

A Visitor Study of The Exhibition of <David Hockney> Using Big Data Analysis which reflects viewing experiences

Ji-Su Kang*, Bo-A Rhee**

*Master Program Student, Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul, Korea

**Professor, College of Art & Technology, Chung-Ang University, Anseong, Korea

[Abstract]

This study aims to analyze the images of Instagram posts and to draw implications regarding the exhibition of <David Hockney>. This study collects and crawl 24,295 images from Instagram posts as a dataset. We use the Google Cloud Vision API for labeling the images and a total of 212,567 clusters of labels are finally classified into 9 categories using Word2Vec. The categories of museum spaces, photo zone, architecture category are dominant along with people category. In conclusion, visitors curate their experiences and memories of physical places and spaces while they are experiencing with the exhibition. This result reproves the results of previous studies which emphasize a sense of social presence and place making. The convergent approach of art management and art technology used in this study help museum professionals have an insight on big data based visitor research on a practical level.

▶ **Key words:** Big Data, Instagram, Visitor experience, Google Cloud Vision API, Word2Vec

[요 약]

본 연구는 <데이비드 호크니>전의 인스타그램 게시물의 이미지를 분석하고, 이에 대한 시사점을 도출하는데 목적을 두었다. 본 연구는 인스타그램 게시물로부터 24,295개의 이미지를 크롤링했으며, 구글 클라우드 비전 API를 활용해서 라벨링을 진행했다. Word2Vec을 통해 총 212,567개의 라벨이 9개로 유형화되었으며, 사람 라벨 유형과 함께 미술관 공간, 포토존, 건축물 등의 빈도수가 높게 제시되었다. 결론적으로 관람객은 전시를 경험하면서 물리적 장소와 공간에 대한 경험과 기억을 큐레이팅했다. 이 결과는 사회적 현존감과 장소 만들기를 강조했던 선행 연구의 결과를 재 입증해 주었다. 본 연구에서 사용된 예술경영과 예술 공학의 융합적 접근방법론은 실무적 차원에서 박물관 및 미술관 전문 인력이 빅 데이터 기반 관람객 연구에 대한 통찰력을 획득하는데 도움을 줄 것으로 기대한다.

▶ **주제어:** 빅 데이터, 인스타그램, 관람 경험, 구글 클라우드 비전 API, Word2Vec

-
- First Author: Ji-Su Kang, Corresponding Author: Bo-A Rhee
 - *Ji-Su Kang (kritik.sowieso@gmail.com), Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University
 - **Bo-A Rhee (boa.rhee@gmail.com), College of Art & Technology, Chung-Ang University
 - Received: 2021. 12. 27, Revised: 2022. 02. 04, Accepted: 2022. 02. 04.

I. Introduction

오늘날 관람객은 소셜 네트워크 서비스(Social Network Service, 이하 SNS)를 통해 관람 경험을 기록 및 공유한다[1]. SNS는 미술관의 핵심적인 가치와 철학을 전달해 주고[2], 미술관의 활동 및 프로그램을 확산시키고[3], 새로운 관람객에게 접근할 수 있으며[4], 미술관과 온라인 커뮤니티간의 대화를 촉진시킴으로써[5] 관람 주체의 능동적 특성을 증진시킨다 [6].

2010년에 설립된 인스타그램은 2021년에 약 14억 3,500만 명의 사용자를 보유하고 있다[7]. 전체 사용자 중 35세 미만이 71%를 차지하는데[8], 미술관의 주요 관람객으로서 이들 연령 계층은 인증 샷을 통해 관람 경험을 타인과 공유하면서 구전효과를 발생시키며[9], 잠재 관람객에게 전시에 대한 흥미와 관람 동기를 유발시킨다[10]. 이는 인스타그램이 다른 SNS에 비해 문화예술 향유의 경험이나 미적 가치를 직관적으로 제시할 수 있다는 것을 의미할 뿐만 아니라[2], 젊은 신규 관람객 유치를 위한 마케팅 도구로서 인스타그램의 잠재력을 입증해준다[11].

뮤지엄 3.0 시대의 도래와 함께 미술관은 SNS를 기반으로 한 디지털 마케팅 전략을 구사하고 있으며[12], 이를 통해 관람 경험에 대한 통찰력을 획득하고 있다[13]. 실제로 SNS 팔로워 순위가 상위권을 점유한 박물관 및 미술관이 관람객 수에서도 상위권을 차지했다[14]. 한편 학술 분야에서 SNS의 영향력에 대한 연구가 이루어지고 있는데, 대부분 빅 데이터로써 이미지보다 해시태그나 지오태그 등 단어 기반의 분석 방법론에 치중되어 있다.

본 연구에서는 서울시립미술관에서 개최되었던 <데이비드 호크니>전의 인스타그램 게시물에 이미지를 빅 데이터로 분석해서, 이미지에 투영된 관람 경험의 특성을 도출하는데 그 목적을 두고 있다. 이를 위해 관람객이 생성한 <데이비드 호크니>전 인스타그램 계정으로부터 데이터를 수집 및 정제했다. 또한 구글 클라우드 비전 API를 연동하여 이미지 데이터를 분석하고, 라벨 유형화를 통해 관람 경험에 대한 시사점을 이끌어내고자 한다. 학술적 관점에서 설문조사나 관찰 등의 전통적인 관람객 연구방법론과는 달리, 본 연구에서는 예술경영과 예술 공학이 융합된 새로운 연구방법론을 기반으로 방대한 관람객 데이터를 분석한다는 차별성을 지닌다. 또한 실무적 관점에서 본 연구는 데이터 기반의 미술관 경영 및 관람객 연구방법론, 그리고 미술관 전문 인력이 데이터를 수집 및 분석해서 실무에 적용하는 방법에 대한 실질적인 이해를 제공해준다.

