



La comunicación gestual con la computadora como una herramienta de ayuda para minusválidos

F. García-Ugalde, V. García-Garduño y D. Gatica-Pérez
*Departamento de Ingeniería Eléctrica, División de Estudios de Posgrado
Facultad de Ingeniería UNAM e IDIAP
E-mail: fgarciau@servidor.unam.mx*

(recibido: junio de 2002; aceptado: octubre de 2002)

Resumen

El uso de los sistemas de captura de imágenes para las computadoras es cada vez más común. Por otro lado, el procesamiento y análisis de video digital puede tener aplicaciones interesantes en el área de la comunicación hombre-máquina. Pensando en este desarrollo científico-tecnológico, este trabajo aborda el problema de la comunicación "amigable" entre la computadora y algunos seres humanos minusválidos que requieren de facilidades, por lo que se desarrolló un lenguaje gestual con el que la persona puede indicar a la máquina algunas acciones simples a seguir por medio de sus manos.

Se definió un alfabeto muy limitado de *gestos* que la máquina puede interpretar y se pensó también en no utilizar guantes instrumentados, ya que son costosos y estorbosos. Por lo tanto, el número de *gestos* tiene que ser reducido y deben mostrarse relativamente burdos; dada la falta de precisión que se tendrá durante su captura con una cámara digital común y corriente.

Una vez realizados y capturados los *gestos*, se procede a su segmentación con el propósito de aislarlos del resto de la escena captada por la cámara y poder efectuar su interpretación de una manera más sencilla. Los *gestos* están provistos de movimiento, de tal modo que para aislarlos se puede explotar esta información recurriendo a un algoritmo de segmentación basada en el movimiento.

Logrando el aislamiento se lleva a cabo su interpretación de una manera más sencilla y rápida al obtener una máscara binaria del *gesto*, la cual se proyecta enseguida en dos ejes ortogonales x y y . Los perfiles del *gesto* obtenidos por la proyección en x y y , se utilizan como una especie de apuntadores; de esta manera se puede buscar en una base de datos previamente estructurada y guardada en memoria.

Los resultados obtenidos hasta ahora son satisfactorios, aunque el número de *gestos* definidos es muy reducido y bastante distintos unos de otros, lo que facilita su interpretación. Por lo anterior este trabajo consiste en ampliar el tamaño del alfabeto y aumentar la complejidad de los *gestos* para tener una aplicación más realista.

Descriptor: interfaz gestual, interacción hombre-máquina, interpretación automática de gestos, segmentación de objetos no rígidos, análisis de movimiento de objetos no rígidos, segmentación de objetos en movimiento.

Abstract

Digital image acquisition systems are currently now a common item for computers. In a complementary way, digital video analysis and processing could have interesting applications in the area of man-machine communication. Thinking in this scientific and technological development, in this work we address the problem of friendly communication between the computer and some people with disabilities, which require some kind of facilities. With this goal we thought in the possibility to develop a hand gesture language for providing a person the facility to direct simple actions to a machine using some simple hand gestures. We define a small alphabet of gestures that the machine could understand and additionally we thought in the necessity to avoid the use of instrumented gloves, which are expensive and cumbersome for most people. Moreover the number of gestures has to be reduced and they have to be roughly defined; because the leak of precision during the capture process by a digital camera in a more natural way.

When the gestures have been realized and then captured by the computer, and because the goal is to isolate them, we proceed to perform a segmentation, in this way their interpretation will be facilitated. The hand gestures are provided of movement so we could use this information for developing a motion based segmentation algorithm.

When the segmentation has been done we start the interpretation, and looking for high speed methods we proceed to obtain a binary mask corresponding to the segmented gesture and proceed to project it on two orthogonal axes x and y . The x and y profiles obtained are then used like pointers in order to find the corresponding gesture in a database previously structured and stored.

The results are satisfactory nevertheless the number of hand gestures defined is too small and the gestures are different enough which facilitates their interpretation. Consequently the natural extension of this work is to have a large alphabet and increasing the complexity of the gestures, in order to have a more realistic application.

