

SISTEMA DE RECONOCIMIENTO DE OBJETOS SOBRE VISION FOVEAL MEDIANTE REALIDAD VIRTUAL



*José Luis Herreros Díaz, Cristina Urdiales García,
Antonio Bandera Rubio*

*Dpto. Tecnología Electrónica, E.T.S.I. Telecomunicación,
Universidad de Málaga, Campus de Teatinos, 29071 Málaga (España)*

Actualmente un número creciente de investigadores ha centrado sus esfuerzos en la robótica, con el objetivo de desarrollar sistemas móviles capaces de desenvolverse en entornos desconocidos, cambiantes, y potencialmente hostiles sin necesidad de la intervención directa de un operador humano. Así, cada vez aparecen más aplicaciones de este tipo, como el PIONEER, desarrollado para la limpieza de los restos de Chernobyl, el DANTE II, para el descenso al volcán Mt. Spurr o el SOUJOURNER, explorador de Marte.

Uno de los principales problemas de estos sistemas es la necesidad de que se comporten al menos de forma parcialmente autónoma, ya que las condiciones en que desarrollan su trabajo imposibilitan un telecontrol exhaustivo. Para trabajar de esta forma, un agente precisa un gran número de módulos trabajando de forma coordinada. Por citar sólo algunos ejemplos, podrían citarse el módulo para evitar colisiones, el módulo para calcular la posición actual o el módulo para desplazarse a un punto determinado. Un problema adicional radica en el hecho de que, incluso aunque el agente fuese completamente autónomo, en la mayoría de los casos se requiere una comunicación con un sistema remoto para enviar los datos recogidos del entorno en que se opera. Podría pensarse en almacenar localmente dichos datos en el robot para recogerlos cuando éste se recupere, pero entonces la pérdida del agente, bastante habitual en estos casos, supondría también la pérdida completa de la información. Habitualmente se recurre a la comunicación vía radio de los datos conforme éstos se van adquiriendo. Dichos datos pueden recogerse a partir de diversos sensores, como táctiles, sonar o infrarrojos, pero indudablemente la visión sigue siendo una de las fuentes más ricas de información para el ser humano.

La mayor desventaja de transmitir con imagen digital vía radio es el enorme volumen de información que debe enviarse a través de un medio ruidoso de ancho de banda limitado, lo que reduce drásticamente el número de imágenes transmitidas por segundo y supone pérdida de información en tanto que el vehículo se va desplazando y dejando atrás escenarios que podrían no haber sido capturados. Para solucionar este problema, existen técnicas de compresión que permiten reducir notablemente el volumen de información

de una secuencia de vídeo, como por ejemplo el estándar MPEG. Sin embargo, la mayoría de estas técnicas implican una pérdida de calidad generalizada que podría imposibilitar el procesado de dichas imágenes si los algoritmos de trabajo requieren un alto nivel de detalle. Una segunda alternativa radica en el uso de visión atencional con campos de resolución no uniforme, como se presenta a veces en el reino animal en ejemplares como las águilas o algunos peces. Este tipo de visión consiste en mantener continuamente un área pequeña de muy alta resolución, conocida como fovea, sobre las áreas de interés de la imagen, mientras que el resto de ésta se presenta con un perfil de resolución decreciente en torno a ésta (Tistarelli y Sandini, 1993) (Camacho et al, 1996). En este caso es necesario definir una estrategia para calcular las áreas de interés de la imagen de forma automática o bien posibilitar su definición desde el extremo receptor, pero a cambio se obtienen reducciones muy considerables del volumen de datos sin perder resolución en dichas áreas.

Este trabajo se centra en el desarrollo de un sistema de control para el módulo de visión de un agente de este tipo, prestando especial atención al desarrollo de una interfaz eficaz hombre-máquina que permita una utilización sencilla e intuitiva de la herramienta. El agente presenta un módulo de navegación que reside en un ordenador remoto (Urdiales, 1999). La comunicación entre el ordenador de a bordo y dicha máquina se lleva a cabo mediante un enlace radio y consiste en una secuencia de lecturas de sensores sonar, infrarrojos y de contacto en dirección al módulo de navegación, que devuelve secuencias de comandos a los motores del agente. Al mismo tiempo, el agente está dotado de un módulo de visión binocular que captura mediante un frame-grabber imágenes de resolución homogénea en escala de grises. Estas imágenes deben transmitirse a una máquina cualquiera a través del enlace radio para que el teleoperador pueda observar el entorno sobre el que se encuentra el robot. Para acelerar la comunicación vía radio, dichas imágenes se fovealizan previamente a su envío. Como aplicación en el sistema remoto se ha implementado un reconocedor de objetos, que requiere que el objeto a analizar se presente a alta resolución. Para ello, el teleoperador debe establecer el área de interés en cada momento y

