

KNN 분류기에 의한 강판 표면 결함의 분류

김철호* (서울산업대 대학원), 최세호(포스코 기술연구소), 주원종(서울산업대 기계설계자동화),
김기범(서울산업대 기계설계자동화)

Classification of Surface Defects on Steel Strip by KNN Classifier

C. H. Kim(Mechatronics. Dept. SNUT), S. H. Choi(Research. POSCO), W. J. Joo(MDAE. Dept. SNUT),
K.B. Kim(MDAE. Dept. SNUT)

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 light and area camera system and the defect images are preprocessed and segmented in real time for feature extraction. 4113 defect samples of cold roll steel strips 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(표면검사), k Nearest Neighbor(KNN, k 최근접 이웃), Synchronization(동기화), IR LED(적외선 조명), Feature Extraction(특징추출)

1. 서론

우리나라는 포스코를 비롯한 경쟁력이 우수한 제철 제강 회사들이 있어 세계적인 철강 강국 대열에 속하여 왔다. 이 중에서 강 코일의 생산이 매우 중요한데 이는 강 코일의 품질과 가격이 우리나라의 주력산업인 자동차 산업이나 가전산업에 직접적으로 영향을 미치기 때문이다. 강 코일을 생산하는 데는 여러 개의 공정을 거쳐야 하는데 대체적인 공정은 1)열연공정(Hot rolling), 2)담금공정(Pickling), 3)냉연공정(Cold rolling), 4)도금공정(Galvanizing), 5)마무리공정(Finishing) 등으로 나누어지고 여기에 열처리 공정(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 등을 정확히 정량적으로 산출하는 것이 선행되어야 한다. 우리나라 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 상하면을 비추는 거울 앞에 서서 두 거울을 번갈아 가며 바라보면서 지나가는 결함들을 검사한 후 검사자의 감(feeling)에 의해 등급을 매긴다.³ 이와 같은 시각에 의한 검사 방법은 부정확하여 통상 30%내의 오 판단율을 가지며 결함에 대한 아무런 정보 데이터를 남겨놓지 않아서 품질향상을 위한 feed-back 자료나 판매를 위한 품질 데이터 자료로도 사용할 수 없다. 이와 같은 사정에도 불구하고 코일 표면의 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 시도되지 않았으며 셋째는 각 공정에 따라 결함의 종류가 10 여 가지 이상이 되며 같은 종류의 결함이라도 코일에 따라 그 형태와 명암이 다양하기 때문에 상업용 장비를 도입할 경우에도 장기간에 걸친 분류기의 친숙화 작업이

요구되며 많은 투자가 지속되어야 한다. 또한 외국 산 검사장비의 성능을 보면 결함검출은 90% 이상이나 분류능력은 75~85% 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 설비 투자를 심각히 생각하지 않을 수 없는 상황이다.

표면 결함검사 장비(Fig. 1)의 주요기능은 고속 이송 중인 강 코일의 표면을 실시간으로 결함의 위치를 파악하여 결함 이미지를 획득하고 고속 이미지 프로세싱 기법을 이용하여 노이즈 제거(Filtering), 영역분할(Area segmentation), 이진화(Binary), 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.

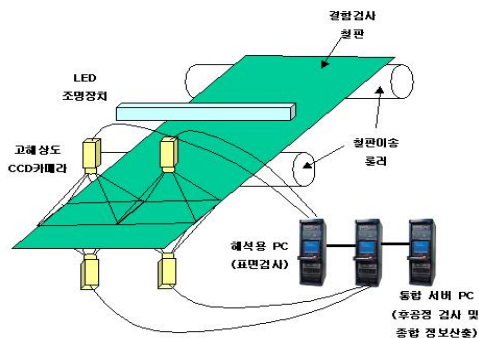


Fig. 1 Steel strip surface inspection system

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 개발하였으며 소프트웨어는 고속 전처리 알고리즘, 특징 선정(Feature selection) 및 추출 알고리즘을 개발하였다. 결함종류판독을 위해서 패턴인식에 많이 사용되는 KNN(k Nearest Neighbor)기법을 이용한 분류기를 개발하였다.

