Sargassum detection and path estimation using neural networks

José A. López-Portillo[a], Iván G. Casasola-Rodríguez[b], Boris Escalante-Ramírez[c,f], Jimena Olveres[c,f], Jaime Arriaga[d], and Christian Appendini[e]

[a]Posgrado en Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico
[b]Posgrado en Ciencia e Ingeniería de la Computación, Universidad Nacional Autónoma de México, Mexico City, Mexico
[c]Facultad de Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico
[d]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[e]Laboratorio de Ingeniería y Procesos Costeros, Instituto de Ingeniería, Universidad Nacional Autónoma de México, Sisal, México
[f]Centro de Estudios en Computación Avanzada, Universidad Nacional Autónoma de México, Mexico City, Mexico

ABSTRACT

Sargassum has affected the Mexican Caribbean coasts since 2015 in atypical amounts, causing economic and ecological problems. Removal once it reaches the coast is complex since it is not easily separated from the sand, damaging dune vegetation, heavy transport compacts the sand and further deteriorates the coastline. Therefore, it is important to detect and estimate the sargassum mats path to optimize the collection efforts in the water. There have been some improvements in systems that rely on satellite images to determine areas and possible paths of sargassum, but these methods do not solve the problems near the coastline where the big mats observed in deep sea end up segregating in little mats which often do not show up in the satellite images. Besides, the temporal scales of nearshore sargassum dynamics are characterized by finer temporal resolution. This paper focuses on cameras located near the coast of Puerto Morelos reef lagoon where images are recorded of both beach and near-coastal sea. First, we apply preprocessing techniques based on time that allows us to discriminate the moving sargassum mats from the static sea bottom, then, using classic image processing techniques and neural networks we detect, trace, and estimate the path of the mat towards the place of arrival on the beach. We compared classic algorithms with neural networks. Some of the algorithms we tested are k-means and random forest for segmentation and dense optical flow to follow and estimate the path. This new methodology allows to supervise in real time the demeanor of sargassum close to shore without complex technical support.

Keywords: Sargassum, detect, estimate, trace, coast, cameras

1. INTRODUCTION

Since 2015 the Mexican Caribbean Shores have been affected by abnormal amounts of sargassum mats. These algae pose economic and ecological repercussion in the region.[1] When sargassum decomposes it releases leachate and organic particles causing severe impacts in the nearby ecosystems.[2] The decomposition also liberates noxious gases and unpleasant smell that discourage tourism which is the principal income in the region.[1] There have been studies for the exploitation of this resource[3] using sargasso as raw material for different products. Although for a better performance it is necessary to get rid of sargassum before it gets to the shore, if it reaches the beach, it becomes more difficult to separate it from the sand, which causes deterioration in the coastline and hinders its processing and transportation.

Quantity and arrival events have proven to be unpredictable. There have been some cases in which it greatly varies its amount by few hundred meters. Some companies like Dakatso offer their services including satellite

and drone monitoring for collection, as well as barriers that contain and redirect the algae.[3] Thus, it is necessary to effectively detect sargassum before it gets to the shore or when the barriers get full.

There are different ways to analyze satellite images to detect and study the sargassum arrivals. In[4] satellite images are processed using 2 and 5 spectral bands to obtain 5 vegetal indices to train a Random Forest algorithm. These methods show their strength when detecting sargassum mats in deep sea, where the mats are bigger.[5,6] These models are complemented with simulated ocean current algorithms,[7] helping to improve their temporal capacity. In[8] it is mentioned that spatial and temporal resolutions in satellite images are inversely proportional. Near to the coast the waves are less affected by the currents and gain influence by the wind, which causes the sargassum mats to segregate. This condition makes it vital to increase the spatial resolution of the image, which in turn derives in loss of temporal definition and the decreases the track capability of the mats. Additionally, the size and layers of the satellite images make it more difficult to store and execute the algorithm in modest systems. At least in cloudy days some images are useless since the cloud blocks away the sargassum mats.

For the above reasons, there are systems that record images from the nearby coast. In the work of Valentini[9] smartphone-based cameras were used to train a pixel-wise convolutional neural network, and were capable to classify coast, sea and sargassum. Rutten et al[10] used 5 years of hourly images captured from a fixed camera system to analyze stranded sargassum. As a result, a cyclic behavior of sargassum was described and the intriguing capability of self cleaning under rough sea conditions, which might be characteristic of reef lined beaches. However, the temporal resolution was not sufficient to infer mat sizes, direction and speed. We propose a passive system capable of analyzing images captured with the same camera to locate, trace and get the size of the sargassum mats near the shore. Making the recollection efforts more efficient and providing a unique set of data required to implement numerical models that can represent the mat dynamics in the nearshore, crucial to test barrier designs.

