

전자해도 수심 밀집도 개선기법 연구

† 오세웅 · 이문진* · 김혜진** · 서상현***

†, *, **, *** 한국해양연구원 해양시스템안전연구소, 해양안전·방재기술연구부

A Study on improvement method of sounding density of ENCs

† Se-Woong Oh · Moon-Jin Lee* · Hye-Jin Kim** · Sang-Hyun Suh***

†, *, **, *** Maritime & Ocean Engineering Research Institute, Korea Ocean Research & Development Institute, Daejeon 305-343, Korea

요 약 : 전자해도는 종이해도 간행을 위한 수치해도를 변환하여 편집되며 그리드 형식의 격자형 체계로 간행된다. 이에 따라 전자해도의 수심정보 밀집도가 일관적이지 못해, 정보 분포에 관한 개선이 요구되어 왔다. 본 연구에서는 위성영상 분류기법 중 K-Means 기법과 ISODATA 기법을 검토 하여, 이를 전자해도 수심정보에 맞게 수정 및 적용 하였다. 전자해도 수심 밀집도 개선기법은 전자해도 로딩 부분, 수심 밀집도 개선 부분, 전자해도 쓰기 부분으로 설계 및 개발하였으며, 개발결과에 조건식에 따른 변수 설정을 달리하여 수심 밀집도 개선 및 변경 결과를 확인하였다.

핵심용어 : 전자해도, 수심정보, 군집화 기법, 밀집도, 품질

Abstract : ENCs are encoded using a numerical charts developed for publishing paper charts and serviced in forms of grid styles. For this reason, the density of ENCs' sounding information is not consistent and that requires improved methods. In this study, K-Means, ISODATA clustering algorithm as classification methods for satellite image was reviewed and adopted to this case study. The designed algorithm includes loading module for ENC data, improvement algorithm of sounding information, writing module of ENC data. According to the results of algorithm, we could confirm the improved result.

Key words : ENC, Sounding, Clustering Method, Density, Quality

1. 서 론

각국 수로국은 해상안전정보 인프라인 전자해도를 서비스 하고 있으며 이를 이용하여 ECDIS 뿐만 아니라 다양한 응용 시스템에 활용하고 있다. 우리나라는 2001년부터 전 연안의 전자해도를 간행 및 서비스 중이며, 품질 개선을 위해 노력하고 있다(KHOA, 2011). 전자해도 개발 초기에는 전자해도 데이터베이스(Database)가 없어 종이해도를 스캔(Scan)하여 벡터라이징(Vectorizing) 작업을 한 후, 속성 입력 작업을 거쳐 전자해도 데이터베이스를 확보하였다. 따라서 전자해도 셀(Cell) 영역은 종이해도와 정확히 일치하여 종이해도와 전자해도를 병행하여 사용하기에는 편리 하였으나, 인접 전자해도 셀 간 등심선이나 해안선이 일치하지 않아 낮은 품질의 주요 요인이 되었다. 전자해도 품질개선을 위해 격자형 간행체계를 도입하였는데, 격자형 간행체계는 축척에 따라 5단계로 구분하고 각 사용 단계별 대표 축척을 지정하여 하나의 격자로 편집하는 방식이다. 이에 따라 편집축척이 상이한 수치해도가

하나의 셀로 편집되어 정보 밀집도 비균일 현상이 발생되었다. 정보 밀집도 비균일 현상에 따라 가장 크게 부각되는 정보는 수심으로, 동일 셀에 수심 밀집도가 높은 부분과 낮은 부분이 동시에 표시되어 정보 표현 일관성이 부족하게 되었다. 전자해도 간행 절차상 종이해도 출력력을 위한 수치해도를 먼저 제작하고, 이를 변환하여 전자해도를 간행하기 때문에 제작 단계상 전처리 공정이 어려운 특징이 있어 이를 보완할 수 있는 방안이 요구되었다.

본 연구에서는 전자해도 수심정보 밀집도 개선을 위해 전자해도 데이터베이스 제작 완료시 후처리 과정으로 적용할 수 있는 수심정보 밀집도 조정기법을 연구하였다. 연구 방법으로 ISODATA(Iterative Self-organizing Data Analysis Technique Algorithm) 클러스터링(Clustering) 기반 알고리즘과 적용 프로그램을 개발 하였으며, 우리나라 전자해도에 적용하여 설정한 변수에 따른 클러스터링 계산 결과와 수심정보 밀집도 개선 결과를 확인 하였다.

