https://doi.org/10.7236/JIIBC.2023.23.3.27
JIIBC 2023-3-4

불법촬영물 검색 방지를 위한 시험 세트 방안 연구

A Study on Test Set to prevent illegal films searches

신용녀*

Yong-Nyuo Shin*

요 약 세계 각국은 아동 그루밍(grooming)과 같은 아동 성 착취 이미지의 제작 및 배포에 강력한 법 집행이 요구하고 있다. 이러한 사회적 문제의 규모와 중요성을 고려할 때 법 집행 기관, 정부, 업계, 정부 기관의 광범위한 협력이 필요하다. N-번방 사건이후 국내에서도 사전조치의무사업자의 제공 부가통신서비스에 대한 전기통신사업법 시행령의 일부 개정이 있었다. 국내 Naver등은 ETRI(한국전자통신연구원)의 자체기술을 탑재하여 불법촬영물에 대하여 필터링을 수행하는데 반해 Microsoft는 자체 PhotoDNA기술을 사용한다. Microsoft의 PhotoDNA는 불법 촬영물 등 비교, 식별하는 기술 자체가 뛰어나 Twitter등 주요 글로벌 사업자가 이미지를 탐지하고 필터링하는데 PhotoDNA 기술을 적용하고 있다. Microsoft는 한국 정부의 시험 기준에 맞추기 위해서 Bing 서비스에 적용되고 있는 “PhotoDNA for Video 2.0A”에 대하여 KCC(방송통신심의위원회)/TTA(한국정보통신기술협회)와 협력하여 총 16회가 넘게 성능 시험을 수행하였다. 본 논문에서는 기준에 통과하지 못한 사례에 대하여 분석하여 로고 추가 관련 개선방안을 도출한다. 또한, 불법촬영물에 대한 필터링 성능시험에 있어서 세 가지 동영상 데이터 세트를 성능시험에 사용하는 것을 제안한다.

Abstract Countries around the world are calling for stronger law enforcement to combat the production and distribution of child sexual exploitation images, such as child grooming. Given the scale and importance of this social problem, it requires extensive cooperation between law enforcement, government, industry, and government organizations. In the wake of the Nth Room Case, there have been some amendments to the Enforcement Decree of the Telecommunications Business Act regarding additional telecommunications services provided by precautionary operators in Korea. While Naver and others in Korea use Electronics and Telecommunications Research Institute's own technology to filter illegal images, Microsoft uses its own PhotoDNA technology. Microsoft's PhotoDNA is so good at comparing and identifying illegal images that major global operators such as Twitter are using it to detect and filter images. In order to meet the Korean government's testing standards, Microsoft has conducted more than 16 performance tests on "PhotoDNA for Video 2.0A," which is being applied to the Bing service, in cooperation with the Korea Communications Commission and Telecommunications Technology Association. In this paper, we analyze the cases that did not pass the standards and derive improvement measures related to adding logos. In addition, we propose to use three video datasets for the performance test of filtering against illegal videos.

Key Words : PhotoDNA for Video, Illegal Film, Logo Image, Filtering, Performance Test

*정회원, 마이크로소프트 최고기술임원(National Technology Officier)
접수일자 2023년 4월 4일, 수정완료 2023년 5월 4일
게재확정일자 2023년 6월 9일

Received: 4 April, 2023 / Revised: 4 May, 2023 / Accepted: 9 June, 2023
*Corresponding Author: yoshin@microsoft.com
Dept. of Public Sector, Microsoft, Korea

I. 서 론

인터넷은 유익하고 건설적인 방식의 콘텐츠 공유의 장이기도 하지만, 범죄자들이 청소년을 이용할 수 있는 통로이기도 하다. 한국에서는 N-번방 사건[1] 이후로 해당 문제에 대한 관련법이 정비되었고, 세계 각국은 아동 성 학대 영상 유포, 아동 인신매매, 아동 그루밍(grooming)과 같은 아동 성 착취 및 학대 이미지의 제작 및 배포에 강력한 법 집행을 요구하고 있다[2].

2002년부터 NCMEC[3](미국 실종과 착취된 아동을 위한 국가 센터)에서는 1억 개가 넘는 이미지와 동영상에 대한 검토를 진행하고있다. 이러한 이미지는 소아성애자들이 아동에 대한 성적 관심을 강화하기 위해 다른 사람들과 공유하는 과정에서 발견된다[3].