Keywords: *hand gesture interface, human-computer interaction, automatic gesture interpretation, segmentation of non-rigid objects, non-rigid motion analysis, segmentation of moving objects.*

Introducción

Generalmente, el desarrollo de interfaces “amigables” hombre-máquina se presenta para algunas aplicaciones como un problema de difícil solución; sin embargo, bajo ciertas restricciones asociadas con una determinada aplicación es posible proponer soluciones con buenos resultados. En este trabajo se presenta una solución para la comunicación hombre-máquina en personas con cierto tipo de discapacidad, la idea es proveerlos de un mecanismo simple de comunicación con la computadora que les permita hacer que la máquina realice cierta acción después de haber efectuado un gesto

con la mano; éste es reconocido e identificado por la máquina dentro de un conjunto que constituye una base de datos, previamente almacenada. Por ahora, el conjunto de gestos es bastante limitado y la complejidad de los mismos muy reducida; sin embargo, la idea ha sido probada. Queda como trabajo por resolver: el aumento del número de gestos y la complejidad de los mismos, para alcanzar finalmente un verdadero lenguaje alternativo de comunicación con la computadora para algunas aplicaciones.

La metodología que se ha elegido para llevar a cabo este trabajo, reposa en el procesamiento de imágenes digitales; y

partiendo de que los gestos efectuados por la mano poseen movimiento, se utiliza un algoritmo de segmentación basada en el mismo, el cual permite aislar el gesto realizado por la mano del resto de la escena. Este algoritmo hace la estimación de un campo denso de vectores de movimiento, el cual es fundamental para llevar a cabo la segmentación del objeto en actividad (dentro del contexto de este trabajo, el objeto en movimiento corresponde al gesto de la mano). La estimación del campo de vectores de movimiento y la segmentación son problemas muy interrelacionados; la calidad de los resultados del primero influye en los resultados de la segunda. La segmentación basada en el movimiento es una etapa crucial para la extracción de información llamada de alto nivel (en contraposición con la información a nivel píxel, llamada de bajo nivel), a partir de las variaciones en el tiempo de la intensidad luminosa de la secuencia de imágenes (García-Garduño, 1995), (François, 1991). Esto representa un cambio cualitativo de un enfoque local, a un enfoque regional o más global para la descripción del movimiento, lo cual mejora su proceso de medida. El algoritmo de segmentación tiene una primera etapa que consta de la estimación del campo de vectores de movimiento, esta estimación forma parte de los métodos llamados: recursivos por píxel (Baaziz, 1991).

En la siguiente etapa se modela la secuencia de imágenes utilizando el formalismo de los *campos aleatorios de Markov* (MRF) y se prosigue la optimización del problema de la segmentación, utilizando un criterio de estimación *Bayesiano: máximo a posteriori*

(MAP), el cual se desarrolla por medio de un método determinístico: *modos condicionales iterados* (ICM). El enfoque probabilístico toma en cuenta el hecho de que un campo de desplazamiento exacto no existe; dado que algunos errores pueden ocurrir en, o alrededor de las fronteras de movimiento y que los mejores resultados se pueden obtener si se conoce un indicador de la calidad del campo de vectores; este indicador se obtiene a partir de los píxeles no compensados (salidos del algoritmo de estimación y compensación de movimiento), así como de los contornos. Una vez que el objeto en movimiento (gesto de la mano) ha sido aislado por medio de la segmentación, se procede a su búsqueda en una base de datos previamente concebida y guardada en memoria. Para que esta operación sea más rápida y sencilla se procede a obtener una máscara binaria del gesto y enseguida se proyecta ésta sobre los ejes ortogonales x y y . Estos perfiles asociados al gesto se utilizan posteriormente a manera de apuntadores, para identificarlos en la base de datos, con lo cual la máquina puede ejecutar la acción asociada correspondiente, ya que también ha sido previamente definida.

