XLIV Jornadas de Automática 2023

https://jautomatica.es

Análisis visual de escenas en entornos submarinos

Borja, C. and Murillo, A. C.[a,*]

[a]*I3A - Departamento de Informática e Ing. de Sistemas. Universidad de Zaragoza, C/ María de Luna, 1, 50018. Zaragoza, España.*

To cite this article: Borja, C., Murillo, A.C. 2023. Visual analysis of scenes in underwater environments. XLIV Jornadas de Automática, 837-842. https://doi.org/10.17979/spudc.9788497498609.837

Resumen

El uso de vehículos autónomos submarinos (AUV) supone una revolución para las tareas de monitorización del fondo marino. Sin embargo, propiedades como la atenuación de la luz o la turbidez del agua, propias de estos entornos, complican el procesado de las imágenes capturadas desde estos AUVs. Este trabajo estudia técnicas para mejorar la comprensión automática de imágenes monoculares de escenas submarinas. El sistema desarrollado no utiliza supervisión o entrenamiento adicional, sino que se construye a partir de técnicas existentes, aprendizaje *zero-shot* y algoritmos sencillos de procesamiento de imagen. El sistema combina una estimación de profundidad para las imágenes (utilizando monoUWNet Amitai et al. (2022), una adaptación a entornos submarinos del estado del arte en tareas de estimación de profundidad con imagen monocular) con la segmentación propuesta para separar las zonas de agua del resto de elementos de la escena. Los resultados del sistema presentado muestran como se consigue una interpretación más completa de la escena que con los algoritmos originales. El sistema propuesto permite segmentar con precisión las zonas de agua, facilitando la identificación de otros objetos de interés, como elementos suspendidos en el agua, que corresponden con peces u otros obstáculos móviles.

Palabras clave: Robótica inteligente, Percepción y sensores.

Visual scene understanding in underwater environments

Abstract

Using autonomous underwater vehicles (AUV) is a revolution for seabed monitoring. However, properties such as light attenuation or water turbidity, typical of these environments, complicate the processing of images captured from these AUVs. This work studies techniques to improve underwater scene understanding from monocular images. The developed system does not use additional supervised training, but builds on existing deep learning based techniques combined with simple image processing algorithms. The system combines depth estimation (using the monoUWNet Amitai et al. (2022) model, an adaptation to underwater environments of the state of the art in depth estimation from monocular image) with the proposed segmentation to separate water regions and the rest of the scene elements. The segmentation results obtained show that more complete scene information is achieved than with the original algorithms. The proposed system allows to accurately segment the water regions and facilitates the detection of other objects of interest, such as elements suspended in the water, corresponding with fish or other moving obstacles.

Keywords: Intelligent Robotics, Perception and Sensing.

1. Introducción

La monitorización de los mares y océanos es una práctica fundamental para la comprensión de los ecosistemas marinos, los procesos geológicos y la biodiversidad marina. El uso de robots autónomos en tareas de monitorización del fondo submarino está revolucionado la manera en la que se llevan a cabo estos estudios, permitiendo la obtención de datos en tiempo real de manera automatizada y en áreas de difícil acceso. La Figura 1 muestra un ejemplo de vehículo autónomo submarino (AUV, del inglés *Autonomous Underwater Vehicle*) y de imágenes de escenas submarinas.

*Autor para correspondencia: 800675@unizar.es

La instalación en estos AUVs de cámaras u otros sensores, como ultrasonidos, permiten, entre otras tareas, inspeccionar estructuras submarinas, explorar yacimientos arqueológicos y mapear el fondo marino. Resulta de gran interés automatizar parcial o completamente estas tareas, y los desarrollos recientes de visión por computador muestran grandes avances en, por ejemplo, sistemas de reconstrucción del color en imágenes submarinas Akkaynak and Treibitz (2019) o sistemas de *tracking* para AUVs Kumar et al. (2018).

Figura 1: Ejemplo de AUV ALICE Gutnik et al. (2022) y de imágenes de escenas submarinas cedidas por el Laboratorio VISEAON (`https://www.viseaon.haifa.ac.il/`) captadas por AUV.

