

KNN 분류기에 의한 강판 표면 결함의 분류

김철호*, 최세호**, 주원종#, 김기범***

Classification of Surface Defect on Steel Strip by KNN Classifier

Cheol Ho Kim*, Se Ho Choi**, Won Jong Joo# and Gi Bum Kim***

ABSTRACT

This paper proposes a new steel strip surface inspection system. The system acquires bright and dark field images of defects by using a stroboscopic IR LED illuminator and area camera system and the defect images are preprocessed and segmented in real time for feature extraction. 4113 defect samples of hot rolled steel strip are used to develop KNN (k-Nearest Neighbor) classifier which classifies the defects into 8 different types. The developed KNN classifier demonstrates about 85% classifying performance which is considered very plausible result.

Key Words : Steel Strip(강 코일), Surface Inspection(표면 검사), k Nearest Neighbor(KNN, k 최근접 이웃), Synchronization(동기화), IR LED(적외선 조명), Feature Extraction(특징 추출)

1. 서론

한국은 포스코를 비롯한 경쟁력이 우수한 제철 제강 회사들이 있어 세계적인 철강 강국 대열에 속한다. 이 중에서 강 코일의 생산이 매우 중요한데 이는 강 코일의 품질과 가격이 주력산업인 자동차 산업이나 가전산업에 직접적으로 영향을 미치기 때문이다.¹ 강 코일을 생산하는 데는 여러 개의 공정을 거쳐야 하는데 대체적인 공정은 1)열연 공정(hot rolling), 2)담금공정(pickling), 3)냉연공정(cold rolling), 4)도금공정(galvanizing), 5)마무리공정(finishing) 등으로 나누어지고 여기에 열처리공정(annealing)도 포함될 수 있다. 강 코일(steel strip)은

표면의 품질에 따라 보통 5 등급(A, B, C, D, E)으로 나뉘는데 A 등급과 E 등급의 가격차는 통상 25~50%까지 이른다.

결함의 정도에 따라 판매되는 코일의 등급(grade)을 정확히 매기기 위해서는 결함의 종류, 개수, 크기, 위치 등이 정확히 정량적으로 산출되어야 한다. 한국에 있는 대부분 철강업체에서의 강 코일 표면 검사는 검사자의 시각에 의존하는 방법을 사용하여 왔다. 즉 검사자가 빠른 속도(약 150~800m/min)로 지나가는 인라인(in-line) 코일 상 하면을 비추는 거울 앞에 서서 두 거울을 번갈아가며 바라보면서 지나가는 결함들을 검사한 후 검사자의 감(feeling)에 의해 등급을 매긴다. 이와 같

☞ 접수일: 2006년 2월 23일, 게재승인일: 2006년 6월 9일

* 서울산업대학교 대학원 메카트로닉스 공학과

** 포스코 기술연구소 계측연구그룹

교신저자: 서울산업대학교 기계설계·자동화공학부

E-mail joo@snut.ac.kr Tel. (02) 970-6340

*** 서울산업대학교 기계설계·자동화공학부

은 시각에 의한 검사 방법은 부정확하여 통상 30% 내의 오판단율을 가지며 결함에 대한 아무런 정보 데이터를 남겨놓지 않아서 품질향상을 위한 피드백(feed-back) 자료나 판매를 위한 품질 데이터 자료로도 사용할 수 없다.¹

이와 같은 사정에도 불구하고 강 코일 표면 검사장비의 자동화가 잘 이루어지지 못한 이유는 다음과 같다. 첫째는 고속의 인라인 코일을 검사할 때 거의 실시간(real-time)으로 이루어져야 하는데 이를 위한 빠른 영상처리 및 영상이해(image processing and understanding) 소프트웨어 개발이 용의하지 않았다. 둘째는 대부분의 결함들이 매우 낮은 영상대비(contrast)를 가지고 있어 배경과 결함과의 구분이 불분명하여 이를 구분하는 매우 정교한 소프트웨어 기술개발이 어려우며, 셋째는 각 공정에 따라 결함의 종류가 10 여가지 이상이 되며 같은 종류의 결함이라도 코일에 따라 그 형태와 명암이 다양하고, 적지 않은 에러를 가지고 있는 특징 값으로부터 결함 종류를 분리하는 분류기(classifier)를 개발하는 것이 매우 어려웠다. 세계 유수 회사의 상업용 장비를 도입할 경우에도 장기간에 걸친 분류기의 친숙화 작업이 요구되며 많은 투자가 지속되어야 한다.^{1, 2} 또한 외국산 검사장비의 성능을 보면 결함검출은 90%이상이나 분류능력은 75~80%정도밖에 되지 않는 것이 투자를 하지 못하고 있는 또 다른 이유이다. 그러나 인건비와 품질 향상의 측면, 고객용 품질 데이터 확보 등, 그 외 다른 이점을 생각하면 설비투자를 심각히 생각하지 않을 수 없는 상황이다.