1.1 Optical flow

Estimating motion vectors field in computer vision is known as optical flow. This task consists of estimating vectors of the apparent motion of pixels in an image along of a sequence. To solve this problem many both traditional and neural networks-based methods have been designed along the years.

Traditionally optical flow estimation has been approached by differential approximation methods and intensity invariance assumption such as Lucas-Kanade algorithm,[11] a more efficient method was proposed by G. Farneback in[12] which is based on polynomial expansion instead of a linear algorithm, such as Lucas-Kanade. Some others used a variational approach derived from the work of Horn and Schunck[13] and turned the problem into an energy minimization one. All of these methods are accurate, however they can be sometimes numerically heavy since their approach is a hand-crafted optimization problem. Some others, are quite slow due to the many iterations they have to perform in order to converge to a minimum.

In the deep learning era, many architectures have emerged during the last few years to tackle the optical flow estimation problem. Early architectures used convolutional neural networks (CNNs) and a correlation layer such as FlowNet,[14] an improved version of this work was proposed by E. Ilg et al.[15] using a stacked architecture achieving more accurate results. In spite of its good results this kind of architectures still had a tremendous amount of parameters which made them quite robust and hard to train. A less robust model was proposed on[16] (PWC-Net) which was smaller in size and easier to train than previous models, this new proposal used a pyramid and warping scheme along with a cost volume, achieving better comparable results to the state-of-the-art. Recent models have far outperformed the PWC-Net and previous models results, such as RAFT,[17] Z. Teed and J. Deng used a correlation cost volume and a gated recurrent unit (GRU) as an iterative mechanism to achieve a better optical flow estimation. Some new models emerged based on RAFT architecture, that is the case of GMA[18] which added a new module to solve the occlusion problem presented on many cases of a sequence. S. Jiang et al. were the first to use transformers in the optical flow task due to its recent success in natural language processing tasks to model long-range dependencies, that is why we are using this architecture to estimate optical flow along the sargassum data and make a comparison with results of a traditional method such as Farneback´s.

2. DATA PROCESSING

Images are collected from a building near the shore of Puerto Morelos, Mexico, reef lagoon. The system was first implemented by the Engineering Institute at UNAM to study deterioration in the coastline. The images are a result of calculating the average of 30 images taken every 2 seconds, the result is a smoother image with most artifacts produced by waves are removed. These parameters were selected to avoid affecting the size and position of the sargassum mats, if images are averaged longer than 2 seconds, sargassum mats may appear larger than they really are.

The first action is to crop useless areas from the images like the beach and the sky. Furthermore, only the second layer of the image in the HSV color space is used. Another average operation takes place to discriminate information from the bottom sea and any other motionless objects in the images. Several minutes of the event are analyzed to ensure the sargassum mats move enough without affecting the result. Then we generate a mask image from the subtraction of the bottom image and the original average image, this mask indicates all the coordinates that move in our interval of time allowing us to focus on these areas. The final results can be seen in Fig. 1.

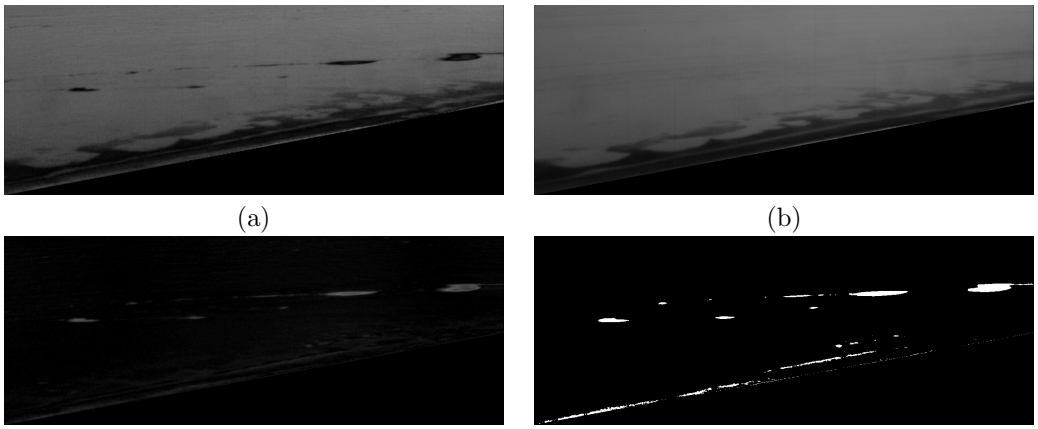


Figure 1. (a) Original image, (b) Average image,(c) Substraction, (d) Final mask

This preprocessing is needed to discriminate large brown stains produced by the so called brown tide from sargassum mats. Brown tide is a phenomenon present in the coast when sargassum decays, the color of bride tide is similar to that of sargassum´s. Without this preprocessing step, the brown tide could be mistaken for moving sargassum mats making subsequent steps more difficult.

