

MÉTODOS DE CORRESPONDENCIA DE NODOS. APLICACIÓN A LA TÉCNICA EGM PARA RECONOCIMIENTO DE OBJETOS

Pablo A. Martínez Ruiz, Miguel Pinzolas Prado, Juan López Coronado
pablo.martinez@upct.es, miguel.pinzolas@upct.es, jl.coronado@upct.es

Departamento de Ingeniería de Sistemas y Automática (**Grupo Neurocor**).
Universidad Politécnica de Cartagena. C/ Doctor Fleming, s/n. 30202. Cartagena (Murcia).

RESUMEN

En este trabajo presentamos un sistema capaz de reconocer una serie de nodos o marcas en objetos presentes en la escena comparándolos con plantillas preasignadas en un modelo. Cada nodo forma una estructura de datos que permite realizar la comparación teniendo en cuenta diversas características intrínsecas (información de color, resultantes de aplicar determinados filtros, etc.). Analizamos dos variantes del método, una basada en filtros de Gabor y otro implantado mediante una técnica de Zooming, con la finalidad de obtener el porcentaje más elevado posible en la correspondencia entre nodos del modelo e imagen a analizar. Uno de los objetivos principales a tener en cuenta es un tiempo de procesamiento reducido, de forma que sea posible aplicarlo en sistemas que requieren una actuación en tiempo real.

*Este análisis de nodos es parte integrante de un método más completo denominado **Elastic Graph Matching (EGM)** cuya función principal es la de identificar el tipo de objeto presente en la escena y estimar su posición y orientación. Dichas estimaciones podrán permitir una adecuada manipulación de objetos mediante una pinza adaptada a un brazo articulado coordinado con un cabezal de visión estereoscópica. Esta plataforma se ha desarrollado y se encuentra totalmente operativa en el laboratorio del Grupo Neurocor de la UPCT.*

Palabras Clave: Elastic Graph Matching, filtros de Gabor, Zooming, reconocimiento de objetos.

1 INTRODUCCIÓN

El método propuesto parte de una descripción genérica del método EGM (Elastic Graph Matching) aplicada en el reconocimiento de gestos de una mano por un lado y de rostros por otro [5][6]. Debido a las

particularidades especiales de esta aplicación, se han analizado planteamientos diferentes en la metodología de correspondencia de nodos. Así mismo, se ha tenido en cuenta la necesidad de una más automatizada incorporación de nuevos modelos al sistema gracias a la autoubicación de nodos.

El proceso seguido por el algoritmo completo se puede resumir en los siguientes pasos:

1. Adquisición de la imagen de la escena conteniendo el objeto a identificar.
2. Preprocesamiento mediante filtro paso-bajo para eliminar ruido. Esta operación también se ha realizado previamente sobre las imágenes que contienen los modelos.
3. Segmentación de la imagen para una adecuada separación entre objeto y fondo de la escena. Se realiza teniendo en cuenta el color del objeto a localizar.
4. Ajuste del tamaño de la imagen que permita una comparación de igual a igual con respecto al tamaño del modelo.
5. Análisis y comparación del entorno de todos los píxeles de la imagen de entrada con cada uno de los nodos de los modelos. Elección del tipo de información pertinente en la comparación.
6. Unión de los nodos mediante enlaces. Estos enlaces vienen caracterizados por su longitud y por los nodos a los que conectan. Esta información es fundamental en el proceso descrito a continuación.
7. Análisis del grado de similitud entre la imagen y todos los modelos almacenados, y elección del objeto (modelo) que más se adecua a lo mostrado en la imagen.

En este trabajo se analizan dos métodos de correspondencia de nodos. Estos métodos se han

seleccionado atendiendo a su adecuación a los objetivos planteados y a la plataforma disponible.

El primero de ellos combina información de color con el filtrado de Gabor sobre la imagen en escala de grises de un entorno del nodo seleccionado. Las imágenes adquiridas, así como los modelos, se encuentran en el formato **YCrCb** [4]. Son tres capas de información en las que **Y** representa la imagen en escala de grises y las capas **Cr** y **Cb** ofrecen la información del color. Al seleccionar un objeto determinado a buscar (cubo, cilindro, etc.) automáticamente se selecciona un color asociado a él. Las capas **Cr** y **Cb** son procesadas mediante un filtro paso-banda que restringe la superficie a buscar en la imagen. Partiendo desde esta simplificación, se aplica un filtro detector de contornos de Gabor con diferentes orientaciones (cuatro) a un entorno del pixel en cuestión que se analiza. Estos mismos filtros se aplican a los modelos almacenados en cada uno de los nodos seleccionados de forma manual o automática.

