proyección en 2D del ambiente en 3D. Al resultado de la proyección de los vectores de velocidad tridimensional se le conoce como campo de movimiento (Horn 1986); mientras que el flujo óptico es definido como un campo de velocidades en el plano de la imagen que describe el movimiento de los píxeles.

Aunque no siempre el flujo óptico corresponde al campo de movimiento, y para algunos puntos no es posible asignar un único vector de velocidad, en muchos casos el flujo óptico es una buena aproximación del campo de movimiento, lo que justifica la utilización de su cálculo.

Las diferentes técnicas para el cálculo de flujo óptico (Fennema & Thompson, 1979; Nagel, 1983; Fleet & Jepson, 1990; Liu et al., 1994; Heeger, 1998; Lucas & Kanade, 1981; Uras et al., 1988; Nesi et al., 1995) asocian las intensidades de brillo dentro de una imagen con el movimiento aparente (Horn & Schunck, 1981). Por otra parte, la mayoría de las técnicas de cálculo de flujo óptico asumen que en un intervalo pequeño de tiempo se conserva la intensidad de brillo de un mismo punto de una imagen a otra; ésto da origen a la Ec. (1) y es conocida como hipótesis de la constancia de brillo:

$$I_x u + I_y v + I_t = 0 \quad (1)$$

donde I_x, I_y, I_t son las derivadas de la imagen en sus coordenadas x y y en el tiempo, respectivamente, mientras que u y v son las componentes del flujo óptico en x y y , respectivamente.

2.1.1 Pre-procesamiento

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

Ya que solo hay interés en analizar el movimiento de la mano en las secuencias de vídeo en un instante de tiempo determinado, se utiliza la información de color presente en las imágenes. Así, se utiliza el modelo HSV y son considerados puntos de color piel los píxeles cuya saturación está por debajo de 0.2, y el matiz entre 0 y 0.12. Estos valores fueron determinados heurísticamente. La elección de umbrales fijos presenta diferentes resultados, dependiendo del color de la piel del demostrador. Sin embargo, funcionan suficientemente bien para los propósitos de este trabajo.

Finalmente, y con el fin de eliminar puntos fuera de la mano, sólo se escogen aquellos puntos que estén concatenados y cuya cantidad sea la mayor. Esto origina una de las restricciones del sistema: la mano debe ocupar el área mayor dentro de las imágenes que forman la secuencia de vídeo.

2.1.2 Flujo óptico afín

El flujo óptico afín, como lo indica su nombre, combina la ecuación de restricción del flujo óptico (Ec. 1) y las ecuaciones correspondientes al modelo de formación de imágenes por proyección perspectiva o modelo afín. Esta técnica realiza un suavizado de los vectores resultantes y ha sido la técnica más apropiada para el caso de un objeto con poca textura como es la mano, en condiciones de iluminación no controladas y para los movimientos que realiza.

El modelo afín es un modelo de transformación rígida de una imagen. Así, el movimiento en un plano queda definido por las Ecs. (2a) y (2b):

$$u(x, y) = u_o + u_x x + u_y y + u_{xy} xy + u_{xx} x^2 \quad (2a)$$

$$v(x, y) = v_o + v_x x + v_y y + v_{xy} xy + v_{yy} y^2 \quad (2b)$$

Las Ecs. (2a) y (2b) para el flujo óptico de una superficie plana pueden ser aproximadas por una transformación afín si se asume que los cambios de profundidad son limitados y el ángulo de vista es pequeño; estas suposiciones permiten descartar las derivadas de segundo orden, usualmente ruidosas. Con estas suposiciones en las Ecs. (2a) y (2b) y usando la Ec. (1), se obtiene la Ec. (3):

$$\begin{bmatrix} I_y & x I_y & y I_y & I_x & x I_x & y I_x \end{bmatrix} \theta = -I_t \quad (3)$$

El flujo óptico afín estará determinado entonces por los valores de,

$$\theta = [v_o \quad v_x \quad v_y \quad u_o \quad u_x \quad u_y] \quad (4)$$

De acuerdo con esto, para resolver el problema bastaría con usar seis medidas de las derivadas espacio-temporales de primer orden para determinar θ . Usualmente, se cuenta con mucho más de seis puntos para su estimación. Cuando se usa un mayor número de puntos, se obtiene una mejor robustez en la estimación de los parámetros.

