

Detección de aleta caudal de ballena jorobada por medio de mapas sobresalientes basado en mediana

Rosa I. Ramos-Arredondo¹, Blanca E. Carvajal-Gómez^{2,4}, Francisco J. Gallegos-Funes³, Miguel Ángel Castillo-Martínez¹

¹ Instituto Politécnico Nacional,
Escuela Superior de Ingeniería Mecánica y Eléctrica,
México

² Instituto Politécnico Nacional,
SEPI Escuela Superior de Cómputo,
México

³ Instituto Politécnico Nacional,
Escuela Superior de Ingeniería Mecánica y Eléctrica,
México

⁴ Instituto Politécnico Nacional,
Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas,
México

{rramosa1300@alumno.ipn.mx, becarvajal@ipn.mx,
fgallegosf@ipn.mx, mcastillom1503@alumno.ipn.mx}

Resumen. El procesamiento de imágenes es una herramienta en diferentes áreas de estudio como la visión computacional que en conjunto aporta en otras áreas como en la biología. En este trabajo proponemos la mejora de mapa sobresaliente basado en mediana para detectar la aleta caudal de la ballena jorobada en imágenes de avistamientos que fueron tomadas en altamar. En el trabajo hacemos un estudio comparativo con los métodos de Saliency Object Segmentation (SOS) y el método de mapa de atención sobresaliente de regiones escasamente salientes en una muestra de 40 imágenes. En donde el método propuesto obtiene una F-Mensure del 81%, siendo 45% mayor en comparación con los otros métodos. La métrica de intersection over union (IoU), intersección sobre la unión que sirve para medir la detección de objetos en problemas de visión computación se logra un promedio de 0.6169, obteniendo un 0.235 menos del SOS y 0.2981 del método de mapa de atención sobresaliente de regiones escasamente salientes.

Palabras clave: Procesamiento de imágenes, extracción de características, detección, ballena jorobada, identificación de animales.

Humpback Whale Tail Fin Detection Using Saliency Maps Based on Median

Abstract. Image processing is a tool in different areas of study such as computational vision that together provides other areas such as biology. In this work, we propose the improvement of the median-based saliency maps to detect

the tail fin of the humpback whale in images of sightings that were taken at sea. In the work, we do a comparative study with the saliency object segmentation (SOS) methods and the outstanding care map method of scarcely salient regions in a sample of 40 images. Where the proposed method obtains an F-Mensure of 81%, being 45% higher in compassion with the other methods. The intersection over union (IoU) metric used to measure the detection of objects in computer vision problems is achieved an average of 0.6169, obtaining 0.235 less than the SOS and 0.2981 of the outstanding care map method of scarcely salient regions.

Keywords: Image processing, feature extraction, detection, humpback whale, animal identification.

1. Introducción

El área de visión de computacional se enfoca en tareas que van en la detección de objetos, seguimiento de objetos, extracción de características, clasificación de elementos de forma automáticas, y para ellos hace uso de herramientas que en el procesamiento de imágenes se especializa en su análisis puntual [1-3]. Por ejemplo, el monitoreo de animales en su habita permite a especialista estudiar factores de comportamiento, población por mencionar y para los cetáceos que son animales de gran tamaño y su habita es inaccesibles controlar; la visión computacional y el procesamiento de imágenes apoya a su estudio [4-6]. Las imágenes que recolecta los biólogos o las personas que realizan observaciones en el mar permite obtenerlas; como es el caso de donde obtuvimos las muestras de imágenes de la ballena jorobada en el sitio de web que cuenta con una interfaz de usuarios para su recolección, análisis, compresión y divulgación de su estudio [7]. El algoritmo que proponemos hace uso de mapa sobresaliente; herramienta que se basa de la manera que la visión humana hace uso de tres detectores de sensibilidad en la retina, que da respuesta a la longitud de onda del campo visible, permitiendo al humano interpretar su entorno. El detector sobresaliente se basa en la habilidad automática de la vista humana de enfocar a elementos sobresalientes. Estos elementos pueden ser desde cambios de color, formas, texturas entre otros para obtener de una imagen (medio de entrada) la detección de elementos sobresalientes [8-11].

El presente trabajo se encuentra dividido en métodos y materiales donde se exponen la obtención de los mapas sobresalientes de los métodos a estudiar, así como el método propuesto en la detección de la aleta de la ballena jorobada. En la sección de resultados y discusión se muestra el estudio con las muestras de imágenes de la ballena jorobada en la detección de la aleta caudal por medio de mapa sobresalientes basada en mediana propuesto y la implementación de los otros métodos.