El segundo método, para segmentar los objetos, utiliza la información de color exactamente igual a como lo hace el método anterior, pero en lugar de aplicar filtros de Gabor para analizar los bordes de los objetos, realiza un promediado de un conjunto de píxeles vecinos (técnica denominada **Zooming**) que permite agilizar enormemente el procesamiento de las imágenes. A partir de la imagen promediada se comparan los entornos de los nodos de la imagen objetivo y del modelo entre sí para computar la similitud.

Los algoritmos se han desarrollado en el entorno Matlab versión 5 ejecutándose sobre un ordenador PC con Pentium II 450 Mhz. En las diferentes imágenes obtenidas para realizar las pruebas se varía significativamente el ángulo de visión de los objetos que aparecen en ellas, de forma que sea viable hacer un análisis de la robustez de los algoritmos ante determinados cambios de perspectiva de dichos objetos.

2 ANÁLISIS DE MÉTODOS PARA CORRESPONDENCIA DE NODOS

En este trabajo hemos analizado dos alternativas para la comparación de nodos, indicando comparativamente porcentaje de éxito y velocidad de ejecución de ambas opciones.

2.1 GABOR + COLOR

La transformación de Gabor de una imagen [1] resulta de convolucionar dicha imagen con un filtro de Gabor. Se trata básicamente de una Gaussiana en dos dimensiones (con varianzas s_x and s_y a lo largo de los ejes x e y respectivamente) modulada por una senoide compleja (con frecuencias centrales U y V a lo largo de los ejes x e y respectivamente) descrita por la siguiente ecuación:

$$g(x, y) = \frac{1}{2\pi s_x s_y} \exp \left[-\frac{1}{2} \left\{ \left(\frac{x}{s_x} \right)^2 + \left(\frac{y}{s_y} \right)^2 \right\} + 2\pi j(Ux + Vy) \right]$$

El dibujo de las partes real e imaginaria de $g(x, y)$ se muestra en la figura 1.

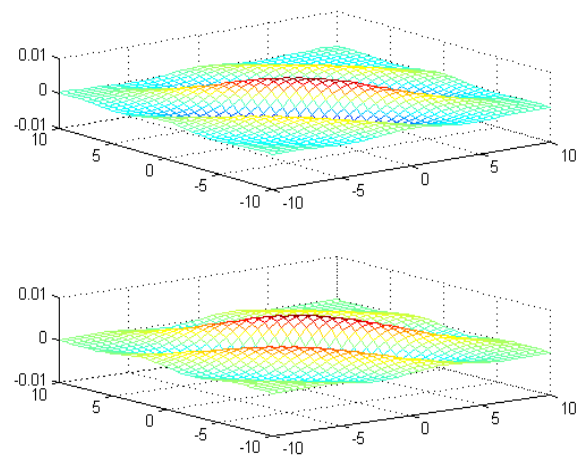


Figura 1: Partes Real (superior) e Imaginaria (inferior) de un filtro de Gabor.

El uso de filtros de Gabor de orden alto permite analizar texturas de la imagen [5]. Sin embargo, debido a las particularidades de nuestra aplicación (requisitos de velocidad de ejecución y predominio de bordes y esquinas en objetos que los describen completamente) hemos considerado más interesante aplicar filtros de primer orden que muestran los contornos de los objetos que son perpendiculares a la dirección del filtro. Así, básicamente actúa como detector de contornos en diferentes orientaciones, con lo que se simplifica el filtro de manera considerable y tiene el aspecto recogido en la figura 2.

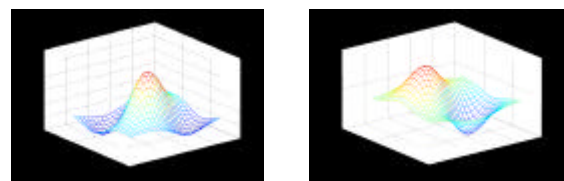


Figura 2: Izda.: Parte Real, Dcha.: parte Imaginaria

Se han utilizado cuatro orientaciones diferentes (45°, 90°, 135° y 180°) para cubrir todo el rango de posibles direcciones de los bordes de los objetos.

