Bagged Fuzzy k-nearest Neighbors for Identifying Anomalous Propagation in Radar Images

Hansoo Lee, Jonggeun Kim, Suryo Adhi Wibowo and Sungshin Kim
Department of Electrical and Computer Engineering
Pusan National University
Busan, 46241, Republic of Korea
Email: {hansoo, wisekim, suryo, sskim}@pusan.ac.kr

***Abstract*—Several advanced observation devices, such as radiosondes, satellites, and radars, are utilized in practical weather prediction. The weather radar is an essential device because of its broad coverage with excellent resolution. However, the radar inevitably observes meteorologically irrelevant signals. An anomalous propagation echo is a nonprecipitating echo generated by significantly refracted radar beam towards ground or sea surface. In the case, the radar misrecognizes the surface as a meteorological phenomenon. The false observation results may decrease the accuracy of weather prediction result. Therefore, we propose a novel classification method for identifying anomalous propagation echoes in the radar data by combining fuzzy k-nearest neighbors and Hamamoto's bootstrapping algorithm. By using actual occurrence cases of anomalous propagation, we confirm that the proposed method provides good classification results.**

Keywords–Fuzzy k-nearest neighbors; bootstrap aggregating; anomalous propagation; weather prediction.

I. INTRODUCTION

There are several advanced devices to observe meteorologically related events in the atmosphere, such as satellite, ground-based weather radar, radiosonde, and so on. The ground-based weather radar is one of the essential devices because of its wide array of advantages, such as high resolution and a wide range of observation [1]. The primary purposes of utilizing the weather radar are locating precipitation echoes and calculating quantitative precipitation estimation.

The radar transmitter should emit intense electromagnetic waves, and the radar receiver should be designed to obtain weak signals due to the following reasons: the intervals and sizes of the expected reflecting objects such as raindrops and snowflakes; exceedingly small amounts of the waves can return to the radar receiver. As a result, the observed outcomes inevitably contain unwanted signals. Furthermore, the ground-based weather radar is frequently affected by return signals that do not originate from the precipitation echoes, such as stationary or moving objects in the atmosphere. Even refracted radar beams towards the ground make significant false signals in the radar image.

In actual weather prediction, hence, there is a quality control process [2] to remove the nonprecipitating echoes. The quality control process highly relied on the expert's knowledge at the beginning. However, currently applied quality control process utilizes data mining courtesy of advances in techniques. For example, an anomalous propagation, which appears by abnormally refracted radar beam towards ground or sea surface, is one of the representative nonprecipitating echoes. There are many successful research results using data mining method for identifying the anomalous propagation: artificial neural network [3][4][5], fuzzy inference system [6][7], Bayesian classifier [8][9] and case study [10].

In this paper, we propose a novel approach of k-nearest neighbor algorithm by combining Hamamoto's bootstrapping method and fuzzy set theory for identifying the anomalous propagation echo. The k-nearest neighbor is one of the most popular data mining techniques because of its simple operation principle and good performance. Also, Hamamoto's bootstrapping method, which is a variant of the bootstrap aggregating method, has already proved its ability to improve classification accuracy by comparative studies.

The rest of the paper organized as follows. In Section 2, we describe the characteristics of anomalous propagation echo. In Section 3, we explain not only the proposed algorithm but its components in detail. After experimental results and analysis in Section 4, we elucidate conclusions and future works.

II. ANOMALOUS PROPAGATION ECHO

The weather radar observes floating objects in the atmosphere by transmitting and receiving intense electromagnetic waves as other kinds of remote sensing devices do. Therefore, pathways of the waves highly depend on the atmospheric conditions, such as temperature, humidity, and so on. The conditions refract the paths to abnormal directions. As shown in Figure 1, the pathways can be classified into four different types: sub-refraction, normal refraction, super-refraction and ducting.

The sub-refraction indicates a radar beam path refracted the opposite direction of the surface more than the normal refraction. And the super-refraction means a radar beam path bent the direction of the surface more. Further, a radar beam can be stuck in a certain atmospheric layer if it refracted more severe than a critical gradient. Considering that the weather radar assumes the altitude of the objects based on the normal refraction, the other types of refracted radar beams can cause a severe error in radar data.