2. Métodos y materiales

En el estudio de cetáceos, el identificarlas es uno de los primeros pasos para determinar su estado de salud, comportamiento individual o en grupo, permitiendo identificar anomalías y esto a su vez un mayor conocimiento en las especies [5-6].

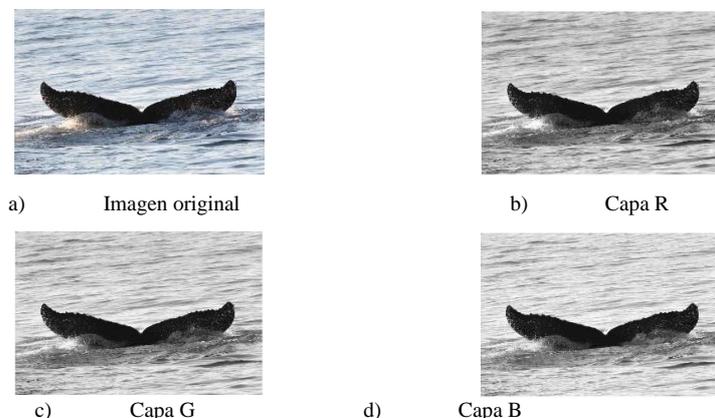


Fig. 1. Representación de una imagen digital RGB.

La ballena jorobada es un cetáceo que entre sus características representativas se encuentra las larga aletas pectorales, las protuberancias llamados folículos pilosos en la cabeza y mandíbula y su aleta caudal. Es un animal marino acrobático por la frecuencia que se impulsa a la superficie del agua deja ver su aleta caudal y eso hace que se una de las características para su identificación [4-5]. Las imágenes para el estudio se tomaron de happywhale, cuenta con una interfaz web para que personas e investigadores suban las fotografías de sus viajes de avistamientos, posteriormente un equipo de especialistas las identifique, permitiendo con ello hacer una red de conocimiento de avistamientos de animales marino [7]. Se tomaron 40 imágenes digitales RGB en plano general y/o primer plano de la ballena jorobada para implementar y realizar las pruebas del algoritmo propuesto, como con los otros métodos de mapa sobresalientes a analizar.

2.1. Imagen digital

Las imágenes obtenidas de la cámara digital son del tipo RGB o true-color; están compuesta por una matriz de pixeles y cada pixel está representado por tres capas y cada una tiene un valor escalar. Adicional a la composición de los pixeles de una imagen digital RGB se tiene la representación espacial por medio de filas y columna.

La representación de una imagen monocromática $I(x, y)$, es un arreglo de valores escalares de dos dimensiones, x, y ; son las coordenadas espaciales en donde contienen la intensidad (niveles de grises). El valor escalar que puede tomar la intensidad de una imagen es de 8bit y está dado entre 0 y 255(28). La dimensión de la imagen está representada por M , número de filas, N número de columnas y el valor de la intensidad es la función $I(x, y)$, (1):

$$I(x, y) = \begin{pmatrix} I(1,1) & I(1,2) & \dots & I(1,N) \\ f(2,1) & I(2,2) & \dots & I(2,N) \\ \vdots & \vdots & \dots & \vdots \\ I(M,1) & I(M,2) & \dots & I(M,N) \end{pmatrix}. \quad (1)$$



Fig. 2. Diagrama de la metodología del estudio.

Cada capa es un plano 2D que puede ser representado como una imagen monocromática; así C_i puede conformarse como $C_i = [R_i, G_i, B_i]$, en donde $0 \leq R_i, G_i, B_i \leq 255$.

2.2. Metodología de estudio

La metodología que proponemos es realizarle una transformación de espacio de color de RGB al espacio de color CIELab con un estándar de iluminación D65 (espacio de color que contempla como interpreta la visión humana). Después de la transformación del espacio de color a CIELab se realiza el estudio para la obtención de mapa sobresalientes. Estos mapas son la salida de los métodos de análisis para obtener las regiones sobresalientes y después se realiza la tarea de segmentación por umbralización. Por último, se realiza el cálculo de las métricas comparando los resultados obtenidos después de la segmentación y la segmentación manual de la base de datos.