II. Preliminaries

1. Theoretical background

1.1 Overview of the exhibition <David Hockney>

2019년 서울시립미술관에서 개최된 <데이비드 호크니(2019.3.22.-8.4)>전은 작가의 아시아 지역 첫 대규모 개인전으로, 총 133점을 선보였다. 이 전시는 미술관 역사상 최대 관람객 규모인 총 375,350명이 관람했으며[15], 전시 만족도 또한 동년에 개최된 전시 가운데 가장 높은 점수(89.5점)를 획득했다[16]. 또한 상기 전시의 경우, 전시실 재입장과 내부 촬영이 불가능했었지만[17], 당해 전시 관람 평균 소요 시간(53.6분)에 비해 상대적으로 높은 체류 시간(평균 92.6분)을 나타냈다.

1.2 Big data technology and museums

빅 데이터 기술은 데이터의 생성, 수집, 저장, 처리, 분배 전달 등을 모두 포괄한 디지털 기술 개념이 내재해 있으며[18], 방대한 원 데이터(raw data)의 관계성을 분석, 고객과 시장에 대한 새로운 통찰력을 제공해준다. 상술한 빅 데이터의 기술적 특성은 미술관 경영에도 유의미한 영향력으로 작용하고 있다. 예를 들어, 오랫동안의 관람객 연구를 통해 관람객에 대한 레거시 데이터를 보유하고 있는 영국의 국립박물관(British Museum) 및 국립미술관(National Gallery)은 빅 데이터 기반의 데이터 수집 및 처리 방법을 적극적으로 수용했다[19].

국립미술관은 머신러닝 알고리즘을 기반으로 관람객 수 예측, 관람 행태, 전시 관람객 시뮬레이션 등 디지털 전략에 대한 통찰력 증진에 집중하고 있다[20]. 한편 국립박물관은 실시간 데이터를 기반으로 박물관 경영에 대한 의사 결정을 실행하기 위해, 트립어드바이저(Tripadvisor)의 리뷰와 평점을 자연어 처리 기술 및 머신러닝 알고리즘으로 도출된 관람객 군집 현상을 비롯한 주요 연구주제를 관람객 연구의 평가도구로 사용하고 있다[20]. 이외에도 메트로폴리탄미술관, 미국자연사박물관(the American Museum of Natural History), 현대미술관(MoMA), 테이트(Tate) 등은 디지털 미디어 조직을 설치하고, 데이터 분석 전문가를 영입해서 빅 데이터 기반의 중장기 경영 전략 수립 및 관람객 데이터 분석을 실행해오고 있다.

빅 데이터 기술의 확산에 따라, 박물관 및 미술관 관련 학술 영역에서도 이와 관련된 연구가 활성화되고 있다. 국내외 미술관 해시태그 관련 선행연구 결과를 종합해보면, 전시 장소, 지역 정보 등 장소성과 관련된 인스타그램 해시태그 유형이 높은 비율을 차지했고[1, 21-23], 온라인

커뮤니케이션 도구로써 해시태그의 잠재력이 확인되었다. 또한 인스타그램에 전시 관련 이미지를 게시하는 행위는 자기표현의 수단의 역할뿐만 아니라 포스팅에 포함된 댓글이나 좋아요 등의 기능을 통해 직관적 커뮤니케이션과 공유를 가능케 한다는 시사점도 제시되었다[9].

하지만 미술관 및 전시 인스타그램 빅 데이터 연구는 해시태그, 지오태그 등 단어 빅 데이터 기반의 연구의 비중이 높아 이러한 선행 연구방법론은 이미지 중심인 인스타그램의 메타 데이터를 분석하는데 한계로 지적된 바 있다 [23]. 뿐만 아니라, Table 1에 제시된 것처럼, 선행 연구의 인스타그램 이미지 데이터 수집 규모는 최소 99장에서 최대 390장 수준이었다. 이에 대해 선행 연구에서는 수기 이미지 데이터 수집 및 분류 방법의 한계로 이미지 표본 규모가 작았다는 점과 이미지에 포함된 관람 경험의 특성을 다양하게 도출하지 못했던 점을 언급하기도 했다[9]. 또는 만 장 이상의 대용량 이미지 수집을 진행한 선행 연구의 경우, 4K stogram 등, 유료 응용 소프트웨어를 사용해서 진행되었는데[1], 이는 미술관 실무 차원에서 이미지 분석에 따른 비용이 추가적으로 요구된다는 한계를 지닌다.

Table 1. Previous research methodology and classification categories

| researcher | research methodology |
|------------------------------|--|
| A. Weilenmann, et al. (2013) | Manually downloading and classifying 123 images and 99 images of interviewees based on geotags |
| K. Budge, et al. (2018) | Using "If This Then That(IFTT)" program, crawling 390 images and metadata based on the Museum of Contemporary Art Australia geotag |
| M. Kim, et al. (2019) | Manually downloading and classifying 200 images of visitor's instagram by exhibition type |
| B. Rhee, et al. (2021) | Using the '4k stogram' application, 14,400 posting crawling and labeling based on Google Cloud Vision API |

이에 본 연구에서는 응용 프로그램이 아닌, 파이썬 셀레늄 기반 크롤링 예제 코드를 사용을 시도하며, 기존의 수백장 데이터 셋의 수준이 아닌 인스타그램 이미지 최소 만 장 이상의 이미지 데이터 셋을 구축하여 분석하고자 한다. 이를 통해 선행 연구보다 세분화 된 관람객 이미지 유형 도출과 이를 기반으로 한 관람 경험의 특성을 도출하는 데 그 목적을 두고 있다.

III. The Proposed Scheme

3. Research Model

3.1 Research Methodology

본 연구에서는 다음과 같은 세 가지의 연구 문제를 설정했다: 인스타그램 이미지 데이터에는 관람의 경험적 특성이 투영되었는가?: 구글 클라우드 비전 API와 워드 임베드 모델 Word2Vec은 관람 경험의 특성을 유형화하는데 적합한가?: 예술경영과 예술 공학이 융합된 분석방법론은 박물관 및 미술관 전문 인력이 관람객 연구의 실무에 적용하기에 용이한가? 상술한 연구 문제를 규명하기 위해 본 연구는 Fig. 1에서 제시된 것과 같이 세 가지의 단계로 구성된다.