La utilización de este tipo de filtros es interesante por cuanto se ha comprobado que el sistema visual humano, en sus fases de más bajo nivel, lo aplica en el procesamiento de la información visual. Sus ventajas más notorias son:

- ⌘ No tiene término en continua, y por tanto sus respuestas son invariantes respecto de desplazamientos uniformes en el nivel de gris de la imagen.
- ⌘ El factor de ganancia compensa el espectro de potencia observado en las imágenes naturales.

El efecto obtenido al aplicar un filtro de Gabor sobre una imagen puede entenderse observando las gráficas de la figura 2. La componente Real presenta un máximo y dos mínimos en la dirección de su eje de orientación. El valor absoluto de este máximo es el mismo que la suma de los dos valores absolutos de los mínimos. Puede por tanto compararse esta parte Real con un filtro gradiente orientado en una dirección que, aplicado sobre una imagen, resalta los bordes de la imagen perpendiculares a la dirección del filtro. El efecto de este filtro se resume así: si el filtro se sitúa sobre una zona de intensidad homogénea, el resultado del filtrado será nulo. En caso contrario, quedarán resaltados los bordes perpendiculares a la dirección del filtro.

La aplicación de la parte imaginaria del filtro permite afinar el proceso por cuanto detecta el sentido de la transición. Sin embargo, el aumento de tiempo de proceso que supone hace no realizable su aplicación en este caso.

Esta aplicación de los filtros de Gabor se realiza sobre la imagen en escala de grises. Las capas de la imagen relacionadas con el color (Cr y Cb) nos permiten restringir la zona sobre la que aplicar todas las operaciones, descartando aquellos píxeles que no pertenezcan a un intervalo determinado en las capas de color. La no utilización de la información de intensidad para la segmentación hace que ésta sea robusta ante cambios de iluminación [4][5].

2.2 ZOOMING + COLOR

Una alternativa que hemos desarrollado al método anterior se basa en reducir la cantidad de información presente en la imagen y modelos mediante el promediado de un entorno del píxel analizado. Las imágenes son reducidas desde el tamaño inicial hasta un tamaño final de 128 x 128, con lo que el proceso de recorrer la imagen es mucho más rápido. Una vez realizada la reducción, hacemos la comparación de

un entorno del píxel elegido con los entornos de los nodos de los modelos, obteniendo un valor mediante el índice de error cuadrático medio que marcará la similitud entre ambos. Este método se aplica sobre las tres capas de la imagen, por haberse comprobado que se refuerzan entre sí.

La información de color y el método de Zooming, en este caso, están íntimamente relacionados entre sí, ya que no se procesarán, dentro de cada entorno, los píxeles que no pertenezcan a los intervalos de color predeterminados mediante los filtros paso-banda. Es decir, si en una máscara de 12 x 12 posicionada sobre una determinada zona de la imagen existen 3 píxeles que no tienen el color del objeto buscado, estos no se consideran para calcular el error cuadrático medio con los correspondientes del entorno del modelo.



Figura 3: Ubicación de nodos en imagen modelo



Figura 4: Imágenes A, B, C (oscuridad alta) y D (sobre fondo verde). De izquierda a derecha y de arriba abajo.

2.3 COMPARACIÓN DE RESULTADOS

Para el estudio de los métodos anteriores hemos partido de cinco imágenes obtenidas en el laboratorio con objetos típicos con formas regulares usados en manipulación. La escena es similar aunque se han cambiado el orden de los objetos, su orientación, la iluminación ambiental y el fondo existente para comprobar el grado de robustez de los algoritmos. Un aspecto a tener en cuenta es que se han comparado métodos entre sí utilizando imágenes obtenidas en diferentes situaciones. No es objetivo del presente

trabajo analizar el grado en el que afectan las variaciones mencionadas anteriormente al éxito del resultado para cada método.

Una de las imágenes se ha utilizado como modelo. En ella se han seleccionado 14 nodos, ubicados la mayoría en zonas de la periferia de los objetos (vértices y bordes). Las otras cuatro se han utilizado como imágenes de prueba a comparar con el modelo.

