

# 이산 웨이블릿 변환과 명암도 동시발생 행렬을 이용한 컬러 레이저프린터 판별 알고리즘

백 지 연<sup>†</sup> · 이 흥 수<sup>†</sup> · 공 승 규<sup>†</sup> · 최 정 호<sup>\*\*</sup> · 양 연 모<sup>\*\*\*</sup> · 이 해 연<sup>\*\*\*\*</sup>

## 요 약

고성능 저가의 디지털 인쇄기의 출현으로 불법적인 위변조가 사회적인 문제로 대두되고 있고, 이를 해결하기 위해서 디지털 포렌식 기술이 필수적이다. 본 논문에서는 컬러 디지털 인쇄기를 판별하기 위한 디지털 포렌식 기술을 제안한다. 컬러 디지털 인쇄기는 제조사마다 인쇄방법이 다르기 때문에, 출력물에 작은 차이가 존재한다. 이와 같은 차이점을 활용하면, 임의의 주어진 출력물에 대해 어떠한 인쇄기로 출력되었는지 구별이 가능하다. 제안하는 방법에서는 차이점을 구별하기 위하여 출력물을 스캔한 디지털 이미지에 대해 이산 웨이블릿 변환을 수행하여 계산한 고주파 영역을 추출한다. 이에 대해 명암도 동시발생 행렬을 계산한 후에 행렬 데이터의 표준편차, 첨도, 왜도, 공분산, 상관계수의 특징을 추출하였다. 추출된 특징을 서포트 벡터 머신 분류기에 적용하여 디지털 인쇄기를 판별하였다. 제안한 알고리즘의 성능을 분석하기 위하여 총 2,597장 이미지와 7대 프린터(HP, Canon, Xerox DCC400, Xerox DCC450, Xerox DCC5560, Xerox DCC6540, Konica)를 가지고, 기존 알고리즘과 비교 분석하였다. 그 결과에 따르면 제안한 알고리즘은 컬러 디지털 인쇄기를 판별하는데 있어서 평균 96.9% 정확률을 보였다.

키워드 : 디지털 포렌식, 이산웨이블릿변환, 명암도 동시발생 행렬, 서포트 벡터 머신 분류기

## Color Laser Printer Identification through Discrete Wavelet Transform and Gray Level Co-occurrence Matrix

Ji Yeoun Baek<sup>†</sup> · Heung-Su Lee<sup>†</sup> · Seung Gyu Kong<sup>†</sup> · Jung-Ho Choi<sup>\*\*</sup> · Yeon-Mo Yang<sup>\*\*\*</sup> · Hae-Yeoun Lee<sup>\*\*\*\*</sup>

## ABSTRACT

High-quality and low-price digital printing devices are nowadays abused to print or forge official documents and bills. Identifying color laser printers will be a step for media forensics. This paper presents a new method to identify color laser printers with printed color images. Since different printer companies use different manufactural systems, printed documents from different printers have little difference in visual. Analyzing this artifact, we can identify the color laser printers. First, high-frequency components of images are extracted from original images with discrete wavelet transform. After calculating the gray-level co-occurrence matrix of the components, we extract some statistical features. Then, these features are applied to train and classify the support vector machine for identifying the color laser printer. In the experiment, total 2,597 images of 7 printers (HP, Canon, Xerox DCC400, Xerox DCC450, Xerox DCC5560, Xerox DCC6540, Konica), are tested to classify the color laser printer. The results prove that the presented identification method performs well with 96.9% accuracy.

Keywords : Digital Forensics, Discrete Wavelet Transform, Gray Level Co-occurrence Matrix, Support Vector Machine Classifier

### 1. 서 론

수많은 디지털 기기의 등장은 대중들에게 많은 편의와 이

익을 가져오고 있다. 최근 발매 되는 디지털 기기는 저렴한 가격과 뛰어난 성능을 가지고 있으며 온라인을 통하여 누구나 쉽게 디지털 기기를 구매 할 수 있다. 그 중 영상 장비는 디지털 기기 중에 가장 큰 부분을 차지하고 있다. 특히 고성능 디지털카메라의 보급으로 선명한 사진을 출력하기를 원하는 사용자가 늘어나고 컬러 레이저프린터의 대중화로 인해 일반인들도 손쉽게 위·변조를 할 수 있게 되면서, 사회

※ 본 연구는 금오공과대학교 학술연구비에 의하여 연구된 논문임  
† 준 회 원 : 국립금오공과대학교 컴퓨터공학부  
\*\* 정 회 원 : 한국과학기술원 전산학과 박사과정  
\*\*\* 정 회 원 : 국립금오공과대학교 전자공학부 교수  
\*\*\*\* 정 회 원 : 국립금오공과대학교 컴퓨터공학부 교수(교신저자)  
논문접수 : 2009년 12월 29일  
수 정 일 : 2010년 2월 24일  
심사완료 : 2010년 3월 12일

적인 위·변조 범죄가 증가하고 있다.

디지털 이미지 위·변조로 인해 몇 년 전부터 디지털 이미지의 신뢰성이 점차 중요해 지고 있다. 또한 최근에 멀티미디어 법이 선고됨에 따라 저작권이 사회적으로 큰 이슈가 되고 있다. 저작권이란 창작자가 가지는 권리로서, 창작자가 창작물에 대한 저작권이 있는 디지털 이미지를 허락 없이 배포 하거나 인쇄를 하게 되면 법적으로 처벌하는 것이다. 이에 따라 저작권을 보호하기 위한 위·변조에 대한 방지 대책과 저작권 범죄 해결에 대한 사람들의 관심이 높아지고 있다. 온라인상으로 배포되는 디지털 이미지는 워터마킹 기술을 통해 저작권 문제를 많이 해결하고 있지만 프린터를 사용하여 인쇄된 후 오프라인을 통해 배포되는 디지털 이미지 콘텐츠에 대해서는 아직 저작권과 관련된 다수의 문제를 가지고 있다.

컬러 레이저프린터의 경우 무분별한 콘텐츠 복제를 방지하기 위하여 문서를 인쇄할 때 인쇄한 날짜와 시간 및 프린터에 대한 시리얼 정보 등을 시각적으로 잘 인지되지 않는 노란색 점으로 부가적으로 인쇄하였다. 이와 같은 기술을 프린터 스테가노그래피라고 한다[1, 2]. (그림 1)에는 프린터 스테가노그래피의 예를 나타낸다. 좌측 영상에는 노란색 점이 찍혀있는 예를 보여주며, 우측 영상은 해당 점들이 의미하는 시간, 장비번호 등을 나타낸다. 그러나 이 기술은 시각

적으로 크게 인지되지 않더라도 출력물에 인쇄가 되기 때문에 부가적인 응용에 있어서 화질 및 인쇄물의 품질 저하가 일어날 수 있다. 따라서 최근에는 기계에 종속적이지 않은 디지털 포렌식 기술 연구를 하는 방법이 연구되고 있다.

