Selective Search Method for Object Localization and Detection using Wavelets and Hierarchical Segmentations

S. Cervantes and R. Pinto

***Abstract*— This article proposes a selective search method for object localization in natural images by applying image multi-segmentation, image scaling, and heuristics. The method increases the number of generated windows that delimitate the area of an object with an accuracy superior to 50%. Over-segmentation is applied on original size images in order to locate small objects, and it is also applied over scaled images because these can still be over-segmented. This process produces less regions on areas with many textures. The over-segmentation was applied using the CIE Luv color model, and using the *H* and the *I* channels of the HSI model. The proposed method is category independent and allows the location of objects with heterogeneous characteristics by using heuristics and hierarchical segmentation. The proposed method produces 9, 366 windows per image covering 96.78% of the objects in the PASCAL VOC 2007 test image collection, increasing in 0.8% the localization results reported in the state of the art.**

***Keywords*— selective search, object localization, object detection, hierarchical segmentation.**

I. INTRODUCCIÓN

LA DETECCIÓN de objetos de interés en imágenes y videos de escenas naturales permite mejorar los sistemas de control de calidad, vigilancia automatizada, sistemas de apoyo para el diagnóstico médico, entre otros. En los últimos años, el reconocimiento de objetos en imágenes de escenas naturales, ha sido un área de investigación en la cual se han realizado grandes avances, entre ellos, el desarrollo de nuevos descriptores como el HOG [1], la creación de métodos para detección de objetos basados en partes [1],[2],[3] y el desarrollo de algoritmos para la localización de objetos que mejoran la detección de estos al crear ventanas que delimitan con mayor precisión las regiones pertenecientes a los objetos [4],[5],[6],[7]. Los eventos internacionales IMAGENET [8] y PASCAL VOC [9] son actualmente puntos de referencia utilizados para comparar resultados en las áreas de clasificación de imágenes, segmentación semántica y reconocimiento de objetos; en dichos eventos se presentan los avances más recientes para el reconocimiento de objetos. De acuerdo a [8] y [9], los trabajos que han mostrado un mejor desempeño en el área de detección de objetos son [1], [2], [3], [10], [11], los cuales en su mayoría utilizan el método de búsqueda exhaustiva conocido como *ventanas deslizantes*. Las ventanas deslizantes generan ventanas de interés que pueden proporcionar soporte espacial (porcentaje del área de un objeto) y contribuyen a la obtención de buenos resultados en la localización de caras [12], personas [1] y vistas frontales o laterales de automóviles [13]. Una desventaja del método de ventanas deslizantes es que en categorías cuyos objetos se presentan en gran variedad de formas (perros, sillas, etc.), los objetos difícilmente pueden ser delimitados de forma correcta por las ventanas producidas al aplicar el método, ya que no toma en consideración las características de la imagen, sino las dimensiones del objeto de interés para establecer las dimensiones de las ventanas. Otra desventaja que presenta el método de ventanas deslizantes, es que cuando existe una gran variedad de formas entre los objetos de una categoría, se produce una gran cantidad de ventanas de interés cuyo objetivo es abarcar las diferentes dimensiones y escalas de los objetos. Con relación a esta desventaja, en los trabajos [5],[11],[14],[15] se consiguió reducir el número de ventanas de interés realizando una preselección de las mismas, sin embargo, dicha preselección ocasionó que la localización de objetos disminuyera y por lo tanto también la detección de éstos. Por otra parte en [4],[7],[14],[16] se han abordado diferentes enfoques de búsquedas que reducen y obtienen ventanas de interés con mejor soporte espacial tomando en consideración las características de la imagen que se procesa. Estos métodos reducen el tiempo necesario para la detección de objetos, sin embargo, el problema radica en que no existe un método que pueda localizar todos los objetos de interés [7], siendo que un objeto que es mal localizado o no es localizado, no puede ser detectado. La Figura 1 muestra la obtención de diferentes apoyos espaciales por medio de ventanas delimitadoras (recuadros azules) para los objetos de la categoría persona.

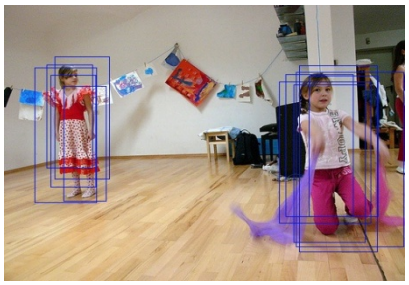

Figura 1. Múltiples ventanas de interés (recuadros en color azul) con diferente apoyo espacial para el reconocimiento de objetos de la categoría *persona*.

S. Cervantes, Centro Nacional de Investigación y Desarrollo Tecnológico, Morelos, México, scervantes@cenidet.edu.mx

R. Pinto, Nacional de Investigación y Desarrollo Tecnológico, Morelos, México, rpinto@cenidet.edu.mx

Debido a que la localización de objetos aún sigue siendo un problema de interés, en este artículo se propone un método de búsqueda selectiva para obtener localizaciones de objetos con diferentes apoyos espaciales, mediante el escalado de imágenes y el uso de heurísticas. Dicho método está basado en la utilización de múltiples segmentaciones que se aplican sobre imágenes escaladas con diferentes modelos de color, lo cual permite generar diversos conjuntos de regiones en una imagen. Las regiones obtenidas durante la segmentación, son descritas en base a su apariencia (textura y color). A través del uso de tres heurísticas se logra la combinación de regiones adyacentes para la generación de segmentaciones jerárquicas (ver Figura 2) que permiten localizar objetos compuestos por regiones con características heterogéneas. Los resultados muestran que nuestro método puede localizar hasta un 96.87% de los objetos en la colección de imágenes de prueba de PASCAL VOC 2007 superando los resultados del estado del arte reportados en [7].

