

<http://dx.doi.org/10.7236/IIBC.2015.15.3.35>

IIBC 2015-3-5

비정상 북스캔 만화 저작물 식별방법에 관한 연구

A Study on Identification System for Abnormally Scanned Cartoon Books

김태현*, 윤희돈*, 강호갑*, 조성환**

Tae-Hyun Kim*, Hee-Don Yoon*, Ho-Gab Kang*, Seong-Hwan Cho**

요약 디지털 음악이나 영상을 식별하기 위한 핑거프린팅 기술과 사진식별 기술로는 북스캔 만화를 식별하기에 성능이 너무 떨어진다는 문제를 해결하고자 한국 저작권위원회에서는 2013년 “북스캔 어문/만화 저작물 식별 및 복제방지 기술” 개발과제를 통해 북스캔 만화 저작물 식별에 대한 기본적인 솔루션을 확보하였다. 1차년도(2013) 과제의 결과물은 정상적인 스캔 만화에 대한 식별 기술을 중심으로 이루어졌고, 2차년도(2014)에는 변형이나 왜곡이 발생된 비정상적인 북스캔 만화와 다양한 공격을 통해 식별 시스템의 기능을 무력화 하려는 시도에 대한 해결방안에 대한 연구가 이루어지고 있다. 본 논문은 비정상적스캔 및 변형공격에 대한 인식률을 높이기 위한 방안으로 비정상 북스캔 만화 저작물 식별 시스템에 대한 모델을 제시하고자 한다.

Abstract In order to solve the problems that fingerprinting technologies and image identification technologies for digital music and videos identification scanned have not yet been enough to identify scanned cartoon books, the copyright project granted from the Korea Copyright Commission(Project title: Identification and Copy Protection Technology of Scanned Text/Comic Books) secured basic solutions for scanned cartoon books. The first result of the research project in 2013 was focused on the identification technology applied to abnormally scanned cartoon books. As subsequent research for the project, studies in the year of 2014 has been to provide solutions to identify cartoon books which abnormally scanned and occurred with transformations and distortions and solutions to protect the system from trials to devitalize functions of the identification system through various attacks. This paper is to propose a model for a system for identifying abnormally scanned cartoon books as a way of enhancing the identification rates for identifying abnormal bookscans and transformation attacks.

Key Words : Abnormal Bookscan, Cartoon book, Copy protection

1. 서 론

PWC 2013년 통계전망 자료에 따르면 2017년을 기준으로 디지털만화 시장 규모는 19억 달러로 인쇄 만화 시장의 28%를 차지할 것으로 예상하고 있다. 이와는 별도로

로 이 보고서는 전자책과 종이책에 대한 출판시장 규모가 역전되는 시기를 2017년으로 예측하고 있다^[1]. 디지털 만화는 최근 5년간 연평균 180%이상의 성장을 보이면서 초고속 성장을 보이고 있기 때문에 이 부분에서도 조만간 종이만화책 시장 규모를 넘어설 가능성이 높아 보인다

*정회원, (주)DRM Inside

**정회원, 금강대학교 (교신저자)

접수일자 : 2015년 2월 4일, 수정완료 : 2015년 4월 26일

게재확정일자 : 2015년 6월 12일

Received: 4 February, 2015 / Revised: 26 April, 2015 /

Accepted: 12 June, 2015

**Corresponding Author: shcho@ggu.ac.kr

School of Liberal Arts & Science, Geumgang University, Korea

다^[2].

그러나 이러한 전자도서 시장의 도래는 출판사나 솔루션 제공자에게 유통 단가를 낮추주는 효과가 있고 사용자에게도 이동의 간편함을 제공하기는 하지만, 저작권자들의 불법복제에 대한 우려는 여전히 해소가 되지 않고 있다. 더욱이 고품질의 고속 칼라 복사기의 사용이 용이해진 최근에는 스캔된 도서들이 웹하드나 토렌트와 같은 P2P 사이트를 통해 불법적으로 유통되고 있어 전자책 도서 시장에서도 과거 2000년 초반의 MP3 시장과 같은 무차별적인 불법복제 콘텐츠들의 난립현상 가능성도 제기되고 있다^[3].

