Marine surveillance has been one of the crucial issues to protect fisheries from illegal fishing activities (Angnew, 2009; Pelich, 2019; Sumaila et al., 2020). Illegal fishing takes approximately 20% of the global fishing amount (Angnew, 2009) and damages 26 to 50 billion USD on the global fishing economy (Sumaila et al., 2020). To counteract this matter, in particular, Republic of Korea operates vessel monitoring systems including Vessel Monitoring System (VMS), Automatic Identification System (AIS), V-Pass and VHF-DSC. However, those methods are not adaptable directly to fishing activity monitoring. The limitation requires more human resources to determine fishing status. Thus, this study proposes a method of estimating fishing activity from V-Pass, fishing vessel position reporting system, using Hidden Markov Model (HMM). HMM is a model to determine status through probability distribution for a sequence of time-series data. First of all, fishing activity status was labeled on V-Pass data. The distribution of speed on fishing activity was computed from the labeled data and HMM was constructed from the data obtained at Socheongcho Ocean Research Station (SORS). The model was first applied to the data of SORS for a test, and then Busan for validation. The model showed 99.4% and 89.6% as test and validation accuracy, respectively. It is concluded that the HMM can be applicable to predict a fishing activity from vessel tracks.

**Keywords**: Fishing Activity, Hidden Markov Model, V-Pass

**1. Introduction**

Marine surveillance has been one of the crucial issues to protect fisheries from illegal fishing activities (Angnew, 2009; Pelich, 2019; Sumaila et al., 2020). Illegal fishing takes approximately 20% of the global fishing amount (Angnew, 2009) and damages 26 to 50 billion USD on the global fishing economy (Sumaila et al., 2020). To counteract this matter, in particular, Republic of Korea operates vessel monitoring systems including AIS (Hong and Yang, 2014; Kim et al., 2016; 2018; Hong et al., 2018; Jeon and Jung, 2018; Jeon and Yang, 2021), V-Pass (Cho and Choi, 2018; Han et al., 2021; Lee et al., 2021) and VHF-DSC and others. But those systems are not yet ready to estimate fishing activity. Subsequently, additional human resources such as vessel traffic service officers, are necessary to determine whether a fishing vessel is engaged in fishing operation by monitoring the movement.

Thus there has been researches and development to counteract the threat, for instance, monitoring ships around coastal waters of each country using Automatic Identification System (AIS) (Hong and Yang, 2014; Kim et al., 2016; Souza et al., 2016; Koodsma et al., 2018; Hong et al., 2018; Jeon and Jung, 2018; Han et al., 2021; Jeon and Yang, 2021). But AIS is only a compulsory requirement for vessels larger than 300 gross tons or carrying passengers in Chapter V of the international convention for the Safety of Life at Sea (SOLAS) and most fishing boats are not obliged to equip the equipment provided that domestic regulation is not reinforced from the international regulation.

On the other hand, V-Pass data, a position reporting system of fishing vessels in Korea, is more effective for both monitoring and researching fishing vessel activities. V-Pass is one of the...
time-series marine traffic monitoring systems developed by and distributed in the Republic of Korea in 2015 (Han et al., 2021). Fishing vessels less than 10 tons must equip this reporting system while those over 10 tons are required to install AIS by the Korean law of fishing vessels. Since the system had been distributed on every fishing vessel, it opened the way to monitor and investigate the exact position and the navigation pattern of fishing vessels around Korean coastal water (Cho and Choi, 2018).

The researches on vessels’ activity using time-series data has been mostly based on AIS and rarely on V-Pass. Hong and Yang (2014) classified ships’ type and flag nearby Ieodo Ocean Research Station (IORS) using AIS. Souza et al. (2016) found fishing patterns from satellite-based AIS, where low orbit satellites receive the report, using data mining methods. Koodsma et al. (2018) processed and trained 4 years of global AIS by CNN and tried to detect fishing activities. Hong et al. (2018) also revealed that a remarkable number of ships do not report correct their flag and type by investigating ships maritime mobile service identity numbers. Jeon and Jung (2018) assessed collision risk quantitatively using AIS while Lee et al. (2021) applied vessel collision warning algorithms in fishing boats from V-Pass data. Han et al. (2021) classified fishing activity by grouping the speed range and course line of fishing boats from V-Pass data and created fishing activity maps. Previous studies based on AIS data have been focusing on activities of large vessels, of course, including large fishing vessels and they cannot represent fishing activities because the large vessels are not engaged in coastal fishing and the number of small fishing vessels is larger than large fishing vessels.