El artículo está estructurado de la siguiente forma: en la Sección II se presenta la descripción de los trabajos relacionados con la localización de objetos, la Sección III presenta el método desarrollado en esta investigación para la localización de objetos en imágenes al cual denominamos método de Jerarquías de Regiones Escaladas (JeReEs), en la Sección IV se presentan los experimentos computacionales y resultados obtenidos, finalmente la Sección V muestra las conclusiones y da algunas guías para trabajos futuros.

II. TRABAJOS RELACIONADOS

Actualmente diversos métodos han surgido para realizar la localización de objetos en base a características de la imagen como la textura, el color, la presencia de bordes, etc. Los trabajos de [17], [18], [19], [20] buscan regiones que delimiten de manera precisa todos los objetos de una imagen clasificando cada uno de los píxeles como correspondientes a la región de algún objeto; este enfoque permite localizar y al mismo tiempo detectar objetos en base a modelos previamente entrenados, sin embargo, son dependientes de la categoría y computacionalmente costosos.

Trabajos como [4],[18],[19],[20] utilizan múltiples segmentaciones para localizar objetos. En [4] se realiza una segmentación múltiple aplicando 3 algoritmos de segmentación sobre una imagen, dichos algoritmos son utilizados con diferentes configuraciones para realizar múltiples segmentaciones que proporcionan diferentes conjuntos de regiones que segmenten la imagen. Las regiones adyacentes obtenidas en cada segmentación son combinadas para generar ventanas de interés que permiten localizar objetos compuestos por regiones heterogéneas.

En [18], [19], [20] las regiones obtenidas de las múltiples segmentaciones son almacenadas en un diccionario, del cual se selecciona un conjunto de regiones que no se solapan, y son clasificadas como pertenecientes a un objeto de interés utilizando modelos entrenados previamente.

Los métodos de [7] y [16] realizan segmentaciones jerárquicas de la imagen donde a partir de las regiones obtenidas en la segmentación, se combinan pares de regiones adyacentes hasta que sólo queda una región que cubre toda la imagen (ver Figura 2).

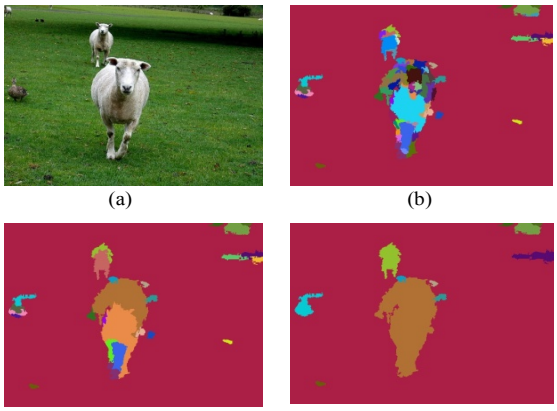

Figura 2. Proceso de fusión de regiones para la obtención de una segmentación jerárquica. (a) Imagen original. (b), (c) y (d) imágenes de las regiones obtenidas en la etapa inicial, media y final (respectivamente) en el proceso de combinación de regiones.

En los trabajos de [4], [5], [7], [14] al igual que en la presente investigación, se busca generar ventanas delimitadoras sobre regiones donde burdamente se localizan los objetos de interés. El método de [7] es el más parecido al que se propone en esta investigación debido a que en ambos se realizan múltiples segmentaciones de una imagen en diferentes modelos de color. Las regiones adyacentes obtenidas en una segmentación se van fusionando de acuerdo a su similitud, obteniendo como resultado una segmentación jerárquica, generando en cada región una ventana de interés donde puede localizarse un objeto. A diferencia del método presentado en [7] que sólo realiza sobre segmentaciones en imágenes de tamaño original, nuestro método realiza sobre segmentaciones en imágenes en su tamaño original y escaladas, aplicando tres heurísticas sobre las regiones obtenidas para obtener segmentaciones jerárquicas.

III. MÉTODO JeReEs PARA LA LOCALIZACIÓN DE OBJETOS EN IMÁGENES NO CONTROLADAS

El método JeReEs está diseñado para localizar objetos en imágenes de escenas naturales las cuales pueden contener objetos con oclusiones parciales, cambios de iluminación, diferente perspectiva, etc. El método JeReEs no sólo obtiene un porcentaje de localización superior al de los trabajos reportados en el estado del arte, sino que también incrementa el número de ventanas de interés, que delimitan el área de un objeto con un porcentaje mayor al 50% (utilizando la ecuación 4), superando a los trabajos del estado del arte reportados en [7]. El método JeReEs consta de 5 etapas principales que se describen en los incisos *b* - *f*. El inciso *a* contiene la descripción de la colección de imágenes utilizada para evaluar el método propuesto.

a) Imágenes de prueba

Colección de prueba PASCAL VOC 2007: cuenta con 4, 952 imágenes y contiene 20 categorías de objetos de interés en donde algunos objetos de una misma categoría presentan grandes diferencias en cuanto a forma y apariencia. Las imágenes constan de 14,734 objetos delimitados manualmente que contribuyen a la evaluación de los métodos de detección de objetos. Las imágenes de esta colección son de escenas naturales por lo que los objetos en ellas pueden estar parcialmente ocluidos, aparecer en diferentes escalas y perspectivas, con diferentes tipos de iluminación, con poco contraste con el fondo, etc.

b) Escalado de imágenes con wavelets

En esta etapa del método JeReEs, se realiza un escalado de las imágenes utilizando la wavelet *Daubichies 3*, la cual genera imágenes de tamaño igual a un cuarto del tamaño original de la imagen. En la Figura 3 se muestra un ejemplo del escalado de una imagen. La Figura 3a muestra una imagen de entrada con el modelo de color RGB, la cual es separada en los canales *R*, *G* y *B* (Figura 3b), en cada canal se aplica la wavelet *Daubichies 3* y se obtienen imágenes escaladas (Figura 3c), dichas imágenes son unidas nuevamente dentro del modelo RGB para obtener una imagen escalada a colores (Figura 3d).