이러한 상황에서 디지털 음악이나 영상의 불법복제 방지를 위한 핑거프린팅 기술은 디지털 만화에 적용할 수 없고 기존에 연구된 단위 사진식별 기술로는 북스캔 만화를 식별하기에 성능이 너무 떨어진다는 문제를 해결하고자 한국 저작권위원회에서는 2013년 “북스캔 어문/만화 저작물 식별 및 복제방지 기술” 개발과제를 통해 북스캔 만화 저작물 식별에 대한 기본적인 솔루션을 확보하였다. 1차년도(2013) 과제의 결과물은 정상적인 스캔 만화에 대한 식별 기술을 중심으로 이루어졌고, 2차년도(2014)에는 변형이나 왜곡이 발생된 비정상적인 북스캔 만화와 다양한 공격을 통해 식별 시스템의 기능을 무력화 하려는 시도에 대한 해결방안에 대한 연구가 이루어지고 있다^[4].

본 논문에서는 동일과제의 2차년도 연구과제인 비정상적 스캔 및 변형공격에 대한 인식률을 높이기 위한 방안으로 1차년도 과제의 결과 시스템에 비해 개선된 비정상 북스캔 만화 저작물 식별 시스템에 대한 모델제시와 구현 결과 및 성능에 대한 내용을 다룬다^[5].

II. 비정상 북스캔 만화 저작물 식별을 위한 성능 요구사항

비정상 북스캔 만화 저작물 식별 시스템은 정상 북스캔 만화 식별 기능에 추가적으로 식별 대상이 되는 만화 이미지에 변형(회전, 이동, 왜곡 현상 등)이 발생했거나 빛의 밝기가 다르거나 페이지가 추가/삭제된 만화 저작물을 식별할 수 있도록 기능을 향상시킨 시스템으로 성능 요구사항은 다음과 같다.

1. 고성능 식별 처리

비정상 북스캔 시스템의 성능은 식별률과 식별시간으로 평가된다. 식별률은 변형유무, 조판방식 및 페이지 추가/삭제 공격에 대한 식별 기능이 추가되는 관계로 정상 이미지를 대상으로 하는 시스템보다 상대적으로 낮은 90% 이상의 성능을 요구하지만, 식별시간은 동일한 조건 하에서 3초/권 이상의 성능을 유지해야 바람직하다.

표 1. 비정상 북스캔 만화 식별 시스템 성능 요구사항
Table 1. Performance requirements of the abnormally scanned cartoon books Identification System

항목	처리 요구 내용	비고
고성능 식별 처리	식별시간 : 3초/권 이내	공통
	식별률 : 90% 이상	
다양한 저장 및 압축포맷 처리 ^[6]	ALZ, ARJ, EGG, RAR, ZIP, PDF	비정상 요구사항 (2차년도)
다양한 변형 형태 처리 ^[3]	회전, 이동, 왜곡, 밝기 처리	
다양한 조판방식 만화 처리 ^[3]	양면/단면, 일본만화, 한국만화 조판 방식 처리	
공격	페이지 추가/삭제 공격 처리	

2. 다양한 저장 및 압축포맷 처리

비정상 북스캔 만화 저작물 식별시스템은 북스캔된 만화들이 저장 형태에 따라 다양한 압축(ALZ, ARJ, EGG, RAR, ZIP)형태와 PDF 파일 포맷으로 구성될 수 있어, 소스 파일 처리과정을 통해 파일의 포맷에 관계없이 처리해 주는 기능이 요구된다. 비정상 북스캔 식별 시스템은 이러한 다양한 저장 및 압축포맷이라 하더라도 하나의 특징점 DB를 통해 신속하게 식별할 수 있어야 한다.