Sin embargo, el uso de muchos puntos hace que se requieran mayores tiempos de cómputo para la estimación de θ ; 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.

2.1.3 Procedimiento para la estimación del flujo óptico afín

A continuación se describen los pasos para el cálculo del flujo óptico afín de forma robusta, aún con presencia de puntos fuera del modelo real (*outliers*) que suelen degradar la estimación.

- a) Se escoge al azar un conjunto de puntos $[I_x, I_y, I_t]^T$ para obtener una primera estimación de θ .
- b) Se calcula el error de acuerdo con la Ec. (3) para el valor de θ estimado.
- c) Se repite el procedimiento de a) a b) hasta que se alcance un error deseado o dicho procedimiento se haya repetido un número de veces dado.

2.2 Codificación del movimiento

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 usadas en (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 la 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 movimientos complejos.

2.2.1 Procesamiento en bajo nivel

Se puede ahorrar tiempo de cómputo sin pérdida de información relevante usando 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%.

La respuesta de los campos receptores fue simulada calculando la media de todos los puntos dentro de círculos fijos de diámetro D píxeles. La matriz de entrada a los RFs corresponde a la magnitud del flujo óptico o a su ángulo. Esto es, dada la matriz de entrada a los campos receptivos $I_{in}(x,y)$, la matriz de salida de los campos receptivos $I_{out}(i,j)$ está definida por la Ec. (5):

$$I_{out}(i, j) = \sum_x \sum_y k(i, j) * I_{in}(x, y) \quad (5)$$

donde

$$k(i, j) = \begin{cases} 1/n, & \text{si } \sqrt{(x-i)^2 + (y-j)^2} \leq D/2 \\ 0, & \text{en otro caso} \end{cases} \quad (6)$$

Las Figuras 3 y 4 muestran 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.

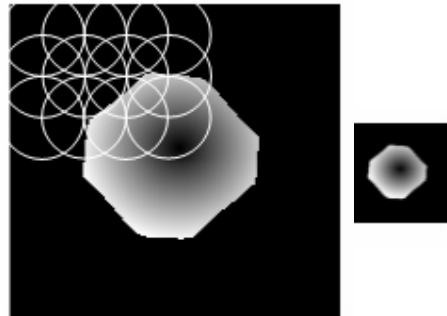


Figura 3. Ejemplo de aplicación de RF a la magnitud del flujo óptico ante un movimiento de rotación en el sentido de las manecillas del reloj. Izquierda: magnitud del flujo óptico como entrada. Derecha: salida de los RF ante la magnitud.

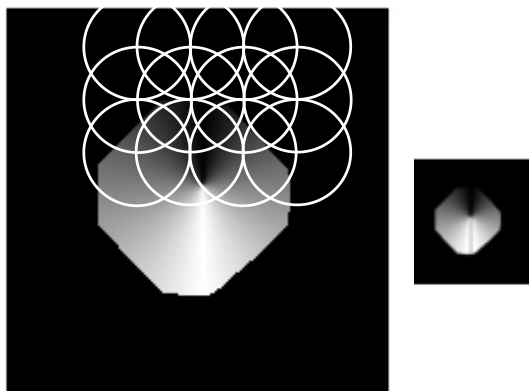


Figura 4. Ejemplo de aplicación de RF al ángulo del flujo óptico ante un movimiento de rotación en el sentido de las manecillas del reloj. Izquierda: ángulo del flujo óptico como entrada. Derecha: salida de los RF ante el ángulo.

2.2.2 Representación del movimiento

La selectividad de las neuronas a una velocidad y dirección particular se simula mediante la multiplicación de la respuesta de dos filtros Gaussianos. 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 filtros Gaussianos afinados a una velocidad con los filtros Gaussianos afinados a una dirección.