2.3. Transformación de espacio de color CIE Lab

La dimensión de este espacio de color está compuesta por L que es la luminosidad, la componente a representa el tono del color y la componente b describen la saturación a lo largo de los ejes verde- rojo y azul-amarillo. La conversión entre CIE XYZ y CIELab está dado por los tres componentes de valores relativos y referenciado por el punto blanco D65:

$$\begin{aligned} L &= 116 \cdot Y' - 16, \\ a &= 50 \cdot (X' - Y'), \\ b &= 0.0193R + (Y' - Z'), \end{aligned}$$

donde, $X' = f_1\left(\frac{x}{x_{ref}}\right)$, $Y' = f_1\left(\frac{y}{y_{ref}}\right)$, $Z' = f_1\left(\frac{z}{z_{ref}}\right)$, (2)

$$f_1(T) = \begin{cases} T^{\frac{1}{3}} & \text{para } T > (24/116)^3, \\ 7.787 \cdot c + \frac{16}{116} & \text{para } T \leq (24/116)^3. \end{cases}$$

2.4. Método para la obtención de mapas de sobresaliente de objetos - Saliency Object Segmentation (SOS)

El método para la obtención de mapas de sobresalientes de objetos en combinación con el modelo de campo aleatorio condicional (CRF). Se basa en el deslizamiento de una ventana a través de la imagen, comparando en cada ventana el contraste entre la distribución de ciertas características en la ventana interna, con la distribución en el contorno de la ventana. La ventana deslizante W está dividida por dos partes; una ventana borde B y una ventana interior K . El mapa F contiene cada punto de característica que son representadas por x . La hipótesis del detector es que los puntos sobresalientes están en K y los puntos de B son partes del fondo. Siguiendo con la hipótesis $x \in K$; W toma valores de una variable aleatoria. Los elementos sobresalientes toman valores entre 0 y 1; la medición se define por una probabilidad condicional [9]:

$$S_o(x) = P(Z \in K | F(Z) \in Q_{F(x)}), \tag{3}$$

$$S_o(x) = \frac{P(F(x)|H_0)P(H_0)}{P(F(x)|H_0)P(H_0)+P(F(x)|H_1)P(H_1)}. \tag{4}$$

La función característica F toma los valores de la imagen en el espacio de color CIELab en el punto x , $L(x)$, $a(x)$, $b(x)$; se mapearía en $F(x)=(L(x), a(x), b(x))$. Para crear el mapa de sobresaliente se construyen tomando a diferentes escalas la ventana

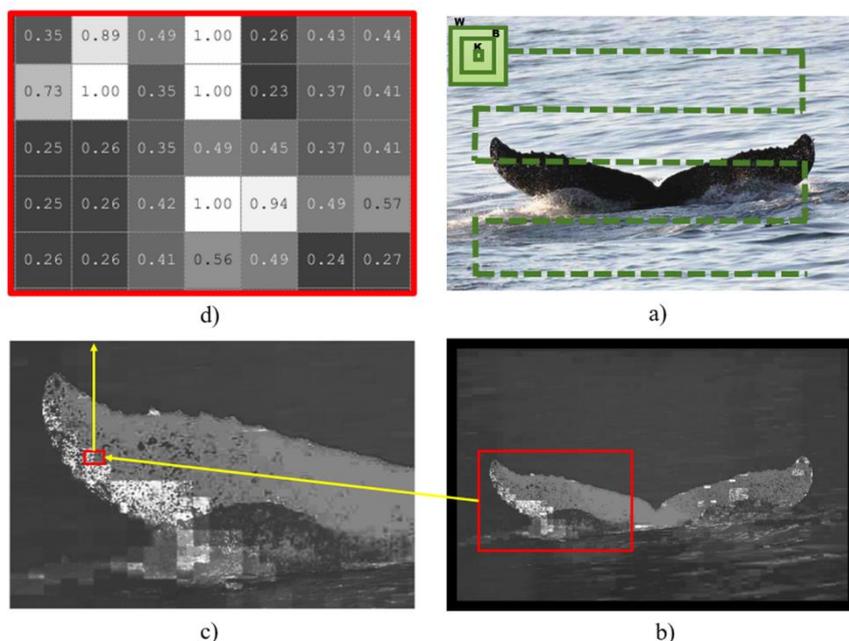


Fig. 3. Mapas sobresaliente de Saliency Object Segmentation (SOS). a) Imagen original b) Mapas sobresalientes c) Acercamiento de mapas sobresalientes d) Valores del acercamiento de mapas sobresalientes.

deslizante de W , por lo tanto, se desplaza en diferentes posiciones dentro de la imagen [8].

