워터마킹을 이용한 PDF문서 보안 시스템 설계 및 구현

손준희, 고성종, 이선호, 이임영
순천향대학교 컴퓨터소프트웨어공학
e-mail:[sjh0109, lunatics, sunho431, imylee]@sch.ac.kr

Design and implementation of PDF document security system using watermarking

Jun-Hee Son, Sung-Jong Go, Sun-Ho Lee, Im-Yeong Lee
Department of Computer Software Engineering, Soonchunhyang University

요 약

인터넷의 급속한 발달로 인해 아날로그 콘텐츠에서 디지털 콘텐츠 서비스로 빠르게 발전하고 있다. 다양한 분야에서 디지털 콘텐츠를 이용하여 오프라인 시스템을 대체하게 되었다. 하지만 디지털 콘텐츠의 특성 때문에 불법 복제 및 무단 유포의 문제점과 저작권 침해의 문제점들이 발생하게 되었다. 따라서 보안 서비스에 대한 요구사항도 급격히 증가하고 있다. 이에 따라 본 논문에서는 워터마크 기술을 기반으로 인터넷 환경에서 사용자가 원하는 워터마킹을 이용하여 불법 복제를 방지함으로써 안전한 문서보안 콘텐츠를 제공하기 위한 시스템을 구현 및 설계 하였다.

1. 서론

인터넷이 발달함에 따라 웹상에서의 정보 및 자료가 많이 만들어 지고 있다. 음악, 영상, 이미지, 문서를 비롯한 많은 정보들이 만들어지는 인터넷이라는 정보의 바다에서 그에 따른 문제점이 발생하고 있다. 인터넷상에서의 자료는 디지털 콘텐츠에 대한 복사물이 원본과 거의 다르지 않다는 점이다. 그에 따라 디지털 콘텐츠에 대한 영상 추출이나, 이미지 및 문서에 대한 복제로 인해 저작권 침해를 비롯하여 그에 따른 금전적 피해 또한 발생하고 있다[1]. 위와 같은 이유로 저작권의 침해를 비롯하여, 디지털 콘텐츠의 유통이 위축되어 수익의 한계를 가지고 콘텐츠 시장의 활성화의 큰 저해 요소가 되고 있다. 이를 해결하기 위해 워터마킹이라는 기술을 사용하여 각종 콘텐츠에 대해 소유권 및 저작권을 인증 하고 있다. 워터마킹은 저작권을 보호하기 위해 눈으로 판독할 수 없는 마크를 삽입하고 필요시 검출을 하여 이에 따른 인증을 할 수 있는 기술이다[2]. 주로 미디어나, 이미지 등에 사용을 하게 되는데 본 연구는 디지털매체나 이미지 대신에 이미지 형식의 문서에 대한 인증을 할 수 있는 워터마킹 기술을 연구하려고 한다. 본 논문은 문서에 대한 신원 확인과 위조 및 변조방지를 요구하기 때문에 데이터의 위, 변조 시도를 막는데 사용되는 강성 워터마킹을 사용하여 문서의 부분적인 영역이 아닌 전체 영역에 워터마킹을 적용한다. 그렇기 때문에 원본의 일부분이 훼손 되더라도 저작권에 대한 인증을 할 수 있다. 또한 워터마킹이 가시적일 경우 이미지를 삭제하여 저작권에 대한 침해를 유발 할 수 있다[3]. 워터마킹이 삽입된 영역에 원본이미지를 커버하여 비가시성을 제공하게 되고 만약 워터마킹이 부분적으로 사용되었을 경우 원본을 포함한 워터마킹의 이미지 자체를 훼손시켜 남은 부분들을 침해 할 수 있다. 그렇기 때문에 문서상에 있는 각각의 텍스트를 검출하여 영역 혹은 블록 단위로 나누어 문서 전체에 있는 텍스트를 부분으로 나눌 수 있는 기술을 사용하여 일부분의 워터마킹이 아닌 문서 전체영역에 마킹을 삽입하게 되면 원본의 일부분과 워터마킹이 훼손되더라도 워터마킹을 검출 할 수 있다. 텍스트상의 영역이나 블록단위의 검출을 하기 위해서 영상처리 기술 중 최소경계사각형 즉 MBR(Minimum Bounding Rectangle)기술을 사용하여 텍스트영역의 좌표 값을 이용하여 특정영역만 변화를 주어 워터마크를 삽입하게 된다[4,5].