† 연회원, osw@moeri.re.kr 042)866-3692

* 연회원, Moonjin.Lee@moeri.re.kr 042)866-3614

** 연회원, hjk@moeri.re.kr 042)866-3649

*** 종신회원, shsuh@moeri.re.kr 042)868-7264

2. 전자해도 및 수심정보

국제수리기구는 차세대 전자해도 표준으로 S-101을 개발 중에 있으며, ECDIS에서 동일 정보 중복 로딩 방지 및 효율적인 관리를 위해 축척 독립(Scale Independent), 축척 종속(Scale Dependent) 개념을 Fig. 1과 같이 도입 하였다. 이에 따라 S-100 표준기반의 다양한 데이터가 전자해도 백터데이터 상에 중첩하여 표현되므로, 정보의 축약 및 요약 기법이 중요할 것으로 판단된다.

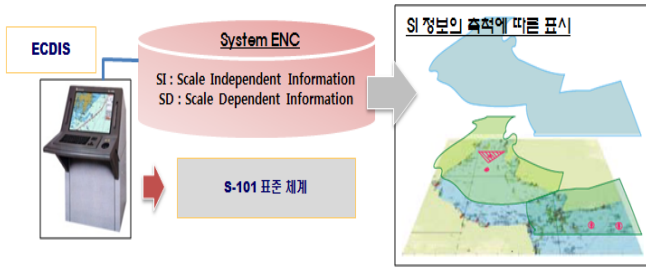


Fig. 1 Concept of SI/SD ENC

한편, 전자해도는 S-57 표준에 따라 항해안전에 관한 정보가 백터 데이터로 기록되며 부호화(Encapsulation) 방식은 ISO/IEC 8211을 따르고 있다. 전자해도 포함 정보 중 가장 중요하며 유일한 3차원 정보인 수심정보는 약어로 SOUNDG로 표기하며, 코드값이 129로 지정되어 있다. 주요 속성 정보로서 수심 위치특성을 나타내는 EXPSON, 수심 측정값 품질인 QUASOU, 수심정보의 시간적 특성인 STATUS로 포함된다. Fig 2는 전자해도 구조와 포함되는 수심의 속성정보와 공간정보를 나타낸 그림으로 속성정보는 속성코드와 속성값으로 구성되고, 공간정보는 3차원인 위도(Y축), 경도(X축), 수심(Z축)으로 구성된다.

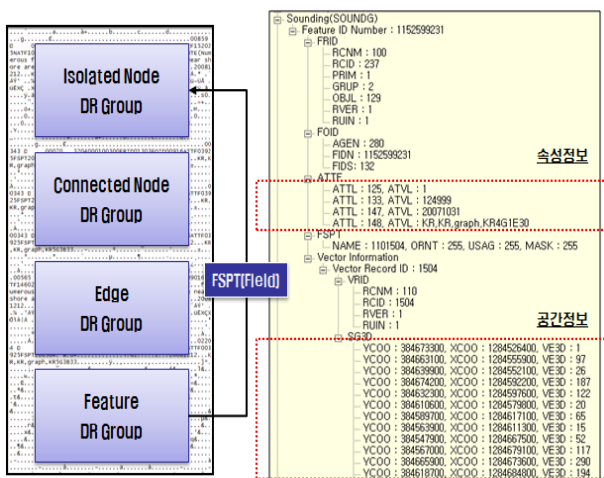


Fig. 2 Structure of Sounding information

전자해도 제작 체계는 Fig. 3과 같이 전자해도 제작 초기에는 종이해도와 동일한 격자로 간행되었고, 2007년에 격자형

전자해도 체계를 도입하여 각 격자에 해당되는 수치해도 정보를 편집 및 조정하여 전자해도를 간행하고 있다. 향후 전자해도 간행기관에서는 수로 데이터베이스를 도입하여 끊김 없이(Seamless) 연결된 간행체계를 도입할 계획이다.

