



## k-means 클러스터링을 이용한 강판의 부식 이미지 모니터링

김범수<sup>a</sup>, 권재성<sup>a</sup>, 최성웅<sup>a</sup>, 노정필<sup>b</sup>, 이경황<sup>c</sup>, 양정현<sup>a\*</sup>

<sup>a</sup>경상국립대학교 기계시스템공학과, <sup>b</sup>경상국립대학교 에너지기계공학과, <sup>c</sup>포스코 철강솔루션연구소

### Corrosion Image Monitoring of steel plate by using k-means clustering

Beonsoo Kim<sup>a</sup>, Jaesung Kwon<sup>a</sup>, Sungwoong Choi<sup>a</sup>, Jungpil Noh<sup>b</sup>,  
Kyunghwang Lee<sup>c</sup>, Jeonghyeon Yang<sup>a\*</sup>

<sup>a</sup>Department of Mechanical System Engineering, Gyeongsang National University, Tongyeong, Gyeongnam, 53064, Korea

<sup>b</sup>Department of Energy Mechanical Engineering, Gyeongsang National University, Tongyeong, Gyeongnam, 53064, Korea

<sup>c</sup>Steel Solution R&D Center, POSCO, Inchen, 21985, Korea

(Received 28 October, 2021 ; revised 29 October, 2021 ; accepted 30 October, 2021)

#### Abstract

Corrosion of steel plate is common phenomenon which results in the gradual destruction caused by a wide variety of environments. Corrosion monitoring is the tracking of the degradation progress for a long period of time. Corrosion on steel plate appears as a discoloration and any irregularities on the surface. In this study, we developed a quantitative evaluation method of the rust formed on steel plate by using k-means clustering from the corroded area in a given image. The k-means clustering for automated corrosion detection was based on the GrabCut segmentation and Gaussian mixture model(GMM). Image color of the corroded surface at cut-edge area was analyzed quantitatively based on HSV(Hue, Saturation, Value) color space.

*Keywords : Corrosion, GrabCut Segmentation, Gaussian Mixture Model, HSV color space, k-means clustering*

## 1. 서 론

탄소강은 산업 전반에 걸쳐 다양한 분야에 사용되고 있는 소재이다. 그러나 부식에 취약한 단점을 가지고 있으며, 부식으로 인한 소재의 열화는 성능저하 및 안전상의 큰 문제를 유발할 수 있어 산업 전반에 걸쳐 큰 위협 요소로 인식되고 있다[1,2]. 특히, 1998년도

에 진행된 미국 부식공업회(National Association of Corrosion Engineers, NACE)의 부식비용 연구에 따르면, 부식으로 인한 손실은 GDP의 3.1%에 달할 정도로 매우 큰 비용이 발생하는 것으로 보고 되었다. 이와 같은 부식을 방지하고자 내식성 소재의 개발 및 표면처리와 같은 방식 기술 개발이 진행되어왔다[2].

부식은 화학적, 전기화학적 반응, 마찰을 비롯한 다양한 요소들에 의해 점진적으로 진행되며 변색 및 부식생성물과 같은 이상 현상으로 소재를 열화시킨다. 특히, 아연도금강판의 경우 표면의 5% 산

\*Corresponding Author: : Jeonghyeon Yang  
Department of Mechanical System Engineering,  
Gyeongsang National University  
Tel: +82-55-772-9107 ; Fax : +82-55-772-9109 ;  
E-mail: jh.yagi@gnu.ac.kr

화철이 보일 때까지의 기간을 최초유지보수시간(Time to first maintenance, TFM)으로 하여 소재의 수명 유지를 위해 보수가 필요한 시간으로 기준을 정해 두었다. 이러한 소재의 부식 진행상태에 대한 적절한 모니터링과 추적 관찰은 효율적인 유지 관리 측면에서 중요한 부분이다.