2. 실험 장치

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

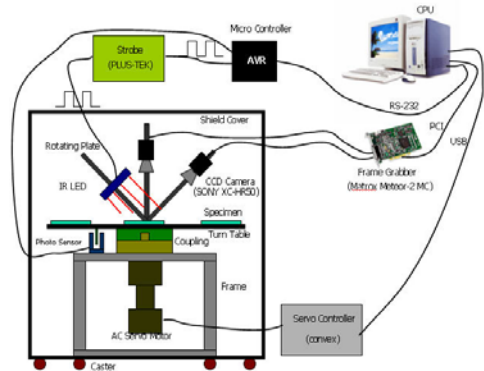


Fig. 2 Prototype surface inspection system

실험장치는 크게 기구부, 제어부, 광학부, 해석 및 데이터 처리부의 4 단계로 구성된다.

2.1 기구부

조명과 카메라를 사용하여 영상 이미지를 획득함에 있어 우선 외부 빛에 의한 노이즈를 차단하는 것이 아주 중요하므로 실험 장치는 차폐막 구조로 하였다. 검사할 시험철판의 크기가 200x200(mm)이므로 회전테이블에 6 개의 철판이 등 간격으로 위치할 수 있도록 고려하여 회전 테이블의 크기($\phi 800\text{mm}$)가 결정되었다. 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. 모터는 회전속도가 항상 일정한 속도를 유지해야 하므로 전용 서보 모터 제어기(Convex社)를 사용하여 컴퓨터로 제어하도록 되어 있으며, 신호 제어부는 ATmel사의 AVR 8bit 마이크로 컨트롤러를 사용하였다. 회전 테이블은 AC 서보 모터에 의해 정해진 정속도로 구동되며 2 대 카메라는 스텝(Step) 모터에 의해 각도를 조절할 수 있다.

2.3 광학부

머신 비전 시스템에서 광학계의 구성은 시스템의 성능을 결정 짓는 중요한 부분이다. 이 실험 장치에서 조명은 노이즈의 영향을 적게 발생하는 장파장의 IR(Infrared, 적외선) LED를 사용하였다.⁵ 결함은 종류에 따라 반사광에 잘 잡히는 결함과 산란광에 잘 잡히는 결함으로 나뉘어지며 이를 위해 두 가지의 카메라 위치를 정하였다. 조명의 정반사 각도에 카메라를 설치하는 Bright field(BF)와 조명의 난반사에 카메라를 설치하는 Dark field(DF)의 구조로 하였다.⁵ 카메라 타입은 진동에 덜 민감한 Area 타입으로 최대 전송률이 120fps, 해상도 640x480인

카메라를 선정하였다.

2.4 해석 및 데이터 처리부

결합 이미지의 해석과 그 결과 데이터를 처리하기 위한 시스템의 구성은 강 코일 폭이 넓은 관계로 결합 이미지의 수집과 특징추출 및 판독을 하는 여러 대의 카메라와 컴퓨터가 커플링으로 구성되어 있으며 각 컴퓨터로 나오는 결과를 하나로 묶어 전체 강 코일의 결합 정보와 위치, 빈도수에 따라 등급을 관리하는 통합 서버 컴퓨터 시스템으로 나누어 구축되어야 한다.⁵ 본 연구에서는 통합 서버 컴퓨터 시스템은 제외되어 있으며 해석용 컴퓨터의 고속 이미지 프로세싱과 결합검출, 특징추출 등의 알고리즘이 구현되어 있다.

3. 동기화

고속에서 영상 이미지가 밀리지 않고 정지한 듯한 영상을 얻기 위해서는 정확한 타이밍 플랜(Plan)에 따라 카메라 노출과 조명을 터트려 이미지를 획득하는 것은 매우 중요하다. Fig. 3은 현 실험장치에서 각 하드웨어 출력 신호들의 동기화 타이밍 차트이다.