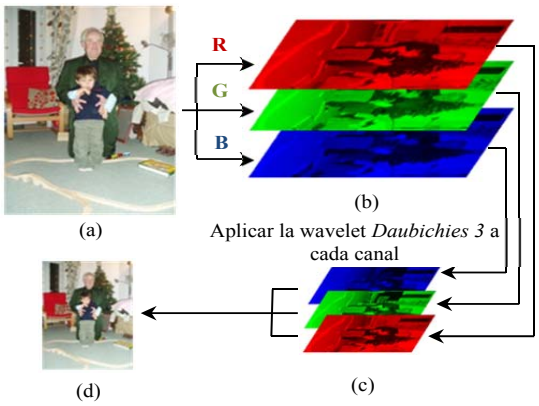

Figura 3. Proceso de escalado de imágenes. (a) imagen original, (b) imágenes de los canales RGB de la imagen original, (c) imágenes obtenidas al aplicar la wavelet a cada canal del modelo RGB, (d) Imagen resultante después de unir los tres canales escalados.

Debido a que las imágenes de la colección de prueba PASCAL VOC 2007 tienen dimensiones que llegan a los 500 píxeles de ancho y alto, la wavelet sólo se aplicó 3 veces para producir una imagen escalada, ya que al aplicarla más veces las segmentaciones obtenidas no proporcionaron regiones útiles para la localización de los objetos.

c) Conversión a diferentes modelos de color

Durante esta etapa se realiza la conversión de las imágenes (tamaño original y escaladas) a los modelos de color CIE Luv y canales *H* e *I* del modelo HSI. La conversión se realiza para resolver el problema de la generación de regiones que no delimitan correctamente a los objetos o que no cubren el área suficiente para su posterior detección.

En la Figura 5c se muestra de izquierda a derecha una imagen en el modelo CIE Luv la cual conserva las características de los colores de la imagen original pero con una representación distinta, posteriormente se presentan las correspondientes imágenes en niveles de gris de los canales *H* e *I* del modelo HSI, dichas imágenes presentan pérdida de información con respecto a la imagen original ya que sólo toman en consideración un canal a la vez, sin embargo, permiten localizar algunas regiones de objetos que sólo se logran identificar a través de su matiz o intensidad.

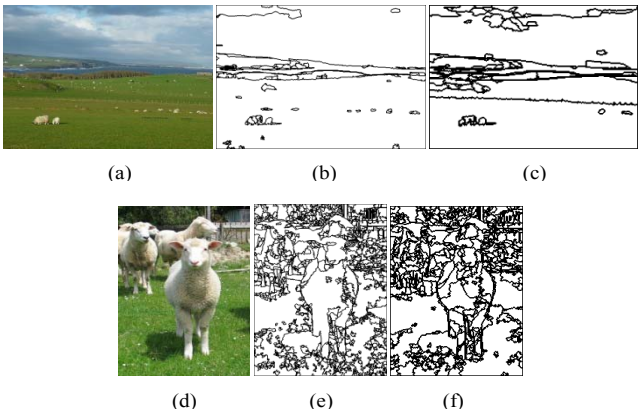

Figura 4. (a) y (d) Imágenes originales. (b) y (e) regiones obtenidas al segmentar las imágenes (a) y (d) respectivamente en su tamaño original. (c) y (f) regiones obtenidas al segmentar la imagen resultante al aplicar la wavelet *Daubichies 3* (como se muestra en la Figura 3) sobre las imágenes de (a) y (d) respectivamente. Las regiones obtenidas sobre el pasto en las imágenes (c) y (f) son menos que las obtenidas en las imágenes de (b) y (e).

d) Generación de múltiples segmentaciones

Esta fase consiste en la generación de múltiples segmentaciones sobre la imagen original y la imagen escalada utilizando diferentes modelos de color. El objetivo es generar diferentes conjuntos de regiones que segmenten la imagen, aprovechando que una imagen escalada requiere menos tiempo para su procesamiento y que al igual que la original puede sobre segmentarse obteniendo menos regiones con diferente soporte espacial. La reducción de regiones se presenta especialmente en áreas que originalmente tienen mucha textura con una escala pequeña y al ser escalada dicha textura parcialmente desaparece (Figura 4). En el método JeReEs se aplica la segmentación una vez sobre la imagen original y 7 veces sobre las imágenes escaladas utilizando diferentes modelos de color.

Para realizar la segmentación se utilizó el algoritmo Mean Shift [21], los valores de los parámetros del algoritmo fueron configurados de tal manera que se pueden obtener sobre segmentaciones. La Figura 5d muestra el resultado de segmentar una imagen de tamaño original utilizando diferentes modelos de color, presentando de izquierda a

derecha las regiones obtenidas en el modelo de color CIE Luv, canal *H* y canal *I* respectivamente. Se puede observar que la segmentación aplicada sobre el modelo CIE Luv genera más regiones que las otras dos, debido a que proporciona mayor información de color, sin embargo, las segmentaciones sobre los canales *H* e *I* son necesarias para obtener regiones con diferente apoyo espacial. La Figura 6 muestra el resultado de la segmentación aplicada a una imagen en diferentes escalas. La Figura 6a corresponde a la imagen original, la 6b presenta las regiones resultantes al segmentar la imagen original y de la 6c a la 6e se presentan los resultados obtenidos al segmentar la imagen escalada aplicando la wavelet *Daubichies 3* una, dos y tres veces, respectivamente.

