Glint Denoising for Low Altitude Ocean Altimetry

Vlad Vandalovsky

Abstract—Ocean surface waves can introduce barriers to accurately estimate altitude for vehicles operating at low heights above the water. Some of these barriers include specular reflections, or glint, appearing on an non-uniform ocean wave state. In this paper, I propose a monocular vision-based method to denoise glints from single camera downward facing images data using the ResNet50 convolutional neural network, along with a temporal minimum filter deboising step to pre-process the image before feeding it into the network. The model is trained on a dataset of artificially generated images of height-labelled "clean" wave states at altitudes of 1 to 5 meters above the surface. Then, each image has artifical glint applied to it, to create the "noisy" dataset, and the algorithms performance is evaluated against both datasets. The results show that denoising the images to reduce glint, and then using a neural network to estimate height above sea level can reduce the impact of glints on altitude estimation and improve accuracy of near-surface ocean altimetry.

Index Terms—ResNet50, convolutional neural network, glint denoising, ocean altimetry

1 Introduction

Height estimatation is a critical measurement task for many vehicles, both piloted and autonomous. Estimation of height above sea level is particularly important for vehicles operating at low altitudes over the ocean, such as small boats, drones, and low-flying aircraft. The main inspiration for this project is the REGENT Viceroy, an electric seaglider designed to take off from the water's surface and fly within one wingspan (5-15 m) abover the ocean [1], where accuate height estimation is critical for human safety and comfort.

However, computing this height estimate can be difficult for such vehicles due to the presence of non-uniform ocean wave states. Solutions in industry and literature have included the use and fusion of multiple sensors, such as radar, ultrasonic, GPS, or lidar in addition to RGB images to estimate height above sea level [2], but these sensors can be expensive and may not be suitable for all applications.

In this paper, based on the learnings from EE367 Computational Imagery techniques, I propose a monocular vision-based method to achieve the height estimatate of a vehicle positioned above the ocean surface. There has been lots of work across autonomus cars, robotics, and related fields on using camera data to estimate depth and height [3], but the presence of sun glints on the ocean surface can cause difficulties for accurate height estimation.

To address this issue, I propose a method to denoise the images using a temporal minimum filter to denoise the glint artifacts from the source images. To map from pixels to real-world distances, I leverage a pinhole camera model, and use the known height of the camera above sea level to compute the focal length of the camera in pixels, which is used to convert from pixel measurements to real-world distances. Then, similar to Arik, et al. [4], train a ResNet50 convolutional neural network to estimate height from filtered images.

In evaluating the performance of this approach, an artifically generated dataset of height-labelled "clean" wave states at altitudes of 1 to 5 meters above the surface is created, and then artifical glint is applied to each image to create a "noisy" dataset. The filter is then tuned, and the model is trained on both datasets, and compared to understand the impact of glint on height estimation, and the effectiveness of the proposed denoising method.

2 Related Work

Oceanic Wave Modeling and Sun Glint Simulation Cox and Munk [5] performed seminal research in 1954 on characterising ocean wave state based on images captured of the ocean with sun glitter. They found a concrete relationship between the statistical distribution of wave state to the wind speed and direction, and sun azimuth and intensity. This sun glint model, combined with the Phillips wave spectrum [6] to help accurately describe wind-generated wave spectra, can be used to generate realistic downward-facing oceanic wave state images.

Sun-glint Removal From Images Similar to the many image denoising techniqes preented in EE367, there are many approaches for removing undesired artifacts from images. Specifically for sun-glint contained sequences of images, Lee et al. [7] propose Fast Adaptive Multivariate Empirical Mode Decomposition (FA-MEMD) to leverage the time relation between images to separate the temporal signal into uniform wave state and glint components. While this project does not use the FA-MEMD method directly, the time-dependent nature of glint is a key motivation for my denoising method.