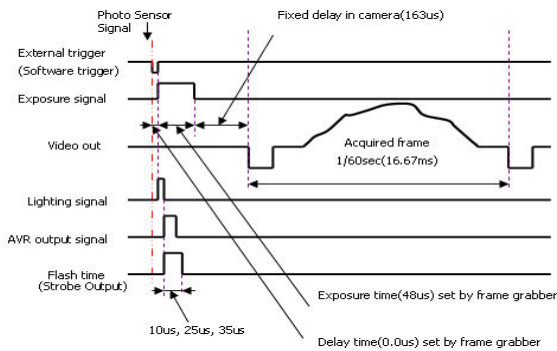


Fig. 3 Synchronization timing chart

4. 결합 분류기

4.1 k Nearest Neighbor(KNN)

k 최근접 이웃(KNN)은 기억 기반 추론기법 중에 하나로 패턴인식에 많이 활용되고 있다. k 최근접 이웃 방법은 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.²

입력패턴과 표준패턴과의 유사도는 유클리디언 거리(Euclidian distance)를 사용하였다. 만약 입력패

턴과 표준패턴이 아래와 같다면

$$X = [x_1, x_2, \dots, x_n], Y = [y_1, y_2, \dots, y_n]$$

두 패턴 간의 Euclidean distance는 다음과 같다.²

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

4.2 KNN 분류 프로시저

1. k 값을 선정한다.
2. 새로운 결합과 표준결합들의 특징 값들을 이용하여 유클리디언 거리를 계산한다.
3. 계산된 거리 값들 중에서 가장 가까운 거리 값의 표준결합을 k 개 만큼 선정한다.
4. k 개만큼 선정된 표준결합들 중 가장 많이 선택된 결합의 타입을 입력 결합의 타입으로 선정한다.

5. 결합 분류

현재 학교 실험실에서는 많은 결합 샘플 이미지를 얻을 수 없는 관계로 포스코로부터 강 코일의 품질에 크게 영향을 미치는 대표적인 8 개 결합들의 샘플 이미지 데이터 4113 개를 제공 받았다. 이 결합 샘플이미지와 특징 데이터를 가지고 KNN 분류기로 결합의 종류를 판별하였다.

5.1 결합종류와 발생원인

본 연구에서는 다루는 결합의 종류와 발생원인은 아래와 같으며 대표적인 샘플이미지는 Fig. 4에 보여진다.

- 1) Hole: Slab의 심한 내부 크랙, 이물 치입
- 2) Roll mark: 압연 및 처리 라인의 각종 Roll에 이물 부착, 롤 타흙 발생
- 3) Scab: Slab 손질 불완전 및 Slab 핸들링 불량
- 4) Blow hole: 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 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: 결합이 아닌데 결합으로 검출되는 흠