El algoritmo de segmentación

En presencia de un movimiento divergente puro (*zoom*) y desde el punto de vista del procesamiento de imágenes, las técnicas basadas en un simple agrupamiento espacial para hacer una segmentación basada en el movimiento, no dan muy buenos resultados (García, 1995). En este caso se tiene que introducir un modelo Θ_M , tanto del movimiento, como de la estructura de

las regiones de la escena. Por lo tanto, el objetivo del proceso de segmentación consiste en asignar cada píxel de la imagen, a una de las diferentes regiones y el resultado dependerá de la precisión que se logre entre cada vector de movimiento y el modelo supuesto. Cada región está caracterizada por un vector de parámetros de movimiento. Las regiones obtenidas podrán quedar asociadas a diferentes partes del mismo objeto, o a diversos objetos de la escena. El algoritmo de segmentación que se utilizó en este trabajo está basado en los *campos aleatorios de Markov* y la teoría de la estimación, utilizando como principio de optimización el criterio de *máximo a posteriori*; tal procedimiento combinado proporciona un marco de trabajo en el que podemos introducir información proveniente de fuentes de diversa naturaleza, se pueden modelar sus interacciones e incorporar propiedades esperadas en la solución. El modelado a través de los *campos aleatorios de Markov* es apropiado para la segmentación basada en el movimiento, dado que se cuenta con un campo de vectores de desplazamiento como la principal información para separar una imagen en regiones, pero como se ha comentado anteriormente, la información de movimiento no siempre es correcta, principalmente en las discontinuidades del mismo (cuando hay dos objetos con movimientos diferentes en la escena). En este caso, con este modelado y para mejorar la solución final, podemos incluir otras fuentes de datos, como los valores de intensidades de gris, los píxeles no compensados durante la estimación del movimiento y las intensidades de los contornos de los objetos. Además, se pueden agregar propiedades físicas al

modelo: (a) un modelo de movimiento Θ_M para cada región de la escena a ser segmentada, (b) continuidad espacial para la segmentación, (c) presencia de fronteras de movimiento, sólo cuando ocurran cambios fuertes de intensidad y (d) algunas formas geométricas esperadas para las fronteras de las regiones.

De acuerdo con la teoría de los *campos aleatorios de Markov* MRF, representamos cada fuente de información como un *campo de observación* y cada resultado de la segmentación como un *campo de etiquetas*. En este caso las observaciones consideradas son:

- Las componentes horizontal y vertical del campo de movimiento d_x y d_y
- El campo binario de píxeles no compensados. Éste puede ser considerado como una manera simple de eliminar algunos defectos, es decir, representa una forma de balance entre el desplazamiento y la información más confiable (valores de intensidad), cuando el campo de movimiento no fue estimado adecuadamente.
- El campo de intensidades de valores de gris.
- El campo binario de intensidades de los contornos, esto favorecerá la coincidencia de las fronteras del movimiento con los gradientes espaciales más fuertes: 0 significa que no hay contorno y 1 significa que sí lo hay.

Por otro lado, las etiquetas deseadas son:

- El campo de etiquetas de segmentación deseado, el cual tiene asociado un

modelo simplificado lineal de movimiento con cuatro parámetros. Este modelo puede describir movimientos combinados de translación, rotación y divergencia de superficies planas, paralelas al plano de la imagen (García, 1995)

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} + \begin{bmatrix} k & -\theta \\ \theta & k \end{bmatrix} \begin{bmatrix} x - x_g \\ y - y_g \end{bmatrix} \quad (1)$$

Donde (x_g, y_g) es el centro de gravedad de cada superficie.

Para mejorar el proceso de segmentación hemos introducido un campo binario de líneas de discontinuidad auxiliar \mathbf{l} , a lo largo del campo de etiquetas de segmentación. Este representa las fronteras del movimiento: 0 (significa que no hay discontinuidad de movimiento) y 1 significa que sí hay discontinuidad de movimiento.

En la figura 1 se muestra la interacción entre las observaciones, las etiquetas y las suposiciones físicas.

Considerando un total de N pixels en la imagen, la segmentación de regiones basada en el movimiento ha sido formulada como un problema de estimación en donde se encuentran simultáneamente los campos de etiquetas $(\hat{\mathbf{e}}, \hat{\mathbf{l}})$ que maximizan la función de densidad de probabilidad *a posteriori* (**pdf**) de las etiquetas, dados los datos observados:

$$(\hat{\mathbf{e}}, \hat{\mathbf{l}}) = \arg \max_{\mathbf{e}, \mathbf{l}} p(\mathbf{e}, \mathbf{l} | \mathbf{d}_x, \mathbf{d}_y, \mathbf{i}, \mathbf{p}, \mathbf{g}) \quad (2)$$