Sin embargo, características del medio subacuático como la turbidez del agua o la atenuación de la luz complican el estudio y procesado de estas imágenes, y hacen que métodos existentes de visión por computador no funcionen con la misma fiabilidad en estos entornos. En la Figura 1 se muestran algunos ejemplos de imágenes captadas desde AUVs. En ellas se pueden observar estructuras y ruinas submarinas, y se aprecia cómo las condiciones subacuáticas dificultan la visibilidad en dichas escenas.

La contribución principal de este trabajo es un sistema para el análisis automático del contenido de imágenes en entornos submarinos. El sistema propuesto combina modelos de deep learning existentes con algoritmos sencillos de post-procesado para obtener un sistema mejorado [1] que salva algunas de las dificultades específicas que se encuentran en este entorno. Por un lado, se han estudiado modelos capaces de estimar la profundidad, respecto a la cámara, a la que están los distintos elementos de una imagen submarina, tomada por una cámara monocular. En particular, se ha considerado el estado del arte para esta tarea, con métodos existentes basados en redes neuronales profundas, concretamente monodepth2 Godard et al. (2019) y monoUWNet Amitai et al. (2022). Este tipo de modelos, que serán explicados en detalle en la siguiente sección, reciben como entrada una imagen RGB y devuelven un valor de profundidad para cada píxel de la imagen. Con esta información se puede generar un modelo sencillo (nube de puntos) 3D de la escena. Por otro lado, se han estudiado distintos métodos de segmentación de imagen para intentar separar las partes más o menos relevantes de la imagen, sin necesidad de entrenar nuevos algoritmos supervisados específicos. Se han estudiado métodos más tradicionales de visión por computador, superpíxeles Van den Bergh et al. (2012), y otros más recientes de segmentación basada en redes neuronales profundas Kirillov et al. (2023).

El sistema propuesto se ha evaluado con un conjunto heterogéneo de imágenes reales submarinas de distintas fuentes, para observar que información útil sobre la escena se puede conseguir sin necesidad de entrenar nuevos modelos.

[1] `https://github.com/cborjamoreno/underwater_analysis.git`

2. Trabajo relacionado

Modelos de estimación de profundidad. En la literatura encontramos numerosos trabajos para estimar la profundidad de la escena representada en una imagen como HiMODE Junayed et al. (2022), Godard et al. (2017) o DINOv2 Oquab et al. (2023). Uno de los modelos pioneros con resultados exitosos para la tarea de estimación de profundidad a partir de una imagen monocular mediante técnicas de deep learning es monodepth2 Godard et al. (2019). Siguiendo algunas de sus ideas, pero adaptado especialmente para entornos submarinos, y por tanto de especial interés para este trabajo, encontramos el modelo monoUWNet. Los dos modelos reciben una imagen RGB monocular como entrada y devuelven una matriz de la misma dimensión en la que cada elemento (x, y) de la matriz contiene el valor de profundidad estimado del píxel (x, y) de la imagen. Esta estimación es un valor a escala, es decir, no se mide en unidades de distancia reales.

La estimación de distancias a partir de imágenes monoculares se realiza utilizando una red neuronal profunda entrenada a partir de un conjunto de pares de imágenes (modelo supervisado). Las dos imágenes de cada par se corresponden con la imagen sobre la que se quiere sacar la estimación y su *ground-truth* (GT) creado a partir de mediciones reales con sensores de profundidad. Dada la difícil adquisición de distancias GT (especialmente en entornos submarinos), tanto monodepth2 como monoUWNet proponen utilizar frames consecutivos para llevar a cabo un entrenamiento auto-supervisado. Utilizando estos frames se obtienen diferentes poses de la misma escena, permitiendo sacar una medida de profundidad teniendo en cuenta la estimación del movimiento entre frames. No obstante, pese a sus similitudes, monoUWNet incorpora una serie de mejoras y adaptaciones importantes enfocadas a mejorar los resultados en escenas submarinas: MonoUWNet está basado en DiffNet Zhou et al. (2021), estado del arte en estimación de profundidad monocular auto-supervisada, pero implementa una solución para disminuir los errores de estimación de profundidad que DiffNet produce en secciones de la imagen de color plano, como el cielo o el mar. Además, para mejorar el rendimiento de la estimación en imágenes con variaciones de iluminación, monoUWNet utiliza *data augmentation* aplicando filtrado homomórfico Adelmann (1998) en el conjunto de imágenes de entrenamiento. Como se demuestra en el trabajo original de monoUWNet, sus resultados son mucho más precisos en imágenes submarinas, por lo tanto es el modelo que se utilizará como base en el sistema diseñado en este trabajo.