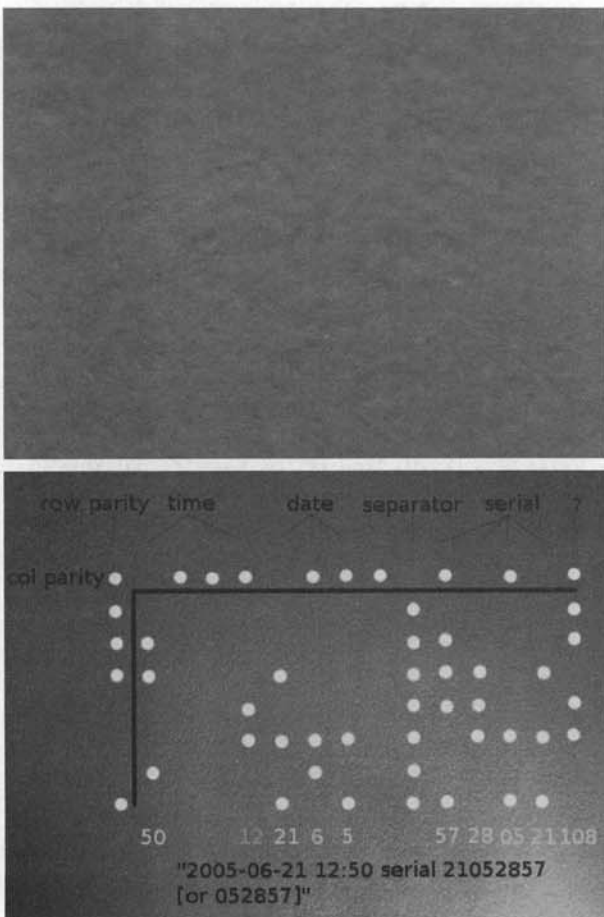
디지털 포렌식 기술은 이러한 문제를 해결하기 위해서 영상에 숨겨져 있는 고유의 특성을 찾아내서 위변조를 탐지하는 것을 목표로 한다. 특히, 디지털 이미지 포렌식 기술이란 디지털 영상에 포함되어있는 고유의 특성을 이용해서 이미지를 만들어 낸 기기를 밝혀낸다는지, 이미지의 위조 여부를 판별하는 것이다.

디지털 이미지 포렌식 기술은 영상 획득 장비 판별, 영상 위조 탐지 기술로 크게 나뉘어져 연구되고 있다. 영상 획득 장비 판별 기술은 주어진 영상으로부터 영상을 획득하는데 사용된 장비가 무엇인지 판단하는 기술이다. 주어진 영상만을 이용해서 영상을 획득하는데 사용한 장비의 제조사 뿐만 아니라 정확한 모델명 수준까지 판별하는 것을 목표로 한다. 영상 위조 탐지 기술은 주어진 영상이 위변조 단계를 거쳤는지 여부를 판별하는 기술이다. 디지털 이미지의 위변조는 눈에 띄는 단서를 남기지 않지만, 영상을 구성하는 통계적 특성에는 변화를 준다. 그러한 변화된 통계적 특성을 찾음으로써 위·변조된 부분을 찾을 수 있다.

본 논문에서는 디지털 이미지 포렌식 중에서 영상 획득 장비에 해당하는 컬러 레이저프린터 판별 기술에 대하여 설명한다. 지금까지 연구되어진 프린터 판별 방법 기술은 주로 흑백 레이저프린터를 판별하는 기술들이 대부분이다. 요즘에는 많은 컬러 레이저프린터 보급률이 높아지면서 컬러 레이저프린터 판별 기술에 대한 적합한 판별기술이 필요하다. 컬러 레이저프린터 판별 기술을 통해 출력된 디지털 이미지를 어떠한 프린터에서 식별되었는지 알 수 있다면 저작권에 위배된 이미지를 인쇄한 범인을 찾는 데 도움이 될 수 있으며, 나아가 위조지폐 및 위조 공문서를 식별도 쉽게 이뤄 질수 있다. 이와 관련하여 2절에서 자세히 알아보도록 하겠다.

제안하는 컬러 레이저프린터 판별 기술은 이산 웨이블릿 변환과 명암도 동시발생 행렬을 사용하여 프린터의 고유한 특징을 추출하였다. 먼저 출력물을 스캔한 디지털 이미지에 대해 이산 웨이블릿 변환을 수행하여 계산한 고주파 영역을 추출한다. 이에 대해 명암도 동시발생 행렬을 계산한 후에 행렬 데이터의 표준편차, 첨도, 왜도, 공분산, 상관계수의 특징을 추출하였다. 추출된 특징을 Support Vector Machine 분류기에 적용하여 디지털 인쇄기기를 판별하였다. 제안한 알고리즘의 성능을 분석하기 위하여 총 2597장 이미지와 7대 프린터(HP, Canon, Xerox DCC 400, Xerox DCC 450, Xerox DCC 5560, Xerox DCC 6540, Konica)를 가지고, 기존 알고리즘과 비교 분석하였다.

본 논문의 구성은 다음과 같다. 2절에서는 프린터 판별과 관련된 기술들에 대하여 설명하고, 3절에서는 이산웨이블릿 변환과 명암도 동시발생 행렬을 활용한 제안하는 프린터 판별 기술을 설명한다. 4절에서는 다양한 프린터를 사용한 실



(그림 1) 프린터 스테가노그래피의 예

험 결과를 제시하고, 5절에서 결론을 짓도록 하겠다.

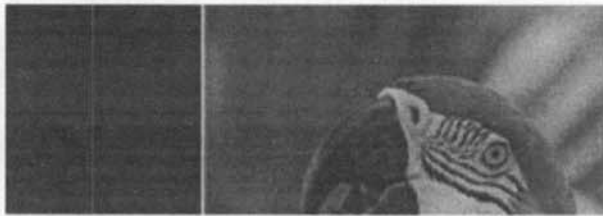
## 2. 관련 연구

### 2.1 줄무늬 주기를 이용한 프린트 판별 기술

프린터의 인쇄 과정 중 광학 광전도체 드럼을 돌리는 기어의 역회전 현상으로 인해 각속도가 일정하지 않게 됨에 따라 발생하는 준 주기적인 변동 때문에 인쇄 방향을 따라서 밝고 어두운 줄무늬가 주기적으로 생긴다. 이렇게 문서에 새겨지는 줄무늬 주기는 프린터의 기계적인 특성을 반영하고 있다고 할 수 있다. Delp et al.는 이와 같은 주기적인 특성을 활용하여 흑백 레이저프린터를 판별하는 알고리즘을 제안하였다[3]. (그림 2)은 레이저프린터 인쇄물에서 나타날 수 있는 줄무늬의 예이다[4].

일반적으로 하프톤 이미지 출력물을 이용하여 줄무늬 주기를 구하는 방법은, 프린터의 줄무늬 주기를 구하는 방법처럼 가로 방향으로 투사한 다음 이진푸리에변환을 이용하는 방법을 그대로 이용하면 된다. 프린터 판별 방식은 어떠한 프린터로 인쇄했는지 모르는 하프톤 이미지 출력물의 줄무늬 주기를 위의 방법으로 구한 다음 그 줄무늬 주기와 일치하는 줄무늬 주기를 갖는 프린터를 찾는 것이다.

이러한 줄무늬 주기는 하프톤 이미지 출력물에는 적용하기 쉽고 결과도 좋게 나오지만 하프톤 이미지 출력물이 아닌 일반 텍스트 출력물에는 적용하기 힘들다는 단점이 있다.

