

DETECCIÓN DE OBJETOS POR SEGMENTACIÓN MULTINIVEL COMBINADA DE ESPACIOS DE COLOR

P. Gil, F. Torres, F.G. Ortiz
Grupo de Automática, Robótica y Visión Artificial.
Dpto. Física, Ingeniería de Sistemas y Teoría de la Señal.
Universidad de Alicante.
Crta. San Vicente s/n. A.P. de correos 99.
pgil@dfists.ua.es

Resumen

En visión artificial, los procesos de segmentación son de vital importancia cuando se trata de detectar objetos en entornos no estructurados. En concreto, la segmentación haciendo uso de información de color adquiere cierta importancia en ámbitos muy diversos, desde el tratamiento de secuencias de videos para la detección de objetos o individuos en escenas móviles hasta la detección para el seguimiento, también conocido como tracking, pasando por la detección de objetos para la manipulación mediante robots.

En este artículo, se presenta un método de detección de objetos haciendo uso de información de color, y tratando de reducir la influencia de sombras y brillos que aparecen formando parte de información cromática presente en las imágenes capturadas. Para ello, se hace uso de las representaciones RGB y HSV, y de una aproximación multiumbral del método de Otsu. El método de detección, que aquí se muestra, se compone de dos fases, por un lado la segmentación en regiones cromáticas y por otro lado la detección de los contornos de esas regiones.

Con este método se pretende, mostrar una primera aproximación en la detección de objetos de una escena para más adelante buscar la existencia o no de solapamiento entre ellos, y por lo tanto la detección de posibles oclusiones parciales entre ellos.

Palabras Clave: Segmentación, espacios de HSV-RGB, umbralización multinivel, detección de objetos.

1 INTRODUCCIÓN

El problema de la segmentación de imágenes ha sido tratado ampliamente en la literatura [10][12], y se han hecho muchas propuestas tanto para segmentación de imágenes en escala de grises

[10][12] como de imágenes en color [10][4][5][13]. También, conviene destacar que típicamente siempre se han hecho dos aproximaciones básicas para resolver el problema de la segmentación, una de ellas basada en la detección de bordes-contornos [6][7] y otra basada en la detección de regiones [4][5][13]. En los métodos basados en bordes se detectan discontinuidades locales y que más tarde se tratan de unir formando un borde que delimita unos objetos de otros. Y en los métodos basados en regiones, se busca determinar las áreas de una imagen que tienen propiedades homogéneas y el borde de estas delimitará unos objetos de otros. Los dos métodos, por lo tanto, son complementarios y dependen en gran medida de la aplicación para la que vayan a ser destinados.

La propuesta que aquí se hace, combina ambos métodos. En una primera fase, se segmenta buscando áreas homogéneas de similar cromaticidad. Y en segundo lugar, se detectan los contornos de las áreas segmentadas y se construyen los bordes para esas regiones detectadas. Con ello, se pretende, mejorar la fiabilidad del proceso de segmentación, evitando falsos contornos que puedan aparecer en la imagen. Cuando se habla de falsos contornos, se entienden como contornos no deseados por que no son representativos o no forman parte de un objeto visible en la imagen. Y por otro lado, únicamente, se detecten los contornos que si forman parte de un borde de objeto con la mayor robustez posible.

Así, en el apartado 2, se comenta cuales son los espacios de color utilizados sobre los que se aplica la técnica de segmentación empleada. En el apartado 3, se presenta la metodología de segmentación basada en los espacios de color comentados en 2 y además se muestran varios experimentos que demuestran la efectividad del método propuesto. Y, finalmente, en el apartado 4 como conclusiones, se comenta brevemente la ventaja de aplicar una detección de bordes sobre las imágenes segmentadas por regiones en función del método aplicado para segmentar.

2 REPRESENTACION DEL COLOR

Antes de abordar, las técnicas de segmentación de color empleadas, conviene destacar que la representación del color no es un problema trivial y que ésta influye notablemente en el modo en que se emplean y su eficacia.

