

# Segmentación Semántica para Imágenes de Paisajes Tropicales

## Semantic Segmentation for Images of Tropical Landscapes

**Corzo, U.G.**

Ejército Nacional de Colombia

germancorzeus@gmail.com

Fecha de recepción: agosto de 2018 / Fecha de aceptación: septiembre de 2018

## Resumen

En el presente trabajo se estableció un método para segmentación semántica de imágenes cartográficas de paisajes tropicales colombianos, a través del análisis supervisado de grupos de píxeles en tres clases: vegetación, fuentes de agua y otros objetos o elementos extraños en la naturaleza del paisaje tropical, que tienen principalmente, influencia del ser humano. El estudio de estas clases, permite reconocer aquellas áreas de píxeles en una fotografía cartográfica, donde pueden existir eventos como deforestación, minería ilegal o cultivos ilícitos, también permite detectar otros fenómenos como zonas inundables o con peligro de erosión y deslizamientos. La técnica planteada permite integrar las características de color, textura y bordes presentes en el contexto de grupos de píxeles y transformar el espacio de múltiples canales, directamente en un espacio vectorial euclidiano. Con ello, se logra mejorar el tiempo de procesamiento de máquina, al tener una representación estadística de los datos, en lugar de una gran cantidad de datos crudos píxel por píxel y permite clasificar los datos, utilizando técnicas de máquinas de aprendizaje. Las imágenes para la investigación provinieron de una base de datos de video, tomadas de forma oblicua, desde plataformas aéreas tripuladas las cuales fueron seleccionadas y clasificadas previamente de acuerdo con la existencia de cada una de las clases definidas, además se requirió del conocimiento experto para la segmentación manual de ciertas imágenes, las cuales sirvieron como contraste para desarrollar el algoritmo y comparar la efectividad en la clasificación de píxeles. Finalmente, se hizo un análisis de error, tanto en el entrenamiento de la máquina como en el proceso de clasificación.

**Palabras clave:** método, segmentación, cultivos, deforestación.

## Abstract

In the present work a method was established for semantic segmentation of cartographic images of Colombian tropical landscapes through the supervised analysis of groups of pixels in three classes: vegetation, water sources and other objects or strange elements in the nature of the tropical landscape that have Mainly human influence, the study of these classes allows recognizing those areas of pixels in a cartographic photograph where there may be events such as deforestation, illegal mining or illicit crops, it also allows detecting other phenomena such as flood zones or with the risk of erosion and landslides. The proposed technique allows integrating the characteristics of color, texture and edges present in the context of groups of pixels and transform the space of multiple channels directly into a Euclidean vector space. This is achieved by improving the machine processing time by having a statistical representation of the data instead of a large amount of raw data pixel by pixel and allows classifying the data using machine learning techniques. The images for the investigation came from a video database taken obliquely taken from manned aerial platforms which were previously selected and classified according to the existence of each of the defined classes, in addition expert knowledge was required for the manual segmentation of certain images, which served as a contrast to develop the algorithm and compare the effectiveness in the classification of pixels. Finally, an-error analysis was made in the training of the machine as well as in the classification process.

**Keywords:** method, segmentation, crops, deforestation.

## Introducción

Para un analista de imágenes, es difícil clasificar y dar conceptos de valor sobre lo que puede determinarse a simple vista de una fotografía o un mosaico de imágenes a color, en la figura 1, se observa una imagen segmentada manualmente, pero el tiempo y exactitud de la clasificación, depende de la experiencia y destreza del analista. En el análisis de imágenes cartográficas, es necesario conocer diferentes objetos que son de interés para determinar múltiples factores con respecto a la demografía, hidrografía, medio ambiente, zonas de producción y vías transitables, entre otras. Luego, identificar, conocer, delimitar, detectar y comparar los objetos de una imagen obtenida con sensores CCD o infrarrojo que son determinantes para dar mayor valor a este tipo de imágenes. Guevara (2015) cita “la técnica más

usada para diferenciar los objetos es la segmentación, existen varios métodos y algoritmos que usan la probabilidad de pertenecer a cierto punto en el espacio de color o el *clustering*, basado en la ubicación de centros de color a cierta distancia”.