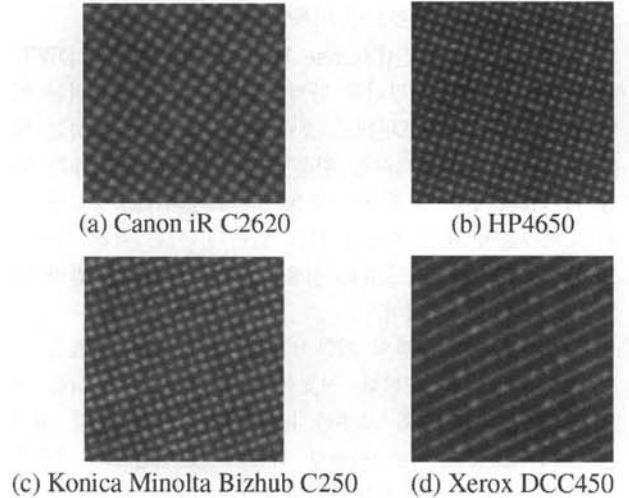


(그림 2) 흑백 레이저프린터에서 인쇄된 출력물의 줄무늬 주기 현상

### 2.2 허프 변환을 이용한 프린트 판별 기술

컬러 레이저프린터가 색을 표현하기 위해서는 CMYK (Cyan, Magenta, Yellow, Black)의 4가지 색상을 조합하여 색을 표현한다. 조합된 4가지 색상을 분리하여 크게 확대해 보면 색상마다 일정한 패턴을 가짐을 확인할 수 있다. (그림 3)에는 프린터별 채널의 다른 주기적인 무늬 특성을 나타내었다. 따라서, 프린터마다 고유한 패턴을 가지기 때문에 이를 이용하여 프린터를 판별 할 수 있다. Ryu et al.은 이와 같은 특성을 활용하여 패턴에 존재하는 선형 특성의 방향성 정보를 허프 변환(Hough Transform)을 이용하여 추출하여 컬러 레이저프린터를 판별하는 기술을 연구하였다[5].

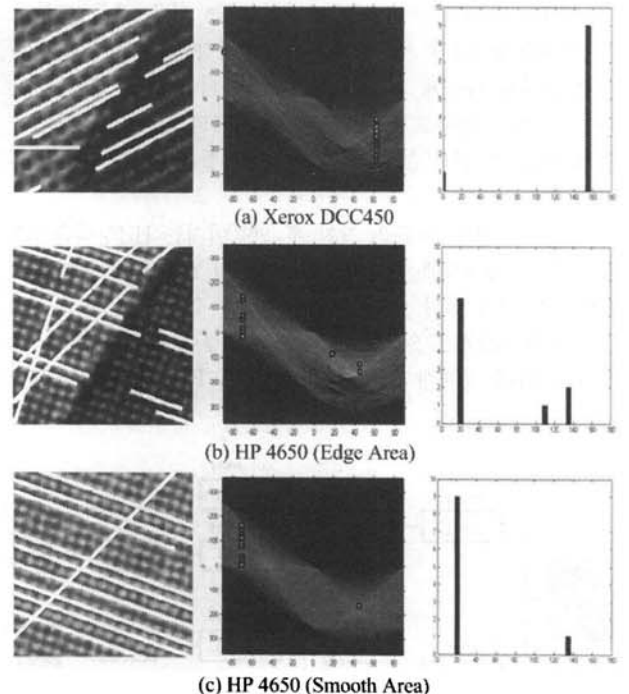
영상이 들어오면 에지 정보를 추출한 후에 이를 이용하여 허프 변환을 통하여 라인 정보를 추출하였다. 프린터마다 추출된 라인 정보는 고유하며, 라인의 각도 정보에 대한 히



(그림 3) 각 프린터별 Cyan채널의 다른 주기적인 무늬

스토그램을 활용하여 프린터 판별을 위한 특징값으로 사용하였다. (그림 4)에는 허프변환 후에 라인의 각도에 대한 히스토그램을 나타내었다. (그림 4b)와 (그림 4c)는 동일한 HP 프린터에서 영상의 특성만 다르게하여 추출한 특징값으로서 (그림 4a)의 Xerox 프린터와 확연한 차이가 있음을 확인할 수 있다.

하지만, 허프 변환을 이용한 프린터 판별 방법은 출력된 이미지의 색상이 가지는 패턴이 확인 가능할 정도의 고해상도 이미지가 필요하다. 따라서 저해상도의 이미지에는 적용하기 힘들다는 단점이 있다.



(그림 4) 프린터 Cyan 채널에 대한 허프변환 후 각도 히스토그램 특징값

### 2.3 이산 웨이블릿 변환을 이용한 프린트 판별 기술

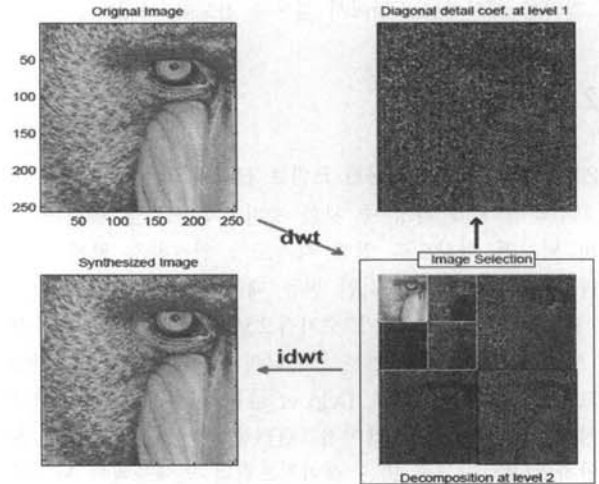
이산 웨이블릿 변환(Discrete Wavelet Transform: DWT)을 이용하여 컬러 프린터를 판별하는 방법은 출력하려는 이미지가 컬러 레이저프린터로 출력될 때 색을 재구성하는 방법이 프린터마다 다르다는 점을 착안한 것이다. Choi는 이산 웨이블릿 변환을 통하여 만들어진 고주파(HH)영역은 컬러 노이즈들에 대한 정보를 담고 있는데, 이는 컬러 레이저 프린터 마다 많이 다르다는 것을 활용하여 프린터 판별 알고리즘을 제안하였다[4].

이산 웨이블릿 변환의 기본적인 동작은 n개의 샘플을 갖는 이산 신호에 적용된다. 신호에 대해 한 쌍의 필터를 적용시켜 저주파 대역과 고주파 대역으로 분리한다. 각 대역은 2라는 요소로 서브 샘플링 되었으므로 n/2개의 샘플을 포함한다. 2차원 이미지의 각 행에 대해 저역 통과 필터와 고역 통과 필터를 적용시키고 2로 다운 샘플링을 수행하면 4개의 서브 이미지 LL, LH, HL, HH가 생성된다. 이러한 4개의 서브 밴드 이미지를 결합하여 원본 이미지와 동일한 샘플의 개수를 갖는 출력 이미지가 만들어진다. 참고로 LL은 원본 이미지에 수평과 수직방향으로 저역 통과 필터를 적용하여 2로 서브 샘플링된 것이고, HL은 수직방향으로 고역 통과 필터를 적용한 것으로 수직 방향의 주파수의 오차 성분을 포함하고 있다. LH는 수평 방향으로 고역 통과 필터를 적용한 것으로 수평방향 주파수의 오차 성분을 포함하고, HH는 수평과 수직 방향에 고역 통과 필터를 적용한 것이다.

이산 웨이블릿 변환의 필터 구조는 (그림 5)와 같으며, Lo\_D는 저역 필터를 거친 데이터를 의미하고 Hi\_D는 고역 필터를 거친 데이터를 의미한다. 이미지 영역을 우선 가로 방향으로 고역 필터와 저역 필터를 통해 두 대역으로 나누게 된다. 그 결과 데이터를 반만 취하는 다운 샘플링 과정을 수행한다. 다시 두 신호를 세로 방향으로 고역 필터와 저역 필터로 나누게 되고 다시 각 데이터는 다운 샘플링 된다. 그렇게 분석하고자 하는 영역을 4개의 영역으로 나누어지게 되고 그 결과 고역 필터를 두 번 걸친 HH영역이 나타나게 된다.

이산 웨이블릿 변환을 거친 후 만들어지는 HH영역은 (그림 6)에서 자세하게 나타나 있다. HH영역은 원본 이미지의 노이즈가 남아 있다.