Monocular Altitude Estimation Monocular vision based height estimation is a very vast area of research, with approachs rooted in structure from motion, defocus-based depth estimation, texture-gradient methods, deep learning based methods [3]. However, the most relevant approach to this project is the work by Arik, et al. [4], which uses a ResNet50 convolutional neural network to estimate height above urban areas from a flying drone using downward facing NADIR images of images 200m to 700m.

3 Approach

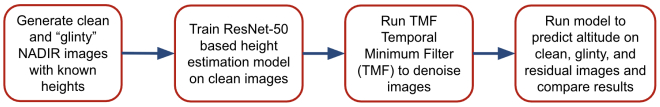

Fig. 1. Summary of Approach

3.1 Synthetic Dataset Generation

To generate 2D wave heightmaps, we use the inverse FFT of the Phillips directional spectrum [6]:

$$S(\mathbf{k}) = A\,\frac{\exp\!\left(-1/(kL)^2\right)}{k^4}\,\max(\hat{\mathbf{k}}\cdot\hat{\mathbf{w}},\,0)^2, \qquad (1)$$

where $\mathbf{k}$ is the wavenumber vector, $L = U_{10}^2/g$ is the dominant wavelength scale set by the 10 m wind speed U_{10}, $\hat{\mathbf{w}}$ is the unit wind direction, and A is an amplitude constant. Multiplying the spectrum by $(\hat{\mathbf{k}}\cdot\hat{\mathbf{w}})^2$ filters only waves aligned in the downwind direction, which is the only direction needed for generating glint patterns.

Surface normals are computed from the gradient of the heightmap above, and are rescaled to match the Cox–Munk empirical RMS using the following equation [5]:

$$\sigma^2_{\text{slope}} = 0.003 + 5.12\times10^{-3}\,U_{10}. \qquad (2)$$

The normalization avoids saturating the images with too many white or too many black pixels, making the images to appear more realistic and improving training accuracy.

Once the wave patterns and surface normals are generated, the RGB image with a base ocean color and sun glint can be generated by using the Cox–Munk model to compute the intensity of the glint at each pixel, which is a function of the surface normal, sun position, and camera position.

Each surface normal per pixel $\mathbf{N}$ is computed as:

$$I = I_{\text{water}} + F(\mathbf{N}\cdot\hat{\mathbf{c}})\,I_{\text{sky}} + G_s(\mathbf{N},\mathbf{H})\,g, \qquad (3)$$

where $F(...)$ is the unpolarised Fresnel reflectance for the air–water interface. The half-vector between sun $\hat{\mathbf{s}}$ and camera is $\hat{\mathbf{c}}$ $\mathbf{H} = (\hat{\mathbf{s}}+\hat{\mathbf{c}})/\|\hat{\mathbf{s}}+\hat{\mathbf{c}}\|$ is . Finally, G_s is a specular lobe with exponent 80, which controls glint severity. In each iteration, the clean and dirty images are generated with the same wave state, but different glint severity g to create a paired dataset of clean and glinty images. For each set of wind, sun, and altitude parameters, the wave state is propogated forward in time for 10 frames to capture a series of temporally correlated images per scenario. This is critical to enable the temporal miniminum filter to efficively find and remove the glint artifacts in each batch.

3.2 CNN-Based Altitude Estimation Model

To estimate height using the source images, a ReNet-50 model is used. However, instead of maintaining the original classifier in ResNet, a simple linear layer is used instead to output a single scalar for height estimation. Furthermore, instead of using mean-squared-error MSE as the reported metric, mean-absolute-error MAE is used instead to penalize large offenders for worst-case performance, and is more

