
Modelo de representación y procesamiento de movimiento para diseño de arquitecturas de tiempo real especializadas

Francisco Flórez Revuelta

Tesis de Doctorado

Facultad: Escuela Politécnica Superior

Director: Dr. Juan Manuel García Chamizo

2001

Tesis Doctoral

Modelo de representación y
procesamiento de movimiento
para diseño de arquitecturas
de tiempo real especializadas

Francisco Flórez Revuelta

Tesis Doctoral

Modelo de representación y
procesamiento de movimiento
para diseño de arquitecturas
de tiempo real especializadas

presentada por:

Francisco Flórez Revuelta

dirigida por:

Dr. Juan Manuel García Chamizo

Programa de Doctorado

“Sistemas industriales, computación y reconocimiento
de formas”

Departamento de Tecnología Informática y
Computación

UNIVERSIDAD DE ALICANTE

Diciembre, 2001

*Para la I chi y
Mari Pulgui*

AGRADECIMIENTOS

Al ponerme a escribir estas páginas me surgen algunas preguntas para determinar quién me ayudó a realizar este trabajo. ¿Cuándo comencé a escribir esta tesis? ¿Por qué este trabajo? ¿Cuándo di con la solución? ¿Qué me llevó a ella? Ni yo mismo sé cuándo di por iniciada la tesis. ¿Quizás cuando acabé los cursos de doctorado? ¿Quizás al empezar el proyecto fin de carrera? Es probable que antes. Muchas personas han pasado por mi lado durante todo este tiempo. Tantas, importantes, que es muy fácil dejar en el olvido a alguna.

En primer lugar, a mi mujer Iciar y mi hija María. Vuestro apoyo y cariño es lo más importante, nada tiene valor sin ellos. Estos últimos meses han sido de los más duros en mi vida, teniendo que estar tanto tiempo separado de vosotras, sabiendo que estabais disfrutando la una de la otra, llegando a casa para que me contarais qué habíais hecho. A partir de ahora, volveréis a ser el centro de mi vida. Os necesito. Os quiero.

A mis padres y a Nane. Podríamos decir que éste es el final de mi etapa de estudiante. Sin vuestro esfuerzo y apoyo no hubiera llegado a ningún lugar. Ahora me toca a mí, enseñar a María, los valores que me inculcasteis y educarla como vosotros habéis hecho conmigo. Os quiero.

Aunque en la portada aparece únicamente mi nombre, este trabajo no sería el que es sin la ayuda, el apoyo, los consejos, las

revisiones, las regañinas,... de Juanma. Has cumplido mucho más allá de lo asociado con tu papel de tutor. Muchísimas gracias.

A José Luis y Jero, compañeros de instituto, de universidad, de piso, de trabajo,... Pero, ante todo, amigos. ¡Cuántos días juntos! ¡Cuántas vivencias! Recordaba hace poco, aquellos días del verano del 93, contigo, Jero, y tu viejo 8086, descubriendo las redes neuronales, intentando reconocer varias letras con una red de propagación contraria. ¡Qué cosas! ¿No tendríamos nada más que hacer?

A mis colaboradores en la etapa final del trabajo, Pepe y Antonio; que me ayudaron en la realización de experimentos, comprobación de resultados, redacción y verificación del documento,...

A la gente del departamento, Paco, Andrés, Antonio, Ginés, Joan Carles, Toni, María Teresa, Sergio, Javi, Paco Mora, Higinio, Dani; por compartir conmigo trabajos, charlas, cafés, fútbol,...

A Ramón, Otto, Domingo, Sco, Faraón, Rosana y resto de gente de i³a, con quienes compartí los inicios de esta tesis. Muchas de las ideas de este trabajo, parten de conversaciones con vosotros.

A Rafa, Pablo, María José, Asun, Maribel, Cande, Andrés, Carlos, Manolo, Tía Lola, María Dolores (sí, ya la acabé), Ángel, Marta, Christian, Laura, Javi, Marta, José Manuel, Yoli, Josele, María José, José Antonio, Mari, Ángel, Andrés, Rosa, Enrique, Armando, José Manuel,...

A todos aquellos que dejo en el olvido.

A todos. Gracias.

