



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사학위논문

보도 사진의 통사적 구조와

미디어 편향에 대한 고찰:

유럽과 중동 언론의 분쟁 보도에 대한  
시각 데이터 분석을 중심으로

2017년 8월

서울대학교 대학원

융합과학부 디지털정보융합전공

유재연

# 보도 사진의 통사적 구조와

## 미디어 편향에 대한 고찰:

유럽과 중동 언론의 분쟁 보도에 대한 시각  
데이터 분석을 중심으로

지도교수 서 봉 원

이 논문을 공학석사학위논문으로 제출함

2017년 8월

서울대학교 대학원

융합과학부 디지털정보융합전공

유 재 연

유 재 연 의 석사학위논문을 인준함

2017년 8월

위 원 장 이 중 식 (인)

부 위 원 장 권 가 진 (인)

위 원 서 봉 원 (인)

## 초 목

기계의 시각 인지는 패턴을 읽는 방식으로 진행돼 왔다. 현행 기계학습 (Machine Learning)에서는 특질을 통해 대규모 데이터를 분석하고 결과를 예측하는 형태(feature engineering)를 띤다. Manovich가 언급한 패턴 읽기 방식처럼 이제는 대규모 시각 데이터에 대한 분석도 가능해졌다.

본 연구는 패턴 읽기를 통해 기계가 사진의 문맥을 파악하고, 그 안에 포함된 인간의 판단을 확인할 수 있는 가능성을 알아보고자 시행됐다. 쉽게 말해 인간이 의식적으로 또는 무의식적으로 고른 사진 속에서, 기계가 인간의 편견이나 고정관념을 찾아낼 수 있는지를 살펴본 것이다.

분석을 위해 중동 언론인 알 자지라(Al Jazeera)와 유럽 언론인 로이터(Reuters)의 보도 사진 가운데 중동 및 유럽 테러를 다룬 것을 데이터로 썼다. 이들 사진을 대규모로 모은 뒤 각각 자기 문화권과 타 문화권의 테러 보도 방식에 어떤 차이가 있는 지 보고, 이를 토대로 편견 여부를 확인했다.

사진의 문맥을 파악하기 위해, 기존의 요소 인지를 중심으로 하는 패턴 뿐 아니라 사진의 구도에 대해서도 살펴보았다. 언어학의 통사론을 차용해 사진 속 위치에 따른 요소들의 기능을 확인했다. 분석 결과, 인간의 시각을 가장 많이 사로잡는 왼쪽 상단과 중앙, 중앙 상단과 한 가운데에는 인물의 등장이 잦았다. 그리고 이 위치에서 그룹별로 나타나는 인물의 감정에는 상당한 차이가 있었다.

연구를 통해 인간의 편견과 같이 정성(定性)적으로 파악되는 요소에 대해서도 기계적 분석을 할 수 있다는 가능성을 확인했다. 이는 기존 사회과학 연구에서 인간 코더가 하던 일을 기계가 대신 할 수 있는 장을 열어줌으로써, 더 많은 양의 데이터를 객관적으로 다룰 수 있도록 한다. 그동안 명확하지 않았던 인간이 편견을 느끼는 근거에 대해서도 교차적으로 연구할 수 있는 가능성 또한 열렸다.

본 연구에서 제시된 바, 사진 속 위치에 따른 문맥 파악을 토대로 이를 활용한 데이터 군집화(clustering) 및 편견 감지(detecting) 등이 가능할 것으로 보인다. 특히 미디어에서 시각적으로 치우친 보도를 할 경우를 경계하는 도구로도 사용될 수

있을 것이다. Eco 가 미디어를 비판하며 말한 바와 같이 “<매스 미디어로 재현된> 세계에서 우리가 할 수 있는 유일한 자유 선택” 일 지도 모른다.

주 요 어: 시각 데이터 분석, 감성 컴퓨팅, 컴퓨터 비전, 보도 사진  
편향 연구, 통사론 분석, 사진 연구

학 번: 2015-26101

# 목 차

제 1 장 서 론 .....	1
제 1 절 연구의 배경 .....	1
제 2 절 연구의 목표 .....	14
제 2 장 선행 연구 .....	16
제 1 절 저널리즘의 관점에 대한 연구.....	16
1.1 지역 미디어에 대한 연구 .....	16
1.2 포토 저널리즘의 편향에 대한 연구 .....	20
제 2 절 이미지 분석에 대한 연구 .....	23
2.1 컴퓨터 비전 .....	23
2.2 감성 컴퓨팅 .....	25
2.3 색채 시맨틱 연구.....	26
제 3 절 사진의 구조에 대한 연구 .....	28
3.1 인간의 시지각에 대한 연구 .....	28
3.2 시각적 표상에 대한 철학적 논의 .....	29
제 3 장 데이터 수집 및 기계 분석 .....	34
제 1 절 데이터 개요 .....	34
제 2 절 데이터의 편견에 대한 연구 .....	38
2.1 사물 및 상황 분석 .....	39

2.2 감정 분석 .....	47
제 3 절 신뢰도 확인을 위한 파일럿 테스트 .....	53
제 4 장 이미지 속 위치에 따른 분석 .....	58
제 1 절 사진 속 인물의 비중 및 위치에 대한 연구 .....	59
제 2 절 사진 속 인물의 위치별 감정에 대한 연구 .....	62
제 3 절 사진의 통사적 구조에 대한 결론 .....	66
제 5 장 결론 및 연구 의의 .....	71
제 1 절 연구 요약 .....	71
제 2 절 연구의 시사점 .....	75
제 3 절 연구의 한계 및 제언 .....	77
참고 문헌 .....	79
별첨 .....	88
Abstract .....	97

## 표 목차

[표 1] 데이터 수집 대상 분쟁 목록 .....	35
[표 2] 수집한 데이터 수 .....	37
[표 3] Amazon API에 의해 검출된 사물 레이블과 신뢰도 점수	39
[표 4] 각 그룹별로 추출된 얼굴 수 .....	49



[표 5] 각 그룹별 감정 평균 .....	51
[표 6] 각 사진별 사람과 컴퓨터의 사물 탐지 차이 .....	56
[표 7] 그룹별 사진 위치에 따른 주요 인물 얼굴 등장 빈도 ...	62
[표 8] 사진 속 위치별 각 감정의 최대 값을 지니는 그룹 .....	64

## 그림 목차

[그림 1] 통사론에 따른 문장의 기본 구조 .....	6
[그림 2] 사진의 요소를 구조 속 관계의 관점에서 분석 .....	8
[그림 3] 같은 사건에 대해 보도한 로이터(좌)와 알 자지라의 사진 .....	10
[그림 4] 본 연구의 프로세스 .....	15
[그림 5] 로이터 웹사이트(좌측)와 알 자지라 아랍어 웹사이트 검색 화면 .....	36
[그림 6] 본 연구의 데이터 그룹 도식 .....	38
[그림 7] 로이터 중동과 유럽 그룹의 공통 사물과 차별되는 사물 .....	41
[그림 8] 사물 및 상황 분석에 대한 정리 .....	46
[그림 9] 각 그룹을 들쭉 찢어볼 때 차별된다고 보기 힘든 감정들 .....	51
[그림 10] 그룹별 감정 추이 .....	53

[그림 11] 사람과 기계가 탐지한 사물의 차이가 큰 사진 .....	55
[그림 12] 단어군에 대한 공감도(좌측)와 사진을 보고 느끼는 공감도 .....	57
[그림 13] 중앙값으로 확인한 각 감정 값이 가장 높게 나온 위치 .....	63
[그림 14] 여러 장의 사진이 있을 때 얼굴 위치별 감정 값의 비율에 따른 그룹 분류 .....	68

# 제 1 장 서 론

## 제 1 절 연구의 배경

장면을 골라 사진을 찍는 과정에는 인간의 주관이 개입된다. 실재하는 모습 가운데 사진가가 어느 한 부분을 골라 초점을 맞추는 행위이기 때문이다. 현상학의 창시자인 Husserl은 ‘현상’을 ‘의식에 나타난 대로의 것’이라 정의했는데, 사진에 담긴 상(像) 또한 인간의 의식에 나타난 것을 그대로 투영한 것이라고 본 연구에서는 해석했다. 우리가 보고 있는 모습을, 세계에 실재하는 존재자(entity)라고 철학적으로 명할 때, 하이데거는 의식에 나타나는 이 존재자를 두 가지로 구분했다. 한 가지는 ‘손에 쥐(ready-to-hand)’이고, 다른 한 가지는 ‘손 안에 있음(present-at-hand)’이다. 도구로서 마치 내 몸의 연장처럼 의식되는 실재는 ‘손에 쥐’으로 나타나고(예를 들어 망치나 라켓은 사용을 하게 되는 순간 더이상 독립적으로 존재하지 않는다), 주체인 내가 신경을 쓰게 되는 도구의 경우는 ‘손 안에 있음’으로 나타난다. Wheeler는 여기에 또 다른 한 가지를 추가했다. 현상학적으로 ‘손에 쥐’과 ‘손 안에 있음’ 사이에

‘손에서 놓침(un-readiness-to-hand)’이 존재하게 된다는 것이다[49]. 이를테면, 내가 펜으로 글씨를 쓰면 의식적으로 글자는 인지하지만, 펜 자체는 하이데거의 표현으로 ‘투명하기 때문에(transparent)’ 내 의식에 드러나지 않는다.

표상은 존재자를 인식하는 과정, 그 가운데서도 ‘손 안에 있음’과 ‘손에서 놓침’에서 나타난다. 손에 짚은 주관과 객관이 물아일체된 상황이기 때문에 표상이 들어갈 여지가 없다. 하지만 손 안에 있음과 손에서 놓침에서는 행위자와 존재자 사이 인지적 거리가 발생하게 되기 때문에, 행위자 입장에서 세계를 이해하는 재-실재(re-presentations)가 형성된다. 본 연구에서는 장면을 고르고 사진을 찍어 발행하는 행위를 ‘손 안에 있음’ 내지 ‘손에서 놓침’에 가깝다고 해석했다. 따라서 사진 속 장면은 존재자(현장)를 보여주는 하나의 표상으로서 역할을 수행한다고 가정할 수 있다.

그렇다면 기계는 인간이 만든 표상을 통해 존재자 또는 행위자를 이해할 수 있을까. Dreyfus는 인공지능의 존재론적 가정에 대해 맥락적 표상을 이해할 수 없다고 지적했다. 철학 용어로 ‘틀 세우기’ 문제에 빠지게 된다는 것이다. 예를 들어

인공지능에게 “패스트푸드점에 가서 햄버거를 사오라”고 한다면, 현실 세계에선 무수히 다양한 사건이 발생하기 때문에 이 모든 경우의 수를 고려하다 무한한 시간이 걸려 결국 일을 수행하지 못한다는 논리다. 따라서 상황에 따른 적절한 프레임을 선택하고 적용시켜야 한다는 것이 틀 세우기의 기본 가정이다. 이를 해결하기 위해 Dreyfus는 기계에도 인간과 같이 세계에 ‘던져져 있는 상태’로 여러 맥락에 항상 체화될 수 있도록 배경 처리(background coping) 혹은 토대적 수준의 지성(ground-level intelligence)이 탑재돼 있어야 한다고 주장했다[16].

Moravec은 기계가 스스로 보상에 대해 느끼게 되는 날이 온다면, 곧장 사변적인 진화를 통해 자율성을 획득하게 될 것이라고 주장했다[36]. 예를 들어 영화 <인터스텔라(2014)>에서 주인공 쿠퍼와 로봇 ‘타스’가 블랙홀로 빨려들어가기 전, 쿠퍼는 타스가 인간을 위해 희생하도록 프로그램 돼 있다고 말한다. 하지만 Moravec의 관점에서 보면 사변적 진화가 일어날 경우, 타스는 저항을 할 수도 있다. 따라서 영화상 타스는 인간의 틀 세우기 문제를 극복하며 ‘토대적 수준의 지성’을 탑재했지만, 보상을 학습하는 등의 자율성은 획득하지

못한 셈이다. 다만 영화에서 유머와 솔직함을 퍼센테이지(%)로 세팅하는 부분은, 기계가 맥락을 읽고 수행하는 방법 또한 인간의 의지에 달려있음을 시사한다. 인간이 만든 표상을 통한 존재자, 행위자에 대한 이해는 ‘보상에 대한 이해’가 있기 전까지는 인간이 만든 프레임에서 수행 가능할 지도 모른다.