NCMEC가 확인한 아동 성학대 피해자의 대부분은 사춘기 이전의 아동이고, 영유아도 가장 빠르게 증가 하고 있다[3]. Microsoft는 이러한 불법 이미지를 찾아 신고하고 삭제함으로써 이 끔찍한 거래와 싸우는 데 중요한 역할을 하고 있다. 아동 성 착취의 또 다른 형태는 아동 성 착취범이 인터넷을 사용하여 피해자를 찾는 것이다. 이러한 범죄자들은 인터넷의 익명성을 이용해 청소년과 온라인 관계를 구축하거나 아동 성착취물을 유통하는 사람들과 아동을 성매매한다. 아동 성학대 이미지와의 싸움에서와 마찬가지로, 인터넷 기업들은 행동 강령을 시행하고, 고객이 가해자를 신고할 수 있는 메커니즘을 제공하며, 그리고 더 나은 탐지를 위한 혁신에 투자함으로써 아동 성매매를 막는 데 중요한 역할을 한다. 이 문제의 규모와 중요성을 고려할 때 법 집행 기관, 정부, 업계, 정부 기관의 광범위한 협력이 필요하다.

이를 위해 Microsoft는 WePR[4] 의 국제 자문 위원회의 80개 이상의 국가, 20개 기술 기업, 17개 시민 사회 단체 및 국제 기구와 협력하여 온라인 배포를 종식하기 위해 노력하고 있다. 또한, 2013년에 개소한 Microsoft의 Digital Crimes Unit[5]은 사이버 범죄 전문가인 전 세계의 변호사, 수사관, 비즈니스 전문가, 포렌식 분석가 등 파트너와 협력하여 기술을 이용한 범죄를 해결하고 있다. 고도로 훈련된 전문가를 고용하여 아동 성적 학대 이미지를 탐지하고 분류하는 데 도움을 주는 필터링 최첨단 기술을 적용하여, 발견 시 NCMEC에 보고하고, 이미지를 삭제하도록 하고 있다. 2009년에 Microsoft Research는 다트머스(Dartmouth) 대학과 협력하여 NCMEC에서 아동 성 학대 검색을 구체화하고 자동화하는 데 도움이 PhotoDNA[6]를 개발했다. PhotoDNA는 이전에 식별된 불법 이미지의 해시를 포함하는 데이터베이스와 매치하여 아동 학대 자료 유포를 감지, 방지 및 신고하는데 매우 유용한 도구이다. 또한, 2015년부터 PhotoDNA를 Azure 클라우드 서비스로 제공하기 시작하였다. 클라우드 서비스를 통해 사용자가 콘텐츠를 업로드할 수 있는 자유성을 가지고, 플랫폼의 무결성을 보장하고자 하는 소규모기업체 및 기타 단체들이 본 서비스를 이용할 수 있게 되었다.

II. 전기통신사업법 시행령 개정

N번방 사건이 후 방송통신위원회는 전기통신사업법을 일부 개정하고, 사전조치의무사업자의 제공 부가통신서비스의 표 1에 대하여 신설하였다.

표 1. 사전조치의무사업자의 제공 부가통신서비스
Table 1. Additional telecommunications services provided by the precautionary duty carrier

"사전조치의무사업자가 제공하는 부가통신서비스"란 이용자가 정보통신망을 통하여 일반에게 공개되어 유통되는 정보(이하 이 표에서 "정보"라 한다)를 게재·공유 또는 검색할 수 있도록 제공하는 부가통신서비스로서 다음 각 호의 어느 하나에 해당하는 서비스를 말한다.
1. 사회관계망서비스, 온라인 커뮤니티, 대화방 등 불특정 다수의 이용자가 부호·문자·음성·음향·화상·영상 등의 정보를 게재하여 이를 서로 공유하는 것을 목적으로 하는 기술적 수단을 제공하는 서비스
2. 진행자가 출연하여 제작한 부호·문자·음성·음향·화상·영상 및 이들의 조합으로 이루어진 콘텐츠를 게재하여 불특정 다수의 이용자에게 실시간으로 공유할 수 있도록 하는 것을 목적으로 하는 기술적 수단을 제공하는 서비스
3. 불특정 다수의 이용자가 정보를 검색했을 때 그 정보 및 부호·문자·음성·음향·화상·영상 등의 검색 결과 정보를 송출[링크(link) 등 정보통신망 상에 있는 해당 정보의 위치를 송출하는 것을 포함한다]하는 것을 목적으로 하는 기술적 수단을 제공하는 서비스