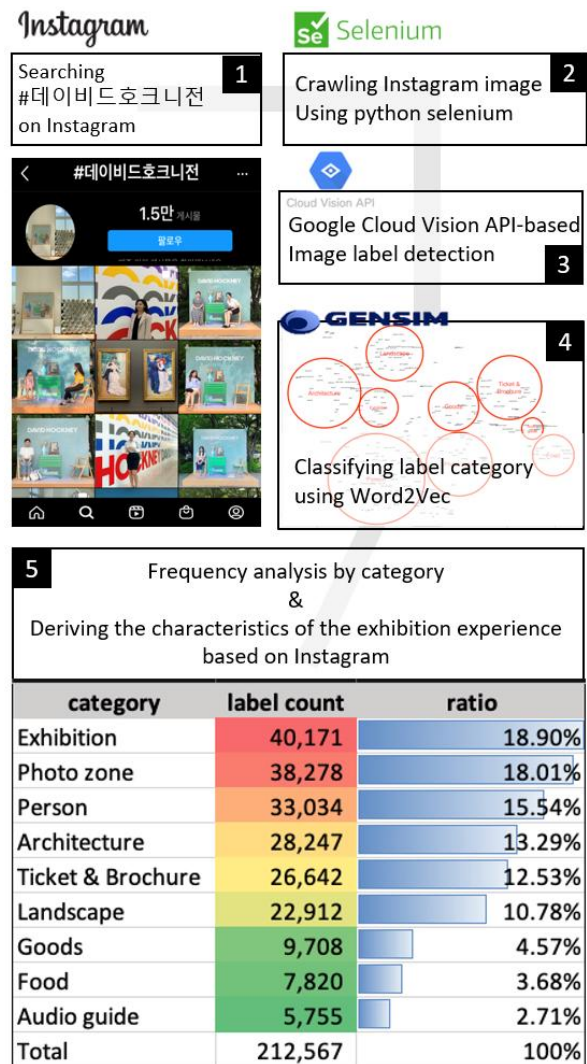


Fig. 1. Research process and methodology

첫 번째 단계는 데이터 수집이다. 본 연구에서는 데이비드 호크니 전시(2019.3.22.- 08.04.) 기간에 사용자 계정(#데이비드호크니전)의 인스타그램에 업로드 된 게시물을 크

를링했다. 빅 데이터 기반의 관람객 인스타그램 데이터 분석을 위해서 Fig. 2와 같이 맥 운영체제(Mac OS)에서 파이썬 버전 3.6.9를 설치하여 코드 편집기로 아이들(IDLE)과 주피터 노트북(Jupyter notebook)으로 환경을 설정했다. 이후 셀레늄 패키지 기반의 크롤링 코드를 활용해서 인스타그램 #데이비드호크니전 검색 결과 게시물을 자동 수집했다. 최종적으로 이미지 24,295장의 데이터 셋이 구축되었다.

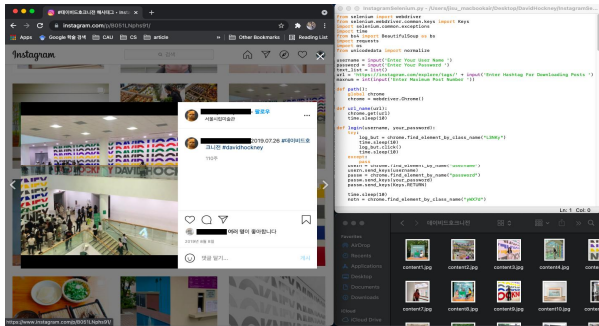


Fig. 2. Instagram crawling

두 번째 단계는 데이터 정제 및 분류이다. 이를 위한 컴퓨터 환경은 구글 클라우드 플랫폼 콘솔 계정을 만든 후, API 키를 JSON 형식으로 받아 로컬 저장소와 연동했다. 이후 총 24,295개 이미지에 대하여 클라우드 비전 API 라이브러리를 활용, 라벨 감지(Label detection)를 요청한다. 라벨 감지 모델은 이미지에서 감지된 객체에 단어 레이블링을 자동으로 처리한다. Fig. 3에서 보이는 것처럼 이미지 1개 당 최대 10개 라벨을 감지될 수 있으며, 구글 클라우드 비전 모델이 이미 학습한 자연어에 기반 하여 명사, 동사, 형용사 단어로 라벨 감지 신뢰도와 함께 제공된다.

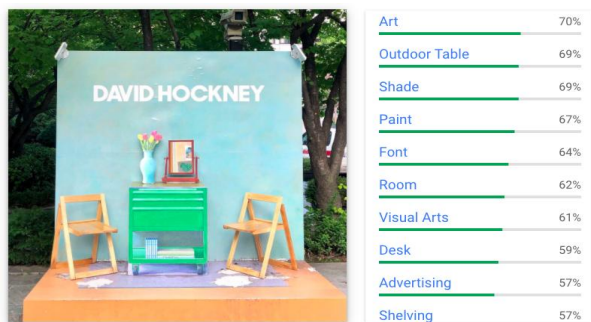


Fig. 3. The result of label detection using Google Cloud Vision API

json 형식으로 출력되는 라벨링 결과는 텍스트(txt)파일 형식으로 변환되어, 라벨의 빈도 분석을 위한 자료로 활용되었다. 의미 기반의 분류는 파이썬 겐심(Gensim) 패키지를 사용, 머신러닝 기반 자연어 임베딩 모델에 해당하는

Word2Vec으로 진행되었다. Word2Vec은 단어의 벡터 간 거리를 비교해 거리가 근접한 것은 의미가 유사하다는 가정을 전제로 학습하는 모델이다[24]. Fig. 4의 시각화 결과에서 나타난 바와 같이, 임베딩 조건 값을 기본으로 하여 전체 라벨을 임베딩 하였을 때, 각 라벨의 의미를 기반으로 유사도가 높은 라벨들이 군집 되었다. 라벨의 간격이 가까울수록 의미 유사도가 높은 것이며, 라벨의 간격이 멀수록 의미 유사도가 낮다는 것을 의미한다. 각 점에는 라벨이 기입되어 있으며, 20만개 이상의 라벨이 의미 기반으로 군집되어 의미 유사도가 큰 라벨들은 군집도가 높다.