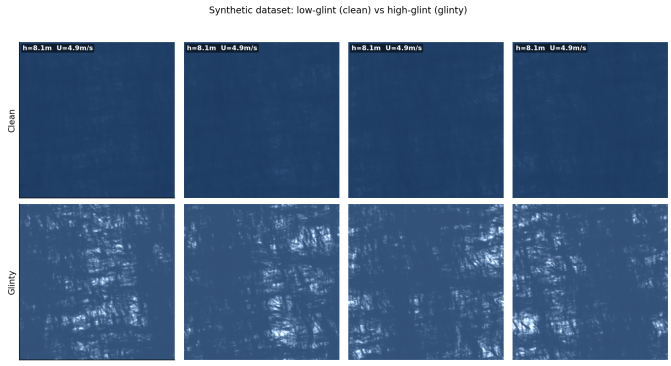

Fig. 2. Four Examples of Artificially Generated Clean vs. Glinty Images. Each image uses the same height above the water and wind speed, but is propogated forward in time to create a different wave state artifical glint.

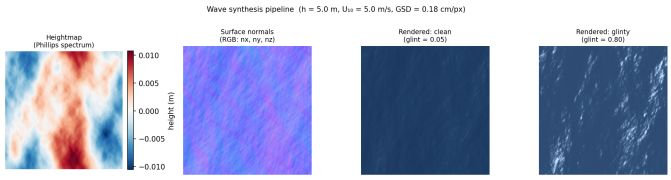

Fig. 3. Wave and Glint Image Generation Pipeline Steps

pertitnent for real world interpretability for human safety concerns.

The model parameters for ResNet are the seen in the table below, and were tuned experimentally and based on trial and error for best results. Given the small size of the dataset, a small learning rate was used, and the limited compute capability meant that I trained the model on my Macbook Air laptop.

TABLE 1
ResNet-50 Hyperparameters and Configuration

Setting	Value
Optimiser	Adam ($\alpha = 10^{-4}$)
Learning Rate schedule	Cosine annealing
Epochs	30
Batch size	32
Augmentation	Random crop, H/V flip, color jitter
Dataset split	70% train / 15% validation / 15% test (seed 42)
Hardware	Apple Macbook Air M3

3.3 Temporal Minimum Filter For Glint Removal

Let the observed image be modeled as a combination of the underlying ocean texture and the transient glint artifacts using the equation below:

$$I(x,y,t) = T(x,y,t) + G(x,y,t), \qquad (4)$$

where T is the steady state wave surface and G is the transient sun glint. Because glint is might only affect some frames in a sequence of 10 captured images in a batch, a given pixel struck by glint in one frame might not be struck in the subsequent ones. Therefore, we can find the minimum

value throughout the 10-frame window to estimate the glint-free baseline:

$$\hat{G}(x, y, t) = \min_{t' \in [t-w/2,\, t+w/2]} I(x, y, t'). \tag{5}$$

The recovered texture is the residual

$$\hat{T}(x, y, t) = \left[I(x, y, t) - \hat{G}(x, y, t) \right] \tag{6}$$

clipped to $[0, 1]$ due to the RGB nature of the image. The figure below shows an example of the TMF decomposition for a single image, where the left image shows the observed dirty image, followed by the estimated surface, then the glint itself, and the corresponding clean identical image from the original dataset. The TMF decomposition is not perfect, and some glint artifacts remain in the estimated surface, but the glint denoiser is better than nothing and improves the final height estimate provided by ResNet50, as will be shown in the results section.

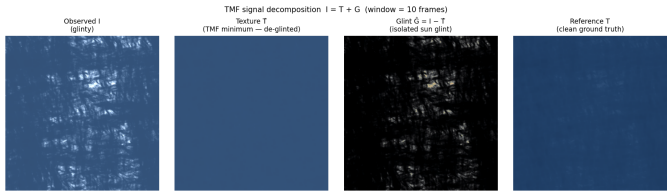

Fig. 4. TMF decomposition Process

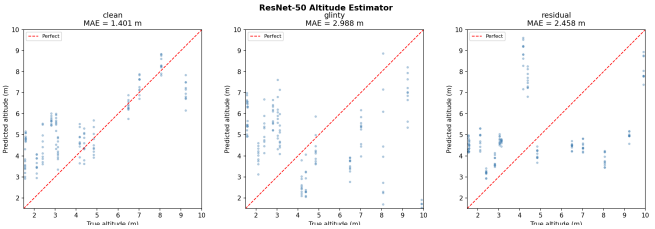

Fig. 5. Predicted vs. actual altitude for each test condition. The dashed line represents perfect prediction. The glinty and residual conditions show a consistent downward bias, while the clean condition clusters closely around the ideal line.

