Clown Fish and Blue Tang Fish Species Classification using Color Moment Feature

Muhammad Fahmi Bin Rosnan*,
Konya Teknik University, Electrical and Electronics Engineering Department, Konya, Turkey
e198221001040@ktun.edu.tr

ABSTRACT

Image vision can be used in the sky, on the ground, and underwater. This research proposed a fish species classification technique using the colour moment feature and a K-NN classifier. The colour moment is based on the probability distribution categorized as the mean, standard deviation, and lack of symmetry in probabilistic distribution (skewness) from the RGB colour model of the images. The dataset collected from Google image search contains 15 clown fish and 15 blue tang fish images. Then, 30 input images are cropped and resized before extracting each colour moment feature, based on the RGB model. The extracted features were stored, and the classification process was performed using a K-NN classifier. The accuracy achieved on the training data was 90.48% and 89% was achieved on the testing data.

Keywords—Image Processing Toolbox, MATLAB, color moment, K-NN, marine habitat

1 INTRODUCTION

Many new publications have recently discussed capturing the marine existence using artificial underwater robot vision [1]. However, due to the unstable environment and various obstacles such as blurry conditions underwater, the images are poorly captured. For this reason, many researchers, especially from the marine sector, take this matter seriously to improve the detection and classification of image processing technology in underwater environments with advanced techniques such as deep learning and machine learning [2]-[4].

2 RELATED WORKS

Image identification and classification has become a hot issue in many sectors, and the marine sector is no exception. The unpredictable conditions when staying underwater cause particular difficulties for the researcher, like capturing the image of marine habitats in cloudy and blurry conditions, and trying to recognize small creatures on the seabed such as scallops.

Fengqiang et al. (2018) attempted to detect and count small marine creatures through faster regions with a convolutional neural network (R-CNN) and kernelized correlation filter (KCF) on the seabed based on the self-collected images captured by an underwater robot in real-time [2]. However, the research does not describe the kind of features extracted from the captured images.

Another publication by Suxia et al. (2020) can detect a variety of fish from the natural underwater environment collected by a built-in camera in an underwater vehicle remotely operated, instead of autonomous underwater robot (AUR), using a convolutional neural network (CNN) [3]. Unfortunately, the image dataset only extracted the features based on various fish species as input for detection with a CNN.

Fenglei et al. (2020) proposed a method to detect and classify tiny marine organisms through videos taken from remote vehicles (ROVs) using deep-CNN (DCNN) and optimized the feature extraction of objects using a Region Proposal Network (RPN) [4]. However, the extracted features only focus on the visibility of the images filmed and during close distance of detection.

Most of the mentioned works succeed in detecting, classifying, and extracting the features of the images captured or filmed. However, they do not discuss how to classify the images captured using the colour moment feature. Thus, this research conducted an experiment using the Image Processing Toolbox in MATLAB [5], proposed a supervised method as described in [6], and classified the images captured using the colour moments as mentioned in [7], without consider kurtosis or data flat at a normal distribution on two well-known fish species: clown fish and blue tang fish, which are popular with other names such as Nemo and Dory.

3 METHODOLOGY

3.1 The Overview of Experiment

The dataset was collected from the Google image search and then cropped and resized using the Image Processing Toolbox software in the MATLAB computing environment [5].

The process starts from the data acquisition and pre-processing until classifying the data using the K-nearest neighbor (K-NN) classifier method [6]. This section answers the research questions about how to classify two different fish species in a supervised manner using the two species' colour moment features [7]. Figure 1 shows the image processing flows, starting from the image gathering and resizing, until classifying the two species using the colour moment feature.