Utilizando la regla de Bayes, la última ecuación se puede expresar como:

$$(\hat{\mathbf{e}}, \hat{\mathbf{l}}) = \arg \max_{\mathbf{e}, \mathbf{l}} p(\mathbf{d}_x, \mathbf{d}_y, \mathbf{i} | \mathbf{e}, \mathbf{l}, \mathbf{p}, \mathbf{g}) p(\mathbf{e}, \mathbf{l} | \mathbf{p}, \mathbf{g}) \quad (3)$$

En el trabajo de Gatica *et al.*, 1997, se ha demostrado que la maximización de la **pdf**

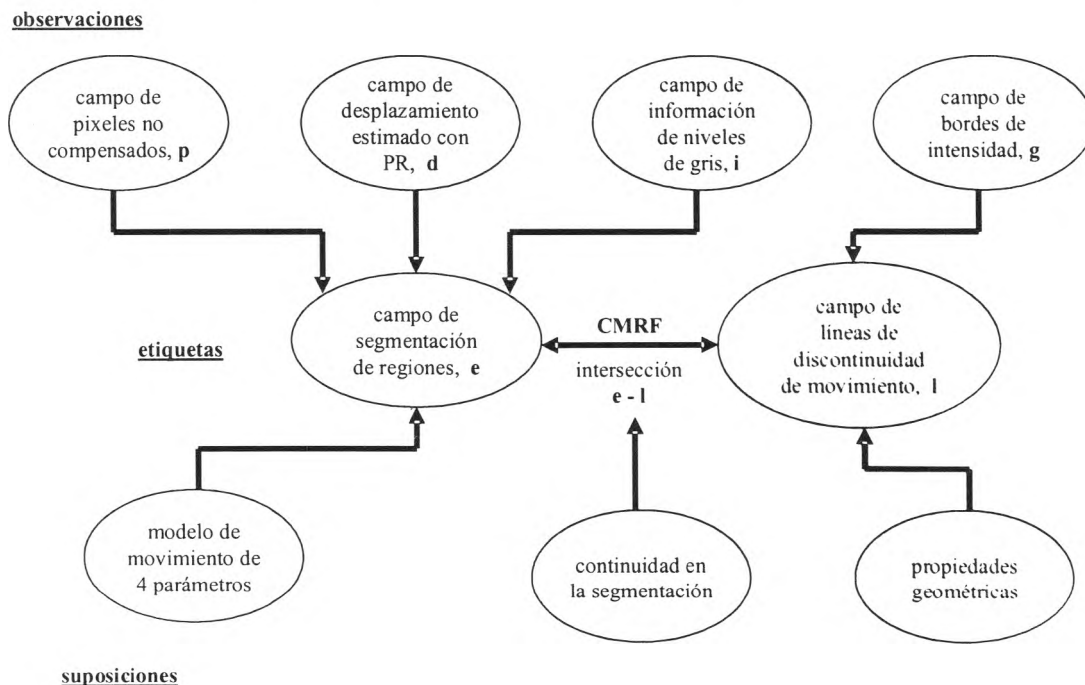


Figura 1. Interacciones del algoritmo de segmentación basada en el movimiento

a posteriori, es equivalente a la minimización de la llamada *función de energía* $U(\mathbf{e}, \mathbf{l}, \mathbf{d}_x, \mathbf{d}_y, \mathbf{i}, \mathbf{p}, \mathbf{g})$ que tiene la forma:

$$U(\mathbf{e}, \mathbf{l}, \mathbf{d}_x, \mathbf{d}_y, \mathbf{i}, \mathbf{p}, \mathbf{g}) = \alpha U_d(\mathbf{d}_x, \mathbf{d}_y, \mathbf{e}, \mathbf{p}) + \beta U_i(\mathbf{i}, \mathbf{e}, \mathbf{p}) + \gamma U_e(\mathbf{e}, \mathbf{l}) + \kappa U_a(\mathbf{l}, \mathbf{g}) \quad (4)$$

Donde α, β, γ y κ son términos de peso. Todos estos términos de energía, han sido definidos en Gatica *et al.*, 1997.