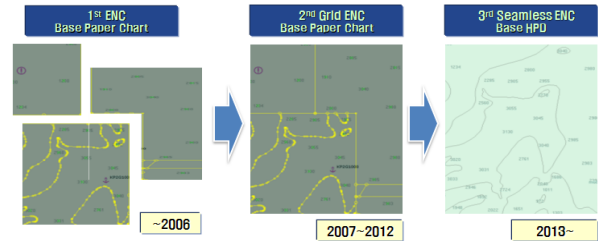


Fig. 3 Production system of Korean ENC

격자형 전자해도의 도입으로 동일 항해목적 내의 상이한 축척의 수치해도 데이터를 편집하여 Fig. 4와 같이 수심정보의 분포가 비균일한 것을 확인할 수 있다.

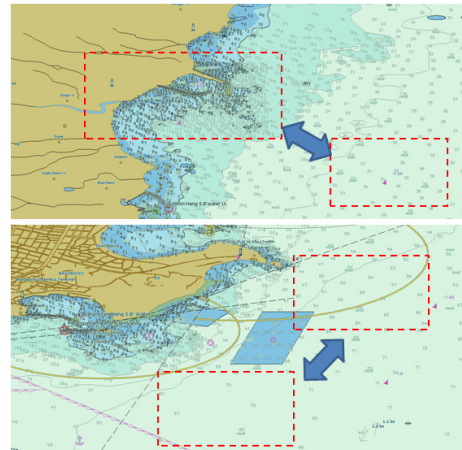


Fig. 4 Pattern of Sounding Information of ENC

전자해도 수심정보 분포의 비균일화는 소축척 전자해도 보다 대축척 전자해도에서 다수 조사 되었으며, 이를 개선할 수 있는 기법 도입이 필요한 것으로 판단된다.

3. 점 사상 일반화 기법

3.1 생략 기법

점 사상을 일반화하는 단순한 기법으로 생략기법이 있으며, Fig. 5와 같이 점 사상의 중요도에 따라 점 사상에 반비례하는 원을 그리고 순차적으로 생략해 가는 Settlement Spacing Ratio 기법과 점사상의 중요도에 따라 점 사상에 비례하는 원을 그리고 생략해 가는 Circle Growth 기법이 있다. 두 기법은 원의 반지름을 이용하여 점사상의 중요도에 따라 반지름을 줄여나가거나 확대해 나가는 기법이나 각 점사상의 상대적 중요도가 도출되어야 하고 점사상의 중요도에 따라 클러스터 그룹이 결정되는 특징이 있다(Li, 2007).

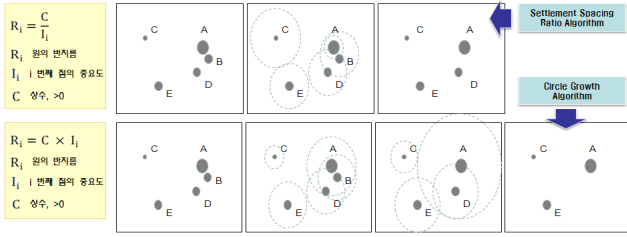


Fig. 5 Algorithms for selective omission of point features

3.2 K-Means 기법

K-Means 기법은 주어진 데이터를 k개의 클러스터로 묶는 기법으로, 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작한다. 본 기법은 간단한 구조를 가지며 일반 환경에 빠르게 수렴하기 때문에 널리 사용되고 있다(유, 1998). K-Means 기법 동작 원리로, 다음과 같이 2차원 공간에 점 이 분산되어 있다고 가정한다.

$$\{X_i, i = 1, 2, \dots, N\} \quad (1)$$

X_i : 2차원 공간에 분포된 점에 관한 변수

N : 점의 총 갯수

K-Means 기법의 경우 클러스터 개수 k 를 설정하고, 임의로 초기 클러스터 중심점을 구한다.

$$M_1^{(0)}, M_2^{(0)}, \dots, M_k^{(0)}, l = 0 \quad (2)$$

M : 클러스터의 중심점

k : 클러스터 개수, l : 클러스터 분할 및 통합 단계

각 샘플과 클러스터 중심 간 거리에 따라 각 샘플을 클러스터 중심으로 포함 시킨다.

$$X \sim w_j, \text{ if } D_L(X, M_j^{(l)}) = \min\{D_L(X, M_i^{(l)}), i = 1, \dots, k\} \quad (3)$$

여기서, w_j 는 중심점이 l 번째 단계에서 $M_j^{(l)}$ 인 샘플의 i 번째 클러스터를 말한다. 다음으로, 다시 클러스터 중심점을 $M_j^{(l+1)}$ 로 업데이트 한다.

$$M_j^{(l+1)} = \frac{1}{N_j} \sum_{X \sim w_j} X, (j=1, \dots, k) \quad (4)$$

$N_j^{(l)}$ 은 l 번째 단계에서 $w_j^{(l)}$ 그룹의 샘플 개수이다.