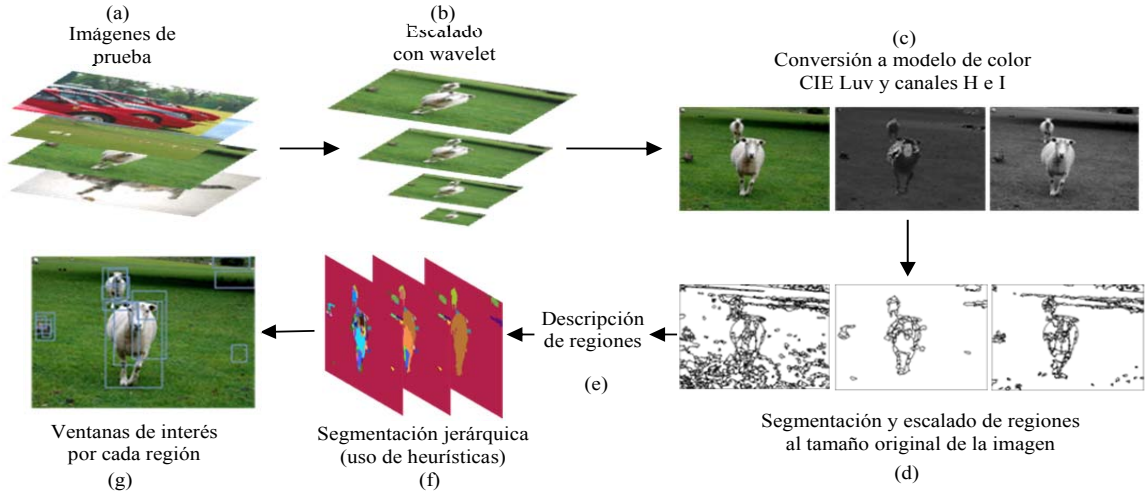

Figura 5. Método para la localización de objetos JeReEs. (a) conjunto de imágenes de prueba de PASCAL VOC 2007. (b) Conjunto de imágenes escaladas de una imagen de prueba. (c) Imágenes en el modelo de color CIE Luv, y los canales *H* e *I* del modelo de color HSI. (d) Regiones obtenidas aplicando Mean Shift sobre las imágenes en diferentes modelos de color. (e) Descripción de las regiones, utilizando textones con el banco de filtros de [17] y los descriptores de color modelos HSI, YCC, CIE Lab y RGB. (f) Resultados de la segmentación jerárquica, combinando regiones adyacentes empleando heurísticas. (g) Ventanas de interés que realizan la posible localización de objetos.

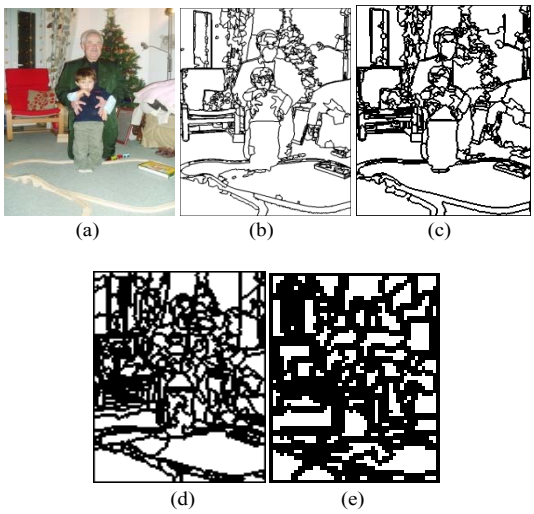

Figura 6. (a) imagen original de 375x500 píxeles, (b) regiones obtenidas de una imagen de 188x250 píxeles obtenida después de aplicar una vez la wavelet, (c) regiones obtenidas de una imagen de 94x125 píxeles obtenida después de aplicar dos veces la wavelet, (d) regiones obtenidas sobre una imagen de 47x63 píxeles obtenida después de aplicar tres veces la wavelet.

e) Descripción de regiones

En esta fase, con el objetivo de incrementar el tamaño de las regiones de imágenes escaladas (Figuras 6c - 6e) a su tamaño original, a las regiones obtenidas durante la segmentación de imágenes escaladas se les aplicó la ecuación 1 *n* veces, donde *n* corresponde al número de veces que se empleó la wavelet partiendo de la imagen original.

$$R_{e\text{-}1}(d_x, d_y) = R_e(d_x*2, d_y*2) \quad (1)$$

donde R_e es la región escalada resultante de la aplicación de la wavelet, R_e-1 es la región aumentada en cuatro veces el tamaño de la región R_e, d_x son los píxeles de una región en el plano *x* y d_y son los píxeles de una región en el plano *y*.

Las regiones obtenidas fueron descritas utilizando textones con el banco de filtros propuesto en [17] y con descriptores de color de los modelos HSI, YCC, CIE Lab y RGB.

f) Segmentación jerárquica

Esta fase recibe como entrada las regiones descritas, obtenidas en la fase *e)*, las cuales contribuyen a la generación de segmentaciones jerárquicas mediante la aplicación de 3 heurísticas (h1, h2 y h3). La función de las heurísticas consiste en combinar regiones adyacentes para obtener nuevas regiones que permitan delimitar un objeto con mayor precisión que la alcanzada en las regiones

originales. En este sentido, un objeto puede estar compuesto por varias regiones que pueden tener o no características en común, de tal forma que la heurística h1 genera ventanas que cubren el área de 2 regiones adyacentes sin considerar sus vectores de características como se describe a continuación: "para cada región x_i, (i = 1, ..., m donde m es el número de regiones que segmenta una imagen), y cada una de sus regiones adyacentes y_j (j = 1, ..., n donde n es el número de regiones adyacentes de la región x_i), se establece una ventana delimitadora que cubre el área de las regiones (x_i, y_j) sin unirlas y generando k ventanas que pueden delimitar un objeto de interés.