Optimización global utilizando el método de modos iterativos condicionales ICM

Con el propósito de reducir el costo computacional, la optimización global de la solución propuesta es lograda utilizando un procedimiento iterativo de relajación determinística, este es el método de modos iterativos condicionales modificado (ICM), basado en una tabla de inestabilidad (Franois, 1991). Los métodos ICM minimizan la energía local ΔU_s en cada píxel $s(x,y)$ de la imagen. Nuestro esquema de minimización considera dos fases en cada iteración (García, 1995), una para la optimización del campo de etiquetas de segmentación, por medio de la minimización de:

$$U_1 = \alpha U_d(\mathbf{d}_x, \mathbf{d}_y, \mathbf{e}, \mathbf{p}) + \beta U_i(\mathbf{i}, \mathbf{e}, \mathbf{p}) + \gamma U_e(\mathbf{e}, \mathbf{l}) \quad (5)$$

Y otra para la optimización del campo de etiquetas de líneas de discontinuidad de movimiento, haciendo la minimización de:

$$U_2 = \gamma U_e(\mathbf{e}, \mathbf{l}) + \kappa U_a(\mathbf{l}, \mathbf{g}) \quad (6)$$

El término $U_e(\mathbf{e}, \mathbf{l})$ representa el enlace entre las dos fases del proceso general de optimización.

Algoritmo completo de segmentación basada en movimiento

El algoritmo completo de la segmentación basada en el movimiento consta de cuatro etapas:

- a) la inicialización.
- b) la numeración, el etiquetado o marcado de cada región en la imagen.
- c) la estimación en cada región de los parámetros del modelo de movimiento y
- d) la optimización de los campos de etiquetas.

Estas etapas se repiten tantas veces como sea necesario hasta que el método alcanza el número máximo de iteraciones, o bien, hasta que la segmentación se vuelva estable. Una de las ventajas de este algoritmo es que el número de regiones en la imagen no es fijo a través del proceso de segmentación, sino que se va adaptando (Figura 2).

Resultados de la segmentación

Los resultados obtenidos sobre las secuencias de prueba: *mano # 1* y *mano # 2*, haciendo la segmentación de las partes en movimiento, se muestran en las figuras 3 y 4, respectivamente. Los campos de segmentación alcanzados utilizando el algoritmo propuesto, se muestran en las figuras 3a y 4a. La superposición de *mano # 1* y *mano # 2*, con sus respectivas regiones

segmentadas se puede observar en las figuras 3b y 4b. De estos resultados se puede apreciar que los objetos en movimiento "manos", han sido relativamente bien segmentados, respecto al restante de las escenas. Estos resultados son cualitativamente correctos y se pueden alcanzar después de solamente dos iteraciones en el algoritmo de segmentación, para cada caso: *mano # 1* y *mano # 2*. Las regiones pequeñas que se quedan alrededor de las regiones principales, se pueden unir en iteraciones posteriores.

Hasta ahora, no se ha hecho algún procesamiento adicional para mejorar las fronteras de las regiones segmentadas. Los campos de vectores de movimiento que se obtuvieron con el método de Baaziz, son relativamente homogéneos y se ajustan adecuadamente a las áreas en movimiento. Los pixels no compensados que quedan con este algoritmo varían en-

tre el 2 y el 5% del total. Una entrada muy importante del algoritmo de segmentación son las regiones segmentadas binarias iniciales, esta inicialización fue obtenida comparando con un umbral, la diferencia entre las imágenes 1 y 2 para *mano # 1* y entre las imágenes 6 y 7 para *mano # 2*, enseguida se trasladó esta diferencia resultante a través de un filtro de mediana de tamaño 3x3 pixels.

Todo esto se logró al tener presente el uso de las regiones segmentadas para ayudar a gente con algún tipo de discapacidad, y así lograra comunicarse con una computadora. Una vez que la segmentación de las *manos* fue completada se definió un diccionario muy simple de *gestos*, los cuales pueden ser utilizados como apoyo de métodos diseñados para el aprendizaje de la lectura.