본 논문에서는 실시간 결함 검출 및 분류 시스템 구축을 위한 기반 요소기술 확보에 중점을 두었다. 최적의 결함 데이터를 분류기에 제공하기 위하여 하드웨어적으로는 최적 조명기술, 비전기술 및 동기화(synchronization)기술 등을 연구개발하였으며 소프트웨어는 고속 전처리 알고리즘, 특징 선정(feature selection) 알고리즘을 적용하였다. 결함종류 판독을 위해서 패턴인식에 많이 사용되는 KNN(k nearest neighbor)기법을 이용한 분류기를 개발하여 분류성능을 평가하였다.

2. 실험장치

Fig. 1 은 현재 상용화된 표면 결함검사 장비의 모습이며 주요기능은 고속 이송 중인 강 코일의

표면을 실시간으로 결함 이미지를 획득하고 고속 이미지 프로세싱 기법을 사용 노이즈 제거(filtering), 영역분할(area segmentation), 이진화(thresholding), 특징추출(feature extraction)을 수행한 후 결함 분류기에 의해 결함 종류를 판별을 하는 것이다. 또한 결함정보 데이터를 저장하여 오프라인으로 각 결함의 심각성에 따른 등급을 매기고 코일 전체의 통계자료를 산출하고 이에 따른 코일의 품질 등급을 매긴다.³

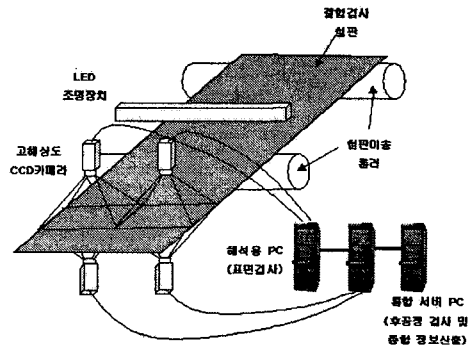


Fig. 1 Steel strip surface inspection system

본 연구에 사용되는 실험장치는 실제 검사 장비인 롤러 이송 장비와 같이 제작하기에는 너무 많은 공간과 고가의 장비들이 사용되므로 현실적으로 불가능하다. 앞서 언급한 단위요소기술을 확보하는데 적합하고 실험실에서 운영할 수 있는 회전 테이블 장비를 Fig. 2 와 같이 제작하였다.

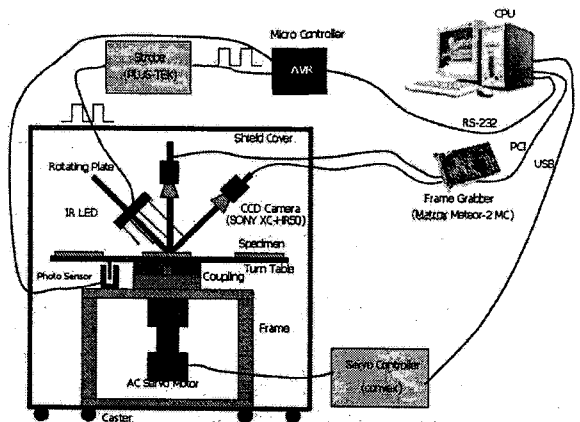


Fig. 2 Prototype surface inspection system

실험장치는 크게 기구부, 제어부, 광학부, 해석 및 데이터 처리의 4 단계로 구성되어 있다.

2.1 기구부

머신 비전(machine vision) 시스템에서 조명과 카메라를 사용하여 영상 이미지를 획득함에 있어 우선 외부 빛에 의한 노이즈를 차단하는 것이 아주 중요하므로 실험 장치는 Fig. 2 와 같이 이동식 차폐막 구조로 하였다. 결합이 있는 시험철판의 크기가 200x200mm 이므로 회전테이블에 6 개의 철판이 등간격으로 위치할 수 있도록 고려하여 회전 테이블($\phi 800\text{mm}$)을 제작하였다. 회전 테이블은 알루미늄 프레임에 고정된 AC 서보 모터에 의하여 최대 200rpm(340 m/min)의 속도로 회전한다.

2.2 제어부

제어부는 크게 모터 제어부와 신호 제어부로 나뉜다. 회전 테이블은 AC 서보 모터에 의해 정속도로 항상 일정한 속도를 유지해야 하므로 전용 서보 모터 제어기(Convex 社)로 제어하며 결합 이미지를 획득하기 위한 2 대 카메라는 스텝(step) 모터에 의해 각도를 조절할 수 있다. 각 하드웨어들의 신호 제어 즉 포토센서의 회전 검출 신호, 조명 제어장치인 스트로브스코프(stroboscope) 출력 신호 제어들은 ATmel 사의 AVR 8bit 마이크로 컨트롤러를 사용하였고 RS-232 및 USB 통신 방식으로 상위 제어기인 컴퓨터에서 VC++6.0 MFC 로 제작된 사용자 프로그램을 이용하여 모든 하드웨어들을 제어할 수 있다.

2.3 광학부

머신 비전 시스템에서 광학계의 구성은 시스템의 성능을 결정짓는 중요한 부분이다.^{3, 5, 7} 실험장치에 사용된 조명은 수명이 길고 노이즈의 영향을 적게 발생하는 장파장의 적외선 조명(infrared led)과 조명을 단속적으로 제어할 수 있는 장치인 스트로브스코프를 사용하였다. 영상신호를 검출하는 카메라는 반사광에 잘 잡히는 결합과 산란광에 잘 잡히는 결합으로 구분하여 두 가지의 카메라 검사위치를 정하였다. 조명의 정반사 영역에 카메라를 설치하는 명영역(bright field) 검사와 조명의 산란영역에 카메라를 설치하는 암영역(dark field) 검사의 구조로 하였다.³ 카메라 타입은 진동에 덜 민감한 area 타입으로 최대 전송률이 120fps, 해상도

640x480 인 카메라를 선정하였다.³ 실험장치의 광학부의 구성은 Fig 3 에 보여진다.