Ponce (2010) afirma que “el agrupamiento de datos (*clustering*), es una técnica común para la clasificación de datos que permite tener un mejor manejo de ellos”; esta técnica se utiliza en diversos campos como en la identificación de estructuras en modelos difusos, compresión de datos y la segmentación de imágenes, donde la distribución de la información puede ser de cualquier tamaño y forma. Las técnicas de segmentación que analizan píxeles individuales son poco confiables para muchos tipos de imágenes, pues los valores de los píxeles pueden ser los mismos en clases diferentes.



**Figura 1.** Izquierda: imagen de prueba. Derecha: imagen segmentada por conocimiento experto con 3 clases: vegetación, fuentes de agua y otros.

## Desarrollo

**Análisis de la técnica de segmentación utilizada.** Actualmente, se usan técnicas que analizan el contexto de los píxeles a clasificar, por esta razón se suelen observar grupos cercanos de píxeles y utilizar técnicas que permiten conocer el valor y la relación estadística con sus vecinos, así se puede conocer si hace parte de un borde, un espacio con textura o si efectivamente hay una superficie plana, en cuyo caso lo más probable es que los píxeles cercanos tendrán valores muy parecidos en el espacio de color.

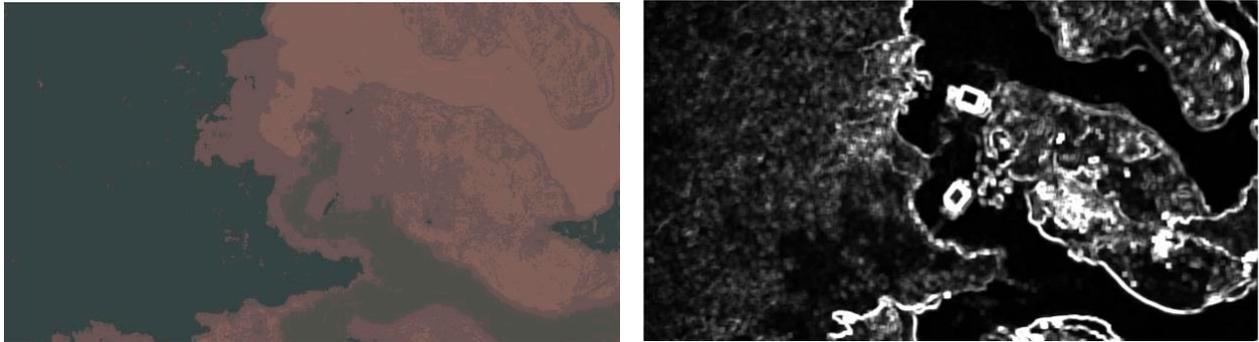
En este artículo se analizó una técnica de clasificación, a través de máquinas de vectores soporte multiclase, el primer problema planteado consiste en presentar los datos de tal forma que puedan ser una entrada para la máquina de entrenamiento y la clasificación de los datos. Para ello, se aplicó una transformación de la matriz de covarianza, a partir de una representación en puntos sigma (Kluckner, 2009), que representa una integración de apariencia, respuesta a filtros y otras características contenidas en el contexto de grupos de píxeles.

Los resultados fueron recogidos en tablas que muestran la probabilidad de error y la matriz de confusión para algunas técnicas de clasificación, a través de máquinas de aprendizaje, luego entrenar y clasificar los datos con una metodología supervisada y realizar el análisis de error en la clasificación, comparándola con las plantillas hechas por el experto y los resultados de un algoritmo de segmentación por píxel.

## Desarrollo del proceso de segmentación de imágenes

**Procesamiento de Datos.** Para el procesamiento de los grupos de píxeles en el entrenador, se realizaron pruebas sobre cinco imágenes de paisajes tropicales seleccionadas de un conjunto de 219 fotografías, tomadas desde una plataforma aérea con sensor CCD e IR. La selección se hizo teniendo en cuenta la presencia de las tres clases requeridas en cada una de ellas y verificando que fuesen homogéneas entre sí, es decir, tomadas en lo posible de forma vertical, a la misma hora del día y en condiciones de sol/nubosidades similares.