Researchers have tried to define the pattern of fishing activities by course and speed of the ship (Souza et al., 2016; Koodsma et al., 2018; Han et al., 2021). But the course itself is not sufficient to confirm fishing activity because fishing vessels have straight course lines outward from a harbor then randomly change their course when they reached fishing areas. The speed itself is also insufficient in a harbor because fishing boats lower their speed to alongside berth. Thus it is proper to analyze fishing activity using speed for data outside of the harbor. In general, the speed does not have an exact threshold value to distinguish whether it is engaged in fishing and non-fishing. Thus threshold method is not proper to analyze fishing activities.

On the other hand, Hidden Markov Model (HMM) can be a solution to remove the ambiguity of the threshold method instead. HMM assumes unknown parameters and classifies the unknown parameters from the observable parameters. The model has been applied in various fields such as bioinformatics, voice recognition, character recognition and furthermore, affected a number of machine learning algorithms (Franzese and Luliano, 2018).

Thus this study classified fishing activities like fishing and non-fishing outside of harbor data by probabilistic model using speed from V-Pass by adopting HMM. The labeling work of data and model generation is explained in Chapter 3. The discussion of HMM performance is in Chapter 4.

2. Area & Data

V-Pass data was collected for two coverages of Socheongcho (124.58-124.76E, 37.36-37.54N) for January 2, 2020, and Busan (128.5-130.0E, 34.5-35.5N) on February 5, 2021, as in Fig. 1. The V-Pass collection system is located at Socheongcho Ocean Research Station (SORS) and Korea Institute of Ocean Science and Technology (KIOST) and both are maintained and operated by KIOST. Because of transmission distance, the locations of fishing vessels are limited to coastal areas within 35 km off the coastline.

The collected data consists of a sequence of time-series data having 9 variables ID, time, longitude&latitude, heading, speed, status type, license type, SOS status, and accident type for every row. Among them, speed is a key element to estimate fishing vessel activity. SORS data contains 706 rows for a single fishing vessel while Busan data contains 16,699 rows for 244 vessels. The interval of V-Pass data is approximately from 1 second to 2 minutes.

Prior to using the V-Pass data, we checked if vessel locations are on land or the distance between land and the vessel is less...
than 1 km by land masking method because fishing activities are held out of harbors. There was no land-masked data on SORS while we found land-masked data in Busan data. Thus we eliminated the data that represents vessels arriving on/departing from harbors. If a ship has less than 10 locations, the ship was also removed from data because the reliability of manual labeling decrease.

3. Methodology

HMM require variables of feature and label to compose a model. In this study, speed is appointed as the feature to distinguish between fishing and non-fishing. Fig. 2 explains the procedure of generation and test of HMM in this study. Firstly, the labeling is made manually with defined patterns which will be explained in Section 3.1. Then each emission and transition probability is calculated from the distribution of speed according to the fishing activity status as in Section 3.2. The two types of probability are applied to generate HMM. The model is tested with the data of the model is generated then validated with the data of the independent area that the model is not generated from as in Chapter 4.

3.1 Labeling

The track of a single fishing vessel of SORS data and multiple fishing vessels in Busan were manually labeled considering dominant patterns during their fishing and non-fishing referring to literature and experts’ advice. Fishing vessels immediately lower their speed when they spread nets and keep their low speed when they heave net (Souza et al., 2016; Han et al., 2021). Then fishing vessels have zig-zag patterns during fishing activity as in Fig. 3. The zig-zag pattern decreases the speed drastically. Thus, we assigned fishing status if the vessels’ course line is a zig-zag pattern. If we could not determine status only by course line, we checked the speed whether it is higher than 3 knots which is the minimum speed of keeping course (Kim et al., 2014). If less than 3 knots, we considered the status as fishing because the vessel is moving without a specific destination. Clusters of points and rapid course change are also considered fishing.

We obtained the speed distribution of fishing boats from the labeled data. It was found that fishing and non-fishing are not exactly distinguished by speed as in Fig. 4(b) although it looks
like to divide them by 6 knots as a threshold value though as in Fig. 4(a).

