

# Un Proceso de Aprendizaje para Reconocimiento de Objetos en Línea en Tareas Robotizadas

Mario Peña-Cabrera, Roman Osorio  
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS-UNAM)  
Circuito Escolar, Cd. Universitaria, D.F. CP 4100, México.  
([mario@leibniz.iimas.unam.mx](mailto:mario@leibniz.iimas.unam.mx))

y

Ismael López-Juárez, Reyes Ríos-Cabrera  
CIATEQ A.C., Centro de Tecnología Avanzada,  
Manantiales 23ª, Fracc. Ind. B.Q., El Marqués, Querétaro. CP 72246. México.  
([ilopez@ciateq.mx](mailto:ilopez@ciateq.mx))

## RESUMEN

El desempeño de tareas de ensamble con robots industriales que trabajan en ambientes no estructurados, puede ser mejorado utilizando percepción visual y nuevas técnicas de aprendizaje. En este trabajo se presenta un método novedoso que utiliza datos en 2D y técnicas simples de procesamiento de imágenes para éste propósito. Un descriptor único de la imagen (CFD&POSE) que contiene también información de profundidad es obtenido y alimentado a una arquitectura de red neuronal del tipo FuzzyARTMAP con propósitos de reconocimiento y aprendizaje. Este vector contiene información de objetos en 3D orientada a realizar tareas de ensamble con robots industriales y es invariante al escalamiento, rotación y la traslación. El método junto con las capacidades rápidas de aprendizaje de las redes neuronales artificiales ART conforman una potente herramienta para las aplicaciones en automatización de tareas de tiempo real en línea.

**Palabras Claves:** visión computarizada, ensamble, imagen, red neuronal artificial, tiempo real, aprendizaje.

## 1. INTRODUCCION

El advenimiento de sistemas roboticos complejos, en diferentes aplicaciones como: manufactura, ciencias de la salud y aeroespaciales, ha desarrollado una demanda para utilizar sistemas de visión artificiales con mejores características y desempeño. Los sistemas de visión deben ser capaces de ver y percibir objetos e imágenes quizá los más parecido a como lo hace el ser humano. Esta apreciación ha llevado a los investigadores a considerar y estudiar el diseño de sistemas de visión artificiales con una apreciación orientada a la morfología neuronal de los sistemas biológicos de visión humana. Diferentes expertos en neuroanatomía y neurofisiología que han realizando experimentos con diferentes especies biológicas, han descubierto hechos fascinantes que nos indican el modo y la trayectoria de las señales de visión, desde los puntos iniciales en la retina hasta puntos cerebrales en la corteza visual. Por otro lado, científicos de disciplinas como ciencias computacionales y matemáticas, han formulado las teorías de las funciones neuronales en la trayectoria visual desde el punto de vista matemático y computacional. Todo esto ha llevado a los

científicos a un mejor entendimiento de cómo las estructuras de redes computacionales y sistemas artificiales de visión deben de ser diseñados mostrando los paradigmas neuronales, modelos matemáticos, arquitecturas computacionales e implementaciones del *hardware* requerido; cuando un sistema contempla todos estos aspectos, podemos hablar de un “Sistema de Neur-Visión” y se puede definir como una máquina artificial con la que se puede ver nuestro mundo visual y crear aplicaciones de nuestra vida diaria.

Basándose en este razonamiento, y para un mejor entendimiento, podemos hablar de dos áreas muy interesantes dentro del campo de la investigación: la morfología visual de los sistemas biológicos de visión y los paradigmas utilizados en las redes neuronales artificiales para el desarrollo de los sistemas de neuro-visión, en donde para su diseño existen preguntas que uno se tiene que hacer como: ¿qué es un sistema de neuro-visión?, ¿que funciones involucra?, ¿que elementos básicos debe tener el sistema para un funcionamiento robusto?, considerando la información de una escena visual, las máquinas desarrolladas tienen que ver, memorizar, comprender lo que ven o han visto y tener un modo de adquisición por naturaleza, en modo remoto. Esto nos lleva a considerar para el diseño de estos sistemas, nuevas arquitecturas con estrategias computacionales que intenten modelar el comportamiento de la visión humana, considerando sin embargo, las restricciones que los dispositivos digitales computacionales existentes presentan [1].

El propósito de esta publicación es ciertamente mostrar un método novedoso y simple que considere el diseño de un sistema de visión robusto, flexible y con atributos como fácil implementación y operación en aplicaciones de tiempo real en tareas de manufactura y ensamble con sistemas robotizados.