2.5. De mapas sobresalientes de regiones escasamente salientes

Este método considera una imagen digital como una firma compuesta por una mezcla de señales espaciales dispersas en un primer plano. El algoritmo analiza la composición de las capas y busca encontrar ciertos puntos y/o regiones de atención de la vista humana. Se propone la reducción de la imagen de estudio para obtener un acercamiento de la distancia perceptual del humano en combinación de la identificación de las regiones y/o puntos de mayor atención. Se define la función de signo de la transformada de coseno discreta (DTC) como un descriptor binario, que es la firma de la imagen en un primer plano:

$$x = f + b, \quad x, f, b \in \mathbb{R}^N, \quad (5)$$

donde, f es la representación del primer plano de la imagen o señal de la firma y b es la representación del fondo de la imagen:

$$\text{ImagenSignature}(x) = \text{sign}(\text{DTC}x). \quad (6)$$

Para obtener el mapa de sobresaliente se suprime el ruido con el suavizado de la imagen por medio de la kernel Gaussiano la cual permite proporcionar el tamaño del objeto de interés, donde, x^i es la representación de cada capa del espacio de color de la imagen y g es el kernel de suavizado Gaussiano:

$$m = g * \sum_i (\bar{x}^i \circ \bar{x}^i). \quad (7)$$

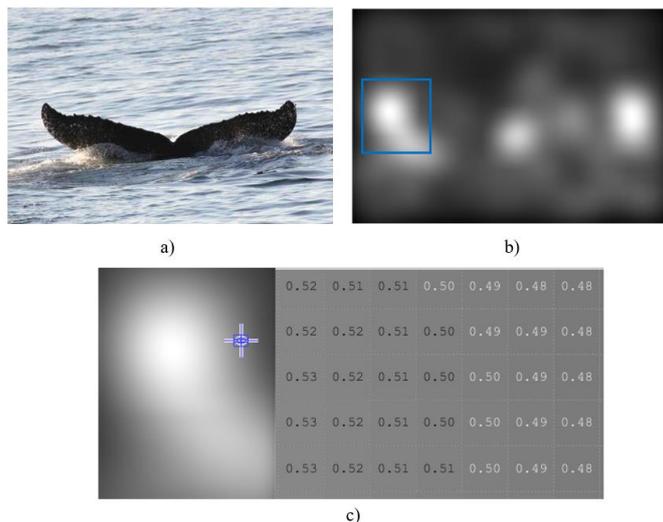


Fig. 4. Representación del método de mapas sobresalientes de regiones escasamente salientes a) Imagen digital original, b) Mapas sobresalientes, c) Acercamiento de mapas sobresalientes.

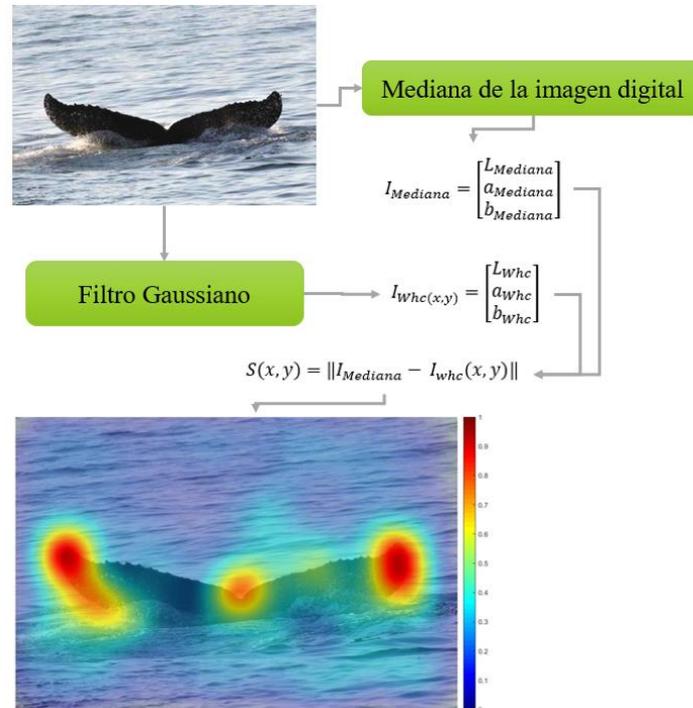


Fig. 5. Diagrama del método de mapas sobresalientes mediante mediana propuesto en una imagen digital.