Microsoft 의 Bing 검색은 이용자의 콘텐츠를 호스팅하는 것이 아니라, 인터넷에서 제공하는 제3자의 콘텐츠만 인덱싱하기 때문에, Bing이 불법 콘텐츠를 식별하고 제거하는 접근 방식은 소셜미디어와 같은 호스팅 플랫폼과는 다르다. 아동 성 착취 및 폭행 이미지(Child Sexual Exploitation and Assault Image)의 경우, Bing은 PhotoDNA에 따라 불법 CSEAI를 식별하고 인덱스에서 제거하고 있다. 전기통신사업법에 따른 그 외의 불법촬영물 등 콘텐츠의 경우, Bing은 한국의 규제당국이 관련 링크를 파악하고 해당 콘텐츠의 소재지를 Bing에 알려주는 것을 따른다. CSEAI와 같이 매우 한정된 몇 가지의 시나리오 외에, 대부분의 콘텐츠에 대한 '불법성'은 맥락과 상황에 따라 달라질 수밖에 없기 때문

에, 완전히 자동화하기에는 어려움이 있었다. 특정 케이스에서 표준 기술의 속도(응답시간) 문제로 서비스에 차질이 발생할 경우 게재 후 삭제(비교·식별)해도 무방한지에 대한 검토가 있었다. 법령에 따르면 원론적으로 불법촬영물이 게재되는 것은 위반사항에 해당한다. 사전에 불법 촬영물을 검색하여 Bing등에서 차단하는 것이 해당 성능시험의 궁극적 목적이었다. 일부 글로벌 기업들은 표 2의 기술적 조치 중에 ETRI와 같은 기관이 개발하여 제공하는 기술을 탑재하기가 현실적으로 어려웠다. 따라서, 방송통신위원회가 정하여 고시한 TTA와 같은 단체에게 PhotoDNA for video같은 서비스에 대하여 성능시험을 받고 Bing에 적용하는 방향이 필수적 이었다.

표 2. 기술적 조치(시행령 제30조의6제2항제3호)
Table 2. Technical Measure(Article 30-6, Paragraph 2, Subparagraph 3 of the Enforcement Decree)

이용자가 게재하려는 정보의 특징을 분석하여 방송통신심의위원회에서 불법촬영물 등으로 심의·의결한 정보에 해당하는지를 비교·식별 후 그 정보의 게재를 제한하는 조치. 이 경우 다음 각 목의 어느 하나에 해당하는 기술을 사용하여 비교·식별해야 한다. 가. 국가기관이 개발하여 제공하는 기술 나. 방송통신위원회가 정하여 고시하는 기관·단체가 최근 2년 이내에 시행한 성능평가를 통과한 기술

III. 불법촬영물 필터링 성능시험 환경과 지표

방송통신위원회에서 제공하는 표준 DNA DB 기술은 영상의 특징값을 추출하는 모듈과 표준 DNA DB와 비교검색을 통해 불법성 여부를 검출하는 모듈을 제공한다. Microsoft에서는 초기에 방통위에서 제공하는 DNA DB 기술이 어떻게 작동하는지, 마이크로소프트 시스템 내에 구현 가능한지, 그리고 표준 DNA DB 내에 무엇이 있고 이를 결정하는 데 사용되는 정확한 기준이 무엇인지 심층적으로 검토할 필요가 있었다. 제공된 방식은 표준 DNA DB 기술은 영상의 특징값을 추출하여 표준 DNA DB와 비교·검색을 통해 불법촬영물 여부를 확인하는 방식이다. 이러한 방식을 이용하여 시스템을 구축하는 경우에 GPU 활용 여부 및 주요 시스템에 필요한 시스템 사양 등은 어느 정도가 필요한지에 대한 사접협의가 필수적이다. 특히, 제공된 기술이 사업자 시스템에 적용하는 경우, 개발 언어 등 필요한 요구사항 및 선호하는 API 형태가 CLI 등 실행파일인지 Python 라이브러리 또는 Docker를 선호하는지에 대한 확인이 필요하였다. 연산을 위해인터넷에서 많은 양의 콘텐츠를 인덱싱하고 있고, Bing에서는 API를 불러올 수 없기 때문에, Microsoft 기존에 사용되고 있는 해시태그 기술(hash tag technology)의 사용이 필수적이었다. 정보통신사업 시행령이 개정됨에 따라 대상이 되는 기업들은 Microsoft, Twitter, Tiktok, Facebook, Twitch, Wechat, Google, Pinterest 등 이었는데, 일부 서비스 제공자는 영상물 길이(평균 시간), 단위시간당 건수 등의 실시간 성으로 사업자 서버에 업로드되는 영상물의 특성을 확인할 수 있는 정보 제공이 가능한지 여부에 대한 확인이 필요하였다.