본 논문의 구성은 2장에서 관련 연구를 기술하고 3장에서는 워터마크 시스템의 보안요구사항을 기술한다. 4장에서는 이를 만족하는 제안방식에 대하여 기술하고, 5장에서는 제안방식의 구현에 대하여 기술하며 마지막으로 6장에서 결론을 맺는다.

2. 관련 연구

본 장에서는 연구를 위해 필요한 내용을 설명한다. 먼저 워터마크의 전반적인 기술을 설명하고 현재 상용화 되고 있는 기술을 설명한다.

2.1 디지털워터마킹(Digital Watermarking)

디지털 워터마킹(Digital Watermarking) 기법은 디지털 콘텐츠에 워터마크 (Watermark)라고 하는 사용자의 ID(Identification)나 자신만의 정보를 삽입시킴으로써 불법적인 복제를 막고, 지적재산권 및 저작권을 보호하며, 소유권을 주장할 수 있는 근거를 제시할 수 있도록 하는 기술이다.

워터마킹이란 용어는 물에 젖어있는 상태에서 그림을 인쇄하는 데서 유래하였다. 지폐의 제작과정에서 위조지폐 여부를 가리기 위해 젖어있는 상태에서 특정 정보를 삽입하고, 말린 후 인쇄를 하여 불빛에 비춰 보았을 때 그림이 보이도록 하는 기술을 말한다. 또 중세기에는 군사적인 목적의 통신문이나 비밀편지에 특수잉크 또는 약품 등을 사용하여, 받는 쪽에서 특별한 처리를 해야만 볼 수 있도록 하였다. 또한 미술작품이나 책의 저자 또는 저작권을 갖고 있는 사람이 자신의 것이라는 것을 표시하기 위해 특별한 방식으로만 볼 수 있도록 실제 작품에 표시해 두는 기술로도 사용하였다. 이때 삽입되는 저작권이나 소유정보, 원본여부를 확인할 수 있도록 숨겨놓은 데이터, 사용권한을 부여 받은 사용자의 ID등의 식별정보를 워터마크라 한다.

2.2 영상처리기술

입출력이 영상인 모든 형태의 정보 처리를 가리키며, 사진이나 동영상을 처리하는 것이 대표적인 예이다. 대부분의 영상 처리 기법은 화상을 2차원 신호로 보고 여기에 표준적인 신호 처리 기법을 적용하는 방법을 사용한다. 일반적으로 디지털 영상 처리는 다양한 방법으로 쓰일 수 있으며 정확하다는 장점이 있고, 아날로그보다 구현하기 쉽기도 하다. 더 빠른 처리를 위해서 파이프라인과 같은 컴퓨터 기술들이 쓰이기도 한다. 워터마킹을 삽입하는 과정에서 공간/시간 영역과 주파수 영역에 삽입하는 두 가지가 있다

이러한 영상처리 기술을 사용하는 데 있어서 중점적으로 사용하는 기술은 MBR 기술을 사용한다. 크게 세 가지로 분류 할 수 있다

RMBR(Relative MBR) : 상대 좌표 값을 이용한 MBR

- 절대 좌표 값을 이용하여 영역을 검출(16바이트), 기준값을 이용한 상대 좌표계 X,Y(4바이트)

QMBR(Quantification MBR) : 정량화에 의한 MBR

- 정량화의 단계가 낮으면 각 좌표값을 1바이트로 표현 가능 기존 MBR에 대해 4배 압축효과를 가지고 있음

HMBR(Hybrid MBR) : MBR의 크기를 고려한 혼합 표현

- MBR의 좌, 하 점은 상대좌표계처럼 표현하되 우, 상 점을 MBR의 크기로 표현하는 혼합 표현법. 정확도는 적절히 유지하면서 데이터의 양을 줄이는 방법

2.3 PDF(Portable Document Format)기술

미국 어도비시스템즈(Adobe Systems)에서 만든 문서파일 유형으로 윈도우, 맥, 유닉스, 구글 안드로이드 등 거의 모든 운영체제에서 읽거나 인쇄할 수 있으며 원본 문서의 글꼴, 이미지, 그래픽, 문서 형태 등이 그대로 유지되고 온라인 및 오프라인 환경에서도 쉽게 문서를 공유할 수 있으면서도 보안성이 높아 공공기관, 연구소 등에서 자료를 배포할 때 많이 사용 무료 프로그램인 아크로뱃리더로 뷰어가 가능하다.

3. 보안 요구사항

본 장에서는 연구의 대한 보안 요구사항을 알아본다.