2.6. Método de mapas sobresalientes basados en mediana

El método propuesto basado en mediana toma de base el método de detección de la región saliente ajustada a la frecuencia; utilizado para la detección de regiones sobresalientes que generen un mapa con límites definidos de los objetos sobresalientes en una imagen. El ajuste a la frecuencia se extiende al análisis de las características como el color y la iluminación. Permitiendo obtener la detección automática de las regiones útiles en aplicaciones y la segmentación de las imágenes [11]. El método de suavizado reduce a la mitad del espectro de frecuencia normalizado de la imagen. Además, de reducir sucesivamente las frecuencias espaciales de la imagen de entrada. Para obtener mapas sobresalientes se ocupa la transformación inversa de la transformada rápida de Fourier, usando un filtro promedio 5X5 esto es un equivalente $R(k)$:

$$R(k) = \ln|X(k)| - g_n * \ln|X(k)|, \quad (8)$$

donde, g_n es un kernel promedio de 5x5. Reduciendo (9) en una dimensión espectral:

$$R(k) = \frac{1}{25} \ln \left[\frac{|X(k)|^2}{|X(k-1)||X(k+1)|} \right]. \quad (9)$$

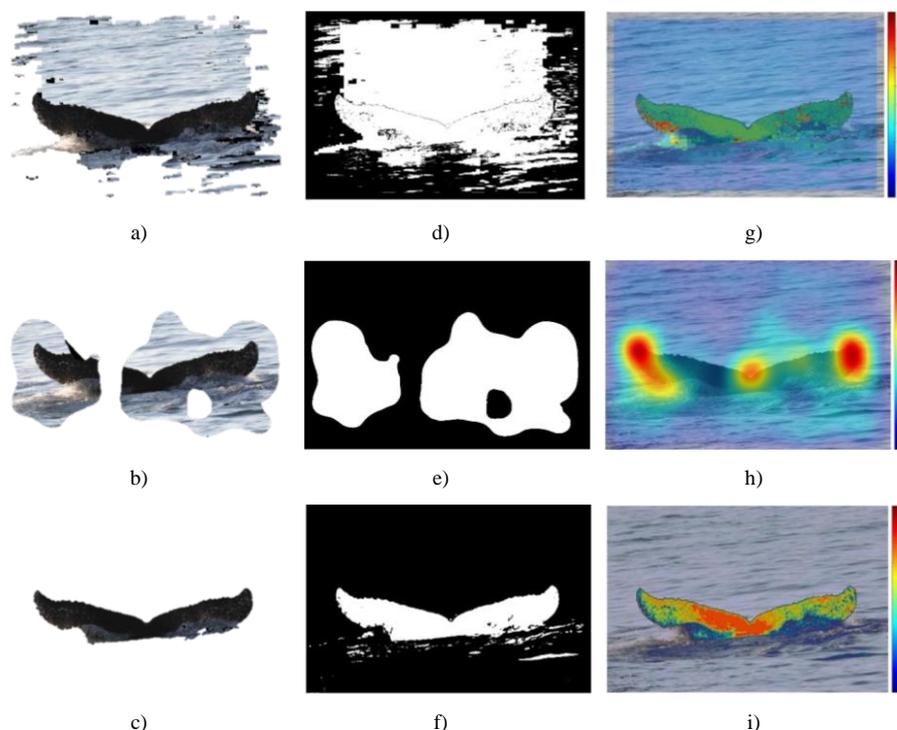


Fig. 6. Resultados de los métodos implementados en el estudio y el método propuesto a) Segmentación con el método SOS, b) Segmentación con los mapas sobresalientes de regiones escasamente salientes. c) Segmentación con mapas sobresalientes basados en mediana, d), e), f) Mascaras binarias después de aplicarles la umbralización promedio global a los métodos estudiados y el propuesto en ese orden g), h) i) Mapas sobresaliente de los métodos estudiados y el propuesto en ese orden.