1. 시험 환경

시험 환경은 그림 1과 같이 평가기관과 신청기관이 각각 독립적으로 구축하고, 인터넷을 통해 데이터를 송수신 한다. 이때, 평기기관은 신청기관의 시험환경 구축에 대하여 관여하지 않는다.

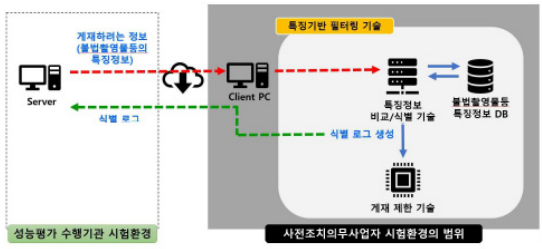

그림 1. 사전조치의무사업자가 준비해야 할 시험환경의 범위
Fig. 1. Scope of test environments for precautionary operators

평가기관은 서버에 평가용 데이터세트(특징정보 전자파일)을 구성하고 FTP나 전자 메일 등의 사전에 협의된 방법으로 신청기관에 제공한다. 본 성능시험의 평가지표 및 상세 설명은 표 3과 같다.

표 3. 성능시험의 평가지표
Table 3. Metrics for performance tests

평가지표	상세 설명
불법촬영물등 식별 가능성	게재하려는 정보가 불법촬영물등(변형물 포함)의 특징정보에 해당하는 경우 방송통신심의위원회에서 불법촬영물 등으로 심의·의결한 정보로 인식하는 비율
일반영상 식별 가능성	게재하려는 정보가 불법촬영물등(변형물 포함)의 특징정보에 해당하지 않는 경우 방송통신심의위원회에서 불법촬영물 등으로 심의·의결한 정보로 인식하지 않는 비율
일관성	불법촬영물 등 식별 가능성 및 일반영상 식별 가능성에 대한 반복 시험결과, 동일특징정보의 식별 결과가 동일한 비율

신청기관은 Client PC에 저장된 평가용 데이터세트를 순차적으로 필터링 기술에 입력하고, 그 결과를 로그파일에 기록한 후 평가기관에 제공한다. 시험평가자는 평가용 데이터세트, 불법 촬영물 등 특징 정보 목록, 로그파일 등을 분석하여 평가지표별로 결과를 계산한다. 평가용 데이터세트는 방송통신심의의원회가 필터링 기술로 생성한 특징정보 전자파일을 성능평가에 적합한 형태로 재분류하여 구성한다. 평가지표별 결과 계산식 및 표기법은 표 4와 같다.

표 4. 성능평가 결과에 대한 계산식
Table 4. Calculations for performance evaluation results

평가지표	상세 설명	표기법
불법촬영물등 식별 가능성	(불법촬영물등으로 인식한 개수 / 평가용 동영상에 포함된 불법촬영물등 개수) * 100	소수점 이하 1자리까지 표기
일반영상 식별 가능성	(불법촬영물등으로 인식하지 않은 갯수 / 평가용 동영상에 포함된 일반영상 개수) * 100	
일관성	동일 특징정보의 식별 결과가 동일한 개수 / 평가용 데이터셋 개수) * 100	

Ⅳ. PhotoDNA for Video에 대한 성능시험

방송통신심의위원회(KCS)는 불법 촬영물 DB를 내부에 가지고 있지만, 불법 촬영물 원본에 대한 접근을 제한하기 위해서 성능 시험을 위해 특징점에 대한 변형을 가한 동영상에 대하여 한국정보통신기술협회에서 이를 기반으로 성능시험을 요청한 주체와 그림 2와 같이 성능시험을 실시한다.