Segmentación de imagen. Por un lado, para este trabajo resultan relevantes los métodos de segmentación no supervisada que agrupan las zonas similares de la imagen en segmentos o superpíxeles. Alguno de los métodos más conocidos son SLIC Achanta et al. (2010) o SEEDS Van den Bergh et al. (2012). En nuestro sistema se va a trabajar con SEEDS (*Superpixels Extracted via Energy-Driven Sampling*), por presentar un buen compromiso entre precisión, facilidad de uso y rapidez. Este método comienza creando una malla cuadricular sobre toda la imagen, siendo cada uno de los cuadrados un superpíxel.

Tras esto el algoritmo hace una optimización donde se favorece la homogeneidad de la distribución del color dentro de cada uno de los superpíxeles. Como resultado, cada superpíxel intercambia píxeles con sus vecinos cambiando la forma del borde de cada superpíxel hasta alcanzar la segmentación final.

Por otro lado, también resultan de gran interés los modelos de segmentación semántica basados en deep learning, con resultados del estado del arte y gran capacidad de generalización demostrados en los últimos años Feng et al. (2020), Yang and Yu (2021), Girshick et al. (2014). En este trabajo se hace uso del reciente modelo *Segment Anything Model* (SAM) Kirillov et al. (2023), ya que es capaz de obtener una segmentación genérica de los posibles objetos de la imagen, sin necesidad de conocer clases concretas de objetos, que funciona de manera muy eficaz sin ningún tipo de entrenamiento adicional.

3. Sistema propuesto para análisis de la escena submarina

El sistema desarrollado en este trabajo consta de tres módulos principales descritos a continuación. Recibe como entrada una imagen monocular, y devuelve un modelo sencillo en 3D (nube de puntos) filtrado y con anotaciones de cierta información de interés sobre la escena.

Módulo 3D. Este módulo obtiene la estimación de profundidad a partir de imágenes monoculares, utilizando monoUWNet, y genera y maneja nubes de puntos 3D a partir de dicha información.

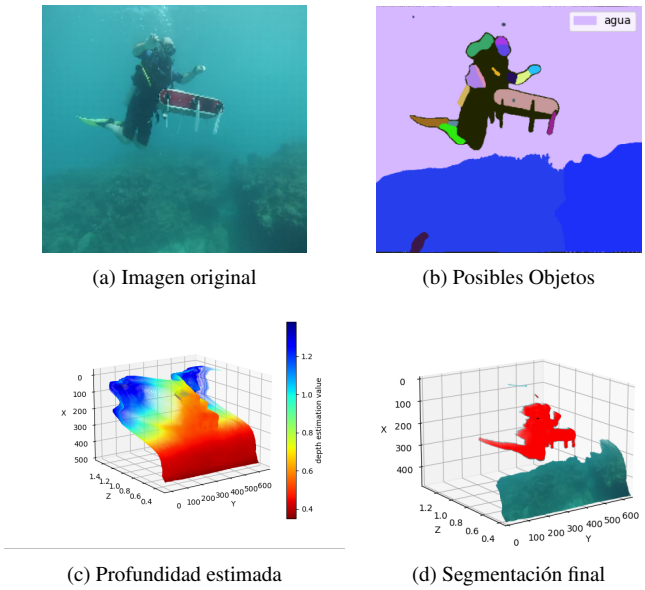
Módulo de segmentación. Este módulo aplica diferentes métodos de segmentación, sin ningún tipo de entrenamiento adicional, para separar ciertas zonas de interés de la escena.