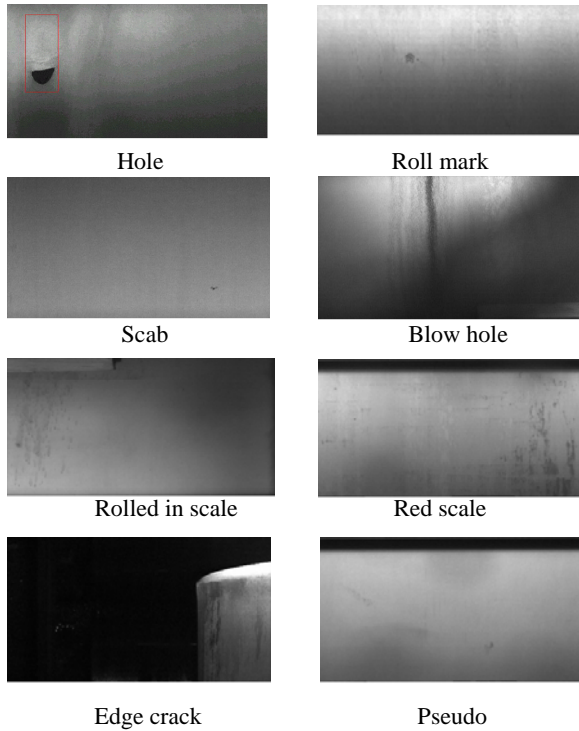


Fig. 4.8 defect types in rolled steel strip

5.2 이미지 전처리

KNN 분류기의 분류율 성능을 시험하기 위해서 표준패턴으로 쓰일 결함 그룹과 테스트용 결함 그룹을 만들기 위해서 샘플 이미지로부터 결함 부위의 영역을 적정크기로 잘라내고 불균일한 조명 영향을 제거한 다음 다중 임계치를 사용하여 이진화를 하였다. Fig. 5 는 실제 검사 장비에서 얻은 Roll mark 이미지에서 결함영역만 잘라내어 조명영향을 제거하고 이진화한 결과이다.

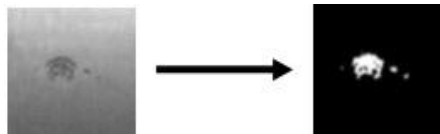


Fig. 5 Original and binary images

5.3 특징 추출

일반적으로 영상처리에 있어서 특징 추출은 어렵고 시간이 오래 걸리는 단계에 속한다. 강 코일 표면 흠 결함 이미지는 여러 가지의 기하학적인 정보와 발생 상황에 따른 명암 대비의 정보를 가지고 있다.^{1,2}

5.3.1 특징 데이터

이번 연구에 쓰인 특징 데이터들 중에서 대표적인 특징들은 다음과 같다.

1) 기하학적 특징(Geometric features)

대상체 (Object)의 윤곽선과 그를 둘러싸는 Box 를 기초로 한 물체의 특징을 말하며, 내용은 결함의 면적(def_area), 결함과 아닌 면적의 비(area_ratio = def_area / box_area), 결함의 응집도(Compactness = $(4\pi * \text{area}) / \text{perimeter}$)등의 정보들이 이에 속한다.⁴

2) 모멘트 특징(Moment features)

모멘트 특징량은 결함의 장단축비, 장축(Major axis) 길이, 단축(Minor axis)길이, 퍼짐(Spreadness)정도가 이에 속하며 흠의 중심 모멘트를 사용하여 구한다.⁴

3) 동시 발생 행렬 특징

흠이 서로 겹치는 경우에 대하여 통계적 특징량 추출기법을 사용하여 특징량을 구한다. 특징량들은 Energy, Entropy, Contrast, Inverse difference moment, Correlation 등이 있다.⁴

5.3.2 특징 데이터 선정

현재 포스코에서 추출한 결함 특징 종류는 378 개 이다. 그러나 유한한 샘플을 사용하는 흠 분류기를 설계할 경우 초기에는 Feature 의 수가 증가함에 따라서 분류기의 오 분류가 감소하다가 어느 점에서 정점에 이르고 Feature 수가 증가함에 따라서 오 분류가 증가한다(Peaking 현상). 또한 Feature 수가 증가함에 따라 요구되는 결함 샘플 수가 기하급수적으로 증가하기 때문에 적절한 feature 수의 선택이 매우 중요하다. 본 연구에서는 특징 데이터의 분석을 다음과 같이 하였다. 첫 번째는 Outlier 를 이용하여 흠에 이상치를 가지는 특징 데이터를 제거한다. 두 번째는 Anova(Analysis of variation) 테스트로 흠 별로 구분할 가능성이 적은 특징 데이터를 제거하고 3 번째는 흠 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 Bottom up method 로 작은 feature set 에서 분류능력이 우수한 feature 를 점진적으로 추가하는 방법이다. SFFS 알고리즘을 사용하여 위에서 선정된 148 개의 데이터로부터 다시 한번 최적의 37 개의 특징 데이터를 선정하였다.

5.4 결함 분류 실험

포스코로부터 획득한 총 4113 개의 결함 샘플 중 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). 각 결함 샘플마다 선정된 148 개의 특징 데이터와 SFFS 에 의해 선정된 37 개의 특징 데이터를 준비한다.