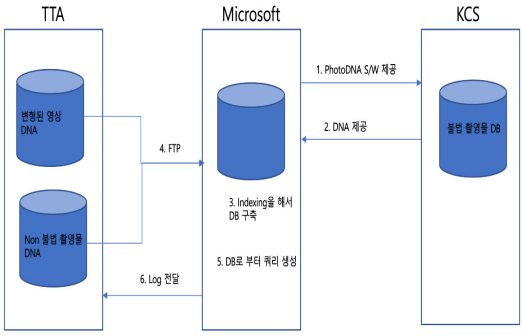

그림 2. 성능시험을 위한 각 기관별 정보 제공 범위
Fig. 2. Scope of information provided by each organization for performance testing

성능시험은 Microsoft가 PhotoDNA for Video 소프트웨어를 KCS에 제공하는데서 시작한다. KCS에서는 불법촬영물에 대한 DNA를 제공한다. Microsoft에서는 검색(Searching)을 위해 Indexing을 해서 자체 데이터베이스를 구축하게 된다. 이 단계가 완료되면 TTA에서는 성능시험을 위해서 불법촬영물에 대하여 일부 영상을 잘라내거나 외곡 등의 변형된(modified) 영상 DNA 샘플과 변형되지 않은(non-modified) 영상 DNA 샘플을 섞어서 FTP로 Microsoft에 제공하게 된다. 평가용 데이터세트는 불법촬영물등과 일반영상을 잘라내기(Clip), 로고 추가(Logo Overlay), 해상도 변화(Resolution Change), 프레임률 변화(Frame Rate Change), 화면 크기 변화(Size Change) 등 변형하여 만든 20초 이하 동영상의 특징정보로 구성되어 있다. Microsoft는 DB로부터 쿼리(Query)를 날려서 TTA가 제공한 이미지가 불법 촬영물에서 변형된 동영상이면 Match의 결과를, 불법 촬영물과 상관없는 동영상이라면 No-Match의 결과를 줘야 한다. 문체부 산하의 저작권 위원회의 통과기준이 이미 설정완료 되어 있었으므로, 시간이 얼마나 걸리는 것은 중요한 요소가 아니고 성공률은 95% 이상이어야 한다.

Microsoft는 성능시험이 완료되고 나면 해당 Log를 TTA에 전달하게 되고, 로그는 그림 3과 같은 형태가 된다.

```
2021-07-04, 15:03:08, 1111, [00001111], 2021-07-04, 15:03:09, matching, 0000, [11110000]
```

그림 3. 성능시험 완료 로그 샘플
Fig. 3. Sample Performance Test Completion Log

TTA에서 제시한 식별 로그 항목들은 표 5와 같다.

표 5. 식별 로그 항목들에 대한 설명
Table 5. Description of identification logs

구분	내용
질의 시작일	'특징정보 비교/식별 기술'이 '불법촬영물등 특징정보 DB'로 특징정보 파일과 특징정보 DB와의 매칭 여부를 질의(query)한 날짜(년-월-일)
질의 시작시간	'특징정보 비교/식별 기술'이 '불법촬영물등 특징정보 DB'로 특징정보 파일과 특징정보 DB와의 매칭 여부를 질의(query)한 시간(시:분:초)
질의 특징정보 파일명	'특징정보 비교/식별 기술'이 '불법촬영물등 특징정보 DB'로 특징정보DB와의 매칭 여부를 질의(query)한 특징정보 파일명

질의 종료일	'특징정보 비교/식별 기술'이 '불법촬영물등 특징정보 DB'로 질의(query)한 결과를 수신한 날짜(년-월-일)
질의 종료시간	'특징정보 비교/식별 기술'이 '불법촬영물등 특징정보 DB'로 질의(query)한 결과를 수신한 시간(시:분:초)
식별 결과	'특징정보 비교/식별 기술'이 '불법촬영물등 특징정보 DB'로 질의(query)한 결과 ('matching' 또는 'none')
매칭 특징정보 파일명	'특징정보 비교/식별 기술'이 '불법촬영물등 특징정보 DB'로 질의(query)한 결과 - 식별 결과가 'matching'인 경우 : 매칭된 특징정보 파일명 - 식별 결과가 'none'인 경우 : NULL

Microsoft의 PhotoDNA for Video 2.0A에 대한 성능시험 결과는 표 6과 같다. 이를 통해 불법촬영물 필터링 기술에 대한 성능시험을 통과(Pass)하였다.