Luego se realizaron pruebas del comportamiento sobre los píxeles de color, a través de algoritmos de clasificación no supervisada como k-means y mezcla de gaussianas, respuesta a filtro promedio y gaussiano, y detección de esquinas, a través del algoritmo de Harris. Así se determinó cómo las diferentes clases responden a los algoritmos que procesan los píxeles, observándose por ejemplo que los grupos de color no reconocen las fuentes de agua, pues esta puede tener colores diferentes. Sin embargo, el filtro promedio, con el cual se puede analizar la varianza entre grupos de píxeles muestra que esta misma clase tiene un comportamiento, en el cual los píxeles adjuntos tienen gran probabilidad de ser iguales, reconociendo el agua sin importar el color que tenga. De forma similar el comportamiento de los algoritmos en diferentes características muestra la importancia del contexto para determinar las clases. Figura 2.



**Figura 2.** Izquierda: clasificación de píxeles por color (RGB) con mezcla de gaussianas. Derecha: textura con filtro promedio.

Los descriptores de covarianza (Tuzel 2008) proveen una representación de características de baja dimensión que puede integrar de forma simple múltiples canales, como color, respuesta a filtros y bordes, entre otros y utiliza la correlación entre ellos. Las matrices de covarianza pueden ser eficientemente calculadas y proveen una integración de varios canales de características. Sin embargo, el espacio de las matrices de covarianza no tiene la forma de espacio vectorial euclidiano; esta representación no puede ser directamente usada para muchas técnicas de aprendizaje de máquina. Para resolver este inconveniente y aprovechar el potencial de los descriptores de covarianza, se propone una técnica que realiza la transformación del espacio vectorial.

#### Algoritmo para presentación de datos.

Para la representación de características se propone un algoritmo basado en puntos sigma (Julier 1996), el cual requiere el vector de medias  $\mu^k$  y la matriz de covarianza  $\Sigma^k$ . Para representar matrices de covarianza individuales directamente en un espacio vectorial euclidiano, es posible realizar una transformación que consiste en encontrar un pequeño conjunto de  $2d + 1$  vectores

específicos si  $\in \mathbb{R}^d$  llamados Puntos Sigma. El cálculo del vector se desarrolló siguiendo los cuatro pasos siguientes:

- » Realizar una regularización simple.

$$\Sigma^k = \Sigma^k + \epsilon I \quad (1)$$

- » Calcular la raíz cuadrada de la matriz tal que

$$\Sigma^k = LL^T \quad (2)$$

- » Construir el conjunto de Puntos Sigma.

## Solución Propuesta

Se propuso implementar la técnica de clasificación, aplicando SVM con kernel lineal, cuadrático, gaussiano y cúbico, los cuales están incluidos en el toolbox de Matlab y compararlo con otras técnicas, desde los resultados de entrenamiento y seleccionar la de menor error. Para

implementar el clasificador se presentaron los datos, aplicando el algoritmo basado en Puntos Sigma para una transformación de la matriz de covarianza en el espacio de color (RGB) de una selección aleatoria de un conjunto de píxeles  $m \times n$ .

Se procesaron conjuntos de 10, 25 y 50 muestras por cada clase, para un total de 150 cuadros aleatorios de  $10 \times 10$  píxeles. Para ingresar los datos al entrenador, se tuvo cada muestra individual etiquetada con un total de 21 características, producto del cálculo de los Puntos Sigma.

El resultado del entrenamiento de máquina mostró que son casi perfectamente diferenciables en todos los casos, mostrando los mejores resultados con kernel cúbico y gaussiano fino, se seleccionó este último para realizar las pruebas con datos provenientes de la misma imagen y de otra similar.

Los datos tanto de entrenamiento, como los de los resultados de clasificación se guardaron en una tabla para su uso posterior, las muestras se tomaron aleatoriamente en forma supervisada y se encontraron los errores por comparación de los resultados de la imagen completa de los datos de prueba, también se revisó la matriz de confusión, con el fin de analizar los resultados frente al error por cada clase.