3. 다양한 변형 형태 처리^[3]

비정상 북스캔 만화 저작물 식별시스템은 북스캔된 만화들이 스캔되는 장비, 방식, 빛의 밝기 형태에 따라 다양한 모습의 이미지 결과물들이 나올 수 있음에 따라 전 처리 과정을 통해 보정해 주는 기능이 요구된다. 변형된 형태의 만화 이미지들은 기준이 되는 특징점 DB 이미지와 비교하여 회전이나 뒤집힘 현상이 있을 수 있고 색의 농도 차이가 있을 수 있으며 상하좌우 공백의 크기도 다를 수 있고, 스캔된 시점의 빛의 정도에 따라 결과물의 밝기도 서로 다를 수 있다. 비정상 북스캔 식별 시스템은 이러한 다양한 변형 형태의 이미지라 하더라도 하나의 특징점 DB를 통해 식별이 가능해야 한다.

4. 다양한 조판방식 만화 처리

비정상 북스캔 만화 저작물 식별시스템은 북스캔 만화들이 조판방식에 따라 만화책의 앞뒤와 좌우의 순서가 반대인 경우가 있을 수 있고 단면 또는 양면으로 구성된 형태, 그리고 양, 단면이 혼합된 형태로 존재할 수 있기 때문에 이들에 대한 처리가 요구된다. 비정상 북스캔 식별 시스템은 이들 양단면에 관련된 상이한 조판 방식도 단일 특징점 DB를 통해 식별할 수 있도록 지원해야 한다.

5. 페이지 추가/삭제 공격 처리

비정상 북스캔 만화 저작물 식별시스템은 식별시스템의 식별기능을 회피하거나 무력화하기 위해 만화의 일부 페이지를 삭제하거나 또는 추가한 만화들에 대한 처리가 요구된다. 비정상 북스캔 식별 시스템은 시스템의 성능을 저하시키지 않는 범위 안에서 추가 또는 삭제된 페이지와 무관하게 단일 특징점 DB를 통해 식별 가능할 수 있도록 지원해야 한다.

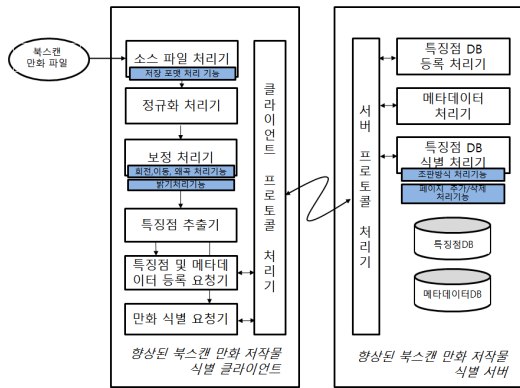


그림 1. 비정상 북스캔 저작물 식별시스템 개념도^[3]
 Fig. 1. Overview of the identification system

III. 비정상 북스캔 만화 저작물 식별 시스템

II 절에서 설명된 비정상 북스캔 만화식별 시스템 성능 요구사항을 만족하기 위해서는 변형된 형태의 만화 처리와 다양한 조판방식 만화 처리 그리고 페이지가 추가 또는 삭제된 만화에 대한 처리가 정상 북스캔 만화식별 시스템에 추가적으로 수행되어야 한다. 정상 북스캔 만화 저작물 식별 시스템에 비정상 처리 기능이 추가된

시스템 구성은 그림 1과 같다.

1. 비정상 식별 시스템 클라이언트

비정상 북스캔 만화 저작물의 식별 시스템 클라이언트는 기존의 정상 시스템의 보정처리기에서 회전, 이동, 왜곡, 밝기 변형에 따른 처리를 수행하는 기능이 추가된다. 다음은 향상된 기능에 대한 설명이다.

가. 저장 포맷 처리기^[3]

저장포맷 처리기기능은 소스 북스캔 만화 파일의 포맷과 종류에 관계없이 실제 만화 식별을 위해 사용되는 부분 이미지 원본만을 메모리로 추출할 수 있도록 내부 구조를 분석하고 유지하는 기능을 제공한다. 그림 2는 ALZ로 압축된 만화 1권과 PDF로 구성된 만화 2권 전체를 RAR로 다시 압축해서 넣은 북스캔 만화 소스 파일을 비정상 북스캔 만화 저작물의 식별 시스템 클라이언트의 저장포맷 처리기가 구조 분석 내용을 보여준다.