enviar las órdenes pertinentes para arrancar el proceso de reconocimiento. Asimismo, si lo desea, podrá nombrar el objeto reconocido. Como interfaz de usuario, siguiendo las últimas tendencias en este tipo de aplicaciones, se ha recurrido a periféricos de realidad virtual, que permiten que el teleoperador se relacione de forma natural e intuitiva con el software de control.

El conjunto se puede dividir en tres bloques muy diferenciados:

- * Extremo robot, incluyendo los motores, sensores, cámaras y un PC para preprocesado de la información.

- * Vía de comunicación, consistente en un enlace radio Ethernet de bajo ancho de banda.

- * Extremo remoto, donde se reciben las imágenes y las órdenes del teleoperador y se lleva a cabo el proceso de reconocimiento.

Existiría un cuarto bloque, la máquina donde corre el módulo sensomotor del agente, pero no se va a contemplar en este trabajo en tanto que no está relacionado con el ciclo de visión.

1.- ARQUITECTURA FÍSICA

El extremo robot consiste en un agente autónomo móvil modelo Nomad 200 capacitado, entre otras muchas tareas, para la captación de imágenes. El sistema de captura se compone de dos cámaras Hitachi 450 y una placa de adquisición tipo Sensus 700 que permite, si se desea, tratamiento de imágenes de resolución homogénea tanto en color como en escala de grises. La tarjeta capturadora admite hasta un total de 4 DSPs tipo C4X trabajando en paralelo, aunque la arquitectura actual sólo incorpora uno de ellos. El trabajar con un único C44 imposibilita la visión estéreo real, ya que no pueden capturarse dos imágenes simultáneamente.

Además del equipamiento anterior, el Nomad 200 cuenta con un PC tipo Pentium a 133 MHz. El DSP se comunica con el PC mediante un bus ISA de 1 Mbyte/seg de ancho de banda. Dicho PC es el encargado de gestionar todo el sistema autónomo, aunque los sensores se preprocesan mediante microcontroladores tipo HC11 de Motorola para restar carga al sistema.

Para comunicarse con el exterior, el agente hace uso de un enlace radio tipo Ethernet cuya velocidad de transmisión media para el protocolo TCP/IP utilizado es de aproximadamente 70 Kbytes/seg. El otro extremo del enlace está conectado a una red local y, mediante ésta, el agente puede comunicarse con cualquier ordenador conectado a dicha red.

En este caso, el extremo remoto soporte para el procesamiento de las imágenes es un Pentium II a 333

MHz, que cuenta con cuatro puertos serie para poder añadir los periféricos de realidad virtual necesarios, que, en este caso, es un guante tipo 5th DATA GLOBE desarrollado por 5th Dimensions Technologies. Este guante de licra es adaptable a cualquier tamaño y consta de dos tipos de sensores:

- De orientación: mediante un mecanismo interno es capaz de cuantificar el giro del guante. Hay dos tipos de medida: *pitch* (grado de inclinación del brazo) y *roll* (giro lateral de la muñeca), que se representan en la Fig. 1.

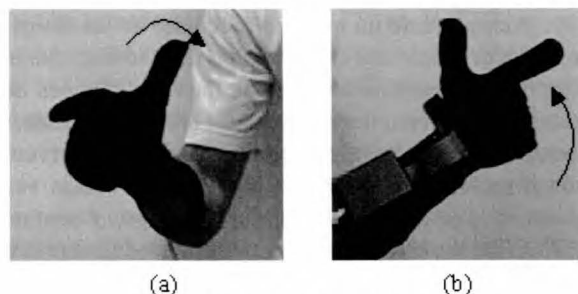


Fig. 1.- a) Roll, y b) Pitch

- De flexión: se usan cables de fibra óptica que se prolongan por cada uno de los dedos. Conforme se flexiona cada uno de ellos, por las características de la fibra, varía el índice de reflexión. A ambos extremos de la fibra se incorporan un fotoemisor y un fotorreceptor respectivamente para medir qué porcentaje de la luz transmitida es recibida, lo que depende directamente del índice de reflexión y, por tanto, del grado de flexión en cada dedo.