3.1 보안 요구사항

본 연구는 다음과 같은 보안 요구사항을 가진다.

- 기밀성 : 사용자는 자신이 원하는 문서에 저작권을 인증하는 워터마킹을 삽입함으로써 워터마크의 정보가 공격자로부터 워터마크의 내용이 노출되지 않도록 기밀성을 제공해야 한다.
- 무결성 : 워터마크 삽입 후 사용자는 워터마킹이 삽입된 문서의 저작권을 가지게 된다. 이 때 문서에 삽입된 워터마크의 정보가 제 3자로부터 악의적 또는 비 악의적 접근에 의해 위/변조가 되지 않았다는 것을 증명할 수 있는 무결성을 제공해야 한다.
- 사용자 인증 : 본 시스템을 사용하고자 하는 사용자나 응용 프로그램의 정보를 확인하여 불법적인 사용자인지 검증할 수 있도록 사용자 인증 기능이 제공되어야 하며, 인증되지 않은 사용자는 워터마크를 인증할 수 없어야 한다.
- 강인성 : 워터마크를 문서 한 부분에 삽입하여 전송이나 저장을 위해 압축할 때 워터마크가 깨지지 않아야 한다. 또한 전송 중에 생길 수 있는 노이즈나 여러 가지 형태의 변형이나 공격에도 추출이 가능해야 한다.
- 명확성 : 추출된 워터마크가 확실한 소유권을 주장할 수 있도록 공격 등에 대해 정확성을 유지해야 한다.
- 원본 없이 추출 : 원본 문서 없이 워터마킹된 문서만으로 워터마크를 검출해야 한다. 이는 워터마킹 기법을 온라인상이나 다양한 응용분야의 적용에 있어, 올바른 소유권자를 구별할 수 있어야 하는 현실성을 고려할 때 반드시 가능해야 한다.
- 보안성 : 관련된 키 값을 알고 있지 않을 경우에 워터마크의 확인이 불가능해야 된다.

4. 제안방식

본 논문에서 제안하는 방식은 문서의 일부분의 영역이 아닌 문서 원본의 텍스트영역을 검출하여 문자에 워터마킹을 삽입함으로써 문서 및 워터마킹의 훼손을 방지하여 저작자의 신원확인 및 문서를 보안하여 유출피해 및 사후 추적관리를 할 수 있는 방식이다.

4.1 시나리오

본 연구에서 제안하는 시나리오는 중점적으로 워터마킹의 삽입과 검출 과정 이를 바탕으로 한 사용자 인증 세 가지 부분으로 나눌 수 있다(그림 1).

Step 1. 사용자가 생성한 문서를 시스템에서 호출한다.

Step 2. 시스템은 사용자를 인증 하게 되는 워터마크를 사용한다.

Step 3. 시스템에서 원본이미지에 워터마킹을 삽입 한 후 커버이미지(원본이미지)를 추가함으로써 워터마킹의 비가시성을 제공한다.

Step 4. 공격자는 비가시적으로 삽입된 워터마킹 문서를 획득한다.

Step 5. 프린터 및 스캐너를 사용하여 위조 및 변조된 문서를 생성한다.

Step 6. 시스템은 위조 및 변조된 문서에서 워터마킹을 검출 하여 원본과 비교 및 인증을 제공한다.

4.2 워터마킹 삽입 과정

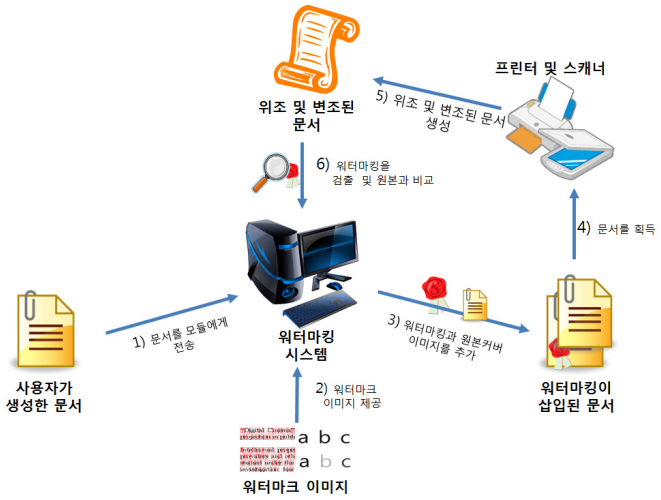
사용자는 회원가입과 사용자인증을 통하여 자신의 워터마크를 생성하게 되고 작성한 문서에 워터마크를 삽입하게 된다. 시나리오는 다음과 같다(그림 2).