3.2 HMM

Hidden Markov Model (HMM) calculates the probability of specific states, called hidden states using emission and transition probability from the time-series and states-labeled data (Franzese and Luliano, 2018). The model is based on the concept that current observation \((O_t)\) is the result of the current hidden state \((H_t)\) and that \(H_t\) is independent of the previous state \((H_{t-1})\). Figure 5 illustrates the structure of HMM for estimating fishing activity on the contiguous sea area of SORS. Here, we defined two hidden states as fishing (F) and non-fishing (NF) and observation as speed. Speed is grouped into four ranges. The probabilities of emission and transition are defined as \(P(O_t \mid H_t)\) and \(P(H_t \mid H_{t-1})\), respectively.

The probability of fishing is computed following Eq. (1). For each time step \((t)\), the emission probability is determined according to the speed group; meanwhile, the transition probability is alternately assigned by fishing activity between fishing and non-fishing. The states that the highest value obtained in a specific activity is determined as the final fishing activity status.

\[
P(H_{t-1} \mid O_{t-1}, r) = P(H_t)P(O_t \mid H_t)\prod_{i=1}^{T} P(O_t \mid H_t)P(H_t \mid H_{t-1})
\] (1)

Emission probability means the probability of observation is calculated from hidden states. The emission probability of a single vessel in SORS data was computed for the four observations in fishing and non-fishing separately. The observation “first” during fishing takes the highest proportion as 86.08% while non-fishing has the highest in the observation “third” as in Table 1 and as already shown in Fig. 4(a).

The transition probability is defined as the probability of a hidden state in the previous step \((t-1)\) become a hidden state in the current step \((t)\). Transition probability requires an initial setting. We presumed that every fishing vessel does not start fishing from an initial point and thus give ratios 0 and 1 for fishing and non-fishing as initial setting values as in Table 2. Table 3 shows the transition probability. The transition probability shows that the crossing transition, fishing to non-fishing or non-fishing to fishing, is seldom.

| Observation sequences | Count(Probability) |
|-----------------------|--------------------|
|                      | Fishing            |
|                      | Non-fishing        |
| First                | 470 (86.08%)       |
| Second               | 74 (74%)           |
| Third                | 2 (0.37%)          |
| Fourth               | 0 (0%)             |
| Sum                  | 546 (100%)         |

Table 1. Emission probability from a single vessel of SORS data

| Ratio(Probability) | Fishing | Non-Fishing |
|-------------------|---------|-------------|
| 0 (0%)            |         | 1 (100%)    |

Table 2. Initial setting of transition probability

| t-1   | t     | Count(Probability) |
|-------|-------|--------------------|
| Fishing | 544 (99.63%) | 2 (0.37%) |
| Non-Fishing | 2 (1.26%) | 157 (98.74%) |

Table 3. Transition probability obtained from a single vessel at SORS

\[ \text{Fig. 5. HMM structure for the SORS data. Hidden states consists of fishing activity status, and observation sequences consist of 4 different ranges of speed. The probabilities of emission and transition computed from labeled data are the key to operate HMM} \]
4. Result

The performance of generated HMM was evaluated by accuracy, precision, recall and f1-score by Eqs. (2)-(5) and the result is described in Table 4. Beforehand, four parameters which are consisted of True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), should be calculated by comparing the label of actual and prediction. TP represents both actual and prediction are positive and TN indicates both are negative. FP means that actual is negative but prediction is positive and FN means that actual is positive but prediction is negative.

Accuracy represents that the prediction is correct among overall cases. Precision, called positive predictive value, is that prediction is positive among the model classified positive. Recall, called sensitivity, shows how much positive model answered among actual is positive. F1-score, harmonized mean of precision and recall, is useful when the number of data labels is unbalanced.

\[
\text{Accuracy} = \frac{TP + TN}{2TP + TN + FP + FN} \times 100 \tag{2}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \times 100 \tag{3}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \times 100 \tag{4}
\]

\[
\text{f1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \tag{5}
\]

The time-series speed and label of a single fishing vessel in SORS data were used for model generation, as in Fig. 5, and test as in Fig. 6(a). The edge parts where the vessel changed its course rapidly and the speed is slower than 6 knots were labeled as fishing points as in Fig. 6(a). The probability of emission and transition were computed using the series of speed and label as in Fig. 4 and Table 1 to Table 3. Then the model classified fishing and non-fishing as in Fig. 6(b). The results were 99.43% and 99.63% in accuracy and f1-score, respectively, as in Table 4. The confusion matrix of fishing activity status classification for the single vessel in SORS data explains the performance, showing only 1 and 3 cases were mismatched as in Table 5. The performance was over 99.43%. The model was needed to be verified with independent labeled data.