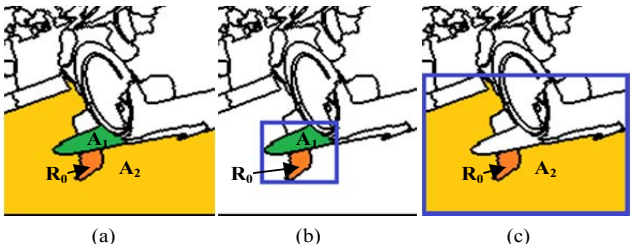

Figura 7. Aplicación de la heurística h1 sobre las regiones de una imagen para la obtención de ventanas delimitadoras.

En la Figura 7 se presenta un ejemplo de generación de ventanas delimitadoras aplicando la heurística h1 y tomando como referencia la región R_0. En la Figura 7a se observa que la región R_0 es adyacente a las regiones A_1 y A_2. La Figura 7b muestra la combinación de la región R_0 con A_1 y la ventana delimitadora que se genera (recuadro azul). Por otro lado, la Figura 7c muestra la ventana que se genera mediante la combinación de las regiones R_0 y A_2 (recuadro azul).

La heurística h2 iterativamente combina 2 regiones adyacentes (recuadro azul Figura 7b) mientras el vector de características de ambas regiones es similar y la suma de sus áreas es la menor. El proceso se realiza hasta que el número de regiones alcanza el valor de 1 y la región cubre toda la imagen. Cada nueva región resulta de la unión de 2 regiones y es delimitada por una ventana. Para seleccionar las regiones a fusionar se aplica la ecuación 2:

$$Fusionar(xi,yj)=MIN(0.5*area(xi,yj)+0.5*diferencia(xi,yj)) \quad (2)$$

donde para cada región x_i (i = 1, ..., m; donde m es el número de regiones que segmenta una imagen) y cada región adyacente y_j (j = 1, ..., n donde n es el número de regiones adyacentes de la región xi) se calculan las funciones *area(xi, yj),* que obtiene el valor del número de píxeles de las regiones xi y yj, y *diferencia(xi, yj),* que representa la disimilitud existente entre el vector de características de las regiones xi y yj utilizando la distancia euclidiana como métrica de disimilitud. El cálculo del vector de características de la nueva región se realiza con la ecuación 3:

$$Vector(x,y) = \frac{(Vector(x) * (area(x)) + (Vector(y) * (area(y))}{(area(x,y))} \quad (3)$$

donde x y y son las regiones que se combinan. Como ejemplo, el resultado de esta segmentación jerárquica se muestra en la Figura 5f.

La heurística h3 al igual que la h1 genera ventanas que cubren el área de 2 regiones adyacentes sin tomar en cuenta sus vectores de características, sin embargo, h3 es dependiente de h2 ya que cada región x_i corresponde a cada una de las nuevas regiones generadas por h2. En la Figura 8 se muestra un ejemplo de la aplicación de la heurística h2 y h3.

La Figura 8a muestra 2 regiones adyacentes R_1 y R_2 que se fusionan aplicando la heurística h2 obteniendo la región R_3 de la Figura 8b, en esta nueva región R_3 se aplica la heurística 3 que, al igual que la heurística h1, considera las regiones adyacentes de la región R_3 (Figura 8c) y las combina una a una con la región R_3 generando ventanas delimitadoras para las dos regiones combinadas (Figuras d - f).

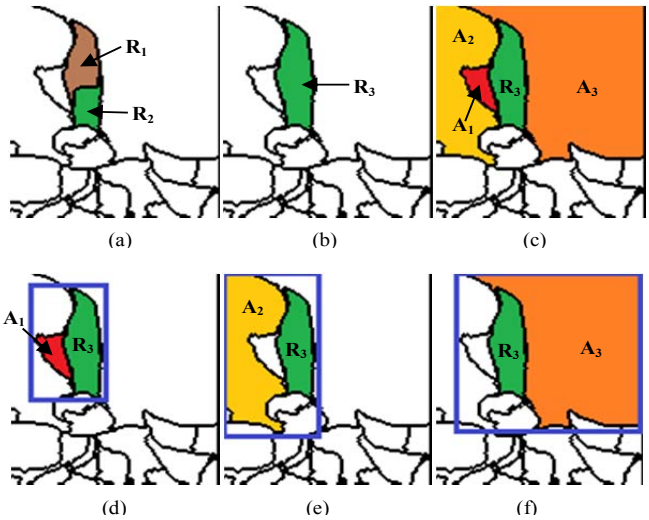

Figura 8. Aplicación de las heurísticas h2 y h3 sobre las regiones de una imagen para la generación de un segmentación jerárquica y la producción de ventanas delimitadoras.

En la Figura 5g se muestran las ventanas delimitadoras producidas por el método JeReEs; se puede observar que varias de las ventanas (recuadros verdes) cubren a los objetos considerando diferentes apoyos espaciales o regiones. En cada una de las ventanas se aplica el método presentado en [2] para detectar la presencia de algún objeto de interés.

IV. EXPERIMENTACIÓN COMPUTACIONAL

Para la realización de los experimentos se utilizó un equipo con procesador de 2.0 GHz y 16 GB en RAM. Las imágenes de prueba corresponden a la colección PASCAL VOC 2007. La colección fue seleccionada ya que contiene imágenes de escenas naturales y es considerada como la más compleja y además es utilizada como punto de comparación por varios trabajos del estado del arte [5],[11],[14],[15],

incluyendo [7] que a la fecha reporta el porcentaje más alto en localización de objetos.