Gran parte de inspiración para el razonamiento de las ideas mostradas en esta publicación han sido obtenidas por haber realizado trabajo de investigación en el comportamiento de bebés humanos, observando la manera en que ellos aprenden un nuevo objeto, y como se comportan al presentarles objetos previamente aprendidos en cuanto a su preferencia por tamaños, formas y colores. De ello podemos decir que existen dos factores importantes que hemos tomado como conclusiones para nuestro trabajo de investigación:

- 1) Tiene que haber un primer punto dentro de la escena de visión, a partir del cual se pueden referenciar o derivar la extracción de características y atributos y parece este ser el

- 2) centro de masa, es decir la parte más pesada dentro del modelo de imagen.
- 3) Una vez que un objeto ha sido presentado y aprendido, la siguiente vez que se presenta para ser reconocido se busca por pistas e indicaciones para un rápido reconocimiento del objeto, utilizando luego la información memorizada en cuanto a características y atributos. Esto nos lleva a pensar en que no es necesario utilizar en aplicaciones de tiempo real enfocados a la industria cada vez un modelo complejo de los objetos, y si en cambio encontrar métodos rápidos para sacar la información previamente memorizada y tomar una decisión de que objeto se está reconociendo con un cierto grado de certidumbre.

## 2. SISTEMAS NEURO-VISUALES

El primer método que la mayoría de los diseñadores de sistemas neuro-visuales utilizan es el llamado *paradigma señal-símbolo*, en este método, el objetivo es tener descripciones significativas de la escena de visión partiendo de datos crudos y utilizando tres niveles de análisis como se muestra en la figura 1,

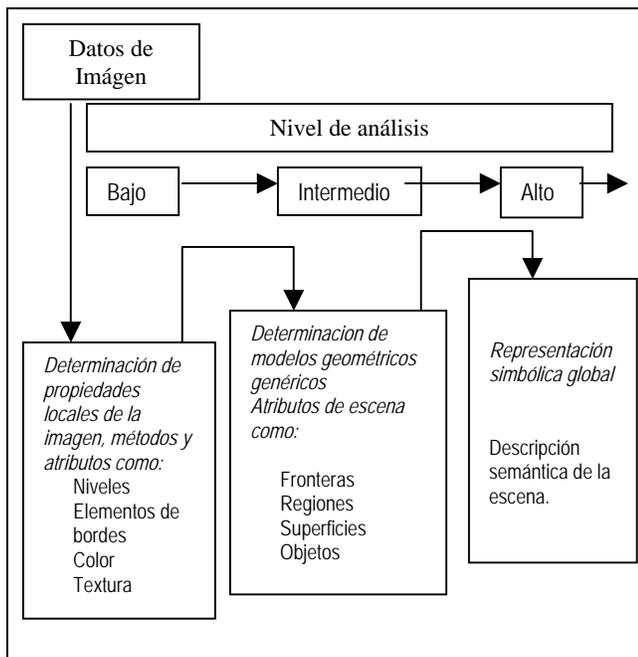


Figura 1. Paradigma señal-símbolo

El costo computacional en este método es alto, y especialmente si se aplican métodos de análisis con variación en el tiempo, se puede hablar de millones de instrucciones realizadas por cada escena, los nuevos procesadores más rápidos en la actualidad resuelven el problema parcialmente pero todavía, esto no es comparable a las velocidades de nuestra propia experiencia visual, con la que obtenemos información significativa en un rango de los 70 a 200 milisegundos y utilizando relativamente pocos elementos procesadores llamados neuronas, necesitando cada una de ellas alrededor de unos 2 milisegundos para generar una respuesta. Considerando que muchos aspectos en la visión biológica primaria se llevan a cabo en solo 18 a 46 pasos de

transformaciones de imágenes [2], parece lógico buscar otro tipo de procesos y métodos para encontrar y describir objetos en escenas visuales utilizando menos recursos computacionales, ya que el objetivo es tratar de emular el comportamiento visual humano, teniendo en cuenta además, que ha sido estimado que el 60% de la información sensorial en los humanos es a través de una trayectoria visual [3].

