

딥 러닝을 이용한 실감형 콘텐츠 특징점 추출 및 식별 방법

A Feature Point Extraction and Identification Technique for Immersive Contents Using Deep Learning

박 병 찬*, 장 세 영*, 유 인 재**, 이 재 청**, 김 석 윤*, 김 영 모*★

Byeongchan Park*, Seyoung Jnag*, Injae Yoo**, Jaechung Lee**,
Seok-Yoon Kim*, Youngmo Kim*★

Abstract

As the main technology of the 4th industrial revolution, immersive 360-degree video contents are drawing attention. The market size of immersive 360-degree video contents worldwide is projected to increase from \$6.7 billion in 2018 to approximately \$70 billion in 2020. However, most of the immersive 360-degree video contents are distributed through illegal distribution networks such as Webhard and Torrent, and the damage caused by illegal reproduction is increasing. Existing 2D video industry uses copyright filtering technology to prevent such illegal distribution. The technical difficulties dealing with immersive 360-degree videos arise in that they require ultra-high quality pictures and have the characteristics containing images captured by two or more cameras merged in one image, which results in the creation of distortion regions. There are also technical limitations such as an increase in the amount of feature point data due to the ultra-high definition and the processing speed requirement. These consideration makes it difficult to use the same 2D filtering technology for 360-degree videos. To solve this problem, this paper suggests a feature point extraction and identification technique that select object identification areas excluding regions with severe distortion, recognize objects using deep learning technology in the identification areas, extract feature points using the identified object information. Compared with the previously proposed method of extracting feature points using stitching area for immersive contents, the proposed technique shows excellent performance gain.

요 약

4차 산업의 주요 기술로 실감형 360도 영상 콘텐츠가 주목받고 있다. 전 세계 실감형 360도 영상 콘텐츠의 시장 규모는 2018년 67억 달러에서 2020년 약 700억 달러까지 증가될 것이라고 전망하고 있다. 하지만 대부분 실감형 360도 영상 콘텐츠가 웹하드, 토렌트 등의 불법 유통망을 통해 유통되고 있어 불법복제로 인한 피해가 증가하고 있다. 이러한 불법 유통을 막기 위하여 기존 2D 영상은 불법저작물 필터링 기술을 사용하고 있다. 그러나 초고화질을 지원하고 두 대 이상의 카메라를 통해 촬영된 영상을 하나의 영상에 담은 실감형 360도 영상 콘텐츠의 특징 때문에 왜곡 영역이 존재하여 기존 2D 영상에 적용된 기술을 그대로 사용하기엔 다소 무리가 있다. 또한, 초고화질에 따른 특징점 데이터량 증가와 이에 따른 처리 속도 문제와 같은 기술적 한계가 존재한다. 본 논문에서는 이러한 문제를 해결하기 위하여 왜곡이 심한 영역을 제외한 객체 식별 영역을 선정하고, 식별 영역에서 딥 러닝 기술을 이용하여 객체를 인식하고 인식된 객체의 정보를 이용하여 특징 벡터를 추출하는 특징점 추출 및 식별 방법을 제안한다. 제안한 방법은 기존에 제안되었던 스티칭 영역을 이용한 실감형 콘텐츠 특징점 추출방법과 비교하여 성능의 우수성을 보였다.

Key words : Immersive Content, Deep Learning, Feature Point Extracting and Matching, Piracy Judgment, OMAF

* Dept. of Computer Science and Engineering, Soongsil University

★ Corresponding author

E-mail : ymkim828@ssu.ac.kr, Tel : +82-2-813-0682

※ Acknowledgment

This research project supported by Ministry of Culture, Sport and Tourism(MCST) and Korea Copyright Commission in 2020(2018-360_DRM-9500).