2.1 ESPACIO RGB

El espacio RGB es el espacio de color más extendido y el que utilizan la gran mayoría de cámaras de video y fotográficas para construir una imagen de color. Y de ahí, su importancia en visión artificial ya que trabajar con el mismo espacio de color con el que trabaja la cámara con la que se capturan las imágenes permite evitar la alteración de las propiedades del color durante el proceso de segmentación, propia de los errores de conversión y transformación, y por otro lado conseguir una mayor velocidad de segmentación por ahorro de esas operaciones de conversión y redondeo.

El espacio RGB se representa como un cubo dónde un color viene definido por la mezcla de valores de intensidad de tres colores primarios, rojo, verde y azul. Un color viene descrito por una tupla de 3-coordenadas en el cubo. El color negro se representa por $(r=0, g=0, b=0)$ y el color blanco por $(r=255, g=255, b=255)$. La gama acromática de escala de grises está representada por la diagonal del cubo.

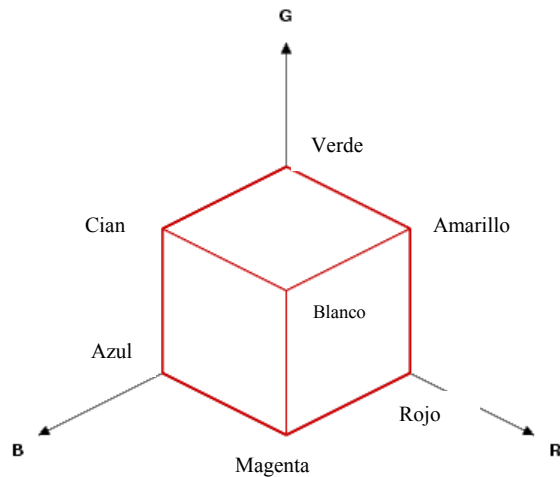


Figura 1: Representación del espacio RGB.

2.2 ESPACIO HSV

El espacio HSV [11] representa uno de los espacios de coordenadas más clásicos e intuitivos existentes en la literatura. Su interpretación geométrica viene determinada por un cono de base quasi-hexagonal. Con esta representación del espacio de color, cada

color trabaja con 3 componentes básicas: matiz, saturación y brillo. El matiz, h_{HSV} , hace referencia al valor de cromaticidad o clase de color. La saturación, s_{HSV} , se refiere a las longitudes de onda que se suman a la frecuencia del color, y determina la cantidad de blanco que contiene un color. Contra menos saturado este un color más cantidad de blanco, y contra más saturado este un color menor cantidad de blanco. En definitiva, la saturación representa la pureza e intensidad de un color. Así, la falta de saturación viene dada por la generatriz en la representación del cono HSV. Esa falta de saturación representa la gama de grises desde el blanco hasta el negro. La luminosancia, v_{HSV} , se corresponde con la apreciación subjetiva de claridad y oscuridad.

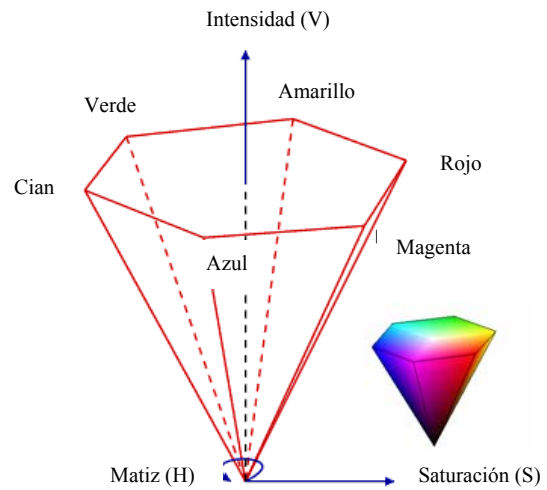


Figura 2: Representación del espacio HSV.