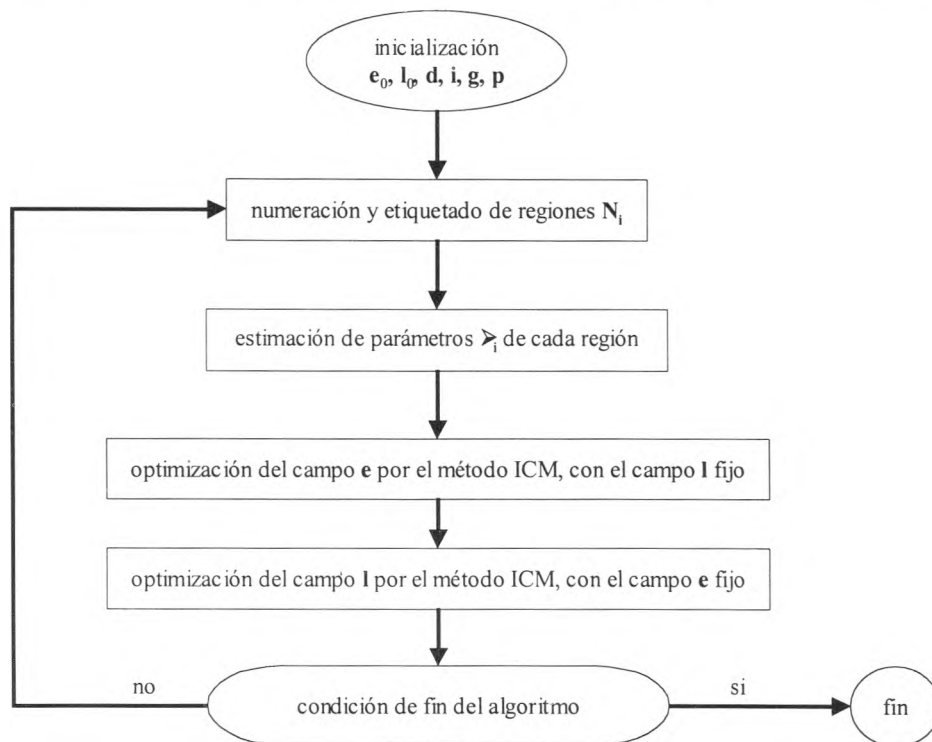


Figura 2. Diagrama general del algoritmo de segmentación basado en el movimiento

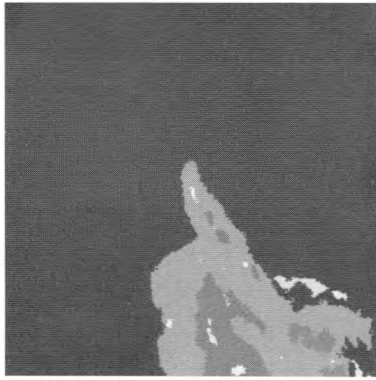


Figura 3a. Campo segmentado *e* de la señal mano #1



Figura 3b. Superposición de la imagen 3 de mano #1 con su campo segmentado *e*

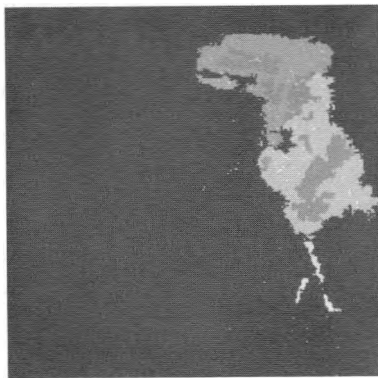


Figura 4a. Campo segmentado *e* de la señal mano #2

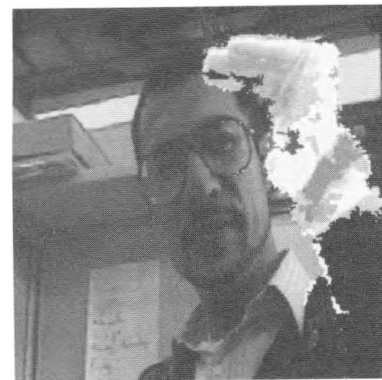


Figura 4b. Superposición de la imagen 7 de mano #2 con su campo segmentado *e*

Clasificación de los gestos

Para clasificar los campos segmentados de los gestos *e*, por ejemplo *mano #1* y *mano #2*, figuras 3a y 4a, se procedió inicialmente a la obtención de una versión binaria de estos campos, por medio de la comparación con un umbral, de tal manera que la clasificación pudiera realizarse rápidamente.