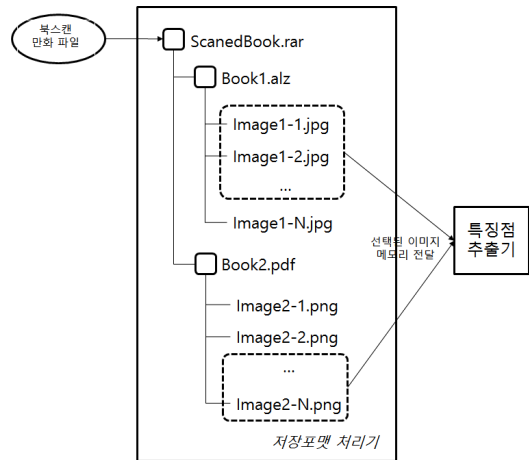


그림 2. 다양한 저장 포맷 처리기 개념도
 Fig. 2. Overview of preprocessing for various image formats

나. 왜곡, 회전, 이동 보정처리기

정규화 과정을 거친 이미지에 대해 왜곡, 회전, 이동과 같은 변형이 있는 경우 이를 정상적인 이미지로 보정해주는 기능을 수행한다. 이러한 변형에 대한 보정과정은 그림 3과 같이 대상 이미지의 4부분에 대한 모서리를 검출한 이후에 정상적인 모서리 값을 기준으로 stretch 과정을 통해 이루어진다.

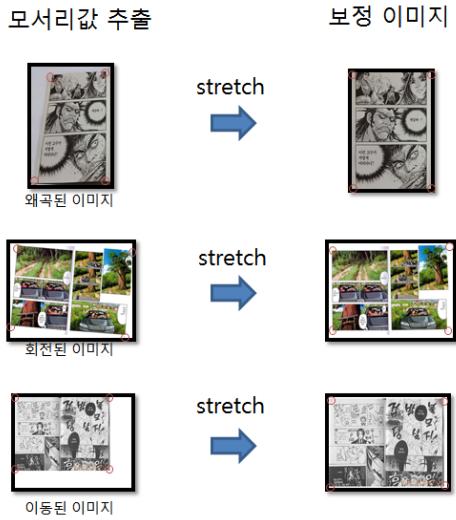


그림 3. 변형 이미지 보정처리 과정
Fig. 3. Overview of preprocessing for transformed images

다. 밝기 보정처리

정규화 과정을 거친 이미지에 대해 복사기를 통한 스캔 시점의 조명 정도나 디지털카메라 촬영시점의 빛의 정도에 따라 발생하는 밝기 변형에 대한 보정을 처리하는 기능을 제공한다. 이러한 변형에 대한 보정과정은 그림 4와 같이 대상 이미지들에 대한 표준 필터 변환 과정을 통해 이루어진다.

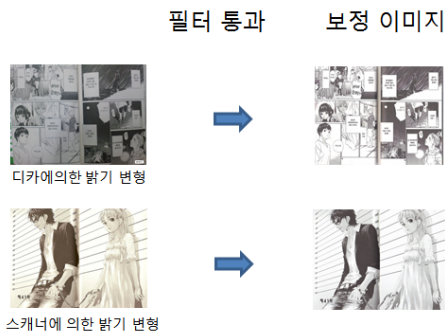


그림 4. 밝기 변형 이미지 보정처리 과정
Fig. 4. Overview of preprocessing for lightness distortion images

라. 그 외 정상 식별 시스템 클라이언트 처리기

보정처리를 통한 변형형태의 이미지 처리기능을 제

외한 나머지 기능처리기들(소스 파일 처리기, 정규화 처리기, 특징점 추출기, 특징점 및 메타데이터 요청기, 만화 식별 요청기, 클라이언트 프로토콜 처리기)은 기존 정상 식별 시스템 클라이언트와 동일하다.