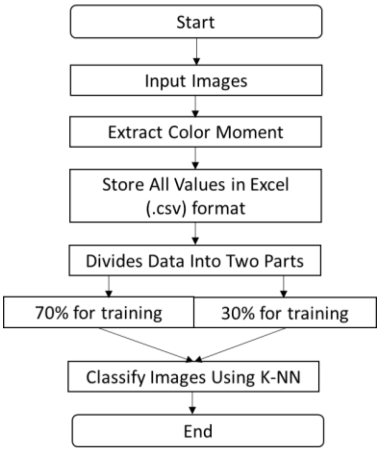


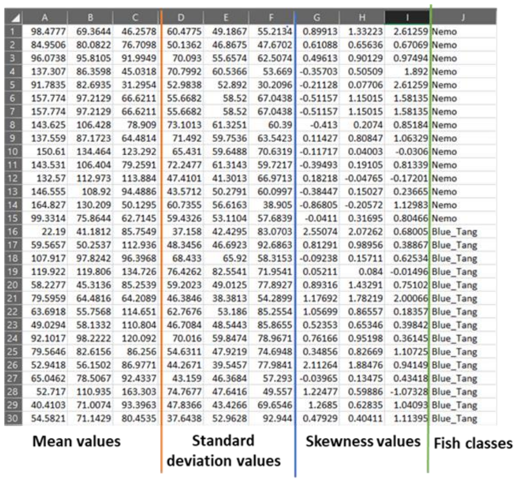
Figure 1: The process flow of image processing, starting from image acquisition until classifying the images.

3.2 Image Acquisition and Pre-Processing

Thirty (30) images were gathered from the Google image search [8] and divided into 15 images for each species. All gathered images were resized to a smaller size, so that a lower amount pixels are need to be extracted and computed. Thus, to avoid spending too much work during the extraction and computation of the images, the images were resized into 100 x 100 pixels so that the preferred region has a smooth process. Figure 2 shows two fish species from the ocean that are well known as Nemo (clown fish) and dory (blue tang fish).

Figure 2: (a) Nemo (Clown fish). (b) Dory (Blue Tang fish) (Retrieved from Google image search [8]).

The colour moment features were extracted by measuring the probability distribution in an image based on mean, standard deviation, and lack of symmetry in probabilistic distribution (or the skewness) from the RGB colour model for all images, as done in [7]. Then, all the extracted colour features were saved in an Excel file (.csv). The computed values were stored in four primary columns. The three main columns were divided by three colour lines, mean values from the first column until the orange line, standard deviation values from the orange line until the blue line, asymmetry of the probabilistic distribution (or the skewness) values from the blue line until the green line, and the last column consisted of the label for each of the species.



	A	B	C	D	E	F	G	H	I	J
1	98.4777	69.3644	46.2578	60.4775	49.1867	55.2134	0.89913	1.33223	2.61259	Nemo
2	84.9506	80.0822	76.7098	50.1362	46.8675	47.6702	0.61088	0.65636	0.67069	Nemo
3	96.0738	95.8105	91.9949	70.093	55.6574	62.5074	0.49613	0.90129	0.97494	Nemo
4	137.307	86.3598	45.0318	70.7992	60.5366	53.669	-0.35703	0.50509	1.892	Nemo
5	91.7835	82.6935	31.2954	52.9838	52.892	30.2096	-0.21128	0.07706	2.61259	Nemo
6	157.774	97.2129	66.6211	55.6682	58.52	67.0438	-0.51157	1.15015	1.58135	Nemo
7	157.774	97.2129	66.6211	55.6682	58.52	67.0438	-0.51157	1.15015	1.58135	Nemo
8	143.625	106.428	78.909	73.1013	61.3251	60.39	-0.413	0.2074	0.85184	Nemo
9	137.559	87.1723	64.4814	71.492	59.7536	63.5423	0.11427	0.80847	1.06329	Nemo
10	150.61	134.464	123.292	65.431	59.6488	70.6319	-0.11717	0.04003	-0.0306	Nemo
11	143.531	106.404	79.2591	72.2477	61.3143	59.7217	-0.39493	0.19105	0.81339	Nemo
12	132.57	112.973	113.884	47.4101	41.3013	66.9713	0.18218	-0.04765	-0.17201	Nemo
13	146.555	108.92	94.4886	43.5712	50.2791	60.0997	-0.38447	0.15027	0.23665	Nemo
14	164.827	130.209	50.1295	60.7355	56.6163	38.905	-0.86805	-0.20572	1.12983	Nemo
15	99.3314	75.8644	62.7145	59.4326	53.1104	57.6839	-0.0411	0.31695	0.80466	Nemo
16	22.19	41.1812	85.7549	37.158	42.4295	83.0703	2.55074	2.07262	0.68005	Blue_Tang
17	59.5657	50.2537	112.936	48.3456	46.6923	92.6863	0.81291	0.98956	0.38867	Blue_Tang
18	107.917	97.8242	96.3968	68.433	65.92	58.3153	-0.09238	0.15711	0.62534	Blue_Tang
19	119.922	119.806	134.726	76.4262	82.5541	71.9541	0.05211	0.084	-0.01496	Blue_Tang
20	58.2277	45.3136	85.2539	59.2023	49.0125	77.8927	0.89316	1.43291	0.75102	Blue_Tang
21	79.5959	64.4816	64.2089	46.3846	38.3813	54.2899	1.17692	1.78219	2.00066	Blue_Tang
22	63.6918	55.7568	114.651	62.7676	53.186	85.2554	1.05699	0.86557	0.18357	Blue_Tang
23	49.0294	58.1332	110.804	46.7084	48.5443	85.8655	0.52353	0.65346	0.39842	Blue_Tang
24	92.1017	98.2222	120.092	70.016	59.8474	78.9671	0.76166	0.95198	0.36145	Blue_Tang
25	79.5646	82.6156	86.256	54.6311	47.9219	74.6948	0.34856	0.82669	1.10725	Blue_Tang
26	52.9418	56.1502	86.9771	44.2671	39.5457	77.9841	2.11264	1.88476	0.94149	Blue_Tang
27	65.0462	78.5067	92.4337	43.159	46.3684	57.293	-0.03965	0.13475	0.43418	Blue_Tang
28	52.717	110.935	163.303	74.7677	47.6416	49.557	1.22477	0.59886	-1.07328	Blue_Tang
29	40.4103	71.0074	93.3963	47.8366	43.4266	69.6546	1.2685	0.62835	1.04093	Blue_Tang
30	54.5821	71.1429	80.4535	37.6438	52.9628	92.944	0.47929	0.40411	1.11395	Blue_Tang

