

저작권 보호를 위한 웹 이미지 크롤러

권오중, 임승원, 오민석, 양보원, 정진우*

동국대학교 컴퓨터공학과

e-mail : {dhendzzang, wing9502, omk730, bowonyang, jwjung}@dongguk.edu

*교신저자: jwjung@dongguk.edu

Web Image Crawler for Copyright Protection

Oh-Jung Kwon, Seung-Won Lim, Min-Seok Oh, Bo-Won Yang, Jin-Woo Jung
Dept. of Computer Science and Engineering, Dongguk University

요 약

디지털 콘텐츠가 넘쳐나는 가운데, 최근 이미지 저작권에 대한 중요성이 대두되고 있다. 블로그 및 SNS 에 쉽게 게시물을 업로드가 가능한 점으로 인해, 이미지 저작권의 침해가 광범위하게 이루어지나, 저작권의 소유자는 침해된 자료를 찾기 위해 일일이 검색하고 적발하는 방식을 사용하고 있어 비용이 많이 들며 저작권 침해 속도에 크게 못 미친다. 이러한 문제를 해결하고 궁극적으로 저작권의 권리 보장과 저작권 침해 예방을 위해 본 연구에서는 키워드 검색으로 얻은 게시물에 게재된 이미지를 저작권 침해 적발 방법을 자동화하는 시스템을 제안하고 실험결과를 제시하였다.

1. 서론

디지털 콘텐츠가 넘쳐나는 가운데, 최근 이미지 저작권에 대한 중요성이 대두되고 있다. 사용자들이 블로그 및 SNS 에 게시물을 작성하면서 이미지를 활용하는 경우가 많지만, 이미지의 저작권을 인식하지 못하여 이를 침해하는 일이 종종 발생하고 있다. 이미지 저작권의 침해는 쉽게 이루어지나, 저작권의 소유자는 침해된 자료를 찾기 위해 일일이 검색하고 적발하는 방식을 사용하고 있어 비용이 많이 들며 저작권 침해 속도에 크게 못 미친다. 이러한 문제를 해결하기 위해 본 연구에서는 이미지 저작권 침해 적발 방법을 자동화하여, 궁극적으로 저작권자의 권리를 보장하고 저작권 침해 예방하는 것을 목적으로 하고 있다.

본 연구에서는 저작권법으로 보호되는 사용자의 이미지를 무단 게재한 게시글을 적발하는 시스템을 설계하고 구현하였다. 웹상에서 키워드를 중심으로 크롤한 이미지에 대해 저작권 이미지와 비교 계산을 하여 유사도가 높은 사진을 골라내어 사용자에게 유사율과 함께 제공한다.

2. 관련 연구

1) 저작권 관련 기존 연구

저작권 기술은 저작권 보호기술과 이용활성화 기술의 두 가지를 중심으로 발전하고 있다[1]. 그 가운데 이미지 저작권 보호 기술은 워터마킹, 핑거프린팅 등의 기술을 대표적이다. 해당 기술은 이미지에 저작권 정보를 삽입하여 무단 사용 여부를 판단하거나 불법 유포자를 추적할 수 있다. 하지만 저작물의 유통이 용이해지고 저작권 피해 규모가 대형화되는 추세를

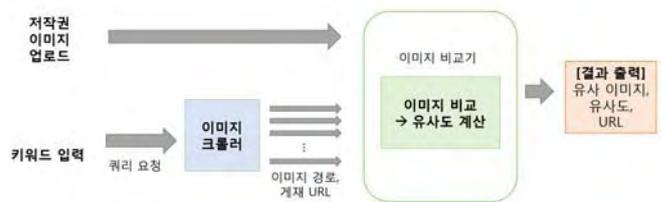
살펴볼 때, 해당 방식으로는 규모를 따라가기 어려운 한계가 있다. 또한 삽입된 정보가 변형에 의해 훼손되었을 경우 불법적으로 사용한 경우임에도 검출이 어렵다.