La información de ambos tipos de sensores se codifica con 8 bits de resolución por sensor, lo que supone 256 bytes en total, y se envía por un interfaz serie RS-232 a la estación de trabajo encargada de procesar la información. De los 7 sensores disponibles se transmiten hasta 200 muestras por segundo.

Al emplear este tipo de periféricos es recomendable usar gestos simples para disparar las distintas acciones, teniendo en cuenta la flexión o extensión de cada uno de los dedos mediante el uso umbrales. En este trabajo se han desarrollado una serie de mecanismos para perfeccionar estas medidas evitando, en la medida de lo posible, los transitorios asociados a una acción cualquiera. El primero de dichos mecanismos es la implementación de un ciclo de histéresis con el fin de evitar la confusión de distintos tipos de gestos. La consecuencia directa de esto es que para extender un dedo, no bastaría con relajar el miembro, sino que se debe extender completamente. Para flexionarlo es necesario operar de forma similar. Otro mecanismo de mejora a este respecto ha sido la introducción del método matemático «media alfa-recortada», que elimina de las medidas los valores extremos y efectúa la media de los restantes. De esta manera se filtran los datos espúreos que falsean las mediciones.

Debido a la simplicidad del método de medida, diseñado para dar velocidad a proceso de toma de muestras, es necesaria una fase de inicialización que va a ser determinante, ya que a raíz de ésta se fijan los ejes imaginarios de coordenadas sobre los que trabajan el usuario y el sistema.

2.- ESQUEMA GENERAL DEL SISTEMA

Tal como se ha comentado previamente, el sistema se ha dividido en una serie de bloques interconectados. El primero de éstos, que es denominado *extremo robot*, realiza la captación de imágenes, un primer tratamiento de los datos y su empaquetamiento para la transmisión. La *vía de comunicación* comunica el extremo robot y el remoto. Se encarga de transmitir los paquetes de datos hacia el PC, donde serán tratados, y además envía los parámetros de control hacia el robot. La comunicación entre bloques se establece a través de dos sockets: uno para la transmisión de datos y otro para el envío de los parámetros de control. El *extremo remoto* captura la información, monta las imágenes recibidas, crea la estructura de datos necesaria para su posterior estudio, efectúa un tratamiento de éstas y, finalmente, presenta los resultados. Dentro del extremo remoto cabe mencionar varios módulos:

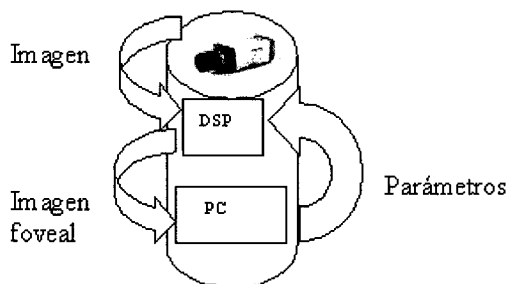


Fig. 2. Esquema de captura de imágenes en el robot

- Tratamiento de la imagen y representación gráfica: incluye en montaje de la imagen, la creación de la estructura de datos y su presentación en pantalla.

- Segmentación y detección de objetos: supone un laborioso tratamiento en la estructura de datos para depurar la información, de manera que se llegue finalmente a la detección de los objetos existentes en la imagen captada.

- Reconocimiento: partiendo de la detección anterior, se conoce el color de cada objeto, lo que permite encajarlo en la zona de máxima resolución de la imagen para analizarlo en detalle y reconocerlo a partir de su contorno.

- Aprendizaje: es la parte encargada de almacenar los datos referentes al reconocimiento e identifica-

ción de los objetos en memoria no volátil. Si se desea, puede asignársele una etiqueta con su nombre. En cualquier caso, la siguiente ocasión en que aparezca un objeto previamente memorizado, éste será reconocido.

- Interfaz guante: es el módulo encargado de captar los gestos y posturas que se adoptan con el guante de realidad virtual e interpretarlos, dotando de significado cada movimiento.