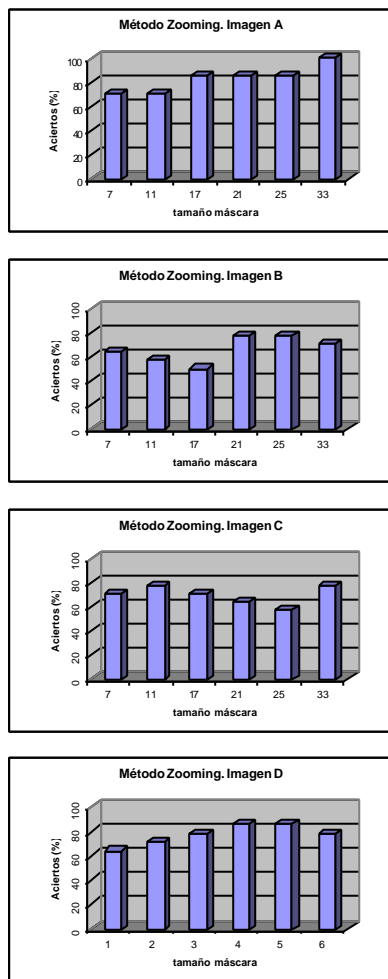


Figura 5: Resultados método Zooming utilizando imágenes **A**, **B**, **C** y **D** (de arriba a abajo). Porcentaje de aciertos frente a tamaño de máscara.

La imagen modelo y la ubicación de los 14 nodos sobre ella se muestra en la figura 3. Las cuatro imágenes de prueba del algoritmo se muestran en la figura 4. Las dos últimas han sido tomadas en condiciones especiales (luminosidad y fondo).

Al realizar la comparación mediante Zooming + Color del modelo con las imágenes de prueba, utilizando diversos tamaños de máscara, obtenemos los resultados representados en la figura 5.

Según se observa en el diagrama de barras correspondiente a la imagen **A** de la figura 5, al aumentar el tamaño de la máscara, aumenta progresivamente el porcentaje de aciertos en el proceso de *matching* de nodos, hasta alcanzar el 100 % de eficacia para máscaras de 33x33. Sin embargo, en las imágenes **B** y **C** no ocurre eso, sino que muestra un comportamiento aparentemente errático que no permite sacar la conclusión válida para la imagen **A**.

Los resultados expuestos aparentan estar en contradicción. Sin embargo, hay una razón por la que en las imágenes **B** y **C**, al aumentar el tamaño de la máscara no siempre aumenta la fiabilidad del método. Esto se debe principalmente a que en la comparación entre dos imágenes con mucha similitud, cuanto mayor sea la máscara menos influyen pequeñas variaciones puntuales, las cuales quedan mitigadas por el resto de la comparación. Sin embargo, si existe una orientación de los objetos significativamente diferente (como ocurre entre el modelo y las imágenes mencionadas), al aumentar el tamaño de la máscara es posible que magnifiquemos la diferencia en orientación de los bordes de los objetos, cosa que no ocurriría si la comparación fuese más local. Sin embargo, como se observa en la gráficas de barras perteneciente a las imágenes **B** y **C** de la figura 5, no existe una regla definida y la evolución sigue un comportamiento variable. Esto es debido a las particularidades de los objetos en cuestión. En este caso particular, los mejores resultados se consiguen con tamaño de máscara de 21x21 para la imagen **B** (78 % de aciertos) y 11x11 para la imagen **C** (78 % de aciertos).

A continuación hacemos el mismo estudio utilizando una combinación de *Zooming* con Gabor. Previamente a filtrar las imágenes, se reducen estas a un tamaño fijo de 128 x 128 píxeles. A partir de este tamaño se aplica el método de Gabor + Color. La razón de experimentar con esta combinación reside en el hecho de que el *Zooming* permite reducir el tiempo de proceso de forma notable al no tener que aplicar filtros de Gabor sobre la imagen completa original. En la figura 6 se muestra el porcentaje de aciertos en la localización. Se observa que mediante este método, el porcentaje de aciertos se reduce enormemente, a pesar de que el comportamiento es más lineal en cuanto a relación aciertos/tamaño máscara en las cuatro imágenes.