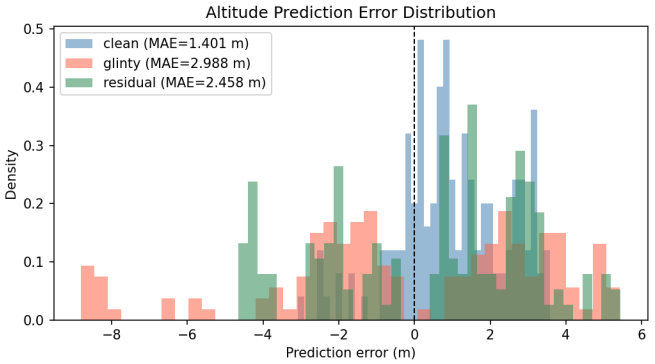

Fig. 6. Distribution of absolute prediction errors across the three test conditions. The clean condition is tightly concentrated at low error, while the residual condition has a long tail reaching nearly 10 m.

4 Experimental Results

To evaluate the performance of the proposed approach, the ResNet50 model trained on clean images is evaluated across three test conditions: clean images (no glint), glinty images (raw glint applied), and residual images (glint removed via TMF). For each condition, the mean absolute error (MAE), root mean squared error (RMSE), prediction bias, and worst-case error are reported in Table 2.

TABLE 2
Altitude Estimation Error by Image Condition ($n = 30$ test samples each)

Condition	MAE (m)	RMSE (m)	Bias (m)	Max Err (m)
Clean	1.82	2.06	−0.06	3.47
Glinty	2.49	2.80	−1.94	3.88
Residual (TMF)	3.12	4.36	−2.58	9.92
Glint penalty	+0.67 m MAE from clean → glinty			

Effect of Glint on Altitude Estimation The clean condition achieves the best performance with a MAE of 1.82 m and RMSE of 2.06 m, with near-zero bias (−0.06 m), confirming that the ResNet50 model is capable of learning the GSD-based altitude signal from clean wave textures across the 1–5 m altitude range. Adding artificial glint to the images degrades performance by 0.67 m MAE and increases the negative bias to −1.94 m, meaning the model tends to underestimate altitude when glint is present. This highlights the fact that glint reflections corrupt the local texture features that the model relies on, making the wave pattern appear coarser than it actually is.

TMF Deglinting Results Based on the results of these simulations, the TMF residual images performs the worst of the three conditions by a small margin, with a MAE of 3.12 m and a worst-case error of nearly 10 m. This is likely because the TMF residual images look different from the clean images the model was trained on. The temporal minimum filter estimates the glint baseline by taking the per-pixel minimum over the 10-frame window, and the residual is the difference between the observed image and this baseline. While this does successfully isolate the glint artifacts as shown in Figure 4, the resulting residual images have most of the interesting texture subtracted away, leaving mostly monotonous features, which is tougher for the model to decipher into a meaningful measurement.

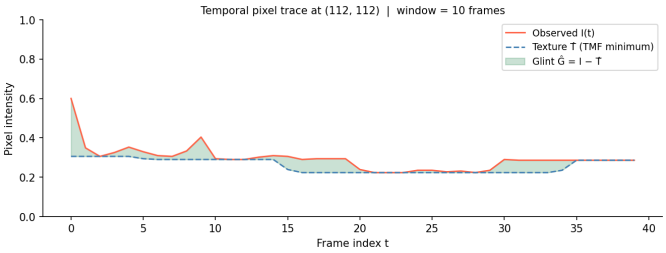

Fig. 7. Per-pixel intensity trace over 10 frames for a sample sequence. The rolling minimum (dashed) tracks the glint baseline, and the residual (shaded) captures the transient glint rather than the underlying wave texture.