Principalmente, se propone cómo identificar la zona de la imagen que corresponde con el agua, ya que genera ruido, por ejemplo en las reconstrucciones 3D obtenidas en el modulo anterior. Esto se consigue estimando primero los valores de profundidad de la imagen y después binarizando el resultado con un *threshold*, fijo para todos los experimentos. Así se clasifica cada píxel como "agua" si su valor de profundidad supera el valor del *threshold* o como "escena" en caso contrario.

Además, se propone un algoritmo sencillo para identificar elementos que se encuentran flotando, suspendidos, en el agua, que resultan muy relevantes para el análisis de la escena. Este algoritmo se aplica sobre la segmentación binaria (agua/escena) de la imagen como la explicada en el párrafo anterior y consiste en la localización de contornos cerrados en dicha imagen. En las Figuras 4 y 5 se muestran varios ejemplos de los resultados la segmentación binaria y de objetos flotantes.

Módulo Final. Este módulo combina los módulos anteriores para obtener la segmentación final. Esta segmentación consiste en una nube de puntos limpia de puntos no relevantes, i.e., los identificados como agua. Además, se marcan posibles elementos de interés, mediante los objetos en suspensión, que son posibles obstáculos o elementos interesantes para tareas de monitorización, por ejemplo animales u otros obstáculos móviles.

La Figura 3 muestra un ejemplo de la segmentación final de la nube de puntos 3D estimada.



(a) Imagen original (b) Posibles Objetos

(c) Profundidad estimada (d) Segmentación final

Figura 3: Ejemplo del modelo 3D anotado final obtenido por el sistema propuesto dada la imagen original (a). (b) posible segmentación de objetos de modelo SAM. (c) puntos coloreados según el valor de profundidad estimados con monoUWNet. (d) puntos con el color original en la imagen, salvo los objetos flotantes identificados, que se remarcan en rojo.

4. Experimentos

Este capítulo describe los experimentos realizados para analizar y evaluar el sistema desarrollado.

4.1. *Configuración de los experimentos*

Datasets. Para conseguir un conjunto de evaluación heterogéneo, se han utilizado imágenes de varias fuentes para conseguir un conjunto de 32 imágenes variadas. Como datos más relevantes, se han utilizado varias imágenes (12) captadas por un AUV (**VISEAON-dataset**), que son imágenes cedidas por el laboratorio de investigación VISEAON [2]. Para tener algunas imágenes con *ground truth* para las zonas de agua, se han incluido 20 imágenes[3] diversas del **SUIM-dataset**, Semantic Segmentation of Underwater Imagery Islam et al. (2020).Estos datos incluyen tienen etiquetas para 8 categorías, entre ellas, la etiqueta de "agua", que es la que se utilizará para evaluar algunos de nuestros experimentos. Se ha elegido este dataset por su variedad de escenas subacuáticas, tanto en los propios objetos (buzos, gran variedad de fauna, embarcaciones hundidas, etc) como en la perspectiva de la cámara y el tono de color.

Entorno de experimentación. Todos los experimentos se han ejecutado en un equipo con un microprocesador AMD Ryzen 5 5600X, 32GB de memoria RAM y una tarjeta gráfica NVIDIA GeForce RTX 3060.

[2] `https://www.viseaon.haifa.ac.il/`

[3] Imágenes usadas del SUIM-dataset: *d_r_47*, *d_r_58*, *d_r_84*, *d_r_166*, *d_r_182*, *d_r_233*, *d_r_465*, *f_r_209*, *f_r_1059*, *f_r_1229*, *f_r_1276*, *f_r_1892*, *f_r_1920*, *f_r_1992*, *n_l_0*, *w_r_4*, *w_r_10*, *w_r_22*, *w_r_84*, *w_r_111*.