Cuando se quiere representar una imagen en color con un espacio de color HSV, es importante determinar como influyen las componentes de color de una cámara de video RGB sobre el espacio HSV. Así, el sistema HSV viene definido por:

$$h_{HSV} = \begin{cases} \frac{g-b}{\max(r, g, b) - \min(r, g, b)} & \text{si } r = \max(r, g, b) \\ \frac{b-r}{\max(r, g, b) - \min(r, g, b)} + 2 & \text{si } g = \max(r, g, b) \\ \frac{r-g}{\max(r, g, b) - \min(r, g, b)} + 4 & \text{si } b = \max(r, g, b) \end{cases} \quad (1)$$

$$s_{HSV} = \frac{\max(r, g, b) - \min(r, g, b)}{\max(r, g, b)}$$

$$v_{HSV} = \max(r, g, b)$$

Existen muchas variantes de espacios de color intuitivos, en función de cómo se modifique su representatividad, cabe destacar otros espacios como el HSI, HLS, y las variantes de éste último según su forma de construirlo [1][8].

El espacio de color HSV representa mejor que HLS la saturación, aunque tiene peor representación de la luminancia. El HLS viene a representarse como un doble cono donde los vértices determinan la máxima y mínima luminancia.

2.3 HISTOGRAMA BIDIMENSIONAL V/S.

El histograma bidimensional V/S, es un histograma de dos variables cartesianas, luminancias y saturaciones. Su finalidad, radica en que proporciona un mecanismo sencillo e intuitivo para separar la información cromática de la acromática. La evaluación de un histograma V/S y la fijación de umbrales sobre él, permiten conseguir la extracción de reflejo y de distintos grados de sombras [1].

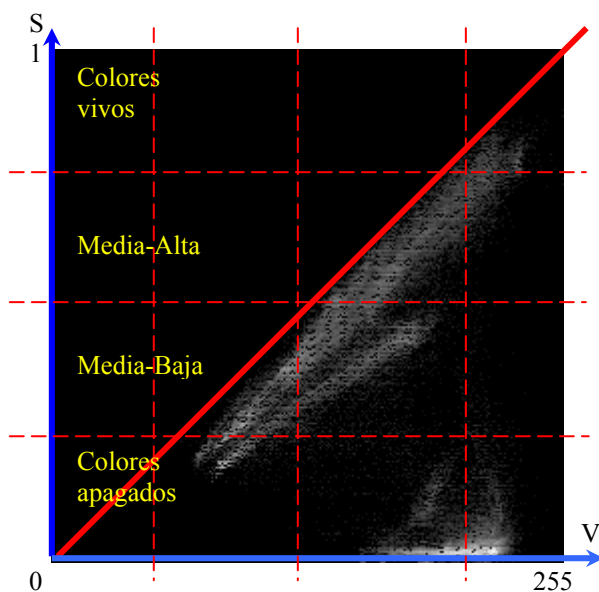


Figura 3: Histograma Bidimensional V/S.

3 METODO DE SEGMENTACIÓN.

La segmentación empleada tanto con RGB como con HSV se fundamenta en la detección de umbrales del mapa de componentes RGB y/o HSV. La umbralización de imágenes de objetos en entornos robotizados muchas veces es compleja por la falta de conocimiento a priori del número de objetos a detectar, por la influencia de elementos no deseados como sombras, brillos, la complejidad de colores, texturas y tamaño de los objetos y la posibilidad de solapamiento de estos, así como las variaciones en el fondo. Esto hace que la elección de un número bajo de umbrales influya en la subsegmentación, se detectan menos regiones de las deseadas, o la elección de un número elevado lo haga en una sobresegmentación, se detectan más regiones de las que interesa. Tanto una como otra, se puede deber a

dos motivos principalmente, la aparición de sombras y brillos por un lado, y por otro la similitud de los colores que tienen los objetos presentes en la escena, principalmente.

La umbralización de cada una de las componentes de color tanto RGB como HSV se ha llevado a cabo mediante la detección automática de multi-umbrales mediante el método de Otsu. Estimándose, el número mínimo de umbrales por componente en 1 y el número máximo de umbrales en 3. El hecho, de que así se haya hecho es porque al menos en toda componente espacial de color es posible separar objeto de fondo, y para ello es necesario 1 umbral. Y el número máximo de 3 viene determinado porque con 3 umbrales por componente se tienen un máximo de 64 posibles clases/región que se consideran más que suficientes para segmentar una imagen. El número de clases/región distintas viene determinado por:

$$nC = \prod_{k=1}^{nc} (nu_k + 1) \quad (2)$$

donde nc indica el número de capas en la imagen, que para las imágenes en color es siempre 3 y corresponde a las componentes básicas: rojo, verde y azul en RGB o matiz, saturación y luminancia en HSV. Y nu_k hace referencia al número de umbrales detectados para cada capa de la imagen.