**Análisis de error por cuadro.** De la imagen o11M.tif se tomaron 50 muestras por cada clase para un total de 150, que son entrenadas con SVM kernel “fine gaussian”. El modelo se aplicó a 30 cuadros aleatorios (10 por cada clase) de la misma imagen (o11M.tif) y de otra imagen (o12M.tif). De esta forma, pudo probarse la base del método, calculando el error cuadro por cuadro, comparándolos según su posición en las imágenes ground truth, así:

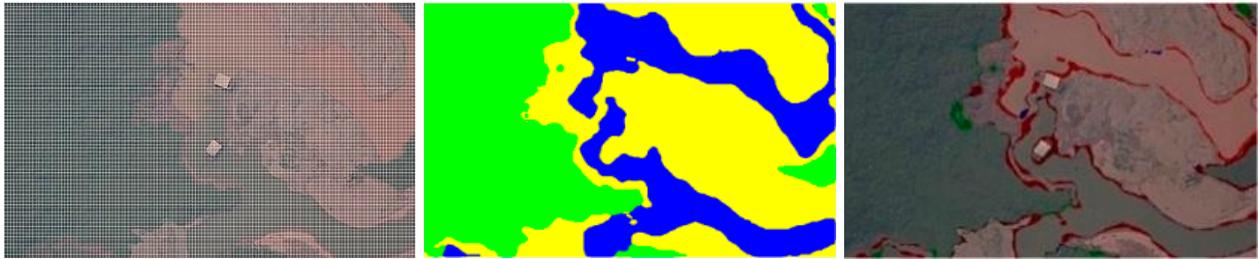
**Tabla 1.** Error de clasificación para cuadros de  $10 \times 10$  píxeles.

Muestra	o12M	o11M	Error	o12M	o11M	Error	o12M	o11M	Error
1	Vegetación	Vegetación	0	Agua	Agua	0	Otro	Otro	0
2	Vegetación	Agua	1	Agua	Agua	0	Otro	Otro	0
3	Vegetación	Vegetación	0	Agua	Agua	0	Otro	Otro	0
4	Vegetación	Vegetación	0	Agua	Agua	0	Otro	Otro	0
5	Vegetación	Vegetación	0	Agua	Agua	0	Otro	Otro	0
6	Vegetación	Vegetación	0	Agua	Agua	0	Otro	Otro	0
7	Vegetación	Vegetación	0	Agua	Agua	0	Otro	Otro	0
8	Vegetación	Agua	1	Agua	Agua	0	Otro	Otro	0
9	Vegetación	Vegetación	0	Agua	Agua	0	Otro	Otro	0
10	Vegetación	Vegetación	0	Agua	Agua	0	Otro	Otro	0

El error para muestras de la misma imagen de donde proviene el entrenamiento es 0 % y el de otra imagen se calcula en el 6,66 %.

**Análisis de error para imagen entrenada.** El desarrollo y puesta a punto del algoritmo de segmentación se hizo con la imagen 012M.tif, el método escogido para hacer el análisis de error consistió en obtener

resultados de la imagen con 10, 25 y 50 muestras supervisadas, con la certeza del experto de pertenecer a cada clase. Luego, se segmentó la imagen completa, recorriéndola en cuadros de 10 x 10 (figura 3). Se emplea esta técnica para reducir el tiempo de procesamiento sin aumentar demasiado el error de clasificación.



**Figura 3.** Derecha: selección de cuadrícula para clasificación. Centro: clasificación de la imagen en 3 clases. Derecha: imagen de píxeles clasificados erróneos (en rojo).