3. DETECTING AND CLASSIFYING SARGASSUM

Even though a lot irrelevant data is discarded, certain spontaneous events, like moving boats and clouds changing the illumination, can affect the process. Two different that ways to eliminate these variables are proposed, the first one uses classic algorithms of classification such as Random Forest and K-means; the second one is a U-Net. To train and evaluate our algorithms we manually labeled a group of images in which these problems are present.

3.1 Classic algorithms

Random forest and a K-means classifiers were used to distinguish between sargassum mats and other invasive elements. The classification results profits from the preprocessing step by executing a pixel-by-pixel classification only in the areas in which movement was detected, this makes a faster and cleaner process. 300 labeled images were used for training with critical events. Different numbers of clusters for the K-means and different maximum depths for the Random Forest were tested.

We are only interested in sargassum mats so the algorithm has some internal freedom to get a better separation for all the different shades and objects that may exist in the image. Random Forest algorithm was set to a

maximum depth of 15 and the K-means was set to 15 clusters. The K-means algorithm presented an accuracy of 83.3939% while the Random Forest presented accuracy of 94.2031%. Results can be seen in Fig. 2 where the best results belong to the Random Forest algorithm that removes most of the brown tide.

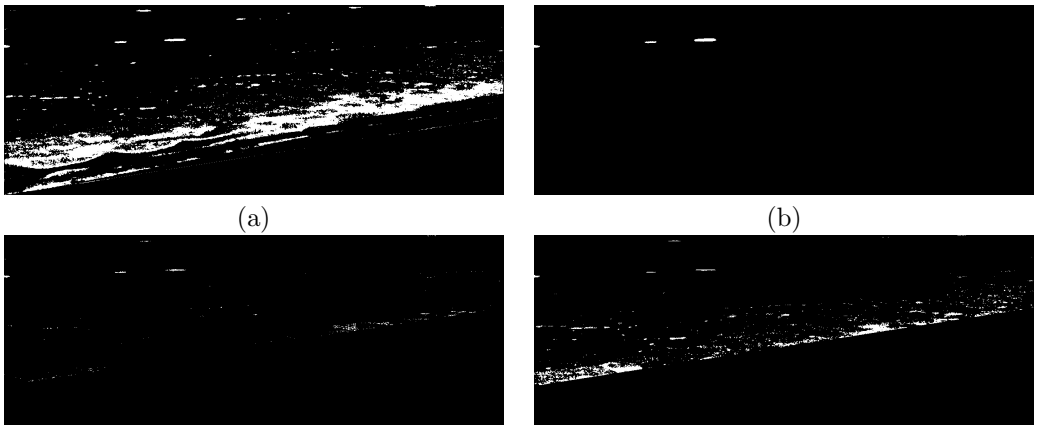


Figure 2. (a) Input image, (b) Hand selected mask, (c) Output from Randon forest , (d) Output from K-means

3.2 UNET

A U-Net convolutional Network, based in Ronneberger et al algorithm,[19] was applied adding small changes like using as input a three layer RGB image. A lot of information was lost in the case of gray-scale images, because the sargassum patches can be easily confused with the brown tide and seagrass. In the training and subsequent prediction of the U-Net, masked images obtained in the preprocessing phase are used to focus the net's attention in the areas of interest. The images must be re-sized to fit the system requirements; in our case an input layer of 256 by 256 was established. The final learning curve can be seen in Fig. 4

At the end, a Gaussian filter is applied to smooth the output image, so the final binary mask looks cleaner. The final result can be seen in Fig. 3 where the input image (a) is shown with a lot of cloud reflection, the desired output is in (b), and finally the raw output (c) and his binarized output.