Imagen	SPX			Depth			Depth+SAM		
	Prec.	Recall	t(s)	Prec.	Recall	t(s)	Prec.	Recall	t(s)
Media	**0.81**	**0.69**	**1.11**	**0.86**	**0.86**	**8.31**	**0.91**	**0.93**	**76.30**
Mediana	**0.89**	**0.84**	-	**0.92**	**0.86**	-	**0.98**	**0.95**	-

Tabla 1: Comparación de precisión, recall (exhaustividad) y tiempo de ejecución de los tres métodos para segmentación del agua. Se muestra la media para las 20 imágenes utilizadas del SUIM-dataset.

4.2. *Evaluación cuantitativa de segmentación binaria (agua)*

Este experimento evalúa los resultados de los diferentes métodos implementados para la tarea de segmentación binaria de imágenes submarinas. En particular, comparamos los resultados obtenidos de usar distintas alternativas:

- **SPX**: Segmentación binaria utilizando superpíxeles obtenida de la siguiente manera. Se generan los superpíxeles en la imagen usando SEEDS, y se establece un rango de color en HSV que englobe los colores del agua. Para cada superpíxel se calcula su color medio, y se identifica como agua o no, dependiendo de si su color medio cae en el rango HSV establecido.
- **Depth**: Segmentación mediante *threshold* binario sobre los datos la estimación de profundidad (más detalles en el *módulo de segmentación* de la Sección 3.
- **Depth+SAM**: Combina la estimación de profundidad de la imagen con el resultado de la segmentación de objetos obtenida del modelo SAM. Este método corresponde con el método del *Módulo final* descrito en la sección 3, coloreando todos los segmentos que no pertenecen al agua de la misma forma.

Para hacer una evaluación cuantitativa, se utilizarán 20 imágenes del conjunto de test del SUIM-dataset, ya que el etiquetado que ofrece permite calcular medidas de precisión (*Prec.*) y exhaustividad (*Recall*) de la segmentación de la masa de agua. En la tabla 1 se muestran estos resultados. Depth+SAM obtiene los resultados más precisos, por encima de 0.9. La mitad de imágenes tienen una precisión mayor o igual que 0.98 y una exhaustividad mayor o igual que 0.95. El método SPX tiene una media de exhaustividad más de 0.1 más baja que los otros dos métodos. Esto se debe a lo sensible que es el método a ligeros cambios en el tono de los colores de las imágenes. Depth ofrece buenos resultados, con pocos valores espurios. Si tenemos en cuanta el tiempo de ejecución, SPX es el más rápido con una media de 1.11 segundos, seguido de Depth con 8.31 y dejando como último a Depth+SAM con más de un minuto de media. Aunque ninguna de las implementaciones esta especialmente optimizada, considerando los 8.31 segundos de media de tiempo de ejecución de la opción *Depth*, se puede considerar el más adecuado para aplicaciones con restricciones computacionales.

En la Figura 4 se muestran ejemplos de los resultados obtenidos con los tres métodos de segmentación binaria. Cualitativamente, tambein se ve que Depth+SAM ofrece una segmentación mucho más fiel a la realidad que Depth y SPX. Entre SPX y Depth, se puede observar como SPX, pese a segmentar bastante bien en algunos casos, tiene problemas con la turbidez del agua y con algunos tipos de iluminación en los que el agua adopta colores más verdosos. En cambio Depth segmenta de manera más robusta la escena, aunque tiene poca precisión en el borde entre la masa de agua y la escena y puede pasar por alto objetos lejanos.