La diferencia de Gaussiana permite la sustracción de una versión borrosa de una imagen original de otra versión menos borrosa de la original. Es implementada para la detección de cambios de iluminación cuando la desviación estándar del Gaussiano se encuentra entre 1:1.6:

$$\sum_{n=0}^{N-1} G(x, y, \rho^{n+1}\sigma) - G(x, y, \rho^n\sigma) = G(x, y, \sigma\rho^N) - G(x, y, \sigma). \quad (10)$$

Cuando σ_1 y σ_2 varia para mantener constante ρ en 1.6. Proporcionando que las regiones sobresalientes estarán cubiertas y no solo los bordes de estas mismas. Para obtener los mapas sobresalientes $S(x, y)$ de una I , imagen digital, se obtiene, $I_{Mediana}$ la mediana de los pixeles de la imagen y I_{whc} es la versión borrosa Gaussiana de la imagen original usando un kernel binomial de 5x5. Permitiendo la eliminación de detalles en la textura [12]. Al considerar las características de color e iluminación se puedes obtener el mapa sobresaliente por medio de (11):

$$S(x, y) = \|I_{Mediana} - I_{whc}(x, y)\|. \quad (11)$$

2.7. Segmentación por umbralización promedio global

Los mapas de sobresalientes de cada método estudiado contienen valores acotados inferior y superiormente; que representa su nivel de sobresaliente de la imagen. Para obtener una máscara que permita seleccionar la región y/u objeto de interés; es necesario definir un umbral que pueda segmentar las regiones de mayor valor de los mapas (11):

$$\begin{aligned} \text{Umbral} &= \mu_{F(x)} \\ \text{Mascara} &= \begin{cases} F(x) = 1 & x > \text{Umbral} \\ F(x) = 0 & x \leq \text{Umbral} \end{cases}, \end{aligned} \quad (12)$$

donde, F(x) es la máscara binaria que permite separar la aleta caudal de la ballena jorobada de la imagen digital.

3. Resultados

Por medio de los métodos de sobresalientes [8-9, 12] y el método propuesto basado en la mediana se realiza un estudio para obtener una región sobresaliente la cual es comparada con la obtenida del proceso de segmentación manual de las muestras de la aleta caudal.

4. Discusión

La muestra de imágenes se le realizó una segmentación manual con el propósito de analizar el método propuesto con los otros métodos a estudiar y con ello medir la respuesta de precisión, de sensibilidad, F-Measure e intersección sobre la unión (intersection over union - IoU) que nos indique cual puede aportar comportamiento para la detección de la aleta caudal de la ballena jorobada.

La precisión se calculó con la cardinalidad de los conjuntos de la máscara de mapas sobresalientes (SF) y la máscara de la segmentación manual (GF), entre la máscara de mapas sobresalientes.

$$\text{Precisión} = \frac{|SF \cap GF|}{|SF|}. \quad (13)$$

La sensibilidad se calculó con la cardinalidad de los conjuntos de la máscara del mapa de atención prominente (SF) y la máscara binaria de la segmentación manual (GF), entre GF:

$$\text{sensibilidad} = \frac{|SF \cap GF|}{|GF|}. \quad (14)$$

La F-Measure nos permite medir la relación entre precisión y sensibilidad, indicando que tan acertado es el método segmentado.

$$F_{\beta} = \frac{(1 + \beta^2) \text{Precisión} \times \text{Recall}}{\beta^2 \times \text{Precisión} + \text{Recall}}. \quad (15)$$

