Segmentación Multimodal para detección de Animales

Jorge Urbón Burgos
777295@unizar.es

Supervisor: Rosario Aragües
raragues@unizar.es

Co-Supervisor: Jesús Bermúdez
bermudez@unizar.es

11 de diciembre de 2024

Resumen

Índice general

Capítulo 1

Introduction

Image segmentation has been a fundamental problem in computer vision tasks since the early days of the field. The main goal of image segmentation is to partition an image into multiple regions or objects

Capítulo 2

Literature Review

Capítulo 3

Metodología

3.1. Data

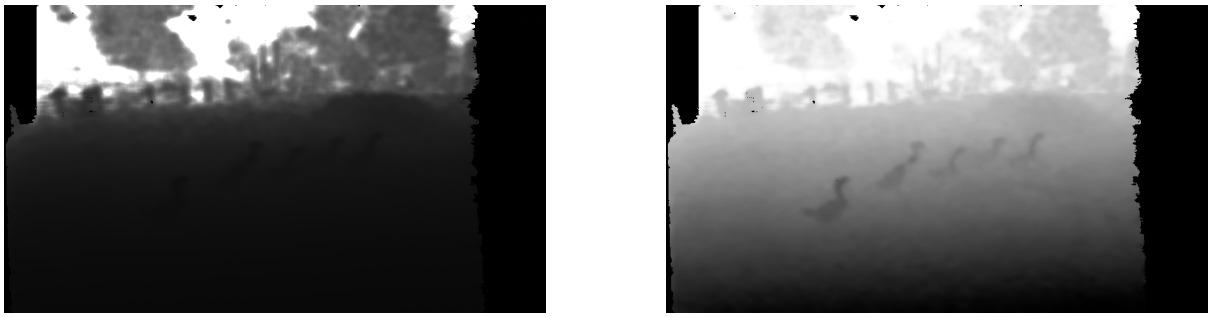
3.1.1. Lindenthal Dataset

3.1.2. Técnicas de Post Procesado

Existe una variedad de técnicas de post-procesado que aprovechan los tres canales de entrada del modelo para así lograr un *input* capaz de aportar mayor cantidad de información al modelo de segmentación. Además, al consistir el *dataset* en imágenes obtenidas con una cámara estéreo `Intel RealSense D435`, la cantidad de ruido en las imágenes es considerable, por lo que toda técnica que ayude a destacar los objetos de interés será importante para el correcto desempeño del modelo.

Equalización del Histograma

La primera técnica de postprocesado a implementar consiste en la equalización del histograma de valores en la imagen. Esto puede suponer una gran mejora en la forma en la que se representan las imágenes debido a que no todo el rango de valores de intensidad es aprovechado dado a que la mayoría de los objetos de interés se encuentran a una distancia relativamente baja de la cámara. Equalizar el histograma puede permitir destacar ciertos objetos que de otra forma no serían tan visibles. Como se puede observar en la figura

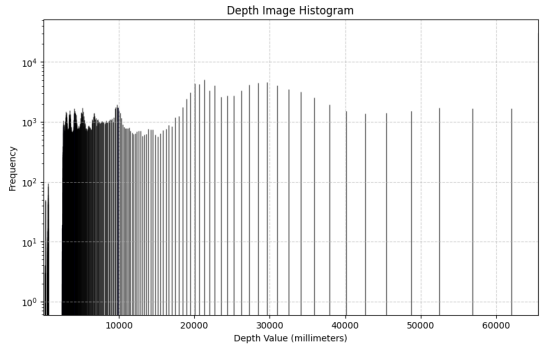


(a) Imagen de profundidad base (b) Histograma equalizado

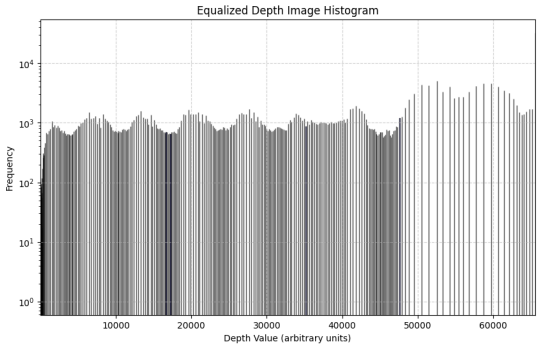
Figura 3.1: Comparación entre la imagen de profundidad base y la imagen con el histograma equalizado

3.2, la ecualización del histograma permite un mayor reparto de los valores de intensidad

en la imagen, lo que aprovecha aquellos rangos menos utilizados en la imagen original permitiendo así un mayor contraste entre los rangos más utilizados en la imagen original como se puede observar en la figura 3.1. Esta técnica, sin embargo, puede no ser adecuada para otras aplicaciones, ya que se pierde la información real sobre la distancia de los objetos a la cámara.