El análisis de error con la matriz de confusión mostró que la mayor probabilidad de error se presenta entre agua y otros, tabla 2. Los objetos que no hacen parte de un paisaje tropical normal como el desvío del cauce del río y deforestación alrededor de maquinaria

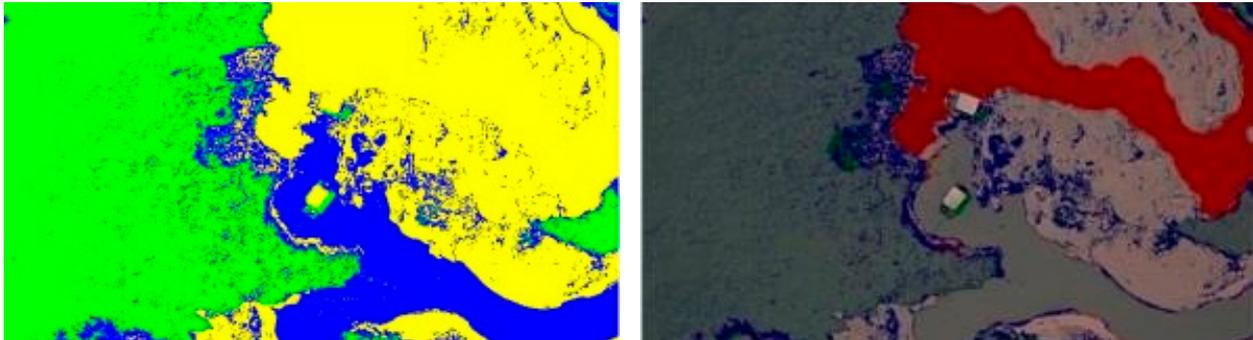
que es utilizada para minería ilegal son vistos en la clase otros y las zonas de la imagen donde ocurre, deben ser identificadas para hacer seguimiento del impacto económico, ambiental y social.

**Tabla 2.** Resultados de la clasificación para 50 muestras de imagen entrenada.

No.	CLASE PREDECIDA			Total x Clase	Píxeles correctos	Píxeles erróneos
	1	2	4			
1	97,1 %	0,0 %	2,9 %	42 %	97 %	3 %
2	0,6 %	85,1 %	14,3 %	25 %	85 %	15 %
4	1,8 %	0,5 %	97,7 %	33 %	98 %	2 %
Total				100 %		

Se hizo una comparación del algoritmo con otros de segmentación no supervisada, se realizó la misma prueba de error con la imagen 012 M.tif, pero esta vez clasificando individuales, a través mezcla de gaussianas, figura 4. En este caso, a cada píxel se asigna

en un grupo  $N$  de distribuciones gaussianas que representan cada clase con una media específica, los píxeles se estiman con la probabilidad de pertenecer a alguna clase cercana a su media.



**Figura 4.** Derecha: resultado de la segmentación con algoritmo no supervisado GMM, Izquierda: imagen de píxeles clasificados erróneos.

La matriz de confusión muestra alta probabilidad de error entre clase agua y otros, ya que el algoritmo reúne solamente

colores y no relación entre colores cercanos. Tabla 3.

**Tabla 3.** Resultados de la clasificación para 60 muestras de 2 imágenes.

No.	CLASE PREDECIDA			Total x Clase	Píxeles correctos	Píxeles erróneos
	1	2	4			
1	94,1 %	5,6 %	0,3 %	42 %	94 %	6 %
2	0,6 %	46,7 %	52,7 %	25 %	47 %	53 %
4	2,5 %	21,7 %	75,8 %	33 %	76 %	24 %
Total				%		

**Análisis de error para nuevas imágenes.**

Uno de los requerimientos de usuario final es poder enriquecer la base de datos para segmentar nuevas imágenes con datos existentes, para comprobar la utilidad, se prueba con 50 muestras obtenidas de otra imagen similar. En este caso, la 011M.tif, con las cuales se entrenó la máquina y aplicó el clasificador en la imagen 012M.tif y de la misma forma se realizó para GMM, escogiendo  $N=3$ . Al final, las muestras de ambas imágenes enriquecerán la población total para evaluación de nuevas entradas.

**Resultados**

**Error de entrenamiento.** Se muestra la recopilación de resultados para las pruebas realizadas con 5 imágenes seleccionadas y comparadas con la plantilla ground truth (tabla 4), clasificada por un experto, también se relaciona el porcentaje de píxeles pertenecientes a cada clase que hay en la imagen evaluada.