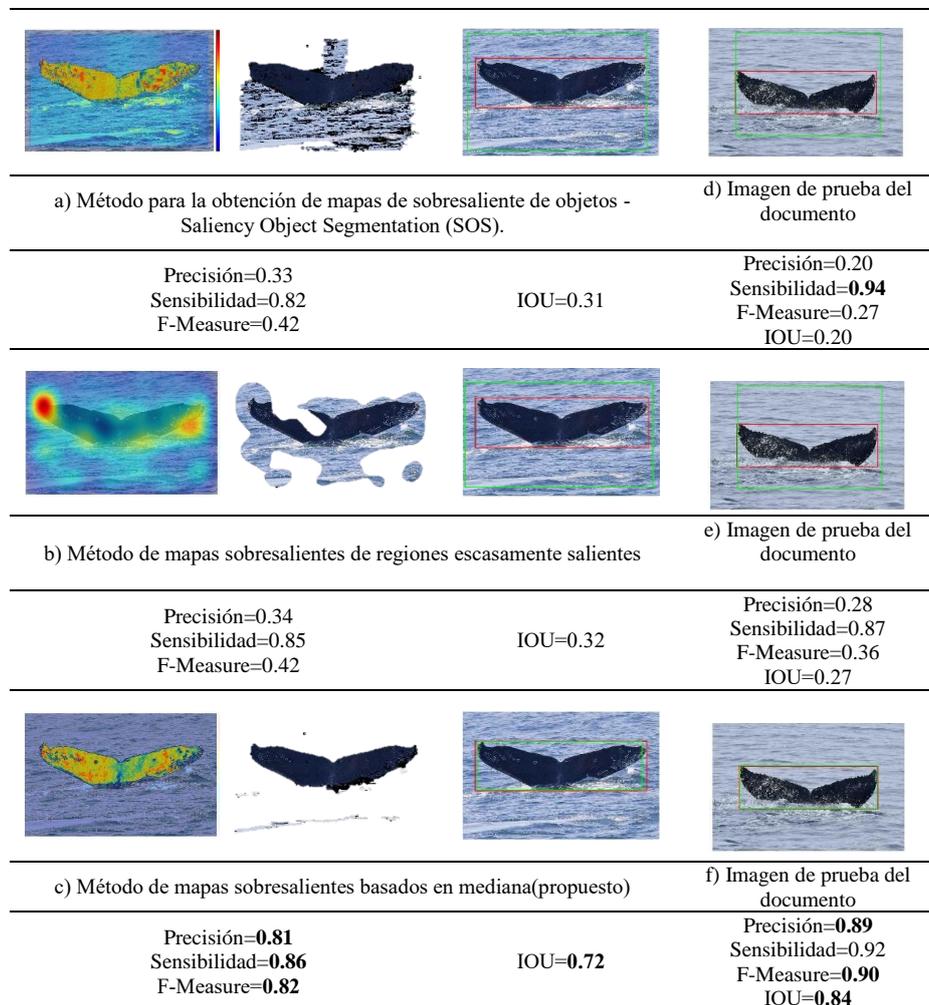


Fig. 7. Estudio de rendimiento entre dos imágenes del estudio. a) d) Método para la obtención de mapas de sobresaliente de objetos - Saliency Object Segmentation (SOS). b), e) Método de mapas sobresalientes de regiones escasamente salientes c) f) Método de mapas sobresalientes basados en mediana.

La intersección sobre la unión nos permite medir la detección de la aleta caudal de la ballena jorobada [13]:

$$IOU = \frac{|SF \cap GF|}{|SF \cup GF|}. \quad (16)$$

En la Fig. 7. se muestra las métricas comparativas de los métodos de mapas sobresalientes después de realizar la umbralización promedio global a dos imágenes. Las del lado derecho muestra de color rojo la detección de la aleta caudal por medio de la segmentación manual y en color verde la detección implementado los métodos del estudio. La columna extrema derecha es el resultado de la imagen con la que se ha

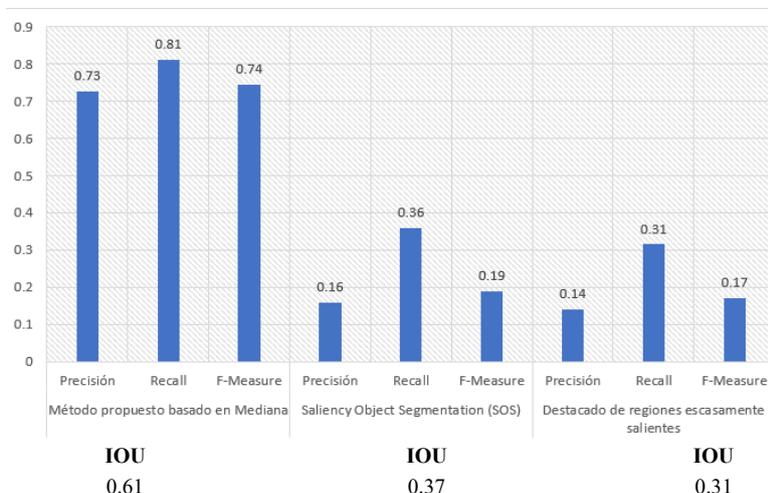


Fig. 8. Resultados de las muestras con relación de cada uno de los métodos.

desarrollado las pruebas visuales del documento por eso ya no fue necesario repetir sus implementaciones más que la detección. Por medio de las métricas los verdaderos positivos son donde SF se relaciona con GF, en caso contrario los falsos positivos del SF esta ubicado en donde no se encuentra la ballena jorobada. En la Fig. 8. se puede observar las comparativas entre los métodos y el propuesto donde el IOU es de 0.6, un valor que permite localizar a la ballena jorobada dentro de las imágenes del estudio, esto puede ser utilizado en el etiquetado de muestras para el entrenamiento de aprendizaje supervisado de redes neuronales convolucionales consume gran tiempo de personal capacitado, así como la calidad de la imagen son primordiales en el desarrollo de modelos de aprendizajes [14-15]. Los métodos comparativos del propuesto muestran un IOU bajo, esto deja que estos métodos se planteen con otro enfoque en donde obtuvieran un mayor rendimiento.