Choi는 HH공간 영역은 컬러 노이즈들에 대한 정보를 활용하기 위해 출력한 문서를 스캔한 후 RGB영역과 YCbCr



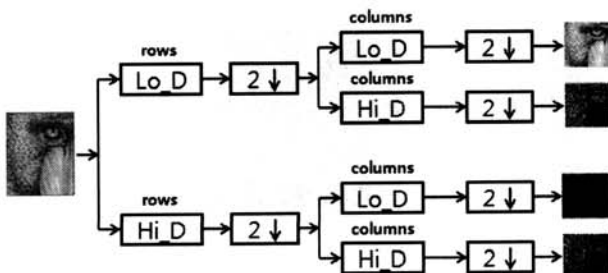
(그림 6) DWT 2레벨의 예제

영역에서 변환 후, 각 영역에 대해 이산 웨이블릿 변환을 수행하여 얻어진 HH영역에 대해 R,G,B 각각의 영역에 대한 표준편차, 왜도, 첨도를 분석하며 R과 G, G와 B, B와 R 영역간의 공분산과 상관계수를 분석한다. YCbCr영역에서도 마찬가지로 Y, Cb, Cr 각각의 영역에 대한 표준편차, 왜도, 첨도를 분석하며 Y와 Cb, Cb와 Cr, Cr과 Y 영역간의 공분산과 상관계수를 분석한다. 이와 같이 추출된 30가지의 특징점을 컬러 레이저프린터를 판별하는데 이용한다. Choi의 방법은 컬러 레이저프린터 제조사를 효율적으로 분류할 수 있지만 같은 제조사 내의 다른 모델은 효율적으로 분류하지 못한다는 단점이 있다. 또한 일부 제조사에 대해서는 낮은 판별률을 보여주었다.

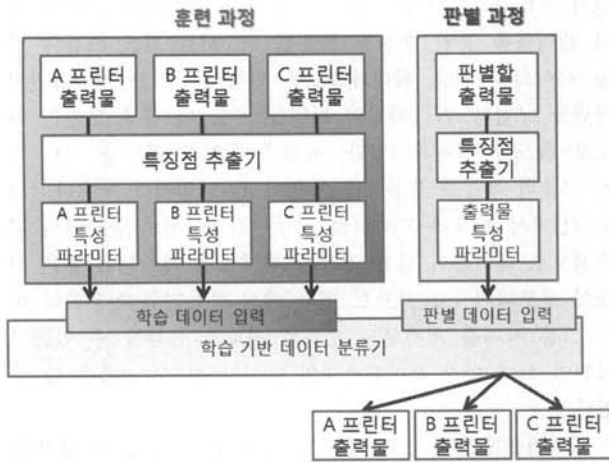
### 3. 제안하는 컬러 레이저프린터 판별 기술

본 논문에서 제안하는 컬러 레이저프린터 판별 기술은 혼련 및 판별의 2단계로 구성되어진다. 제안하는 기술의 대략적인 수행 과정은 (그림 7)에 도시되어 있다. 혼련 과정에서는 판별하고자 하는 프린터로 인쇄한 이미지를 사용하여 각 프린터의 고유한 특징을 추출하고, 학습기반 데이터 분류기의 파라미터로 입력하여 혼련을 수행한다. 판별 과정에서는 임의의 데이터가 들어왔을 때, 특징을 추출하고, 학습기반 데이터 분류기에 입력하여, 어떠한 인쇄기로 출력을 했는지 판별을 수행한다.

제안하는 판별 기술의 성능은 프린터에 존재하는 고유한 특징에 대한 정확한 추출과, 추출된 특징의 학습 분류기를 통한 올바른 분류 기능에 의하여 좌우된다. 본 연구에서는 프린터에 존재하는 고유한 특징을 추출하기 위하여 다양한 특징점에 대한 연구를 수행하였고, 가장 최적의 결과를 나타내는 이산 웨이블릿 변환 후의 고주파(HH) 성분에 대한 명암도 동시발생 행렬을 이용한 특징의 추출 방법을 제안한다. 특징 추출 방법에 대하여 3.1절에서 설명하고, 3.2절에서는 학습기반 데이터 분류기에 대해서 설명한다.



(그림 5) DWT 1레벨의 필터구조

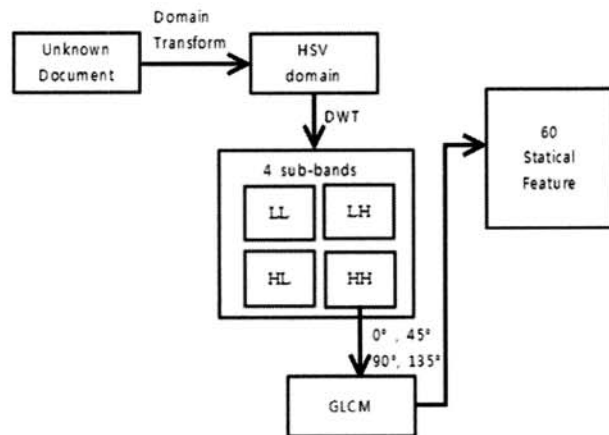


(그림 7) 제안하는 컬러 레이저프린터 판별 과정: 훈련과정 및 판별과정

### 3.1 명암도 동시발생 행렬을 통한 특징 추출

컬러 레이저프린터에 존재하는 고유한 특징을 추출하기 위하여, 먼저 컬러 레이저프린터로 인쇄된 디지털 이미지를 스캐너로 스캔하여, RGB 영역에서 HSV 영역으로 변환 시킨 다음 이산 웨이블릿 변환을 수행하였다. 이산 웨이블릿 변환을 수행하여 변환된 영역 중에서 컬러 노이즈의 정보를 갖고 있는 HH영역에 대해 명암도 동시발생 행렬로 변형 후, 이에 대하여 질감분석을 위한 대조, 상관계수, 합계, 유사성, 공분산을 계산하여 특징으로 사용한다. 명암 동시발생 행렬은 0도, 45도, 90도, 135도의 4가지 방향성에 대하여 계산을 수행하였다[5]. 이와 같은 이산 웨이블릿 변환과 명암도 동시발생 행렬을 이용한 제안한 알고리즘의 특징추출 과정은 (그림 8)과 같다.

사용하는 도메인 영역에 대하여 RGB 영역보다 HSV 영역에서 프린터마다 많이 차이가 나타나는 것을 발견하였다. HSV 영역은 사람이 컬러 감각을 경험하고 묘사하는 방식에 대해 RGB 시스템보다 훨씬 더 가깝다. H영역은 색상(hue),



(그림 8) 이산 웨이블릿 변환과 명암도 동시발생 행렬을 활용한 특징 추출 과정

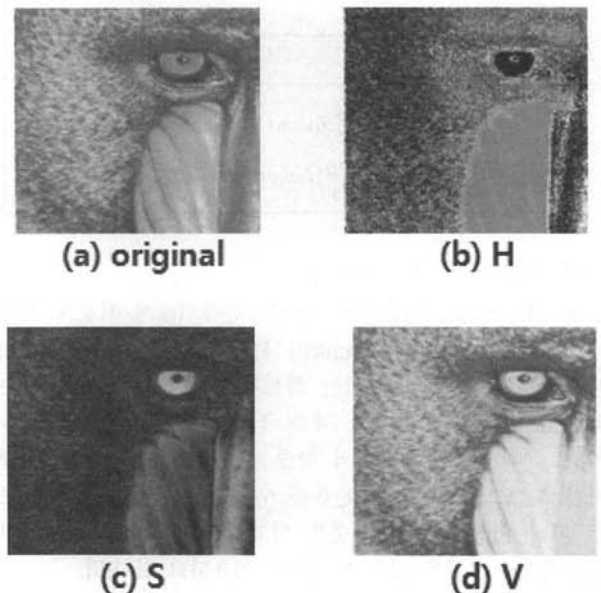
채도(saturation), 값(value)은 대략적으로 tint(밝은 색조/흰색과 혼합된 순수 색조/색), shade(어두운 색조/검정과 혼합된 순수 색조/농도), tone(회색조/검정과 흰색 둘 다 혼합된 순수 색조/농도)를 칭한다. HSV 영역으로 표현한 예제는 (그림 9)과 같다.