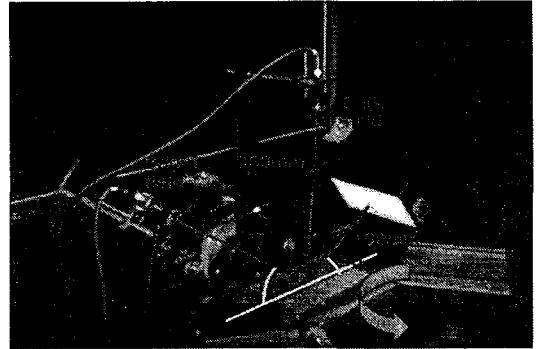


Fig. 3 Optical part of experimental equipment

2.4 해석 및 데이터 처리

결합 이미지의 해석과 그 결과 데이터를 처리하기 위한 시스템의 구성은 강 코일의 폭이 넓은 관계로 결합 이미지의 수집과 특징 추출 및 판독을 하는 여러 대의 카메라와 컴퓨터가 커플링으로 구성되어 있으며 각 컴퓨터로 나오는 결과를 하나로 묶어 전체 강 코일의 결합정보와 위치, 빈도수에 따라 등급을 관리하는 통합 서보 컴퓨터 시스템으로 나누어 구축되어야 한다. 본 연구에서는 통합 서보 컴퓨터 시스템은 제외되어 있으며 해석용 컴퓨터의 고속 이미지 프로세싱과 결합검출, 특징추출 등의 알고리즘이 구현되어 있다. 본 논문에서 다루는 영상획득에서 결합분류까지의 소프트웨어 순서도는 Fig. 4 에 보여진다.

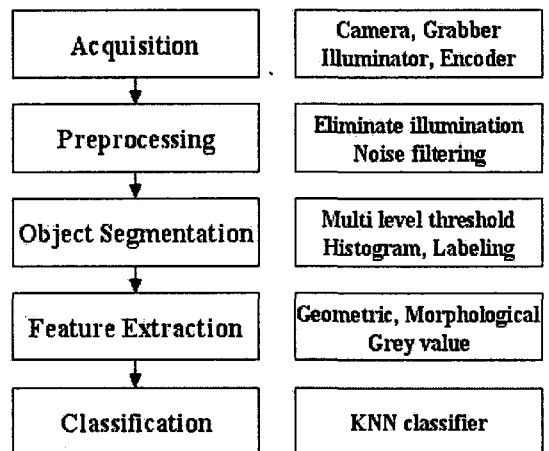


Fig. 4 Flow chart of software procedure

3. 영상처리

3.1 이미지 획득

고속 이동하는 물체표면의 이미지를 밀리지 않고 정지한 듯한 영상을 얻기 위해서는 정확한 타이밍 플랜(plan)에 따라 카메라 노출과 조명을 터뜨려 이미지를 획득해야 한다. 고속으로 이송하는 물체에 연속적인 조명을 사용하여 영상을 획득하면 뚜렷하지 못하고 흐릿한 영상을 얻게 된다. 정지한 듯한 영상을 획득하기 위해서는 카메라 노출시간(exposure time) 동안에 스트로브스코프를 사용하여 조명을 터뜨려 영상을 획득해야 한다. 고속으로 회전하는 철판 표면의 영상을 짧은 노출시간에 조명 신호를 정확하게 일치시키는 과정인 동기화 작업이 필요하다. Fig. 5 는 여러 번의 실험을 통해 결정된 고속에서 최적의 이미지를 획득하기 위한 소프트웨어와 하드웨어 출력신호들 사이의 동기화 타이밍 차트(timing chart)이다.

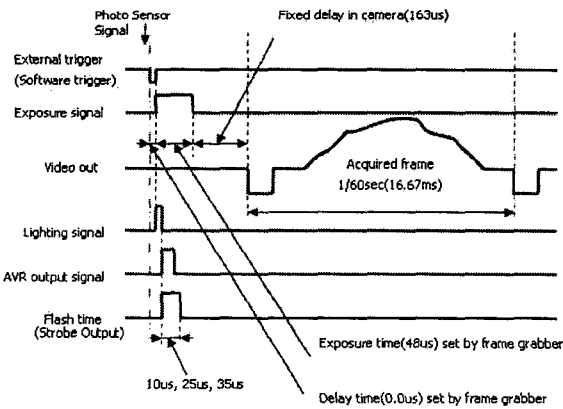


Fig. 5 Synchronization timing chart

카메라의 노출시간은 48us 이고, 조명의 플래쉬 시간은 10~35us 로 결정되었다. 시험장비 턴테이블의 회전속도가 200rpm 이라면 1 회전당 300ms 가 걸리고 6 장의 시험철판의 이미지를 획득해야 하므로 50ms 마다 위의 신호를 한번씩 출력하여 시험철판 표면의 이미지를 획득한다.