Luego, al basarnos en la transformada de Radon (Lim, 1990):

$$p(t) = \int_{u=-\infty}^{\infty} e(t_1, t_2) \Big|_{t_1=t \cos \theta - u \sin \theta, t_2=t \sin \theta + u \cos \theta} du \quad (7)$$

se han obtenido las sumas proyectadas sobre los ejes *x*, *y* de las intensidades binarias: 0° y 90° (Figuras 5 y 6) para *mano #1* y *mano #2*.

Por último, para reducir el número de datos a procesar, se realizó un filtrado pasa-bajas (semi-banda) y un submuestreo de orden 2 de las sumas proyectadas de las intensidades para cada orientación, de esta manera se redujo a la mitad el número de puntos o coeficientes necesarios para representar cada orientación. Esta última información fue utilizada para clasificar los gestos por medio de un entrenamiento de nuestro sistema con el diccionario definido y calculando el coeficiente de correlación

entre los que corresponden a las secuencias de entrenamiento y aquellos que se obtienen durante la operación continua del sistema.

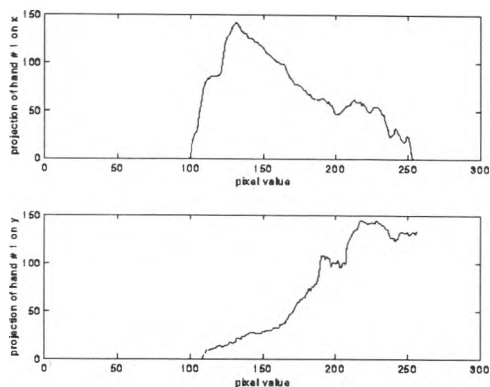


Figura 5. Sumas proyectadas de las intensidades binarias del campo segmentado e de mano #1, a 0° y 90° , ejes x , y respectivamente

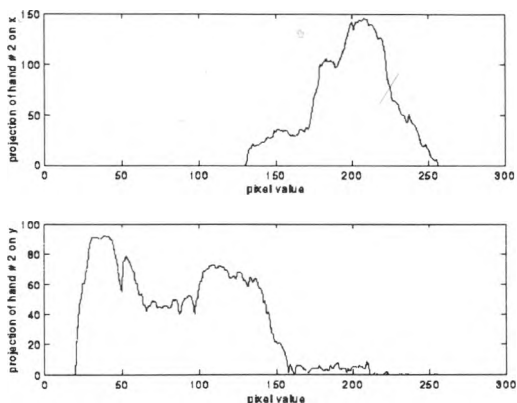


Figura 6. Sumas proyectadas de las intensidades binarias del campo segmentado e de mano #2, a 0° y 90° , ejes x , y respectivamente

Conclusiones

Los resultados logrados para la segmentación y clasificación de *gestos* de la mano como un medio de apoyo para gente con cierto tipo de discapacidad, son muy esperanzadores.

Para este diccionario simple de *gestos* se obtuvo un éxito del 100%. Cabe señalar que a medida que el diccionario se complete, tomando en cuenta *gestos* más complicados y a la vez parecidos unos a otros, la clasificación será más difícil y el por ciento de éxito será menor.

Este sistema es muy simple y fácil de obtener, ya que solo requiere de una pequeña cámara de video conectada a una computadora. Sin embargo, nuestra mayor desventaja es el tiempo de procesamiento: no estamos en condiciones de efectuar la ejecución de los programas en tiempo real. Comparando el sistema con otros desarrollados previamente (Starner y Pentland, 1995), (Quek *et al.*): se tiene un sistema más general, capaz de reconocer mayor número de *gestos* de la mano, pero aún no pueden operarse en tiempo real.

Reconocimientos

Este trabajo fue apoyado por la Universidad Nacional Autónoma de México (UNAM) y por el Consejo Nacional de Ciencia y Tecnología (CONACyT).

Referencias

- Gatica-Pérez D., García-Ugalde F. y García-Garduño V. (1997). Segmentation Algorithm for Image Sequences From a Pel-Recursive Motion Field. *Proc. of the VCIP, SPIE*, Vol. 3024, 1152-1163.
- François E. (1991). Interpretation Qualitative du Mouvement a Partir d'une Séquence d'images. Ph. D. Thesis, Université de Rennes I, France, June.