화소 기반의 통계적 Texture영상 생성을 위한 기본 개념은 Haralick et al.에 의해 제안된 명암도 동시발생 행렬 (Gray Level Co-occurrence Matrix)에 통해 체계화되었다 [6, 7]. Texture 영상의 응용 연구에서는 영상 분류 단계에서 Texture 영상을 추가적으로 적용하는 경우, 정성적인 분류결과와 정량적인 분류 정확도의 향상에 Texture 영상이 유용한 자료임이 여러 사례연구에서 밝혀진 바 있다. 그러나 Texture 영상이 유용함에도 불구하고, Texture 영상의 선택 기준에 대한 연구 결과는 실제적으로 제시된 바가 거의 없다.

명암도 동시발생 행렬은 영상 내에서 픽셀간의 밝기 값의 관계를 나타내는 방법으로 원본 이미지 (i, j) 위치에 있는 인접한 두 픽셀의 값을 좌표 값으로 이용하여 그러한 좌표 값이 얼마나 자주 나오는지 세어 놓은 행렬이다. 두 점을 어떻게 선택할 것인가는 방향과 두 점 사이의 거리로 (그림 10) 과 같이 표현할 수 있다.

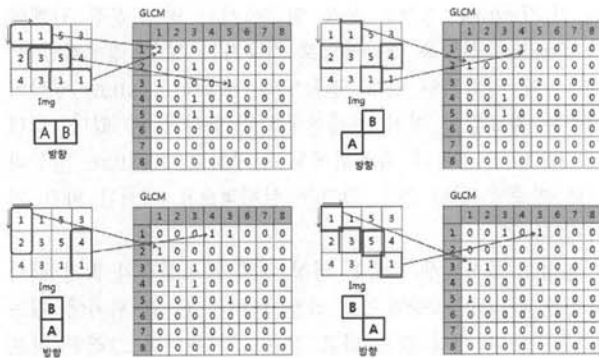
명암도 동시발생 행렬으로 변형된 영역에 대하여 <표 1>에 나타난 것과 같이, 유사성, 대조, 합계, 상관계수, 공분산의 통계적인 특성을 사용하여 각 특징을 추출하였다.

요약하면, 출력한 문서를 스캔한 후 HSV영역으로 변환 후, 각 영역에 대해 이산 웨이블릿 변환후의 4가지 방향성에 대한 명암도 동시발생 행렬 변환을 수행하여 얻어진 HSV 각각의 영역에 대한 유사성, 대조, 합계를 분석하며 H와 S, S와 V, V와 H 영역간의 공분산과 상관계수를 분석한



(그림 9) HSV 영역으로 표현된 이미지

다. 이와 같이 추출된 60가지의 특징점들, 3 채널(HSV) \* 4 방향성(0도, 45도, 90도, 135도) \* 5 질감분석(유사성, 대조, 합계, 상관계수, 공분산) = 60, 하나의 벡터로 합친다. 각 이미지마다 생성된 벡터를 프린터 제조회사별로 다시 벡터를 합친다. 이렇게 프린터 제조회사별로 생성된 벡터를 학습기 반 데이터 분류기를 이용하여 훈련과 판별에 사용하여, 컬러 레이저프린터를 판별하는데 이용한다.



(그림 10) 명암도 동시발생 행렬 방향성의 예제

<표 1> 명암도 동시발생 행렬에 대한 특징점

특징점	식	설명
유사성	$homogeneity = \sum_{i,j} \frac{p(i,j)}{1+ i-j }$	행렬 내 각 화소들 간의 균일한 정도를 표현
대조	$contrast = \sum_{i,j}  i-j ^2 p(i,j)$	화소간의 명암도 차이를 표현
합계	$energy = \sum_{i,j} p(i,j)^2$	화소들의 합을 표현
상관계수	$correlation = \frac{\sum_{i,j} (i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j}$	화소간의 상관관계를 표현
공분산	$Cov(I_1, I_2) = \frac{1}{N} \sum_{x,y=1}^N I_1(x,y) \times I_2(x,y) - \frac{1}{N} \sum_{x,y=1}^N ((I(x,y) - \bar{I}(x,y))^3)$	전체 이미지에서 두 픽셀이 연관된 정도를 표현

3.2 학습기반 데이터 분류기

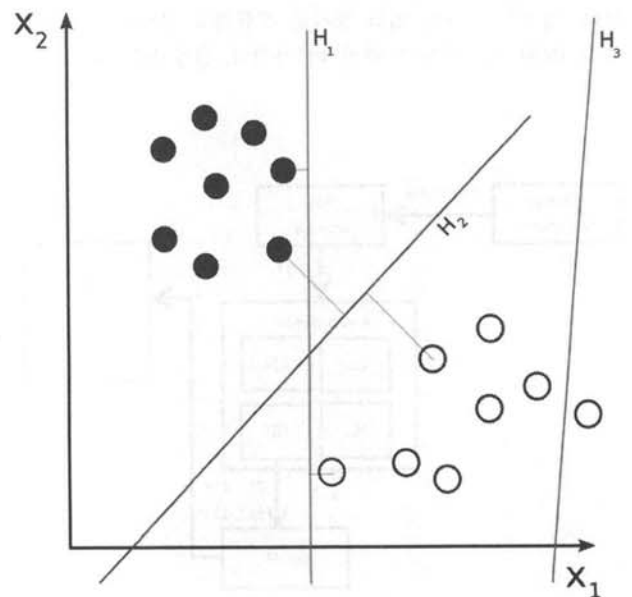
본 연구에서는 학습기반 데이터 분류기로서 서포트 벡터 머신(Support Vector Machine) 분류기를 사용하였다[8, 9]. 서포트 벡터 머신 분류기는 분류와 회귀 분야에서 쓰이는 학습 방법들 중 하나이다. 예를 들어 주어진 학습 세트들을 서포트 벡터 머신 분류기의 학습 알고리즘을 이용하여 학습시킨 후 모델을 만들고, 분류를 수행할 판별 세트들을 서포트 벡터 머신 분류기의 분류 알고리즘과 모델을 이용하여 최종적으로 분류를 하는 방식으로 이용되고 있다[9].

서포트 벡터 머신 분류기에서는 모든 학습 세트들과 판별 세트들을 점으로 취급한다. 서포트 벡터 머신 분류기는 N차

원의 벡터공간에서 분류하고자 하는 공간 사이에 그룹에 따라 레이블을 붙인 점들을 분포한 뒤, 서로 다른 그룹의 점들 사이의 거리를 최대화하도록 만들 수 있는 N-1차원의 평면을 구한다. N-1차원의 평면을 구할 때 분류 오차는 최소로 하도록 설계되어있다. 이렇게 구해진 평면을 서로 다른 그룹의 점들을 분류해주는 하이퍼플레인이라 부른다. (그림 11)에서 X<sub>1</sub>축과 X<sub>2</sub>축으로 이루어진 2차원의 공간에서 검은색으로 레이블이 붙은 점들과 흰색으로 레이블이 붙은 점들이 분포되어있고 서포트 벡터 머신 분류기를 이용하여 이 두 그룹 사이를 최소한의 오류로 최대한 분류할 수 있는 1차원의 최대 마진 하이퍼플레인 H<sub>2</sub>를 구할 수 있음을 알 수 있다.