The fishing activities of multiple fishing vessels in Busan were labeled as in Fig. 7(a). The labeled data were used in the verification of the model. The model classified the status of vessels as in Fig. 7(b). The accuracy and f1-score were 89.60% and 92.82% as in Table 4. The confusion matrix of status classification for multiple vessels in Busan explains the performance as in Table 6. The performance was approximately 90% and revealed the generated model is proper to estimate fishing activity.

| Area | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) |
|------|--------------|---------------|------------|--------------|
| SORS | 99.43        | 99.45         | 99.81      | 99.63        |
| Busan| 89.60        | 91.86         | 93.80      | 92.82        |

Table 4. Performance of HMM-Based fishing activity prediction

| Actual     | Fishing | Non-Fishing |
|------------|---------|-------------|
| Fishing    | 157     | 3           |
| Non-Fishing| 1       | 545         |

Table 5. Confusion matrix of a single vessel at SORS

Fig. 6. Fishing activity status of a single vessel from the SORS data: (a) actual (b) prediction. Blue and red dots represent fishing and non-fishing, respectively. Prediction accuracy is 99.43% and it represents that the model is overfitted with SORS data.
In this study, we propose to classify fishing activity status from V-Pass using HMM. The fishing statuses first were labeled on V-Pass data manually after defining fishing patterns for two independent areas: a contiguous sea area of SORS and Busan. Then the speed distribution of fishing and non-fishing from the labeled data revealed that a threshold method is not sufficient to distinguish fishing activities. The probabilities of emission and transition were calculated from test data of SORS and then HMM was generated. The model was tested and showed its accuracy over 99%. For the validation of the model, to check whether it has overfitted, it was applied to another independent data on Busan. The performance was around 90% and revealed that the use of HMM is applicable to determine fishing activity status.

For the improvement of this study result, there is a need to define more patterns of fishing activities according to the types of fishing gear because the model is only based on the speed of a single vessel trace on SORS data. Although the verification result was satisfactory there were still many mismatches between prediction and actual. To compensate for this, the fishing pattern will be investigated from literature review and expert counseling. Then the model will perform a more accurate classification. In addition, the study area was limited to 35 km off from coastline.

Although the use of V-Pass data contributes to enhancing the analysis of coastal fishing, it does not represent overall fishing off the coast of Republic of Korea. Thus the locations of vessels where their type is not identified in satellite AIS will be investigated whether they are engaging in fishing using HMM.

5. Conclusion and Future Work

In this study, we propose to classify fishing activity status from V-Pass using HMM. The fishing statuses first were labeled on V-Pass data manually after defining fishing patterns for two independent areas: a contiguous sea area of SORS and Busan. Then the speed distribution of fishing and non-fishing from the labeled data revealed that a threshold method is not sufficient to distinguish fishing activities. The probabilities of emission and transition were calculated from test data of SORS and then HMM was generated. The model was tested and showed its accuracy over 99%. For the validation of the model, to check whether it has overfitted, it was applied to another independent data on Busan. The performance was around 90% and revealed that the use of HMM is applicable to determine fishing activity status.

For the improvement of this study result, there is a need to define more patterns of fishing activities according to the types of fishing gear because the model is only based on the speed of a single vessel trace on SORS data. Although the verification result was satisfactory there were still many mismatches between prediction and actual. To compensate for this, the fishing pattern will be investigated from literature review and expert counseling. Then the model will perform a more accurate classification. In addition, the study area was limited to 35 km off from coastline.

Although the use of V-Pass data contributes to enhancing the analysis of coastal fishing, it does not represent overall fishing off the coast of Republic of Korea. Thus the locations of vessels where their type is not identified in satellite AIS will be investigated whether they are engaging in fishing using HMM.

Acknowledgements

This research is a part of the projects entitled “Development of satellite based system on monitoring and predicting ship distribution in the contiguous zone”, funded by the Korea Coast Guard, and “Establishment of the ocean research station in the jurisdiction zone and convergence research”, funded by the Ministry of Oceans and Fisheries, Korea.