Manuscript received Apr. 8, 2020; revised Jun. 23, 2020; accepted Jun. 24, 2020.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. 서론

4차 산업 혁명의 주요 기술로 실감형 콘텐츠가 주목받고 있다. 실감형 콘텐츠는 다양한 센서를 이용해서 영상 안에 현실감 있게 재현하여 픽셀 위주의 콘텐츠가 아닌 사람의 제스처, 모션, 음성 등 실제 현장을 몰입할 수 있도록 경험을 제공한다. 또한, 가상의 물체를 실제의 물체처럼 조작할 수 있게 하는 디지털 콘텐츠이다[1]. 최근 몇 년간 실감형 콘텐츠에 대한 관심이 높아지고 있으며 관련 기업들은 실감형 콘텐츠 기업 또는 개인이 직접 제작할 수 있는 360도 카메라와 실감형 콘텐츠를 지원하는 HMD(Head Mounted Display)를 출시하고 있다[2]. 하지만 이러한 실감형 콘텐츠가 본격적으로 유통됨에 따라 DRM을 해체하여 웹하드, 토렌트 등을 통해 불법 유통되어 심각한 저작권 침해를 하여 피해를 보는 사례가 나타나고 있다. 기존 저작권 보호 기술은 2D 영상에 주요하고 있어 실감형 콘텐츠에 적용하기에 어려움이 있다. 실감형 콘텐츠의 특징은 2개 이상의 카메라가 하나의 영상으로 합쳐 있으며 HMD를 통하여 사람이 보는 영상은 전체 실감형 콘텐츠의 일부분만 보기 때문에 전체 실감형 콘텐츠의 해상도가 매우 높아야 한다. 또한, 전 방향 영상을 2D로 표현하기 위해서 프로젝션을 해야 한다. 2개 이상의 카메라를 하나의 영상으로 합쳐진 것이기 때문에 어느 특정한 영역에서 왜곡이 발생할 수 있으며 프로젝션 타입에 따라 왜곡 영역도 다르다.

본 논문에서는 저작권 보호 기술을 적용하기 위해 기존의 실감형 콘텐츠에서 특징점을 추출하기 위한 영역을 지정하고 고해상도의 영상을 리사이즈한다. 그리고 딥 러닝 기술을 이용하여 리사이즈된 영역에서 객체를 추출한 다음, 객체에서 특징점을 추출하는 방법을 제안한다. 본 논문의 구성은 다음과 같다. 2절에서는 관련 연구로서 실감형 콘텐츠의 프로젝트 타입 중 하나인 ERP와 본 논문에서 적용할 딥 러닝 기술인 YOLO 알고리즘에 대해서 기술한다. 3절에서는 본 논문에서 제안하는 딥 러닝을 이용한 실감형 콘텐츠의 특징점 추출 및 식별 방법을 기술한다. 4절에서 결과를 제시하고, 5절에서 결론으로 마무리한다.

II. 관련 연구

1. ERP(Equirectangular Projection)

실감형 콘텐츠를 제작하는 과정으로 두 대 이상의 카메라 또는 어안렌즈 카메라로 현장을 촬영한다. 촬영된 장면은 3차원 구형(Rectangle) 영상으로 촬영되며 한 장면으로 합치기 위해 스티칭하고 2차원 영상으로 변형하기 위한 프로젝션을 한다. 특히, 3차원의 영상을 일반 모니터로 표현할 방법이 없기 때문에 프로젝션 과정이 필요하다. 프로젝션 방법 중 가장 많이 사용하는 방법은 ERP(Equirectangular Projection)이며, Fig. 1과 같다.

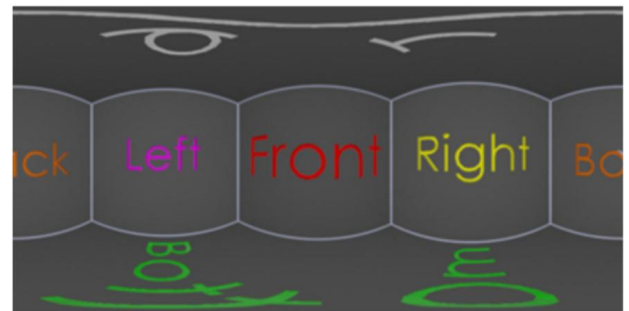


Fig. 1. ERP Representation.

그림 1. ERP의 표현

실감형 콘텐츠의 ERP 표현 방법은 전방향(앞, 뒤, 위, 아래, 오른쪽, 왼쪽)을 한 영상에 모두 담았기 때문에 왜곡 영역이 존재하며, 특히 위, 아래 부분이 가장 심한 특성을 가지고 있다[3-5].