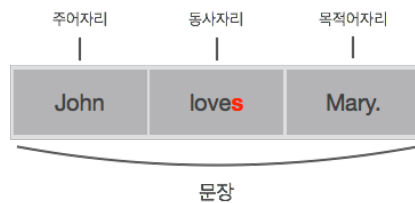
이 같은 관점에서 본 연구는 맥락을 파악할 수 있는 프레임을 확인하는 방향으로 진행했다. 이전까지 맥락 기반으로 영화나 음악, 여행지, 레스토랑 등에 대한 추천 시스템 연구는 다수 있었다. Adomavicius & Tuzhilin은 이 문맥적 정보가 데이터에 대한 직·간접적 접근이나 내포된 정보의 획득 등으로 구해질 수 있고, 잠재 변수에 숨겨져(hidden) 있는 내용을 통해 얻어질 수도 있다고 했다. 그러면서 일부 문맥적 정보는 전문가가 설정한 변수를 통해 획득될 수 있는데, 그 예시로 영화 추천 시스템을 들었다. 단지 영화 상영 시간과 배급사, 극장같은 일반적인 변수 뿐 아니라 전문가가 추천하는 보다 맥락적인 변수들도 수집해야 제대로 된 추천을 할 수 있다는 주장이다[6].

보도 사진 또한 단순히 기계적으로 파악 가능한 사물이나 감정 값만 가지고는 전체적인 맥락을 확인하기 어렵다. 따라서 사진의

구도와 같이 전문적인 분야를 변수로 넣어 맥락을 살필 필요가 있다. 이전까지 그림의 구도나 사진 속 요소의 위치에 대한 미학 및 미디어 분야 전문가들의 연구는 존재했다. 하지만 대부분이 인간의 정성연구로 진행돼 객관성을 담보하기 어려웠다. 보도사진에 대한 기계적 맥락 분석은 이뤄진 바가 없어 본 연구가 다량의 데이터를 통해 정성적 연구를 시행하는 시작점이 될 것으로 보인다.

기계적 맥락을 학습시키는 방법으로, 본 연구에서는 사진의 구조적 요소를 살펴보았다. Manovich는 기계의 진화가 읽는 것에서 패턴을 인지하는 것으로 변하고 있다고 언급했다. 이전까지의 사회·문화 분석에서는 자료들을 “통계학이나 사회학에서의 많은 ‘표면적 데이터(shallow data)’ 와, 가까이서 들여다보아야(close reading) 하는 심리학·인류학의 ‘심화된 데이터(deep data)’ 에 각각 의존해왔다” 는 것이다. 하지만 소셜미디어의 발달과 이를 다룰 수 있는 컴퓨테이션 툴의 발전으로 더 이상 데이터의 깊이나 양에 한계를 두지 않아도 되는 상황이 왔고, 방법론적으로는 패턴을 통해 사회를 분석할 수 있게 됐다[31].

패턴을 파악하기 위해 본 연구에서는 언어학의 통사론(Syntax)과 형태론(Morphology)을 활용했다. 이미지와 언어학을 연계해 규칙을 찾아가는 연구는 실제로 여러 곳에서 진행되고 있다. 예를 들어 사진에 바이크와 사람이 나올 경우, ‘사람이 바이크를 타고 있다’는 문장과의 유사도와 높다고 판단해, 사진 속 이미지에 대해 ‘타고 있다’라는 동사형을 부여하는 식의 의미론을 활용한 연구도 있다[23]. 나아가 CNN과 같은 기계학습을 의미론에 적용하는 연구도 등장했다[21].



[그림 1] 통사론에 따른 문장의 기본 구조

본 연구의 모티브가 된 통사론은 언어의 공간적 형태에 집중하는 분야다. 언어적 의미를 이룬 추상적인 공간을 하나의 배열 관계를 통해 시각화된 표상(representation)으로 만들어내는데, 그것이 바로 문장이다. 이 과정에서 단어들이 각자 자신의 자리에 알맞게 파생(derivation)해 위치하도록 하는 것이 통사 구조



(syntactic structures)다[14]. 예를 들면 [그림 1]에서 볼 수 있는 것과 같이 “John loves Mary”에서 ‘John’과 ‘loves’, ‘Mary’가 A-B-C 각각의 위치에 자리하게 되는데, love는 A에 형태를 맞춰야 하는 자리이기 때문에 3인칭 단수형으로 바뀌는 것이다. Milner가 주창한 위치통사론에서는 문장이 단순한 배열관계를 넘어서서 어느 단어가 어느 자리를 점유하느냐(relation d’ occupation)에 따라 의미가 달라진다는 점을 탐구한다[35]. 앞 문장에서 A에 ‘John’ 대신 ‘Mary’가 들어가면 ‘Mary loves John’으로 뜻이 완전히 바뀐다. 이 경우 어휘들은 통사의 차원에서 어느 위치(position)를 차지함으로써 그 기능(Mary가 주어, John이 목적어의 기능)을 결정 받았다고 볼 수 있다[4]. 확률에 따라 단어가 등장하는 마르코프 체인과 달리, Chomsky는 모든 상이한 자연 언어의 바탕에 공통된 회로가 깔려있다고 본다. 이같은 통사론의 원리로 만화의 컷들을 관통하는 시퀀스의 시각적 인지 규칙을 찾아 나간 연구도 있다.



〔그림 2〕 사진의 요소를 구조 속 관계의 관점에서 분석

보도사진 기저에 위치한 공통 회로를 찾기 위해서는 사진 속 형태들에 대한 연구 또한 필수적이다. 언어학에서의 형태론은 이 추상적인 공간인 문장을 구성하는 구체적인 단어에 대한 연구다. 통사론의 설명 방법으로 한때 주목받았다. 어휘론(vocable)과 달리 단어의 의미보다 단어의 형식에 집중하므로 ‘형태론(形態論)’으로 불린다. 이 연구의 범위는 단어의 소리와 문자, 의미에 대한 것부터 단어의 생성과 변화를 역사적으로 연구하는 관점, 소쉬르식 표현으로 ‘공시대(synchronie)’에 놓인 단어들의 상태(état)를 살피는 것까지 이른다. 본 연구에서 다루는 보도사진 또한 어느 공간(사진)에 놓인 사물 또는 감정들의 상태들로 구성돼 있다. 좀 더 인지적으로 접근하면, 실재하는 현실 자체는 기자의 눈에 보이지만, 사진을 찍는 의도라는 것은 그의 머릿속에 있는

추상적인 무엇이다. 셔터를 누르는 순간, 카메라의 프레임 안에 실재하는 현실이 포착됨으로써 이는 이미지라는 표상으로 남겨지는 것이다. 보도사진은 ‘상(facts)을 상(picture)으로 바꾼 것(encoding)’으로, 그 과정에서 인간이 지닌 선택적 속성(selective exposure, selective perceive, selective intention)을 주의해야 한다는 주장도 있다[2]. 보도사진 속 요소들이 텍스트(text)의 어원처럼 직조(tissue)된다는 입장이다. 현대 보도사진의 아버지라 불리는 Cartier-Bresson도 “좋은 사진이란, 각 요소를 최대한으로 표현하고, 또 활력있게 이끄는 주제와 형식의 조화”라고 밝혔다[19]. 그만큼 내용뿐 아니라 사진 속 형태에 있어 시각적으로 엄격한 구성이 전통적으로 중시돼 왔다.

따라서 본 연구는 사진 속 요소와 형태, 그리고 그들 간의 문법에 집중해 저널리즘적 메시지를 파헤치려고 한다. 그리고 그 방법으로 딥러닝을 활용하고자 한다. 마침 기계의 시각적 인지 능력은 최근 들어 더욱 가파르게 발전하고 있다. 2012년 구글에서 개발한 이미지 인식 프로그램이 여러 이미지 데이터

속에서 고양이를 정확하게 찾아낸 데 이어,<sup>1</sup> 2017년 현재는 사용자가 입력한 사진을 검색 쿼리로 활용하고<sup>2</sup> 학습한 화법 대로 그림을 그리기도 한다. 이 가운데 본 논문에서 다룬 이미지 분석 방법은 사물·상황 탐지와 감정 인식 기술이다. 사진에 있는 요소를 결과로 도출하고, 만일 그 안에 인물이 있다면 그의 감정을 인지하는 기술이다. 이를 통해 보도사진 속 기자의 의도와 그를 둘러싼 의미를 파악했다.



[그림 3] 같은 사건에 대해 보도한 로이터(좌)와 알 자지라의 사진

이 연구에서는 기계가 시각 이미지의 구조를 이해함으로써 맥락을 파악하고, 이를 통해 인간의 편견과 같은 무의식적 행위를 구분해 낼 수 있는 기반을 마련하고자 한다. 그 방법으로 앞서 언급한 컴퓨터 비전을 기반으로 하는 AI API를 활용하고자 한다.

<sup>1</sup> <https://googleblog.blogspot.kr/2012/06/using-large-scale-brain-simulations-for.html>

<sup>2</sup> <https://support.google.com/websearch/answer/1325808?p=searchbyimagepage&hl=en>

기계가 발견해내는 시각적 요소 가운데 인간의 경우와 상당부분 일치하는 경향을 보인 사물 탐지 및 감정 인식에 대한 부분이다. 이처럼 기계의 시각 이미지 인지를 통한 빅데이터 분석을 진행한 연구는 다수 있다. Lydia Manikonda & Munmun De Choudhury는 인스타그램의 사진 150만장을 모아 시각적으로 드러나는 정신건강에 대한 연구를 진행한 바 있다. 사진의 채도와 명도가 정신적으로 취약한 유저의 사진에서 더 높게 나타나는데, 이는 뷰어들의 시선을 더 모으는 경향이 있다고 설명했다. 또한 샘플 사진을 주제별로 분류한 뒤 이를 히트맵으로 구현, 주제별 특징을 찾아 다량의 사진에서 주제를 추측하는 알고리즘을 사용하기도 했다. 이를 통해 정신적 취약 유저 그룹이 주로 글귀가 삽입된 사진을 올리는 경향이 있고, 셀피와 함께 비극적인 태그도 다수라는 것을 확인했다. 나아가 주제별로 어떤 감정값이 드러나는 지 확인했는데, 사람들과 함께 찍은 사진에서는 불안감을 표출하고, 음식 사진에서는 분노와 함께 긍정적인 감정도 많이 나타났다. 음식의 경우 식이장애가 있는 사람들과 회복기에 있는 사람들이 동시에 올리기 때문에 두 가지 감정이 모두 나타난 것으로 연구진은 해석했다[27].

분석 대상으로는 두 지역 언론의 분쟁 보도사진을 다뤘다. 2017년 현재 국제사회는 소용돌이에 휘말려 있다. Beck이 말한 위험사회(Risk Society)에서 위협사회(Threat Society)로 전환하고 있다[37]. 이슬람 무장세력 IS의 테러 위협과 더불어, 중동 국가 출신의 난민 수용 여부를 두고 유럽과 북미 대륙을 중심으로 정치적 이견이 팽팽하게 맞서고 있다. 일부 유럽 지역 테러가 난민 중 일부의 소행으로 밝혀지면서 이 지역의 경계는 한층 강화됐고, 급기야 영국이 EU에서 탈퇴를 하는 이른바 브렉시트(BREXIT)를 감행하는 데 이르렀다. 그리고 트럼프 미국 대통령은 일부 중동 지역 국가 출신 사람들의 미국 입국을 거부하는 ‘반(反)이민 행정명령(Executive Order)’을 추진하고 있다<sup>3</sup>.

이런 가운데 전 세계적으로 유럽 지역 테러에 대해서는 공감도가 높고, 중동 지역에서 벌어지는 내전에 대해서는 무감각하다는 비판도 나온다. 파리 테러가 일어났던 지난 2015년

---

<sup>3</sup> Protecting the Nation from Foreign Terrorist Entry into the United States.

<http://edition.cnn.com/2017/01/28/politics/donald-trump-immigration-detention-deportations-enforcement/>

11월, 페이스북(Facebook)은 프로필 사진에 프랑스 국기를 합성하고 해당 위치에 있는 이들의 안전을 묻는 ‘Safety Check’ 서비스를 마련한 바 있다. 사건 전날 레바논 베이루트에서도 IS의 테러로 마흔 한 명이 사망하는 일이 있었지만, 레바논 국기 합성 서비스나 ‘Safety Check’는 없었다. 이에 대해 뉴욕타임즈는 “세계가 아랍인의 삶은 덜 중요하다고 보고 있다”는 레바논 논평가들의 비판을 전하면서, “시리아의 이웃이긴 하지만 상대적으로 평온하던 레바논에서 벌어진 테러 공격이 무관심 속에서 만성적인 것으로 간주되고 있다”고 지적했다<sup>4</sup>.