부식 모니터링에는 여러 가지 기법들이 사용되고 있으며, 그중에서 전반적인 부식 진행상태에 대해 간단하게 평가를 진행할 수 있는 육안관찰 방법이 주로 사용되고 있다. 그러나 육안관찰의 경우, 검사자의 눈에 의한 시각적 판단에 의한 결과로써 객관적이고 정량적 평가를 기대하기는 힘든 실정이다 [2].

이러한 문제점으로 인해 최근에는 영상처리기술을 이용한 이미지 분석법을 적용하여 부식에 대한 평가를 진행하고 있다. 특히, 영상처리기술은 비파괴 분석 방법으로 다양한 분야에서 소재 표면 분석을 위해 사용되고 있으며, 빠르고 정확한 데이터 값을 제공할 수 있다[3-5]. 또한, 사물의 라인 감지와 같은 형태학적 구분, 색상 분석 및 이미지 분할 등이 가능하여 다양한 분야의 모니터링에 영상처리기술이 이용되고 있다[6].

이전 연구에서는 부식 시험편에 대해 GrabCut[7]을 이용한 이미지 분할과 HSV를 이용한 영상 이미지로부터 부식 감지에 초점을 맞춘 연구결과를 보고하였다.

본 연구에서는 POSCO에서 내식성 평가를 위해 진행한 두께 1.6mm, 도금량 180g/m<sup>2</sup>의 용융 아연도금(Galvanized steel, GI)강판의 절단면(cut-edge section) 부식 이미지를 이용하여 k-means 클러스터링(clustering)을 이용한 부식 모니터링을 진행하였다. 시험편은 서산 한국건설 생활 시험연구원 옥외 폭로 시험장에서 5년 동안 노출 시킨 이미지이며, 클러스터링 기법을 이용하여 1, 2, 3, 5년 절단면에서의 부식 진행 과정을 분석하였다. GI 강판의 부식 특성을 효율적으로 분석하기 위해서 GrabCut과 k-means 클러스터링을 이용해서 시험편을 여러 개의 영역으로 분할하고 각 영역에서 히스토그램, 색상 범위, 색상 평균과 표준편차 그리고 영역 비율을 구하였다. 입력 영상에서 사용자가 추출하고자 하는 객체 분리에는 영상 분할기법 중 하나인 GrabCut과 가우시안 혼합 분포(GMM)를 이용 하였으며, 영상처리에는 특정 색상에 대한 이미지처리가 뛰어난 HSV의 컬러 공간을 이용하였다. 일반적으로 부식은 그 형태가 매우 불규칙하고

다양한 계통의 색상이 혼재되어 있으므로 특정 색상이 분포된 영역을 검출하여 분석할 필요가 있다.

또한, 클러스터링은 주어진 데이터에서 유사한 데이터값을 나타내는 것들을 그룹으로 묶어 분석하는 매우 유용한 방법이다. 특히, 비지도 학습(Unsupervised Learning)기법 중 하나인 k-means 클러스터링은 데이터의 유사성과 특성에 따라 k개의 클러스터로 분할하여 분석하는 것이 가능한 방법이다[8,9].

## 2. 접근방법

### 2.1 영상 분할

GI 시험편의 절단면 영상에서 시험편의 전체 픽셀 수, 색상 등 시험편 특성을 정확하게 분석하기 위해서는 주어진 영상에서 시험편을 분리하였다. 영상에서 추출하고자 하는 객체(전경, foreground)를 배경으로부터 분리하는 방법이 많이 소개되고 있으나[6-8], 대부분 사용자와의 상호작용이 필요하다. 따라서 주어진 영상에서 배경과 객체 화소를 구분하여 지정하였다. GrabCut도 객체와 배경 영역을 화소의 부분 집합으로 할당하여 모든 영역의 최소값을 찾아 객체를 분리하는 알고리즘으로 사용자가 정확하게 객체를 추출하는 것이 가능한 방법이다.