2. YOLO

YOLO는 객체 탐지 알고리즘으로써 원본 이미지를 동일한 크기의 그리드로 나누고 각 그리드에 대해 그리드 중앙을 중심으로 미리 정의된 형태(Perdefined Shape)로 지정된 경계박스의 개수를 예측하여 이를 기반으로 신뢰도를 계산한다. 이미지에 객체가 포함되어 있는지 또는 배경만 단독으로 있는지에 대한 여부가 포함되며, 높은 객체 신뢰도를 가진 위치를 선택해 카테고리를 파악한다. k-평균 알고리즘에 의해 객체에 바운딩 박스를 만들 수 있으며, 테이터셋의 객체 크기와 형태에 대한 사전 정보를 확보한다. 각각의 박스는 각기 다른 크기와 형태의 객체를 탐지하도록 설계되어 있다.



Fig. 2. Bounding Box.
그림 2. 바운딩 박스

3. A Method for Extracting Feature Point Using Stitching Area

기존 연구의 하나로써 스티칭 영역을 이용한 특징점 추출방법이 제안되었다[6]. 두 대 이상의 카메라가 촬영한 영상을 하나로 합치는 과정을 스티칭 과정이라고 하는데, 편집에 따라 스티칭 영역이 확실히 드러난다.

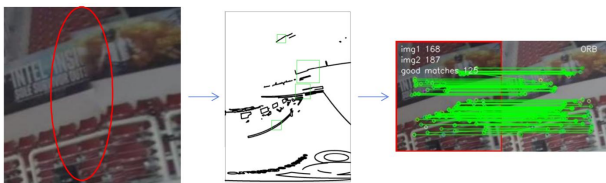


Fig. 3. Feature point extraction process in stitching area.
그림 3. 스티칭 영역에서 특징점 추출 과정

[6]에서는 이러한 스티칭 영역을 이용하여 특징점을 추출하는 과정을 ORB 알고리즘을 이용하여 제안하였다. 그러나 이러한 방법은 편집 방법에 따라 스티칭 영역이 드러나지 않는 경우가 있으므로 모든 실감형 콘텐츠의 특징점 추출방법으로 적용하는 것은 무리가 있다.

III. 딥 러닝을 이용한 실감형 콘텐츠의 특징점 인식률 향상 방법

실감형 콘텐츠는 UHD 이상의 화질을 지원하는 초고화질의 특징을 가지고 있어 특징점 데이터의 증가로 인해 특징점 추출 및 인식속도가 느려질 수 있는 문제를 해결하기 위한 방법은 Fig. 4와 같다.

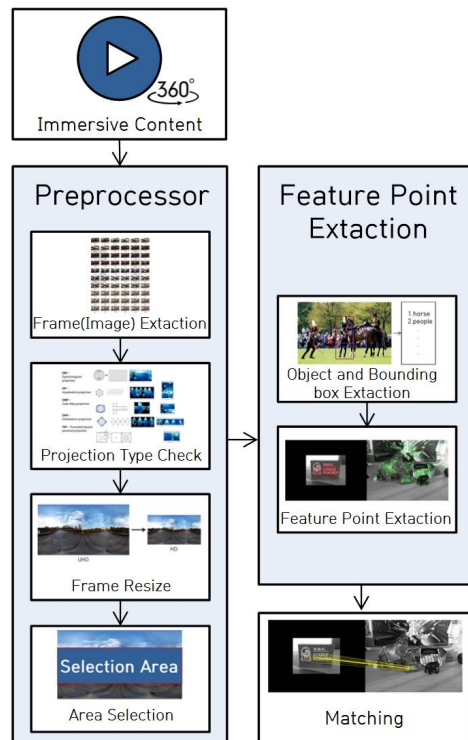


Fig. 4. A Feature Point Identification Method for Immersive Contents.

그림 4. 실감형 콘텐츠의 특징점 식별 방법

본 논문에서 제안하는 딥러닝을 이용한 실감형 콘텐츠의 특징점 추출 및 인식속도 개선 방법으로 크게 전처리 과정, 특징점 추출 과정, 매칭 과정으로 3개의 영역으로 나눈다[8]. 전처리 과정에서는 먼저 실감형 콘텐츠가 입력되면 콘텐츠의 프레임을 추출한다. 추출방법은 GOP(Group of Picture)의 I-frame을 이용하여 추출한다[9]. 추출된 프레임에서 입력된 실감형 콘텐츠의 프로젝션 타입을 확인한다. 본 논문에서 필요한 프로젝션 타입은 ERP임으로 다른 타입의 프로젝션이 사용된 콘텐츠는 ERP로 바꾼다.