Es conocido que las arquitecturas de los sistemas visuales biológicos y considerando el recorrido de la trayectoria de la señal visual esta masivamente paralelizada y utiliza un proceso de información jerárquica [4]. La transformación y reorganización de los datos de visión en representaciones abstractas es similar al llamado paradigma de señal-símbolo que se ha expuesto anteriormente, y que efectúa cómputo de datos en paralelo involucrando procesamiento de la información visual en el espacio (plano X-Y) y el tiempo. Desde una perspectiva ingenieril, no se ve necesario y además sería prácticamente imposible emular los aspectos electrofisiológicos detallados del proceso biológico de visión, sin embargo, es deseable tener estructuras computacionales neuronales inspiradas en la visión biológica que contemplen el procesamiento, almacenamiento e interpretación de la información visual desde el punto de vista espacio-temporal [5].

## 3. NEURONA Y SU MODELO BIOLÓGICO

La neurona es el elemento computacional biológico básico, típicamente existen alrededor de  $10^{11}$  neuronas en el sistema nervioso central de un ser humano, la neurona está compuesta básicamente de un procesador llamado *soma*, múltiples entradas llamadas *dendritas* y una sola salida llamada *axon*, cada neurona puede tener hasta unas  $10^4$  conexiones de entrada pero una sola salida.

El diagrama y modelo simplificado de una neurona biológica desde un punto de vista del procesamiento de la información se muestran en las figuras 2 y 3, la *sinapsis* contempla dos funciones importantes: conversión de frecuencia de pulsos a voltaje y memoria de largo término (LTM).

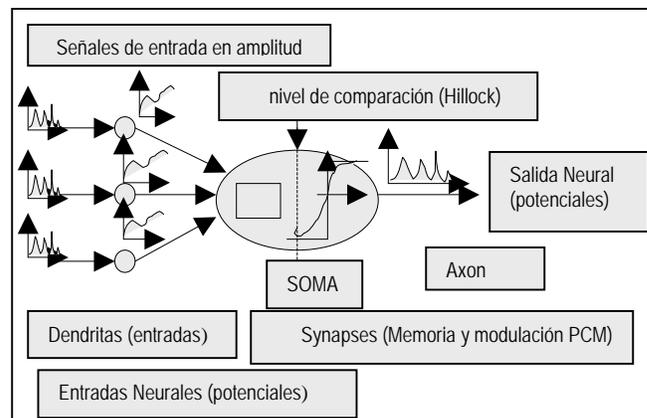


Figure 2. Diagrama simplificado de neurona biológica con perspectiva de procesamiento de información.

Estudios experimentales en neurofisiología han demostrado que la acción potencial de una neurona es una variable aleatoria [6], y utiliza un ancho de banda de 500 hz. o menos, de tal manera,

que para llevar a cabo procesamiento neuronal eficiente y rápido con elementos impredecibles, se necesitan emplear un gran número de ellos con el objeto de establecer computación en paralelo.

#### 4. REDES NEURONALES COMPUTACIONALES

Desde un punto de vista computacional, capas individuales de neuronas como la retina, pueden ser conceptualizadas como uno o mas arreglos de dos dimensiones (redes neuronales) de neuronas realizando alguna operación específica con las señales visuales. Las características estructurales primarias de una red neuronal computacional (CNN) son:

- 1) Una morfología organizada conteniendo muchas neuronas distribuidas en paralelo.
- 2) Un método de codificación de información dentro de las conexiones sinápticas de las neuronas.
- 3) Un método para retraer información cuando se presente un patrón de entrada como estímulo.

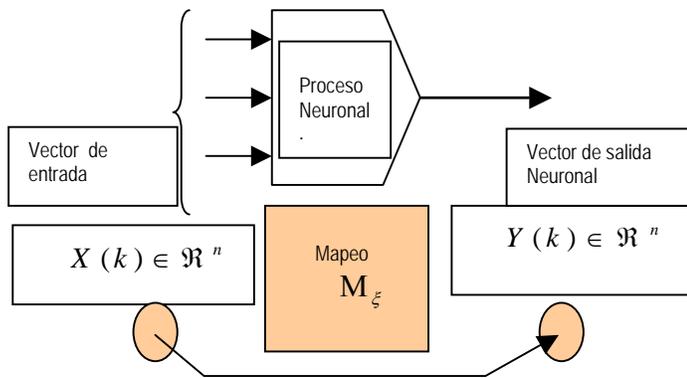
Desde un punto de vista de procesamiento de señales, la neurona biológica tiene dos elementos claves la *sinapsis* y el *soma*, y son responsables de efectuar tareas computacionales como: aprendizaje y adquisición de conocimiento (memoria de experiencias pasadas y reconocimiento de patrones). En términos simples, una neurona puede ser descrita como un elemento de procesamiento de información que recibe un vector neuronal de entrada n-dimensional (ecuación 1), que representa la señal que se transmite desde la neurona n-vecinal sensorial y es recibida por la neurona en cuestión.