시험편을 추출은 Fig. 1 (a)에 나타낸 것과 같이 주어진 영상 테두리에서 상하좌우로 일정한 간격(offset)을 갖는 사각형을 지정하였다. 간격은 5픽셀로 하였으며 배경(Background area)과 객체(Detection area)를 구분하였다. Fig. 1 (b)에는 GrabCut을 이용하여 입력 영상 데이터로부터 GMM을 이용하여 구축된 이미지를 나타내었다.

입력 영상 데이터 벡터  $z = [z_1, z_2, \dots, z_N]$ 로 부터 GMM을 이용하여 데이터를 구축하였다. (N : 전체 픽셀수,  $z_i$  :  $i$ 번째 픽셀의 컬러 값)

전경과 배경에 독립인 모든 GMM은 K 차원을 갖는 공분산(covariance)으로 볼 수 있으므로 GMM을 최적화하기 위해  $k = [k_1, k_2, \dots, k_N]$ ,  $k_i \in 1, 2, \dots, k$ 을 도입하였으며,  $k = 5$ 로 설정하였다. 각 픽셀에서 GMM을 구하기 위해 객체와 배경의 GMM 분포는 식 (1)과 같이 정의하였다.

$$\theta = \{ \pi(\alpha, k), \mu(\alpha, k), \Sigma(\alpha, k) \} \quad (1)$$

(  $\pi$  : 가중치(weight),  $\mu$  : 평균,  $\Sigma$  : 공분산)

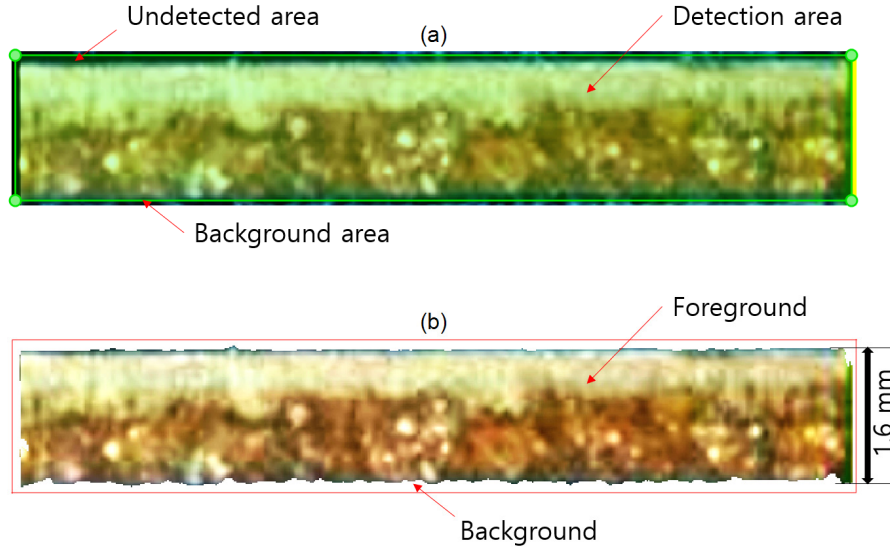


Fig. 1. (a) Original image with superimposed program input parameter(when offset is 5 pixels). (b) GrabCut image after image segmentation.

입력 영상  $z$ 를 분할 하는 문제는 에너지를 최소화하는 문제로 귀결된다. 최적화 과정을 거쳐 전경과 배경으로 분할된 영상은  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]$  벡터에 저장되며, 분할 벡터의  $i$  번째 픽셀  $\alpha_i \in \{0, 1\}$  은 바이너리 값으로 0은 배경으로, 1은 전경으로 정의하였다.

에너지 함수는 식 (2)와 같이 컬러 영상에 관련된 데이터 항  $U$ 와 구배(gradient)를 고려한 평활도(smoothness) 항  $V$ 로 구성하였다.