Durante la experimentación, se realizaron 3 pruebas para medir la efectividad en la localización de objetos del método JeReEs. En la primera prueba se realizó una segmentación, la cual produjo pocas ventanas de interés por imagen, pero con un porcentaje de localización superior al alcanzado por los trabajos [11] y [14]. En esta prueba y en la segunda se utilizó la métrica de [9] (ecuación 4) que mide el grado de solapamiento entre dos ventanas, lo cual permite establecer si una ventana de interés localiza correctamente un objeto.

$$Ao = \frac{area(Vp \cap Vr)}{area(Vp \cup Vr)} \quad (4)$$

donde $Vp \cap Vr$ denota la intersección de las ventanas propuesta por el método JeReEs y las ventanas que delimitan el objeto respectivamente; $Vp \cup Vr$ denota la unión de las áreas de la ventana propuesta y la ventana que delimita el objeto. Se considera un objeto correctamente localizado cuando el porcentaje de solapamiento entre las ventanas es mayor o igual al 50%. En la prueba 2, se muestra el resultado de aplicar múltiple segmentación, lo cual contribuye a superar el porcentaje de localización de objetos de la prueba 1 y además el reportado en [7].

La tercera prueba muestra cómo al incrementar el número de ventanas de localización de un objeto, también se incrementa el porcentaje de localización de objetos en una imagen. Aplicando la ecuación 4 y el método basado en partes de [2] para la detección de objetos, se muestra que JeReEs obtiene un 80% de exactitud en la detección de objetos. A continuación se detalla las pruebas realizadas con JeReEs.

Prueba 1. Método JeReEs aplicando una segmentación

El objetivo de esta prueba es mostrar que el método JeReEs con una sola segmentación puede obtener mejores resultados que algunos de trabajos del estado del arte reportados en [7], aún cuando el número de ventanas producidas por imagen es menor. La segmentación se realizó utilizando el método Mean Shift [21] con la siguiente configuración: $e = 7$, $r = 6.5$ y $mp = 50$, siendo aplicada sobre la colección de imágenes en su tamaño original y utilizando el modelo de color CIE Luv. Con esta configuración se generan en promedio 11 ventanas de interés por objeto localizado.

La Tabla 1 presenta en la columna "Trabajo" los nombres de los trabajos relacionados que se comparan, en la columna "Ventanas por imagen" el número de ventanas producidas por imagen. La columna "Número de segmentaciones" indica el número de segmentaciones que se aplican en cada trabajo, en este sentido, los trabajos de [11], [14] y [5] no requieren de la segmentación. La columna "localización" presenta el porcentaje de localización obtenido por cada trabajo.

En la Tabla 1 se observa que JeReEs utiliza menos ventanas por imagen, esto tiene la ventaja de que se pueden utilizar descriptores más complejos para describirlas sin requerir un tiempo computacional excesivo, además se observa que los trabajos de [7] y [14] obtienen mejores resultados en cuanto al porcentaje de localización de objetos, sin embargo, en [14] se produce la mayor cantidad de ventanas. En cuanto JeReEs, en esta prueba sólo necesita una segmentación, por tanto requiere menos tiempo de procesamiento para su ejecución y reduce también el tiempo requerido para la detección de objetos, siendo posible la aplicación de este método en sistemas que exigen una rápida respuesta, por ejemplo, los sistemas de clasificación de imágenes Web basados en el contenido de la imagen [22].

Trabajo	Ventanas por imagen	Número de segmentaciones	Localización (%)
Ventanas deslizantes [11]	4,000	-	83.00
Jumping Windows [14]	200,000	-	94.00
Objectness [5]	10,000	-	82.40
Búsqueda selectiva [7]	1,536	8	96.70
JeReEs 1 segmentación.	1,338	1	86.68

TABLA I. RELACIÓN ENTRE EL NÚMERO DE VENTANAS GENERADAS POR IMAGEN Y LA EFICIENCIA DE LOCALIZACIÓN EN LA COLECCIÓN DE PRUEBA PASCAL VOC 2007.

Prueba 2. Método JeReEs aplicando múltiples segmentaciones

Esta prueba tiene como objetivo incrementar el porcentaje de localización de objetos al aumentar el número de ventanas cuyo porcentaje de solapamiento sea mayor o igual al 50% (aplicando la ecuación 4).

A diferencia de la prueba 1, donde se aplica una sola segmentación sobre la imagen, en esta prueba se aplicaron 10 segmentaciones cuyos resultados contribuyen a incrementar el porcentaje de localización de objetos. Durante la aplicación de la segmentación, el Algoritmo Mean Shift fue configurado utilizando los parámetros que se presentan en la Tabla II. La columna "No." indica el número de la segmentación, la columna "Modelo de color" hace referencia al modelo de color se utilizado (CIE Luv, canales H, canal I). La columna "Aplicación wavelet" indica el número de veces que fue aplicada la wavelet sobre la imagen original para obtener una imagen escalada, un valor de cero significa que no se aplicó escalado sobre la imagen. Las columnas "Dominio espacial" y "Dominio de rango" presentan los valores del dominio espacial y del rango del algoritmo Mean Shift. La variación de los valores en los parámetros de dominio y rango generan diferentes conjuntos de regiones lo cual incrementa la posibilidad de localizar los objetos contenidos. La columna "Min. de píxeles" indica cuál es el mínimo de píxeles que debe tener una región, si una región no tiene una cantidad igual o mayor a la especificada, la región en cuestión es combinada con una región adyacente.

TABLA II. MODELOS DE COLOR Y PARÁMETROS DEL ALGORITMO DE SEGMENTACIÓN MEAN SHIFT.