하이퍼플레인은 H<sub>1</sub>도 있을 수 있고 H<sub>3</sub>도 있을 수 있지만, 이 둘은 최대 마진 하이퍼플레인이 아니다. 서포트 벡터 머신 분류기의 목표는 분류를 할 때 최소한의 오류, 최대한의 분류 거리를 둘 수 있는 최대 마진 하이퍼플레인을 구하는 것이고, 이렇게 구해진 최대 마진 하이퍼플레인을 이용하여 새로 입력되는 점들의 위치를 판단하여 그 점이 어떠한 그룹에 속하는지를 판단하고 그 결과를 통해 점들을 분류한다. (그림 11)의 경우 새로 입력된 점이 최대 마진 하이퍼플레인 H<sub>2</sub>의 상단에 위치하게 되면 그 점은 검은 점 그룹으로 분류되고 입력된 점이 H<sub>2</sub>의 하단에 위치하게 되면 그 점은 흰 점 그룹으로 분류된다. 실제 그 점의 속성과 분류된 속성이 일치하는지 아닌지를 판단하여 오분류 행렬(confusion matrix)을 만들며 정확하게 분류된 점들이 전체 점들의 몇 %인지를 결과로 기술하는 것이 일반적이다.

서포트 벡터 머신 분류기는 수학적으로 표현할 수 있다. 주어진 학습 세트에 레이블을 붙여서 한 쌍의 벡터 (x<sub>i</sub>, y<sub>i</sub>), i = 1, ..., l where x<sub>i</sub> ∈ R<sup>n</sup> and y<sub>i</sub> ∈ {1, -1}로 표현한다. 이 때 서포트 벡터 머신 분류기는 다음과 같은 최적화 문제



(그림 11) 서포트 벡터 머신 분류기를 이용한 서로 다른 두 그룹의 분류 예

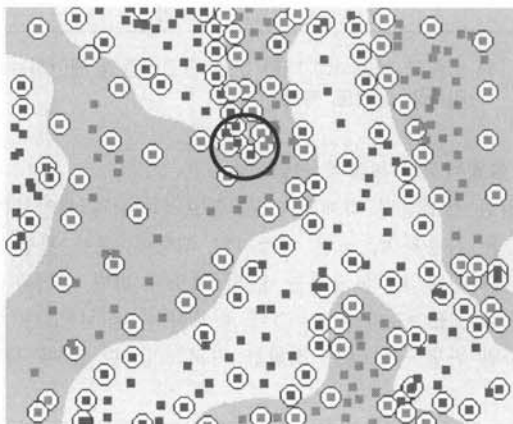
의 답을 구한다[10].

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0. \end{aligned}$$

학습 세트 벡터  $x_i$ 는 함수  $\phi$ 에 의하여 자신의 차원보다 한 차원 높은, 혹은 무한대 차원의 공간에 놓인다. 상기 식에 대한 답을 통하여 서포트 벡터 머신 분류기는 최대 마진 하이퍼플레인을 찾는다. 여기서, 오류를 최소로 하려면  $C > 0$  이어야 한다. 또한 위 문제에서  $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ 는 커널 함수라고 정의되는데, 일반적으로 커널 함수는 다음과 같이 Linear, Polynomial, Radial basis function, Sigmoid의 네 가지 종류가 있다. 여기서  $\gamma, r, d$ 는 각 커널에서의 파라미터를 의미한다.

- Linear:  $K(x_i, x_j) = x_i^T x_j$
- Polynomial:  $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- Radial basis function(RBF):  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
- Sigmoid:  $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

본 논문에서 분류해야 하는 공간은 입력되는 출력물을 인쇄한 프린터의 수로서  $N$ 차원이므로, 다중 클래스 분류를 할 수 있는 LIBSVM 서포트 벡터 머신 분류기를 사용하였다 [10, 11]. LIBSVM 서포트 벡터 머신 분류기의 커널로 RBF(Radial Basis Function)을 사용하였다. LIBSVM 서포트 벡터 머신 분류기는 unbalanced 클래스에 대한 penalty 부가 기능을 가지고 있어서, 임의의 데이터가 들어왔을 때 어느 클래스에 속하는지 정확히 분류할 수 있다. (그림 12)은 두 종류의 특징점들에 대하여 RBF 커널을 적용한 서포트 벡터 머신 분류기에 의해 분류를 수행한 후에, 이를 통



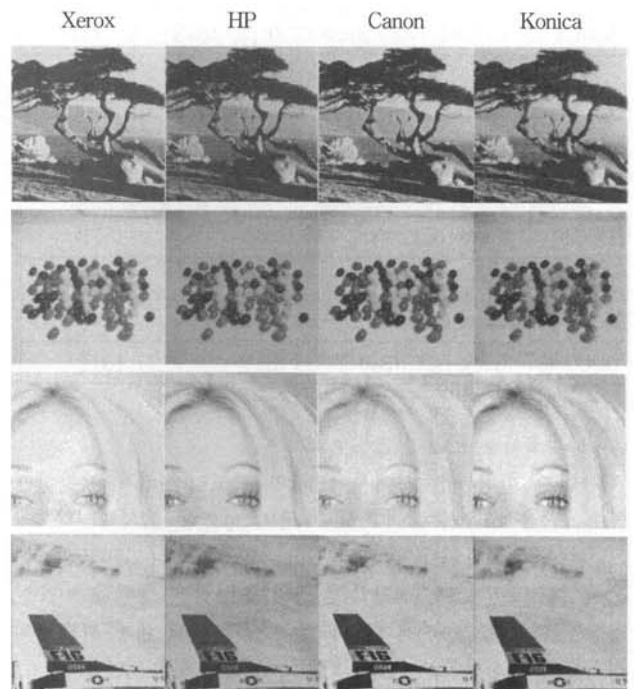
(그림 12) SVM 분류기에서 RBF커널로 분류한 벡터들의 집합 예

해 분류된 특징점 및 하이퍼플레인의 모습을 나타낸다. 원으로 표시한 영역에 포함되어 있는 어두운 점들은 분류 오류에 해당된다.

#### 4. 실험 결과

제안한 알고리즘의 성능을 분석하기 위해 4개 제조사 (Xerox, HP, Canon, Konica)에서 생산된 총 7 대의 프린터를 사용하였다. 또한 관련 연구에서 제시한 Choi의 방법과 비교 분석을 수행하였다[4]. 이미지 데이터를 취득하기 위하여 371장의 이미지를 각 컬러 레이저 프린터마다 인쇄하였다. 그 다음 스캐너로 스캔하여 총 2597(371×7대)장의 이미지를 디지털화하였다. 실험에 사용된 스캐너는 EPSON Perfection 2400이고, 300 dpi의 해상도로 스캔하였다. 이미지의 크기는 256×256 크기로 맞추어 실험하였다. 실험에 이용한 이미지 세트 중 일부는 (그림 13)과 같다.