3.- DESARROLLO DEL SISTEMA

3.1. Extremo robot

I. Captura de imágenes.

Para capturar imágenes el robot hace uso del sistema que se muestra en la Fig. 2. El DSP HECFG44 (versión del TMS320C40 de Texas Instruments) proporciona tratamiento de vídeo. Aunque dicho DSP está capacitado además para el tratamiento de imagen, éste se efectúa en el PC encargado del control ya que numerosas pruebas mostraron que encargar de dicha tarea al DSP ralentizaba el proceso global. En la captura, el DSP actúa como esclavo del PC, que ordena secuencialmente la captura de imágenes.

La comunicación PC-DSP utiliza un protocolo de comunicaciones llamado *RPC* y, dentro de éste, el modo *call-back*, que permite la ejecución paralela de dos procesos. De esta forma puede llevarse a cabo simultáneamente la captura de la imagen en el DSP, el tratamiento de la misma en el PC y la transmisión de los datos preprocesados. La consecuencia directa de este enfoque fue la duplicación de la velocidad en esta parte del proceso.

II. Tratamiento de la imagen

Para transmitir una secuencia de imágenes a través de un enlace de ancho de banda reducido, tal como se comentó anteriormente, es necesario comprimirlas. La compresión utilizada consiste principalmente en la fovealización y el empaquetamiento de los datos.

La fovealización es la conversión de una imagen de resolución uniforme en una multirresolución. Se basa en definir un área de máximo nivel de detalle denominada fovea en la zona de interés, mientras que la resolución que va disminuyendo progresivamente al alejarse de ésta. De esta forma, se reduce el volumen de datos de la imagen manteniendo constante el campo de visión y obteniendo un nivel de resolución alto en las áreas importantes. La visión foveal presenta dos geometrías básicas:

- Geometría Logaritmo Polar (Fig. 3.a) (Tistarelli y Sandini, 1993). Similar al mecanismo de visión de

algunos animales. El muestreo que se efectúa es «no uniforme» y los campos receptivos están ubicados en círculos concéntricos. Otra característica es que los sensores no son de tamaño homogéneo. Este mecanismo no se ha empleado en el trabajo actual debido a la necesidad de algoritmos de procesamiento muy específicos para adecuarse a las estructuras de datos resultantes de un muestreo de este tipo. Además, en la Fig. 3.a puede observarse la existencia de zonas muertas entre el cuadrado de la fovea y el primer círculo de elementos sensores.

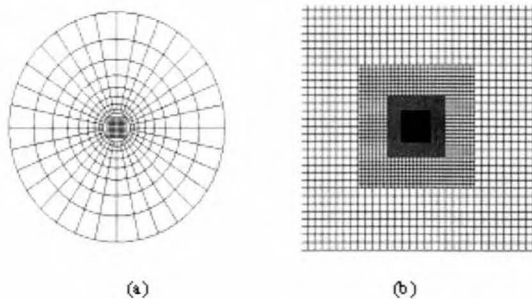


Fig 3. Geometrías foveales: a) Log-Polar; b) cartesiano-exponencial

- Geometría cartesiano-exponencial (Fig. 3.b) (Camacho et al, 1996). Esta geometría presenta discontinuidades en su perfil de resolución. El tamaño de los campos receptivos es siempre un número entero en relación con el tamaño de los sensores de la fovea.

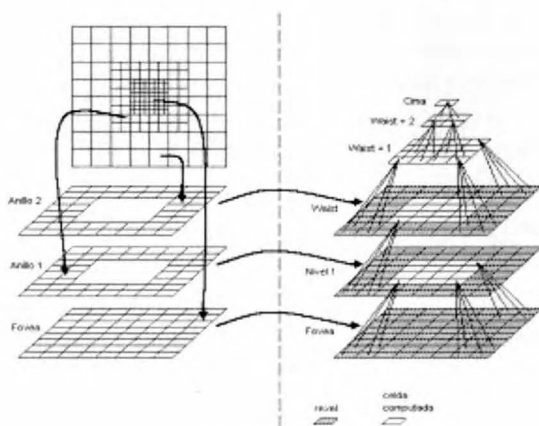


Fig. 4. Polígono foveal

Dado que los algoritmos de visión para imagen uniforme son fácilmente adaptables a este tipo de imágenes, es la que finalmente se empleará en el sistema implementado.

La estructura de datos 3D que permite manejar la geometría cartesiano-exponencial se denomina *polígono foveal* (Fig. 4) y es una estructura compuesta por capas, donde la primera se corresponde con la fovea y el resto con la combinación de cada nivel de resolución con la capa previa (Bandera, 1994) (Camacho et al, 1996).