2) 이미지 비교 알고리즘

이미지 비교는 기존에 많은 연구가 진행되었다. 비교 방식은 기하학적 특징 기반 방법, 템플릿 매칭 방법, 주성분 방법 등 다양한 방식이 존재한다. 특징 기반 방식은 SIFT, SURF, FAST 등이 있으며 색깔, 빛, 변형이 이루어진 경우 인식이 가능하다[2]. 템플릿 매칭 방법은 빠른 연산이 가능하다는 장점이 있어[3] 다량의 이미지 데이터를 비교연산 하는 본 시스템에서 활용하였다.

3. 제안 시스템

2 의 1)에서 제시된 워터마크 혹은 핑거프린트 등의 방식으로는 표준화된 방식의 부재로 대규모로 처리하기 어려울 뿐 아니라 이미지 변형으로 인해 삽입된 저작권 정보의 훼손 시 적발에 한계가 있다. 본 연구에서는 이러한 점을 반영하여, 저작권 정보를 이미지에서 추출하는 것이 아니라, 이미지 비교 프로세스를 통해서 저작권 침해 게시물을 찾을 수 있는 시스템을 제안하였다.



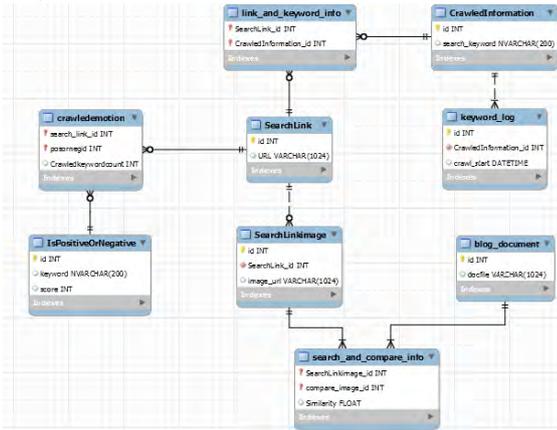
(그림 1) 시스템 구조도

*교신저자: jwjung@dongguk.edu

1) 시스템 구조도

그림 1 은 시스템 구조도를 표현한 도식이다. 사용자 인터페이스에서 사용자는 저작권 이미지를 업로드한 후, 이미지와 관련된 키워드를 입력한다. 입력 받은 키워드를 바탕으로 관련 게시물을 검색하고 게시물 내의 이미지를 크롤한다. 이미지 비교기로 크롤링한 다수의 이미지와 사용자가 업로드한 저작권 이미지를 비교하여 각각 유사도를 계산한다. 최종적으로 사용자에게 유사도가 높은 순서대로 이미지와 해당 이미지가 게재된 URL 을 출력한다. 이미지 비교 프로세싱은 C++을 활용한 OpenCV 를 사용하여 구현되었다.

2) 데이터베이스 스키마



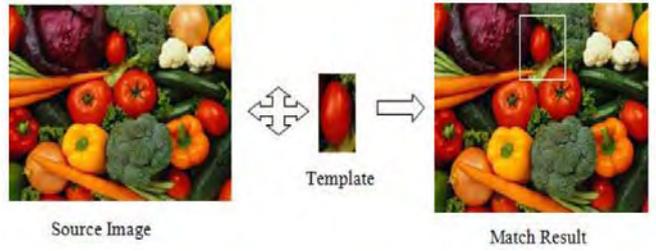
(그림 2) 데이터베이스 스키마

그림 2 는 해당 시스템을 위한 데이터베이스 스키마로, 사용자 이미지 및 크롤 기록을 저장할 수 있도록 구성하였다. 자료를 효율적으로 저장할 수 있고 중복을 방지한다. 예를 들어 첫 번째 검색어로 사이트 A 를 찾았고, 두 번째 검색어로도 사이트 A 를 찾을 경우가 있다. 이 경우 중복체크가 가능하여 두 번째 검색어로 사이트 A 를 찾을 경우에는 저장하지 않는다.