**Tabla 4.** Píxeles correctamente clasificados en 5 imágenes evaluadas con información de sí mismas.

INFORMACION DE LAS IMÁGENES			ERROR DE ENTRENAMIENTO			
CLASE	IMAGEN	Píxeles X Clase	10	25	50	GMM
	000MO.bmp	70%	55,53%	76,00%	88,25%	95,70%
	011M.tif	61%	91,73%	93,45%	95,07%	91,70%
VEGETACIÓN	012M.tif	42%	95,11%	97,58%	97,10%	94,10%
	030M.tif	88%	62,01%	95,88%	98,60%	75,50%
	031M.tif	54%	88,57%	92,67%	93,99%	82,30%
	000MO.bmp	11%	87,96%	73,86%	73,99%	95,50%
	011M.tif	22%	65,76%	79,61%	84,10%	95,60%
AGUA	012M.tif	25%	86,44%	81,28%	85,10%	46,70%
	030M.tif	9%	84,34%	84,99%	85,55%	98,50%
	031M.tif	26%	70,06%	91,28%	92,20%	99,80%
	000MO.bmp	19%	95,50%	96,93%	97,24%	55,60%
	011M.tif	17%	97,84%	97,03%	95,77%	61,50%
OTROS	012M.tif	33%	97,34%	97,32%	97,73%	75,80%
	030M.tif	4%	96,19%	92,63%	86,35%	3,30%
	031M.tif	20%	97,92%	97,72%	97,02%	0,40%

Se comprobó que con un porcentaje muy pequeño de muestras para entrenamiento en cada imagen para segmentación semántica, se pueden obtener resultados correctos, mediante una técnica simple de entender para un usuario final. El algoritmo permite además, tomar mayor cantidad de muestras para obtener precisión en la clasificación por encima del 74%, en todas las clases, segmentando eficientemente las imágenes por completo y guardando las muestras para análisis futuros, mejorando la precisión.

**Error de clasificación.** La técnica requiere poner a prueba imágenes nuevas con información previa, de tal forma que sobre las zonas de duda se haga un análisis más profundo. Se comparan los resultados para el caso de la segmentación semántica vs la segmentación por píxeles GMM. En este caso, cabe destacar que el análisis para ambos casos fue supervisado, pues

se tomaron decisiones a partir de la información previa.

En la tabla 5 se observaron mejores resultados en segmentación semántica, en la clase denominada 'otros' y se dieron resultados aceptables en las otras clases, teniendo en cuenta que la información de entrenamiento es escasa, sin embargo, la última columna muestra que teniendo mayor cantidad de muestras en todas las clases, el algoritmo alcanza excelente precisión en todas las clases, siendo la base de la inteligencia artificial una ventaja adicional para el método propuesto. Por estas razones, el algoritmo semántico se adapta mejor a las condiciones de cambio, en cuanto a la toma de las fotografías, ya que estas difieren en muchos factores, además es más simple de usar y tiene mayor efectividad, clasificando los grupos de desconocidos.

**Tabla 5.** Píxeles correctamente clasificados en 5 imágenes con información de otra imagen similar.

INFORMACIÓN DE LAS IMÁGENES			ERROR DE CLASIFICACIÓN CON NUEVAS IMÁGENES		ERROR CON TODOS LOS DATOS
CLASE	IMAGEN	Píxeles x Clase	50 MUESTRAS DE OTRA	GMM DE OTRA	750 MUESTRAS
	000MO.bmp	70%	21,31%	97,72%	95,45%
	011M.tif	61%	95,90%	99,79%	97,10%
VEGETACIÓN	012M.tif	42%	93,60%	96,53%	96,80%
	030M.tif	88%	18,70%	98,14%	98,99%
	031M.tif	54%	43,50%	22,79%	98,19%
	000MO.bmp	11%	24,95%	7,91%	84,13%

	o11M.tif	22%	82,40%	0,00%	92,56%
AGUA	o12M.tif	25%	32,30%	43,45%	86,37%
	o30M.tif	9%	38,70%	34,52%	88,52%
	o31M.tif	26%	0,40%	99,94%	93,67%
	oooMO.bmp	19%	94,77%	0,00%	89,94%
	o11M.tif	17%	95,10%	17,80%	91,73%
OTROS	o12M.tif	33%	99,20%	11,87%	97,76%
	o30M.tif	4%	99,90%	90,92%	84,76%
	o31M.tif	20%	96,90%	1,36%	94,32%