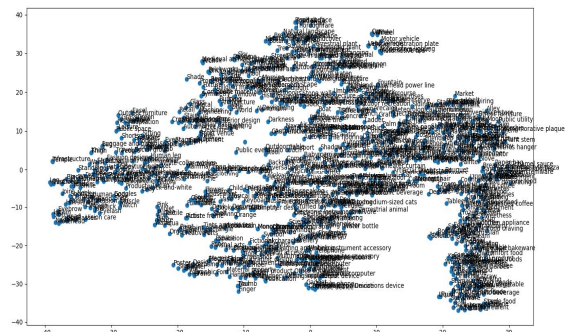


Fig. 4. the first visualization result using Word2Vec

본 연구는 라벨의 군집화 결과의 가시성을 높이기 위해 Word2Vec의 최종 임베딩 조건 값을 조정했다. 차원은 50, 단어의 최소 빈도수는 100 이상으로 설정하여, 전체 이미지 데이터셋을 기준으로 100 미만으로 도출된 라벨은 제외했다.

데이터 분석이 이루어지는 마지막 단계에서는 레이블의 유형과 하위 라벨의 빈도와 그 빈도에 대한 시사점이 도출되고, 인스타그램 게시물에 제시된 관람 경험 공유의 특성과 의미가 종합적으로 분석된다. 라벨 유형화의 경우, Table 2에서 보여지는 것처럼 본 연구에서는 예술전시, 전시물, 건축물, 셀피(selfie), 음식, 사람, 풍경 등의 유형화를 다룬 선행 연구의 유형화 방식이 일부 차용되었다[1]. 인스타그램 이미지 빅 데이터 기반 <우미의 세포들>특별전 관련 연구[25]의 경우, 전시장 내부 촬영이 가능했기 때문에 라벨 빈도 분석 결과 ‘사람과 전시물’, ‘전시물’ 유형의 빈도가 높았다. 관람객이 전시물과 상호작용을 통해 직접 촬영할 수 있는 인스타그램버전 전시 특성이 전시 만족도에 긍정적 영향을 미치는 것이 입증된 바 있다.

Table 2. Previous research methodology and classification categories

| researcher | category |
|------------------------------|---|
| A. Weilenmann et al. (2013) | subject choice, aesthetics, captioning creating cover pages and collections |
| K. Budge & A. Burness (2018) | objects only, people and objects, people only. others |
| P. Vitale et al. (2020) | exhibition space, exhibited object, selfie, close-up. figure objects |
| B.A. Rhee et al. (2021) | art exhibition, artifact, architecture, selfie, food, human body, landscape |

IV. Finding and Discussion

4.1 Image data analysis results

구글 클라우드 비전 API를 연동하여 라벨 디텍션을 진행한 결과, 이미지 당 평균 라벨 개수는 약 8.7개로, 총 212,567개 라벨이 도출되었다. 라벨의 의미에 따라 Fig. 5와 같이 건축물, 미술관 공간, 풍경, 사람, 굿즈, 포토존, 티켓 & 브로슈어, 오디오 가이드, 음식과 관련된 라벨 벡터가 가깝게 모여 있는 것이 확인되었다. 이 결과는 ‘사람과 전시물’, ‘전시물’ 유형의 빈도가 높았던 선행연구와는 큰 차이를 드러냈는데, 그 이유는 <데이비드 호크니>전에서는 전시장 및 전시물 촬영 불가 규정으로 인해 선행연구의 라벨 유형에서 예술전시와 전시물이 제시되지 않았기 때문이다.

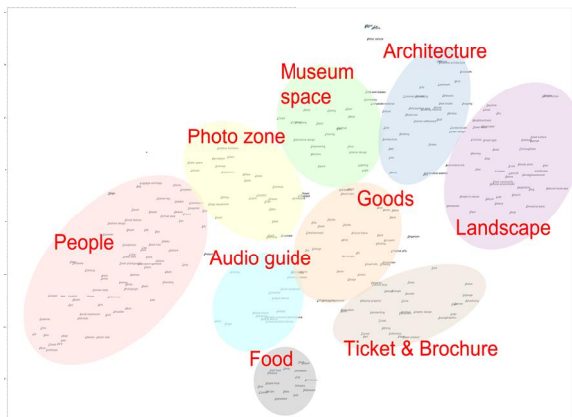
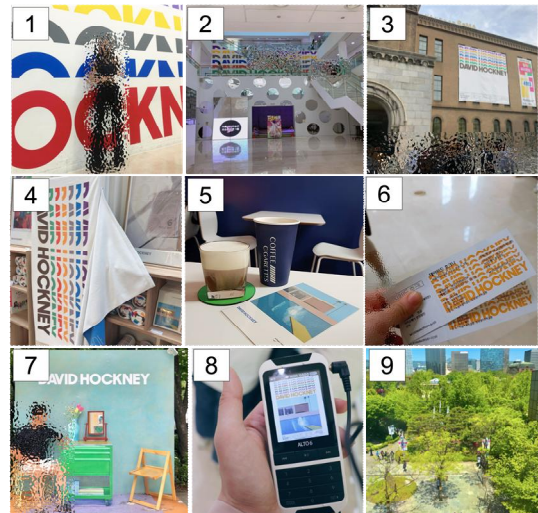


Fig. 5. the final visualization result using Word2Vec

상술한 라벨의 군집화 결과를 기반으로, 본 연구에서는 Fig. 6의 예시 이미지처럼 총 9개의 주요 라벨 유형이 확정되었다.