$$X(k) = [x_1(k), x_2(k), \dots, x_i(k)] \in \mathfrak{R}^n \quad (1)$$

Matemáticamente, la habilidad de procesamiento de información de una neurona, puede ser representada como una operación de mapeo no-lineal  $M_\xi$ , desde el vector de entrada

$X(k) \in \mathfrak{R}^n$  al escalar de salida  $Y(k) \in \mathfrak{R}^n$  esto es:

$$M_\xi : X(k) \in \mathfrak{R}^n \rightarrow Y(k) \in \mathfrak{R}^n \quad (2)$$

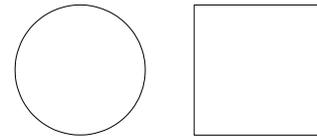


**Figure 3.** Modelo simplificado de una neurona biológica (perspectiva procesamiento de información).

#### 5. NUESTRA PERSPECTIVA

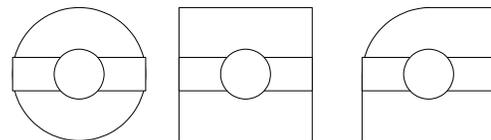
- 1) Pensamos que es posible obtener información rápida y confiable con un análisis simple pero enfocado a lo que del objeto se debe mostrar como la más primitiva y necesaria información para tener un conocimiento substancial y robusto de lo que se esta viendo, posteriormente almacenar los aspectos más importantes de la escena (le hemos llamado "pistas"), las cuales más tarde pueden ser utilizadas para retraer los aspectos memorizados del objeto sin tener que retraer todas las características detalladas. En cierta manera nosotros los humanos hacemos ese proceso una vez que ya se ha visto y aprendido de un objeto por primera vez.
- 2) Creemos que aprendiendo formas canónicas dentro del proceso inicial de conocimiento, mas tarde es posible reconstruir el conocimiento del objeto utilizando los aspectos primitivos y perceptuales de grupo como en las Leyes de Gestalt acerca de proximidad de grupos, y factores de similaridad y simplicidad. Se muestra esta idea original con partes simples que se estan utilizando en la experimentación con aplicaciones de ensamble dentro de una celda de manufactura inteligente.

Para el aprendizaje inicial de conocimiento, se consideran un círculo y un rectángulo como las dos formas canónicas que se muestran en la figura 4.



**Figure 4.** Formas canónicas para el aprendizaje inicial de conocimiento.

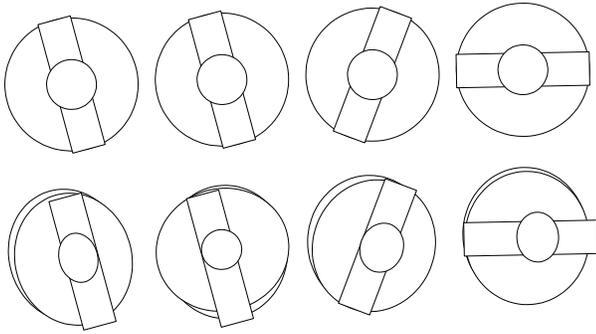
Si nuestro sistema puede aprender estas formas canónicas y se implantan al conocimiento inicial (conocimiento *a priori*), creemos que es posible representar información de objetos reales en 3D para fines de ensamble con imágenes de 2D que representan las piezas a ser ensambladas y a las que hemos llamado: circular, rectangular y combinada como se muestra en la figura 5.



**Figure 5.** Representación de piezas de ensamble en 2D.

Estas piezas, pueden ser contruídas con las formas canónicas agrupadas de diferentes maneras y conforme al conocimiento *a*

priori, considerando estas formas de agrupación como “pistas”, estas pueden ser codificadas y representadas por un *Vector Descriptivo* que contenga toda la información necesaria para implementar la idea.