As shown in Figure 1, there is a chance to miss the precipitation echo when the sub-refraction occurs. Even if the refracted radar beam can detect the precipitation echo, the miscalculated altitude of the echo causes erroneous observation results. Also, when the super-refraction or ducting occur, the

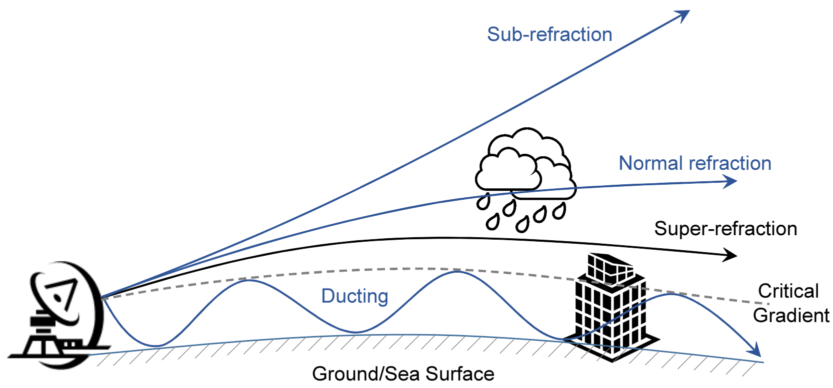

Figure 1. Anomalous propagation echo

radar beam faces on the surface as shown in Figure 1. Then the weather radar will get return signals that do not originate from the precipitation echoes, such as geographic features or meteorologically unrelated floating objects in the atmosphere. Usually, the refracted radar beams towards the surface make severe false observation results in the radar data. For example, the wrong rainfall estimation like an overestimation of rainfall quantity. Furthermore, it also can miss the precipitation echo.

In weather prediction process, meteorologists utilize several complex rules to identify the anomalous propagation echo in radar images. The representative rules are listed as follows.

1) The echo has near-zero Doppler velocity.
   a) On ground surface $= 0$ m/s.
   b) On sea surface ≈ 0 m/s.
2) The echo has discontinuous reflectivity distribution in vertical and horizontal directions.
3) The echo usually locates at low altitude which makes difficult to separate precipitation echoes at the similar region.

According to the list, it is reasonable to consider Doppler velocity, reflectivity, and altitude as classification attributes. In this paper, we chose six classification inputs based on the features: minimum and average Doppler velocity; minimum, maximum and average reflectivity; centroid altitude.

III. Bagged Fuzzy k-Nearest Neighbors

The nearest neighbor algorithm, first introduced in [11], is a nonparametric method for pattern classification based on instances. It has become an active research area in machine learning since proposed. Its popular variant, called the k-nearest neighbor algorithm, is selected as one of the top ten algorithms in data mining [12]. The primary advantages of the k-nearest neighbor algorithm are its simplicity to use and also its often good performance. However, it has also some drawbacks: the necessity of storage, low efficiency of the computation of the decision rule, low tolerance to noise, and high dependency on the given instances [13]. Therefore, lots of researches have been conducted to solve the problems.

The fundamental improvements over k-nearest neighbors are as follows. The first is applying weights in k-nearest neighbors, called as wk-nearest neighbors [14]. The second is generating an artificial training set by using a bootstrapping method. The classic bootstrapping, which changes the training set slightly, has a weak influence to the k-NN because the algorithm is stable. However, bootstrapping method by [15] have a positive effect on improving performances of the k-nearest neighbors classifier. The third is using fuzzy set theory in k-nearest neighbors, called as fuzzy k-nearest neighbors [16].

We select the second and third approaches to improve a performance of k-nearest neighbors algorithm in this paper. The rest of this section organizes as follow. First, we describe the Hamamoto's bootstrapping method and the fuzzy k-nearest neighbor algorithm in sequence. After that, we elucidate our proposed method.

A. Hamamoto's Bootstrapping Method

Bootstrap aggregating, called as bagging, is one of ensemble methods, which uses random sampling methods to improve the performance of the classifier by allowing the classifier to utilize newly created training samples [17]. The classical bagging uses random sampling with replacement to generate samples.