Step 1. 사용자는 회원가입을 하게 되면 키 값을 생성하게 되고 DB에 같이 저장이 된다. 로그인이 이루지게 되면 키 값을 불러와 워터마크인 이진수를 생성한다.

Step 2. 사용자는 워터마크를 할 문서를 호출한다.

Step 3. 불러온 문서는 문자영역을 검출하기 위해 이진화를 통하여 비트맵형식의 파일로 변환을 하게 되고 문자영역을 검출한다.

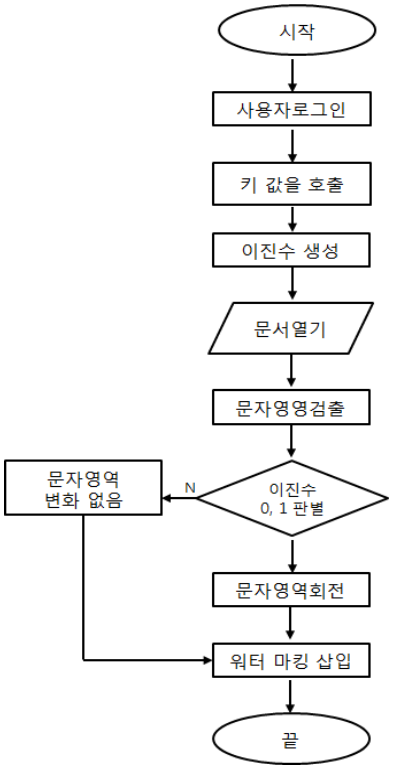
Step 4. 검출된 문자영역과 키 값으로 생성한 이진수를 사용하여 워터마크를 삽입하게 되는데 이진수가 0일 경우 영역의 변화는 없고 이진수가 1일 경우 영역의 회전을 주게 된다.



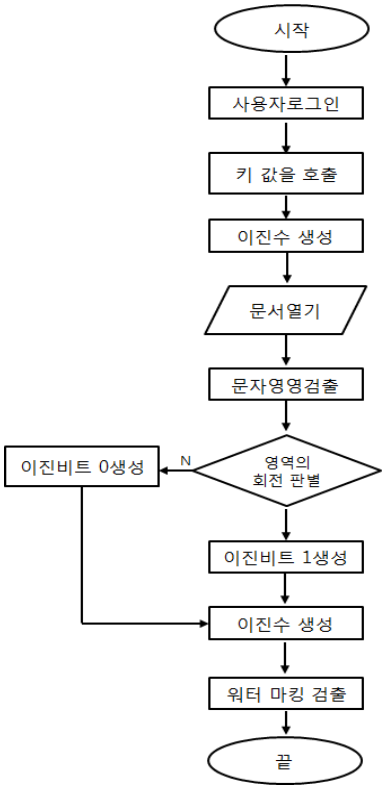
(그림 1) 프로그램 시나리오

4.3 워터마킹 검출 과정

사용자는 로그인을 통하여 자신의 워터마크를 DB로 부터 전송받게 되고 워터마크가 삽입된 문서를 호출하여 검출한다. 이에 따른 시나리오는 다음과 같다(그림 3).



(그림 2) 워터마킹 삽입 시나리오



(그림 3) 워터마킹 검출 시나리오

Step 1. 사용자의 로그인 과정을 통해 워터마크정보인 이진수를 DB로부터 받는다.
Step 2. 사용자는 워터마크가 삽입된 문서를 호출한다.
Step 3. 호출된 문서에서 문자영역을 검출하여 회전여부를 판단한다.
Step 4. 문자영역을 검출하여 영역의 회전여부를 판단하게 되는데 영역의 회전이 이루어 졌으면 1을 반환하게 되고 회전이 이루어지지 않았으면 0을 반환하여 워터마크인 이진수를 생성한다.