La respuesta de un filtro bidimensional $G(s_k, \theta_p)$ ajustada a una velocidad s_k y a una dirección de movimiento θ_p está determinada por la Ec. (7), en la cual 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 son las desviaciones estándares del filtro ajustado a la velocidad para cada uno. Entonces:

$$G(s_k, \theta_p) = e^{-\frac{(I_s - s_k)^2}{\sigma_s}} e^{-\frac{(I_\theta - \theta_p)^2}{\sigma_\theta}} \quad (7)$$

La Figura 5 muestra en detalle los vectores de flujo óptico para un octágono que rota en el sentido de las manecillas del reloj (izquierda) y la respuesta de un filtro bidimensional ajustado a una velocidad de 0.7 píxeles / trama y un ángulo (θ) de

0° (derecha). Observe que la respuesta del filtro es en efecto más brillante para aquellos vectores con valores cercanos a los de sintonización del filtro (vectores dentro del óvalo).

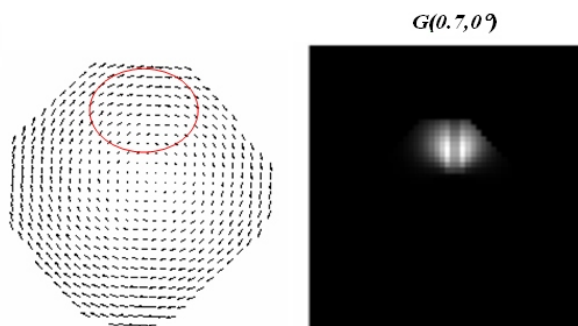


Figura 5. Ejemplo para un octágono que gira en el sentido de las manecillas del reloj. Izquierda: vectores del flujo óptico afin. Derecha: respuesta del filtro bidimensional ajustados a 0.7 píxeles/trama y $\theta = 0$.

En este momento, se tiene una representación del movimiento inspirada biológicamente, pero no es robusta a cambios del punto de vista. Para alcanzar esta robustez en la representación del movimiento, en lugar de usar la dirección de movimiento por sí sola, se utiliza el ángulo entre los vectores de flujo óptico y el gradiente de la magnitud del flujo óptico (α).

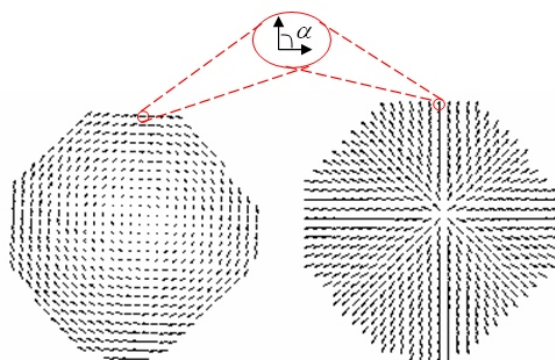


Figura 6. Ejemplo para un octágono que rota en sentido de las manecillas del reloj. Izquierda: vectores de flujo óptico. Derecha: gradiente de la velocidad.

La Figura 6 muestra un ejemplo de los vectores de flujo óptico a la izquierda y de los vectores correspondientes al gradiente de la velocidad a la derecha. En el óvalo que está en la parte superior de la figura, se puede apreciar el ángulo α entre

estos vectores para el caso de un movimiento de rotación en el sentido de las manecillas del reloj, y tomando como referencia los vectores de flujo óptico, ese ángulo corresponde a 90° . Así mismo, se puede verificar que para una rotación en el sentido inverso este ángulo es 270° , para un movimiento de expansión es 0° y 180° para compresión. Valores angulares diferentes corresponden a una mezcla entre estos movimientos básicos.

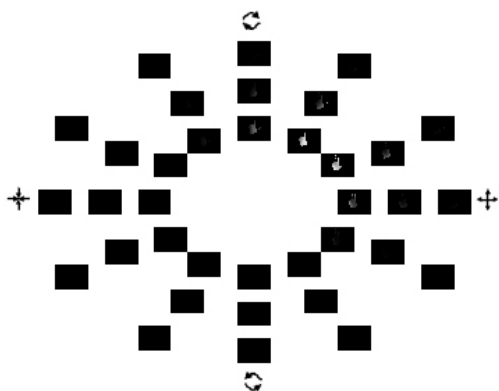


Figura 7. Conjunto completo de respuestas neuronales ante la rotación de una mano en el sentido de las manecillas del reloj.

En la Figura 7 se presenta el conjunto completo de respuestas neuronales en el caso de una mano que realiza un movimiento de rotación en el sentido de las manecillas del reloj. El anillo interior corresponde a la velocidad más baja y se 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. 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 (+).