These results suggest that, while the TMF is a reasonable method for visualising and isolating glint, this simulation is not an ideal test of its effectiveness for improving altitude estimation, since the residual images do not have enough non-glint features for the model to leverage.

As a future improvement, if given more time and resources, I would try this approach on real-world imagery data. The Phillips spectrum produces smooth, periodic wave patterns, and the glint is rendered with a fixed specular exponent and sun angle, meaning the model only needs to learn a limited set of texture variations to estimate altitude. In real-world conditions, the wave surface would contain richer multi-scale structure, foam, breaking waves, varying sea states, and more complex glint distributions driven by changing sun position and wind direction throughout a flight. The poor residual performance seen here may be partly an artifact of the simulation, and real-world data could tell a different story.

5 Conclusion

In this paper, I proposed a monocular vision-based approach to altitude estimation above the ocean surface, motivated by the needs of ground-effect vehicles like the REGENT Viceroy operating at 1 to 15 m above the water. The approach leverages the fact that the scale of wave texture in a nadir image encodes altitude through the ground sampling distance (GSD), and trains a ResNet50 regression model to recover this altitude estimate from a single image.

The experiments confirm that ResNet50 can learn the GSD-altitude relationship from clean synthetic wave images, achieving a MAE of 1.82 m across a 1–5 m altitude range. Adding glint degrades this by 0.67 m MAE and introduces a systematic downward bias, confirming that sun glint is a real obstacle for this approach.

However, the TMF deglinting step did not improve on the glinty baseline within this simulation, and in fact made accuracy worse. The core issue is that the synthetic dataset is intentionally simple, ie. smooth Phillips-spectrum waves with a fixed optical reflection model. This implies that the TMF residual strips away most of the limited texture the model relies on. In real-world conditions, the ocean surface contains richer multi-scale structure, foam, breaking waves, and more complex glint distributions, which could leave substantially more useful texture in the residual even after deglinting.

Going forward, the most promising next steps would be to evaluate the approach on real nadir ocean imagery, retrain on TMF minimum images to better align training and inference distributions, or explore a learnable deglinting front-end co-trained with the altitude estimator end-to-end. Despite the mixed simulation results, this work demonstrates that monocular altitude estimation from wave texture is a tractable problem, and that glint handling deserves careful attention as a first-class concern in any future deployable system.

Acknowledgments

The authors would like to thank Professor Gordon Wetzstein and TA Minseo (Sonia) Kim for their continued support and guidance not just on this project, but throughout the entire class.

References

[1] REGENT Craft, "Viceroy seaglider," https://www.regentcraft.com/seagliders/viceroy, 2025, accessed: 2026-03-13.

[2] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Addison-Wesley, 1999.

[3] R. Mao and Y. Zhang, "Basic theories and methods of target's height and distance measurement based on monocular vision," *International Journal of Network Dynamics and Intelligence*, 2025, monocular vision based estimation.

[4] A. E. Arık, "Vision-based UAV altitude estimation using deep learning: A ResNet50 approach on nadir images," *Academic Platform Journal of Engineering and Smart Systems*, vol. 14, no. 1, pp. 46–54, 2026.

[5] C. Cox and W. Munk, "Measurement of the roughness of the sea surface from photographs of the sun's glitter," *Journal of the Optical Society of America*, vol. 44, no. 11, pp. 838–850, 1954.

[6] O. M. Phillips, "The equilibrium range in the spectrum of wind-generated waves," *Journal of Fluid Mechanics*, vol. 4, no. 4, pp. 426–434, 1958.

[7] J. Lee *et al.*, "FA-MEMD: Fast adaptive multivariate empirical mode decomposition for sun-glint removal in ocean remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, 2025, motivation for temporal glint-removal approach.