An attempt of combining the bagging with k-nearest neighbors already conducted, but the outcomes were not satisfactory because the k-nearest neighbors is a stable algorithm [17]. In other words, small changes in the training samples do not lead to improving the performance of the classifier significantly. However, Hamamoto's bootstrapping methods [15], one of variant bagging methods, showed remarkable classification results with k-nearest neighbors. We selected the Hamamoto's bootstrapping II method among four different suggestions to creating training samples. The main reason why we chose the way among them is that all the original training samples participate in generating bootstrap samples by the locally weighted sum.

Figure 2 describes the Hamamoto's bootstrapping II method when $k = 3$ in a binary class problem. Let us assume that there are given samples as shown in Figure 2(a). As a first step of the bootstrapping method, it separates the given samples according to its class which is expressed as white

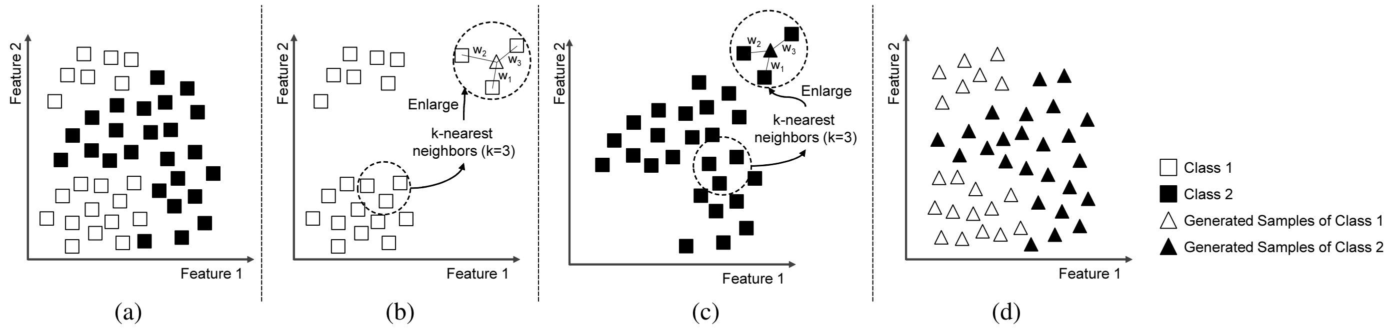

Figure 2. Hamamoto's bootstrap II method: (a) original samples, (b) class 1 data, (c) class 2 data, (d) generated samples

squares in Figure 2(b) and black squares in Figure 2(c). After separation process, it derives nearest neighbor samples of each training sample by utilizing k-nearest neighbor including the selected one. Then, the bootstrapping sample is created using the selected samples and locally weighted sum as shown in (1).

$$\begin{aligned}\mathbf{x}_i^b &= \sum_{j=0}^{r} \omega_j \mathbf{x}_{i,j} \\ &= \omega_0 \mathbf{x}_{i,0} + \omega_1 \mathbf{x}_{i,1} + \cdots + \omega_r \mathbf{x}_{i,r}\end{aligned} \tag{1}$$

where $\mathbf{x_i^b}$ means the i-th bootstrap sample, and $\mathbf{x_{i,j}}$ indicates the j-th nearest neighbor sample of the i-th original sample. The ω_j means weight derived by (2).

$$\omega_j = \frac{\Delta_j}{\sum_{c=0}^{r} \Delta_c}, \qquad 0 \le j \le r \tag{2}$$

where Δ_j is chosen from a uniform distribution. As shown in Figure 2(b) and Figure 2(c), the bagging samples are represented as triangular-shape, and their color indicates class information. Finally, it is possible to obtain bagging samples, as shown in Figure 2(d), by repeating the mentioned process until all of the original data is selected.

B. Fuzzy k-Nearest Neighbors

The fundamental principle of the fuzzy k-nearest neighbors is to assign membership as a function of the selected sample's distance from its nearest neighbors and memberships of the neighbors in the possible classes [16]. The scheme is similar to the k-nearest neighbors in the sense that there is a search process for the training sample set. However, the class assign process differs significantly from the search process. The membership of the sample x is computed by (3).

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij} \left(\frac{1}{\|x - x_j\|^{\frac{2}{(m-1)}}} \right)}{\sum_{j=1}^{k} \left(\frac{1}{\|x - x_j\|^{\frac{2}{(m-1)}}} \right)} \tag{3}$$

where k is the number of nearest neighbors, i indicates class, j is an index of the nearest neighbors, and m is a parameter to determine a type of distance. The most frequently used distance is the Euclidean distance ($m = 2$) when the attributes of the samples are normalized.