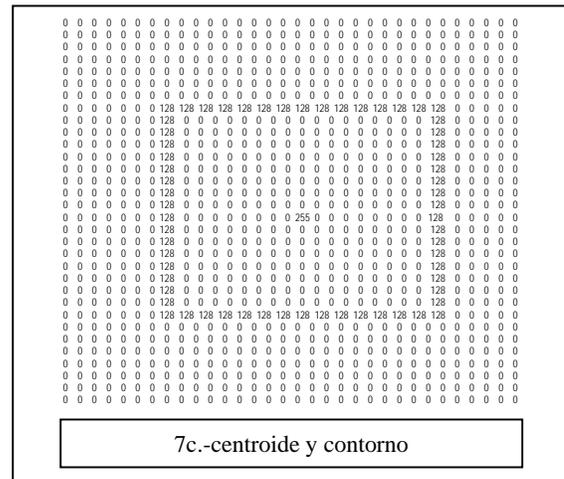
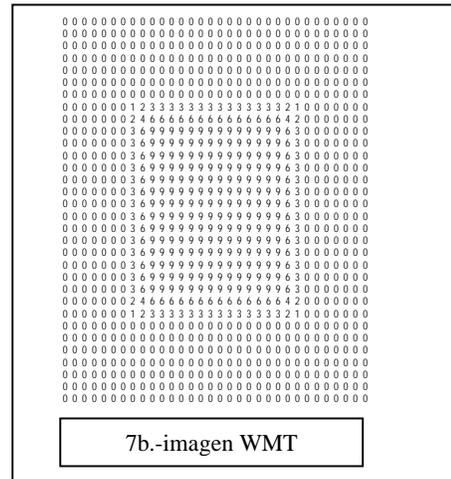
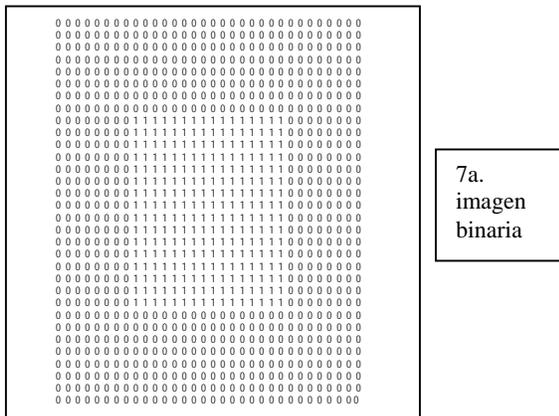


**Figure 6.** Familia de Vector Descriptivo para piezas de ensamble.

Teniendo un *Vector Descriptivo* como el mencionado, es posible entrenar a una red neuronal artificial (ANN) en línea con el mismo sistema de visión, y es de esperarse que pueda obtener un conocimiento incremental para conformar familias de vectores descriptivos como se muestra en la figura 6. Con muchos procesos de aprendizaje, el sistema incrementa su conocimiento y cada vez el proceso se hace más rápido porque actúa solamente en *proceso de recuerdo* teniendo así, un beneficio en costo computacional para implementar el algoritmo y crear el *Vector Descriptivo*, este proceso (creación del vector descriptivo del componente-pieza de ensamble) llega a ser un mecanismo autónomo y la red ARTMAP mandará cada instancia a su respectivo grupo para clasificarlo.

#### Generación del vector descriptivo

El algoritmo para generar el Vector Descriptivo, que hemos llamado [CFD&POSE] se muestra para una de las formas canónicas (rectángulo), primeramente, una imagen binaria es generada de la imagen original adquirida utilizando técnicas de segmentación con operadores de umbral y análisis de histogramas 1D y 2D (figura 7a). Aplicando el algoritmo CFD&POSE se crea una imagen con información del contorno y centroide del objeto (figura 7c), se utiliza un método original para obtener su codificación a partir de un conjunto de pares numéricos en la imagen WMT que se obtienen de una transformación de pesos (figura 7b).



**Figura 7.** Imágenes en el proceso del algoritmo CFD&POSE. a).- imagen binaria, b).- imagen WMT c).- imagen CFD&POSE.

La Matriz de Transformación de Pesos ( $H_{Wf}$ ) genera la imagen WTM mostrada en la figura 7, en donde se encuentra un conjunto de pares numéricos (números de peso):

$$\text{Número } Nw_f \rightarrow [\text{bin} \{ \text{de coordenadas numéricas} \}] \quad (3)$$

donde los valores generados son :

$$Nw_f \min \leq \sum (\mathbf{I}'s) \in (kxk) \text{ KernelPixek}(i, j) \leq Nw_f \max \quad \forall i, j = 1, M \quad (4)$$

de la imagen WMT, se observa que  $Nw_f \max$  calculado denota el centroide del objeto y los  $Nw_f \min$  los puntos frontera del contorno. El centroide calculado se obtiene para cada coordenada  $X_c$  y  $Y_c$  de la suma de los  $Nw_f \max$  entre el número de ellos en la imagen como se muestra:

$$X_c = \frac{1}{No.NWf_{max-X}} [\sum NWf_{max-X}]$$

$$Y_c = \frac{1}{No.NWf_{max-Y}} [\sum NWf_{max-Y}] \quad (5)$$

y para los puntos frontera tenemos:

$$vectorX(m) = \{x_0, x_1, x_2, \dots, x_n\}_{NWf_{min}}$$

$$vectorY(m) = \{y_0, y_1, y_2, \dots, y_n\}_{NWf_{min}} \quad (6)$$

estos conjuntos de puntos codificados se estandarizan para un tamaño de rejilla angular  $A\gamma(\theta)$  interpolando valores, donde el objeto es referenciado. El centroide y coordenadas de los puntos frontera, son obtenidos también como un vector de donde es posible extraer características de las formas de los objetos. Se obtiene una función de frontera del objeto (BOF) calculando las distancias de bordes a centroide. Para el caso de una rejilla angular de tamaño 16 se tiene:

$$\{P_1, P_2, P_3, \dots, \max P(n)\} \text{ para } n=16 \quad (7)$$

y para fines de demostración otorgando valores numéricos correspondientes a la forma canónica del rectángulo el vector de distancias sería:

$$\{20, 22, 28, 22, 20, 22, 28, 22, 20, 22, 28, 22, 20, 22, 28, 22\}$$

las distancias para obtener el BOF (función de frontera del objeto) son calculadas con :

$$D_n = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (8)$$

$$\forall 0 \leq n \leq \text{tamaño de rejilla angular}$$

ya la Función de Frontera del Objeto se define como:

$$BF = f(D_n, No.angular / grid)$$

dependiendo de la rejilla angular que se escoga, la resolución del ángulo para fines de rotación es:

$$\text{Angulo } \Phi = 360^\circ / No.angular / grid$$

el cual determina el número de puntos frontera que se escogen y la referencia de orientación obtenida en el orden de las manecillas del reloj (CW) ej.: rejilla para 8:

N	NE	E	SE	S	SW	W	NW
D0	D1	D2	D3	D4	D5	D6	D7

El vector descriptivo es normalizado y construido de acuerdo a la rejilla angular escogida, proporcionando los pasos de incrementos angulares que determinan los movimientos de rotación y con lo que se obtiene la invariancia rotacional con el método de codificación. Cada  $D_n$ , el centroide  $C(x,y)$  y ID/clues's son codificados en un formato con dígitos BCD [bbb.f] y los "clues" se estructuran de manera que se indica las formas canónicas y las secuencias de agrupación necesarias para la definición del objeto conformándose el vector descriptivo como se muestra en la figura 8.

$[CFD\&POSE]_{vector} =$

D0 / 00041200 D1 / 00047000 D2 / 00041200 D3 / 00047000 D4 / 00041200 D5 / 00047000 D6 / 0..... DF / 00047000 Cxy / 00160000 Φ.... / 00000001 ID0 / 44404000 ID1 / 00001200	BOF : codificación de función de frontera del objeto
	Centroide & Orientación
	Profundidad, nombre de objeto ID e información de secuencia grupal y "clues".

Figura 8. Estructura del vector descriptivo CFD&POSE

para la forma canónica "rectángulo" las imágenes BOF y el vector CFD&POSE se muestran en la figura 9.

$[CFD\&POSE]_{vector}$

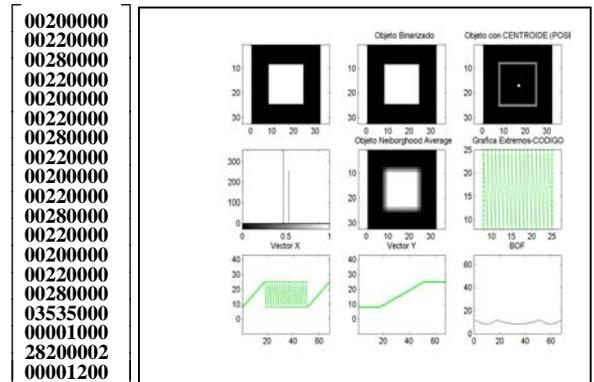


Figure 9. CFD&POSE y función de frontera para la forma canónica rectangular.