2. 비정상 식별 시스템 서버

비정상 북스캔 만화 저작물의 식별 시스템 서버는 기존 기본 시스템의 특징점 DB 식별 처리기에서 조판방식 처리기능과 페이지 추가/삭제 처리기능이 추가된다. 다음은 향상된 기능에 대한 설명이다.

가. 조판방식 처리기

조판방식은 만화책의 앞뒤와 좌우의 순서가 반대인 경우와 단면 또는 양면으로 구성된 형태에 무관하게 DB 식별처리를 할 수 있는 기능을 수행한다. 조판방식에 대한 처리는 그림 5와 같이 비교 대상 이미지들의 왼쪽 페이지에 대해서만 오른쪽으로 한 페이지 이동 시키면서 비교하는 방법을 통해 검출되도록 한다. 조판방식 처리기는 정상 식별시스템 서버의 특징점 DB 식별 처리기를 통과한 이후에 검출이 안 된 경우에 한해서 수행되는 관계로 기존 DB 검색 시간 보다 평균 2배의 시스템 처리 시간을 요구한다.

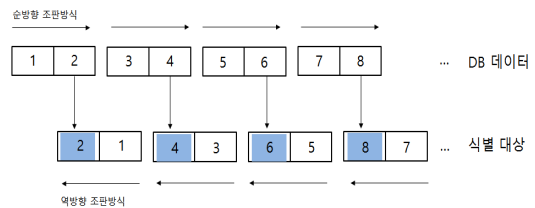


그림 5. 상이한 조판방식에 대한 식별 과정
Fig. 5. Identification process for different reading directions

나. 페이지 추가/삭제 처리기

페이지 추가/삭제는 스캔된 만화의 원본에 특정 페이지를 추가하거나 삭제를 하더라도 DB 식별처리를 할 수 있는 기능을 수행한다. 페이지 추가/삭제에 대한 처리는 그림 6과 같이 비교 대상이 되는 페이지 집합의 크기를 줄이면서 추가되거나 삭제된 페이지에 영향을 받지 않는 페이지 단위를 찾아내는 과정을 통해 검출되도록 한다. 페이지 추가/삭제 처리기는 정상 식별시스템 서버의 특징점 DB 식별 처리기를 통과한 이후에 검출이 안 된 경

우에 한해서 수행되는 관계로 기본적으로 최소 평균 2배의 시스템 처리 시간이 요구된다.

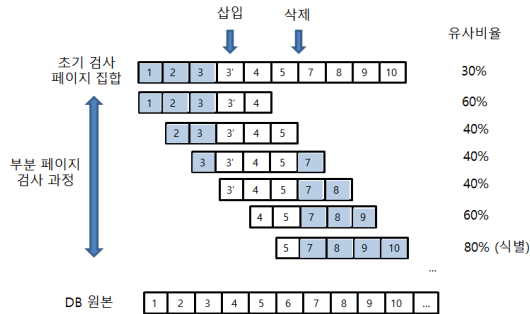


그림 6. 삽입/삭제된 만화에 대한 식별과정
 Fig. 6. Identification process for inserted/deleted cartoons

다. 그 외 정상 식별 시스템 서버 처리기

특징점 DB 식별처리기에서 조판방식과, 페이지 추가/삭제를 위한 예외 처리 기능을 제외한 나머지 기능처리기들(특징점 DB등록 처리기, 메타데이터 처리기, 정상 특징점 DB 식별처리기, 특징점 및 메타데이터 DB)은 기존 정상 식별 시스템 서버 기능을 모두 필요로 한다.

IV. 구현 및 성능

복스캔 만화 저작물 식별 시스템은 식별 요청 및 결과 수신을 네트워크를 통해 검증을 할 수 있도록 C/S 환경으로 구현되었다. 구현된 시스템의 개발 환경은 표 2와 같다.