Para que dicha capa presente la misma resolución que el anillo de trabajo y, por tanto, pueda combinarse con éste en un nivel de resolución uniforme, la resolución de ésta puede reducirse mediante diezmo, tomando una muestra aleatoriamente de cada conjunto de 2x2 celdas, promediado, llevando a cabo la media aritmética de cada conjunto de 2x2 celdas, o mediante filtros complejos, pero esta última opción ralentiza considerablemente el proceso si se efectúa mediante software. Al existir distintos niveles de resolución en el polígono foveal, se puede trabajar en el más adecuado para cada tarea. Así, para la detección de objetos interesa una capa alta de la estructura de datos, para trabajar con el menor número de celdas posibles. Sin embargo, el reconocimiento, que requiere un buen nivel de detalle, se llevará a cabo en la fovea.

A pesar de las ventajas mencionadas, la geometría cartesiano-exponencial clásica requeriría movimientos de cámara de alta precisión para centrar el área de máxima resolución en el punto de interés. Dado que serían lentos y costosos, se ha optado por utilizar variaciones de la geometría clásica denominadas *geometrías multirresolución de fovea desplazable*, evaluándose tres tipos (Fig. 5) (Arrebola, 1998):

- Básicas: donde el desplazamiento entre niveles constante, así como el número de celdas. La estructura de datos es el *polígono foveal asimétrico*, que es una versión del polígono foveal donde la región de celdas computadas no tiene una ubicación fija (ni centrada como en el caso previo).

- De movimiento generalizado: en este caso el desplazamiento entre niveles de resolución sucesivos no tiene que ser igual. Cada nivel se desplazará los réxeles (denominación que se le da a una celda en las capas de la estructura de datos) necesarios para obte-

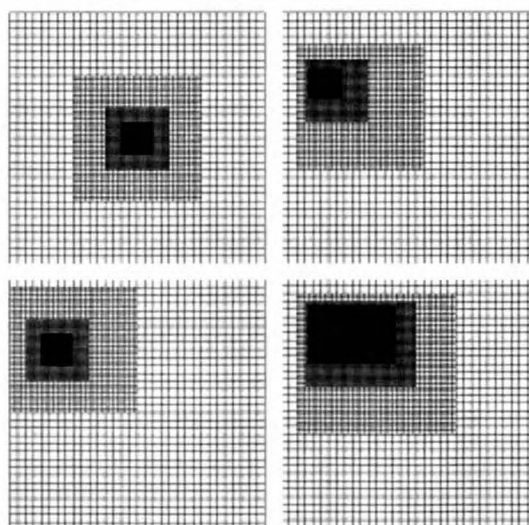


Fig. 5. Geometrías foveales cartesiano-exponenciales: a) de fovea centrada; b) básicas; c) de movimiento generalizado; d) de tamaño adaptativo

ner la fovea en la posición deseada. El tamaño de cada nivel de resolución es fijo.

- De fovea de tamaño adaptativo: aquí el tamaño de la fovea y, por tanto, del resto de niveles, puede variar y adaptarse al objeto de interés de la escena. Una vez se han determinado las dimensiones de cada nivel de resolución, el desplazamiento entre estos es constante.

Esta última geometría es la empleada finalmente. La motivación esencial es que la fovea se adapta al objeto a estudiar mejor que en los casos anteriores, con lo que el volumen de datos resultante queda optimizado para cada objeto.

3.2 Transmisión de imágenes

El agente autónomo móvil requiere un interfaz radio para comunicarse sin que su movilidad se encuentre limitada por un cable. El sistema de transmisión aéreo disponible es un enlace radio Ethernet que, tras montar sobre él un protocolo TCP/IP de comunicación, trabaja a una velocidad de aproximadamente 70 KBytes/seg, lo que obliga a reducir al máximo la carga de datos a transmitir para tener una tasa de imágenes por segundo aceptable.

El protocolo de transmisión empleado es de tipo *cliente-servidor*, donde el robot actúa de servidor de imágenes para cualquier usuario que se le conecte, mientras que el PC remoto es el cliente que captura los datos. Así, el servidor envía la imagen fovealizada, y el cliente la captura remitiéndole los parámetros de la próxima ubicación de la fovea, que el usuario dispondrá mediante el guante de realidad virtual sobre el área que le interese.