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z) \quad (2)$$

식(2)에서 GMM 데이터 모델은 다음과 같이 정의하였다.

$$U(\alpha, k, \theta, z) = \sum_{i=1}^N D(\alpha_i, k_i, \theta, z_i) \quad (3)$$

여기서

$$D(\alpha_i, k_i, \theta, z_i) = -\log p(z_i | \alpha_i, k_i, \theta) - \log \pi(\alpha_i, k_i) \quad (4)$$

이다. 그리고  $p(\cdot)$  는 가우시안 확률 분포 함수이며,  $\pi(\cdot)$ 는 혼합 가중 계수이다. 에너지 함수 (2)에서 평활도(smoothness) 항  $V(\alpha, z)$ 은 다음과 같이 컬러 공간에서 유클리드 놈(Norm)을 사용하여 정의하였다.

$$V(\alpha, z) = \gamma \sum_{(m,n) \in C} [\alpha_n \neq \alpha_m] e^{-\beta \|z_m - z_n\|^2} \quad (5)$$

여기서  $C$ 는 영상 내 모든 이웃 픽셀 쌍들의 집합이며,  $\gamma$ 는 윤곽선 가중치,  $\beta = (2E[z_m - z_n]^2)^{-1}$  이고  $E[\cdot]$ 은 기대값을 나타내었다.

따라서 에너지 함수 (2)를 최소화하는  $\alpha^*$  는 다음과 같이 전역 최소값을 찾는 문제로 귀결된다.

$$\alpha^* = \arg \min_{\alpha} E(\alpha, k, \theta, z) \quad (6)$$

여기서 구해진  $\alpha^*$  는 영상 분할을 의미한다.

프로그램에서 설정한 사각형을 기준으로 GrabCut 알고리즘은 주어진 입력 영상의 픽셀을 배경 영역과 객체 영역, 미지정 픽셀 영역이 혼재한 두 개의 그룹으로 구분한 후 다음과 같은 단계를 거쳐 객체를 분리하였다.

step 1. 미지정 픽셀 영역의 모든 화소  $n$  에 대해서

$$k_n = \arg \min_{k_n} D(\alpha_n, k_n, \theta, z_n)$$

을 구한다.

step 2. 영상 데이터  $z$ 으로부터 GMM 파라미터를 갱신한다.

$$\theta = \arg \min_{\theta} U(\alpha, k, \theta, z)$$

step 3. 식 (6)에서  $a^*$ 를 구한다.

step 4. 에너지 함수가 수렴할 때까지 step 1~3을 반복한다.

### 2-2. k-means 클러스터링

비지도(Unsupervised learning) 머신 러닝 알고리즘 중 데이터 분류에 사용되는 k-means 클러스터링은 구현이 비교적 간단하여 광범위하게 사용되고 있다[9,10]. k-means 클러스터링은 주어진 입력 데이터를 특성에 따라 k 개의 군집으로 분류하기 위해 중심점(centroid) k 개를 임의로 설정한 후 각 데이터로부터 유사성을 바탕으로 그 데이터가 속한 군집의 중심점까지 평균 거리를 최소화시키는 과정을 일정 오차 범위로 수렴할 때까지 반복한다. 이 과정을 유사코드(pseudo code)로 나타내면 다음과 같다[11].

step 0. 입력 데이터 및 군집수 설정

데이터 set :  $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}, x^{(i)} \in R^n$

목표 군집 수 : k

step 1. 초기화: k 개의 중심점(centroid)을 임의 선정

$$\mu_1, \mu_2, \dots, \mu_k \in R^n$$

step 2. 수렴할 때 까지 반복

for i=1 to m

$$c^{(i)} := \arg \min_j \|x^{(i)} - \mu_j\|^2$$

for j=1 to k

$$\mu_j := \frac{\sum_{i=1}^m r_{ij} x^{(i)}}{\sum_{i=1}^m r_{ij}}$$

여기서  $\|x^{(i)} - \mu_j\|$  는 데이터 포인트  $x^{(i)}$  와 클러스터 중심점  $\mu_j$  간의 유클리드 거리이고,  $c^{(i)}$ 는  $x^{(i)}$ 에 가장 가까운 클러스터 중심점의 인덱스, 그리고  $\mu_k$ 는 클러스터 k의 중심점이다. 그리고

$$r_{ij} = \begin{cases} 1 & \text{if } j = \arg \|x^{(i)} - \mu_j\| \\ 0 & \text{otherwise} \end{cases}$$

와 같이 정의된다.