1) People / 2) Museum Space / 3) Architecture
 4) Goods / 5) Food / 6) Ticket & Brochure
 7) Photo zone / 8) Audio guide / 9) Landscape
 *Blurred to protect users' personal information

Fig. 6. Examples of image types classified by label clustering

라벨에 대한 유형별 빈도 분석 결과에 의하면, 미술관 공간(museum space, 18.90%)이 가장 높은 비율을 점유했다. 이 결과는 인스타그램 해시태그를 분석한 선행연구와 일치한다. 관람객들은 자기표현의 수단으로써 물리적인 전시 장소 및 공간에 대한 경험과 기억을 큐레이션 하고 관계를 형성하고 있다는 것을 의미한다[26]. Fig. 7의 그래프에서 보듯, 미술관 공간 유형이 가장 높은 빈도 결과로 제시되었다. 각 유형에 해당하는 세부 라벨의 빈도를 정리한 Table 3에서 제시된 바와 같이 공간에 표현된 장식물을 ‘art(n=5,435)’로 인식했고, 공간의 구조와 관련된 라벨(flooring, wall)들도 비교적 상위를 점유했다.

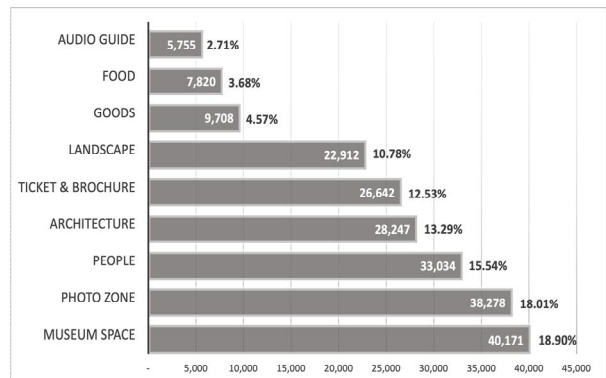


Fig. 7. Frequency and ratio by image type

Table 3. Top 5 Labels in each label category

| | | | | |
|----------------------|-------------------------|------------------------|-----------------------|----------------------------|
| museum space (18.9%) | photo zone (18.01%) | people (15.54%) | architecture (13.29%) | ticket & brochure (12.53%) |
| art (5,435) | plant (5,558) | sleeve (2,869) | building (6,388) | product (1,821) |
| rectangle (5,052) | leisure (4,044) | street fashion (2,819) | window (4,769) | publication (1,740) |
| event (4,533) | grass (2,859) | travel (2,499) | facade (3,552) | Thumb (989) |
| flooring (2,803) | advertising (2,624) | smile (2,111) | wood (3,029) | paper (894) |
| wall (1,925) | interior design (2,079) | gesture (2,010) | architecture (1,168) | Finger (819) |

| | | | |
|--------------------|---------------------------|-------------------|-------------------------|
| landscape (10.78%) | goods (4.57%) | food (3.68%) | audio guide (2.71%) |
| tree (5,203) | material property (1,827) | tableware (1,204) | Electric blue (2,403) |
| sky (4,238) | publication (1,740) | food (984) | gadget (414) |
| city (2,672) | book (412) | ingredient (912) | technology (312) |
| cloud (1,706) | office supplies (135) | recipe (737) | display device (255) |
| natural land (585) | writing implement (114) | cuisine (624) | electronic device (238) |

미술관 공간과 관련, 건축물(architecture)은 라벨 유형 가운데 네 번째로 빈도수(13.29%)가 높았으며, 건축물 자체(building)와 건축물 구성요소(window, facade, wood) 등이 상위 빈도를 나타냈다.

특히 미술관 공간 이미지 유형에서는 Fig. 8에서 보이는 바와 같이 관람객들이 군집된 모습이 'event(n=4,533)'로 인지되었으며, 이와 관련해서 'crowd(n=399)'도 함께 라벨로 생성되었다. 상술한 바와 같이 이 두 가지 라벨의 경우 대부분 미술관 내부 공간에서 입장을 기다리고 있는 관람객의 집객 상황이 담긴 이미지로부터 생성된 것이었다. 또한 서울시립미술관의 경우, 출생연도가 1980년대부터 2000년대 사이인 밀레니얼(millennials) 세대와 Z세대(generation Z)를 통칭하는 MZ 세대[27]가 관람객의 40% 이상을 차지하고 있었는데, 대부분 이미지에서 MZ 세대 관람객의 모습뿐만 아니라, 대기 시간 동안 스마트 폰을 사용하고 있는 관람객 모습이 확인되었다.

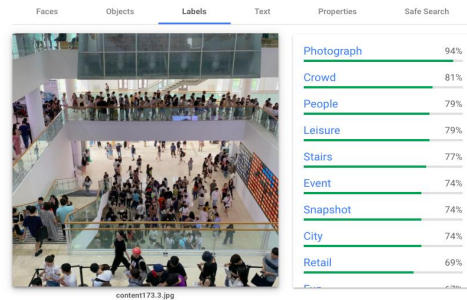


Fig. 8. Example images of the museum space

상술한 바와 같이, <데이비드 호크니>전은 내부 사진 촬영이 금지되었기 때문에, 관람객들이 전시 관람 인증 샷을 촬영할 수 있는 장소가 상대적으로 매우 한정적이었다. 전시장 내부 촬영이 불가능한 전시의 경우, 포토존의 설치가 관람객의 전시 관람 인증 및 참여를 유도할 수 있는 효과적인 장치로 작용했다. 이에 두 번째로 높은 빈도를 차지한 라벨 유형은 포토 존(photo zone, 18.01%)이었다. 미술관이 설치한 포토존뿐만 아니라, 관람객이 포토 존으로 활용한 장소는 2층 전시실 앞 벽, 티켓 부스 옆, 기념품점 앞 등 세 곳으로 나타났는데, 이 결과는 관람객이 사진을 촬영하기에 적합한 장소를 선택했다는 것을 의미한다. 포토 존 이미지 유형의 경우, 포토 존에 설치된 'plant(n=5,558)'가 가장 높은 빈도수를 나타냈으며, 이와 함께 'grass(n=2,859)'라벨도 비교적 높은 비율을 차지했다. 또한 포토 존과 관련된 라벨에서는 이미지의 속성으로 'leisure(n=4,044)'와 'advertising(n=2624)', 'sitting(n=171)', 'standing(n=777)' 등이 생성되었다. 이러한 라벨은 관람객의 자세 및 옷차림이 여가활동이나 광고에 등장하는 이미지의 속성과 유사하다고 판별되어 감지되었다.