Hasta el momento se tiene una representación instantánea del movimiento. Sin embargo, para lograr el reconocimiento de gestos, es necesario integrar dicha información en el tiempo, para ello, se calculó la energía del movimiento de la imagen y la historia del movimiento de la imagen. Ya que estas nuevas imágenes contienen mucha información redundante y son de un tamaño muy

grande, esta información es reducida mediante el cálculo de histogramas y el uso de la técnica de análisis de componentes principales (PCA). Con los vectores resultantes, ahora es posible reconocer los gestos.

3. 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.

Para el reconocimiento de gestos, fue necesario incluir otros módulos que no serán descritos, estos corresponden a la integración temporal, disminución de la resolución de la información y a la estrategia de clasificación.

Los resultados del reconocimiento se obtuvieron utilizando la estrategia de clasificación del vecino más próximo y se muestran a través de la matriz de confusión de la Tabla 1.

Tabla 1. Matriz de confusión

	Gesto 1	Gesto 2	Gesto 3	Gesto 4
Gesto 1	12	0	0	0
Gesto 2	0	9	0	3
Gesto 3	0	2	10	0
Gesto 4	0	0	0	12

Se cuenta con 24 secuencias de vídeo para cada uno de los cuatro gestos, de estas, 12 de cada gesto fueron usadas para el entrenamiento y las 12 restantes para prueba. Lo ideal en una matriz de confusión es que los valores de su diagonal correspondan al número de datos de validación por clase (12), y que el resto de las casillas sea cero. El valor de 3 en dicha tabla, indica que el gesto 2 (número de la fila) fue confundido con el gesto 4 (número de la columna). El porcentaje de buena clasificación para este caso corresponde al 89.6%; este valor se obtiene mediante la suma de las componentes de la diagonal en la matriz de confusión dividida por el valor de la suma de todas sus casillas.

Adicionalmente, se realizaron pruebas con el objeto de medir la robustez al cambio de punto de vista con 8 nuevas secuencias para cada gesto, en ellas, el punto de vista fue variado entre -60° y 60° . La Tabla 2 presenta la matriz de confusión para este caso. El porcentaje de buena clasificación con diferente punto de vista fue de 96.9%. Es importante resaltar que durante el entrenamiento no se incluyeron secuencias con diferente punto de vista.

Tabla 2. Matriz de confusión para secuencias con diferentes puntos de vista.

	Gesto 1	Gesto 2	Gesto 3	Gesto 4
Gesto 1	8	0	0	0
Gesto 2	0	8	0	0
Gesto 3	0	0	8	0
Gesto 4	0	1	0	7

Observe que en los dos casos el sistema tiende a confundir el gesto 2 con el 4 o viceversa; ésto es debido 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.

4. Conclusiones

Se ha presentado un algoritmo de codificación de movimiento con inspiración biológica que ha sido aplicado en el reconocimiento visual de cuatro gestos. Los resultados de las pruebas iniciales indican una buena representación de los mismos obteniéndose un buen porcentaje de clasificación (89.6%) usando sólo la información de movimiento. Así mismo, dicha representación es robusta al cambio de punto de vista; el porcentaje de clasificación en este caso fue del 96.9% con cambios de punto de vista entre -60° y 60° .

El trabajo futuro incluye ampliar la base de datos tanto para el entrenamiento como para probar el sistema, ampliar el número de gestos observados, 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.

5. Agradecimientos

Agradecemos al programa de apoyo a doctorados de Colciencias, a la Universidad del Valle y al Instituto Técnico Superior (IST), Portugal, por el soporte a este trabajo. Un reconocimiento especial al profesor José Santos-Victor del IST, Portugal por su orientación, consejo y apoyo en este proyecto.

6. Referencias bibliográficas

Bruce, V., & Green, P. R. (1990). *Visual perception: physiology, psychology & ecology*. Second Edition. Nottingham, England: Lawrence Erlbaum Associates.

Cohn, J.F., Zlochower, A.J., Lien, J.J., & Kanade, T. (1998). *Feature-point tracking by optical flow discriminates subtle differences in facial expression*. Proceedings of the Third International Conference on Face and Gesture Recognition, p. 396-401.