### 2-3 분석 방법

일반적으로 컬러 영상은 Red(R), Green(G), Blue(B) 채널의 혼합으로 표현되나, 이 RGB 영상의 경우, 각 채널별로 처리할 때 색상의 왜곡 및 인간의 색상에 대한 직관성과 다소 거리가 있다. 반면 색상(Hue), 채도(Saturation), 밝기(Value)로 구성된 HSV의 컬러 공간은 특정 색상에 대한 처리가 상대적으로 뛰어나다.

HSV 모델에서 색상은 생상각(Hue angle)으로 표현하며, 범위는  $0^\circ \sim 360^\circ$ 이다. 채도는 0~255의 구간을 나타내며, 숫자가 클수록 색이 선명하다. 밝기 구간은 0~255이고, 0은 가장 어두운 상태를 나타내며 255는 가장 밝은 상태를 나타낸다.

Fig. 2에는 부식 분석 방법을 도시하였으며, 다음과 같은 순서를 분석하였다.

1. 히스토그램 평활화(Histogram Equalization)를 통해 입력 영상의 명암 대비를 개선
2. GrabCut을 통해 영상에서 시험편을 분할한다. 분할 후 배경은 모두 검은색으로 대치
3. 위의 RGB 영상을 HSV 색 공간으로 변환
4. HSV 공간 데이터에서 Hue 데이터를 기준으로 k-means 클러스터링(k = 5)
5. 각 클러스터링에 대한 histogram과 픽셀수를 계산

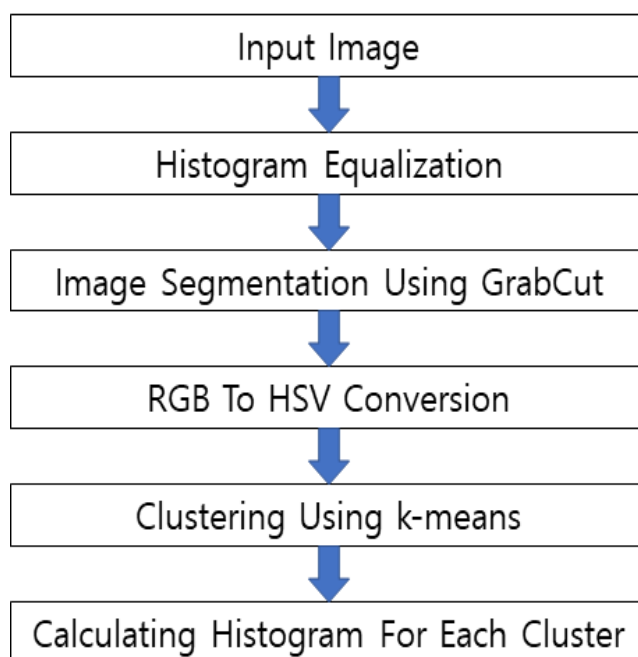
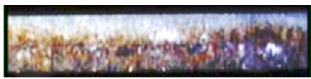
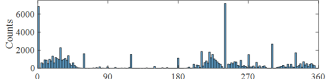
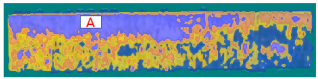

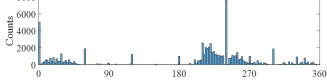
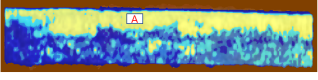

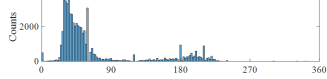
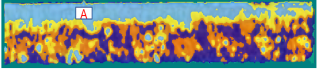

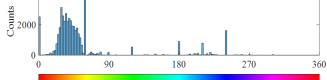
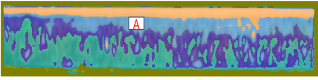


Fig. 2. Analysis flow diagram.