사람 이미지(people, 15.54%)의 경우, 'sleeve(n=2,869)', 'street fashion(n=2,819)'과 같이 의상과 관련된 라벨이 높은 비율을 차지했다. 또한 관람객의 얼굴 표정(smile, n=2,111)이나 몸짓(gesture, n=2,010)과 관련된 라벨은 비교적 높게 제시되었다. 하지만, 셀피(selfie, n=10) 라벨은 유의미한 값이 도출되지 않았고, 오히려 동반 관람객이 찍어준 전시 인물 사진이 다수를 차지했다.

티켓 및 브로슈어(ticket & brochure) 라벨의 경우, 가장 높은 빈도수를 획득한 'Font(n=9,552)'는 기존의 선행 연구와 마찬가지로 제외되었다. 그 결과, 'Product(n=1,821)'와 'Publication(n=1,740)'이 높은 비율을 획득했다. 티켓 및 브로슈어 유형의 하위 라벨의 경우, 'Thumb(n=990)'와 'Finger(n=830)' 등의 속성이 감지된 것은 Fig. 9에서 보듯이 티켓이나 브로슈어를 손으로 잡고 사진을 찍었기 때문이다.

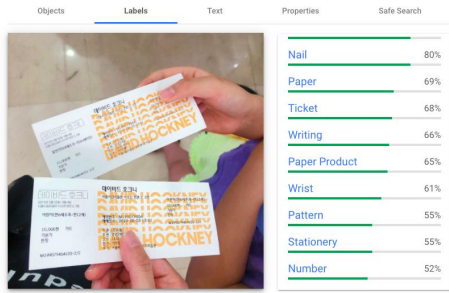


Fig. 9. The result of label detection of ticket and brochure using Google Cloud Vision API

하지만 티켓 및 브로슈어에 해당하는 군집에 속한 라벨 기준으로 Word2Vec 유사도 상위 10개 라벨을 도출한 결과, ‘Thumb(0.570)’나 ‘Nail(0.593)’, ‘Finger(0.457)’는 누락되는 현상을 확인할 수 있다. 부연하면, Word2Vec의 유사도 값은 단어 간 의미가 유사할수록 1의 값에 가까워지며, 0에 가까울수록 유사하지 않다는 의미로써, ‘Thumb’, ‘Nail’, ‘Finger’ 라벨은 Fig. 10처럼 Word2Vec의 군집화 결과와 유사도 값만을 기준으로 분류한다면 티켓과 브로슈어 유형에서 제외될 수밖에 없다. 하지만 상술한 문제는 본질적으로 구글 클라우드 비전 API의 이미지 분류 모델에 기인한 문제점이기 때문에, 본 연구에서는 선행연구 결과와 동일하게 상술한 세 가지의 라벨이 포함되었다[1, 25].

```
In [9]: Ticket_Brochure = model.vw.most_similar('Publication',
print(Ticket_Brochure)
[('Book', 0.8869412541389465), ('Stationery', 0.8602216
243743896), ('Writing implement', 0.7945601344108582),
('Office supplies', 0.7909629940986633), ('Paper', 0.77
4738073348999), ('Writing', 0.7570151686668396), ('Pape
r product', 0.7151000499725342), ('Handwriting', 0.7066
382169723511), ('Material property', 0.703041672706604)
, ('Linens', 0.6793814301490784)]
```

Fig. 10. The vectorized similarity values of publication label using Word2vec

이외에 미술관 주변의 풍경(landscape, 10.78%) 유형의 라벨의 경우, 선행연구 결과와 동일하게 ‘tree(n=5,203)’와 ‘sky(n=4,238)’, ‘city(n=2,672)’, ‘cloud(n=1,706)’ 등의 라벨이 도출되었다. 굿즈(goods, 4.57%) 유형의 라벨에서는 도록 등의 단행본과 문구류가 많았기 때문에 ‘Material property(n=1,827)’, ‘Publication(n=1,740)’, ‘Book(n=412)’, ‘Office supplies(n=135)’ 등이 상대적으로 높은 빈도로 제시되었다. 음식(food, 3.68%)의 경우 ‘tableware(n=1,204)’, ‘food(n=984)’ 순으로 라벨이 감지되었다. 관람객이 관람 경험을 인스타그램에 게시할 때, 식음료에 대한 소비 기록도 함께 게시하지만, 그 빈도가 높지 않았던 선행연구[1]와

마찬가지로, <데이비드 호크니>전에서도 음식 유형의 빈도는 낮게 도출되었다.

전체 라벨 유형 가운데 오디오 가이드 라벨은 가장 낮은 빈도를 획득했는데, 이는 <데이비드 호크니>전 관람객은 해석 매체에 대한 의존도가 낮고, 작품 감상 자체에 집중했다는 사실을 시사해 준다. 또한 오디오 가이드 유형에 포함된 하위 라벨에는 ‘Audio equipment(n=137)’뿐만 아니라, ‘Output device(n=152)’, ‘Gadget(n=414)’, ‘Mobile Phone(n=158)’ 등이 포함되었다.

V. Conclusions

본 연구는 미술관 관람객 연구 및 빅 데이터에 대해 융합적 관점에서 접근하기 위해 <데이비드 호크니>전의 인스타그램 게시물 이미지(#데이비드호크니전)를 데이터 셋으로 구축했다. 또한 구글 클라우드 비전 API를 사용해서 24,295개의 이미지가 크롤링 되었으며, Word2Vec을 기반으로 212,567개의 라벨에 대한 군집화 및 유형화가 이루어졌다. 라벨 유형화의 결과를 종합해보면, 미술관 공간, 포토존, 사람, 건축물, 티켓 및 브로슈어 등이 상위를 점유했으며, 이외에 풍경, 굿즈, 음식, 오디오 가이드 등도 라벨 유형에 포함되었다.