$\sum_{j=1}^k N_j^{(l)} = N$, w_j 그룹의 모든 점에서 신규 중심점까지 거리의 합을 최소화 하도록 반복한다.

$$\sum_{X \sim w_j^{(l)}} D_L(X, M_j^{(l+1)}) \rightarrow \min, (j = 1, \dots, k) \quad (5)$$

계산 값이 하나의 값으로 수렴할 경우 알고리즘을 종료한다.

K-Means 기법의 진행 절차는 Fig. 6과 같다.

$$M_j^{(l+1)} = M_j^{(l)}, (j = 1, \dots, k) \quad (6)$$

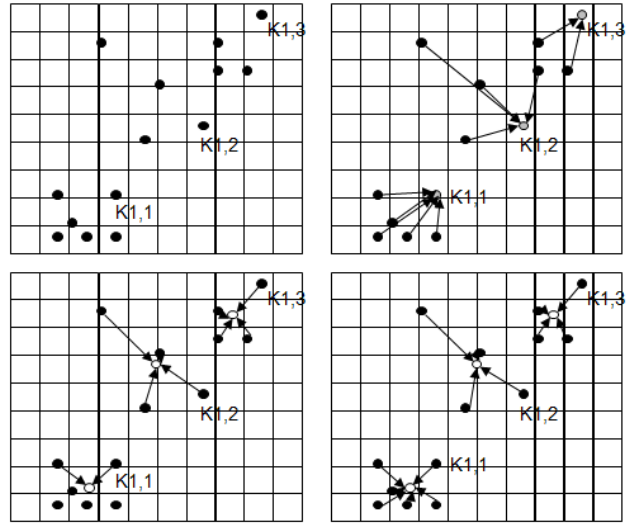


Fig. 6 Process of K-Means Algorithms

본 기법은 최초 설정한 그룹에 크게 의존하므로 초기 클러스터 설정에 따라서는 실제 최적 값보다 나쁜 값을 얻을 수 있다. 이를 방지하기 위해 서로 다른 초기값으로 여러 번 시도하여 가장 좋은 결과를 사용하는 기법 등을 사용할 수 있다. 또한, 구하려는 클러스터의 개수를 미리 설정해야 하며, 클러스터 개수가 너무 많으면 큰 클러스터가 여러 개로 나뉘는 결과를 얻을 수도 있다(곽, 1995).

3.2 ISODATA 기법

ISODATA 기법은 K-Means 기법을 기반으로 한 클러스터링 기법으로, K-Means 기법처럼 초기에 클러스터의 수 k 를 파라미터로 입력 받아서 실행되는 것이 아니라 클러스터의 크기에 따라 분할 및 병합하는 기법이다(Li, 2007). 본 기법에서 적절한 초기 파라미터를 설정하면 분할에 사용되는 클러스터 중심에서 각 축으로의 최대표준편차를 따라서 분할되고, 병합될 최대 클러스터 쌍의 최대 수에 따라 병합된다.

ISODATA 기법은 다음의 변수가 요구된다.

- K : 클러스터 개수, I : 알고리즘 반복 횟수
- P : 통합 시 허용되는 최대 요소 개수
- Θ_N : 클러스터 최소 요소 개수
- Θ_S : 분할을 위한 표준편차 지정값
- Θ_C : 통합을 위한 표준편차 지정값

데이터셋 $\{X_i, i = 1, 2, \dots, N\}$ 에서 임의의 클러스터 개수 k 와 M_1, M_2, \dots, M_k 를 지정한다. 각 N 샘플에 가장 근접한 중심점을 위치시킨다.

$$X \sim w_j, \text{ if } D_L(X, M_j) = \min\{D_L(X, M_i), i = 1, \dots, k\} \quad (7)$$

만약 클러스터에 포함되는 요소의 수가 θ_N 보다 작을 경우 if, $N_j < \theta_N$, 해당 클러스터 w_j 는 삭제하며 $k \leftarrow k-1$ 로 한다. 다음으로 각 클러스터의 중심점을 다음과 같이 조정한다.