3.2 이미지 전처리

카메라와 조명을 사용하여 철판 표면의 영상을 획득하면 영상은 각종 잡음과 조명의 영향을 포함하고 있다 Fig. 6(a)의 조명을 포함한 이미지의 이

진화는 Fig 6(b)에서처럼 조명의 영향으로 인하여 결함을 추출하는 것이 불가능하다. 이러한 잡음과 조명의 영향이 존재하는 영상으로부터 결함영역을 추출하기 위해서는 조명과 잡음을 제거하고 흠을 배경과 더욱 명확하게 분리하여 이미지를 향상시키는 이미지 전처리(image preprocessing) 과정이 필요하다.^{2, 4, 5} 우선 조명의 영향을 제거하는 기존의 방법들은 결함 이미지로부터 조명의 프로파일을 추출하여 원 이미지와 차연산을 하여 조명 영향이 제거된 이미지를 얻는다. 이런 과정들은 어렵고 복잡한 계산과정으로 인해 많은 시간을 소비하므로 실시간 프로세싱에는 적합하지 않다. 본 연구에서 제안된 방법은 결함이 없는 시험철판의 영상인 Fig. 7(a)를 기준이미지로 하여 결함이 있는 영상과 차연산을 통하여 조명영향을 제거하였다.

$$I(x, y) = |I_{ref}(x, y) - I_{orig}(x, y)| \times \alpha \quad (1)$$

$I(x, y)$ 는 조명이 제거된 이미지이고, $I_{ref}(x, y)$ 는 결함이 없는 시험 철판으로부터 얻는 조명을 포함한 기준 이미지, $I_{orig}(x, y)$ 는 현재 결함이 있는 시험 철판의 이미지이고, α 는 대비계수로 결함과 배경의 차이를 더욱 증가시키는 역할을 한다. 위의 방법으로 Fig. 7(b)의 조명영향이 제거된 이미지를 얻었다.



Fig. 6 Image with illumination and its binary image

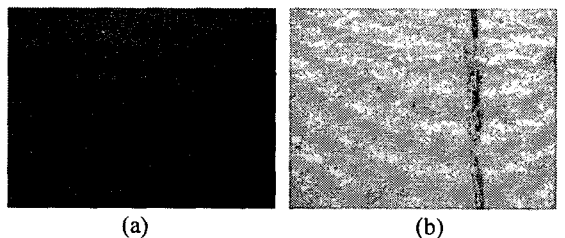


Fig. 7 Reference image without defect and difference image after subtraction

조명이 제거된 이미지를 얻은 다음 노이즈의 형태에 따라 적절한 필터를 사용하여 노이즈를 제거하거나 이미지의 상태에 따라 히스토그램 평활화, 명세화, 명암대비 스트레칭 등의 방법을 사용하여 이미지를 향상시키는 작업이 필요하다.

3.3 결함 분리

결함 분리(object segmentation) 과정은 전처리가 끝난 이미지에서 결함과 배경을 분리시켜 이진화하고 유사한 결함별로 분리(segmentation)하여 관심영역(region of interest: ROI)을 설정하는 과정이다.^{3,7} 이 과정에서 가장 중요한 이진화를 할 때 임계값(threshold value)를 설정하는 데에는 여러 가지 다양한 방법들이 있는데 본 연구에서는 히스토그램 분석법을 사용하여 임계값을 설정하였다.⁷ Fig. 8(a)는 Fig. 7(b) 이미지의 히스토그램이고 Fig. 8(b)는 이진화된 이미지이다. 이진화된 이미지에서 결함들을 라벨링(labeling)하여 각 결함별로 분리하여 ROI를 설정한다.^{3,4,5,7}

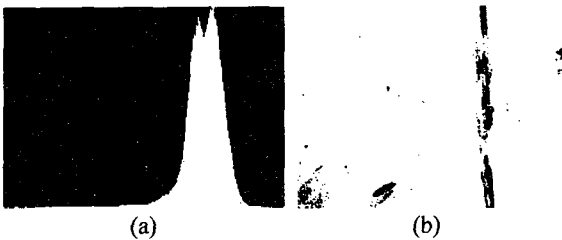


Fig. 8 Histogram and thresholding via histogram

3.4 특징 추출

전처리 및 결함 분리가 끝나면 ROI 영역에 있는 결함에서 특징 값을 추출한다. 일반적으로 영상 처리에 있어서 특징 추출은 어렵고 시간이 오래 걸리는 단계에 속하므로 실시간으로 정확하게 결함을 분류하기 위해서는 적절한 특징데이터와 개수의 선정에 많은 연구와 시간이 필요하다.² 실험에서 사용되는 특징 데이터는 크게 아래의 두 가지로 구분할 수 있다.