Table 1. GI materials with 1,2,3, and 5 years exposure. The first column shows section images, the second column is corresponding Hue histogram, and the third column shows the result of the k-means cluster, where k=5.

Exposure years	Image	Histogram	k-means clustering
1			
2			
3			
5			

### 3. 실험 및 분석

Table 1에는 GI 시험편을 1년, 2년, 3년, 5년 동안 노출한 후에 찍은 단면 영상 (a), 단면 영상을 GrabCut으로 전경을 분리하고 전경을 HSV 색 공간으로 변환한 Hue의 히스토그램(b), 그리고 전경을 k-means 클러스터링한 결과(c)를 나타낸다. 여기서 k는 5로 설정하였으며, 각 이미지의 테두리 부분은 배경에 속한다.

GI강판의 1년 경과 후 이미지에서는 히스토그램 Hue 값이 4~60(붉은색 및 짙은 노란색 계열), 199~255(짙은 청색 및 보라, 붉은색)에 주로 분포하고 있음을 보여준다. 2년 경과 후 시험편의 히스토그램에서는 Hue 값이 0~50, 193~289 에 주로 분포함을 보여주고, 3년 경과 후 시험편의 히스토그램은 Hue 값이 22~90, 152~222 사이에, 5년 경과 후 시험편의 히스토그램에서는 Hue 값이 10~60 사이

에 집중적으로 분포하는 것을 확인할 수 있었다.

Fig. 3 ~ 6에는 1년, 2년, 3년, 5년 동안 노출된 시험편을 클러스터 개수 5개로 하였을 때의 k-means 클러스터링 결과이다. Figure에서 좌측 열은 클러스터를 우측 열은 해당 히스토그램을 나타낸다. Fig. 3의 (a-2), Fig. 4의 (a-1), Fig. 5의 (a-2), Fig. 6의 (a-3)는 배경에 해당되는 클러스터이다. 배경은 검은색으로 치환하였기에 해당 히스토그램은 시험편 대비 Hue 픽셀 수 비율이 거의 0에 근접한 수치를 나타내었다.

Fig. 3의 (a-4), Fig. 4의 (a-3), Fig. 5의 (a-5), 그리고 Fig. 6의 (a-5)는 각각 Table 1 k-means 클러스터링에서 클러스터 A의 영역을 나타내었다. 이 영역들은 시험편 단면의 위쪽 부분으로 노출 시간에 따라 Fig. 3의 (b-4), Fig. 4의 (b-3), Fig. 5의 (b-5), 그리고 Fig. 6의 (b-5)에 도시된 히스토그램을 통해서 부식 색상의 분포 범위 및 비율을 구할 수

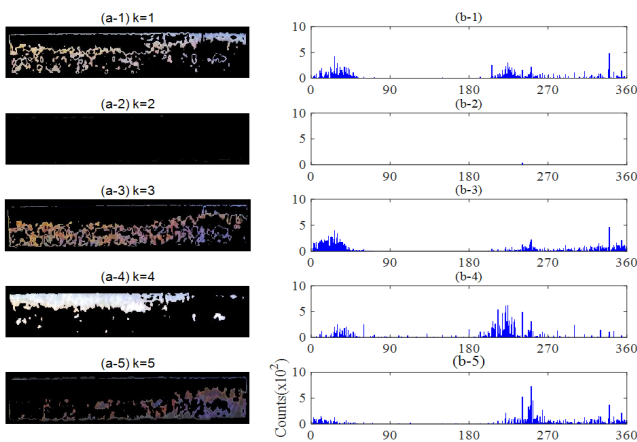


Fig. 3. GI material. 1 year exposure. The left column shows the clustering and the right column displays their corresponding histograms.