Se ha optado por el método basado en información estadística y espacial de histogramas, como es el método de Otsu [9], fundamentado en distribuciones Gaussianas de partes del histograma. Y el cálculo de umbrales viene determinado por una aproximación para automatizar el método de Otsu [3].

3.1 UMBRALIZACIÓN MULTINIVEL

El método de Otsu multinivel permite discriminar los píxeles de una imagen en varias clases/región $C_0 = \{1, \dots, u_1\}, C_1 = \{u_1 + 1, \dots, u_2\}, \dots, C_n = \{u_n, \dots, L - 1\}$, donde L es el valor máximo de un píxel, y donde cada clase agrupa todos los píxeles hasta un umbral detectado por Otsu, siendo $U = \{u_1, u_2, \dots, u_n\}$ el conjunto de umbrales detectados. El modo en que Otsu selecciona los umbrales, U , para crear el conjunto de clases/región, se fundamenta en la maximización de la varianza entre clases/región, σ^2 , $U = \arg \max_{1 \leq u_1 < u_2 < \dots < u_n < L} \{\sigma^2\}$. Y la varianza entre clases se puede calcular como:

$$\sigma^2 = \sum_{k=1}^n \omega_k (\mu_k - \mu_T)^2 \quad (3)$$

donde:

$$\mu_k = \sum_{i \in C_k} i \cdot p_i / \omega_k \quad (4)$$

$$\omega_k = \sum_{i \in C_k} p_i, \omega_1 = \sum_{i=1}^{u_1} p_i, \dots, \omega_n = \sum_{i=u_{n+1}}^{L-1} p_i \quad (5)$$

$$\mu_T = \sum_{i \in C_k} \mu_k = \sum_{i=1}^{L-1} i \cdot p_i \quad (6)$$

siendo μ_K , la media de la clase C_K , ω_K la probabilidad de que los píxeles pertenezcan a la clase C_K , μ_T la media global de la imagen y p_i los píxeles con valor i .

3.2 SEGMENTACIÓN EN RGB

La segmentación de una imagen RGB se realiza segmentando individualmente cada una de las tres componentes de color de las que consta la imagen. Así, se aplica el método de Otsu comentando en el apartado anterior a cada una de las tres componentes de color [13], previamente suavizadas para provocar mejoras en el histograma que permitan obtener los umbrales con un mayor éxito. Se obtendrán, por lo tanto tres conjuntos de umbrales. Y de este modo, el número de clases/región detectadas en la imagen de color vendrá determinada por la ecuación (2). Así, ya se dispone de una segmentación de color inicial, donde se asignan valores de intensidad, $\tilde{p}_{ij,k}$, para cada componente, k , en función del número de umbrales obtenidos, nu_k .

$$\tilde{p}_{ij,k} = \begin{cases} 0 & \text{si } p_{ij,k} \leq u_{1,k} \\ \frac{L}{nu_k} \cdot n & \text{si } u_{n,k} < p_{ij,k} \leq u_{n+1,k}, n = 1 \dots nu_k \\ L-1 & \text{si } p_{ij,k} > u_{nu_k,k} \end{cases} \quad (7)$$

De este modo, si fusionamos cada una de las tres componentes de color de la imagen, previamente umbralizadas y posteriormente coloreadas de acuerdo al criterio de la ecuación (7), cada píxel habrá sido codificado como un código $C = \{c_1, c_2, c_3\}$ donde c_k puede tomar únicamente los valores de $\tilde{p}_{ij,k}$. Y cada una de estas posibles codificaciones, determinan una posible clase/región $\{C_k\}_{k=1}^{nC}$.

Una vez, se dispone de una imagen previamente segmentada, donde el número de regiones detectadas

vendrá dado por nc , se busca reducir su número que en el peor de los casos será de 64 usando 3 umbrales por componente, uniendo regiones, empleando una técnica de clusterización basada en el concepto de varianza entre clases de Otsu [13], al que se incorpora información de un número nC indeterminado de umbrales detectados.