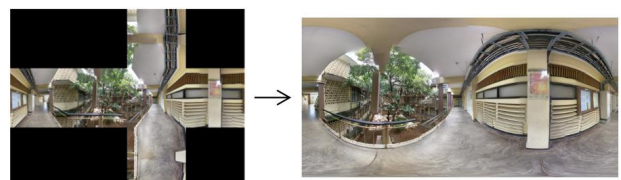


Fig. 5. Example of CMP to ERP Conversion.

그림 5. CMP에서 ERP 변환 예

추출한 프레임이 ERP로 변환되었거나 확인이 되었으면 프레임의 크기를 줄이고 각 실감형 콘 리사

이즈를 실시한다. 실감형 콘텐츠는 작게는 HD급(1280×640) 영상에서 크게는 8K UHD급(7680×4320)을 지원하기 때문에 동일한 해상도로 리사이즈를 해야한다. 영상의 화질이 크면 클수록 데이터셋의 크기도 커지기 때문에 최대한 줄이기 위해서는 작은 크기로 리사이즈를 해야 한다. 리사이즈 크기는 HD급(1280×640)으로 한다.

Table 1. Frame(Image) Resize.

표 1. 프레임(이미지) 리사이즈

Quality	Value
HD	1280×640
FHD	1920×960
QHD	2560×1280
4K UHD	3840×1920
8K UHD	7680×4320
Resize	1280×640

다음으로 객체 추출을 위한 영역을 지정한다. 전 방향을 표시하는 ERP에서 왜곡 영역이 가장 심한 위, 아랫부분을 제외한 나머지 부분에서 객체를 추출한다. 객체 식별 영역 선정 방법은 구 형상(Sphere)인 프레임을 평면 형상으로 펼치고 상하 45도 기준으로 4분할하고 좌우 90도를 기준으로 4분할하여 분할된 각각의 영역에 대해 왜곡 보정을 수행한 후, 평면 형상의 되도록 분할된 영역을 순서대로 병합하여 상하기준 45도 내지 -45도에 해당하는 중앙 영역을 객체 식별 영역으로 선정한다.

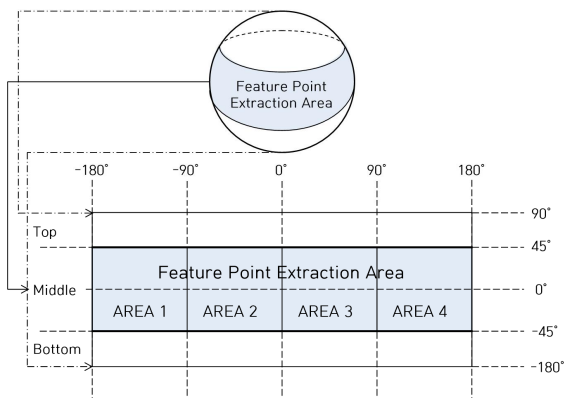


Fig. 5. Object Identification Area Range of ERP.

그림 5. ERP의 객체 식별 영역 범위

영역이 선정되고 리사이즈 된 HD급 프레임에서 특징점 추출을 위한 선택 영역에서 4개 영역 픽셀은 Table 2와 같다.

Table 2. Specify Area.

표 2. 영역 지정

Quality	AREA 1		AREA 2		AREA 3		AREA 4	
	Row	Column	Row	Column	Row	Column	Row	Column
Rato	0-90	30-70	91-180	30-70	181-270	30-70	271-360	30-70
HD	1280×640							
	0-30	193-448	321-640	193-448	641-960	193-668	961-1280	193-448

특징점 추출 과정에서는 먼저 딥 러닝을 이용하여 객체를 추출한다. 객체를 추출하기 위하여 ImageNet에서 제공하는 이미지를 데이터셋으로 1,000개의 카테고리 코드로 객체를 학습하였다. 객체 추출 과정은 Fig. 6과 같다.



Fig. 6. Object Recognition Using YOLO.

그림 6. YOLO를 이용한 객체 인식

추출 과정에서 바운딩박스가 생성되며 객체가 추출되게 된다. 각 프레임에서 여러 객체가 추출될 수 있다. 추출한 객체에서 ID, 해당 프레임 내에서 해당 객체가 차지하는 면적 비율 정보 및 해당 객체에서 추출된 인접한 특징점간 벡터 정보가 포함된 원본 데이터셋을 복수 개의 프레임별로 각각 생성하는 데이터셋이 생성한다.