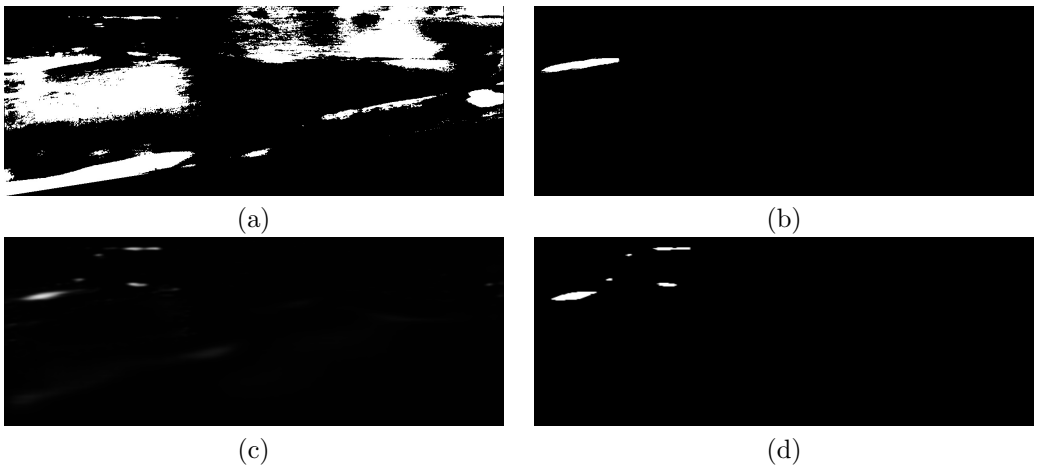
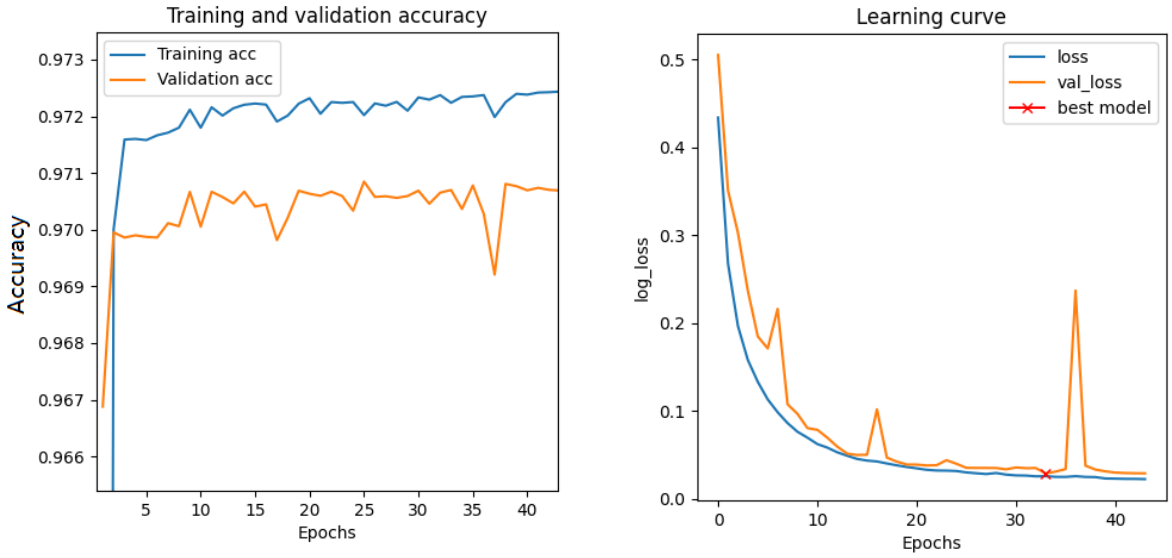


Figure 3. (a) Input image, (b) Hand selected mask, (c) Output from the unet, (d) Mask generated



(a) (b)

Figure 4. (a) Accuracy curve for the U-NET, (b) Loss curve for the U-net .

4. SARGASSUM MOTION ESTIMATION

We used the generated masks as data input to estimate optical flow, both implementations of Farneback method and GMA architecture were implemented in order to compare results of a traditional method and a neural networks based one. We also evaluated both methods with a dataset taken from[20] since the ground truth is provided within this dataset, so that way we can determine how accurate a method is.

4.1 Farneback method

One requirement for this algorithm is that input images must be in gray scale. Default parameters for Farneback method were used.

4.2 GMA architecture

This implementation was based on source code from the original paper of GMA. However, ours was trained with MPI Sintel dataset along with synthetic echocardiograms processed dataset where we have completed the ground truth with interpolation methods. Fig. 5 shows the training process and the end-to-end point error (EPE) which is a common evaluation metric for optical flow. It can be noticed that the validation datasets behaved similarly and the EPE slightly fluctuated but decreased. The same behavior is shown in the entire training EPE. Tab. 1 shows the achieved EPE values in validation datasets at the end of the training process.