표 2. 구현 시스템의 개발 환경
 Table 2. Development environment of the implemented system

항목	사용 기술/환경
식별 요청 클라이언트	C++, MFC, WinSock, ZipArchive, Libxml2, Opencv
식별 서버	C++, WinSock, Libxml2
DBMS	PostgreSQLv9.3
OS	Windows 8
개발툴	Visual Studio 2013

1. 비정상 식별 시스템 클라이언트

비정상 식별 시스템 클라이언트는 1차년도 연구개발

결과인 정상 식별 시스템 클라이언트에 보정처리기 기능을 추가하였으며 DB에 구축되어 있는 정상 스캔만화와 검사대상이 되는 비정상 스캔만화(변형된 만화, 상이한 조판방식의 만화, 페이지 삽입/삭제된 만화)를 시각적으로 비교할 수 있는 화면을 추가로 제공한다. 그림 7은 비교화면이 추가된 비정상 식별 시스템 클라이언트 응용프로그램의 구현 결과를 보여준다.

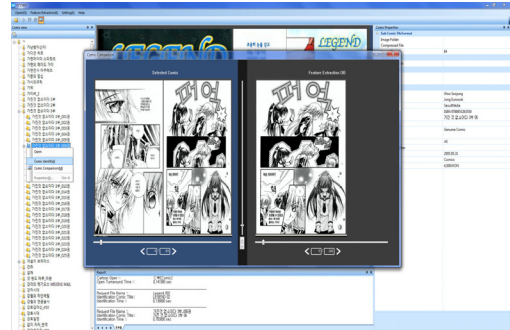


그림 7. 식별 시스템 클라이언트 응용프로그램
 Fig. 7. Client application program of the identification system

2. 비정상 식별 시스템 서버

비정상 식별 시스템 서버는 1차년도 연구개발 결과인 정상 식별 시스템 서버 기능에 조판방식 처리와, 페이지 삽입/삭제된 만화들에 대한 처리를 위한 기능을 추가하였다.

3. 성능

식별 시스템의 식별 성능을 측정하기 위해 웹하드를 통해 유통되고 있는 만화들 중에서 4만여 권을 확보하고, 이 중에서 중복된 만화들을 분류한 후 스캔 상태가 양호한 2만2천여 권을 대상으로 특징점 DB를 구축하였다. 변형된 스캔 만화의 실험대상은 특징점 DB에 구축된 만화를 기준으로 동일한 만화가 스캐너를 이용한 캡처와 디지털카메라를 이용한 캡처 등과 같이 서로 상이한 캡처본이 존재하는 만화들 5,560권을 실험대상으로 선정하였다. 본 논문에서 구현한 시스템의 식별 성능은 표3과 같다. 식별에 성공한 만화는 95.56%로 변형된 만화에 대해서 매우 높은 식별 성공률을 보임을 확인할 수 있으며, 오식별 0.45%, 미식별 3.99%로 나타났다. 평균 특징점 추출시간은 0.3654초, 평균 식별시간은 0.0336초로 만화 식

별에 소요되는 전체 시간은 평균 0.3990초로 측정되었다.

표 3. 식별 시스템 인식률 성능

Table 3. Identification performance of the system

구분	식별성공	오식별	미식별	전체
만화 권수 (비율)	5,313권 (95.56%)	25권 (0.45%)	222권 (3.99%)	5,560권

V. 결론 및 향후 과제

본 논문은 한국저작권위원회의 저작권기술 R&D 연구로 수행중인 “복스캔 어문/만화 저작물 식별 및 복제방지 기술” 연구 과제의 2차년도 연구결과로, 비정상 복스캔 만화 식별 시스템의 구현 모델을 제시하였다. 제안된 모델은 현재 연세대학교 다차원융합 연구실과 공동으로 개발이 이루어지고 있다. 전체 연구는 2013년 5월부터 2015년 2월까지 3년간에 걸쳐 진행되고 있으며, 현재 2차년도 개발에 대한 마무리 작업이 진행되고 있다. 2차년도의 개발은 1차년도의 결과에 다양한 압축 및 저장 포맷, 다양한 변형, 다양한 조판방식, 양면만화 그리고 페이지 삽입/삭제가 이루어진 비정상 형태의 복스캔 만화에 대한 처리기능이 추가되어 구현되어 있다. 향후 3차년도 과제에서는 복스캔 도서에 대해서도 식별이 가능한 기술 개발을 할 예정이다.