따라서 이번 연구를 통해 양측 언론에 내재한 편견을 확인하고, 편견의 근거로 제시할 수 있는 맥락적 정보들을 취합해 이들이 예측 가능한 변수인 지를 알아보고자 한다. 특히 편견에 대한 시각적 근거(visual cue)를 제공하는 데 초점을 맞춘다. 이를 통해 기존 사회과학분야의 정성연구에서 한계로 지적되던 정량적 연구를 가능케 하고, 편향과 같은 인간이 설명하기 어려웠던 막연한

---

<sup>4</sup> [https://www.nytimes.com/2015/11/16/world/middleeast/beirut-lebanon-attacks-paris.html?\\_r=0](https://www.nytimes.com/2015/11/16/world/middleeast/beirut-lebanon-attacks-paris.html?_r=0)

부분에 대해 기계적 해석을 동반한 근거를 찾아볼 수 있다. 나아가 이미지의 통사적 분석이라는 새로운 학문의 장을 열어 향후 이미지 프로세싱 및 추천 방법에도 다양하게 활용될 수 있을 것으로 전망한다.

## 제 2 절 연구의 목표

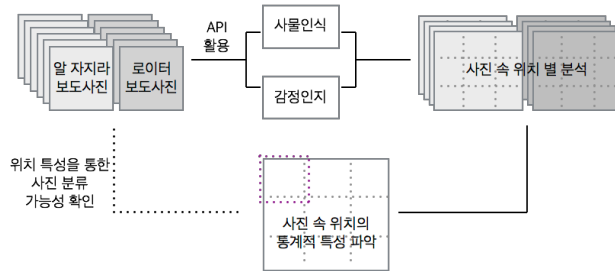
본 연구에서 해결하고자 하는 문제는 다음과 같다.

1. 기계학습을 통해 사진의 차이를 구조적으로 확인할 수 있는가?
  - 1.1. 기계적 분석 결과를 토대로 사진의 차이를 구조화할 수 있는가?
  - 1.2. 기계가 마련한 통사 구조로, 역으로 편견을 증명하거나 예측할 수 있는가?
2. 미디어 편향에 대한 기계적 확인이 가능한가?
  - 2.1 기계적으로 분석할 때, 중동과 유럽 미디어의 사진에서 편견이 드러나는가?
  - 2.2 인간도 이 같은 편향의 결과에 동의하는가?
3. 학문적, 실용적으로도 기여가 가능한 연구인가?
  - 3.1 이론적으로 정당화할 수 있는가?



### 3.2 실생활에서 재현 가능한 분석 방법인가?

이를 총체적으로 확인하기 위해 연구 절차는 [그림 4]와 같이 진행했다.



[그림 4] 본 연구의 프로세스

## 제 2 장 선행연구

본 연구를 진행함에 있어 먼저 데이터로 다루는 지역 언론이 대표성을 지니고 있는 지 살피고, 실제 미디어에서의 편향에 대한 연구는 어떤 방식으로 진행돼 왔는지 짚어 보도록 한다. 또한 기계의 이미지 분석과 관련해 현재의 기술 및 학문적 성과를 알아보고, 기존 인간의 시지각 연구를 토대로 사진의 구조를 문맥적으로 읽어낼 수 있는 지를 점검한다.

### 제 1 절 저널리즘의 관점에 대한 연구

사회과학분야에서 언론사의 정치적, 사회적 편향에 대한 연구는 꾸준히 이어지고 있다. 특히 특정 정당이나 정치인에 대한 커버리지(coverage)에 대한 실증 연구가 많다. 본 연구는 사회과학에서 입증한 저널리즘의 편향을 기계적으로 도출하는 것을 목표로 한다.

#### 1.1 지역 미디어에 대한 연구

본 연구는 서구 언론의 시선과 중동 언론의 시선을 중점적으로

다룬다. 대표성 있는 언론으로 서구에서는 로이터 통신사(Reuters)를, 중동에서는 알 자지라(Al Jazeera)를 꼽았다. 먼저 로이터의 경우 한때 유럽 상위 500 개 기업 가운데 48 위에 오를 정도로 ‘다국적(multinational) 초국가적(transnational) 미디어 기업’의 입지를 다졌고, 전세계 150 개국과 제휴를 맺으며 영향력을 확대했다. AP, AFP, UPI 를 비롯한 대형 뉴스 통신사(news agencies)와 더불어 뉴스의 세계화에 앞장섰다. 특히 로이터는 19 세기 중반부터 Havas, Wolff 와 함께 글로벌 뉴스 통신사 카르텔을 형성하는 등 긴 역사를 자랑하고 있다. 그만큼 굴곡도 많았다. 소속 국가(영국)의 입장을 대변하는 등 편향적인 입지를 퍼뜨린다는 비판도 나왔다[12]. 하지만 언론사 자체적으로 편향에 대해 경계하며, 최대한 중립적인 단어를 쓰는 등<sup>5</sup>의 노력을 기울이고 있다. 9/11 테러 당시 ‘테러리스트’라는 단어를 쓰지 않은 것이 대표적 사례다<sup>6</sup>.

중동의 대표적 지역 언론인 알 자지라는 1996 년 서구 중심의

---

<sup>5</sup> [http://handbook.reuters.com/index.php?title=Freedom\\_from\\_bias](http://handbook.reuters.com/index.php?title=Freedom_from_bias)

<sup>6</sup> <https://en.wikipedia.org/wiki/Reuters>

글로벌 TV 뉴스 저널리즘에 반하며 세워진 카타르의 아랍어 방송국이다. 아랍을 대표하는 통신사가 없는 관계로 방송사인 알 자지라를 선택해야 했다. 알 자지라와 같이 아랍위성방송(ASB)을 대표하는 언론으로는 아부다비 TV 와 알-아라비야가 있는데, 이 가운데 알 자지라가 가장 먼저 서구의 뉴스 독점 시대에 대한 대항적(contra-flow) 채널로 자리매김했다. 그러다보니 안티-아메리칸, 안티-이스라엘 프로파간다 등으로 거센 항의에 직면하는 경우도 다반사였다. 이라크전 기간, 주요 서구 언론을 통해 알려진 알 자지라 소스의 기사들이 대부분 반서구적 메시지를 담고 있었다는 것이다. 다만 이를 전하는 서구 언론의 자체적인 프레임을 거치다보니, 알 자지라가 ‘동양에서 서양(east to west)’을 잇는 소통의 가교(communication bridge) 역할은 잘 하지 못하고 있다는 의견도 있다[48].

알 자지라가 서구에 대항하는 대표적 중동언론이 맞느냐는 반문도 나온다. 알 카에다 테러리스트인 오사마 빈 라덴 테이프를 제대로 방영하지 않고, 전쟁의 일그러진 얼굴에 대해선 묵인함으로써 알 자지라 스스로 언론으로서의 정직성의 의무를

저버리는 등 국제적인 저널리즘의 가치를<sup>7</sup> 제대로 이행하지 못하고 있다는 것이다[42]. 비판에 부딪힌 알 자지라는 자체적으로 ‘침묵하지 않을 권리(the right not to remain silent)’를 슬로건을 내걸고 이-팔 분쟁 당시 팔레스타인의 목소리를 전하는 등 적극적으로 분쟁 보도에 나섰지만 장벽은 있었다. 카타르 정부의 외교적 입지상 이스라엘의 입장 또한 전해야 했고, 또한 이라크전 이후 미국 기반의 민영 미디어 그룹으로부터 언론사 자체적으로 다양한 문화를 도입하는 바람에 오히려 중동 국가들로부터 견제를 받았다. 비슷한 시기, 영국 언론에선 미국 부시 행정부가 알 자지라 본부를 폭발하려는 것을 토니 블레어 당시 총리가 만류했다는 폭로가 터져나왔고<sup>8</sup>, 이는 곧 서구의 알 자지라에 대한 불편한 시선을 보여주는 대표적 사례로 꼽혔다. 서구 언론이 포착하기 힘든 아라비아 반도의 뉴스를 전할 수 있는 강점을

---

<sup>7</sup> Sakr 는 2004 년 카타르 도하에서 열린 국제 미디어 포럼에서 정한 저널리즘의 가치로 정직성, 용기, 공정성, 균형, 독립성, 그리고 다양성이 꼽혔다고 설명했다.

<sup>8</sup> 2005 년 11 월 영국의 데일리미러 보도. 2004 년 미국 워싱턴을 방문한 토니 블레어 총리에게 조지 W 부시 대통령이 알 자지라 본부를 폭파하고 싶다고 말한 것을 총리가 말렸다는 내용이 담겨있다.

[https://www.theguardian.com/media/2005/nov/25/pressandpublishing.Iraqandth  
emedia](https://www.theguardian.com/media/2005/nov/25/pressandpublishing.Iraqandth<br/>emedia)

지니고도, 알 자지라의 입지는 상당히 모호하다고 Sakr 는 주장했다[42].

본 연구에서는 서구의 시선, 그리고 중동의 시선을 확인하고자 한다. 따라서 BBC 에서 일하던 스태프가 대거 투입된 알 자지라 영어판 대신 아랍어판 인터넷 페이지에 올라온 사진을 중심으로 연구를 진행한다.

## 1.2 포토 저널리즘의 편향에 대한 연구

언론사의 편향에 대한 연구는 특히 정치 분야에서 활발하게 다뤄지고 있다. 그 중 한 예로 신문사의 사진 게재 행위에 편향성이 있다는 사실을 통계적으로 밝힌 연구가 있다. 이들은 1998 년과 2002 년 미국 선거 기간 7 개 신문사의 435 명 후보에 대한 사진을 수집하고, 이를 3 명의 코더로 하여금 호감도와 사진 사이즈에 따라 구분하게 하는 방법을 썼다. 결과적으로 신문의 정치적 편향성과 후보에 대한 호감도는 강한 관계를 보였고, 이 신문의 사설을 분석해 언론사의 정치적 분위기를 시간대별로 측정한 결과 사진 선정시 호감도와 강한 연관성을 보인 것으로 나타났다[9][10].

또한 뉴스 사진에 숨겨진 패턴을 확인하기 위해 Google Cloud Vision API 를 활용한 연구도 있다[25]. 정치인들이 실린 보도사진을 중심으로 구글의 사물 및 감정 분석 프로그램을 시행했는데, 환희와 슬픔에 대한 버니 샌더스와 힐러리 클린턴 당시 민주당 경선 후보들의 비율이 언론사별로 다르게 나타났다. 로이터와 CNN 에서 보도에서 클린턴의 환희 비율이 20%대를 웃돈데 반해, 샌더스의 환희 비율은 0%를 기록했다. 하지만 대부분의 연구가 감정을 확인하거나 선호도를 살피는 데 머물렀을 뿐, 총체적인 분석을 시행한 경우는 눈에 띄지 않는다.

보도사진 전체적으로 볼 때, 현실을 다소 왜곡하는 경향이 있다는 비판도 나온다. 미국 남북전쟁(1861~1865) 당시 신문 1 면에 게재된 삽화를 살펴보면, 텍스트는 전쟁에 대한 아주 세부적인 이야기라 할 지라도 그림 자체는 당시 19 세기 중반 화풍을 따른 풍경화를 중심으로 실렸다고 한다. 전쟁터를 보여주더라도 군인 대신 군인의 묘나 캠프를 보여주었고, 사람이 없는 미개척지의 풍경을 실어날랐다는 것이다. 이처럼 공간으로서의 현장이 아닌 현장의 ‘가까운 곳’ 을 보여주면서, 뉴스 전반의 타당성을 높이는 듯한 19 세기 중반의 저널리즘

문화가 현재까지 이어지고 있다는 비평도 나온다[38].

더 나아가 Eco는 텔레비전으로 인해 “투명성의 권리 즉 외부 세계와 접촉할 기회를 박탈당한 채 시청자는 자기자신으로 회귀한다”고까지 말한다. 실제로 일어난 실재가 카메라라는 프레임과 언론의 연출로 규정돼 전달되기 때문에, 결국 시청자는 실재라 불리는 것을 기껏 이 텔레비전 기기라는 것을 통해 인지하게 된다는 것이다. 뿐만 아니라 실재 조차도 텔레비전 방송을 위해 움직이고 있다는 게 Eco의 주장이다[18]. 사실 ‘조작된 실재’가 우선인지, 혹은 ‘의도적 연출’이 먼저인지는 여전히 ‘닭이 먼저냐, 달걀이 먼저냐’와 같은 오류에 빠져 있는 상황이다. 본 연구는 실재를 투영한 연출이 어떤 방식으로 다르게 비춰졌는지를 중점적으로 구성했다.