Paco

Alicante, a 1 de Noviembre de 2001

RESUMEN

La investigación realizada aborda el problema del seguimiento de objetos y análisis del movimiento en escenas. Este trabajo se enmarca dentro del proyecto de investigación "Sistema de visión para navegación autónoma" (CICYT TAP98-0333-C03-03). El objetivo principal de este proyecto es el desarrollo de un dispositivo de visión artificial para la ayuda a la navegación autónoma.

Dado que se plantea la integración del módulo de procesamiento de visión en el dispositivo, es necesario dotar a este último de potencia computacional suficiente para procesar las imágenes a frecuencia de video. Este hecho establece la necesidad de diseñar arquitecturas de alto rendimiento, con capacidad de refinamiento de la respuesta en función del tiempo disponible, especializadas para los problemas de visión; todo ello con enfoque de miniaturización y empotrabilidad. Para una mejor comprensión del problema y cómo abordarlo, se han revisado numerosos trabajos relacionados con él y se han considerado los diversos enfoques de la visión activa.

Se ha propuesto el empleo de redes neuronales auto-organizativas para la representación de los objetos, por su cualidad de preservadoras de la topología, lo que proporciona el Grafo Preservador de la Topología (*GPT*). El mismo tiene alta

capacidad expresiva, es sencillo de obtener, y resulta robusto. Se aporta, asimismo, una técnica de síntesis de los objetos a partir de dicha representación. Las características que se extraen del *GPT* simplifican las operaciones de clasificación y reconocimiento posteriores al evitar la alta complejidad de la comparación entre grafos. Como ejemplo de aplicación de la caracterización de los objetos, se propone un problema de clasificación de formas a partir de su contorno.

El modelo de caracterización de los objetos permite refinar la calidad de representación en función del tiempo disponible para su cálculo, de modo que sirva como base para el diseño de arquitecturas de visión de alto rendimiento que operen bajo restricciones de tiempo real.

Se extiende la aplicación de los *GPT* al tratamiento de secuencias de imágenes y, más concretamente, al seguimiento de objetos y análisis del movimiento. Debido al carácter dinámico de estas redes, se modeliza la evolución de los objetos como las modificaciones que sufre la red a lo largo de toda la secuencia, entendiendo dichas modificaciones como variaciones en la topología de la red de interconexión. La modelización del movimiento como la dinámica de las neuronas, evita el problema de correspondencia, uno de los más costosos en la mayoría de técnicas de seguimiento de objetos. Para comprobar sus prestaciones, se aplica al reconocimiento de gestos de la mano.

Aunque no ha sido destacado especialmente en la memoria, este modelo permite la caracterización y seguimiento de múltiples objetos presentes en la escena. Asimismo, se puede realizar el seguimiento de un objeto que se disgregue en varias partes o que se fusione.

La propuesta es suficientemente generalista para dar cobertura a técnicas de tratamiento de una amplia diversidad de problemas de visión; si bien, por cuestiones de extensión del trabajo, la investigación se ha concretado en establecer la base conceptual sobre la que después se materializarán las arquitecturas.

ABSTRACT

The research performed approaches the problems of tracking objects and motion analysis in scenes. This work belongs to the research project "Sistema de visión para navegación autónoma" (CICYT TAP98-0333-C03-03). The main goal of this project is the development of an artificial vision device for the aid to autonomous navigation.

Since we are considering to integrate the vision processing module into the device, it is necessary to give to this one enough computational power to process the images at video rates. This fact establishes the necessity to design high performance architectures, with capacity of answer refinement of the answer based on the availability of time, specialized for vision problems; with miniaturization and empotrability approach. For a better understanding of the problem and how to approach it, we have reviewed a great deal of related works, considering the different approaches of active vision.

We have proposed the use of self-organizing neural networks to represent the objects, due to its quality of topology preservation, which provides the Topology Preserving Graph (*GPT*). This one has a high expressive capacity, it is simple to obtain and it is robust. We contributed, also, a synthesis of synthesis of the objects from this representation. Features that are extracted from the *GPT* simplify the later operations of classification and recognition, avoiding the high complexity of comparisons between

graphs. As an example of the application of the characterization of objects, we present the problem of shape classification from its contour.