## Conclusiones

Los resultados demuestran que el clasificador funciona con baja probabilidad de error, sacando pocos datos de entrenamiento. Por otro lado, el algoritmo consume pocos recursos computacionales y más, teniendo en cuenta que se transformaron 928 800 datos de color con características en sus componentes RGB, en 150 muestras con 21 características obtenidas de su respuesta a los descriptores de covarianza en el entrenamiento y luego, aplicándolos en la clasificación. Los resultados fueron superiores al 80 % en segmentación correcta de las clases, haciendo que al algoritmo aplicado en la metodología de segmentación fuera relevante y funcional para el usuario final.

La técnica es aplicable y eficiente para la segmentación de imágenes de paisajes tropicales, además permite integrar otras características que ayudan a diferenciar más clases y con mayor precisión.

Una ventaja es que siempre se pueden mejorar los resultados, en la medida que se tenga mayor número de muestras de entrenamiento, dando la posibilidad de obtener resultados más exactos de los objetos contenidos en la imagen. Esta posibilidad no es aprovechable en los algoritmos de segmentación por píxel.

## Referencias Bibliográficas

- Arbeláez, P., Hariharan, B., Gu, C., Gupta, S., Bourdev, L. & Malik, J. *Semantic Segmentation using Regions and Parts*. Recuperado de [https://people.eecs.berkeley.edu/~sgupta/pdf/ahggbm\\_cvpr2012.pdf](https://people.eecs.berkeley.edu/~sgupta/pdf/ahggbm_cvpr2012.pdf).
- Avendaño, J., Parra, J. & Bayona, J. (2014). Segmentación y clasificación de imágenes SAR en zonas de inundación en Colombia, una herramienta computacional para prevención de desastres. *Universidad Antonio Nariño – Revista de Facultades de Ingeniería*, 4( 8) 24-38.

- Carmona, E. (2014). *Tutorial sobre máquinas vectoriales de soporte (SVM)*. Recuperado de [http://www.ia.uned.es/~ejcarmona/publicaciones/\[2013-Carmona\]%20SVM.pdf](http://www.ia.uned.es/~ejcarmona/publicaciones/[2013-Carmona]%20SVM.pdf)
- Castilla, G. (2003) *Object-oriented analysis of remote sensing images for land cover mapping: conceptual foundations and a segmentation method to derive a baseline partition for classification*. (Tesis Doctoral). Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros de Montes, España.
- Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke., Roth, U. & Schiele, B. *The Cityscapes Dataset for Semantic Urban Scene Understanding*, Recuperado de <https://arxiv.org/pdf/1604.01685.pdf>.
- Guevara, E. & Sánchez, O.A (2015). Segmentación de imágenes a color mediante técnicas de agrupamiento de datos empleando los algoritmos k-means y c-means. *Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas-IPN*, 51.
- Julier, S. & Uhlmann, J.K. (1996) *A general method for approximating nonlinear transformations of probability distributions*. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford.
- Kluckner, S., Mauthner, T., Roth, P. & Bischof, H. (2009). *Semantic classification in aerial imagery by integrating appearance and height information*. Reuperado de <http://www.icg.tugraz.at>.
- P, Ponce. (2010). *Inteligencia Artificial con Aplicaciones a la Ingeniería* (1ª edición) México: Alfaomega. ISBN: 978-607-7854-83-8.
- Pacheco, C., Del Valle, H. & Delrieux, C.(2014) *Revisión de algoritmos de segmentación desde una perspectiva orientada al sensado remoto de imágenes ópticas y de radar*. Recuperado de <http://www.researchgate.net/publication/269631987>.
- Tuzel, O., Porikli, F. & Meer, P. (2008). *Learning on lie groups for invariant detection and tracking*, CVPR. Recuperado de <https://pdfs.semanticscholar.org>
- Ventura, C. (2016). *La segmentación semántica y sus benchmarks*. Recuperado de <http://informatica.blogs.uoc.edu/2016/05/26/la-segmentacion-semantica-y-sus-benchmarks/>.