$$M_j = \frac{1}{N_j} \sum_{X \sim w_j} X, \quad (j = 1, \dots, k) \quad (8)$$

각 클러스터 w_j 의 중심점과 각 요소 간의 평균 거리 D_j 를 계산한다.

$$\bar{D}_j = \frac{1}{N_j} \sum_{X \sim w_j} D_L(X, M_j), \quad (j = 1, \dots, k) \quad (9)$$

또한, 전체 거리의 평균값을 구한다.

$$\bar{D} = \frac{1}{N} \sum_{j=1}^k N_j D_j \quad (10)$$

만약 $k \leq K/2$ 이면 클러스터 수가 너무 적기 때문에 분할 작업을 반복하고, $k > 2K$ 이면 클러스터 수가 많기 때문에 통합작업을 반복하게 된다. 클러스터 분할과 관련하여, 다음과 같이 표준편차를 계산한다.

$$\sigma_j = \sqrt{\frac{1}{N_j} \sum_{X \sim w_j} (x_i - M_j)^2}, \quad (i = 1, \dots, n, j = 1, \dots, k) \quad (11)$$

여기서, M_j 는 클러스터의 중심점이고, σ_j 는 j 번째 클러스터의 표준편차이며, N_j 는 클러스터 w_j 의 요소 개수 이다.

각 클러스터 w_j 에서 가장 큰 표준편차 σ_j^{\max} 를 구한다.

만약 σ_j^{\max} , ($j = 1, \dots, k$)가 다음을 만족한다면,

$$- \sigma_j^{\max} > \theta_s, D_j > D, N_j > 2\theta_N \quad (12)$$

클러스터 중심점 M_j 에 해당 σ_j^{\max} 를 이용한 $\pm \delta$ 를 적용하여 M_j^+ 와 M_j^- 로 분할하며, δ 는 $\alpha \sigma_j^{\max}$ ($\alpha > 0$)로 계산한다. 이때, 클러스터 개수를 나타내는 k 는 $k \leftarrow k+1$ 로 조정한다.

알고리즘 종료 조건은 실험자가 설정한 반복 횟수에 도달하면 종료하도록 설정한다. ISODATA 기법의 진행 절차는 Fig. 7 과 같다.

4. 수심정보 클러스터링 기법 설계 및 적용 결과

4.1 수심정보 클러스터링 기법 설계

본 연구에서는 K-Means 기법을 기반으로 한 ISODATA 기법을 이용하여 전자해도 수심정보의 클러스터링 기법을 설계하였다. 알고리즘 수행 과정은 Fig. 8과 같이 전자해도 읽기 부분, ISODATA 적용부분, 전자해도 쓰기 부분으로 구성된 다.