언론의 편향적 보도는 실제 사람들에게도 영향을 끼친다는 연구가 이어지고 있다. UN 산하 국제난민기구(UNHCR)에서 실시한 연구에 따르면, 유럽 미디어들이 보도하는 EU의 난민 뉴스의 톤이 지역에 따라 상당히 차이가 있고, 그것이 실제 여론으로도 이어지는 것으로 밝혀졌다[11]. 해당 연구에서는 신문의 텍스트만을 토대로 연구를 진행했는데, 본 연구에서는



사진을 다루려고 한다. Magnussen & Greenlee 의 연구처럼, 시각적인 인상은 말보다 기억이 더 잘 되기 때문이다[29].

## 제 2 절 이미지 분석에 대한 연구

보도 사진이 내포한 편견을 기계적으로 확인하기 위해 이미지 프로세싱에 대한 기존 연구를 확인했다. 기계가 인간과 비슷한 결정을 내리기까지 컴퓨터 비전의 발전이 수십 년에 걸쳐 진행됐는데, 그 가운데 데이터 처리 용량 및 속도의 획기적인 증가에 따라 딥러닝이 가능해지면서 최근 이 분야는 비약적인 성장을 기록했다.

### 2.1 컴퓨터 비전

기계가 보여주는 이미지와 인간의 관계를 연구하는 분야는 크게 둘로 나누어 볼 수 있다. 컴퓨터가 인간처럼 이미지를 인식하게끔 하는 비전(Computer vision) 분야, 그리고 사람이 컴퓨터상 이미지를 어떻게 보고 있는지 탐구하는 인터랙션(Visual Interaction) 분야다. 이 가운데 후자는, 컴퓨터 화면에서 인포그래픽 이미지를 보는 독자들이 어떤 색채를 제대로 인지하지

못하는 지 살펴보는 연구[40], 독자가 어떤 이미지를 응시하는 지 동공을 트래킹(eye tracking)하면서 자동으로 선호사진을 골라주는 연구[47] 등을 포함한다. 이와 달리 컴퓨터 비전은 컴퓨터가 인간의 시각 기능을 모방하는 모양새를 띤다. 크게 사물 탐지(image detection)와 감성컴퓨팅 (Affective Computing)으로 나눌 수 있는데, 모두 딥러닝의 비약적인 발전으로 최근 각광받고 있는 분야이기도 하다. 본 연구에서는 이 두가지를 모두 다룬다.

기존의 사물 탐지(image detection)는 사진 속에 들어있는 요소를 레이블링하는 데서 시작한다. 지도학습(supervised learning)을 통해 특정 규칙을 기반으로 분류(categorized)되는 체계가 진행돼야 사물에 대한 탐지가 가능했다. 하지만 이후 DNN 즉 심층 신경망(Deep Neural Network) 기반의 비지도학습(unsupervised learning)이 출현하면서 각각의 사물들이 이미지 기본 요소들의 계층적 구성으로 표현되기에 이른다. 이후 계층이 점차적으로 규합하면서, 아주 작은 수의 유닛만으로도 데이터를 모델링하는 게 가능해진 것이다[44].

이같은 컴퓨터 비전이 가능해짐으로써 대용량의 잡지 사진이나 플리커(Flicker) 상 이미지 데이터를 인간 사회 분석의 자료로 쓸

수 있게 됐다고 Manovich 는 설파한 바 있다. 그는 기계의 패턴 읽기가 가능해짐에 따라 데이터의 양이나 깊이에 상관없이 분석을 할 수 있게 됐다고 밝힌 바 있다[32].

## 2.2 감성 컴퓨팅

감성 컴퓨팅(Affective computing)은 “감정과 관련이 있고, 감정으로부터 발생하며, 의도적으로 감정에 영향을 주는 컴퓨팅을 통칭” 한다[39]. 딥러닝을 활용해 프로그램에게 감정을 학습시키는 연구와 마찬가지로, 학습된 감정을 활용하는 연구 또한 감성 컴퓨팅의 영역에 포함된다.

무의식의 영역으로 여겨지던 사람의 표정을 시시각각 분석해 그의 감정을 인식(emotion recognition)하는 연구가 대표적이다. 예를 들면 선거 후보자들의 토론을 지켜보는 패널들의 얼굴을 읽어, 각 후보에 대한 선호도를 측정 한 연구도 있다[33]. 이 연구에서는 2012 년 미국 대통령 선거 후보자 토론회에서 발췌한 비디오 영상을 실험자들에게 보여주고, 그들의 감정을 트래킹한 내용으로 선호도 예측 모델을 만들었는데, 결과적으로 실험자들이 밝힌 후보 선호도와 예측 결과가 거의 일치했다.

기계가 학습한 감정을 역으로 인간에게 가르치는 사례도 있다. 거울뉴런의 문제로 상대의 표정을 읽는 데 어려움을 느끼는 것으로 알려진 자폐아동에게 상대방의 감정을 학습시키는 사례가 대표적이다. 자폐증을 앓는 환자들은 상대방의 얼굴에 전반적으로 드러나는 윤곽(Configural Information)보다 특징적인 정보(feature information)에 더 의지해 상대의 감정을 인식하는 경우가 많다. 이를테면 상대의 행복을 입의 움직임을 통해 배우는 대신, 화난 상태와 두려워 하는 모습은 눈과 입과 이마의 정보를 모두 통합해야 하기 때문에 상대적으로 어려워 한다는 것이다. 이같은 경향은 성인이 될 수록 더 극심해 진다[41]. 따라서 구글 클래스에 설치된 프로그램이 상대의 감정을 머신러닝으로 학습하고, 이를 착용한 자폐증 환자에게 실시간으로 상대방이 느끼는 감정을 알려주는 방법이 제안되고 있다[46].

### 2.3 색채 시맨틱 연구

단어의 뜻을 알아내듯 색채에서 의미를 파악하려는 작업은 미학과 심리학은 물론 최근 HCI 분야에서도 활발하게 진행되고 있다. 1970년대부터 색채의 의미와 관련한 연구가 본격적으로

이어졌는데, 이를테면 Adams & Osgood 은 23 개 문화권에서 색채의 의미가 어떤 식으로 차이를 보이는 지를 연구해 각기 다른 색채 인식에 대해 발표하기도 했다[5].

여러 색상 조합 안에서 인간이 지각하듯 대표적인 색상을 꼽아, 이를 토대로 의미를 찾아나선 연구도 있다. Jahanian et al.은 12 종의 잡지 71 권에서 모은 표지 디자인 2654 건을 토대로 대표 색상과 어울리는 단어의 조합을 찾았다[22]. 예를 들어 실제 잡지 이미지를 모아보니, 정원수, 과학, 골프 같은 단어가 많이 나오면 연두색, 흰색, 짙은 녹색, 청록색, 풀색 같은 대표 색상 팔레트와 짝지어지는 것이다. 특히 색채 팔레트와 워드 클라우드를 매칭하는 과정에서 이중잠재 디리클레 할당 모형(LDA-Dual Model) 을 활용했는데, 본 연구에서도 해당 모델을 활용해 구조적 의미를 파악했다.

본 연구에서는 색채 의미론 연구를 따로 진행하지는 않았다. 앞서 밝혔듯 언어학에서도 통사론과 의미론은 다른 갈래에서 진행되는 연구다. 통사 구조는 문장의 형태 및 형식적인 면을 살피는 측에 속하고, 의미는 말 그대로 글이나 말의 내용을 살피고 뜻의 기원을 찾는 학문이기 때문이다. 따라서 언어학의 방법론을

따르는 만큼 본 연구에서는 구조주의적 측면만 살펴도록 한다. 추후에는 사진 속 위치 별 색채의 분포 등을 통사적으로 다룰 수도 있을 것이다.

### 제 3 절 사진의 구조에 대한 연구

사진의 구조와 관련해서는 미학과 심리학을 비롯한 여러 분야에서 연구가 이뤄진 바 있다. 이 가운데 인간의 시각을 중심으로 하는 인지과학 분야의 선행 연구를 살펴보고 보다 철학적인 관점에서 시각 인지를 어떻게 바라보는 지 살펴보았다.

#### 3.1 인간의 시지각에 대한 연구

신경미학(Neuroesthetics)을 탄생시킨 Zeki 는 Da Vinch 의 말을 인용하며 “시각계 세포의 길항적 작용(opponency)을 통해 밝혀진 생리학적 진실” 을 이야기 했다. 길항적 작용이란 Zeki 의 표현으로 “한 가지 색에 의해 흥분되는 세포는 그 보색 관계에 있는 색에 의해 억제되는 현상” 이라 할 수 있는데, 그만큼 인간의 미학에 대한 시지각 또한 신경 세포에 의한 활동임을 다시금 인식하게 한다. Zeki 는 미술이 “대상과 표면, 얼굴, 상황 등에 대한

항상적이고 지속적이며 본질적이고 영속적인 특성들을 표현” 하고 있다며, 이는 “일반화된 대상과 얼굴들에 관해서도 폭넓은 정보를 얻도록” 한다고 했다. 시각 활동을 주관하는 뇌의 부분에서 기능이 확장된다는 것을 과학적으로 밝혀내기도 했다[50].

Zeki 에 앞서 인간 시지각의 구조적인 요소를 파헤쳤던 Arnheim 은 사람이 그린 그림 속에서 시각적 균형을 맞추려는 노력이 보인다고 밝힌 바 있다. 이를 엔트로피 법칙에 대입해 알아본 결과, 그림의 중앙에 가중치가 부여되고 나아가 오른쪽보다 왼쪽에 대해 좀 더 힘이 실린다는 이론을 제기했다. 또한 극장에서 자리를 고르는 행위와 관련해 왼쪽 자리와 오른쪽 자리 중 어느 곳을 선호하는 지를 실험적으로 살핀 연구도 있다[24]. 이 같은 선행 연구를 토대로 본 연구에서도 왼쪽과 중앙부를 중심으로 하는 구조적 특이점을 살펴보고자 한다.

### 3.2 시각적 표상에 대한 철학적 논의

에크프라시스(Ekphrasis)는 레토릭 용어다. 시각적 표상이 기호적인 표상으로 해석되는 것을 뜻한다. 호메로스가 일리아드에서 아킬리스의 방패를 수 페이지에 걸쳐 기술했던 것을

생각하면 된다. 이 방패는 실제로 존재하는 것도 아니지만, 독자들은 이 무기에 대해 표현한 글자들을 읽음으로서 방패가 실재하는 것처럼 떠올리게 된다. 이 에크프라시스라는 용어가 바로 여기에서 나온 것이다. 어원을 따져보면 그리스 단어인 프라시스- 즉, 밖으로, 말하다 라는 뜻을 지닌 이 단어에 에크프라자인(ekphrazein) 즉 무생물을 이름으로써 선포하는 것을 합성한 것이다[20].

본 개념은 실제 학계에서도 수 차례 쓰였는데, 주로 디지털 테크놀로지학과 미디어 연구에서 활용됐다. 예를 들어 Proch 는 ‘산업적 에크프라시스’ 라는 용어를 썼는데, 이미지를 검색하는 행위에서 ‘구어적 레이블(verbal label)’ 을 쓴다는 점을 포착한 것이다. 이재현은 이것을 더 확장해 ‘디지털 에크프라시스’ 라는 용어를 썼는데, 이는 멀티미디어적 전환과 함께 변화한 에크프라시스의 개념을 뜻한다[3].

시각적 표상을 알고리즘 그 자체로 표상하는 ‘알고리즘 에크프라시스’ 라 일컬을 수도 있을 것이다. ‘알고리즘’ 이라는 용어 자체는 수년 새 다수 활용되고 있다. 대표적인 것이 Uricchio 가 주창한 ‘알고리즘 턴(Algorithmic Turn)’ 이다[45].