## 6. RESULTADOS EXPERIMENTALES

Se preparó un área de trabajo con una cámara progresiva e iluminación apropiados, se instalaron los algoritmos CFD y las rutinas de adquisición de imágenes en una computadora PC. Se obtuvieron los vectores descriptivos en línea y las funciones *Boundary object Function* para los tres tipos diferentes de piezas mostradas en la figura 5. Con las funciones se entrenó a un modelo de red neuronal artificial del tipo FuzzyARTMAP y luego se probó su clasificación con los resultados mostrados en las figuras 10, 11 y 12, las piezas fueron presentadas 10 veces cada una para diferentes localizaciones, rotaciones y

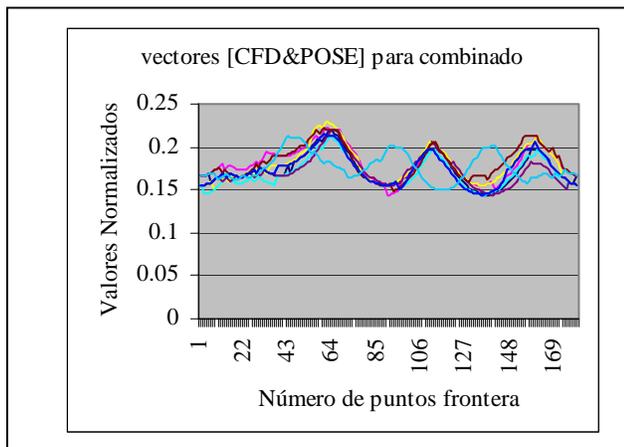
escalamientos, siempre con una altura constante ( $z=fija$ ), la combinación utilizada en el experimento y la codificación Para clasificar los objetos, en la red neuronal artificial de las piezas de trabajo se muestra en la tabla 1

$\phi$ ángulo rotado	0	45	90	135	180	225	270	315	360
50 patrones circulares	x	x	x	x	x				
40 patrones cuadrados	x	x	x	x					
80 patrones combinados	x	x	x	x	x	x	x	x	
<b>total: 170 patrones</b>									

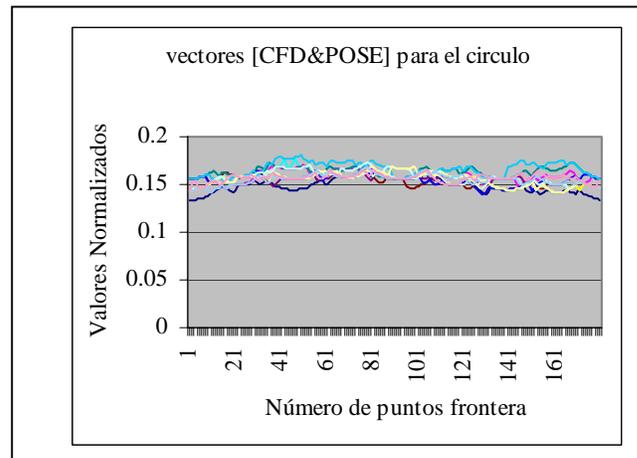
Codificación para clasificación neuronal	
1000=	Círculo
1100=	Cuadro
1010=	Combinado

**Table 1.** Distribución de las piezas de trabajo para el experimento y codificación en la red neuronal.

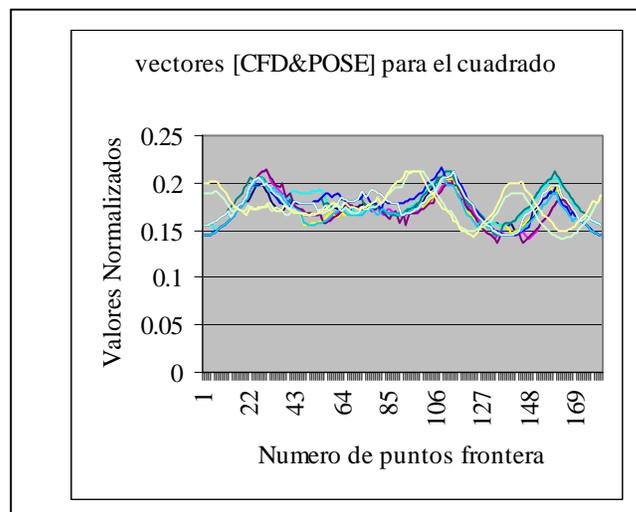


**Figura 10.** vector descriptivo para pieza combinada.

Al final del proceso, el sistema obtiene un vector descriptivo [CFD&POSE], que contiene la información de la identificación de la pieza de trabajo y su posición y orientación, la orientación se obtiene de analizar en la función de frontera del objeto, la distancia mayor y se establecen los números de corrimiento de puntos frontera con respecto a el punto de referencia norte-norte (nn), relacionando la cantidad de grados por punto frontera desplazado y de acuerdo a un patron de rejilla angular que se establece en la programación del algoritmo.