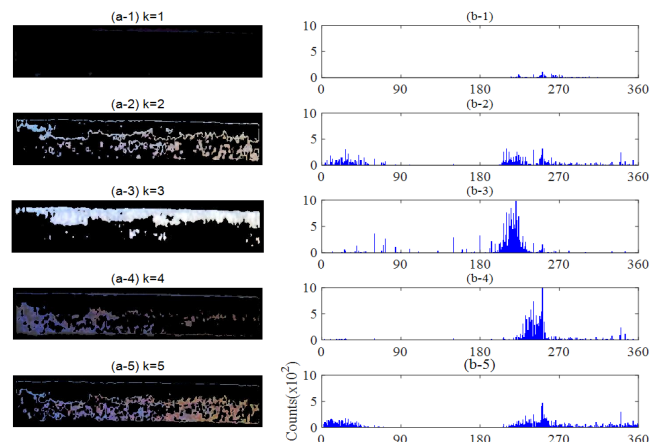


Fig. 4. GI material. 2 years exposure. The left column shows the clustering and the right column displays their corresponding histograms.

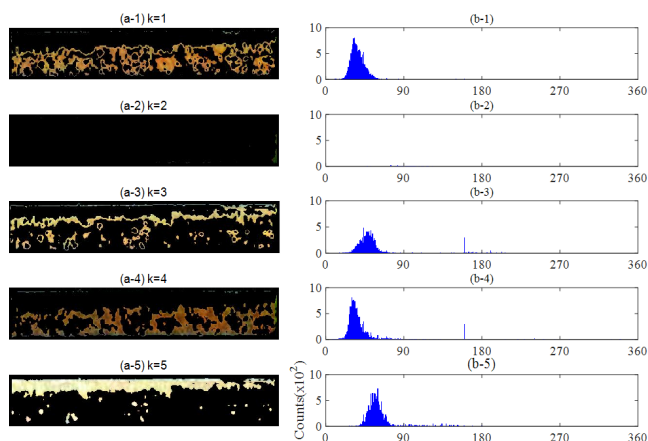


Fig. 5. GI material. 3 years exposure. The left column shows the clustering and the right column displays their corresponding histograms.

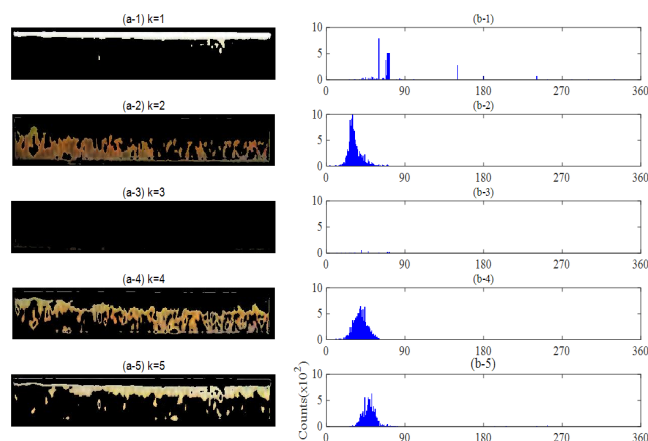


Fig. 6. GI material. 5 years exposure. The left column shows the clustering and the right column displays their corresponding histograms.

Table. 2. Hue range and ratio for cluster A.

Exposure years	Ratio (%)	Hue range	Mean	Std.
1	25.55	23~51, 193~263	186.3	86.8
2	29.59	193~253	200.2	49.7
3	25.62	40~75	62.6	21.4
5	23.19	35~60	50.3	16.9

있으며, Table 2와 같이 값을 표시하였다.