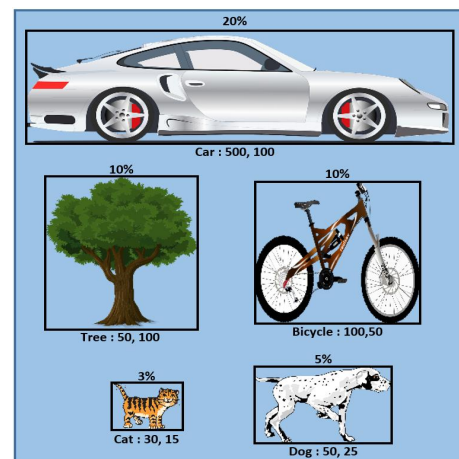


Fig. 7. Object Extraction for Dataset Creation.

그림 7. 데이터셋 생성을 위한 객체 추출

생성된 바운딩박스의 각 모서리를 특징점으로 하여 각 객체 마다 복수의 특징점을 추출한다(a). 그리고 K-means 알고리즘을 이용하여 추출된 특징점들을 군집화하고 주성분 분석 알고리즘을 이용하여 군집화된 특징점들의 차원을 감소하여 인접한 벡터 정보를 추출한다(b). 특징점 간의 관계에 대한 벡터의 내적을 구하는 방법은 식(1)과 같다.

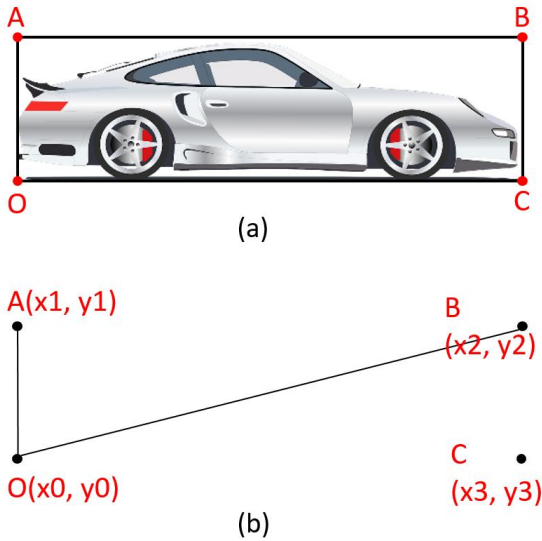


Fig. 8. (a) Feature Point Display (b) Coordinate Display.
그림 8. (a) 특징점 표시 (b) 좌표 표시

$$\vec{OA} \cdot \vec{OB} = \vec{a} \cdot \vec{b} = |\vec{a}| * |\vec{b}| * \cos\theta \quad (0 \leq \theta \leq \pi) \quad (1)$$

식별된 각 객체의 입력된 영상에서 프레임 간 변화율이 설정 비율 이상인 프레임을 설정값에 따라서 추출하되, 2D 뷰포트 형태의 프레임으로 추출한다.

IV. 실험 및 결과

본 논문에서 제안한 딥 러닝을 이용한 실감형 콘텐츠의 특징점 추출 및 식별을 위하여 다음과 같은 PC를 구비하여 실험하였다.

Table 3. PC Specification.

표 3. PC 사양

	Specification
CPU	Intel(R) Core(TM) i7-6700 3.40GHz
RAM	8 GB
HDD	512 GB
NIC	Ethernet 100 Mbps * 1 Port
OS	Windows 10 Pro 64bit

또한, 실험에 사용할 실감형 콘텐츠는 유튜브의 실감형 콘텐츠를 사용하였으며, 총 200개의 콘텐츠로 콘텐츠의 길이는 짧게는 약 1분에서 길게는 약 5분의 콘텐츠로 구성되어 있다.



Fig. 9. 360-degree Immersive Contents Used in Experiment.
그림 9. 실험에 사용될 360도 실감형 콘텐츠

200개의 실감형 콘텐츠를 이용하여 특징점 추출과 식별을 확인하기 위하여 각 콘텐츠 당 특징점 추출하고 추출된 특징점으로 같은 콘텐츠로 식별되는지 실험을 하였다. 본 논문에서 제안하는 방법을 (A), 기존에 제안되었던 방법을 (B)로 표시하였다. (B)에서 제안된 특징점 추출방법은 ORB 알고리즘을 이용하여 특징점 추출과 특징점 식별 및 인식속도를 측정하였다.