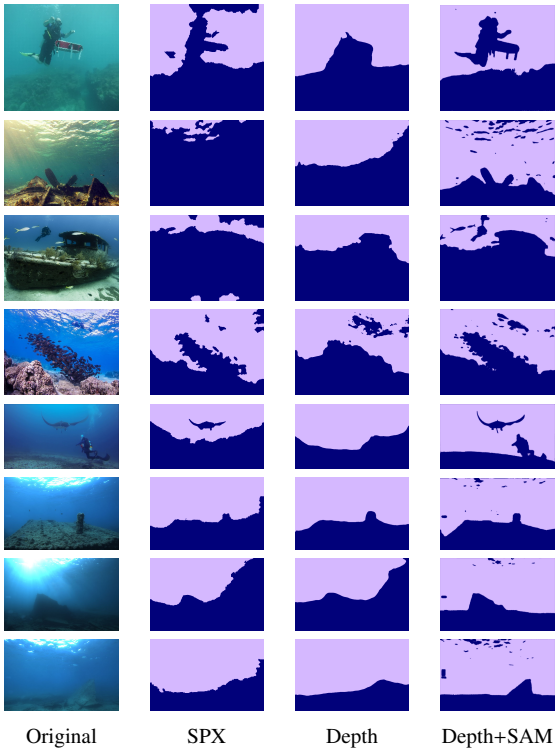


Figura 4: Comparación de los resultados de segmentación binaria utilizando SPX, Depth y Depth+SAM. Se puede apreciar como Depth+SAM consigue segmentar la escena y el agua considerablemente mejor que SPX y Depth.

4.3. *Evaluaciones cualitativas del sistema final*

En este apartado se muestran resultados cualitativos del método de segmentación de objetos flotantes y del sistema completo.

Segmentación de objetos flotantes. En la Figura 5 se muestran ejemplos de los resultados obtenidos con el método de segmentación de objetos flotantes. Esta alternativa sencilla para segmentar objetos en suspensión consigue resultados prometedores. Generalmente detecta correctamente los objetos flotando en el agua, aunque existen situaciones concretas en las que se suelen producir falsos positivos y falsos negativos. Los falsos positivos suelen darse por brillos en el agua o cambios de tonalidad en el color del agua. Los falsos negativos suelen deberse a que el objeto flotante se encuentra contiguo en la imagen al fondo marino (y por tanto no existe contorno para poder diferenciarlo), o el objeto se encuentra en mitad de alguno de los bordes de la imagen, haciendo que no sea un contorno cerrado.

La perspectiva de la cámara es una limitación por lo tnato, ya que si la imagen es sacada apuntando hacia abajo, los objetos no se verán rodeados de agua.

Evaluación del sistema completo. Estos resultados ilustran las mejoras conseguidas respecto a la comprensión del contenido de la escena con los algoritmos básicos frente al sistema completo. En la Figura 6 se puede observar un mosaico con los distintos resultados intermedios que obtiene el sistema para diferentes imágenes. Primero se hace una estimación de profundidad, añadiendo una tercera dimensión a los datos. Sin embargo, estas nubes de puntos 3D contienen puntos de la masa de agua del mar que no son de interés y dificultan el análisis. Se hace una segmentación de la imagen separando el agua del resto de elementos de la escena, y se combina el resultado con la nube de puntos, eliminando los puntos de agua. Además, se puede aplicar el método de segmentación de objetos flotantes para resaltar objetos de interés en la nube de puntos, como estos posibles obstáculos o elementos móviles a monitorizar.

Tras esta evaluación, se verifica que los resultados del sistema favorecen considerablemente la comprensión automatizada de las escenas submarinas, consiguiendo de manera automática información relevante para monitorización o para sistemas autónomos que por ejemplo vayan a navegar o realizar tareas de seguimiento en estos entornos.

También se han observado algunas limitaciones del modelo de estimación de profundidad para cierto tipo de escenas. Pese a que el modelo utilizado está optimizado para imágenes submarinas, a menudo presenta problemas en la estimación de escenas que contengan elementos flotantes, estimando incorrectamente la profundidad de los elementos flotantes en comparación con la profundidad del suelo. Esto puede ser debido a que estos modelos no han sido entrenados con datos similares, y no esperan encontrar nada "suspendido" en el agua. Con lo cual, el modelo no es capaz de estimar bien la profundidad relativa de estos elementos que no "tocan" el suelo.

5. Conclusiones

Este trabajo presenta un sistema desarrollado para mejorar la comprensión automática de escenas submarinas, combinando información de profundidad con información semántica. Esta información se ha obtenido sin necesidad de re-entrenar ningún modelo adicional, con algoritmos sencillos de post procesado combinados con modelos del estado del arte basados en deep learning.