1) 형상 특징(shape features)

대상체의 윤곽선과 그를 둘러싸는 Box를 기초로 한 물체의 특징들이다. 이진화된 이미지에서 추출할 수 있는 특징 데이터들이며 종류는 결함의 면적(area), 면적 중심(center of area), 결함의 둘레(perimeter), 원형도(compactness), 결함과 배경의 면

적비(area ratio), 1, 2차 모멘트(moments), 주 2차 모멘트(principal 2nd moments), 2차 모멘트 최대, 최소 각(orientation) 등의 정보들이 이에 속한다.^{2,7}

2) 명암도/히스토그램 특징(intensity/histogram)

대상체와 배경의 명암 값을 기초로하는 특징 값들이며, 결함영역의 명암 값과 히스토그램을 사용하여 추출할 수 있는 특징 데이터들이다. 종류는 평균(mean), 표준편차(standard deviation), 왜도(skew), 굽힘 에너지(bending energy), 엔트로피(entropy), 대비(contras), 첨도(kurtosis) 등이 있다.^{2,7}

4. 결함 분류기

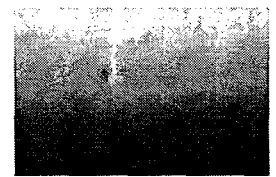
4.1 결함 선정

본 논문에서는 분류기의 성능평가를 위하여 열연공정에서 발생하는 대표적인 결함 8 개를 선정하였고 그 결함들의 종류와 발생원인은 아래와 같으며 대표적인 샘플이미지는 Fig. 9에 보여진다.

- 1) **Hole**: 슬랩(stab)의 심한 내부 크랙, 이물 치입
- 2) **Roll mark**: 압연 및 처리 라인의 각종 Roll에 이물 부착, 롤 타흙 발생
- 3) **Scab**: 슬랩 손질 불완전 및 슬랩 핸들링 불량
- 4) **Blowhole**: 표면 가까이 있는 기포의 미압착 및 대형 개재물에 기인
- 5) **Rolled in scale**: 고온 조건에서 압연될 경우 발생 스케일 제거 불량이나 혼입된 경우 생성
- 6) **Red scale**: 강 중에 실리콘(Si)이 강 표면으로 고온 확산되어 페라이트계 산화물을 형성하여 제거되지 않고 냉각수와 응결되어 산화됨으로서 생성
- 7) **Edge crack**: 강 코일 측면에 크랙이 압연 중 에지부로 전진, 고 탄소강의 압연 및 냉각 시 에지부 온도저하로 저온변태 조직인 베이나이트 생성에 따라 발생
- 8) **Pseudo**: 결함이 아닌데 결함으로 검출되는 흠



Hole



Roll mark

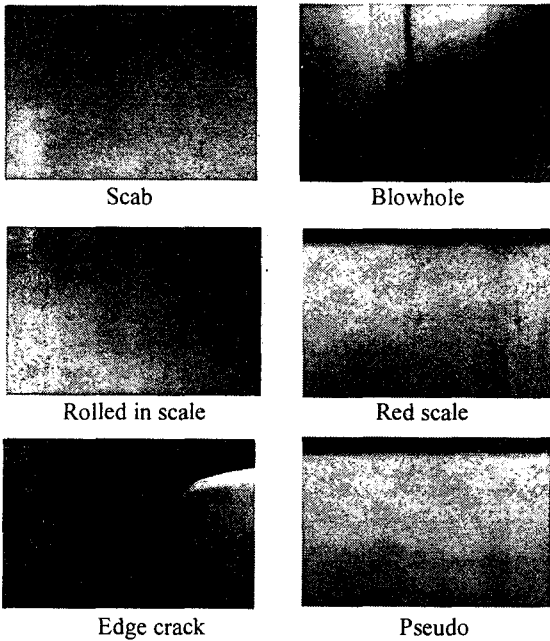


Fig. 9 Eight defect types in hot rolled steel strip

4.2 특징 데이터 선정

포스코에서 실제 검사 공정에서 사용하는 특징 데이터의 종류는 거의 400 여개 이고 필요에 따라서 추가를 하고 있다. 물론 모든 특징 데이터를 사용하는 것이 아니고 해당 공정의 특성에 따라 적절한 특징 데이터와 개수를 선정하여 사용한다.

본 논문의 결함 분류기의 성능을 평가하기 위해 포스코로부터 받은 샘플 결함의 특징 데이터의 종류는 378 개이다. 그러나 유한한 샘플을 사용하는 홈 분류기를 설계할 경우 초기에는 특징 데이터의 개수가 증가함에 따라서 분류기의 오분류가 감소하다가 어느 점에서 정점에 이르고 이후 데이터 개수가 더 증가할수록 오분류가 다시 증가한다 (peaking 현상). 또한 특징 데이터 수가 증가함에 따라 요구되는 결함 샘플 수가 기하급수적으로 증가하기 때문에 적절한 특징 데이터 개수의 선택이 매우 중요하다. 본 연구에서는 특징 데이터의 분석을 다음과 같이 하였다. 첫 번째는 Outlier 를 이용하여 홈에 이상치를 가지는 특징 데이터를 제거한다. 두 번째는 Anova(analysis of variation) 테스트로 홈 별로 구분할 가능성이 적은 특징 데이터를 제거하고 세 번째는 홈 별로 상관관계(correlation)가

높은 특징 데이터들, 즉 유사한 값을 가지는 특징 데이터들을 제거하였다. 총 230 개가 제거되고 최종적으로 특징 데이터는 148 개가 선정되었다.