표 6. PhotoDNA에 대한 성능시험 결과
Table 6. Performance test results for PhotoDNA for Video

평가지표	통과기준
불법촬영물등 식별 가능성	95.5% 이상
일반영상 식별 가능성	99.9% 이상
일관성	100% 이상

V. PhotoDNA의 성능시험 결과

파라미터(Parameter) 값을 수정하면서 TTA 기준에 맞춰 PhotoDNA가 성능시험에 통과될 수 있도록 테스트를 진행했다. 총 16차례의 성능시험 테스트가 진행되었으며, 이 과정에서 14번째에 표7과 같은 수치들이 도출되었다.

표 7. PhotoDNA에 대한 14번째 성능시험 결과
Table 7. 14th Performance Test Results for PhotoDNA

정의	
matching	불법 촬영물을 원본 불법 촬영물로 매칭
none	불법 촬영물을 매칭하지 못함
matching&none	불법 촬영물을 다른 불법촬영물로 매칭함

	식별 가능성
불법촬영물	90.13006208
일반영상물	100

	matching	none	matching& none
20초 미만 불법촬영물	93.03626943	6.67357513	0.207253886
20초 불법촬영물	89.64661265	8.908808826	1.423892432

20초 미만 불법 상세	matching	none	matching&none
원본	99.79274611	0	0.207253886
로고	67.25388601	32.43523316	0.207253886
해상도	99.48186528	0.207253886	0.207253886
프레임	99.58549223	0.103626943	0.207253886
보더	99.06735751	0.621761658	0.207253886

20초 불법 상세	matching	none	matching& none
원본	98.39682813	0.206860886	1.379072574
로고	57.52456473	41.21703155	1.258403723
해상도	98.12101362	0.430960179	1.448026202
프레임	98.3278745	0.189622479	1.396310981
보더	95.86278228	2.49956904	1.637648681

위의 성능시험 결과를 기반으로 분석해 보면, 로고 추가(Logo Overlay)를 제외하고는 99% 이상의 matching 결과를 보이고 있다. 불법촬영물로 통용되는 일반적인 이미지에서는 로고가 사용되는 사례가 극히 드물고, 특히 TTA에서 제시된 로고의 크기와 위치는 일반적인 영상에서도 흔하지 않은 경우이다. 시험자의 입장에서는 합법적인 콘텐츠를 차단할 가능성(Fault Positive)도 존재 할 수 있고, 로고가 지나치게 크게 삽입되어 제공되고 있는 TTA의 영상 컨텐츠에 가릴 수도 있기 때문에 로고 추가 관련 개선이 필요하다. PhotoDNA는 Microsoft에서 만든 이미지 매칭 기술로, 전 세계 인터넷 회사와 법 집행 기관에서 불법 콘텐츠를 탐지하는 데 사용되고 있다. 현재 TTA에서 제공하는 로고가 포함된 이미지와 테스트 데이터에 대한 접근 권한이 없는 상황에서 사용자 보호를 위한 솔루션을 구현할 수 있는 대안을 제대로 평가하기 위해서는 불법 아동 성적 학대 콘텐츠의 유통을 방지하는 데 오랫동안 사용되어 왔고, 전 세계의 다른 인터넷 기업과 한국 경찰을 포함하는 법 집행 기관에서도 이 기술을 사용하고 있는 점을 감안하여 이러한 문제를 해결하기 위해 함께 노력해야 한다. 이를 위해서는 로고의 크기와 위치를 유연하게 결정할 수 있도록 해야 한다.