References

Agnew, D. J., Pearce, J., Pramod, G., Peatman, T., Watson, R., Beddington, J. R. and Pitcher, T. J. (2009). “Estimating the worldwide extent of illegal fishing.” *PLoS ONE*, Vol. 4, No. 2, e4570. https://doi.org/10.1371/journal.pone.0004570

Cho, S. J. and Choi, H. J. (2018). “Recent trends and their Implications of marine activities mapping for marine spatial planning.” *J. Korean Soc. Marine Environment & Energy.*, Vol. 21, No. 4, pp. 270-280. https://doi.org/10.7846/JKOSMEE.2018.21.4.270 (in Korean).

De Souza, E. N., Boerder, K., Matwin, S. and Worm, B. (2016). “Improving fishing pattern detection from satellite AIS using data mining and machine learning.” *PloS one.*, Vol. 11, No. 7, e0158248. https://doi.org/10.1371/journal.pone.0163760

Franzese, M. and Luliano, A. (2018). “Reference Module in Life Sciences: Hidden Markov Models.” *Science Direct*, pp. 756-792. https://doi.org/10.1016/B978-0-12-809633-8.20488-3
Han, J. R., Kim, T. H., Choi, E. Y. and Choi, H. W. (2021). “A Study on the Mapping of Fishing Activity using V-Pass Data-Focusing on the Southeast Sea of Korea.” J. Kor. Asso. Geographic Information Studies., Vol. 24, No. 1, pp. 112-125. https://doi.org/10.11108/kagis.2021.247.1.112

Hong, D. B. and Yang, C. S. (2014). “Classification of Passing Vessels Around the Ieodo Ocean Research Station Using Automatic Identification System (AIS): November 21-30, 2013.” J. Korean Soc. Marine Environment & Energy., Vol. 17, No. 4, pp. 1-9. https://dx.doi.org/10.7846/JKOSMEE.2014.17.4.1 (in Korean).

Hong, D. B., Yang, C. S. and Kim, T. H. (2018). “Investigation of Passing Ships in Inaccessible Areas Using Satellite-based Automatic Identification System (S-AIS) Data.” Korean J. Remote Sens., Vol. 34, No. 4, pp. 579-590. http://dx.doi.org/10.7780/kjr.2018.34.4.1

Jeon, H. K. and Jung, Y. C. (2018). “Development of a Collision Risk Assessment System for Optimum Safe Route.” J. Kor. Soc. Marine Environment & Safety., Vol. 20, No. 2, pp. 193-201. http://dx.doi.org/10.7837/kosomes.2014.2.2.193 (in Korean).

Kim, T. H., Jeong, J. H. and Yang, C. S. (2016). “Construction and Operation of AIS System on Socheongcho Ocean Research Station.” J. Coastal Disaster Prevention., Vol. 3, No. 2, pp. 74-80. http://dx.doi.org/10.20481/kscdp.2016.3.2.74 (in Korean).

Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerder, K., Ferretti, F., Wilson, A., Bergman, B., White, T. D., Block, B. A., Woods, P., Sullivan, B., Costello, C. and Worm, B. (2018). “Tracking the global footprint of fisheries.” Science., Vol. 359, No. 6378, pp. 904-908. http://doi.org/10.1126/science.aao5646

Lee, M. K., Park, Y. S., Park, S. W., Lee, E. K., Park, M. J. and Kim, N. E. (2021). “Application of Collision Warning Algorithm Alarm in Fishing Vessel’s Waterway.” Applied Science., Vol. 11, 4479. https://doi.org/10.3390/app1104479

Pelich, R., Chini, M., Hostache, R. Matgen, P., Lopez-Martinez, C., Nuevo, M., Ries, P. and Edien, G. (2019). “Large-Scale Automatic Vessel Monitoring Based on Dual-Polarization Sentinel-1 and AIS Data.” Remote Sens., Vol. 11, 1078. doi:10.3390/rs11091078

Sumaila, U. R., Zeller, D., Hood, L., Palomares, M. L. D. and Pauly, D. (2020). “Illicit trade in marine fish catch and its effects on ecosystems and people worldwide.” Science advances., Vol. 6, pp. 1-7. https://doi.org/10.1126/sciadv.aaz3801