SFFS(sequential forward feature selection) 알고리즘은 상향식 기법(bottom up method)으로 작은 특징 값들에서 분류능력이 우수한 특징 값들을 점진적으로 추가하는 특징 데이터 선정기법으로 분류율에 큰 영향을 미치는 특징 데이터들을 선정하는데 다소 많은 시간을 요구하지만 분류율의 성능을 저하시키지 않는 최적의 특징 데이터를 선정할 수 있다. SFFS 알고리즘에 의해 선정된 특징 데이터들의 분류율 성능 검증을 위하여 평가 함수로 KNN 분류기를 사용하였다. SFFS 알고리즘을 사용하여 위에서 선정된 148 개의 데이터로부터 다시 한번 최적의 37 개의 특징 데이터를 선정하였다.

4.3 k Nearest Neighbor(KNN)

k 최근접 이웃(KNN)은 기억 기반 추론기법 중에 하나로 패턴인식에 많이 활용되고 있다.^{6,8} k 최근접 이웃방법은 입력패턴이 주어졌을 때 유사도 함수에 기반하여 미리 저장 되어있는 표준패턴들의 집합에서 입력패턴과 가장 유사한 k 개의 표준패턴들을 선택한다. 선택된 k 개의 표준패턴들의 클래스들을 확인하여 가장 많이 선택된 표준패턴의 클래스를 입력패턴의 클래스로 결정한다.

입력패턴과 표준패턴과의 유사도는 유클리디언 거리(euclidian distance)를 사용하였다. 만약 입력패턴과 표준패턴이 아래와 같이 주어 진다면

$$\begin{aligned} X &= [x_1, x_2, \dots, x_n] \\ Y &= [y_1, y_2, \dots, y_n] \end{aligned} \quad (2)$$

여기서 X는 새로운 입력결함이고 x_1, x_2, \dots, x_n 은 X 결함의 특징 값들이다. Y는 표준패턴 결함에 속한 결함이고, y_1, y_2, \dots, y_n 은 Y 결함의 특징 값을 의미한다. 두 결함 사이의 유클리디언 거리는 다음 (3)의 식으로 구한다.

$$d_e(X, Y) = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

여기서 $d_e(X, Y)$ 는 입력결함과 표준결함 사이의 유클리디언 거리, N은 특징 값의 전체 개수이다.

4.4 KNN 분류 프로시저

1. k 값을 선정한다.
2. 새로운 결함과 표준결함들의 특징 값들을 이용하여 유클리디언 거리를 계산한다.
3. 계산된 거리 값들 중에서 가장 가까운 거리 값의 표준결함을 k 개 만큼 선정한다.
4. k 개만큼 선정된 거리 값들 중 가장 많이 선택된 결함의 타입을 입력 결함의 타입으로 선정한다.

5. 결함 분류 실험

현재 학교 실험실에서는 많은 결함 샘플 이미지를 얻을 수 없는 관계로 포스코로부터 강 코일의 품질에 크게 영향을 미치는 대표적인 8 개 결함들의 샘플 이미지 데이터 4113 개를 제공 받았다. 각 결함 별 샘플 수는 괄호 안의 수치로 다음과 같다. Hole(56), Roll mark(482), Scab(83), Blow hole(492), Rolled in scale(920), Red scale(1257), Edge crack(7), Pseudo(816) 이 샘플 데이터를 가지고 KNN 분류기로 결함의 종류를 판별하였다.

5.1 샘플 수 조정

각 결함 별 샘플 수가 일정하지 않으면 분류기에 많은 샘플을 가진 결함의 영향력이 커지므로 이를 방지하기 위하여 통계적인 데이터 조작을 통해 결함 별 샘플 데이터의 수를 대략 500 개로 조정하였다. 결함 샘플 수가 적은 Hole, Scab, Edge crack 은 노이즈 데이터를 첨가하여 결함 수를 증가하였고 결함 수가 많은 Rolled in scale, Red scale, Pseudo 는 랜덤하게 선정하여 필요한 결함 수를 조절하였다. 이렇게 조절된 결함들은 Table 1 에 보듯이 학습용(training set)과 테스트용(test set)으로 구분하였다. 각 결함은 특징 값들을 포함하고 있다.

Table 1 Regulated sample data

Class	Defect name	Samples	Modified	Training/Test
1	Hole	56	504	254 / 250
2	Roll mark	482	482	237 / 245
3	Scab	83	498	241 / 257
4	Blowhole	492	492	247 / 245
5	Rolled in scale	920	500	245 / 255
6	Red scale	1257	500	260 / 240
7	Edge crack	7	504	259 / 245
8	Pseudo	816	500	247 / 253
		4113	3980	1990 / 1990

5.2 데이터 정규화

샘플 결함들은 타입별로 특징 값들이 큰 차이를 보인다. 특징 값들의 차이가 너무 크면 작은 값을 가진 특징 값들은 분류기에 아무런 성능을 미치지 못하고 상대적으로 큰 특징 값들이 분류기에 큰 영향을 미쳐 분류기의 성능이 많이 떨어지게 된다. 이런 특성을 방지하기 위하여 특징 값들의 분포는 그대로 유지하면서 일정 범위 안으로 특징 값들의 크기를 조정하여야 한다. 즉 데이터를 정규화(normalization) 하는 과정이 필요하다. 표준편차를 이용하여 특징 값들을 정규화하였다. 표준편차 S 는 (4)의 식으로 구한다.