Also, it is necessary to define $u_{i,j}$ because the parameter determines class membership as shown in (4).

$$u_{i,j} = \begin{cases} 0.51 + (n_j/k) * 0.49, & \text{if } j = i \\ (n_j/k) * 0.49, & \text{if } j \neq i \end{cases} \tag{4}$$

where n_j is the number of the neighbors which belong to the jth class. This method makes the samples fuzzified by considering the labels of the samples and its neighbors. By utilizing (3), (4), and inverses of the distances from the nearest neighbors, the class of given sample with unknown label can be derived.

C. Proposed Approach

In this paper, we propose a novel nearest neighbor method named bagged fuzzy k-nearest neighbors classifier to improve classification performance by combining two techniques, fuzzy k-nearest neighbors and Hamamoto's bootstrapping II method. Figure 3 shows an overview of the proposed system.

At first, a hierarchical clustering categorizes a given radar data for deriving input attributes. Theoretically, there are a lot of data points to consider in the radar data due to its wide range of observation: for example, over 9 million points should be considered if a radar has 240km observation radius along 10km altitude. Therefore, for deriving attributes efficiently, we applied a hierarchical clustering. After the clustering process, it is possible to use the clusters as training data. From the clusters, we derived six attributes for classification: centroid altitude of the cluster; mean and maximum reflectivity; minimum, maximum and mean Doppler velocity. The reason why we selected these attributes is to reflect expert's knowledge mentioned the previous section.

The training data is used to generate the N number of the artificial training dataset. Simultaneously, the parameter k is derived by k-fold cross validation method using the original training dataset. Including the initial and artificial training datasets, it is possible to implement the bagged fuzzy k-nearest neighbors by utilizing the $(N + 1)$ datasets.

 ISBN: 978-1-61208-576-0

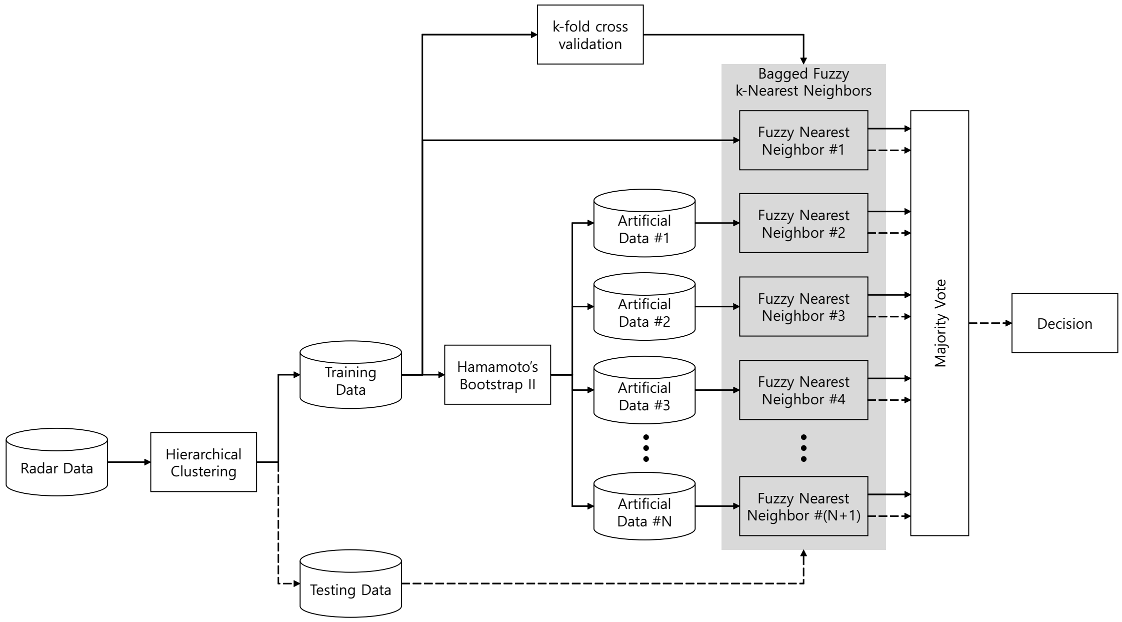

Figure 3. Overview of proposed method

Finally, a sample with an unknown class from the testing dataset can obtain its class by the majority vote process of the proposed classifier as shown in (5).

$$\text{Decision}(X) = \arg\max \sum_{j=1}^{N+1} I\left(\text{FNN}_j(X) = i\right) \qquad (5)$$

where X is an attribute vector from learning data, which consists of six elements mentioned above. And $I(\cdot)$ is an indicator function and FNN_j is a j-th fuzzy k-nearest neighbors classifier. Note that the N should be even to avoid a tie result of the majority voting process.