Table 1. End-to-end point error (EPE) in validation datasets

Dataset	EPE
Clean (Sintel)	0.84
Final (Sintel)	1.33
Echocardiograms	0.23

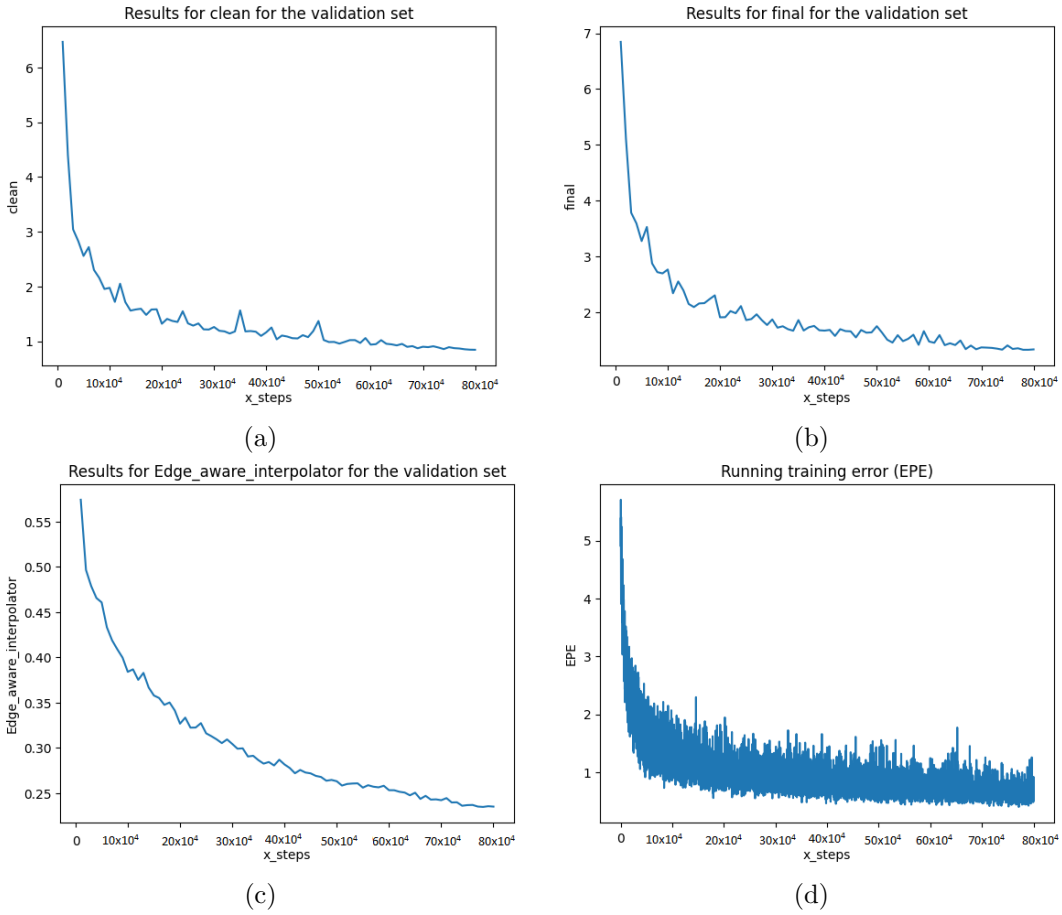
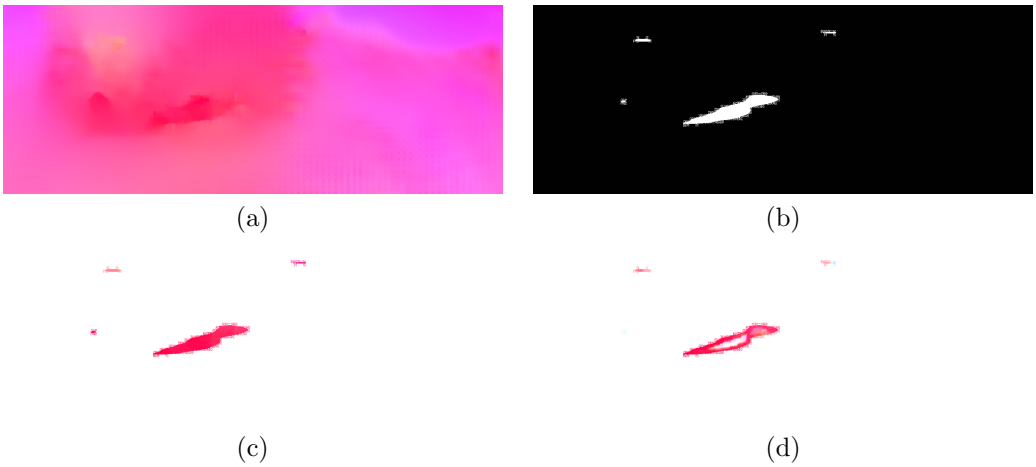


Figure 5. (a) Clean MPI Sintel validation dataset, (b) Final MPI Sintel validation dataset, (c) Echocardiograms dataset with interpolated ground truth, (d) GMA training error.

4.3 Post processing

Optical flow estimation of the entire scene is multiplied by the corresponding mask (binarized), in order to keep motion vectors only on regions of interest (sargassum regions). This post processing is also applied to Farneback estimation. Fig 6 shows the final sargassum motion estimation.



(a) (b)

(c) (d)

Figure 6. (a) Complete estimation of GMA, (b) Corresponding image mask, (c) Post processed optical flow of GMA, (d) Post processed optical flow of Farneback.