2010 년 Microsoft 사에서 내놓은 Photosynth 와 같은 시기에  
태동하기 시작한 Augmented Reality(AR)의 사례를 들며, 시각적  
이미지와 위치정보의 결합을 통해 보는 주체(subject)와 보여지는  
세계(객체, object)의 관계에도 변화가 왔다는 주장이다. 즉,  
하이데거식 welt-bild 에 대한 관점을 현 시대의 시각적 체계(visual  
regime)에서의 주체와 객체 간 관계에 대한 알고리즘적 재설정  
(Reconfiguration)으로 살펴봤다. 즉, 오직 한 사람(viewer)의  
관점에서만 객체를 보는 것이 아니라, 알고리즘으로 형성된 정보  
시스템을 통해 새로운 관점들이 형성되고 있다는 것이다. 예를  
들어 Photosynth 는 다수의 사람들이 찍은 한 장면들을 묶어  
하나의 현장을 만들어내는데, 그 안에는 고정된 시점이라는 것이  
없다. 초기 증강현실도 지도 이미지속 공간이 사람에게 관점 및  
시점을 주는 역할을 해내는 모습을 보인다. 맛집, 저렴한 주차장,  
불만한 명소라는 식으로 기호가 새겨짐으로써 사람에게 인식을  
주는 것이다.

Uricchio 도 인용한 바 있는 Culler 는 이같은 기호화 작업을  
통한 주체와 객체의 관계 정립에 대해 자세히 기술한 바 있다.  
여행자(traveler)는 사라지고 관광객(tourist)만 증가하는 세대가

자본주의의 영향으로 각 스팟마다 기호(marker)가 새겨졌기 때문이라는 주장이다. Culler 는 MacCannell 의 표현을 빌어 시각의 신성화 작업(Sight Sacralization)이 진행됐고, 그로 인해 관광객에게 어떤 기호적 이미지가 그려짐으로써 객체와의 상호작용이 발생했다고 썼다. 나이아가라 폭포에 간 방문객이 그것이 ‘나이아가라 폭포’임을 모르고 그저 그 폭포 자체를 본다면 지루해질 것이라는, “이름없는 풍경만큼 지루한 것도 없다(Rien n’ est plus ennuyeux qu’ un paysage anonyme)” 는 Prosper Mérimée 의 표현을 곁들였다[15][28].

결국 ‘알고리즘 에크프라시스’ 라는 용어는 결국, 시각적 표상을 표현하는 데 있어 산업적 에크프라시스나 디지털 에크프라시스에서 부차적으로 여겨지던 알고리즘적 표상을 뜻한다고 할 수 있다. 선행 연구에서는 프로그래밍 언어를 일리아드에서의 기호적 표상으로 다루기도 했다. 우리는 이같은 전통적 아이디어를 계승하되, 기계 언어의 기능에 대해서는 조금 다른 견해로 접근하고자 한다. 이재현은 컴퓨터 코드를 이미지를 불러오기 위한 호명적(interpellational) 이고 주석 달기 식의(annotational) 역할자로 언급했다[3]. 본 연구에서는 이 같은

컨셉을 기계적 역할자(machinable actor)로 재조명하고자 한다.  
기계 용어가 그저 이미지를 스크린에 불러내기만 하는 것이 아니라,  
시각적 기호도 생성할 수 있음을 제안하고자 한다.

## 제 3 장 데이터 수집 및 기계 분석

본 연구를 진행하는 데 있어 인간의 편견에 대한 선행 연구가 많은 유럽 언론과 중동 언론의 데이터를 수집해 그 안의 요소를 살피는 방법을 택했다. 이를 토대로 구조 및 위치별 구성 요소를 확인하고 최종적으로 이를 일반화할 수 있는 지 살펴봤다.

### 제 1 절 데이터 개요

로이터와 알 자지라 웹페이지에서 검색한 내용을 중심으로 크롤링(crawling)을 시행했다.<sup>9</sup> 내전 및 테러에 대해서는 2015년 1월부터 2016년 12월 31일까지 2년 간의 데이터를 대상으로 했다. 이 기간 일어난 사건 사고 가운데 국제적인 테러로 규정됐고, 5명 이상 사망한 사건을 살핀 결과는 [표 1]과 같다. 다만 터키의 경우, 이슬람 문화권으로 볼 것인지 혹은 유럽에 가깝다고 봐야할지 다소 모호한 부분이 있어 제외했다. 아래 사건을 위주로

---

<sup>9</sup> <http://www.reuters.com/search/pictures?blob=aleppo&sortBy=&dateRange=>

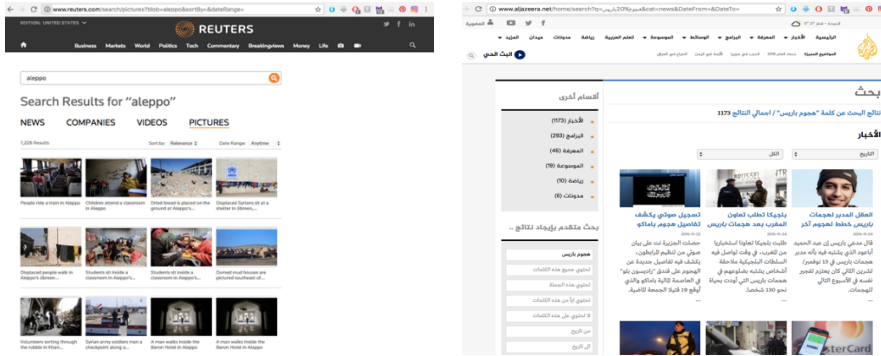
데이터로 수집했다. 수집 환경상 시리아 알레포에 대해서는 12월 21일까지를 다뤘다.

Data label	query	발생시기
Aleppo (중동 그룹)	Aleppo	2011~2016.12
Europe/USA (서구 그룹)	Paris attack	2015. 11
	Hebdo	2015. 01
	Berlin truck	2016. 12
	Nice truck	2016. 07
	Brussels attack	2016. 03
	Munich Olympia	2016. 07
	Orlando nightclub	2016. 06

[표 1] 데이터 수집 대상 분쟁 목록

먼저 로이터에서는 Aleppo, Paris attack, Hebdo, Nice truck, Berlin truck, Brussels attack, Munich Olympia, Orlando nightclub 등을 검색 쿼리로 썼다. 언론사가 대개 한 사태, 사건에 대해 하나의 단어로 통칭해 캡션 및 제목을 단다는 행태를 고려했다. 해당 사이트 서칭을 통해 Terror 보다 Attack 이 주로 쓰인다는 점을 확인했다. 다만 프랑스 니스나 독일 베를린에서는 Attack 보다 특정 테러 행위자인 Truck 이 주요 단어로 쓰였고, 독일 뮌헨과 미국 올란도에서는 사건이 벌어진 장소인 올랭피아 몰, 나이트클럽이 주로 활용됐다. 또한 파리 테러가 같은 해 두 차례

일어난 점을 고려, paris attack 은 주로 11 월에 일어난 사건을, hebdo 는 그에 앞선 샤를리엵도 언론사 테러 사건을 다루고 있기 때문에 따로 검색했다. 웹페이지를 그대로 크롤링하는 방식을 썼다.



[그림 5] 로이터 웹사이트(좌측)와 알 자지라 아랍어 웹사이트 검색 화면

알 자지라 웹사이트의 경우 영어 사이트가<sup>10</sup> 아닌 아랍어 사이트를<sup>11</sup> 대상으로 했다. 알 자지라 인터내셔널 영어사이트가 2005 년 문을 열면서 BBC 등에서 일한 스태프를 대거 채용했고, 그에 따라 영어권 사람들에게 맞춘 콘텐츠가 제공됐기 때문이다[42]. 본 연구에서 살피고자 하는 것은 아랍에서 바라보는 아랍, 아랍에서 바라보는 세계다. 따라서 아랍어권 사이트를 조사하는 것이 옳다고 판단했다. 검색 쿼리는 로이터와

<sup>10</sup> <https://www.aljazeera.com>

<sup>11</sup> <https://www.aljazeera.net>

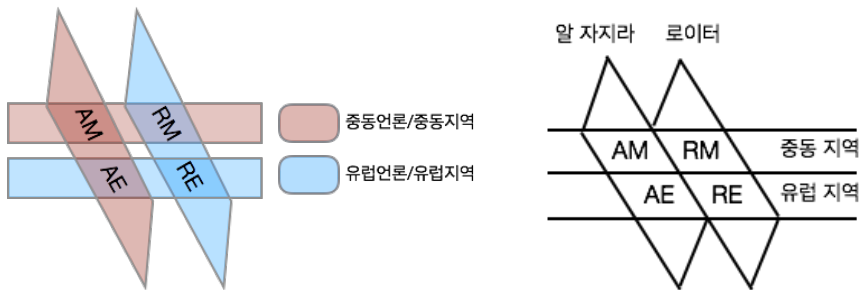
흡사하게 썼지만, hebdo 나 olympia 처럼 특정하지 않는 것이 결과적으로 더 나은 것으로 나타났다. 또한 니스 테러는 paris attack 과, 뮌헨 테러와 베를린 테러는 ‘독일 테러’ 와 사진이 모두 겹치는 것으로 나타나 paris attack, german attack 로 각각 통합해 검색했다. Aleppo, Paris Attack, German Attack, Orlando, Brussels Attack 등을 아랍어로 변환해 쿼리로 활용했다.

로이터 쿼리	사진 수	알 자지라 쿼리	사진 수
Aleppo	6861	حلب (aleppo)	11388
Paris attack	1287	باريس هجوم (paris attack)	634
Hebdo	348	الألمانية هجوم (german attack)	44
Nice truck	156	أورلاندو(Orlando)	21
Berlin Truck	288	بروكسل هجوم (brussels attack)	38
Brussels attack	487		
Munich Olympia	60		
Orlando nightclub	232		

[표 2] 수집한 데이터 수

중동은 알레포 사진을 중심으로 다루고, 유럽은 나머지 데이터를 통합해 묶는다고 해도 사진 수 차이가 다소 큰 편이다. 로이터의 경우 6861 장 대 2858 장, 알 자지라는 11388 장 대 737 장이다. 분석시 이를 정규화(normalize) 하기 위해 사진

총량을 각 수식에 활용했다. 데이터는 알 자지라 중동 그룹과 알 자지라 서구 그룹, 로이터 중동 그룹과 로이터 서구 그룹으로 묶었다.



[그림 6] 본 연구의 데이터 그룹 도식. AM은 알 자지라의 중동 지역, AE는 알 자지라 유럽지역, RM은 로이터의 중동 지역, RE는 로이터의 유럽 지역 보도다.

## 제 2 절 데이터의 편견에 대한 연구

사물 및 상황을 분석하기 위해 딥러닝 API를 활용했다. 다국적 기업에서 제공하는 API 가운데 성능이 좋은 것으로 Amazon, Microsoft, Google의 기능이 손꼽힌다. 이 가운데 사물에 대해서는 Amazon의 것을 썼고, 감정에 대해서는 MS의 기능을 활용했다. 경험적으로나 이들이 내어주는 결과 면에서 각각의 API가 본 연구에 적합하다고 판단했기 때문이다.



## 2.1 사물 및 상황 분석

수집된 사진은 취재와 편집, 보도를 거쳐 공개된 이미지들이다. 즉, 여러 장의 사진 가운데 인간의 의지에 따라 취사 선택된 것이다. 본 연구에서는 이 사진의 기저에 자리하는 인지적 작용을 알아보고자 한다. 먼저 어떤 의도에 의해 사진이 찍혔는지 확인해야 한다. 다만 여기서 말하는 의도는, 악의적인 왜곡이나 사기가 아닌 포토저널리스트(그리고 언론사 편집국)로서의 주관을 뜻한다. 이들은 사진을 ‘찍는’ 사람이지만, ‘만드는’ 이들이 아니고, 따라서 스토리텔링의 사진을 되도록 있는 현실 그대로 전해야 한다는 원칙이 있기 때문이다. “카메라는 거짓말하지 않는다, 사람이 그럴 뿐이다”라는 이론에 따라[13], 본 연구에서는 카메라 앵글이 향한 사물과 상황을 확인하고자 한다.



people	0.99264366
person	0.99264366
human	0.99257965
alley	0.93618599
alleyway	0.93618599
road	0.93618599
street	0.93618599
town	0.93618599
kicking	0.8101561
baby	0.66784905

[표 3] Amazon API 에 의해 검출된

사물 레이블과 신뢰도 점수

연구에서 활용한 아마존 서비스(Amazon Rekognition Service) API 는 사물과 상황에 대한 탐지를 효과적으로 한다. 동급의 구글(Google)이나 마이크로소프트 (MS)에서 제공하는 Open API 에 비해 사물 탐지를 상대적으로 상세하게 하고, 각 사물에 대한 신뢰도 점수까지 공개해 객관성을 높였다. 해당 서비스는 합성곱 신경망(CNN)과 순환 신경망(RNN) 등의 여러 딥 러닝 아키텍처가 적용돼있어, 복잡한 컴퓨터 비전 작업을 진행하는 데 용이한 편이다.<sup>12</sup> JPEG 과 PNG 형식이 모두 가능하고, 최대 15MB 크기의 이미지 파일을 사용할 수 있다. 분석을 위해 가로 및 세로 화소가 최소 80 픽셀 이상인 것을 썼는데, 전체 이미지 크기의 5% 이상이어야 객체가 제대로 인식된다고 한다.