The model of characterization of the objects allows to refine the quality of the representation based on the time available for its calculation, so that it will be the basis for the design of high performance realtime vision architectures.

We extend the application of the *GPT* to the processing of image sequences and, particularly, to tracking objects and motion analysis. Due to the dynamic character of these networks, we model the evolution of the objects as the changes that the network suffers throughout the sequence, understanding these changes as variations in the topology of the interconnection network. Modelling motion as neuron dynamics avoids correspondence problem, one of the most expensive in tracking techniques. In order to verify the model, we apply it to the problem of hand gesture recognition.

Although it has not been specially outstanding in this document, this model allows the characterization and tracking of multiple objects in the scene. Also, it can perform the tracking of an object that disgregates in several parts or that merges.

The proposal is generalist enough to cover processing techniques of a wide diversity of vision problems; although, by work extension, the research has been focused in establishing the conceptual basis on which the architectures will be developed.

CONTENIDO

AGRADECIMIENTOS	VII
RESUMEN	IX
ABSTRACT	XI
CONTENIDO	XIII
INDICE DE FIGURAS	XV
INDICE DE TABLAS	XIX
INTRODUCCIÓN: OBJETIVOS Y ESTADO DEL ARTE	21
1.1 MOTIVACIÓN Y OBJETIVOS	22
1.2 ESTADO DEL ARTE	26
1.3 PROPUESTA DE SOLUCIÓN.....	40
MODELOS NEURONALES DE REPRESENTACIÓN	45
2.1 REDES AUTO-ORGANIZATIVAS	46
2.2 GASES NEURONALES.....	50
2.3 PRESERVACIÓN DE LA TOPOLOGÍA.....	57
2.4 INCORPORACIÓN DE RESTRICCIONES DE TIEMPO REAL.....	61
2.5 CONCLUSIONES.....	69

REPRESENTACIÓN Y CARACTERIZACIÓN DE OBJETOS	71
3.1 REPRESENTACIÓN DE UN OBJETO 2D CON UN GAS NEURONAL	72
3.2 APLICACIÓN AL ANÁLISIS DE IMÁGENES	75
3.3 APLICACIÓN A LA SÍNTESIS DE IMÁGENES	92
3.4 EJEMPLO DE APLICACIÓN: CLASIFICACIÓN DE OBJETOS	94
3.5 CONCLUSIONES.....	99
SEGUIMIENTO DE OBJETOS Y ANÁLISIS DEL MOVIMIENTO.....	103
4.1 SEGUIMIENTO DE OBJETOS.....	104
4.2 ANÁLISIS DEL MOVIMIENTO.....	108
4.3 EJEMPLO DE APLICACIÓN: RECONOCIMIENTO DE GESTOS	111
4.4 CONCLUSIONES.....	124
CONCLUSIONES	127
5.1 APORTACIONES	128
5.2 LÍNEAS DE CONTINUACIÓN	130
ESTUDIO DE LA PRESERVACIÓN DE LA TOPOLOGÍA.....	133
REFERENCIAS	143

INDICE DE FIGURAS

Figura 1.1.	Arquitecturas en malla.	31
Figura 2.1.	Inserción de neuronas durante el aprendizaje de la Growing Cell Structure.	49
Figura 2.2.	División de la red GCS en varias subredes al eliminar neuronas.	50
Figura 2.3.	Estados inicial, intermedio y final del aprendizaje de la Neural Gas.	51
Figura 2.4.	Estados inicial, intermedio y final del aprendizaje de la Growing Neural Gas.	53
Figura 2.5.	Triangulación de Delaunay inducida	58
Figura 2.6.	Adaptación de diversos modelos auto-organizativos a un espacio de entrada.	59
Figura 2.7.	Comparación del error de cuantización entre mapas de Kohonen y Neural Gas.	60
Figura 2.8.	Adaptación incompleta de la Neural Gas.	61
Figura 2.9.	Adaptación de la Growing Neural Gas con límites de tiempo.	64
Figura 2.10.	Adaptación de la Growing Neural Gas modificando el número de patrones por iteración λ .	65
Figura 2.11.	Adaptación de la Growing Neural Gas modificando el número de neuronas insertadas por iteración.	67