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^n (m - f_i)^2} \quad (4)$$

m 은 특징 데이터들의 평균 값이고 f_i 은 원 특징 데이터(f₁ ~ f_n), N 은 특징 값의 전체 개수이다. (4)의 식으로 구한 표준편차 값을 이용하여 각 특징 데이터들을 (5)식을 이용 정규화하였다.

$$f'_i = (m - f_i) / s \quad (5)$$

여기서 f'_i ~ f'_n 은 정규화된 특징 데이터들이다.

5.3 분류 실험 결과

4113 개의 샘플 데이터에서 결함별 샘플 데이터 수를 조절하여 최종 3980 개의 결함 데이터를 구성하였다. 이를 훈련용(training set) 1990 개, 성능 평가용(test set) 1990 개로 반반씩 나누어 분류율 실험을 하였다. 먼저 148 개, 227 개의 특징 값을 사용하여 분류율 실험을 하고 다음으로 SFFS 알고리즘을 이용하여 선정된 37 개의 최적 특징 값을 가지고 실험을 하였다. Fig. 10 은 k 값의 변화에 따른 오분류율(error rate)를 나타내며 오분류율의 오차가 가장 적은 k 가 7 인 경우에 각 특징 개수별 오분류율을 살펴보면 37 개는 15.779%, 148 개 16.181%, 227 개는 15.126%로 나타난다. 전체적으로 특징 값 37 개는 특징 값 148 개 보다 분류율의 성능이 향상되었으며 특징 값 227 개일 때의 분류율과 거의 비슷한 성능을 보인다. 하지만 전체 테스트 결함들의 분류시간은 37 개는 18 분, 148 개는 67 분, 227 개는 134 분으로 큰 차이를 보였다. 실제 검사 시

특징 값 추출 및 분류는 수학적 계산으로 많은 연산시간을 필요로 하게 되므로 분류율의 성능에 영향을 미치지 않으면서 특징 값의 개수를 줄이는 것이 분류기의 복잡성을 단순화하고 처리 속도를 빠르게 할 수 있다는 결과를 보여주므로 37 개의 특징 값만 가지고도 분류율 성능을 유지하면서 결함의 분류가 가능하다는 것을 알 수 있다. Fig. 10의 분류율 결과는 검사자의 시각에 의한 것 보다 매우 우수하며 실제 검사 장비에 사용되는 독일 파시텍(Parsytek) 회사의 C5 분류기와 대등한 성능을 보여준다.

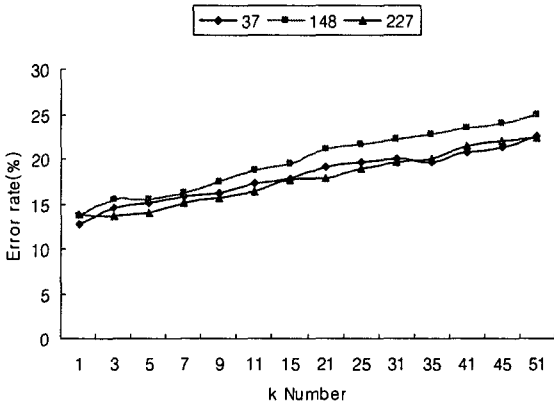


Fig. 10 Error rates of KNN Classifier

Fig. 11(a), (b), (c) 는 특징 값의 개수 37, 148, 227 개에 대하여 k 값의 변화에 따른 각 결함별 분류율의 결과를 보여준다. 그림에서 보면 특징 값 개수의 변화에 따라 특정 결함들의 오분류율이 급격히 증가하는 것을 볼 수 있는데 각 결함별 분류율의 결과가 차이가 크지 않도록 적합한 특징 데이터의 종류와 개수가 수반되어야 하는 것을 알 수 있다.

KNN 분류기의 k 값을 선정하는데 있어서 각 결함별 분류율의 오차가 적어야 신뢰성 있는 결과를 기대할 수 있다. 이 논문에서 KNN 분류기의 적정 특징 값 개수로 선정된 Fig. 11(a)의 37 개에 대한 분류율 결과를 살펴보면 k가 7일 때 각 결함별 분류율의 오차가 가장 적고 k가 17 이상인 경우 Red scale 과 Pseudo 의 오분류율이 급격히 증가하고 있다. k 값을 17 이상으로 할 경우 이 두 가지 결함의 오분류율이 높아져 신뢰할 수 없게 되므로 k의 상수 값을 7로 선정하는 것이 분류기의 신뢰성을 높인다.