5. METHODS EVALUATION

5.1 Classic classifiers and U-NET

As observed, classic and net classifiers perform differently, the best classifier being the U-net followed by Random Forest as shown in Tab. 2.

Table 2. Classifiers performance

Classifier	Accuracy
K-means	83.3939%
Random Forest	94.2031%
U-Net	97.0702%

Performance can also be assessed in two specific cases. Fig. 7 shows how these classifiers deal with moving boats, random forest and K-means struggle to eliminate the ship while the U-net does not.

In Fig. 8 a raindrop in the camera lens is evident. Its effect in the image is a heavy blur in a specific area, so some of the pixels of a real mat can be lost in this region. Again, Random Forest and K-means struggle to discern the difference between the real mat and the raindrop while the U-net eliminates the latter.

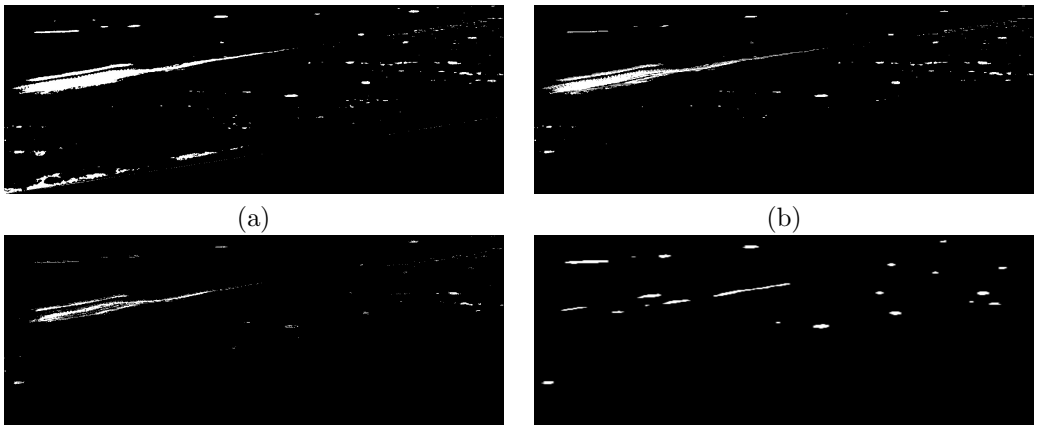


Figure 7. (a) Input image of a boat in image, (b) K-means classifier, Random forest (c), Mask generated (d)

Figure 8. Input image of raindrop in camera lens (a), K-means classifier (b), Random Forest (c), U-net (d)

5.2 Farneback and GMA architecture

5.2.1 Evaluation dataset

As mentioned before, a different dataset was used to evaluate each method and verify accuracy of the estimate optical flow in the sargassum dataset. The sequences contained in the evaluation dataset consist of Grove2, Grove3, Urban2, Urban3 and Venus. Fig. 10 shows the estimated optical flow of each method with ground truth encoded with the color wheel shown in Fig. 9. Tab. 3 confirms that GMA implementation presents a lower EPE.

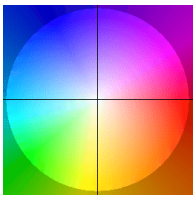


Figure 9. Color wheel to encode optical flow.