3.3. Extremo remoto

I. Segmentación y detección

Una vez recibida la información procedente del extremo robot y antes de procesarla, el receptor debe montar la estructura de datos denominada *polígono foveal asimétrico*, que no es más que una adaptación del *polígono foveal* (Fig. 4) para poder trabajar con la geometría multiresolución de fovea desplazable y tamaño adaptativo.

La segmentación es un proceso de partición de una imagen en un conjunto de regiones no solapadas cuya unión es la imagen completa. La detección es la determinación de qué regiones de la segmentación corresponden a objetos físicos existentes en la escena. Para ser detectadas como objetos, las regiones han de presentar características homogéneas frente a una aplicación de alto nivel (Haralick y Shapiro, 1992). Una labor esencial en esta parte del procesado es la identificación y etiquetado de todas y cada una de las regiones disponibles.

Dado que la segmentación de imágenes reales es un proceso muy complejo, para llevarlo a cabo se han estudiado diversas alternativas:

- Segmentación por mezclado: consiste en recorrer la imagen píxel a píxel y comparar el valor de brillo de cada punto con el de las regiones existentes. Caso de parecerse a alguna de éstas, se engloba en la región oportuna y en caso contrario, se genera una nueva. Presenta el problema de que no se tiene en cuenta la conectividad entre regiones (Pitas, 1993).

- División y mezclado: sigue una filosofía descendente (*top-down*) y parte de la premisa de que toda imagen es homogénea. Caso de no ser así, la imagen se divide en cuatro subimágenes. El estudio se repite recursivamente dentro de cada subimagen hasta obtener regiones uniformes. Una vez concluido este proceso comienza la fase de mezclado, ya que habrán surgido una infinidad de clases, la mayoría pertenecientes a unas pocas. Se unen aquellas clases en contacto cuyo color sea lo suficientemente parecido. Su mayor desventaja es una excesiva lentitud (Pitas, 1993).

- Segmentación jerárquica multiclase: que ha sido finalmente seleccionado por adaptarse perfectamente a una estructura del tipo del polígono foveal y porque es el que mayor robustez presenta para imágenes captadas en entornos reales. Es un algoritmo iterativo aplicado al polígono foveal asimétrico que se basa en estabilizar esta estructura nivel a nivel mediante el principio de enlace adaptativo. Así, cada celda del nivel $K-1$ busca la más similar entre las 2×2 celdas situadas inmediatamente sobre ella en el nivel K , denominadas padres. Cuando ha finalizado la este proceso, cada celda padre toma el nivel de gris promedio de las celdas hijas que han enlazado con ella. La estabilización continúa mientras existan celdas padre cuyo nivel de gris cambie (Arrebola, 1998).

Una vez se ha estabilizado toda la estructura de datos, tiene lugar la *generación de clases*. Para ello se elige un nivel que se denomina de propagación. Cada nodo de dicho nivel está asociado a un conjunto de celdas de la base a través de los enlaces establecidos entre niveles. La estabilización adaptativa garantiza que dichos nodos componen una región homogénea, ya que se ha obtenido la agrupación progresiva de los nodos cercanos más similares. Adicionalmente, el proceso aporta un suavizado generalizado y un realce de bordes.

Para localizar las regiones susceptibles de ser un objeto, se calcula su *bounding-box*, que es rectángulo mínimo que engloba a la región a estudiar. Una vez se tienen las *bounding-boxes*, para evaluar potenciales objetos se emplea un criterio de compacidad, que comprueba que las *bounding-boxes* encontradas tie-

nen un área, como máximo, tres veces mayor que la celda asociada en el nivel de propagación. Este criterio permite calcular todas manchas de la imagen que destacan sobre un fondo. Una vez localizadas todas las posibles manchas, es necesario un tratamiento para depurar las detecciones y dejar únicamente las correspondientes a objetos realmente existentes. Se llevan a cabo cuatro mecanismos de depuración:

a) Test de compacidad. Si se detecta más de una clase, se mide el área de cada una. Si el porcentaje de ocupación de la celda es mayor del 80 % la clase se toma como válida y el objeto es detectado. Si no es así se descarta.

b) Fusión de raíces. Se comprueba si alguna raíz potencial es susceptible de unirse con otra (bounding-box solapadas, niveles de gris similares y celdas hijas han de formar una región conectada). Si cumple las condiciones adecuadas se unen en una sola.

c) Test de conectividad. Se recorren las celdas secuencialmente, y las que tengan un nivel de gris similar al del objeto se etiquetan con un número. Pasadas sucesivas asignan la misma etiqueta a las celdas adyacentes. La etiqueta más numerosa es la que se toma como válida.

d) Mecanismo de contraste. Se estudia si el objeto se distingue de su entorno más cercano. Para ello se compara el nivel de gris del objeto con el fondo.