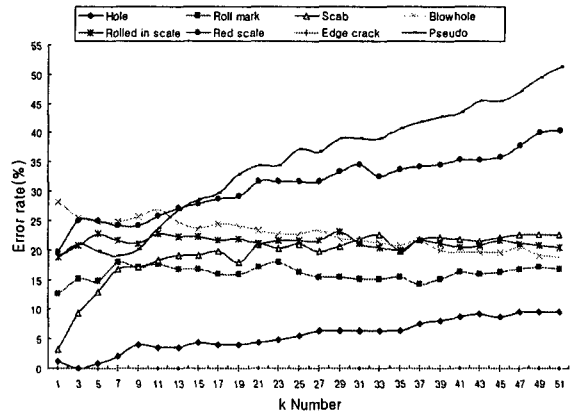


Fig. 11(a) Error rates of each defect with 37 features

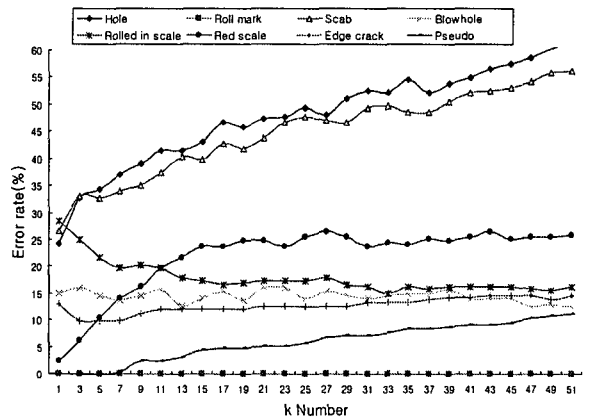


Fig. 11(b) Error rates of each defect with 148 features

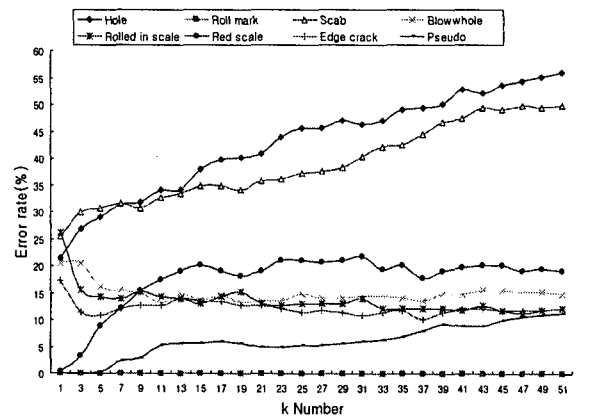


Fig. 11(c) Error rates of each defect with 227 features

6. 결론

본 연구에서는 강 코일의 표면 결함 검사 자동화를 위한 하드웨어 제작 기술 및 소프트웨어 알고리즘, 전체 시스템의 통합에 관하여 연구하였다.

하드웨어적으로는 실시간 강 코일 표면 결함 검사를 하기 위한 검사장비의 설계와 제작에 있어서 검사 공정과 결함 종류에 따른 최적 조명 기술과 카메라 선정과 검사법 설정 등의 단위 요소 기술을 확보하였다. 또한 고속에서 각 장비들간의 동기화를 하드웨어 제어 프로세서인 마이크로 컨트롤러와 컴퓨터의 인터페이스로 구현하였다.

소프트웨어 측면에서는 고속에서의 이미지 획득과 백그라운드 노이즈 제거, 이진화, 결함 검출 및 특징 추출 등의 알고리즘을 개발하였다. 결함 분류기의 성능을 높이기 위한 특징 데이터의 선정과 특징 데이터의 개수를 결정 짓는 다양한 통계적인 방법들을 시도하였으며, 선정된 특징데이터와 개수를 가지고 KNN 분류기에 의하여 샘플 결함 데이터를 이용하여 분류 성능을 평가하였다. 개발된 KNN 분류기는 분류율의 성능은 만족할 만한 수준이나 최적의 특징 종류와 개수가 분류기의 성능을 좌우하므로 이에 대한 깊은 연구가 필요하다.

후 기

이 논문은 한국산업기술재단(KOTEF)의 지역 혁신 인력 양성 사업 연구비 지원에 의하여 연구되었습니다.

참고문헌

1. Song, S. J., Kim, H. J., Choi, S. H. and Lee, J. H., "Classification of Surface Defects on Cold Rolled Strips by Probabilistic Neural Networks," Journal of the Korean Society for Nondestructive Testing, Vol. 17, No. 3, pp. 162-173, 1997.
2. Kim, K. M., Park, G. T., Park, J. J., Lee, J. H., Jung, J. Y. and Lee, J. K., "Development of Surface Defect Inspection Algorithms for Cold Mill Strip," Journal of Control Automation and System Engineering, Vol. 3, No. 2, pp. 179-186, 1997.
3. Lee, S. W., Park, W. J., Joo, J. I., Jung, J. Y., Lee, J. H. and Park, C. H., "Development of Surface Inspection System for #1 PCM," Technical Research Lab. POSCO, 2000.
4. Gonzalez, R. C. and Woods, R. E., "Digital Image Processing," Addison-Wesley, 1992.
5. Randy, C., "A simplified approach to image processing," Prentice Hall, 1996.
6. Gose, E., Honsonbaugh, R. and Jost, S., "Pattern Recognition and Image Analysis," Prentice Hall, 1996.
7. Umbaugh, S. E., "Digital Image Analysis and Processing," CRC Press, 2005.
8. Choi, Y. I., Park, K. H. and Kee, C. D., "Rejection Scheme of Nearest Neighbor Classifier for Diagnosis of Rotating Machine Fault," Journal of the Korean Society of Precision Engineering, Vol. 19, No. 3, pp. 52-58, 2000.