Figure 3: The average, standard deviation and the lack of symmetry or skewness saved in excel file (.csv).

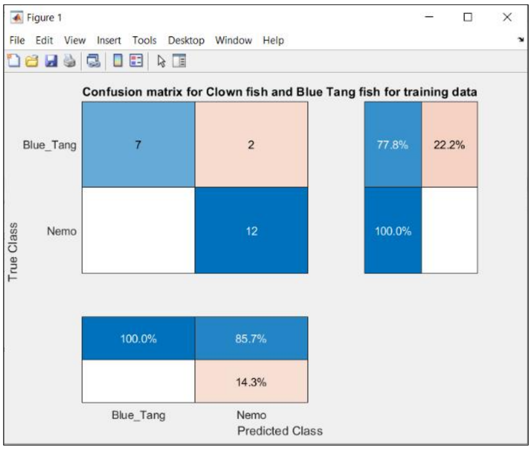
Before proceeding with the classification process, the data were divided into two categories, training and testing. In this research, the 70% of the data was used for training, and 30% was used for testing. The classification process was demonstrated by the K-NN classifier method for classifying the two species into their respective classes [6]. The K-NN model was implemented based on predictor X and Y, and the indicator value used for the closest neighbours was 10. Figure 4 shows the K-NN model code in MATLAB.

```
% build a K-NN model 10 closest neighbour
model = fitcknn(TrainFeatures,TrainClass,'NumNeighbors',10);
```

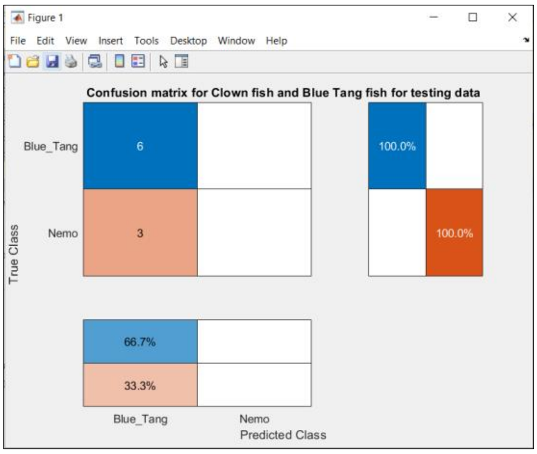
Figure 4: The model code of K-NN from MATLAB.