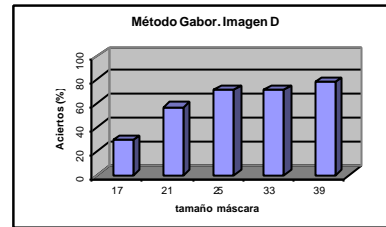
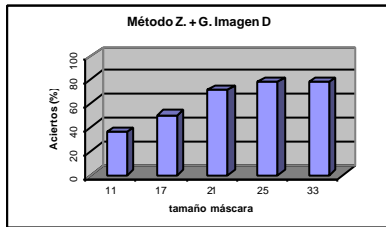
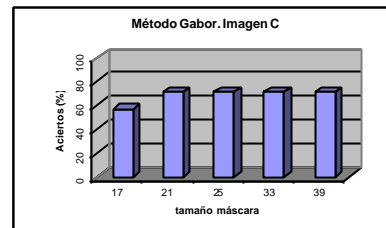
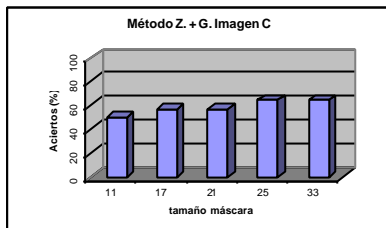
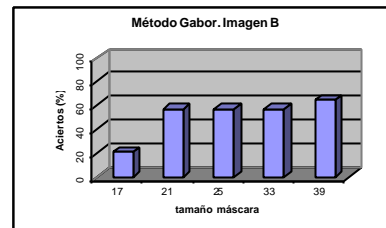
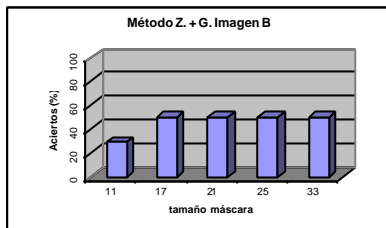
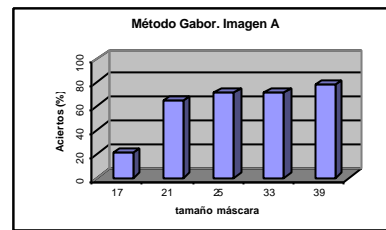
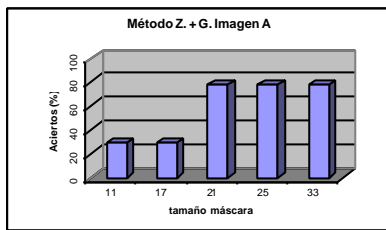


Figura 6: Resultados método Zooming + Gabor utilizando imágenes **A**, **B**, **C** y **D** (de arriba a abajo). Porcentaje de aciertos frente a tamaño de máscara.

Figura 7: Resultados método Gabor utilizando imágenes **A**, **B**, **C** y **D** (de arriba a abajo). Porcentaje de aciertos frente a tamaño de máscara.

Por último, aplicamos el método de Gabor + Color, el cual es el que se suele aplicar de forma estándar en EGM, sin reducir previamente las imágenes. Esto produce dos resultados remarcables. Por un lado, el tiempo de ejecución se dispara; por otro, la eficacia no aumenta. Para que el número de aciertos aumente ligeramente es necesario aumentar el tamaño de máscara significativamente, por lo que resulta un método inviable para aplicaciones exigentes en cuanto a velocidad de respuesta. Nuevamente se deduce de las gráficas de la figura 7 que al aumentar el tamaño de la máscara, siempre aumenta o como mínimo se mantiene el número de aciertos con las cuatro imágenes **A**, **B**, **C** y **D**.

En la tabla 1 se puede comparar la velocidad de ejecución de los diferentes métodos expuestos en el presente trabajo para los mejores resultados posibles en cada uno de ellos.

Los métodos de Zooming y Zooming + Gabor muestran una velocidad de ejecución similar, tan sólo se observa que Zooming resulta un 10% más lento que Zooming + Gabor.

Sin embargo, sí se ha encontrado una gran diferencia con respecto a Gabor, que ha tardado 9 veces más en ejecutarse que cualquiera de los otros dos anteriores. Este inconveniente se une a la relativamente baja tasa de aciertos (sólo ha mostrado un 78 % de aciertos para el mejor caso, con el tamaño de máscara más apropiado y utilizando la imagen que más se parece al modelo), comparable a la del método Zooming + Gabor pero sensiblemente inferior a la de Zooming (donde se llega a alcanzar el 100 % de aciertos).

En las cuatro imágenes utilizadas de prueba se han obtenido los mismos resultados cualitativos, por lo que no se ha considerado necesario ampliar el conjunto de imágenes. También se ha puesto de manifiesto que la variación del fondo de la imagen no altera la eficacia de los diferentes métodos, ya que los objetos son segmentados en base a su crominancia y el resto es desechado en el procesamiento. Por lo que cualquier otro fondo complejo, siempre que no coincida exactamente con el color de algún objeto a identificar, será ignorado por el algoritmo.