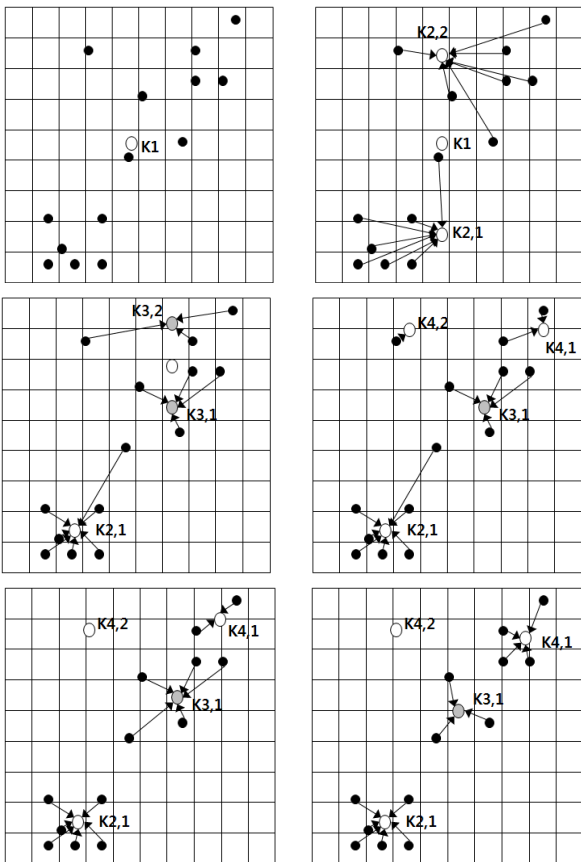


Fig. 7 Process of ISODATA algorithm

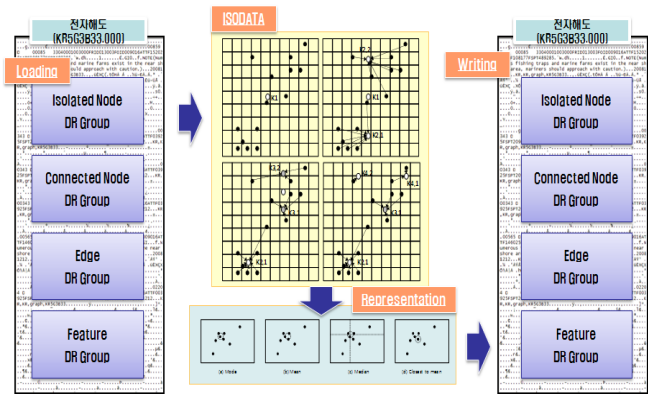


Fig. 8 Process of Sounding clustering algorithm

수심정보 클러스터링 기법의 세부 내용은 Fig. 9와 같다. 먼저 전자해도 로딩 부분으로, 전자해도 데이터에 몇 개의 수심 정보 레코드가 포함되어 있는지 확인하고 각 수심정보 레코드에 연결되어 있는 수심 벡터 레코드를 통합하여 하나의 수심 정보 클래스로 저장한다. 다음으로 초기 클러스터를 생성하고 X축, Y축 표준편차를 계산하여 기준 설정값과 비교한 후 X축으로 분할하지 Y축으로 분할할지를 결정한다. X축 혹은 Y축으로 클러스터를 분할하고 각 그룹의 중심점을 계산한다. 각 클러스터의 중심점을 기준으로 다시 표준편차 및 분할을 반복함으로써 최적의 클러스터 개수를 계산할 수 있다. 전자해도 저장 모듈에서는 각 클러스터의 대표점을 선택하고 각 대표점

만을 수심정보 레코드로 저장하며, 전자해도 로딩 모듈과 역순으로 전자해도를 저장 한다.

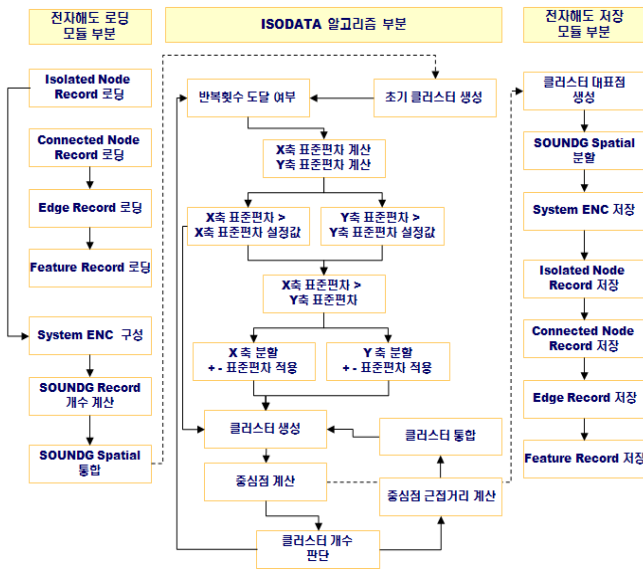


Fig. 9 Structure of Sounding clustering algorithm

각 클러스터의 대표점 추출 방법은 Fig. 10과 같은데, 먼저 주위 점에 비해 보다 높은 밀집도를 가지는 점을 선택하는 방법(Mode), X축과 Y축 기준 평균값을 취하는 방법(Mean), X축 기준 중심점과 Y축 기준 중심점의 교점을 취하는 방법(Median), 그리고 마지막으로 평균값(Mean)에서 가장 근접점을 취하는 방법(Closest to mena) 등이 있다. 본 연구에서는 기존 연구에서 가장 대표성이 높다고 분석된 평균값에 대한 근접점 방법을 채택하였다(Li, 2007).