IV. Experimental Results

For evaluating and verifying the proposed method, we used actual anomalous propagation echo occurrence cases. As mentioned earlier, we derived six attributes as inputs for classification according to experts' knowledge: centroid altitude of the cluster; mean and maximum reflectivity; minimum, maximum and mean Doppler velocity.

Figure 4 shows a complicated example of anomalous propagation echo appearance. A significant precipitation echo exists on the left-upper side of Figure 4(a). And the central region shows the anomalous propagation echo. By using the proposed method, we could obtain the successful classification results as shown in Figure 4(b) and (c).

Figure 5 shows another example of the anomalous propagation echo case. Almost all of the observation area is distorted by significant anomalous propagation echo as shown in Figure 5. By using the proposed method, we also could obtain the successful classification results as shown in Figure 5(b) and (c), respectively.

For comparing the performance with other nearest neighbors classifier, we conducted evaluations using accuracy as shown in (6).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (6)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. Also, in this paper, the true indicates the anomalous propagation echo, and the false indicates the non-anomalous propagation echo, respectively.

We compared the proposed method with four kinds of nearest neighbors classifiers: 1-NN, k-NN, fuzzy k-NN, bagged k-NN. 1-NN showed the worst classification accuracy: 84.96%. k-NN showed 87.61%, and bagged k-NN showed 89.52% accuracy. Fuzzy k-NN showed better accuracy than k-NN: 89.05%. And the proposed method derived the best accuracy: 92.38%. From the experimental results, we can conclude that the proposed method can classify the anomalous propagation echo successfully.

V. Conclusion

An anomalous propagation echo is a nonprecipitating echo generated by significantly refracted radar beam towards the surface. The false observation results may decrease the accuracy of weather prediction. Therefore, we proposed a novel approach of k-nearest neighbor algorithm by combining Hamamoto's bootstrapping II method and fuzzy k-nearest neighbors. The fuzzy k-nearest neighbor proves its remarkable performance with simple operation. Also, Hamamoto's bootstrapping II method has demonstrated its ability to improve classification accuracy by comparative studies. By experiments with actual anomalous propagation echo cases, we proved that the proposed method could classify the echo from radar data successfully.