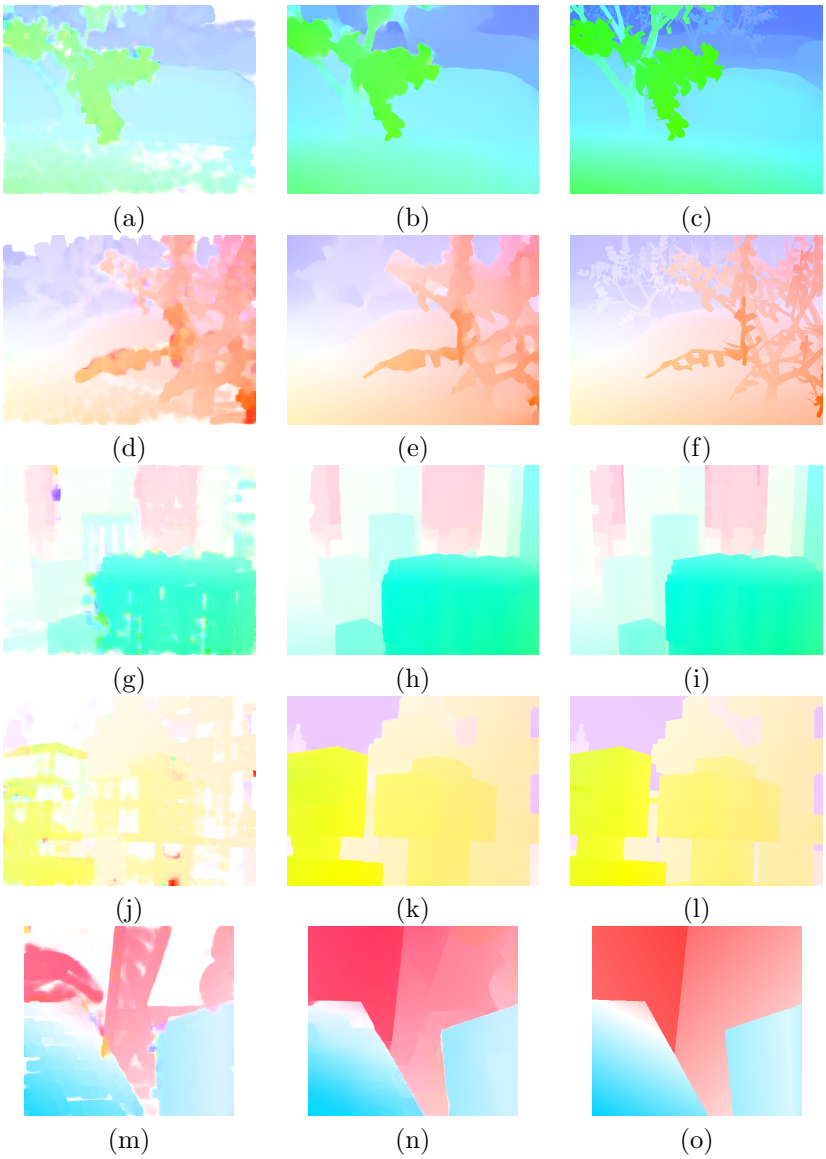


Figure 10. (a) Grove2 Farneback estimation , (b) Grove2 GMA estimation (ours), (c) Grove2 ground truth, (d) Grove3 Farneback estimation , (e) Grove3 GMA estimation (ours), (f) Grove3 ground truth, (g) Urban2 Farneback estimation , (h) Urban2 GMA estimation (ours), (i) Urban2 ground truth, (j) Urban3 Farneback estimation , (k) Urban3 GMA estimation (ours), (l) Urban3 ground truth, (m) Venus Farneback estimation , (n) Venus GMA estimation (ours), (o) Venus ground truth.

5.2.2 Sargassum dataset

Finally, both methods have been validated and their accuracy calculated. Based on the above GMA estimation is preferred over Farneback´s. Fig. 11 shows the optical flow estimation in four consecutive frames of sargassum data, showing red colors, meaning that sargassum is shifting to the right. Fig. 12 displays the actual frames in which optical flow is estimated.

Table 3. EPE on evaluation dataset

Sequence	Farneback	GMA
Grove2	0.58	0.26
Grove3	1.33	0.71
Urban2	1.41	0.28
Urban3	2.97	0.35
Venus	1.44	0.26

(a) (b)

(c) (d)

Figure 11. Optical flow estimation: (a) Frame 1, (b) Frame 2, (c) Frame 3, (d) Frame 4.

(a) (b)

(c) (d)

Figure 12. (a) Frame 1, (b) Frame 2, (c) Frame 3, (d) Frame 4.

6. DISCUSSION

This work focuses on near to real-time detection of sargassum mats near the coast with static camera images. It is a complement of approaches based on satellite image analysis, as mentioned before. The proposed preprocessing stage proves to be useful to reduce the data quantity to analyze, eliminating most of the undesired objects, but it is also needed to ease the classification task in charge of discriminating moving objects like ships and sudden

changes like raindrops and cloud illumination. The best performing classifier was the U-net which better removed these elements very quickly, but it requires more computational resources and some refinement to correctly detect small patches that get lost when the image is resized. Classic classifiers do a better job with smaller patches, but they were not able to totally eliminate the major objects.

Regarding the optical flow estimation, we have shown that neural networks and traditional methods produce reliable results for this particular case of sargassum data. GMA was the best model to estimate optical flow. It allowed to trace sargassum up to the shore once it was detected by any of the classification methods used in this work.

ACKNOWLEDGMENTS

This work was sponsored by UNAM PAPIIT grants TA101121, TA100420 and IV100420 and CONACYT. The authors acknowledge Servicio Académico de Monitoreo Meteorológico y Oceanográfico (SAMMO) for their support with the cameras used for this work.

REFERENCES

[1] Chávez, V., Uribe-Martínez, A., Cuevas, E., Rodríguez-Martínez, R. E., van Tussenbroek, B. I., Francisco, V., Estévez, M., Celis, L. B., Monroy-Velázquez, L. V., Leal-Bautista, R., Álvarez-Filip, L., García-Sánchez, M., Masia, L., and Silva, R., "Massive influx of pelagic sargassum spp. On the coasts of the mexican caribbean 2014–2020: Challenges and opportunities," *Water (Switzerland)* **12**(10), 1–24 (2020).