Con todos estos mecanismos se logra tener localizados los posibles objetos englobados en la imagen tratada, listos para el siguiente tratamiento.

II. Reconocimiento.

La detección sirve para señalar posibles objetivos a estudiar, facilitando la labor de búsqueda y, por tanto, economizando recursos. Gracias a este proceso, se obtienen unas zonas señaladas como potenciales objetos. Debido a que el reconocimiento se va a efectuar con exclusividad en la fóvea, el proceso anterior permite orientar dicha zona de máxima resolución, de forma que se centre en un área ya marcada. Todo esto se conoce como *mecanismo atencional* (Bandera et al, 2000). Llegado este punto, el teleoperador es quien decide la zona a estudiar. Para ello hay dos posibilidades:

1. Señalar una zona marcada, lo cual requiere tener activada la detección automática de objetos.
2. Reconocer una zona, independientemente de las detecciones.

Una vez localizada la zona a estudiar tiene lugar la *caracterización* de la imagen contenida en la fóvea. Para ello es imprescindible que el sistema haya detectado un objeto, y que éste se encuentre totalmente

contenido en la zona de máxima resolución. En este procesado se efectúan las operaciones que a continuación se detallan.

En primer lugar se lleva a cabo una binarización objeto-fondo, de manera que se marca con diferente nivel de gris cada una de esas partes. A continuación se aplican unos algoritmos de limpieza para depurar el proceso anterior:

- Dilatación.- Elimina el ruido interior del objeto.

- Erosión.- Es el algoritmo complementario a la dilatación. Aquí se trata de eliminar los puntos externos que, teniendo un color similar al objeto, se encuentran fuera del mismo.

Con todo esto se logra una compactación de colores por zonas y un suavizado del borde. A continuación tiene lugar una nueva segmentación por niveles de gris, de manera que se marca con otro color el contorno del objeto. Finalmente tiene lugar la eliminación de aristas de la periferia.

Tras limpiar y preparar la imagen se lleva a efecto el estudio de la misma, que se inicia con el cálculo del *código cadena* del contorno del objeto (Freeman, 1977). Este código cadena se usa para calcular la curvatura del contorno. Para ello, se emplea el algoritmo de código cadena incremental de longitud variable, que asigna a cada punto del contorno una dirección que permitiría seguir la silueta (Bandera et al, 2000b).

III. Reconocimiento

La función de curvatura obtenida se compara con otros vectores predeterminados que componen una base, la cual ha sido extraída del estudio de unos modelos prefijados. A través de la correlación de la función de curvatura con cada vector de la base, se obtiene un vector de características. Con estos datos se efectúa la clasificación utilizando distancias euclídeas y un umbral: si la distancia el vector obtenido y los previamente almacenados supera dicho umbral se crea una nueva clase. En caso contrario se añade a la clase ya creada más parecida.

IV. Interfaz de usuario

El usuario tiene en la pantalla el interfaz que se presenta en la Fig. 6. En la esquina superior izquierda de la pantalla se puede ver la imagen recibida en tiempo real y a su derecha la versión ya segmentada de ésta, con las detecciones positivas englobadas en su *bounding-box*. La imagen superior derecha corresponde a la fóvea que se capta en cada momento.

La interfaz incluye un guante de realidad virtual que, mediante un conjunto de gestos, permite llevar a cabo las siguientes funciones:

- Desplazamiento de un puntero, que aparece como un punto naranja en la Fig. 6, por la pantalla para seleccionar distintas regiones de la imagen o pulsar botones de la interfaz.

- Centrado de la fovea en la imagen captada.