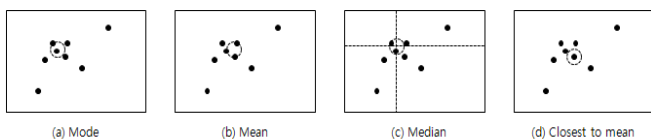


Fig. 10 The most representative point of a point class cluster

4.2 적용 프로그램 개발 및 결과 검토

본 연구에서는 설계한 수심정보 클러스터링 기법 적용을 위한 프로그램을 개발 하였다. 본 프로그램은 기법에서 설계한 전자해도 읽기 기능, 수심정보 클러스터링 기능, 전자해도 쓰기 기능이 포함되고 해독한 전자해도를 문자열로 입력된 텍스트 파일 저장 기능을 개발 하였다. 개발환경은 마이크로소프트 비주얼 스튜디오 2008을 이용하였으며 개발언어는 C#을 사용하였다. Fig. 11의 좌측 그림은 적용 프로그램이며, 전자해도에 포함된 각 피쳐정보를 종류별로 구분하고 각 피쳐정보의 속성과 벡터 정보를 표시하는 트리노드를 개발한 그림이다. 우측 그림은 원 전자해도 파일을 전자해도 뷰어 프로그램으로 확인한 그림이다.

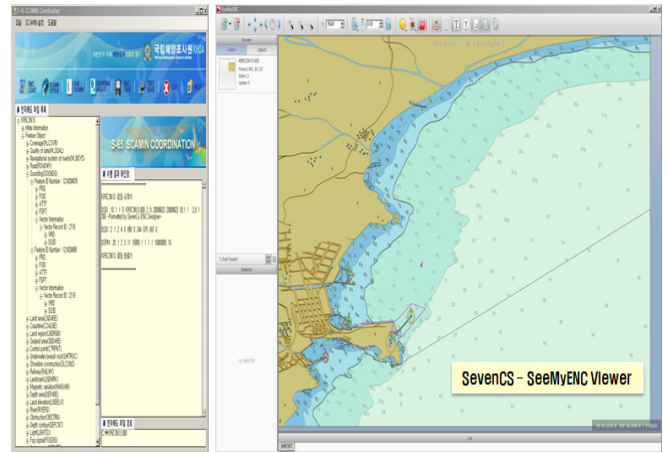


Fig. 11 Image of the original ENC file

전자해도 수심정보 조정 프로그램에서 수심정보 조정 기능을 실행하면 본 연구에서 설계한 기법을 통해 로딩한 전자해도 수심정보를 재조정하고, 저장기능을 실행하여 전자해도 수심정보 조정이 완료된다. Fig. 12의 우측 그림은 전자해도에 포함된 수심정보가 조정된 그림이다.