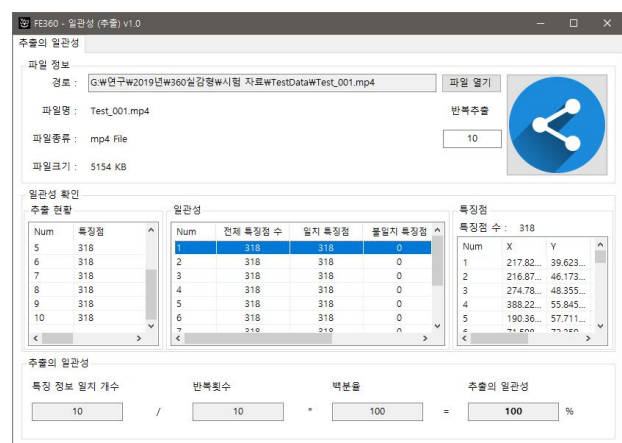


Fig. 10. Feature Point Extraction Process for Immersive Contents.

그림 10. 실감형 콘텐츠 특징점 추출 과정

Test_001부터 Test_010의 각 실감형 콘텐츠의 10회 실험 동안 추출된 특징점 평균 수와 평균 시간 데이터셋은 Table 4와 같다.

Table 4. Dataset.

표 4. 데이터셋

Content Name	(A)		(B)	
	Feature Point	Extraction Time(ms)	Feature Point	Extraction Time(ms)
Test_001	318	216.8	3,753	192.0
Test_002	217	155.6	1,353	142.6
Test_003	282	191.1	8,648	202.1
Test_004	255	177.4	4,682	151.6
Test_005	219	158.4	1,368	172.4
Test_006	286	194.4	8,648	192.1
Test_007	114	97.3	5,212	101.5
Test_008	80	76.3	4,987	100.7
Test_009	263	185.5	5,789	147.5
Test_010	136	106.9	9,878	114.4
Average of 200 Extraction Time		155.97	-	151.12

또한, 각 특징점에 대한 객체들의 X, Y의 값이 벡터의 내적 값으로 저장되었다. 이렇게 추출된 데이터셋을 바탕으로 동일한 콘텐츠로 식별되는지 총 10회(1회, 200개 x 1번)를 반복하여 실험하였으며, 식별률은 식(2)로 하여 계산하였고, 결과 값은 Table 5와 같다. 200개의 실감형 콘텐츠의 평균 식별률도 표시하였다.

$$\frac{\sum \text{인식결과갯수}}{\sum \text{반복시험횟수}} * 100 \quad (2)$$

Table 5. Identification Rate and Speed.

표 5. 식별률 및 속도

	(A)		(B)	
	Identification Rate(%)	Identification Speed(s)	Identification Rate(%)	Identification Speed(s)
Average of 200	100	1.1175	93	1.5457

Test_001에서 Test_200, 총 200개의 실감형 콘텐츠의 특징점 추출 및 식별 실험 결과 특징점 추출 과정에서 특징점 추출 속도는 두 과정이 비슷하였으나, 특징점 추출 개수가 (A)가 더 적어 100%의 식별률을 확인할 수 있었으며, 특징점 추출 개수가 많은 (B)는 식별률이 (A)보다 다소 떨어진 것을 확인할 수 있어 본문에서 제안하는 방법이 좀 더 특징점 추출에서 유리하다는 것을 확인할 수 있다.

IV. 결론

본 논문에서는 실감형 콘텐츠의 ERP의 특징인 왜곡이 가장 심한 위, 아랫부분을 제외하여 중간 영역에서 객체 식별 영역을 지정하고, 딥 러닝 기술을 적용한 객체 추출과 벡터 내적을 이용한 특징점을 식별하여 특징점을 추출하는 방법을 제안하였다. 실험을 통하여 특징점 추출의 일관성을 확인하였으며, 벡터 내적을 이용한 데이터셋 추출과 그 데이터셋을 이용한 실감형 콘텐츠 식별에서 정확한 식별 결과의 일관성을 입증하였다. 이 방법은 기존 방법보다 빠른 특징점 추출 속도를 갖는 잇점이 있다.