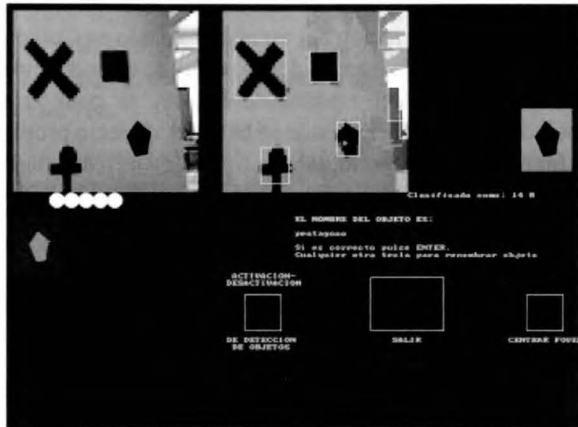


Fig. 6. Pantalla de la interfaz de usuario

- Activación/desactivación de la detección automática de objetos activando con el puntero los botones correspondientes.

- Posicionamiento de la fovea sobre un objeto cualquiera. Si la detección es automática, basta desplazar el cursor hasta cualquiera de las cajas detectadas y seleccionarla. En caso contrario, hay que arrastrar la fovea hasta el área deseada y redimensionarla manualmente.

- Solicitud del reconocimiento del objeto contenido en la zona de máxima resolución. En este caso, aparece una imagen en la esquina inferior izquierda, correspondiente a la caracterización de la fovea: binarización, segmentación y seguimiento del contorno para su reconocimiento. Una vez se ha completado este ciclo con éxito se obtiene el resultado de la clasificación e identificación, donde se indica el nombre que se le asigna a dicho objeto.

4.- CONCLUSIONES

En este trabajo se presenta la implementación de un sistema de reconocimiento supervisado completo para un agente autónomo móvil que se desenvuelve en entornos cerrados y capta imágenes mediante una cámara analógica y un *frame grabber*. Las principales novedades que presenta el sistema descrito son: i) el empleo de un algoritmo realimentado de detección de objetos, que permite descartar de forma efectiva el

fondo de la imagen, y extraer correctamente los objetos de la imagen; ii) la integración de los algoritmos que tratan con imágenes multirresolución en un sistema complejo de transmisión y compresión de imágenes; y iii) la efectiva caracterización de objetos para su posterior reconocimiento.

REFERENCIAS

(Tistarelli y Sandini, 1993) Tistarelli, M. y Sandini, G., «On advantage of polar and log-polar mapping for direct estimation of time-to-impact from optical flow», *IEEE Trans. on Pattern Analysis and Machine Intell.*, 15 (4), pp. 401-410, 1993.

(Camacho et al, 1996) Camacho, P., Arrebola, F. y Sandoval, F., «Shifted fovea multiresolution geometries», *IEEE Int. Conf. on Image Processing (ICIP'96)*, 1, pp. 307-310, Laussana-Suiza, 1996.

(Urdiales, 1999) Urdiales, C., *Arquitectura de control de movimiento y exploración para un agente autónomo*, Tesis Doctoral, Dpto. Tecnología Electrónica, Universidad de Málaga, Málaga-España, 1999.

(Bandera, 1994) Bandera, C., *Structures and algorithms for foveal machine vision*, Amherst Systems, Technical Report, Buffalo-EEUU, 1994.

(Arrebola, 1998) Arrebola, F., *Sistemas de visión basado en imágenes multirresolución de fovea desplazable*, Tesis Doctoral, Dpto. Tecnología Electrónica, Universidad de Málaga, Málaga-España, 1998.

(Haralick y Shapiro, 1992) Haralick, R.M. y Shapiro, L.G., *Computer and robot vision*, Addison-Wesley: New York, 1992.

(Pitas, 1993) Pitas, I., *Digital image processing algorithms*, Prentice Hall: New York, 1993.

(Bandera et al, 2000) Bandera, A., Urdiales, C., Herberos, J.L. y Sandoval, F., «Implementación de un mecanismo atencional sobre un agente autónomo móvil», *Actas del XV Simp. Nacional de la Unión Cient. Int. de Radio (URSI'00)*, pp. 323-324, Zaragoza-España, 2000.

(Freeman y Davis, 1977) Freeman, H. y Davis, L.S., «A corner finding algorithm for chain-coded curves», *IEEE Trans. on Computers*, C-26, pp. 297-303, 1977.

(Bandera et al, 2000b) Bandera, A., Urdiales, C., Arrebola, F. y Sandoval, F., «Corner detection by means of an adaptively estimated curvature function», *Electronics Letters*, 36 (2), pp. 124-126, 2000.