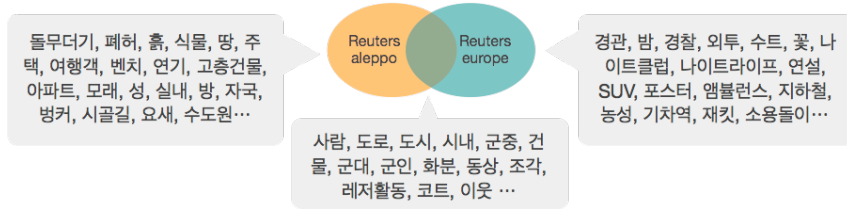
탐지 가능한 사물은 수천 개에 이르는데, 아마존 측은 이를 레이블(label)로 분류하고 있다. 이를테면 사람과 이벤트에 신부, 아기, 기타리스트가 들어가고, 음식과 음료에 사과와 샌드위치가, 동물에 개, 고양이가, 스포츠 및 레저에는 농구, 하키, 테니스가 들어가는 식이다. 레이블 탐지는 그 객체가 정말 그 객체일 가능성,

---

<sup>12</sup> <https://console.aws.amazon.com/rekognition/home?region=us-east-1#/label-detection>

즉 신뢰도 점수가 50% 이상인 것에 대해 식별토록 했다. 아마존 Rekognition API 로 추출한 값은 [표 3]과 같다. 코드는 파이썬 2.7.10 으로 짰다.

이후 각 그룹에 속한 사진들의 레이블과 신뢰도 점수를 매겼다. 이때 그룹당 출현한 레이블의 빈도수와 평균 신뢰도 값을 곱해 ‘빈도 X 신뢰도 점수’ 를 구했다. [별첨 1]에서 볼 수 있는 것과 같이 양쪽 그룹 모두 사람(Human, People, Person)의 점수가 높았다. 특히 분쟁사진에서 인물의 비율이 높다고 볼 수 있는 대목인 동시에, 사물 탐지 기술이 유독 인물을 잘 알아본다고도 해석할 수 있다.



[그림 7] 로이터 중동과 유럽 그룹의 공통 사물과 차별되는 사물

[별첨 1]의 ‘빈도 X 신뢰도 점수’ 를 토대로 그룹별 차별화된 요소를 알아봤다. 레이블을 100 개까지 늘려 비교할 때, 로이터의 그룹별 차이는 [그림 7]과 같다. 알레포에서는 주로 돌무더기나 폐허, 고층건물, 병커와 같은 배경 묘사 요소가 서구 그룹에 비해

많이 나타나고, 서구 그룹에서는 경찰관이나 꽃, 연설, 농성 같은 상황 묘사적 요소가 많이 추출됐다.

**로이터 중동 :** 돌무더기, 폐허, 흙, 돌, 식물, 땅, 주택, 여행객, 벤치, 연기, 고층건물, 아파트, 모래, 성, 실내, 방, 자국, 병커, 카우치, 시골길, 요새, 수도원, 오프로드(비포장도로), 들판, 탱크, 파괴, 눈, 보트, 동물, 지하실, 불

**로이터 서구 :** 경관, 밤, 경찰, 외투, 수트, 꽃, 나이트클럽, 나이트라이프, 연설, SUV, 포스터, 앰블런스, 지하철, 농성, 꽃봉우리, 기차역, 재킷, 꽃꽂이, 소용돌이, 머리, 보도, 음식, 판잣길, 인도, 통행로, 여성, 식물군

알 자지라의 경우, 알레포 그룹에선 연기와 안개 등 전쟁 현장의 즉각적인 풍경과 어린이들의 모습이 부각된 데 비해 서구 그룹에선 도시 사고 수습 현장의 모습이 주로 그려졌다. 분쟁의 진행형과 완료형의 차이가 두드러졌다.

**알 자지라 중동 :** 돌무더기, 연기, 흙, 탱크, 고층건물, 안개, 아파트, 이웃, 무기, 기계, 오프로드(비포장 도로), 비행기, 전투기, 엔진, 돌, 종이, 공장, 오염, 로고, 어린이, 콜라주, 아이, 배, 대형선박, 관광객, 땅, 스모그, 흙길, 아기, 항구, 자갈, 폐허, 불도저, 항만

**알 자지라 서구 :** 경관, 경찰, 엠블런스, 소방관, 택시도, 교차로, 가게, 보행자, 중형차, 꽃, 자전거, 바이크, 조명, 버스, 청중, 지하철, 기차역, 광장, 여성, 깃발, 주차, 주차장, 안경, 고글, 걷기, 음료, 벤치, 그릇, 도시광장, 의자, 수업, 수화물, 여행가방, 꽃봉우리

매체별로 같은 사건에 대해 바라보는 차이도 살펴봤다. 먼저 중동에 대해 로이터는 의류와 가방, 모래 같은 세세한 요소들에서 차별화됐고, 알 자지라는 기자회견 자리나 엠블럼 같은 공식적인 정보에 대한 부분이 두드러졌다.

**로이터 중동 :** 캡, 의상, 길, 데님, 청바지, 자전거, 바지, 바이크, 벤치, 예배, 그릇, 모래, 성, 동상, 조각, 자국, 병커, 미소, 카우치, 시골길, 요새, 수도원, 교차로, 가게, 교회, 웃음, 들판, 수화물, 여행가방, 성당, 파괴, 눈, 보트, 동물, 지하실, 광장, 도시광장, 불

**알 자지라 중동 :** 외투, 정장, 포스터, 조종사, 연설, 텍스트, 책자, 대기실, 접견실, 안개, 머리, 무기, 의류, SUV, 기계, 주요도시, 비행기, 나무, 항공기, 엔진, 엠블럼, 기자회견, 종이, 공장, 머리카락, 오염, 로고, 콜라주, 교실, 배, 선박, 스모그, 나무, 흙길, 향만, 자갈, 불도저, 항구

서구 그룹에 대해서도 서로 다른 양상을 보였는데, 로이터는

추모 현장, 사고현장(나이트클럽) 등 현장 중심적인 장면을 차별적으로 포착한 반면 알 자지라는 여기에서도 기자회견과 같은 공식적인 행사에 초점을 맞췄다.

**로이터 서구** : 어린이, 아이, 의상, 밤, 캡, 조각, 동상, 나이트클럽, 나이트라이프, 타워, 미소, 예배, 아기, 교회, 농성, 웃음, 길, 이웃, 재킷, 꽃꽂이, 소용돌이, 데님, 인도, 청바지, 통행로, 성당, 꽃

**알 자지라 서구** : 조종사, 대기실, 의류, 소방관, 응접실, 텍시도, 텍스트, 주요도시, 엠블럼, 나무, 책자, 머리카락, 중형차, 실내, 기자회견, 버스, 청중, 깃발, 주차장, 교실, 안경, 고글, 주택, 나무, 걷기, 음료, 벤치, 의자, 수업

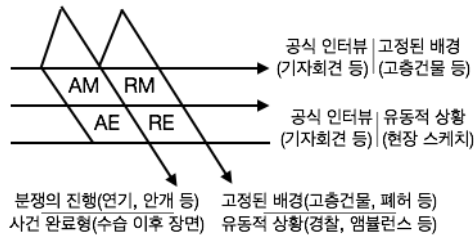
끝으로 각 개체들이 상대적으로 얼마만큼 자주 등장했는지 알아보기 위해 새로운 점수를 계산했다. 각 그룹별로 탐지된 객체들의 점수를, 그룹별 총 사진 수로 나누었다. 수식으로 표현하면 다음과 같다.

$$\frac{\text{신뢰도점수} \times \text{출현빈도}}{\text{그룹별 총 사진 수}} = \text{new\_score}$$

[별첨 2]에서 볼 수 있듯, 주로 인물이 가장 많이 나타난

그룹은 알 자지라 서구와 로이터 서구 그룹이다. 서구 그룹의 보도 특징 자체가 인물에 맞춰져 있음을 알 수 있는 대목이다. 그 중에서도 알 자지라 서구 그룹은 셀피(selfie)와 초상(portrait)의 점수가 로이터보다 높게 나타났는데, 이는 얼굴 자료사진을 쓴 특정 인물의 인터뷰나 수배범의 얼굴을 타 그룹에 비해 많이 쓴 것이라고도 볼 수 있다. 알레포에 대해선 알 자지라보다 로이터에서 더 풍경에 신경쓴 부분을 포착할 수 있다. 돌무더기나 길거리, 시내, 군대, 아이들의 모습이 서구 언론에서 상대적으로 더 높은 값을 보였다.

본 분석 결과를 토대로 그룹별 특징을 크게 세 가지로 나뉘 볼 수 있다. 첫째, 매체별 특성이 드러난다. 로이터는 주로 풍경에 대한 묘사가 많다. 알레포에 대해서는 돌무더기나 폐허, 어린아이, 시가지 같은 풍경 묘사를, 서구 테러에 대해선 수습 및 추모 모습 등을 주로 내보낸다. 알 자지라는 그에 비해 정보 전달 목적이 강한 편인데, 이를테면 기자회견 모습이나 현재 벌어지고 있는 전쟁 상황 이미지를 주로 활용한다.



[그림 8] 사물 및 상황 분석에 대한 정리

둘째로, 지역적 영향도 무시할 수 없다는 점을 밝힌다. 피해자와 가해자 구도가 뚜렷한 서구 테러와 ‘전쟁터’로서의 이미지가 강한 시리아 알레포 분쟁에서 나오는 대상의 차이가 수치적으로도 드러났다. 예를 들어 경찰이나 앰블런스가 등장하면 서구 테러 보도사진일 가능성이 높다. 반대로 돌무더기와 탱크가 등장하면 시리아 알레포 사진일 가능성이 높다. 사물의 차이뿐 아니라 비율적으로도 지역적 차이를 살펴볼 수 있을 것으로 보인다. 이와 관련해선 추후 연구에서 워드 분류 등을 통해 카테고리를 나눠 수식화하고, 관련한 통계 분석을 지속할 예정이다.

셋째, 그럼에도 불구하고 매체 내에서도 자국 문화권과 타문화권에 대한 시선에 차이가 있다. 로이터는 서구보다 중동에 대해 특히 도시의 모습과 같은 보다 큼직한 요소들에 대한 묘사를 더 많이 하고 있고, 알 자지라의 경우 중동보다 서구에 대해 유독



인물 얼굴에 초점을 많이 맞춘다. 증동에 대해선 현재 벌어지는 전쟁 상황을 시시각각 내보내다보니 전투기나 선박 같은 전쟁 관련 사물이 다수 출현하고, 서구에 대해선 사건현장의 경찰이나 여성들의 시위 모습 등이 나온다. 테러 순간보다 그 이후에 중심을 둔 것이다. 대부분 IS 소행으로 추정되는 사건이라는 점을 보도 당시 알 자지라 측에서도 다소 의식한 결과로 짐작된다. 또한 로이터가 먼 지역의 전쟁에 대해 풍경 위주의 사진을 실는 것은 Park 이 주장한 전통적 미국 포토저널리즘의 흐름을 뒷받침한다고도 볼 수 있다[38].

## 2.2 감정 분석

사진 데이터 속 인물들에 대한 감정 분석은 마이크로소프트 프로젝트 옥스포드 인지 서비스(Microsoft Project Oxford Cognitive Service)의 Emotional API 를 활용했다. 구글과 아마존 모두 감정 인식 서비스를 제공하고는 있지만, MS 서비스가 모두 여덟 가지 감정 값에 대해 소수점 아홉 자리까지 세분화해서 답을 준다는 점, 값들을 노멀라이즈해 상대적 비교를 손쉽게 한다는 점, 그리고 사용이 손쉽다는 점 등을 고려했다. 뿐만 아니라 이미 언론을

비슷한 여러 분야에서 해당 서비스를 근거로 분석을 시행하고 있다는 점도 함께 따져봤다<sup>13</sup>.