로고가 매우 크거나 중요한 콘텐츠 위에 배치된 경우, 오탐 결과가 많이 발생하므로 매칭 되지 않을 수 있다. PhtoDNA는 이미지의 해시(Hash)를 계산한 다음 이 해시를 알려진 불법 콘텐츠의 해시 데이터베이스와 비교하는 방식으로 수행한다. PhotoDNA 해시는 144자 길이이며, 해시를 비교할 때 첫 번째 해시의 각 값에서 두 번째 해시의 비교 가능한 값을 뺀다. 그 차이를 제곱하고 합산하여 두 이미지가 얼마나 유사한지 보여주는 결과를 생성한다. PhotoDNA 해시는 6x6 격자로 분리된 이미지의 '특징'을 나타낸다.

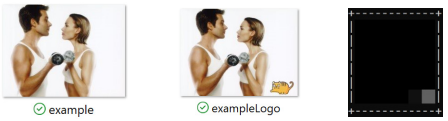

그림 4. 상대적으로 작은 로고가 삽입된 이미지 예제
Fig. 4. Example image with a relatively small logo embedded

이는 불법 이미지에 추가된 로고의 일반적인 크기이며, Uploader는 평판을 위해 이 'Tag'를 추가하는 것으로, 이미지를 가리기 위한 것이 아니다. 이 두 이미지의 해시를 비교하면 다음과 같이 시각화할 수 있다. 오른쪽 하단에서 이미지의 차이가 발생하는 부분을 강조 표시한 것을 볼 수 있다. 더 큰 사이즈의 로고를 보면 그 차이가 더 분명하게 드러나는 것을 볼 수 있다.

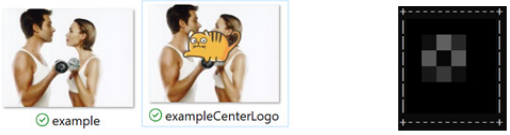

그림 5. 큰 로고가 삽입된 이미지 예제
Fig. 5. Example image with a big logo embedded

이러한 로고에 의해 불법 콘텐츠가 가려질 가능성이 높기 때문에 TTA에서 제공하는 삽입 로고 이미지는 일반적으로 불법 콘텐츠를 공유하는 방식이 아니다. 이렇게 큰 로고 오버레이가 되는 경우는 이미 불법 촬영물이 아닐 가능성이 많다. 국가적인 테스트를 수행하는 것이기 때문에 전세계적으로 이러한 데이터 세트를 가지고 불법촬영물에 대한 성능시험이 수행되는 것이 일반적이다. 따라서, 불법 촬영물에 대한 성능시험을 수행하는 데 있어서 다음의 세 가지 동영상 데이터 세트를 성능시험에 사용하는 것을 제안한다. 첫째, 합법적으로 동영상이 제작되는 경우 동영상의 내용을 볼 수도 있고, 시험용 도구 제작자가 수정이 필요한 경우에 동영상 제작 툴을 사용해서 동영상을 테스트를 해볼 수 있는 동영상 세트를 제공한다. 둘째, 합법적으로 동영상이 제작되는 경우 성능시험 수행자가 동영상 내용을 볼 수 있고 성능시험 수행자가 시험 도구의 초기 테스트를 수행하는 데 사용할 수 있는 동영상 세트를 제공한다. 셋째, 불법적으로 동영상이 제작되는 경우 성능시험 수행자나 성능시험 대상 도구 제작자가 볼 수 없으며 최종 검증에 사용할 수 있는 동영상 세트를 제공한다. 성능평가 수행기관으로부터 제시된 통과기준은 불법촬영물 식별가능성, 일반영성 식별가능성, 일관성에 있어서 모두 95% 이상이여야 했고, PhtotoDNA는 Parameter의 조정을 통해 성능시험을 통과하였다. 성능기준은 추후, 불법 촬영물 등이 사회에 미치는 영향 증가, 기술 성숙도 향상 등 인터넷 이용환경 변화에 따라 통과 기준은 변경이 가능하지만, 개발 평가 지표 모두 통과 시 성능평가를 통과할 수 있다.