박물관 및 미술관 관련 인스타그램 해시태그 분석을 다른 선행연구와 마찬가지로, 인스타그램 이미지 빅 데이터 분석 결과, 전시 관람 경험의 특성이 도출되었다. 관람객은 인스타그램에 미술관 공간을 가장 많이 게시했으며, 또한 높은 비율로 포토존과 미술관 선호 공간에서의 인증 샷을 게시했다. 인스타그램을 사용하는 관람객은 자기표현의 수단으로서 전시 장소에 대한 물리적 경험과 기억을 큐레이션 하며 관계를 형성하고 있다는 것을 의미한다. 특히 관람객은 전시 관람 중 ‘장소 만들기(place making)’[26], 즉 공간과 장소에 대한 인상을 기록하고 공유함으로써, 사회적 현존감을 생성한다는 선행연구의 결과와도 일치한다[1, 10].

인스타그램 이미지 빅 데이터 분석을 통한 관람객의 관람 경험의 특성과 의미를 해석하는 작업은 기존에 미술관이 사용해 오던 관찰이나 설문조사 등의 관람객 연구방법론에 내재해 있던 문제점을 보완해줄 수 있다. 예를 들어, 관찰의 경우, 표본의 수가 매우 제약적이고, 설문조사의 경우, 인구통계학적 정보, 관람만족도, 미술관 시설 이용, 전시 정보 획득 경로 등 대부분 관람 경험의 본질을 파악하는 데는 한계를 지닌 문항들로 구성된다. 이러한 관점에서 보면, 인스타그램 이미지 빅 데이터 분석은 관찰이나

설문조사의 표본의 규모, 분석 결과의 정확도나 신뢰도, 실용성 등을 개선해 줄 뿐만 아니라 관람 행태를 비롯하여 기존의 방법론에서 획득할 수 없는 선호 장소, 관람 인증 행태 등, 관람 경험에 대한 새로운 정보를 제공해준다.

본 연구에서의 한계는 두 가지로 요약될 수 있다. 첫 번째로, 상술한 미술관 촬영 정책에 따라 전시실 내의 관람 경험에 대한 이미지 유형은 도출되지 않았다. 전시실 외 미술관 공간 및 미술관 주변 환경에서 촬영된 이미지 분석은 관람객이 전시를 인증하는 방식, 전시 관람 전과 후의 행태 등에 대한 통찰력을 제공해줌으로써, 선행 연구보다 세분화가 이루어진 관람 경험을 유형화했다. 하지만 전시실 내 관람 경험을 촬영할 수 없었기 때문에, 이미지 데이터만으로는 선호 작품이나 관람객의 작품에 대한 상호작용의 분석은 본 연구에서 제외되었다.

두 번째 한계는 구글 클라우드 비전 API에 있다. 본 연구에서는 인스타그램 이미지 데이터 셋을 구축해서 구글 클라우드 비전 API로 라벨링이 진행되었다. 이 라벨링이 진행되는 과정에서 구글 클라우드 비전 API는 '티켓 및 브로슈어'처럼 전시 관람이라는 상황이 적절하게 반영되지 않았고, Word2Vec의 군집화 결과와 유사도 값을 기준으로 라벨에 대한 판단이 이루어졌다. 이러한 한계를 극복하고 더욱 방대한 규모의 데이터를 자동으로 처리하기 위해서는 이미지 데이터뿐만 아니라 관람객 게시 글 등, 비정형의 이종 데이터 분석을 병행하여 AI 모델의 분석 결과의 정교함을 증진시키는 방안을 고려할 수 있다.

관람객 관련 빅 데이터 구축 작업은 상당한 시간과 예산이 필요하며, 구조뿐만 아니라 실무 측면에서 미술관 내부의 혁신적인 변화가 필수적으로 요구된다. 예를 들어, 관람객 연구 조직의 설치, 미술관 레거시 데이터의 확보, 빅 데이터 관련 전문 인력의 수급, 온·오프라인 데이터 수집 방법 및 경로 등 일련의 체계적인 데이터 시스템 구축이 이루어져야 한다. 문화체육관광부(2019)는 박물관 및 미술관 고객관리 선진화를 위해[28], 빅 데이터형 관람객 연구 프로그램 개발과 관람 패턴에 대한 심도 있는 빅 데이터 수집을 지원할 계획이다.

본 연구는 향후 데이터 기반의 미술관 디지털 전환을 위한 국·공립 박물관·미술관의 지능형 고객관리시스템 구축 등 미술관 관람객 연구에 대한 실무 적용에 대해 유의미한 시사점을 제공해주며, 궁극적으로 수요자인 관람객 중심의 프로그램 개발에 새로운 기회를 제공해 줄 것으로 기대한다.

ACKNOWLEDGEMENT

This research was supported by the Chung-Ang University Research Scholarship Grants in 2021