No.	Modelo de color	Aplicación wavelet	Dominio espacial	Dominio de rango	Min. de pixeles
1	CIE Luv	0	7	6.5	25
2	CIE Luv	1	6	5	15
3	CIE Luv	2	5	4.5	15
4	CIE Luv	3	4	2.5	15
5	H	0	5	2	100
6	H	1	5	2	50
7	H	2	5	2	20
8	H	3	5	2	10
9	I	0	6	5	25
10	I	1	6	5	20

La columna "No." indica el número de la prueba realizada, la columna "No. ventanas" muestra el número de ventanas generadas por imagen, la columna "No. segmentaciones" indica cuántas segmentaciones fueron requeridas, la columna "Ventanas por objeto" presenta el número de ventanas que se obtuvieron por objeto, la columna "Localización de objetos" presenta el porcentaje de localización de objetos alcanzado al considerar un 50% de solapamiento.

TABLA III. RESULTADOS OBTENIDOS DE LOCALIZACIÓN DE OBJETOS EN LA PRUEBA 1 Y 2.

No. Prueba	No. ventanas	No. Segmentaciones	Ventanas por objeto	Localización de objetos
1	1,338	1	11	86.68%
2	9,366	10	74	96.87%

A pesar de que la prueba 2 supera la localización de objetos obtenida en la prueba 1 en más de un 10%, la prueba 1 genera 7 veces menos ventanas de interés por imagen y sólo requiere de una segmentación haciéndola útil para sistemas que requieran mayor velocidad de respuesta a costa de exactitud en la localización.

Prueba 3. Porcentaje de localización de objetos del método JeReEs

Para validar la hipótesis de que "la detección de objetos en imágenes naturales puede ser incrementada si se dispone de un mayor número de ventanas con diferente apoyo espacial por cada objeto localizado", dadas las ventanas obtenidas en la prueba 2, se consideraron aquellas cuyo porcentaje de solapamiento fuera mayor al 80%. Con el filtrado de las información, el número de ventanas se redujo a 22 en promedio por imagen y a 12 por objeto de interés, lo cual contribuyó a incrementar el porcentaje de localización hasta un 61.59%. Para el proceso de reconocimiento se utilizó el método basado en partes de [2]. Los resultados obtenidos con JeReEs son comparados contra los resultados obtenidos en [7].

La precisión promedio (average precision - AP) obtenida para cada una de las 20 categorías se presenta en la Tabla IV, donde se puede observar que las categorías de autobús y de auto obtuvieron las precisiones promedio más altas, siendo estas de 0.633 y 0.626 respectivamente.

TABLA IV. PRECISIÓN PROMEDIO OBTENIDA EN CADA CATEGORÍA.

No.	**Categoría**	**AP**	**No.**	**Categoría**	**AP**
1	Pájaro	0.091	11	Aeroplano	0.453
2	Bote	0.266	12	Bicicleta	0.520
3	Gato	0.182	13	Sofá	0.545
4	Maceta	0.091	14	Caballo	0.434
5	Botella	0.273	15	Motocicleta	0.545
6	Perro	0.180	16	TV/Monitor	0.541
7	Oveja	0.345	17	Tren	0.633
8	Silla	0.557	18	Mesa	0.535
9	Persona	0.452	19	Auto	0.626
10	Vaca	0.182	20	Autobús	0.455

Se obtuvo una media de precisión promedio (Mean Average Precision - MAP) de 0.3953 que supera el 0.2960 obtenido en [7] con una mejora de 0.0993, sin embargo, el trabajo de [7] obtuvo una precisión promedio superior en las categorías *ave* y *bicicleta*.

En la figura 7 se presenta de forma gráfica el porcentaje de la precisión promedio obtenida para cada una de las categorías, se observa que en las categorías auto y autobús se obtuvieron los porcentajes más altos, por encima del 60%, en las categorías motocicleta, tren, TV/monitor, sofá y mesa obtuvieron porcentajes arriba del 50%, para las categorías vaca, aeroplano, persona y bicicleta se obtuvo un porcentaje superior al 40%, para la categoría silla se obtuvo un porcentaje de 34.5%, en las categorías perro y oveja se alcanzaron resultados cercanos al 30%, en las categorías maceta, gato y botella se obtuvieron porcentajes cercanos al 20% y para las categorías pájaro y bote se obtuvieron resultados por debajo del 10%.

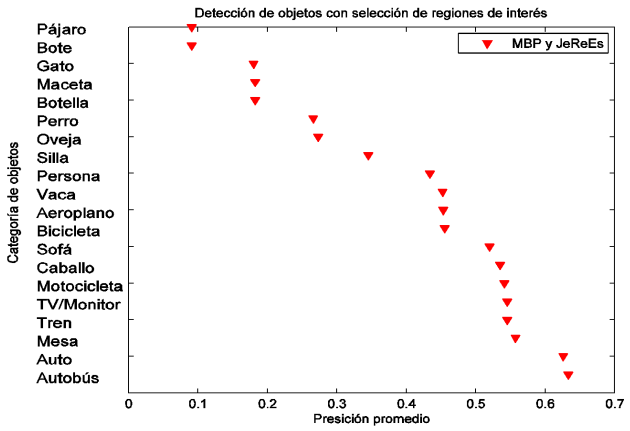

Figura 7. Resultados obtenidos en el reconocimiento de 20 categorías de objetos utilizando la colección de imágenes de prueba del evento PASCAL VOC 2007.