5.4.1 샘플 수 조정 및 일반화

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 결함별 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 조정하였다. 또한 각각 특징 값의 크기 범위(order of magnitude)가 Euclidean distance 산출 시 큰 값을 갖는 특징 값이 큰 영향을 주게 된다. 이를 방지하기 위하여 각 특징 값마다 표준편차(Standard deviation)를 구하여 실제 값으로 나누어 주고 Normalizing 작업을 하였다.

5.4.2 실험 결과

먼저 148 개의 특징 값을 사용하여 실험을 하고 다음으로 SFFS 알고리즘을 이용하여 선정된 37 개의 최적 특징 값을 가지고 실험을 하였다. 4113 개의 샘플 데이터에서 결함별 샘플 데이터 수를 조절하여 최종 3980 개의 결함 데이터를 구성하였다. 이를 훈련용 Training set 와 성능 평가용 Test set 로 반반씩 나누어 실시한다. Fig. 6 은 k 값의 변화에 따른 분류 오차율(Error rate)를 나타내며 가장 오차율이 적은 k=3 일 때의 분류율을 살펴보면 특징 값 148 개인 경우의 분류율은 84.572%이고 특징 값 37 개 일 때의 분류율은 85.427%를 나타냈다. 이 분류율은 검사자의 시각에 의한 것 보다 매우 우수하며 독일 Parsytek 사의 C5 분류기와 대등한 성능을 보여 준다.

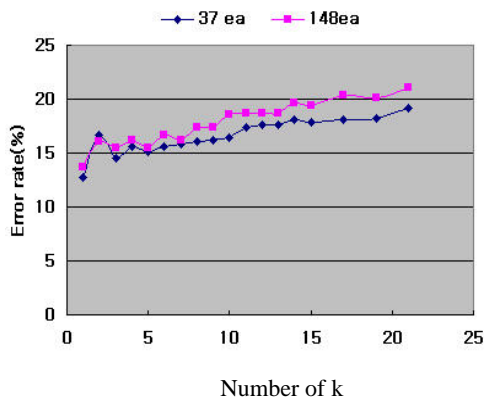


Fig. 6 Error rate of KNN Classifier

특징 데이터가 148 개인 경우보다 SFFS 알고리즘에 의해 선정된 37 개가 전체적으로 k 값이 10 이하인 경우에는 약 1% 정도의 분류율 성능이 높으

며 k 값이 10 이상부터는 점진적으로 더 큰 분류율의 성능차이를 보인다. 분류기의 성능은 결함의 특징 수를 많이 한다고 성능이 우수해 지지 않으며 최적의 특징종류와 개수를 찾는 것이 중요하다. 본 실험에서 보듯이 적절한 특징 수가 보다 우수한 성능과 계산시간 단축을 가져온다.

6. 결론

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, 다중 임계치를 이용한 이진화 기법, 결함 검출 및 특징 추출 등의 알고리즘을 개발하였다. 또 결함의 종류를 판별하기 위한 KNN 분류기를 직접 설계하여 실험하였다. 개발된 KNN 분류기는 분류율의 성능은 만족할 만한 수준이나 최적의 특징종류와 개수가 분류기의 성능을 좌우하므로 이에 대한 깊은 연구가 필요하다.

후 기

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

참고문헌

1. R. C. Gonzalez, R. E. Woods, "Digital Image Processing," Addison-Wesley, 1992.
2. E. Gose, R. Honsonbaugh, S. Jost, "Pattern Recognition and Image Analysis," Prentice Hall, 1996
3. 송성진, 김학준, 최세호, 이종학, "확률신경회로 망에 의한 냉연 강판 표면결함의 분류," 비파괴 검사 학회지, Vol. 17, No. 3, 1997.
4. 김경민, 박귀태, 박중조, 이종학, 정진양, 이주강, "냉연 표면 흠 검사 알고리즘 개발에 관한 연구," 제어 자동화 시스템공학 논문지, 제 3 권, 제 2 호, 1997.
5. 이시우, 박옥재, 주진일, 정진양, 이종학, 박창현, "1 냉연 #1PCM-PL 표면결함 탐상기 적용 결과," (주)포스코 기술연구소, 2000.