Método	Núm. de nodos acertados		Tpo. de ejecución	Calidad del método
	Mejores resultados	Resultados medios		
Zooming	14 de 14 (100 %)	11 de 14 (78 %)	110	Muy buena
Z. + G.	11 de 14 (78 %)	9 de 14 (64 %)	Índice de referencia (100)	Buena
Gabor	11 de 14 (78 %)	9 de 14 (64 %)	930	Regular

Tabla 1: Comparativa de resultados

3 CONCLUSIONES Y FUTUROS TRABAJOS

Después de analizar los 3 métodos anteriores, se puede concluir que la máxima eficacia y eficiencia se obtiene aplicando Zooming + Color, no sólo teniendo en cuenta el número de correspondencias correctas (que en algunos casos llega hasta el 100%) sino también por el tiempo de cómputo necesario, el cual es sensiblemente inferior al método de Gabor + Color, que puede llegar a tardar nueve veces más en el procesamiento. Los resultados de este último método podrían mejorarse si incorporáramos la información de fase del filtro de Gabor y si utilizáramos filtros de Gabor de mayor orden. Sin embargo, esto se traduciría directamente en un aumento del tiempo de procesamiento inviable para el tipo de aplicación para la que se requieren estos algoritmos. Es interesante remarcar que en todas las imágenes de prueba se ha obtenido el mismo orden de eficacia para los métodos probados.

La simplificación que supone partir de imágenes reducidas para aplicar el método de Gabor no ha dado resultados comparables a Zooming, pues aunque el tiempo de procesamiento se asemeja al de Zooming, la eficacia es menor. Cuando se comparan los dos métodos en los que intervienen filtros de Gabor, resulta interesante la simplificación de información inicial que supone reducir las imágenes a un tamaño fijo más pequeño que el original, lo cual permite un notable ahorro de tiempo de procesado sin perder precisión.

Este estudio realizado forma la base de la aplicación de una herramienta de reconocimiento de objetos denominado Elastic Graph Matching (EGM), en la que los nodos pertenecientes a un objeto forman el esqueleto a partir del cual es posible identificar dicha figura analizando la posición relativa de los nodos y la información contenida en ellos. Es una herramienta eficaz que permite el reconocimiento con un amplio margen de disimilitud entre modelo e

imágenes de prueba, lo que caracteriza su robustez. Uno de los puntos donde interesa una especial dedicación es en la ubicación de nodos en el modelo, de manera que pueda realizarse de modo automático prescindiendo del operador humano al añadir un nuevo modelo. Hasta ahora hemos colocado los nodos de forma manual. La transformada de Hough [3] es una herramienta interesante para conseguir este propósito de forma automatizada. Las ventajas de este método están relacionadas con su facilidad y robustez para detectar líneas rectas incluso en imágenes con ruido, así como con la capacidad para discriminar frente a formas no reales. Una vez obtenidas las líneas que definen el objeto, se calcula la intersección de ellas entre sí para determinar los picos de los objetos donde es interesante la ubicación de un nodo por contener información relevante.

Debido a las necesidades de funcionamiento en tiempo real del sistema, será necesario optimizar el algoritmo e implantarlo en un lenguaje de programación compilado (Visual C++) que ofrezca unas prestaciones adecuadas. Una de las actuaciones más significativas será la incorporación de una tarjeta de procesamiento de imágenes (convoluciones) en el propio ordenador, lo que disminuirá de manera significativa el tiempo de proceso de las imágenes y permitirá dedicar la potencia del ordenador en el resto de operaciones.

Agradecimientos

Nos gustaría agradecer el apoyo recibido del resto de los miembros del Grupo de Neurotecnología, Control y Robótica (NEUROCOR) del Departamento de Ingeniería de Sistemas y Automática de la Universidad Politécnica de Cartagena (UPCT). Así mismo, queremos resaltar que la investigación realizada ha sido financiada en parte con fondos del Proyecto Europeo de Investigación SYNERAGH (BRE2-CT980797 BRITE EURAM- BASIC Research).

Referencias

- [1] Gabor D., (1946) Theory of Communication, J. IEE, Londres, 93:429-457.
- [2] González, R. C., Woods R. E., (1994) Digital Image Processing, Addison-Wesley Longman.
- [3] Jain R., Kasturi R., Schunck B. G., (1995) Machine Vision, McGRAW-HILL INTERNATIONAL EDITIONS.
- [4] Poynton Ch., (1999) Frequently Asked Questions about color, www.inforamp.net/poynton.

- [5] Triesch, J., (1999) Vision-Based Robotic Gesture Recognition, Tesis Doctoral, Universidad de Bochum.
- [6] Wiskott L., Fellous J.-M., Krüger N., Malsburg C., (1997) Face Recognition by Elastic Bunch Graph Matching, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):775-779.