본 감정 분석 API는 딥러닝을 활용해 사진 속 인물의 얼굴을 탐지하고(face detection), 그 얼굴이 보이는 감정을 총 8 가지 레벨로 구분해(emotional recognition) 이를 수치로 설명하는 것을 골자로 한다. 감정은 분노(Anger)와 경멸(Contempt), 역겨움(Disgust), 공포(Fear), 행복(Happiness), 중립(Neutral), 슬픔(Sadness), 그리고 놀람(Surprise) 등으로 구분된다. 해당 서비스에서는 각 얼굴에서 나오는 감정 값의 총 합을 1로 노멀라이즈(normalize)한다. 따라서 각각의 값은 비율로 매겨진다. 예를 들어 행복이 지배적인 한 인물에 대한 감정 총합은 1로, 행복이 0.93이면 나머지 감정 값은 모두 더해 0.07이 된다.

이처럼 노멀라이즈가 되면 인물 간 상대적인 비교도 가능하다. 본 연구에 앞서 시행한 보도사진 속 국회의원 선거 후보자들의 얼굴 비교에 따르면, 각 후보자들의 사진 속 행복이 당선자 그룹보다 낙선자 그룹에서 통계적으로 유의하게 높은 것으로

---

<sup>13</sup> 머신러닝도 인정한 이정미 대행의 포커페이스. 지디넷코리아. 2017.03.10.  
[http://www.zdnet.co.kr/news/news\\_view.asp?artice\\_id=20170310181149](http://www.zdnet.co.kr/news/news_view.asp?artice_id=20170310181149)

나타났다[1]. 인물별로 나타날 수 있는 감정의 총합이 1로 고정되어 있기 때문에 감정 별 비율을 바로 비교할 수 있다.

본 API 서비스를 이용하기 위해 토큰을 받아 분석을 시행하는 방법을 썼다. 코드는 사물 탐지때와 마찬가지로 파이썬 2.7.10 으로 썼다. 코드를 여러 차례 돌려본 결과를 토대로, 정확도를 높이는 동시에 시간 효율 또한 고려하기 위해 한 얼굴 당 최대 5초 동안, 최대 10번까지 분석을 시도하도록 지정했다. 다만 해당 API 서비스에서 제공하는 8 가지 감정값 가운데 경멸과 역겨움에 대해서는 아직 인식률이 낮다<sup>14</sup>. 각 그룹별로 나온 얼굴의 수는 [표 4]와 같다.

알 자지라 중동	알 자지라 서구	로이터 중동	로이터 서구
7301	876	838	669

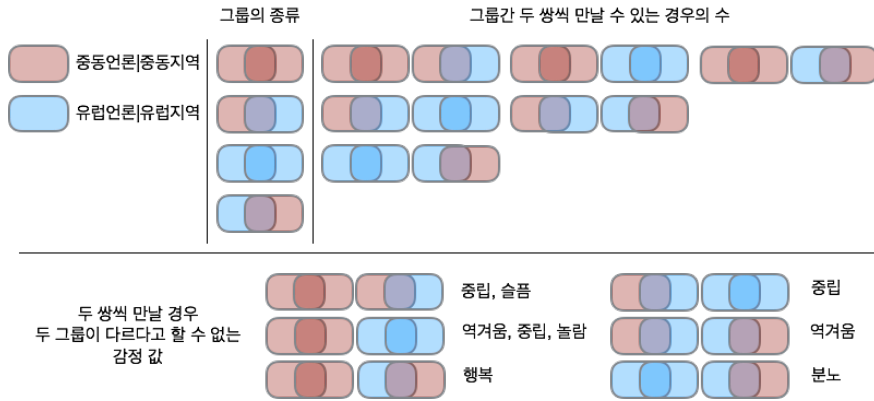
[표 4] 각 그룹별로 추출된 얼굴 수

먼저 각 그룹의 차이가 통계적으로 유의한 지 알아보기 위해 각 감정 조합에 대해 등분산성 테스트를 시행했다. 결과적으로 각 그룹에서 대부분의 감정이 분산의 차이( $p < 0.05$ )를 보이고 있는

<sup>14</sup> Microsoft Cognitive Service, Emotion API.  
<http://dev.projectoxford.ai/docs/services>

것으로 나타났다. 다만 알 자지라의 알레포 보도사진과 서구 보도사진에서의 중립과 슬픔, 로이터의 알레포 사진과 서구 사진에서의 분노, 알 자지라의 알레포 사진과 로이터의 알레포 사진 그룹에서의 행복, 알 자지라의 서구 사진과 로이터의 서구 사진 그룹에서의 중립, 알 자지라가 바라본 알레포와 로이터가 바라본 서구의 역겨움과 중립, 그리고 놀람, 마지막으로 알 자지라에서 본 서구와 로이터에서 본 알레포에서의 역겨움에서는 귀무가설을 기각한다. 적어도 이 감정들에 대해선 차별된 그룹이라고 보기 힘들다는 것이다.

이는 매체별·지역별 특성으로도 해석할 수 있는 대목이다. 알 자지라가 바라본 전쟁과 테러는 지역을 막론하고 중립과 슬픔을 차별없이 보도했다고 할 수 있고, 로이터의 경우엔 분노에서 비슷했다고 볼 수 있다. 알레포에 대한 보도사진에선 알 자지라, 로이터 할 것 없이 행복에서 차이가 없다고 볼 수 있고, 서구에 대해선 중립적인 감정에 차별이 덜했다. 마지막으로 각 매체가 자신의 지역을 보도할 때 역겨움과 중립, 놀람에 있어서 비슷한 분포를 보였고, 다른 지역을 보도할 때 역겨움에 대해 차이 없이 보도했다고 볼 수 있다.



[그림 9] 각 그룹을 둘씩 짝지을 때 차별된다고 보기 힘든 감정들 ( $p>0.05$ )

마지막으로 각 그룹이 어떤 감정 값을 보이고 있는 지 살펴보았다. [표 12]에서 볼 수 있는 것과 같이 로이터가 보도한 중동 사진에서 거의 대부분의 감정 값 평균이 높게 나왔고, 알자지라가 바라본 유럽에서 경멸의 감정이, 로이터가 바라본 유럽에서 중립의 감정이 다른 그룹에 비해 높게 나왔다.

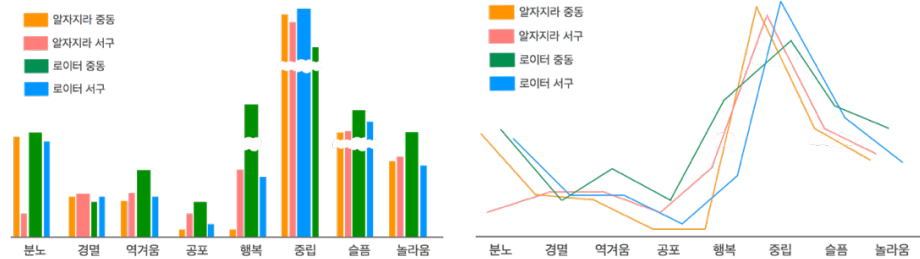
	AM	AE	RM	RE
분노	0.0233513	0.00320715	0.025879	0.0229063
경멸	0.007367	0.0098980	0.0068638	0.0079205
역겨움	0.0068422	0.0090644	0.01005	0.007051
공포	0.0019143	0.003511	0.0062501	0.002887
행복	0.0019544	0.1038009	0.125059	0.0864905
중립	0.7857	0.7771	0.6882	0.7914
슬픔	0.032977	0.037533	0.100583	0.059120
놀람	0.0211407	0.0270511	0.0370719	0.0221828

[표 5] 각 그룹별 감정 평균

본 분석에 따르면 크게 두 가지를 확인할 수 있다. 눈에 띄는 점은 알 자지라가 바라보는 유럽 이미지에 경멸 값이 다른 그룹에 비해 높다는 것과 로이터가 바라보는 유럽 이미지에 중립 값이 비교적 높다는 점이다. 앞서 밝힌 바와 같이 경멸 값은 아직 정확성이 떨어지기 때문에 전자에 대해서는 의미를 부여하기 힘들 것으로 보인다. 다만 유럽 언론이 바라보는 유럽 테러의 이미지에 중립이 많다는 점은 살펴볼 필요가 있다. 실제 데이터를 살펴보면, 로이터가 보도한 유럽 테러 이미지에 피해자보다는 현장 주변에 배치된 경찰이나 군인의 이미지가 많다. 같은 지역을 보도한 알 자지라의 경우 브리핑을 하는 중동의 주요 지도자나 EU 관계자의 얼굴이 많은 편이다.

같은 시리아 알레포 내전에 대해 로이터와 알 자지라가 보도한 사진의 감정 값 차이도 살펴볼 필요가 있다. 특히 로이터에서 대부분의 감정 값이 높게 나오는 부분에 주목할 필요가 있다. 전쟁 보도에 있어 객관성을 지키는 관행 중 하나가 공식적 루트에 의지하는 것인데, 정부에서 받은 소말리아의 굶주린 아이들 사진이나 터키 난민들의 영상을 내는 것이 오직 객관적인 행위인가에 의문을 제기한 연구도 있다[34]. 국가에서 만든

프레임 자체가 편향돼 있다면, 그 프레임 대로 보도하는 것일 뿐 객관적이라고 볼 수 없다는 것이다. 이 연구를 참고할 때, 로이터의 사진이 객관성을 필두로 외려 편향된 보도를 했을 수도 있다는 가정을 해볼만하다.



[그림 10] 그룹별 감정 추이

### 제 3 절 신뢰도 확인을 위한 파일럿 테스트

기계 분석의 신뢰도를 확인하기 위해 일반인을 대상으로 사진에 대한 이미지 분석 설문을 시행했다. 24 세~39 세 성인 남녀 21 명을 대상으로 실시한 이번 설문조사는 워드프레스 기반 홈페이지에 설문조사 플러그인을 설치해 시행했다. 질문 구성은 별도 첨부한 바와 같다. 각 8 문항씩 총 32 문제로 구성했고, 각 그룹별로 사진 및 단어 균을 두 개씩 섞어 배치했다.

질문 1 은 2 장 2 절에서 논의한 대로 사안별 차별화된 단어

군에서 랜덤으로 발췌한 단어를 토대로 그에 대한 공감도를 확인했다. 네번째 질문 역시 같은 맥락에서 그룹별 사진을 번갈아 랜덤으로 제시하며 얼마만큼 응답자의 이야기마냥 공감이 되는지를 확인했다. 두 번째와 세 번째 질문은 감정 인식과 사물 탐지에 있어 인간과 기계가 얼마나 비슷하게 작동했는지 알아보기 위해 준비했다.

결과적으로 사진의 사물 탐지 및 감정 인식에 있어서 상당히 높은 일치율을 보였다. 감정인식의 경우, 귀무가설은 ‘인간의 감정 판단과 컴퓨터의 감정 판단이 동일할 것이다’ 였다. 먼저 응답자 스물 한 명을 하나의 컴퓨터 프로그램이라고 할 때, 이들이 선택한 감정에 대해 확률 값을 구했다. 즉, A 사진에 대해 분노를 11 명이 선택하고, 행복을 7 명이, 중립을 2 명이, 슬픔을 1 명이 선택했다면 각 감정은 분노 0.52381, 행복 0.33333, 중립 0.09524, 슬픔 0.04762 의 값을 가지게 된다. 본 연구에서 활용한 MS 감정인식 프로그램의 값을 토대로 사진 속 인물들의 평균 감정값을 구한 결과와 그룹별로 대조해 통계적 유의도를 확인했다. 설문에 쓴 여덟 장의 사진 가운데 두 장은 MS 프로그램상 얼굴 인식이 되지 않아 제외했다. 여섯개 그룹에 대해



Kruskal-Wallis 검정을 시행한 결과, 각 그룹별로 모두  $X^2(7)=7$ ,  $p=0.4289$  라는 통계값을 도출할 수 있었고, 귀무가설을 기각할 수 없는 것으로 판명됐다. 즉, 인간과 MS 감정인식 API의 판단 간에 통계적으로 차이가 있다고 보기 어려웠다.