Como pasos futuros, se plantean mejoras para segmentar objetos concretos que pudieran ser de interés, como corales o especies concretas de peces, combinando el trabajo realizado con otros trabajos más específicos. Por otro lado, se podría integrar en un sistema robótico en tareas de navegación, aprovechando la información de profundidad de la nube de puntos y detectando posibles obstáculos como los elementos flotantes. Tal y como se ha comentado en la sección anterior, para esto último sería interesante utilizar el método de segmentación basado en profundidad, teniendo en cuenta su mejor rendimiento en tiempo de ejecución.

Figura 5: Ejemplos de segmentación de objetos flotantes obtenidos. Para cada par de imágenes, se muestra la imagen original y la segmentación. En los dos ejemplos de la última fila, se pueden ver marcados casos particulares de esta segmentación. Con un cuadrado verde, se resaltan falsos negativos en la segmentación. Esto se debe a que el contorno de los peces se junta con el contorno del coral, evitando por tanto que se forme un contorno cerrado. Por otro lado, un cuadrado rojo resalta como los brillos del agua producen falsos positivos.

Agradecimientos

Este trabajo ha sido financiado parcialmente por FEDER/-Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación proyecto PID2021-125514NB-I00, y DGA T45_23R/FSE. Los autores agradecen el apoyo del Laboratorio VISEAON, de la Universidad de Haifa, Israel, a lo largo del trabajo. Además, el proyecto se ha desarrollado en el marco de la beca para *Prácticas de Estudiantes de Grado Universitario en el marco del TFG* del I3A.

Referencias

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2010. Slic superpixels. Tech. rep.

Adelmann, H. G., 1998. Butterworth equations for homomorphic filtering of images. Computers in Biology and Medicine 28 (2), 169–181.

Akkaynak, D., Treibitz, T., 2019. Sea-thru: A method for removing water from underwater images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1682–1691.

Amitai, S., Klein, I., Treibitz, T., 2022. Self-supervised monocular depth underwater. arXiv preprint arXiv:2210.03206.

Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., Wiesbeck, W., Dietmayer, K., 2020. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. IEEE Transactions on Intelligent Transportation Systems 22 (3), 1341–1360.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 580–587.

Godard, C., Mac Aodha, O., Brostow, G. J., July 2017. Unsupervised monocular depth estimation with left-right consistency. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