Quek F.K.H., Mysliwiec T. y Zhao M. *Finger mouse: A freehand pointing interface*. Vision Interfaces and Systems Lab. (VISLab). The University of Illinois at Chicago.

Lim J.S. (1990). *Two Dimensional Signal and Image Processing*. Englewood Cliffs, NJ: Prentice Hall.

Baaziz N. (1991). *Approches d'estimation et de Compensation de Mouvement Multiresolutions Pour le Codage de Séquences d'images*. Ph. D. Thesis, Université de Rennes I, Francia, Octubre.

Starner T. Y Pentland A. (1995). *Real-time American Sign Language Recognition From Video Using Hidden Markov Models*. MIT Media Lab. *Perceptual Computing Section Technical Report* No. 375.

García-Garduño V. (1995). *Une Approche de Compression Orientée-Objets par Suivi de Segmentation Basée Mouvement pour le Codage de Séquences d'images Numériques*. Ph.

D. Thesis, Université de Rennes I, Francia, Mayo.

Bibliografía sugerida

Netravali A.N. y Robbins J.D. (1979). *Motion-Compensated Television Coding: part I. The Bell System Technical Journal*, Vol. 58, No. 3, Marzo, 631-670.

García-Ugalde F., Savage-Carmona J., Furness T.A. III, Gatica-Pérez D. y García-Garduño V. (1997). *Segmentation of Moving Human Body Parts by a Modified MAP-MRF Algorithm*. *Proc. of the VSMM, IEEE*, 197-205.

Daubechies I. (1988). *Orthonormal Bases of Compactly Supported Wavelets*. *Comm. Pure and Applied Mathematics*, Vol. 41, 909-996.

Biamond J., Looijenga L., Boekee D.E. y Plompen R. (1987). *Pel-Recursive Wiener-Based Displacement Estimation Algorithm*. *Signal Processing*, Vol. 13, No. 4, diciembre, 399-412.

Semblanza de los autores

Francisco García-Ugalde. Obtuvo el título en 1977 como ingeniero mecánico electricista en el área de ingeniería de sistemas eléctricos y electrónicos en la Facultad de Ingeniería, UNAM. Asimismo, finalizó la maestría en ciencias en el área de sistemas de Comunicaciones/Electrónicas del Estado Sólido de la Ecole Supérieure d'Electricité de Francia (SUPELEC) en 1980 y el doctorado en ingeniería en el área de procesamiento de la información de la Universidad de Rennes I de Francia en 1982. Es profesor desde 1983 en la División de Estudios de Posgrado de la Facultad de Ingeniería de la UNAM. En 1991, fue profesor visitante en el IRISA/INRIA-Rennes, Francia y en el HITLab del Departamento de Ingeniería Eléctrica de la Universidad de Washington, Seattle, U.S.A. en 1997. Es miembro del SIN y Catedrático UNAM. Sus áreas de interés son: codificación de imágenes y video, análisis de imágenes, teoría y aplicaciones de códigos correctores, codificación conjunta fuente-canal, turbo codificación, análisis del diseño de filtros digitales, arquitectura de computadoras y procesamiento en paralelo.

Víctor García-Garduño. Obtuvo el título de ingeniero mecánico electricista en 1986 y la maestría en ingeniería eléctrica en 1990, ambos en la Facultad de Ingeniería de la UNAM. Asimismo, en 1995 terminó el doctorado en Procesamiento Digital de Señales y Telecomunicaciones por la Universidad de Rennes I Francia. Su área de interés es el análisis de video digital para fines de transmisión y manipulación.

Daniel Gatica-Pérez. Se tituló en 1993 como ingeniero en electrónica en la Universidad de Puebla, posteriormente como maestro en ingeniería eléctrica en la División de Estudios de Posgrado de la Facultad de Ingeniería en la UNAM en 1996. Así también, concluye el doctorado en ingeniería eléctrica de la Universidad de Washington, Seattle, U.S.A. en el año de 2001. Actualmente es investigador en el IDIAP de Suiza y sus áreas de interés son análisis de video, máquinas inteligentes y morfología matemática.