3.3 SEGMENTACIÓN EN HSV

El fundamento empleado para segmentar imágenes es similar al usado para el espacio de color RGB. Así, se transforma inicialmente, la imagen RGB a HSV, y se separan las tres componentes de color: matiz, saturación y luminancia, para analizarlos y trabajar con ellos.

Haciendo una primera aproximación, para detectar objetos de color en una imagen en entornos robotizados, se ha partido de la suposición que la falta de color en las imágenes adquiridas hace referencia principalmente a dos situaciones básicas zonas de sombras y zonas de brillos en la imagen; información que facilita el histograma bidimensional V/S.

Es importante, sobre todo para las regiones de sombras que estas no sean añadidas como área de los objetos presentes en la imagen, y que por otro lado, no supongan que una región con distintas zonas de sombra sea segmentada como regiones distintas. De ahí, que el espacio de color HSV proporcione una información de falta de cromaticidad más intuitiva y más sencilla de segmentar, de la que proporciona el espacio RGB. Siguiendo un esquema similar al usado en la segmentación RGB.

El proceso que se ha seguido es el siguiente:

1. Transformar de RGB a HSV. Separar cada uno de los canales H, S y V y suavizarlos para eliminar discontinuidades.
2. Detectar los nu_k umbrales de Otsu para cada componente $k = \{H, S, V\}$.
3. Seleccionar los umbrales más adecuados para la segmentación cromática de entre los detectados por Otsu haciendo uso del histograma bidimensional V/S.
4. Codificar cada píxel con un código $C = \{c_1, c_2, c_3\}$, donde c_k toma valores de la ecuación (3).
5. A partir de las distintas clases/región en las que se ha segmentado, agruparlas buscando regiones homogéneas para reducir el número de regiones detectadas. Para ello, se aplica, únicamente, como criterio de clusterización medidas de varianza y de desviación típica respecto al valor medio de

cromaticidad dado por la componente polar de matiz H. Para cada región detectada en 4.

6. Colorear las nuevas regiones clusterizadas, buscando asignar como color, la media de valor de matiz de los píxeles de la imagen original H del paso 2.

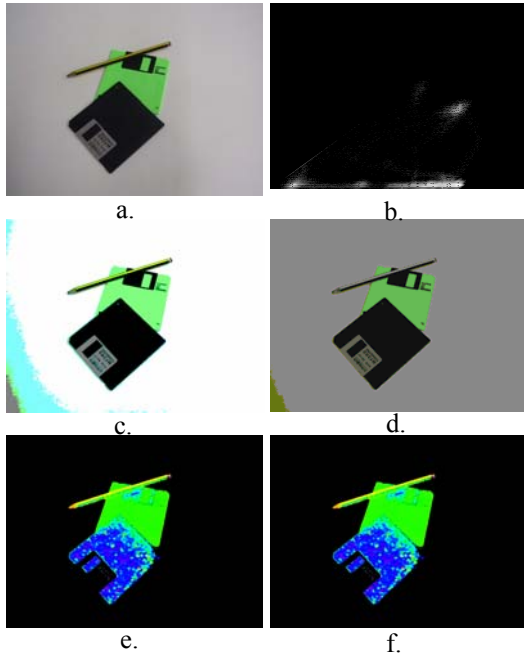


Figura 4. a). Original, b) Histograma V/S. c) Segmentación RGB, d) Clusterización RGB, e) Segmentación HSV, f) Clusterización HSV.