5. Conclusión

Los métodos con lo que se compara el método propuesto de mapas sobresalientes basados en mediana obtuvieron resultados menores en el estudio de 40 imágenes de la aleta caudal de la jorobada. Presentando este comportamiento en imágenes donde el mar se encuentra con una mayor uniformidad en el color y en los casos donde se tiene un mayor contraste entre los valores de las capas de color estos métodos tienen una mejor respuesta en la información que contiene los mapas de sobresaliente. En la detección de la aleta caudal se consigue obtener entre los métodos, pero el único que alcanza a pasar el 0.5 que es considerado una correcta detección es con el método propuesto. Para trabajos futuro se propone etapas en el preprocesamiento antes de obtener los mapas de sobresalientes porque la captura de las imágenes en los avistamientos no pueden ser homologadas y este factor beneficia en los métodos de mapas sobresalientes. Se queda abierto en la utilización como una herramienta en el

etiquetado de muestras que suelen ser necesario en el aprendizaje supervisado de redes neuronales.

Agradecimientos. Los autores agradecen al Instituto Politécnico Nacional de México (IPN) por la ayuda y el apoyo al proyecto de investigación con número 20200736.

Referencias

1. Petrou, M., Petrou, C.: Image processing: the fundamentals. John Wiley & Sons (2010)
2. Acharya, T., Ray, A.K.: Image processing: principles and applications. John Wiley & Sons (2005)
3. Burger, W., Burge, M.J., Burge, M.J., Burge, M.J.: Principles of digital image processing. pp. 221, London: Springer (2009)
4. Rangelova, E., Huiskes, M., Pauwels, E.J.: Towards computer-assisted photo-identification of humpback whales. In: Image Processing, (ICIP'04), International Conference on, 3, pp. 1727–1730 (2004)
5. Urian, K., Gorgone, A., Read, A., Balmer, B., Wells, R.S., Berggren, P., Hammond, P.S.: Recommendations for photo-identification methods used in capture-recapture models with cetaceans. *Marine Mammal Science*, 31(1), pp. 298–321 (2015)
6. Hoyt, E., Iñiguez, M.: Estado del avistamiento de cetáceos en América Latina. In: WDCS, Chippenham, 60 (2008)
7. Cheeseman, T., Southerland, K., Burns, D., Calambokidis, J., Friedlaender, A., Gabriele, C., Harrison, P., Jones, L., Neilsen, J., Schwarz, L., Straley, J., Clapham, P.: Tired eyes no more: a fully automated, high-accuracy photo-identification matching system for humpback whales. <https://happywhale.com> (2020)
8. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), pp. 1254–1259 (1998)
9. Rahtu, E., Kannala, J., Salo, M., Heikkilä, J.: Segmenting salient objects from images and videos. *Computer Vision–ECCV*, pp. 366–379 (2010)
10. Li, J., Gao, W.: Visual saliency computation: A machine learning perspective. LNCS 8408, Springer (2014)
11. Ko, B.C., Nam, J.Y.: Object-of-interest image segmentation based on human attention and semantic region clustering. *Journal of Optical Society of America A*, 23(10), pp. 2462–2470 (2006)
12. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: *Computer Vision and Pattern Recognition, IEEE conference on*, pp. 1597–1604 (2009)
13. Rezaeifighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: A metric and a loss for bounding box regression. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 658–666 (2019)
14. Ochoa-Ruiz, G., Angulo-Murillo, A.A., Ochoa-Zezzatti, A., Aguilar-Lobo, L.M., Vega-Fernández, J.A., Natraj, S.: An asphalt damage dataset and detection system based on retinanet for road conditions assessment. *Applied Sciences* 10(11), pp. 3974 (2020)
15. Dodge, S., Karam, L.: Understanding how image quality affects deep neural networks. In: *Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6 (2016)