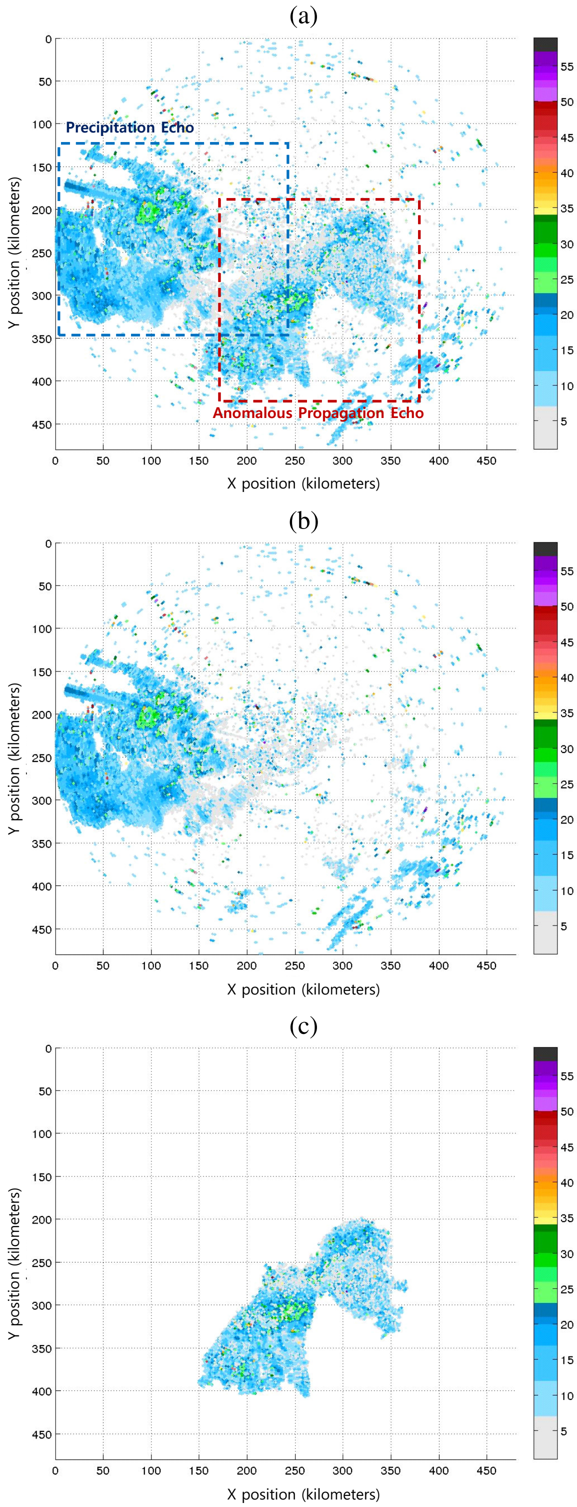

Figure 4. Experimental result, case 1: (a) original radar image, (b) image without classified anomalous propagation echo, (c) classified anomalous propagation echo

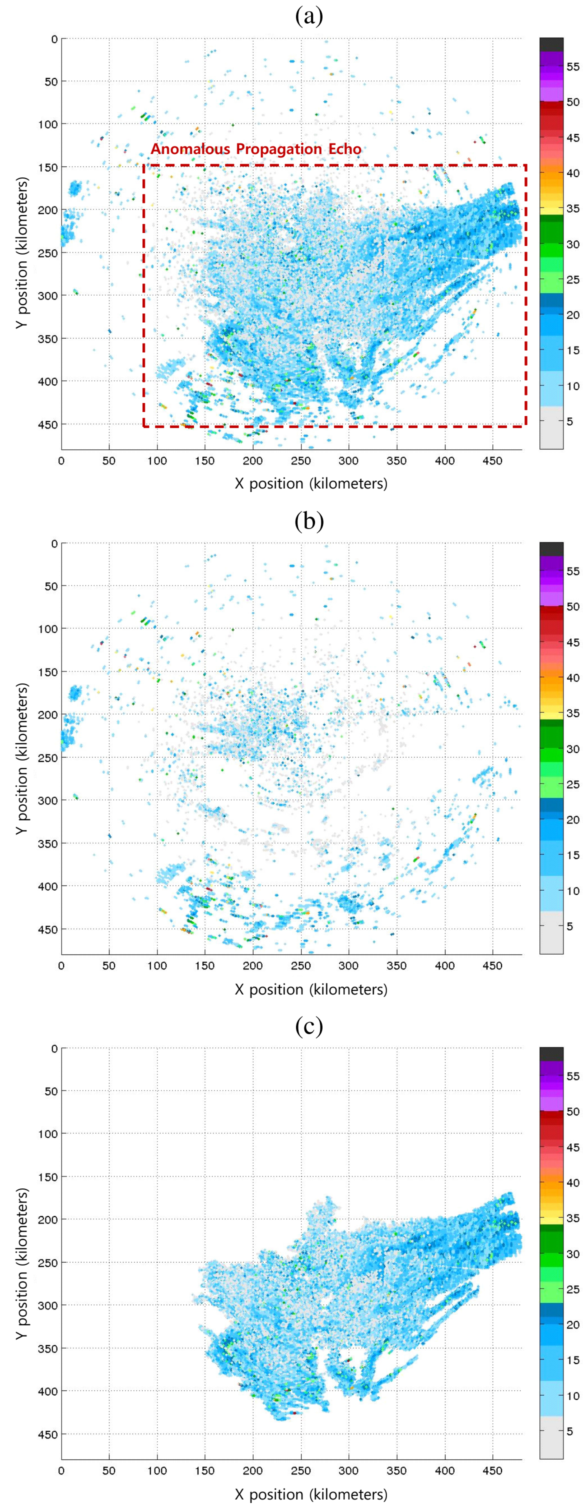

Figure 5. Experimental result, case 2: (a) original radar image, (b) image without classified anomalous propagation echo, (c) classified anomalous propagation echo

In future works, we will try to improve the accuracy of the proposed method. We consider replacing the fuzzification algorithm such as interval type-2 fuzzy logic. Also, we think it is possible to obtain better classification results by combining clustering methods. Further, we will apply the proposed method to the recognition of other echoes, such as chaff echo, interference patterns, and so on.