REFERENCES

- [1] B. Rhee, F. Pianzola, and G. Choi, "Analyzing the Museum Experience Through the Lens of Instagram Posts," *Curator: The Museum Journal*, 64(2), pp. 1-19, April 2021. DOI: 10.1111/cura.12414
- [2] V. G. Vrana, D. A. Kydros, E. C. Kehris, A. I. T. Theocharidis, and G. I. Kavavasilis, "Top museums on Instagram: A network analysis," *International Journal of Computational Methods in Heritage Science*, 3(2), pp.18-42, July 2019. DOI: 10.4018/978-1-7998-9020-1.ch052
- [3] M. V. Ciasullo, M. Gaeta, G. Monetta, and L. Rarità, "E-cultural value co-creation. A proposed model for the heritage management," *Proceedings of 18th Toulon-Verona International Conference*, pp.139-158, Palermo, Italy, August 2015.
- [4] A. Fletcher, and M. J. Lee, "Current social media uses and evaluations in American museums," *Museum Management and Curatorship*, 27(5), pp.505-521, September 2012. DOI: 10.1080/09647775.2012.738136
- [5] P. Capriotti, and H. P. Kuklinski, "Assessing dialogic communication through the Internet in Spanish museums," *Public relations review*, 38(4), pp.619-626, November 2012. DOI: 10.1016/j.pubrev.2012.05.005
- [6] B. Koo, "A Study on Meaning of Technology Utilization in the Field of Museums," *Korean Journal of Arts Education*, 18(1), pp.47-68, March 2020.
- [7] M. Iqbal, *Instagram Revenue and Usage Statistics(2021)*, Buisness of Apps, <https://www.businessofapps.com/data/instagram-statistics/>
- [8] S. Kemp, *TikTok hits 1 Billion Users - Faster than facebook*, Hootsuite, <https://blog.hootsuite.com/simon-kemp-social-media/>
- [9] M. Kim, et al., "The Effects of Exhibition Types on Exhibits Behavior of Instagram Users - focusing on mediating effects of Narcissism, Moderating Effect of advertisement promotion," *Journal of Basic Design & Art*, 20(6), pp.67-88, December 2019. DOI: 10.47294/KSBDA.20.6.6
- [10] K. Budge, & A. Burness, "Museum objects and instagram: Agency and communication in digital engagement," *Continuum*, 32(2), pp.137-150, June 2018. DOI: 10.1080/10304312.2017.1337079
- [11] National Museum of Modern and Contemporary Art Korea, "2020 Annual Report," National Museum of Modern and Contemporary

- Art Korea, pp.138-139, June 2021.
- [12] H. Kang, et al., "The Influence of Instagram Posts on Exhibition Participation: Focusing on Generation Z," Digital Contents Society, 21(4), pp.731-739, April 2020. DOI: 10.9728/dcs.2020.21.4.731
- [13] S. Shin, "A Study on the Impact of Digital Technology on the Organization of Museums and Art Museums," Ministry of Culture, Sports and Tourism, pp.172, May 2017.
- [14] A. Dawson, Which museums have the biggest social media followings?, The art newspaper, www.theartnewspaper.com/analysis/museum-masters-of-the-social-media-universe
- [15] Seoul Museum of Art, "Statistics on the number of visitors per day in 2019," Seoul Museum of Art, January 2020.
- [16] Southern Post, "2019 Seoul Museum of Art Exhibition Satisfaction Survey," pp.3-70, Southern Post, 2020.
- [17] Seoul Museum of Art, "David Hockney 190322-190804," Seoul Museum of Art, <https://sema.seoul.go.kr/ex/exDetail?exNo=316623&glolangType=KOR&searchDateType=SOON>
- [18] S. Park, "New Economic Indicators and Gross Data Production in the 4th Industrial Revolution Era," KISTI Issue Brief, 8, pp.2-5, April 2019. DOI: 10.22810/2019KIB008
- [19] E. Nofal, R. Stevens, T. Coomans, and A. V. Moere, "Communicating the Spatiotemporal Transformation of Architectural Heritage via an In-Situ Projection Mapping Installation," Digital Applications in Archaeology and Cultural Heritage, 10(0), pp.1-15, November 2018. DOI: 10.1016/j.daach.2018.e00083
- [20] B. Rhee, "Art-tech 4.0: the convergence of art and technology in the era of the fourth industrial revolution," Book Korea, 352p, 2020.
- [21] J. Park, et al., "An Analysis of Visitor Responses Based on Instagram Hashtags," Proceedings of the Korean Society of Computer Information Conference, 26(2), pp.369-372, July 2018.
- [22] E. Lee, et al., "A Study on the Characteristics of Social Presence of Museum Experience through the Lens of Instagram Posts : Case Studies of Art Museums in New York Area including Metropolitan Museum of Art, Museum of Modern Art and Solomon R. Guggenheim Museum," Journal of Museum Studies, 40, pp.201-226, June 2021. DOI: 10.22884/joksms.40.202106.008
- [23] E. Lee, et al., "Instagram Hashtag Analyses Related to Museum Experience -A Comparative study of British Museum and Tate Modern-," Journal of Museum Studies, 38, pp.237-259, June 2020. DOI: 10.22884/joksms.38.202006.010
- [24] J. Lee, et al., "Book Recommendation System using Word2vec," The Korean Institute of Information Scientists and Engineers, 2019(6), pp.922-923, June 2019.
- [25] B. Rhee, J. Choi, W. Hyung, and F. Pianzola, "Computer vision algorithm for the analysis of museum visitors' pose," Under review. 2021.
- [26] M. Georgalou, "Placemaking and place identity in social media: Snapshots from Facebook," Proceedings of the 3rd international biennial conference, pp.361-370, National and kapodistrian university of athens, Greece, September 2015.
- [27] Y. Lee, "MZ Generation: Culture and Arts Trends," The Korean Journal of Arts Studies, 28(0), pp.341-350, June 2020. DOI: 10.20976/kjas.2020..28.016
- [28] Ministry of Culture, Sports and Tourism, "Museums and art galleries that enrich life with culture. - Mid- to Long-term Plan for Promotion of Museums and Art Museums (2019-2023)-," pp. 19-20, June 2019.

Authors



Ji-Su Kang received the B.A. degree in German from Hankuk University of Foreign Studies in 2016. She received the M.A. degree in Art & Technology in GSAIM Chung-Ang University in 2022.

Her research interests include online exhibition and museum visitor analysis using big data technology.



Bo-A Rhee received the B.A. degree in Library Science from Sung Kyun Kwan University in 1987, M.A. degree in Art Studies from Graduate School of Sung Kyun Kwan University in 1990 and Ph.D. degree

in Art Management from Florida State University, U.S., in 1997. Dr. Rhee is currently a professor at the College of Art and Technology, Chung-Ang University, South Korea. She is a museum technology and informatics researcher. Her works have focused on transdisciplinary approaches, hyper-connected museum, mediation between digital surrogates (i.e. digital exhibitions, VR and AR) and viewers, social media, and museum 3.0.