본 논문에서 제안하는 특징점 추출 및 식별 방법은 해상도를 변형시키거나 다른 프로젝션 타입 사용 등 여러 변형된 콘텐츠에서 강인성을 지니며 향후 실감형 콘텐츠에서 적용 가능한 저작권 보호 기술이 될 수 있을 것이다.

향후 연구로서 추출된 특징점을 기반으로 한 강인성 테스트와 불법복제 판단을 위한 비교 등의 세부적 연구가 필요하다고 할 수 있다.

References

[1] H. W. Chun, M. K. Han, and J. H. Jang, "Application trends in virtual reality," 2017 *Electronics and Telecommunications Trends*, 2017.

[2] J. Y. Kim, "Design of 360 degree video and VR contents," *Communication Books*, 2017.

[3] R. Kijima and K. Yamaguchi, "VR device time Hi-precision time management by synchronizing times between devices and host PC through USB," *IEEE Virtual Reality(VR)*, 2016. DOI: 10.1109/VR.2016.7504723

[4] Y. S. Ho, "MPEG-I standard and 360 degree video content generation," *Journal of Electrical Engineering*, 2017.

[5] W16824, Text of ISO/IEC DIS 23090-2 Omnidirectional Media Format (OMAF).

[6] Y. H. Won, J. S. Kim, B. C. Park, Y. M. Kim and S. Y. Kim "An Efficient Feature Point Extraction

Method for 360° Realistic Media Utilizing High Resolution Characteristics,” *Journal of The Korea Society of Computer and Information*, vol.24, no.1, pp.85–92, 2019.

DOI: 10.9708/jksci.2019.24.01.085

[7] B. C. Park, J. S. Kim, Y. H. Won, Y. M. Kim and S. Y. Kim “An Efficient Feature Point Extraction and Comparison Method through Distorted Region Correction in 360-degree Realistic Contents,” *Journal of The Korea Society of Computer and Information*, vol.24, no.1, pp.93–100, 2019.

DOI: 10.9708/jksci.2019.24.01.093

[8] W. J. Ha and K. A. Sohn, “Image classification approach for Improving CBIR system performance,” *2016 KICS Conf. Winter*, pp.308–309, 2016.

DOI: 10.7840/kics.2016.41.7.816

[9] H. J. Jung and J. S. Yoo, “Feature matching algorithm robust to viewpoint change,” *J. KICS*, vol.40, no.12, pp.2363–2371, 2015.

DOI: 10.7840/kics.2015.40.12.2363

BIOGRAPHY

Byeongchan Park (Member)



2015 : BS degree in Dept. of Computer Engineering, The Academic Credit Bank System

2018 : MS degree in Dept. of Computer Science and Engineering, Soongsil University.

2018~current : PhD degree in Dept. of Computer Science and Engineering, Soongsil University.

Sayoung Jang (Member)



2018 : BS degree in Dept. of Computer Engineering, The Academic Credit Bank System

2019~current : MS degree in Dept. of Computer Science and Engineering, Soongsil University.

Injae Yoo (Member)



2017 : BS degree in Dept. of Software Engineering, The Cyber University of Korea

2019~current : MS degree in Dept. of Computer Science and Engineering, Soongsil University

2015~current : Senior Researcher of Research Institute, Beyondtech Inc.

Jaechung Lee (Member)



1996 : BS degree in Dept. of Computer Science and Engineering, Seoul National University of Science and Technology

2017~current : Head of Research Institute, Beyondtech Inc.

Seok-Yoon Kim (Member)



1980 : BS degree in Dept. of Electrical engineering, Seoul National University

1990 : MS degree in Dept. of ECE, University of Texas at Austin.

1993 : PhD degree in Dept. of ECE, University of Texas at Austin.

1982~1987 : Research Member, ETRI

1993~1995 : Senior Staff Engineer, Motorola Inc, TX

1995~current : Professor, Dept. of Computer Science and Engineering, Soongsil University

Youngmo Kim (Member)



2002 : BS degree in Dept. of Computer Engineering, Daejeon University.

2004 : MS degree in Dept. of Computer Engineering, Daejeon University.

2011 : PhD degree in Dept. of Computer Engineering, Daejeon University.

2014~current : Professor, Dept. of Computer Science and Engineering, Soongsil University