4 RESULTS AND DISCUSSIONS

This part discusses the results after the prediction process. The results were stored in the form of a confusion matrix to calculate the accuracy for the training and testing data. For training data, 2 out of 21 data from blue tang fish were wrongly predicted, while 12 out of 21 data from the clown fish was correctly predicted. Figure 5 shows the confusion matrix for both species using the K-NN classifier method [7]. For the testing data, 6 out of 6 from blue tang fish were correctly predicted, while 3 out of 6 data from clown fish was wrongly predicted.



(a)



(b)

Figure 5: (a) Training data of confusion matrix. (b) Testing data of confusion matrix.

The results from the confusion matrix were calculated for obtaining the accuracy for the testing and training data. Equation 1 shows the formula of how to obtain the accuracy of both data. The accuracy for testing is 0.8889~89%, and for training is 0.9048~90.48%. All of the predicted results were computed by MATLAB.

$$Accuracy = \frac{Predicted_{Clown\_fish} + Predicted_{Blue\_fish}}{Total\ of\ data} \quad (1)$$

```
Command Window
>> KNNclassification

TrainingAccuracy =

    0.9048


TestingAccuracy =

    0.8889


ans =

     7     2
     0    12


ans =

     6     0
     3     0
```

Figure 6: The accuracy results of train data and test data using the K-NN classifier method.

5 CONCLUSION

In this work, two types of fish species were classified using the colour moments feature with the K-NN classifier method. A total of 30 images were gathered from Google image search, and divided into two species, namely, clown fish and blue tang fish. The colour moment of the images was extracted using the RGB colour model by extracting the mean, standard deviation, and lack of symmetry (skewness) in probabilistic distribution or the skewness of each colour medium. Next, the data were divided into two parts for training and testing. Finally, the data were classified by the K-NN method according to their classes. The results show that the supervised method successfully classified the fish species according to their classes with an accuracy of 90.48% for training data and 89% for testing data. However, this research only focused on the supervised method, but an unsupervised method can be considered for future research.

REFERENCES

1. Dinc, M., & Hajiyev, C. (2015). Integration of navigation systems for autonomous underwater vehicles. Journal of Marine Engineering & Technology,14(1),32–43. https://doi.org/10.1080/20464177.2015.1022382
2. Xu, F., Ding, X., Peng, J., Yuan, G., Wang, Y., Zhang, J., & Fu, X. (2018). Real-Time Detecting Method of Marine Small Object with Underwater Robot Vision. 2018 OCEANS - MTS/IEEE Kobe Techno-Oceans(OTO).Published.https://doi.org/10.1109/oceanskobe.2018.8558804
3. Cui, S., Zhou, Y., Wang, Y., & Zhai, L. (2020). Fish Detection Using Deep Learning. Applied Computational Intelligence and Soft Computing, 2020, 1–13. https://doi.org/10.1155/2020/3738108
4. Han, F., Yao, J., Zhu, H., & Wang, C. (2020). Marine Organism Detection and Classification from Underwater Vision Based on the Deep CNN Method. Mathematical Problems in Engineering, 2020, 1–11. https://doi.org/10.1155/2020/3937580
5. L. (2016). STUDY OF OBJECT DETECTION IMPLEMENTATION USING MATLAB. International Journal of Research in Engineering and Technology, 05(08), 109–114. https://doi.org/10.15623/ijret.2016.0508020
6. Sari, Y. A., & Suciati, N. (2014). Flower Classification using Combined a* b* Color and Fractal-based Texture Feature. International Journal of Hybrid Information Technology, 7(2), 357–368. https://doi.org/10.14257/ijhit.2014.7.2.31
7. Chen, Z., Wang, F., Zhang, P., Ke, C., Zhu, Y., Cao, W., & Jiang, H. (2020). Skewed distribution of leaf color RGB model and application of skewed parameters in leaf color description model. Plant Methods, 16(1). https://doi.org/10.1186/s13007-020-0561-2
8. nemo and dory real fish - Google zoeken. (n.d.). Retrieved June 7, 2021, from https://www.google.com/search?q=nemo+and+dory+real+fish&sxsrf=ALeKk03Dxajc9kJWg76QQXnLbF_rkNwosg:1623081030390&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjB5oeC8IXxAhX8zDgGHRfsB9oQ_AUoAXoECAEQAw&biw=1536&bih=722&dpr=1.25