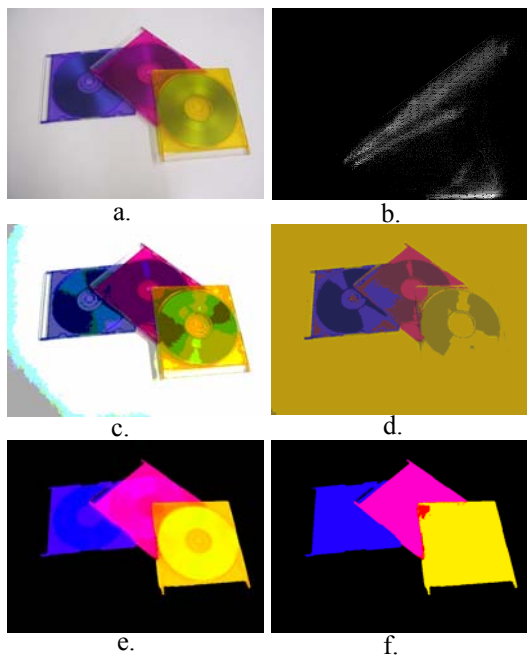


Figura 5. a). Original, b) Histograma V/S. c) Segmentación RGB, d) Clusterización RGB, e) Segmentación HSV, f) Clusterización HSV.

3.4 SEGMENTACION COMBINADA RGB-HSV

La ventaja que ofrece el método de segmentación multi-umbral RGB se basa principalmente en que es capaz de detectar información cromática con tonos de color muy diversos incluyendo los conocidos acromáticos (escala de grises). Esto permite detectar objetos cuyo color es una tonalidad de gris (Figura 4.a-5.a). Sin embargo, también, obliga a que se detecten como objetos regiones de sombras y brillos, que en esta segmentación serán considerados como objetos de un determinado color grisáceo (Figura 4.c-4.d,5.c-5.d).

Por otro lado, el método multi-umbral HSV mostrado, permite detectar objetos de color sin detectar por ello regiones de sombras y brillos (Figura 5.a) que al ser acromáticas, generalmente, son desechadas como objetos válidos (Figura 5.e). Como inconveniente, reseñar que un objeto en escala de grises no será detectado correctamente, al considerarse o bien un brillo o bien una sombra, observar la chapa metálica en las figuras 4.e y 4.f. Además, tiene el inconveniente que trabajar con las componentes desacopladas de HSV, hace que se pierda la generalidad de la información de color muy útil para la segmentación (Figura 4.e-4.f).

Si se compara la segmentación HSV de los objetos representados en la Figura 5, se observa que es mucho más precisa que la realizada para los objetos de la Figura 4. Una sencilla explicación se puede llevar a cabo realizando un pequeño análisis de los histogramas V/S (Figura 4.b-5.b), donde se comprueba que los objetos de la figura 4 destacan principalmente por su falta de saturación, dificultando la segmentación HSV, que viene delimitada por el umbral más pequeño detectado por Otsu para la componente S.

Vistos los problemas que acarrearán uno y otro método de segmentación, se ha optado por probar una combinación de ambos. De modo que inicialmente, se busca detectar los objetos presentes en la imagen haciendo uso de la segmentación RGB vista en el apartado 3.2. Y posteriormente, haciendo uso de algunas de las informaciones que proporcionan las componentes del espacio de color HSV de la imagen original, realizar una post-segmentación de la imagen segmentada mediante RGB. Así, se busca detectar todos los objetos en la imagen incluidos los que carecen de cromaticidad y por otro lado, eliminar en la medida de lo posible los brillos y las sombras, buscando conseguir homogeneización de las áreas detectadas.

Como se puede observar en la figura 6, el resultado de combinar ambos espacios de color, garantiza obtener colores más realistas y no tan artificiales

como se obtenían en el modelo HSV, cuando la clusterización se llevaba a cabo únicamente con la cromaticidad. En el modelo RGB-HSV la clusterización se lleva a cabo con las tres componentes de color del espacio RGB. De este modo se proporcionan colores menos sintéticos y más parecidos a los colores de la imagen sin segmentar. Por otro lado, el usar componentes del espacio HSV para corregir algunas situaciones de iluminación, brillos de fondo y leves sombreados, permite conseguir segmentaciones más precisas. Además la combinación de ambos espacios permite mejorar el proceso de clusterización, evitando la fusión de regiones no-homogéneas que podía ocurrir en la segmentación multinivel RGB (Figuras 4.d-5d).