앞서 설명한 것과 같이 제안하는 알고리즘 및 Choi의 알고리즘은 서포트 벡터 머신 분류기를 사용하므로, 훈련 과정과 판별 과정으로 구성되어진다. 성능 판별의 객관성 및 정확도를 위하여 훈련 이미지와 판별 이미지는 무작위로 선택하였다. 총 371장의 영상을 이용하였으므로, 훈련 이미지는 각 프린터 당 186장씩 사용하였고, 판별 이미지는 185장을 이용하였다. 본 논문에서 사용한 LIBSVM 분류기는 RBF커널을 사용하여 최대마진 하이퍼플레인을 계산하는데, 파라미터로  $c$ 와  $r$ 을 8과 0.5로 설정하였다[11]. 또한, 실험의 오류를 최소화하기 위해 Cross-validation 이라는 기법을 사용하여 훈련 데이터와 판별 데이터를 바꾸어 판별 데이터를



(그림 13) 실험에 쓰인 영상의 예

훈련 데이터로 쓰고 훈련 데이터를 판별 데이터로 사용하여 실험을 하였다. 이와 같이 2가지 실험에 대한 평균값을 최종 결과로 계산하였다. 스캔한 이미지는 RGB영역으로 표현된다. 이 원본이미지를 HSV 영역으로 변환하여 이산 웨이블릿 변환을 수행한 후 HH 영역을 얻는다. 좀 더 정확한 결과 값을 도출하기 위해 HH영역을 다시 명암 동시발생 행렬로 변환하여 최종적으로 특징값을 구하는데 사용하였다.

실험은 프린터 제조사 판별과 동일 제조사의 제품 판별의 2가지 분석을 수행하였다. 프린터 제조사 판별은 다른 제조사의 제품들을 얼마나 잘 검출하는지 분석하는 것이 목적이며, 동일 제조사의 제품 판별은 동일 제조사의 제품들의 결과를 얼마나 잘 구분하는지 분석하는 것이 목적이다.

프린터 제조사의 판별을 위해서 사용된 컬러 레이저프린터의 제조사는 <표 2>에 정리되어 있다. 각 제조사 프린터로 취득한 371장의 영상에 대하여 임의로 186장을 선택하여 학습기반 분류기의 훈련에 사용하고, 훈련에 사용되지 않은 185장을 선택하여 프린터 판별 성능을 분석하였다.

동일 제조사의 제품 판별을 위해서 사용된 컬러 레이저프린터의 제품은 <표 3>에 정리되어 있다. Xerox의 각 프린터로 취득한 371장의 영상에 대하여 임의로 186장을 선택하여 학습기반 분류기의 훈련에 사용하고, 훈련에 사용되지 않은 185장을 선택하여 프린터 판별 성능을 분석하였다.

<표 2> 프린터 제조사 판별 실험에 사용한 컬러 레이저프린터

Label	Printer Model	훈련 이미지 수	판별 이미지 수
Xerox	Xerox DCC 400	186장	185장
HP	HP 4650	186장	185장
Canon	Canon iR C2620	186장	185장
Konica	Konica Minolta bizhub C250	186장	185장

<표 3> 동일 제조사의 제품 판별 실험에 사용한 컬러 레이저프린터

Label	Printer Model	훈련 이미지 수	판별 이미지 수
DCC400	Xerox DCC 400	186장	185장
DCC450	Xerox DCC 450	186장	185장
DCC5560	Xerox DCC 5560	186장	185장
DCC6540	Xerox DCC 6540	186장	185장

4.1 제조사별 판별 성능분석

동일한 실험환경 하에서 Choi의 알고리즘[4]과 제안하는 알고리즘을 사용하여 제조사별 판별 성능을 분석하였다. 이산 웨이블릿 변환을 이용한 Choi 알고리즘의 결과는 <표 4>에 제시하였으며, 제안하는 명암 동시발생 행렬을 이용한 실험 결과는 <표 5>에 나타나있다. 제조사에 대한 판별율의 정확도에 대한 비교는 (그림 14)에 도시되어 있다.

실험 결과, 제안하는 방법의 프린터별 사이의 분류 결과

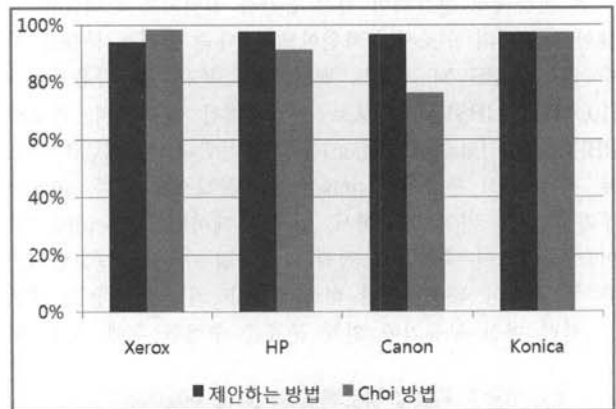
는 96.9%로 이산 웨이블릿 변환을 이용하는 Choi 알고리즘의 분류 결과인 90.82% 보다 6% 정도 높게 나왔다. 또한 Choi 알고리즘보다 제시하는 방법이 HP 프린터를 더 안정적으로 구별함으로써 4가지를 평균적으로 다 높게 판별함을 볼 수 있다. 따라서, 이미지가 프린터에 의해 출력되는 방법이 프린터 제조회사별로 조금씩 다른데, 제안하는 방법이 이미지가 출력될 때 생기는 컬러 노이즈 특성을 올바르게 식별할 수 있음을 확인할 수 있다.

<표 4> Choi 알고리즘의 컬러 레이저프린터 제조사 판별 결과

	Xerox	HP	Canon	Konica
Xerox	98.3%	0.7%	0.5%	0.6%
HP	7.0%	91.4%	1.6%	0%
Canon	15.6%	1.1%	76.3%	7.0%
Konica	2.2%	0.2%	0.2%	97.3%

<표 5> 제안한 방법의 컬러 레이저프린터 제조사 판별 결과

	Xerox	HP	Canon	Konica
Xerox	94%	1.3%	3.7%	0.8%
HP	0%	97.5%	1.8%	0.5%
Canon	1.8%	0.8%	96.7%	0.5%
Konica	0%	2.4%	0.26%	97.3%



(그림 14) 프린터 제조사에 대한 Choi 알고리즘과 제안하는 알고리즘의 정확도 분석

4.2 동일제조사 제품의 판별성능

동일한 실험환경 하에서 Choi 알고리즘[4]과 제안하는 알고리즘을 사용하여, 동일 제조사의 제품에 대한 판별 성능을 분석하였다. 실험 결과는 4.1의 제조사 판별 실험과 같은 방법으로 <표 6>과 <표 7>에 비교하여 나타내었다. 같은 제조사별 판별율에 대한 정확한 비교는 (그림 15)에 나타나 있다.

Choi 알고리즘의 경우 평균 67.93%의 정확도를 갖고 있었으나 제안하는 알고리즘은 평균 79.23%의 정확도를 나타



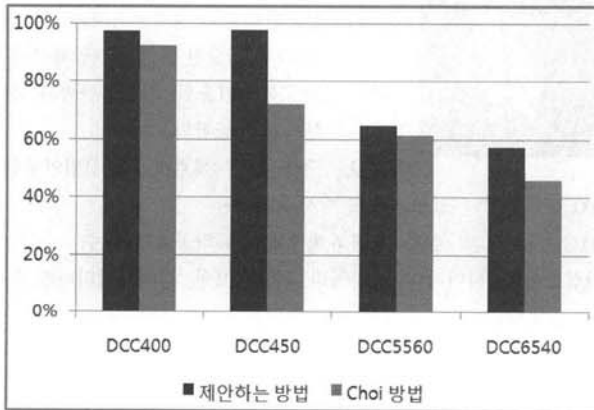
내었다. 이와 같은 결과를 통하여 제안하는 알고리즘이 동일 제조사의 프린터일지라도 영상에 존재하는 특징을 좀 더 정확히 구분할 수 있음을 확인할 수 있다. DCC5560과 DCC6540에 대해서 제안한 방법이나 Choi 알고리즘 모두 낮은 판별율을 보이고 있다. 이는 DCC5560과 DCC6540의 인쇄 프로세스가 유사하기 때문으로 추정된다. 이들을 좀 더 잘 판별하기 위해서는 프린터 인쇄과정의 미세한 차이를 구분할 수 있는 최적화된 특징점이 필요하며, 이에 대한 지속적인 연구가 필요하다.