V. CONCLUSIONES

Este trabajo tiene cinco aportaciones. La primera es que se obtiene un porcentaje de localización de 96.87% siendo el mejor resultado obtenido en la colección de prueba PASCAL VOC 2007. La segunda es mostrar que el combinación de imágenes escaladas con diferentes modelos de color proporciona conjuntos de regiones que pueden localizar objetos. La tercera es que la segmentación aplicada a imágenes escaladas permite obtener sobre segmentaciones y se reducen las regiones generadas en áreas con mucha textura. La cuarta consiste en la utilización de heurísticas que permiten localizar objetos compuestos por regiones heterogéneas. La quinta aportación es la generación una mayor cantidad de ventanas con diferente apoyo espacial que delimitan a un objeto para mejorar la detección del mismo; para la evaluación de esta aportación se utilizó el método de [2] obteniendo una media de precisión promedio de 0.3953 superando en un 0.0993 el resultado de [7]. Por último, de acuerdo a los resultados de [7] y a los obtenidos en la prueba 3 por el método JeReEs, se observa que ambos métodos podrían ser complementarios.

El método JeReEs podría ser mejorado realizando: 1) Un análisis de las características de los objetos que no fueron localizados, creando nuevas heurísticas que permitan incrementar su eficiencia y 2) El análisis de las características de las ventanas generadas en [7] y las generadas por JeReEs, para incorporar las ventajas de ambos métodos en un sólo método de localización.

REFERENCIAS

[1] N. Dalal and B. Triggs. “Histograms of oriented gradients for human detection”. In IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886 - 893, (2005).

[2] P. Felzenswalb, R. Girshick, D. McAllester, and D. Ramanan. "Object Detection With Discriminatively Trained Part Based Models". In IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1627 - 1645, (2010).

[3] L. Zhu, Y. Chen, A. Yuille and W. Freeman. "Latent Hierarchical Structural Learning for Object Detection". In Computer Vision and Pattern Recognition, pp. 1062 - 1069, (2010).

[4] T. Malisiewicz and A. A. Efros. "Improving Spatial Support for Objects via Multiple Segmentations". In British Machine Vision Conference, pp. 1 - 10, (2007).

[5] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. "Multiple kernels for object detection". In International Conference on Computer Vision, pp. 606 - 613, (2009).

[6] I. Endres and D. Hoiem. "Category independent object proposals". In European Conference on Computer Vision, pp. 575 - 588, (2010).

[7] K. van de Sande, J. R. Uijlings, T. Gevers and A. Smeulders. "Segmentation as Selective Search for Object Recognition". In IEEE International Conference on Computer Vision, pp. 1879 - 1886 (2011).

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. "ImageNet: A large-scale hierarchical image database". In IEEE Computer Vision and Pattern Recognition, pp. 248 - 255, (2009).

[9] M. Everingham, L. van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. "The Pascal Visual Object Classes (VOC) Challenge". on International Journal of Computer Vision, vol. 88, pp. 303-338, (2010).

[10] J. Zhang, K. Huang, Y. Yu and T. Tan. "Boosted Local Structured HOG-LBP for Object Localization". In Computer Vision and Pattern Recognition, pp. 1393 - 1400, (2011).

[11] H. Harzallah, F. Jurie, and C. Schmid. "Combining efficient object localization and image classification". In International Conference on Computer Vision, pp. 237 - 244, (2009).

[12] P. Viola and M. Jones. "Robust real-time face detection". In International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, (2004).

[13] H. Schneiderman and T. Kanade. "A Statistical Method for 3D Object Detection Applied to Faces and Cars". In IEEE Conference of Computer Vision Pattern Recognition, vol. 1, pp. 746 - 751, (2000).

[14] B. Alexe, T. Deselaers, and V. Ferrari. "What is an object?". In Computer Vision and Pattern Recognition, pp. 73 - 80, (2010).

[15] C. H. Lampert, Blaschko M. B. and T. Hofmann: "Efficient Subwindow Search: A Branch and Bound Framework for Object Localization". In IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2129- 2142, (2009).

[16] C. Gu, J. J. Lim, P. Arbeláez, and J. Malik. "Recognition using regions". In Computer Vision and Pattern Recognition, pp. 1030 - 1037, (2009).

[17] J. Winn, A. Criminisi and T. Minka. "Object Categorization by Learned Universal Visual Dictionary". In International Conference on Computer Vision, vol. 2, pp. 1800 - 1807, (2005).

[18] S. Gould, R. Fulton, and D. Koller. "Decomposing a scene into geometric and semantically consistent regions". In International Conference on Computer Vision, pp. 1 - 8, (2009).

[19] S. Gould, T. Gao, and D. Koller. "Region-based segmentation and object detection". In Neural Information Processing Systems, pp. 655 - 663, (2009).

[20] M. Pawan Kumar and D. Koller. " Efficiently Selecting Regions for Scene Understanding". In Proceedings of Conference on Computer Vision and Pattern Recognition, pp. 3217 - 3224, (2010).

[21] D. Comaniciu and P. Meer. "Mean Shift: a robust approach toward feature space analysis". In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, (2002).

[22] P. R. Kalva, F. Enembreck and A. L. Koerich. "WEB Image Classification using Combination of Classifiers". In IEEE Latin America Transactions, vol. 6, pp. 661 - 671, (2008).

Salvador Cervantes Alvarez recibió el grado Maestro en Ciencias en Ciencias Computacionales por el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) y actualmente está inscrito en el programa doctoral en Ciencias Computacionales en el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET). Sus áreas de interés incluyen el reconocimiento de patrones y el procesamiento digital de imágenes.

Raúl Pinto Elías recibió el grado Maestro en Ciencias en Ciencias Computacionales por el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) y el grado de Doctor en Ciencias en Ingeniería Eléctrica del Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN). Es profesor Titular en el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET). Sus área de interés incluyen la visión por computadora, el reconocimiento de patrones, la visión robótica y sus aplicaciones a procesos de inspección y manufactura.