**Figura 11.** vector descriptivo para pieza círculo.



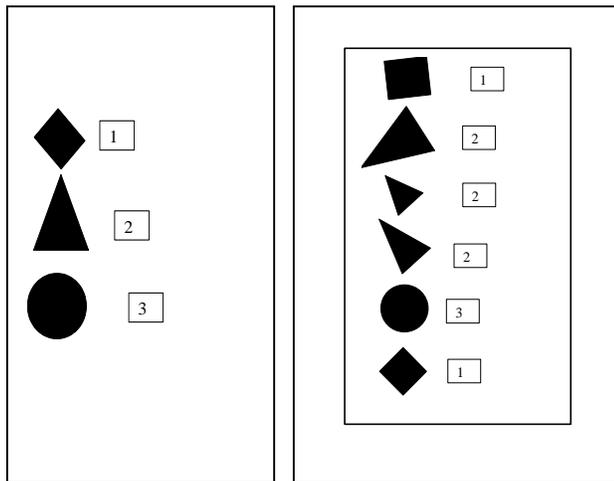
**Figura 12.** vector descriptivo para pieza cuadrado.

## 7. RESULTADOS Y CONCLUSIONES

Se ha obtenido un método para codificar información de imágenes para el reconocimiento de objetos basándose en el análisis de formas canónicas. Hasta ahora se han hecho experimentos con ejemplos simples con el objeto de entrenar a una red neuronal artificial y así obtener resultados que validen nuestra investigación, observando el buen desempeño y desarrollo de este método novedoso, esperamos mejores resultados con objetos más complejos. El vector descriptivo ha sido definido de una manera estática hasta ahora, la construcción de un vector descriptivo dinámico que pueda adaptar sus dimensiones a cada objeto puede ser hecho en un trabajo futuro con la finalidad de expandir las aplicaciones. Resultados del método para clasificación invariante exitosa se muestran en la figura 13, en donde se aplicaron los vectores CFD&POSE para formas canónicas rectángulo, círculo y un triángulo (no canónico) a una red del tipo FuzzyARTMAP, en

primera instancia se presentaron a la red para ser aprendidas y luego se volvieron a presentar primeramente bajo las mismas condiciones de escalamiento, rotación y traslación y se observaron resultados impresionantes al tener un buen desempeño y rapidez aceptables (milisegundos) para las aplicaciones en tiempo real en línea, la clasificación fue correcta posteriormente con las formas canónicas presentandolas trasladadas, rotadas o escaladas. Los resultados experimentales mostraron la factibilidad del uso de esta metodología para aplicaciones de manufactura inteligente realizando tareas de ensamble con robots en tiempo real [7].

7. López-Juárez, M. Howarth. Learning Manipulative Skills with ART. IEEE/RSJ, **Proc. International Conference on Intelligent Robots and Systems (IROS'2000)**, Takamatsu, Japan, Vol 1, pp 578-583 ISBN 0-7803-6351-5. 2000.



Patrones entrenados	Patrones probados
FuzzyARTMAP	Complement Coding
Rho map 0.7	Learning Time $\leq 1.0$ ms.
Learning Rate =0.1	Pentium IV, 2.6 GHz Procesor

**Figura 13.** Resultados de clasificación invariante con 6 diferentes vectores CFD&POSE

## 8. REFERENCIAS

1. A. Rosenfeld,. Computer Vision: A Source of Models for Biological Visual Processes, **Biomedical Engineering**, 36,1, 1989, pp. 93-96.
2. L. Uhr.. Highly parallel, hierarchical, recognition cone perceptual structures. **Parallel Computer Vision**, L. Uhr Ed., 1987, pp. 249-292.
3. R.E. Kronauer, Y. Zeevi. ,Reorganization and Diversification of Signals in Vision. **IEEE Transactions on Systems, Man and Cybernetics.**, SMC-15,1,1985, pp. 91-101.
4. L. Uhr.. Psychological motivation and underlying concepts . **Structured Computer Vision**, S. Tanimoto ,A. Klinger Ed. , 1-30, 1980.
5. Douglas G. Granrath . The Role of Human Vision Models in Image Processing. **Proc. IEEE** 69 ,5, 1981, pp. 552-561.
6. R.J. McGregor, R.Lewis , Neural Modelling Electrical Signal Processing in the Nervous System. **Plenum Press**, 1977.