〈표 6〉 Choi 알고리즘의 동일제조사 컬러 레이저프린터 판별 결과

	DCC400	DCC450	DCC5560	DCC6540
DCC400	92.2%	1%	3.2%	3.5%
DCC450	0.8%	72%	14.6%	12.7%
DCC5560	6%	4.6%	61.5%	28%
DCC6540	7%	11.3%	35.6%	46%

〈표 7〉 제안한 방법의 동일제조사 컬러 레이저프린터 판별 결과

	DCC400	DCC450	DCC5560	DCC6540
DCC400	97.3%	2.7%	0%	0%
DCC450	1.9%	97.6%	0.3%	0.3%
DCC5560	0%	0.8%	64.6%	34.8%
DCC6540	0%	0.3%	42.3%	57.4%



(그림 15) 동일 제조사에 대한 Choi 알고리즘과 제안하는 알고리즘의 정확도 분석

### 5. 결론 및 향후 과제

디지털 기기의 급격한 발전과 각 가정으로의 보급은 포렌식 기술의 중요성을 증대시키고 있다. 그 중 컬러 레이저프린터로 인쇄된 이미지와 여러 문서에 대한 저작권 문제가 대두되고 있다. 그래서 본 논문에서는 위의 문제를 해결하기 위해 출력된 문서에 대하여 어떠한 프린터에서 출력 되

었는지 판별하는 기술을 제안하였다. 프린터를 식별하기 위해 제안하는 기술로는 영상 이미지에 대해 RGB영역을 HSV 영역으로 변환시킨 다음 이산웨이블릿변환을 수행하여 HH영역을 구하였다. HH영역에 대하여 다시 명암 동시발생 행렬을 구하여 특징값을 분석하였다. 프린터 제조회사 별로서의 기술을 동일하게 적용하여 특징점을 추출하였다. 그 특징점 들을 프린터 제조회사 별로 벡터로 분류하여 서포트 벡터 머신 분류기를 통해 훈련과 판별을 거쳐 인쇄물을 출력하는데 사용된 컬러 레이저프린터를 판별하였다. 프린터 제조회사 별로 분류 결과는 96.9%로 신뢰성 있는 결과를 얻었다. 그러나 조금씩 생기는 오차는 프린터 제조회사별로 유사하게 출력되는 것도 알 수 있었다.

앞으로의 향후 연구 방향은 프린터 제조회사 별로 정확히 구별할 수 있는 더 좋은 특징점을 추출하는 것이고, 더 많은 프린터 제조회사에 대해 연구 결과를 적용시켜볼 뿐만 아니라 더 다양한 이미지에도 적용시켜 실험을 진행할 것이다.

### 참 고 문 헌

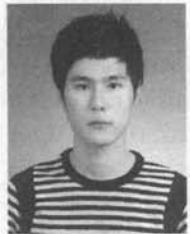
- [1] Wikipedia, "Printer steganography," [http://en.wikipedia.org/wiki/Printer\\_steganography](http://en.wikipedia.org/wiki/Printer_steganography).
- [2] Electronic Frontier Foundation, "DocuColor Tracking Dot Decoding Guide," <http://w2.eff.org/Privacy/printers/docucolor/>
- [3] A. K. Mikkilineni, G. N. Ali, P.-J. Chiang, G. T.-C. Chiu, J. P. Allebach and E. J. Delp, "Signature-embedding in printed documents for security and forensic applications," Proc. of the SPIE Int. Conf. on Security, Steganography and Watermarking of Multimedia Contents, Vol.5306, pp.455-466, 2004.
- [4] J.-H. Choi, "Color laser printer identification using color noise features," MCS 09030, MS thesis, Korea Advanced Institute of Science and Technology, 2009.
- [5] S.-J. Ryu, H.-Y. Lee, D.-H. Im, J.-H. Choi, and H.-K. Lee, "Electro photographic Printer Identification by Halftone Texture Analysis," Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2010.
- [6] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. on Systems, Man, Cybernetics, Vol.3, No.6, pp.610-621, 1973.
- [7] A. K. Mikkilineni, P.-J. Chiang, G. N. Ali, G. T.-C. Chiu, J. P. Allebach and E. J. Delp, "Printer identification based on graylevel co-occurrence features for security and forensic applications," Proc. of the SPIE Int. Conf. on Security, Steganography and Watermarking of Multimedia Contents, Vol.5681, pp.430-440, 2005.
- [8] S.-J. Ryu, H.-Y. Lee, I.-W. Cho, and H.-K. Lee, "Document Forgery Detection with SVM Classifier and Image Quality Measure," Lecture Notes in Computer Science, Vol.5353, pp. 486-495, 2008.

- [9] Wikipedia, "Support Vector Machine", [http://en.wikipedia.org/wiki/Support\\_vector\\_machine](http://en.wikipedia.org/wiki/Support_vector_machine)
- [10] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, 2001.
- [11] C.-W. Hsu, C.-C. Chang, and C.-J Lin, "A practical guide to support vector classification," <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>s



### 백지연

e-mail : jiyoun1128@naver.com  
 2007년~현 재 국립금오공과대학교 컴퓨터공학부 재학 중  
 관심분야: 멀티미디어, 디지털포렌식



### 이흥수

e-mail : lhslife@kumoh.ac.kr  
 2004년~현 재 국립금오공과대학교 컴퓨터공학부 재학 중  
 2009년~현 재 삼성전자소프트웨어멤버십 회원  
 관심분야: 멀티미디어, 디지털포렌식



### 공승규

e-mail : kkongs86@gmail.co.kr  
 2005년~현 재 국립금오공과대학교 컴퓨터공학부 재학 중  
 2009년~현 재 삼성전자소프트웨어멤버십 회원  
 관심분야 : 멀티미디어, 디지털포렌식



### 최정호

e-mail : jhchoi@mmc.kaist.ac.kr  
 2007년 한국과학기술원 전산학과(학사)  
 2009년 한국과학기술원 전산학과(공학석사)  
 관심분야: 멀티미디어, 영상처리, 콘텐츠 보안, 디지털포렌식 등



### 양연모

e-mail : yangym@kumoh.ac.kr  
 1990년 한국과학기술원 전기전자공학과(학사)  
 2006년 광주과학기술원 정보기전공학부(공학박사)  
 2005년 노스캐롤라이나 주립대학 박사후연구원  
 2006년~2008년 대구경북과학기술원  
 2008년~현 재 국립금오공과대학교 전자공학부 전자통신  
 관심분야: WSNs, PONs, URC, 위치인식기술 융합기술



### 이해연

e-mail : haeyeoun.lee@kumoh.ac.kr  
 1997년 성균관대학교 정보공학과(학사)  
 1999년 한국과학기술원 전산학과(공학석사)  
 2006년 한국과학기술원 전자전산학과 전산학전공(공학박사)  
 2001년~2006년 (주)세트랙아이 선임연구원  
 2006년~2007년 코넬대학교 박사후연구원  
 2008년~현 재 국립금오공과대학교 컴퓨터공학부 교수  
 관심분야: 멀티미디어, 영상처리, 콘텐츠보안, 디지털워터마킹 등