해안가에 1년 노출 시킨 시험편은 Hue 색상 분포가 크게 23~51과 193~263 에 걸쳐 분포되어 있다. 표준편차는 86.8로 상당히 큰 편이며, 2년 경과 후에는 Hue 색상 분포가 193~253에 집중되어 있지만, 표준편차는 49.7로서 푸른색 계열에 넓게 분포하고 있는 것을 확인할 수 있었다. Fig. 4의 (b-3)에서 색상이 하얀색 계열로 보이는 것은 밝기 값이 크기 때문이다. 3년 경과 후에는 Hue 색상 분포가 40~75 에 집중되어 있고 평균은 62.6, 표준편차는 21.4로써 노란색과 주황색 계열에 분포하고 있음을 확인할 수 있었다. 5년 노출 후에는 Hue 색상 분포가 35~60값에 집중되어 있고 평균은 50.3, 표준편차는 16.9로 주황색 계열에 집중적으로 분포하고 있음을 확인할 수 있었다.

#### 4. 결 론

본 연구에서는 GrabCut과 k-means 클러스터링을 이용해서 장기간 옥외 폭로시켜 부식된 시험편의 영상 분석 결과를 나타내었다. 영상 이미지로부터 시험편 표면의 다양한 색상들에 대해 여러 개의 클러스터로 분할 하는 것이 가능하였으며, 특정 영역에서의

부식에 의한 색상 변화를 HSV 공간 모델을 이용하여 모니터링 및 정량적인 수치로 표현하는 것이 가능한 것을 확인할 수 있었다.

#### References

- [1] Juhee Yun, Dooyoul Lee, Sungryul Park, Min-Saeng Kim, and Dongsu Choi, The Effect of Aircraft Parking Environment on Atmospheric Corrosion Severity, Corros. Sci. Tech., 20 (2021) 94-104.
- [2] Beomsoo Kim, Jaesung Kwon, Yeonwon Kim, Myeonghoon Lee, and Jeonghyeon Yang, Formation of Calcareous Deposit on Steel Plate by using Waste Oyster Shell, J. Korean Inst. Surf. Eng., 50 (2017) 531-535.
- [3] Beomsoo Kim, Yeonwon Kim, and Jeonghyeon Yang. Detection of corrosion on steel plate by using Image Segmentation Method. J. Korean Inst. Surf. Eng., 54 (2021) 84-89.

- [4] Bondada, Venkatasainath, Dilip Kumar Pratihari, and Cheruvu Siva Kumar. Detection and quantitative assessment of corrosion on pipelines through image analysis. *Procedia Comput. Sci.*, 133 (2018) 804-811.
- [5] Kuo-Wei Liao, and Yi-Ting Lee, Detection of rust defects on steel bridge coatings via digital image recognition, *Autom. Constr.*, 71 (2016) 294-306.
- [6] Hiba Ramadan, Chaymae Lachqar, and Hamid Tairi, A survey of recent interactive image segmentation methods. *Comput. Vis. Media.*, (2020) 1-30.
- [7] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, "GrabCut":interactive foreground extraction using iterated graph cuts. *ACM Tran. Graph.*, 23 (2004) 309-314.
- [8] Kristina P. Sinaga and Miin-Shen Yang, Unsupervised k-means clustering algorithm, *IEEE Access*, 8 (2020) 80716-80727.
- [9] Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam, The k-means Algorithm: A Comprehensive Survey and Performance Evaluation, *Electronics*, 9 (2020) 1295.
- [10] Xinjian Chen and Lingjiao Pan, A survey of graph cuts/graph search based medical image segmentation, *IEEE Rev. Biomed. Eng.*, 11 (2018) 112-124.
- [11] Chunhui Yuan and Haitao Yang, Research on K-Value Selection Method of k-means Clustering Algorithm, *J. Multidisciplinary Scientific Journal*, 2 (2019) 226-235.