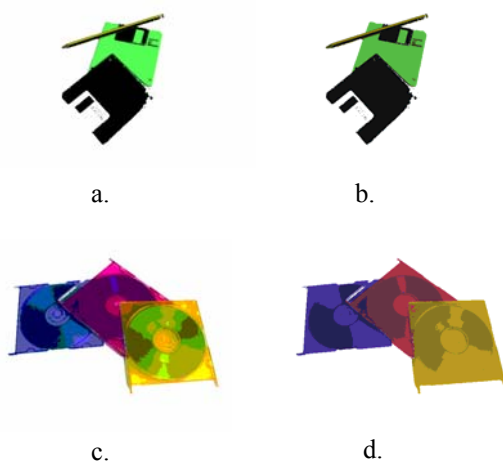


Figura 6: a) Segmentación combinada Figura 4a, b) Clusterización, c) Segmentación combinada Figura 5a, b) Clusterización,

4 CONCLUSIONES

Las ventajas que ofrece el método multinivel de segmentación combinada quedan de manifiesto como ya se comentó en el apartado 3. No obstante, una detección de bordes mediante el método de Canny [2], permite resaltar la eficacia del método cuando se trata de comprobar cuales son las regiones detectadas previamente. Así, en la figura 7, se puede comprobar que los contornos de los objetos segmentados por el método combinado son más precisos.

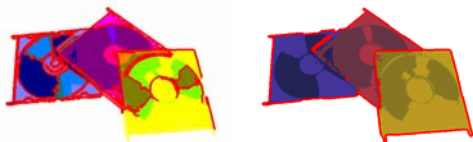


Figure 7: a) Detección de contornos de imagen segmentada RGB, b) Detección de contornos de imagen segmentada por combinación RGB-HSV.

Agradecimientos

Este trabajo ha sido financiado por el proyecto español MCYT, “DESAURO: Desensamblado Automático Selectivo para Reciclado mediante Robots Cooperativos y Sistema Multisensorial” (MCYT (DPI2002-02103).

Referencias

- [1] Angulo, J., (2003). “Morphologie mathématique at indexation d’images couleur. Application à la microscopie en biomedicine”. *Tesis doctor de la Ecole Des Mines de Paris*.
- [2] Canny, J.F., (1986) “A computational approach to edge detection”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol 8, No 6, pp. 112-131.
- [3] Cheriet, M., Said, J.N, Suen, C.Y., (1998) “A recursive thresholding technique for Image Segmentation”, *IEEE Transactions on Image Processing*, Vol. 7, No 6, pp 918-921.
- [4] Comaniciu, D., Meer, P., (1997). “Robust analysis of feature spaces: Color image segmentation”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp 750-755.
- [5] Deng Y., Manjunath, (2001). “Unsupervised Segmentation of Color Texture Regions in Images and Video”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 23, No 8, pp 800-810.
- [6] Fan J., David K., Yan Y., (2001). “Automatic Image Segmentation by Integrating Color-Edge Extraction and Seeded Region Growing”. *IEEE Transactions on Image Processing*. Vol 10, No 10, pp 1454-1466.
- [7] Kovese P., (2002). “Edges are not just steps”. *Proceedings of ACCV2002 The Fifth Asian Conference on Computer Vision*, Melbourne Jan 2002, pp 822-827.
- [8] Ortiz, F.G., (2002). “Procesamiento morfológico de imágenes en color. Aplicación a la reconstrucción geodésica”. *Tesis doctor de la Escuela Politécnica Superior. Universidad de Alicante*.
- [9] Otsu, N., (1978). “A threshold selection method from grey-level histograms”. *IEEE*

Transactions Systems man, Cybernetics, Vol. 8,
pp 62-66.

- [10] Pal, N.R, Pal, S., (1993). "A review on image segmentation techniques", *Pattern Recongnition*, Vol 26, pp 1277-1294.
- [11] Smith, A.R., (1978). "Color gamut transformation pairs". *Computer Graphics*, Vol 12, No 3, pp 12-19.
- [12] Trier, O.D., Jain, A.K., (1995). "Goal-directed evaluation of binarization methods". *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol 17, pp. 312-315.
- [13] Yingzi, D., Chein-I C., Thouin P.D., (2004). "Unsupervised approach to color video thresholding", *Optical Engineering*, Vol. 43, No 2, pp 282-289.