[2] van Tussenbroek, B. I., Hernández Arana, H. A., Rodríguez-Martínez, R. E., Espinoza-Avalos, J., Canizales-Flores, H. M., González-Godoy, C. E., Barba-Santos, M. G., Vega-Zepeda, A., and Collado-Vides, L., "Severe impacts of brown tides caused by Sargassum spp. on near-shore Caribbean seagrass communities," *Marine Pollution Bulletin* **122**(1-2), 272–281 (2017).

[3] López Miranda, J. L., Celis, L. B., Estévez, M., Chávez, V., van Tussenbroek, B. I., Uribe-Martínez, A., Cuevas, E., Rosillo Pantoja, I., Masia, L., Cauich-Kantun, C., and Silva, R., "Commercial Potential of Pelagic Sargassum spp. in Mexico," *Frontiers in Marine Science* **8**(November) (2021).

[4] Cuevas, E., Uribe-Martínez, A., and Liceaga-Correa, M. d. l. Á., "A satellite remote-sensing multi-index approach to discriminate pelagic Sargassum in the waters of the Yucatan Peninsula, Mexico," *International Journal of Remote Sensing* **39**(11), 3608–3627 (2018).

[5] Wang, M. and Hu, C., "Automatic Extraction of Sargassum Features From Sentinel-2 MSI Images," *IEEE Transactions on Geoscience and Remote Sensing* **59**(3), 1–19 (2020).

[6] Nguyen Thi Thu Hang, Nguyen Thai Hoa, Tong Phuoc Hoang Son, and Lam Nguyen-Ngoc, "Vegetation Biomass of Sargassum Meadows in An Chan Coastal Waters, Phu Yen Province, Vietnam Derived from PlanetScope Image," *Journal of Environmental Science and Engineering B* **8**(3), 81–92 (2019).

[7] Maréchal, J. P., Hellio, C., and Hu, C., "A simple, fast, and reliable method to predict Sargassum washing ashore in the Lesser Antilles," *Remote Sensing Applications: Society and Environment* **5**(May 2016), 54–63 (2017).

[8] Ody, A., Thibaut, T., Berline, L., Changeux, T., André, J. M., Chevalier, C., Blanfuné, A., Blanchot, J., Ruitton, S., StigerPouvreau, V., Connan, S., Grelet, J., Aurelle, D., Guéné, M., Bataille, H., Bachelier, C., Guillemain, D., Schmidt, N., Fauvelle, V., Guasco, S., and Ménard, F., "From in Situ to satellite observations of pelagic Sargassum distribution and aggregation in the Tropical North Atlantic Ocean," *PLoS ONE* **14**(9), 1–30 (2019).

[9] Valentini, N. and Balouin, Y., "Assessment of a smartphone-based camera system for coastal image segmentation and Sargassum monitoring," *Journal of Marine Science and Engineering* **8**(1), 1–21 (2020).

[10] Rutten, J., Arriaga, J., Montoya, L. D., Mariño-Tapia, I. J., Escalante-Mancera, E., Mendoza, E. T., van Tussenbroek, B. I., and Appendini, C. M., "Beaching and Natural Removal Dynamics of Pelagic Sargassum in a Fringing-Reef Lagoon," *Journal of Geophysical Research: Oceans* **126**(11), 1–16 (2021).

[11] Lucas, B. D. and Kanade, T., "An iterative image registration technique with an application to stereo vision," (1981).

[12] Farnebäck, G., "Two-frame motion estimation based on polynomial expansion," (2003).

[13] Horn, B. and Schunck, B., "Determining optical flow," *ARTIFICIAL INTELLIGENCE* , 1,2,3,4 (1981).
[14] Alexey Dosovitskiy, Philipp Fischer, E. I. P. H. C. H. and Golkov, V., "Flownet: Learning optical flow with convolutional networks," *IEEE International Conference on Computer Vision* (2015).
[15] Eddy Ilg, Nikolaus Mayer, T. S. M. K. A. D. and Brox, T., "Flownet 2.0: Evolution of optical flow estimation with deep networks," *Conference on Computer Vision and Pattern Recognition* (2018).
[16] Deqing Sun, Xiaodong Yang, M.-Y. L. and Kautz, J., "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," (2018).
[17] Teed, Z. and Deng, J., "Raft: Recurrent all-pairs field transforms for optical flow," (2020).
[18] Shihao Jiang, Dylan Campbell, Y. L. H. L. and Hartley, R., "Learning to estimate hidden motions with global motion aggregation," (2021).
[19] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation.," (2015).
[20] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael Black, Richard Szeliski, "Datasets for optical flow." https://vision.middlebury.edu/flow/data/ (2022).