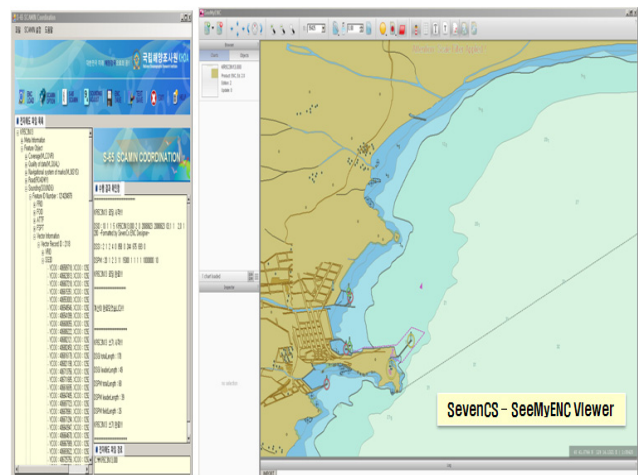


Fig. 12 Clustering result of the original ENC file

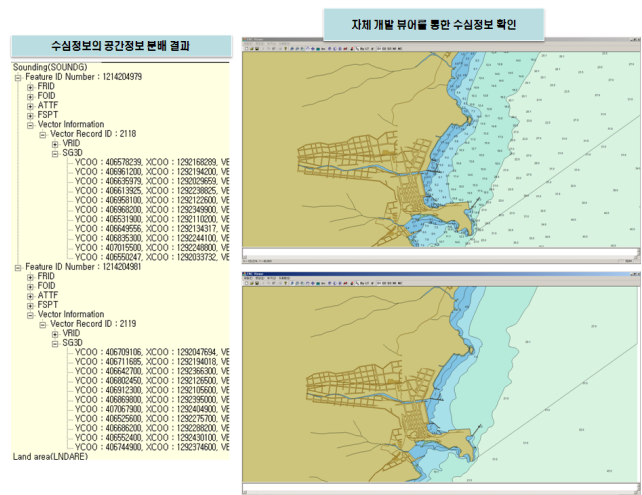


Fig. 13 Clustering result based on self developed

Fig. 11과 Fig. 12는 전자해도 뷰어 전문 프로그램으로 확인한 화면이며, Fig 13은 본 연구에서 자체 개발한 전자해도 뷰어를 통해 수심정보를 확인한 화면이다.

5. 결 론

전자해도는 제작 체계 특성 상 종이해도 간행을 위한 수치해도를 기반으로 제작되며, 격자형 전자해도 제작을 통해 수심정보의 밀집도가 비균일하게 되는 특징이 있다. 본 연구에서는 전자해도 수심정보 밀집도 개선을 위해 위성영상 분석에서 주로 사용되는 ISODATA 기법을 이용하여 수심정보 밀집도 개선 기법을 개발하였으며, 전자해도에 적용하여 그 결과를 검토 하였다. 수심정보 조정 기법을 비균일하게 분포되어 있는 전자해도 수심정보를 설정한 클러스터 수와 반복횟수에 따라 적정 수준의 클러스터 개수를 구하고 대표점을 도출하여 수심정보 밀집도 개선 결과를 확인할 수 있었다. 전자해도의 수심 분포는 통상적으로 해안선에 근접할수록 수심 밀집도가 높고, 해안선에서 멀어질수록 수심 밀집도가 낮은 고유의 특징이 있으므로, 해도에 적합한 수심 밀집도 조정 방법이 개발되어야 할 것으로 사료된다.

후 기

본 논문은 한국해양연구원에서 수행중인 “해양안전정보시스템 국제협력방안 연구(PMS2290)”과 “오염물질의 해양 유출사고 대응지원 기술개발(PG47643)”의 지원으로 수행 되었습니다.

참 고 문 헌

- [1] 광강울(1995), 소축척 지도제작을 위한 데이터 감축 기법에 관한 연구, 한국측지학회지, Vol. 13, No. 2, pp. 77~83.
- [2] 김감래(1992), 지도 일반화에 따른 단순화 알고리즘의 평가에 관한 연구, 한국측지학회지, Vol. 10, No. 2, pp. 63~71.
- [3] 김남신(2004), 규칙기반 모델링에 의한 하계망 일반화에 관한 연구, 대한지리학회지, Vol. 39, No. 4, pp 633~642.
- [4] 김감래(1996), 일반화 기법을 이용한 소축척 지도의 자동 생성 및 정확도 평가에 관한 연구, 한국측지학회지, Vol. 14, No. 1, pp. 27~38.
- [5] 박경식(2001), 수치지도 일반화 위치정확도 품질평가, 한국측량학회지, Vol. 19, No. 2, pp. 173~181.
- [6] 박우진(2009), 수치지도 작성을 위한 건물외곽선 단순화기법 연구, 한국측량학회지, Vol. 27, No. 1, pp. 657~666.
- [7] 유근배(1998), 점패턴분석을 이용한 수치지형도의 점사상 일반화, Journal of GIS Association of Korea, Vol. 6, No. 1, pp. 11~23.

- [8] 이제기(2000), 수치지도 일반화시스템 개발, 한국지형공간정보학회지, Vol. 8, No. 2, pp. 35~46.
- [9] 최병길(2001), 수치지도 일반화에 있어서 단순화에 관한 연구, 한국측량학회지, Vol. 19, No. 2, pp. 199~208.
- [10] 홍현기(1995), 디지털지도의 일반화 모델에 관한 연구, 서울산업대학교 산업대학원 논문집, Vol. 3, pp. 44~50
- [11] KHOA(2011), 국토해양부 국립해양조사원 홈페이지, www.khoa.go.kr
- [12] Zhilin Li(2007), Algorithmic Foundation of Multi-Scale Spatial Representation, CRC Press, pp. 75-80.

원고접수일 : 2011년 10월 26일
 심사완료일 : 2011년 12월 22일
 원고채택일 : 2011년 12월 22일