ACKNOWLEDGMENT

This research was supported by the MOTIE (Ministry of Trade, Industry & Energy), Korea, under the Industry Convergence Liaison Robotics Creative Graduates Education Program supervised by the KIAT (N0001126).

REFERENCES

[1] I. S. Merrill, "Introduction to radar systems," Mc Grow-Hill, 2001.

[2] V. Lakshmanan, A. Fritz, T. Smith, K. Hondl, and G. Stumpf, "An automated technique to quality control radar reflectivity data," Journal of applied meteorology and climatology, vol. 46, no. 3, 2007, pp. 288–305.

[3] R. B. da Silveira and A. R. Holt, "An automatic identification of clutter and anomalous propagation in polarization-diversity weather radar data using neural networks," IEEE transactions on geoscience and remote sensing, vol. 39, no. 8, 2001, pp. 1777–1788.

[4] M. Grecu and W. F. Krajewski, "An efficient methodology for detection of anomalous propagation echoes in radar reflectivity data using neural networks," Journal of Atmospheric and Oceanic Technology, vol. 17, no. 2, 2000, pp. 121–129.

[5] ——, "Detection of anomalous propagation echoes in weather radar data using neural networks," IEEE transactions on geoscience and remote sensing, vol. 37, no. 1, 1999, pp. 287–296.

[6] M. Berenguer, D. Sempere-Torres, C. Corral, and R. Sánchez-Diezma, "A fuzzy logic technique for identifying nonprecipitating echoes in radar scans," Journal of Atmospheric and Oceanic Technology, vol. 23, no. 9, 2006, pp. 1157–1180.

[7] Y.-H. Cho, G. W. Lee, K.-E. Kim, and I. Zawadzki, "Identification and removal of ground echoes and anomalous propagation using the characteristics of radar echoes," Journal of Atmospheric and Oceanic Technology, vol. 23, no. 9, 2006, pp. 1206–1222.

[8] J. Pamment and B. Conway, "Objective identification of echoes due to anomalous propagation in weather radar data," Journal of Atmospheric and Oceanic Technology, vol. 15, no. 1, 1998, pp. 98–113.

[9] J. R. Peter, A. Seed, and P. J. Steinle, "Application of a bayesian classifier of anomalous propagation to single-polarization radar reflectivity data," Journal of Atmospheric and Oceanic Technology, vol. 30, no. 9, 2013, pp. 1985–2005.

[10] W. F. Krajewski and B. Vignal, "Evaluation of anomalous propagation echo detection in wsr-88d data: A large sample case study," Journal of Atmospheric and Oceanic Technology, vol. 18, no. 5, 2001, pp. 807–814.

[11] E. Fix and J. L. Hodges Jr, "Discriminatory analysis-nonparametric discrimination: consistency properties," DTIC Document, Tech. Rep., 1951.

[12] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, and S. Y. Philip, "Top 10 algorithms in data mining," Knowledge and information systems, vol. 14, no. 1, 2008, pp. 1–37.

[13] J. Derrac, S. García, and F. Herrera, "Fuzzy nearest neighbor algorithms: Taxonomy, experimental analysis and prospects," Information Sciences, vol. 260, 2014, pp. 98–119.

[14] S. A. Dudani, "The distance-weighted k-nearest-neighbor rule," IEEE Transactions on Systems, Man, and Cybernetics, no. 4, 1976, pp. 325–327.

[15] Y. Hamamoto, S. Uchimura, and S. Tomita, "A bootstrap technique for nearest neighbor classifier design," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 1, 1997, pp. 73–79.

[16] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," IEEE transactions on systems, man, and cybernetics, no. 4, 1985, pp. 580–585.

[17] L. Breiman, "Bagging predictors," Machine learning, vol. 24, no. 2, 1996, pp. 123–140.