4.4 워터마크 검증 과정

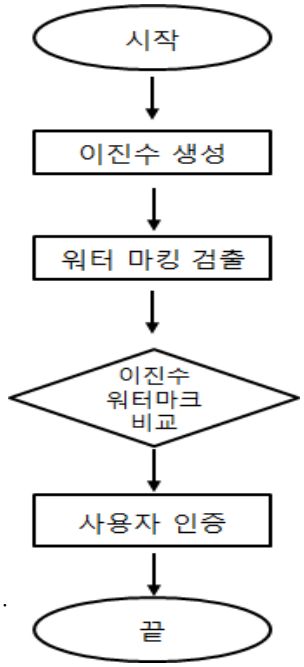
사용자는 로그인을 통하여 자신의 워터마크를 DB로 부터 데이터를 받게 되고 검출된 워터마크와 비교하여 인증을 하게 된다. 시나리오는 다음과 같다(그림 4).
Step 1. 로그인을 통하여 사용자의 워터마크 정보인 이진수를 호출하게 된다.
Step 2. 워터마크 검출을 통하여 생성된 이진수와 사용자의 이진수를 비교한다.
Step 3. 두 이진수를 비교하여 문서에 삽입, 검출된 워터마크로부터 사용자인증과정을 수행하게 된다.

5. 제안방식 구현

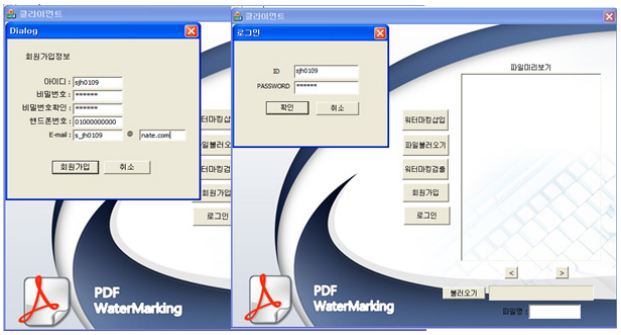
4장에서 설계한 내용을 기반으로 문자영역 검출을 통한 워터마킹 솔루션을 구현하였다.

5.1 클라이언트 구현

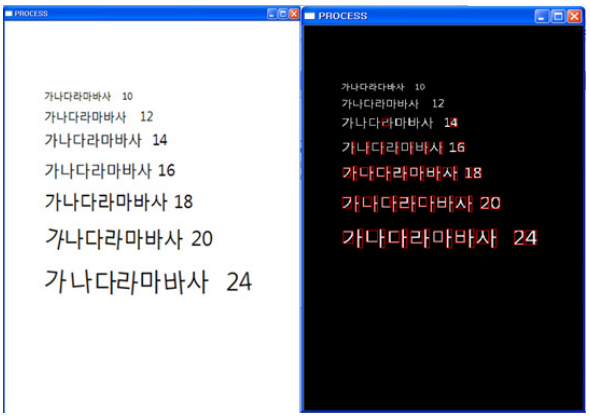
클라이언트는 사용자의 회원가입, 로그인 과정을 수행하고 생성된 키 값을 사용하여 워터마크로 사용될 이진수를 생성하게 된다. 이진수로 생성된 워터마크의 삽입, 검출 과정을 나타내는 시나리오이다(그림 5,6).



(그림 4) 워터마크 검증



(그림 5) 회원가입과 로그인 화면



(그림 6) 이진화를 통한 워터마킹 삽입화면

6. 결론

이미지 상으로 보여 지고 있는 텍스트가 위조 혹은 변조가 되었을 경우에 본 워터마킹기술은 마크가 문서전체에 삽입이 되어 있기 때문에 그러므로 텍스트 기반의 PDF문서를 생성 시 이와 같은 기술을 사용하여 워터마킹을 할 수 있게 되고 웹상에서의 저작권과 무단유출, 복제 시 저작권에 대한 인증과 무결성을 제공하여 개인의 저작권 및 디지털 콘텐츠를 보호 할 수 있다. 추후 다양한 방법으로 워터마킹을 할 수 있는 연구가 진행 될 수 있을 것이며 다양한 매체들의 워터마킹을 연구하여 디지털 콘텐츠 보안에 대해 다양한 연구가 시도 될 것이다.

참고문헌

[1]디지털타임즈 “디지털저작권 관리와 워터마킹”, 2012.01.25
[2] DATENET "씨케이엔비 이북 불법복제 막는다.“, 2011.05.12.
[3] 김진호, 서영호, 이홍규 “워터마크 공격 및 평가 기술 동향”, 학술지 전자통신동향분석 제19권 제4호, 2004. 8
[4] 공영민, 추현곤, 최종욱, 김회율 “이진 문서 영상에서의 특징 기반 텍스트 워터마킹”, 대한전자공학회 2002년도 하계종합학술대회, 2002
[5] 공영민 “문서 이미지에서의 특징 기반 텍스트 워터마킹”, 한양대학교 학위논문(석사), 2002.12