Davis, J.W., & Taylor, S. R. (2002). *Analysis and recognition of walking movements*. Proceedings of the 16th International Conference on Pattern Recognition, Vol. 1, p. 315-318.

DeValois, R.L., Morgan, H.C., & Snodderly, D.M. (1974). Psychophysical studies of monkey vision. Part III. Spatial luminance contrast sensitivity tests of macaque and human observers. *Vision Research* 14, 75-81.

Dufaux, F., & Konrad, J. (2000). Robust, efficient and fast global motion estimation for video coding. *IEEE Transactions on Image Processing* 9 (3), 497-501.

Duffy, C. J., & Wurtz, R.H. (1997). Medial superior temporal area neurons respond to speed patterns in optic flow. *Journal of Neuroscience* 17 (8), 2839-2851.

Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing* 9 (4), 301-315.

- Fleet, D. J., & Jepson, A. D. (1990). Computation of component image velocity from local phase information. *International Journal of Computer Vision* 5(1), 77-104.
- Graziano, M. S. A., Andersen, R. A., & Snowden, R. J. (1994). Tuning of MST neurons to spiral motions. *Journal of Neuroscience* 14(1), 54-67.
- Hang, H.-M., Chou, Y.-M., & Cheng, S.-C. (1997). Motion Estimation for Video Coding Standards. *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology* 17 (2-3), 113-136.
- Heeger, D. J. (1998). Optical flow using spatio-temporal filters. *International Journal of Computer Vision* 1(4), 279-302.
- Horn, B.K.P., & Schunck, B.G. (1981). Determining Optical Flow. *Artificial Intelligence* 17, 185-203.
- Horn, B. K. P. (1986). *Robot Vision*. Cambridge: MIT Press.
- Keller, Y., & Averbuch, A. (2003). Fast gradient methods based on global motion estimation for video compression. *IEEE Transactions on Circuits and Systems for Video Technology* 13 (4), 300-309.
- Liu, H., Hong, T. H., Herman, M., & Chellappa, R. (1994). *A generalized motion model for estimating optical flow using 3-D hermite polynomials*. Proceedings of the 12th International Conference of Pattern Recognition, p. 361-366.
- Lucas, B.D., & Kanade, T. (1981). *An iterative image registration technique with an application to stereo vision*. Proceedings of the 7th International Joint Conference on Artificial Intelligence, p. 674-679.
- Nagel, H.H. (1983). Displacement vectors derived from second order intensity variations in image sequences. *Computer Graphics and Image Processing* 21, 85-117.
- Nesi, P., Del Bimbo, A., & Ben-Tzvi, D. (1995). A Robust algorithm for optical flow estimation. *Computer Vision and Image Understanding* 62(1), 59-68.
- Nestares, O., & Heeger, D. (2000). Robust multiresolution alignment of MRI brain volumes. *Magnetic Resonance in Medicine* 43 (5), 705-715.
- Nicolescu, M., & Medioni, G. (2003). *Motion segmentation with accurate boundaries - a tensor voting approach*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Vol. 1, p. 382-389.
- Orban, G.A., Kennedy, H., & Bullier, J. (1986). Velocity sensitivity and direction sensitivity of neurons in areas V1 and V2 of the monkey: Influence of eccentricity. *Journal of Neurophysiology* 56(2), 462-480.
- Pomplun, M., Trujillo-Martinez, J., Simine, E., Liu, Y., Treue, S., & Tsotsos, J. K. (2002). *A neurally-inspired model for detecting and localizing simple motion patterns in image sequences*. Proceedings of the 4th Workshop on Dynamic Perception, p. 47-52.
- Santos-Victor, J., & Sandini, G. (1996). Uncalibrated obstacle detection using normal flow. *Machine Vision and Applications* 9(3), 130-137.
- Treue, S., & Andersen, R.A. (1996). Neural responses to velocity gradients in macaque cortical area MT. *Visual Neuroscience* 13 (4), 797-804.
- Uras, S., Girosi, F., Verri, A., & Torre, V. (1988). A computational approach to motion perception. *Biological Cybernetics* 60(2), 79-87.
- Zucchelli, M., Santos-Victor, J., & Christensen, H. (2002). *Constrained structure and motion estimation from optical flow*. Proceedings of the 16th International Conference on Pattern Recognition, Vol. 1, p. 339-342.