현재 진행 중인 연구가 마무리되게 되면 현재 급속도로 증가되고 있는 불법 만화 복제물들에 대한 정확한 식별과 저작권정보의 연동을 통하여 복스캔 만화에 대해서도 저작권을 보호할 수 있는 환경이 구축될 것으로 기대한다.

References

[1] “Content Industry Trends Analysis Report”, Korea Creative Content Agency, 2012

[2] “Big companies leaving e-book business”, Korea Electronic Times, 2013.12.30.
 [3] B. Ahn, T. Kim, H. Kang, S. Cho and K. Lee, “A Study on An Identification System for Scanned Cartoon Book”, The Journal of The Institute of Internet, Broadcasting and Communication, vol. 14, no.1, pp.131-137, 2014
 [4] H. Yoon, T. Kim, H. Kang and S. Cho, “A Study of Performance Evaluation Standard for Scanned Cartoon Book Identification Technology”, The Journal of The Institute of Internet, Broadcasting and Communication, vol. 15, No. 1, pp.29-35, Feb. 28, 2015
 [5] Sang-Hoon Lee, Nakyeon Choi, Sanghoon Lee, “An Identification and Feature Search System for Scanned Comics”, Journal of KIISE Database vol. 41, no 4, pp.199-208, 2014.8
 [6] Wikipedia, List of archive formats, 2013

저자 소개

김 태 현(정회원)



- 1993년 : 중앙대학교 전자계산학과 졸업(학사)
- 2011년 : 성균관대학교 대학원 전기전자 및 컴퓨터공학과(공학석사)
- 1992년 ~ 2000년 : (주)삼성SDS 정보기술연구소
- 2000년 ~ 2004년 : (주)과수닷컴 개발실장

• 2005년 ~ 현재 : ㈜디알엠인사이드 CTO

<주관심분야 : 저작권보호기술(DRM), 정보보안기술, 전자책>

※ 본 논문은 문화체육관광부의 저작권기술개발사업에 의거 한국저작권위원회의 정부지원금을 받아 연구되었습니다.

(This research project was supported by Government Fund from Korea Copyright Commission.)

윤 희 돈(정회원)



- 1996년 : 성균관대학교 전자공학과(학사)
- 2001년 : 성균관대학교 대학원 전자공학과(공학석사)
- 2008년 : 성균관대학교 대학원 전자공학과(공학박사)
- 2009년 ~ 현재 : (주)디알엠인사이드 상무

<주관심분야 : 전자책, 이미지 식별기술, 저작권보호기술>

강 호 갑(정회원)



- 1985년 : 성균관대학교 전자공학과(학사)
- 1988년 : 성균관대학교 대학원 전자공학과(공학석사)
- 2010년 : 성균관대학교 대학원 전자공학과(공학박사)
- 1991년 ~ 2000년 : (주)삼성SDS 정보기술연구소 책임연구원

- 2000년 ~ 2003년 : (주)파수닷컴 연구소장
- 2005년 ~ 현재 : (주)디알엠인사이드 CEO

<주관심분야 : 저작권보호기술, 정보보안기술, 전자책>

조 성 환(정회원)



- 1980년 : 성균관대학교 전자공학과(학사)
- 1982년 : 성균관대학교 대학원 전자공학과(공학석사)
- 1991년 : 성균관대학교 대학원 전자공학과(공학박사)
- 1982년 ~ 1985년 : 해군사관학교 전기 및 전자공학과 전임강사

- 1997년 : 미국 Columbia 대학 CATT Visiting Scholar
- 1985년 ~ 2002년 : 동서울대학 컴퓨터공학과 부교수
- 2002년 ~ 현재 : 금강대학교 교수

<주관심분야 : 영상통신, 무선네트워크, 저작권보호기술(DRM)>