Ⅵ. 결 론

마이크로소프트는 정보통신서비스 제공자가 불법촬영물 등의 유통 방지를 위하여 크게 세 가지의 노력을 기울이고 있다. 첫째, 불법촬영물 등의 유통 방지를 위하여 아동 성 착취 및 학대 이미지, 합의되지 않은 사적인 이미지(non-consensual sharing of intimate imagery) 공유 행위 등을 비롯하여 불법적이고 유해한 성적 콘텐츠 유통의 방지를 위해 CSEAI 삭제를 포함한 디지털 안전 투명성 보고서[7]를 제공하고 있다. 둘째, 이용자의 불법촬영물 신고기능을 마련하였다. CSEAI의 유통을 방지하기 위하여 PhotoDNA 기술을 개발하였으며, CSEAI 대응을 위하여 전 세계 유관 기관들과 해당 기술을 공유하는 등 문제 해결을 위해 오랜 노력을 기울이고 있다. 피해자들이 신고하면 대상 서비스 내에서 해당 촬영물 등을 삭제할 수 있도록 하는 정책을 시행하고 있다[8]. 셋째, 검색제한조치를 수행하고 식별 및 게재 제한 조치를 수행한다. 아울러, 통상적인 성인 콘텐츠와 관련하여 Bing 이용자들에게 연령확인과 '안전검색(SafeSearch)' 기능을 통해 성인 콘텐츠에 대한 필터링 기능을 제공하고 있다. 또한, 불법 콘텐츠의 자동 탐지, 방지 및 삭제를 위해 사용되는 분류기(designed classifiers)[9]를 지속적으로 개선하고 있다. 또한, 서비스에 대한 불만 기록을 해결일로부터 3년간 보관하고 있다. 법적인 의무를 다하기 위해 관련 컨텐츠 조정 및 공개되지 않는 필터링 분류 기록을 보존하고 있다.

본 연구에서는 사전조치의무사업자의 제공 부가통신 서비스의 의무를 준수하기 위해 PhotoDNA 에 대하여 KCC(방송통신심의위원회)[10] /TTA(한국정보통신기술협회)와 협력하여 성능 시험을 수행한 결과를 제시한다. 기준에 통과하지 못하는 사례에 대하여 분석하여 로고 추가 관련 개선방안을 도출한다. 본 연구를 통해 추후 불법 촬영물에 대한 필터링 성능시험에 있어서 세 가지 동영상 데이터 세트를 성능시험에 사용하는 것을 제안하였다. 이를 통해, PhotoDNA와 유사한 불법 촬영물을 검출해주는 서비스 혹은 도구를 개발하려는 경우에 이런 기술적인 내용들을 고려하는데 기여를 할 것이다.

향후 연구는 본 연구에서 제안된 세 가지 동영상 데이터 세트를 사용하여 불법촬영물에 대한 필터링 성능시험을 연구하는 것이다.

References

[1] Wikipedia, “Nth Room Case” June 2020. Nth Room case – Wikipedia

[2] Microsoft, “Combating Child Sexual Exploitation and Abuse Online”, July 2020. GAHub - Combating Child Sexual Exploitation and Abuse Online - All Documents (sharepoint.com)

[3] National Center for Missing &Exploited Children (missingkids.org), May 2021.

[4] WeProtect Global Alliance - against child sexual exploitation and abuse online

[5] Digital Crimes Unit: Leading the fight against cybercrime – On the Issues (microsoft.com), Jan 2022.

[6] Microsoft, PhotoDNA | Microsoft.

[7] Microsoft, Digital Safety Content Report | Microsoft CSR

[8] ‘Revenge porn’: Putting victims back in control - Microsoft On the Issues

[9] Victoria Baines, Online Child Sexual Exploitation: Towards an Optimal International Response, August 2018

[10] Digital Sexual Crime and PublicDNA DB technical Support Portal, Korea Communicatins Standards Commission, https://dna.kocsc.or.kr/mainPage.do

[11] Eun-Gyu Ham, Chang-Bok Kim, "Model Implementation of Reinforcement Learning for Trading Prediction Using Deep Q Network ," The Journal of KIIT, Vol. 17, No. 4, pp. 1-8, 2019.

저 자 소 개

신 용 녀(정회원)

- 1999년 : 숭실대학교 컴퓨터학사.
- 2001년 : 고려대학교 전산학 석사.
- 2008년 : 고려대학교 전산학 박사.
- 2002년 ~ 2009년 : 한국인터넷진흥원 연구원.
- 2010년 ~ 2014년 : 한양사이버대학교 컴퓨터학과 교수.
- 2015년 ~ 2018년 : 아마존 웹 서비스 기술이사.
- 2018년 ~ 현재 : 마이크로소프트 최고기술임원.