(a) Histograma de la imagen de profundidad base



(b) Histograma de la imagen con el histograma equalizado

Figura 3.2: Comparación entre los histogramas de la imagen de profundidad base y la imagen con el histograma equalizado

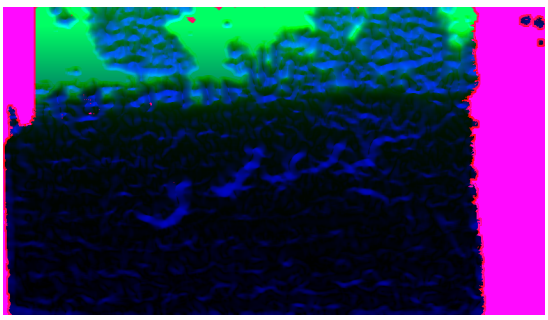
HHA Encoding

Esta técnica, propuesta por [1] usa los tres canales de la imagen de entrada para codificar las siguientes tres características:

- Altura sobre el suelo
- Disparidad horizontal
- Ángulo con respecto a la gravedad

Estas colorización, además de implementar características que difícilmente serían aprendidas por el modelo si no se codificaran en la imagen de entrada, aprovecha los tres canales de entrada al *encoder*, por lo que los pesos ya preentrenados en *ImageNet* pueden ser empleados para una mejor comprensión de la escena.

(a) Imagen de profundidad base



(b) Profundidad colorizada con HHA

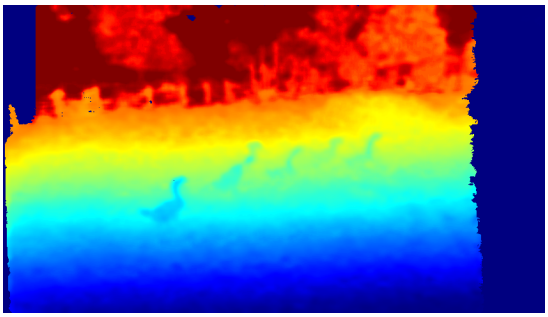
Figura 3.3: Comparison between the base depth image and the HHA encoded image

Colorizacón Jet

Otra posible técnica es la colorización de la imagen de entrada usando el esquema de color *Jet* mostrado en la figura 3.4, lo que implica la asignación de un valor *RGB* a cada píxel dependiendo de su valor de intensidad. Esto, de forma similar a 3.1.2, permite al modelo aprovechar sus pesos ya entrenados en *ImageNet* para identificar con mayor facilidad los objetos a segmentar.

Figura 3.4: *Jet* colormap

(a) Imagen de profundidad base

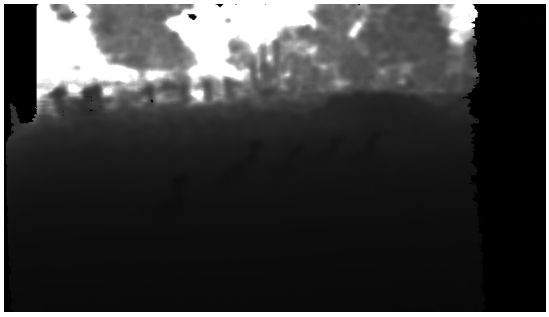


(b) Profundidad colorizada con *Jet*

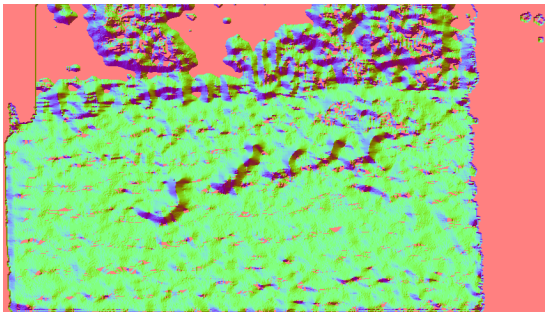
Figura 3.5: Comparación entre la imagen de profundidad y la colorización *Jet*

Normales

La última técnica de postprocesado a implementar consiste en la codificación de las normales en la imagen de entrada. Esta técnica consiste en calcular el vector normal de la superficie de cada pixel partiendo de la imagen de profundidad y codificarlo en los tres canales del *input* del modelo. Esto permite al modelo aprender características de la escena que de otra forma serían difíciles de aprender, como la orientación de los objetos en la escena.



(a) Imagen de profundidad base



(b) Normales codificadas en la imagen de profundidad

Figura 3.6: Comparación entre la imagen de profundidad y la codificación de las normales

Comparación de Técnicas

Las distintas técnicas anteriores presentan diferentes resultados a la hora de ser aplicadas al *input* del modelo. La tabla 3.1 muestra una comparación entre las distintas técnicas de postprocesado.

Técnica	Equalización	HHA	Jet	Normales
Deer IoU	0.000	0.000	0.000	0.000
Goat IoU	0.000	0.000	0.000	0.000
Donkey IoU	0.000	0.000	0.000	0.000
Goose IoU	0.000	0.000	0.000	0.000
Mean IoU	0.000	0.000	0.000	0.000

Cuadro 3.1: Comparación entre las distintas técnicas de postprocesado

Capítulo 4

Results

Capítulo 5

Discussion

Capítulo 6

Conclusion

Bibliografía

[1] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, "Learning rich features from rgb-d images for object detection and segmentation," 2014. [Online]. Available: https://arxiv.org/abs/1407.5736