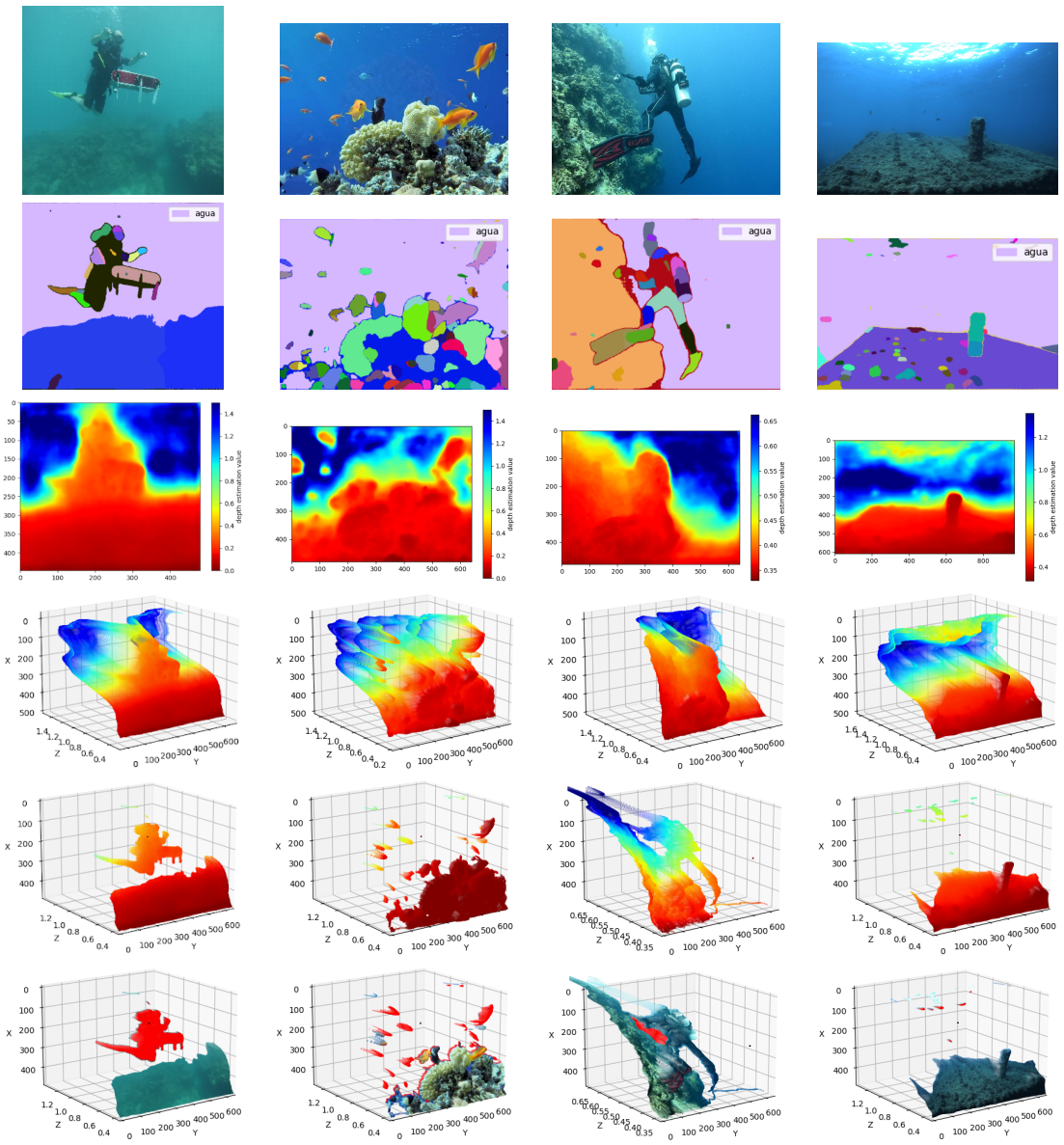


Figura 6: Resultados obtenidos por el sistema de análisis de escenas presentado, en distintos pasos intermedios, para distintas imágenes de los datasets utilizados. Cada fila (de arriba a abajo) corresponde con: foto original, segmentación final, estimación de profundidad, nube de puntos, nube de puntos eliminando puntos de agua, nube de puntos destacando objetos flotantes (puntos en rojo) y con colores de la imagen original, respectivamente.

Godard, C., Mac Aodha, O., Firman, M., Brostow, G. J., 2019. Digging into self-supervised monocular depth estimation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3828–3838.

Gutnik, Y., Avni, A., Treibitz, T., Groper, M., 2022. On the adaptation of an auv into a dedicated platform for close range imaging survey missions. Journal of Marine Science and Engineering 10 (7), 974.

Islam, M. J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., Enan, S. S., Sattar, J., 2020. Semantic segmentation of underwater imagery: Dataset and benchmark. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 1769–1776.

Junayed, M. S., Sadeghzadeh, A., Islam, M. B., Wong, L.-K., Aydın, T., 2022. Himode: A hybrid monocular omnidirectional depth estimation model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5212–5221.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al., 2023. Segment anything. arXiv preprint arXiv:2304.02643.

Kumar, G. S., Painumgal, U. V., Kumar, M. C., Rajesh, K., 2018. Autonomous underwater vehicle for vision based tracking. Procedia computer science 133, 169–180.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al., 2023. Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193.

Van den Bergh, M., Boix, X., Roig, G., de Capitani, B., Van Gool, L., 2012. Seeds: Superpixels extracted via energy-driven sampling. ECCV (7) 7578, 13–26.

Yang, R., Yu, Y., 2021. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. Frontiers in oncology 11, 638182.

Zhou, H., Greenwood, D., Taylor, S., 2021. Self-supervised monocular depth estimation with internal feature fusion. arXiv preprint arXiv:2110.09482.