3) 이미지 비교 알고리즘

이미지 비교 모듈은 템플릿 매칭 알고리즘을 이용하였다. 템플릿 매칭은 그림 3 과 같이 비교하고자 하는 이미지(Source Image)를 템플릿 이미지로 만든 후, 비교 대상 이미지와 거리 또는 상관관계를 이용하여 유사성을 가진 영역을 찾아내는 방식이다[4]. 제안 시스템에서는 이미지를 여러 슬라이드로 잘라서 아래의 수식과 같이 두 이미지의 상관계수를 결과 값으로 비교한다[5]. R 은 결과값, T 는 템플릿 이미지, I 는 비교 이미지에 해당하며, 에 대한 합연산을 한다. 타 알고리즘에 비해 빠른 비교 속도를 가진 것이 특징이다.

$$R(x, y) = \frac{\sum_{x', y'} T'(x', y') \cdot I'(x + x', y + y')}{\sqrt{\sum_{x', y'} T'(x', y')^2 \cdot \sum_{x', y'} I'(x + x', y + y')^2}}$$



(그림 3) 템플릿 매칭

4. 한계 및 개선 방향

제안 시스템에는 알고리즘의 불완전성이라는 한계를 가지고 있다. 먼저, 이미지 프로세싱의 템플릿 매칭 알고리즘은 수행 속도가 빠르다는 장점이 있지만 크기, 형태, 위치 등이 변형을 거칠 경우 정확도가 떨어진다. 이 경우, 다양한 알고리즘을 복합적으로 사용하여 개선하는 방법이 있다. 또한 실행 결과로 출력된 유사도가 높은 이미지에 대해 합법적 게시물인지 불법인지 판단하기 위해서는 저작권자가 거래 정보를 확인해야 한다는 한계가 있다. 현 시스템은 저작권 정보를 DB 에 저장하지 않고 있어서 현 시스템만으로는 확인이 어렵다는 한계가 있다.

이러한 한계점을 극복하기 위해서는 먼저 알고리즘을 다방면으로 적용하여 저작권 이미지 불법 사용자가 이미지에 변형을 가하더라도 찾아낼 수 있어야 한다. 또한, 저작권 유통회사가 보유한 이미지 저작권 데이터베이스를 활용한다면 제시된 시스템에 워터마킹 및 핑거프린트를 추출하여 최초 유포자를 추적하는 모듈을 추가하여 할 수 있다.

5. 결론

본 연구를 통해 저작권 침해 적발을 자동화하는 시스템을 제안하였다. 대규모 저작권 침해 환경을 살펴볼 때 기존의 워터마크 및 핑거프린트 없이 대중적으로 사용 가능하다는 점에서 유의미한 접근이라고 할 수 있다. 추후에 저작권 유통회사에서 이와 같은 프로그램을 적용할 경우, 더욱 광범위한 저작권 침해 적발이 가능할 것으로 보인다.

Acknowledgement

본 연구는 미래창조과학부 및 정보통신기술진흥센터 서울어코드활성화지원사업의 연구결과로 수행되었음 (IITP-2015-R0613-15-1147)

참고문헌

[1] 남현우, “이미지 저작권 보호 및 이용 활성화를 위한 개선방안에 대한 연구,” *Korea Science & Art Forum*, vol. 15, pp. 233-246, Sep. 2014.
 [2] M. Guerrero, “A Comparative Study of Three Image Matching Algorithms: Sift, Surf, and Fast”, M. S. thesis, Dept. Civil and Environment Eng., Utah State Univ., Logan, UT, 2011.
 [3] M. Gharavi-Alkhansari, “A Fast Globally Optimal Algorithm for Template Matching Using Low-Resolution

- Pruning,” IEEE Transactions on Image Processing, Vol. 10, pp. 526-533, 2001.
- [4] A. Kumar, A. Joshi, A.Kumar, A. Mittal, and D. R. Gangodkar, “Template Matching Application in Geo-Referencing of Remote Sensing Temporal Image”, International Journal of Signal Processing, Image Processing and Pattern recognition, vol.7, No. 2, pp. 201-210, 2014.
- [5] http://docs.opencv.org/modules/imgproc/doc/object_detection.html