Figura 2.12.	Adaptación de la Growing Neural Gas completando la red una vez finalizado el tiempo.	68
Figura 3.1.	Descripción global del sistema de obtención del Grafo Preservador de la Topología de un objeto.	74
Figura 3.2.	Diferentes adaptaciones del gas neuronal a un mismo objeto.	74
Figura 3.3.	Extracción del GPT del contorno de un objeto.	76
Figura 3.4.	Obtención de los diversos contornos de un objeto.	76
Figura 3.5.	Curvaturas de los contornos.	79
Figura 3.6.	Simplificación del GPT_{∇} mediante la detección de esquinas.	80
Figura 3.7.	Simplificación del GPT mediante la obtención de aristas.	80
Figura 3.8.	Firmas.	81
Figura 3.9.	Códigos de cadena.	81
Figura 3.10.	Grafos Preservadores de la Topología de diversos objetos bidimensionales.	83
Figura 3.11.	Representación de varios objetos.	83
Figura 3.12.	Subgrafos del GPT .	86
Figura 3.13.	Histograma de vecindad.	88
Figura 3.14.	Contorno del GPT .	91
Figura 3.15.	Diagrama de Voronoi del contorno del GPT .	92
Figura 3.16.	Esqueletos.	92
Figura 3.17.	Síntesis a partir de un GPT .	94
Figura 3.18.	Sistema de clasificación de formas.	95
Figura 3.19.	Obtención y simplificación del grafo del contorno.	96
Figura 3.20.	Clasificación de los objetos a partir del GPT .	98
Figura 3.21.	Clasificación de los objetos según su firma.	100
Figura 3.22.	Clasificación de los objetos según su curvatura.	101
Figura 4.1.	Etapas del problema del análisis del movimiento.	104
Figura 4.2.	Posición del GPT antes de la predicción.	107
Figura 4.3.	Grados de libertad de la mano.	112

Figura 4.4.	Localizaciones típicas de los puntos de interés de la mano.	115
Figura 4.5.	Sistema de reconocimiento de gestos.	116
Figura 4.6.	Adaptación de la Growing Neural Gas a diferentes posturas de la mano.	117
Figura 4.7.	Extracción del radio estándar de la palma.	118
Figura 4.8.	Adaptación dinámica de las Growing Neural Gas a un gesto completo.	119
Figura 4.9.	Normalización de las trayectorias del gesto.	121
Figura 4.10.	Posturas de la mano inicial y final de cada uno de los gestos.	123
Figura A.1.	Vecindades en el espacio de entrada y en la red.	134
Figura A.2.	Espacios de entrada.	136
Figura A.3.	Distancia geodésica.	137

INDICE DE TABLAS

Tabla 2.1.	Tiempos de aprendizaje de los diferentes modelos auto-organizativos.	60
Tabla 2.2.	Cálculo del producto topográfico de Neural Gas con distintas t_{max} .	62
Tabla 2.3.	Cálculo del producto topográfico de Growing Neural Gas con restricciones temporales.	63
Tabla 2.4.	Cálculo del producto topográfico de Growing Neural Gas (variación 1).	65
Tabla 2.5.	Cálculo del producto topográfico de Growing Neural Gas insertando diferente número de neuronas (variación 2).	66
Tabla 2.6.	Cálculo del producto topográfico de Growing Neural Gas completando la red una vez finalizado el aprendizaje (variación 3).	68
Tabla 3.1.	Longitud, diámetro y longitud normalizada de los contornos.	78
Tabla 3.2.	Momentos de los contornos.	82
Tabla 3.3.	Información extraída del número y longitud de las aristas.	87
Tabla 3.4.	Información extraída de las vecindades.	88
Tabla 3.5.	Momentos del <i>GPT</i> .	90
Tabla 4.1.	Tasa de acierto en el reconocimiento.	124

Tabla A.1.	Cálculo de la preservación de la topología (100 neuronas).	138
Tabla A.2.	Tiempos de aprendizaje de los diferentes modelos auto-organizativos.	139
Tabla A.3.	Preservación de la topología de la Neural Gas deteniendo su aprendizaje a los 12 segundos.	140
Tabla A.4.	Preservación de la topología de la Neural Gas modificando sus parámetros para que finalice a los 12 segundos.	141