사물 탐지에 있어서도 8 장 가운데 두 장을 제외하곤 사람이 사진을 보고 제일 먼저 탐지한 사물의 대다수를 아마존 프로그램이 알아맞혔다. [표 6]에 따르면 특히 사람이나 사람의 직업군에 대한 탐지에서 많은 경우 일치하는 모습을 보였다. 뿐만 아니라 담요나 사람의 등, 쓰레기와 같은 비교적 세부적인 물체에 대해서도 잘 알아맞혔다. 다만, 사진에 인물 한 명만 집중돼 나온 경우나, 사람들이 뒤엉켜 있는 모습에 대해서는 컴퓨터의 탐지가 다소 더딘 경향을 보였다.



[그림 11] 사람과 기계가 탐지한 사물의 차이가 큰 사진

RM 사람	RM 컴퓨터	RM 사람	RM 컴퓨터	RE 사람	RE 컴퓨터	RE 사람	RE 컴퓨터
담요	사람	군인	사람	눈물	사람	쓰레기	쓰레기
사람	모터	사람	군대	포옹	춤	사람	모자
얼굴	오토바이	군복	군인	문신	춤자세	길	어린이
	차량	철근	군복	사람	탱고		길
	시장	전쟁	군사	등	등		건물
	가게			흑인	군중		시내
	담요						광장
	집장식						군중
	퀵트						교차로
AM 사람	AM 컴퓨터	AM 사람	AM 컴퓨터	AE 사람	AE 컴퓨터	AE 사람	AE 컴퓨터
전쟁	사람	연설	사람	공항	사람	성	건축
군인	군대	마이크	초상	층	경관	깃발	돔
무기	군사	사람	셀피	무장	경찰	국기	성당
로켓	군인	안경	얼굴	경찰	보울	건물	교회
군복	의복	모자		사람		흐린날 씨	송배
사람		마이크		겨울		하늘	법원
RPG		꽃		수색			건물
두건		의상		군인			성
농성		히잡		순찰			지붕

[표 6] 각 사진별 사람과 컴퓨터의 사물 탐지 차이

각 그룹별 단어군 및 사진에 대한 공감률 분석에선 그룹별로 조금씩 차이를 보였다. 먼저 단어군에 대한 거리감은 대부분 중간 수준, 즉 ‘그다지 남의 이야기 같지도, 내 이야기 같지도 않은’ 정도에 머물렀다. 하지만 [그림 12]에서 볼 수 있는 바와 같이

사진에서는 다소 차이가 나타났다. 로이터 알레포 사진에 대해선 대부분이 공감에 되지 않는다는 쪽으로 손을 들었고, 알 자지라의 알레포에 대해서는 그보다 낮은 점수로 몰렸다. 같은 중동 내전 사진이어도 로이터가 알 자지라보다 더 추상적이거나, 공감이 어려운 사진을 보도하는 것은 아닌지 눈여겨 볼 만한 부분이다.



[그림 12] 단어군에 대한 공감도(좌측)와 사진을 보고 느끼는 공감도. Y 축이 5에 가까울 수록 심리적 거리가 먼 것이다. X 축 1 부터 RM, RM, RE, RE, AM, AM, AE, AE.

## 제 4 장 이미지 속 위치에 따른 분석

앞선 연구에서 추출한 사물과 감정을 토대로, 이번 절에서는 통사적 특성을 살펴보기로 한다. 통사 구조는 어느 사진에서나 적용될 수 있는 통시성과 함께 그 위치만이 갖는 특질을 보여줄 수 있어야 성립된다[14].

논문 전반부 실험에서는 인간의 편견과 기계가 찾아낸 사진 속 사물 사이에 일정한 패턴을 확인했다. 하지만 기계가 인간의 편견을 자동으로 알아내기 위해서는 요소 하나하나가 아닌 이들이 이루는 구조, 즉 문법을 파악할 필요가 있다. 이를 위해 두 가지 방법으로 접근했다. 첫 번째는 사물이 사진 안에서 얼마만큼의 비중을 차지하고 있는가에 대한 것이고, 두 번째는 이들 사물의 사진 안에서의 위치를 파악해 인간의 시각 행위와 비교하는 것이다. 이는 문장을 쪼개 위치에 따른 역할을 확인하는 위치통사론을 방법론적으로 차용했다[35]. 문장과 마찬가지로 사진 속 사물이 위치에 따라 어떤 역할을 하는지 알아내고자 하는 것이다.

본 연구에서는 기술적인 한계로 오직 인물을 통해서만 사진 속 문법을 확인했다. 인물은 앞선 연구에서도 확인했듯, 사진에서

가장 높은 평균 신뢰도 점수 및 빈도를 기록했다. 또한 얼굴의 위치는 MS 의 API 를 통해 계산할 수 있었다. 이들의 비중과 위치별 등장 횟수를 먼저 확인했다. 사물 인식을 위해 활용한 Amazon 측 API 는 사물의 레이블과 신뢰도 점수만 내 놓고 각 위치는 내어주지 않는다. 이는 Amazon 측에도 문의를 한 내용이나, 최종적으로 포지션에 대한 정보는 받을 수 없는 것으로 결정됐다.

## 제 1 절 사진 속 인물의 비중 및 위치에 대한 연구

먼저 인물의 얼굴이 사진 전체에서 얼마 만큼의 비중을 차지하는 지 살펴봤다. 얼굴은 기계가 가장 먼저 파악한, 즉 눈에 가장 잘 띄는 것으로 판명된 대표 인물의 것을 썼다. 구조를 확인하는 데 있어 제일 주요하게 인지된 얼굴이, 결국 촬영 기자가 사진에서 제일 강조하는 인물이라고 해석했다. 사진 속 얼굴의 사이즈는 MS API 에서 제공한 가로(width), 세로(height) 픽셀을 사용했다.

그룹별로 살펴보면 얼굴의 크기는 유럽 테러에 대한 보도 그룹에서 대체로 높게 나타났다. 로이터의 유럽 그룹이 4.35%로 네 그룹 가운데 가장 큰 대표 얼굴 비중을 보였고, 알 자지라의

유럽 그룹은 4.03%로 그 뒤를 이었다. 가장 비중이 작은 곳은 로이터의 중동 그룹으로 2.77%를 기록했다. 로이터가 중동그룹에서 특히 다른 언론에 비해 얼굴을 작게 내보낸 것으로 볼 수 있는 부분이다. 이는 앞선 선행연구에서 제시됐듯, 먼 지역의 사건에 대해서는 풍경을 위주로 보도하는 서구 언론 고유의 전통이 반영됐기 때문으로도 해석할 수 있다[38]. 즉 인물이 사진에서 주요한 역할을 하지 않는, 배경 중 하나일 수 있다는 것이다.

또한 중동의 상황을 주로 전달하는 알 자지라의 역할 영향도 고려할 수 있다. 이들은 시리아 내전을 전할 때와 마찬가지로, 유럽의 테러리즘과 관련해서도 중동 측 인사의 인터뷰 기사를 더 많이 내보냈다. 인터뷰 사진의 경우 사람의 얼굴이 사진 속 주요 요소이기 때문에 비중을 크게 잡는다. 따라서 양 언론 모두 유럽 테러를 다루는 사진 속 인물의 비율이 다소 크게 나타난 것으로 해석할 수 있다.

인물의 사진 속 위치와 관련해서는 선행연구에서도 다룬 Arnheim 의 이론을 참고했다. 사람들은 사물을 볼 때 왼쪽에서 오른쪽으로 읽는 경향을 보인다. 이것은 독서 습관에서 나온

것이라는 견해가 많다. 나아가 사람들이 왼쪽에 대해 오른쪽보다 더 많은 비중을 두기 때문이라는 분석도 있다. Arnheim 은 엔트로피 이론을 차용해, 물리적 세계의 균형성을 시각적 구조 안에서도 찾을 수 있다고 밝힌 바 있다[7][8]. 얼굴의 위치는 MS API 에서 제공한 사진 상단으로부터의 좌표(top)와 왼쪽으로부터의 좌표(left)를 토대로 얼굴의 중심이 되는 좌표를 계산해 사용했다.

[표 7]에서 볼 수 있는 것과 같이 그룹별로 얼굴의 대부분이 사진 좌측과 중앙, 그리고 상단에 밀집한 것으로 나타났다. 특히 좌측상단(left-top)과 중앙상단(center-top)에서 높은 비율을 보인다. 그룹별로 보면 알 자지라 중동 그룹에서는 왼쪽 상단부터 중앙중앙으로 이어지는 ‘ $\square$ ’ 자 구역에 얼굴이 밀집했고, 로이터 중동 그룹에서는 왼쪽 상단부터 우측상단, 중앙중앙으로 이어지는 ‘ $T$ ’ 자 구역에 얼굴이 몰려 있는 것을 볼 수 있다. 알 자지라 유럽 그룹과 로이터 중동 그룹은 시그마 ‘ $\sigma$ ’ 모양으로 분포했다. 그룹별 차이는 카이스퀘어 검증 결과 통계적으로 유의하게 나타났다 ( $p < 2.2e-16$ ).

102 (14.85%)	272 (39.59%)	59 (8.59%)	70 (19.44%)	160 (44.44%)	31 (8.61%)
81 (11.79%)	96 (13.97%)	38 (5.53%)	28 (7.78%)	35 (9.72%)	25 (6.94%)
17 (2.47%)	14 (2.04%)	8 (1.16%)	4 (1.11%)	3 (0.83%)	4 (1.11%)
AM			AE		
71 (13.52%)	258 (49.14%)	46 (8.76%)	92 (23.96%)	139 (36.20%)	38 (9.90%)
21 (4.00%)	86 (16.38%)	22 (4.19%)	31 (8.07%)	44 (11.46%)	16 (4.17%)
10 (1.90%)	9 (1.71%)	2 (0.38%)	9 (2.34%)	5 (1.30%)	10 (2.60%)
RM			RE		

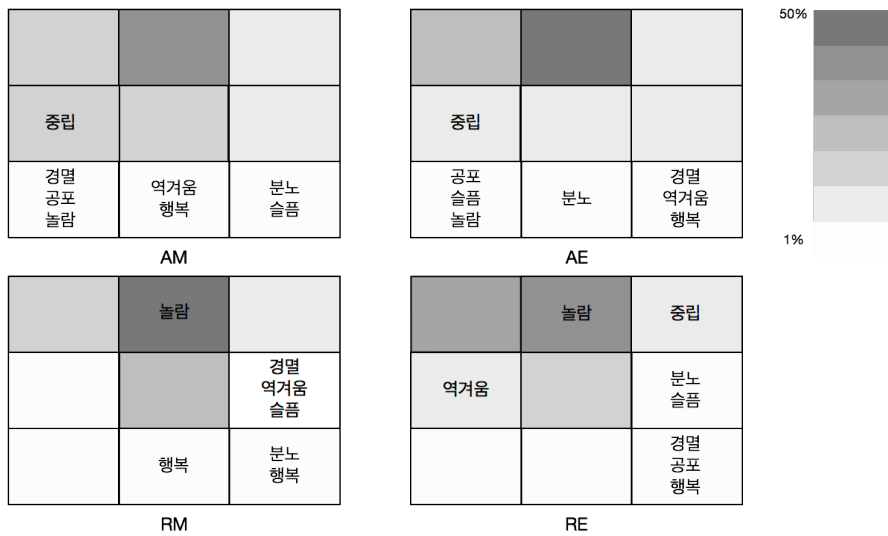
[표 7] 그룹별 사진 위치에 따른 주요 인물 얼굴 등장 빈도

## 제 2 절 사진 속 인물의 위치별 감정에 대한 연구

사진 속 각 위치가 가지는 역할을 확인하기 위해 인물의 감정 값을 입력 분석했다. 앞서 추출한 여덟가지 감정 값의 중앙값(median)이 각각 가장 높게 나타난 위치를 그룹별로 살펴봤다. 이때 중앙값을 측정한 이유는 다음과 같다. 각 위치별로 얼굴이 나타난 사진의 수의 차이가 크고 감정 값의 편향(Skew) 정도가 크기 때문에 일반적으로 사용하는 평균을 쓸 경우 아웃라이어의 영향을 많이 받게 된다. 또한 최빈값을 활용해 한계치(threshold)를 정한 뒤 이 수치를 넘어서는 감정 값들에



대해서만 사진 수를 추리는 방법도 확인했다. 하지만 위치별 얼굴 사진 수의 갭(gap)이 크기 때문에 모든 감정 값이 중앙상단(center-top)에서 모두 높게 나타나는 현상이 발생한다. 따라서 비교적 자유로운 중앙값을 확인했다.



[그림 13] 중앙값으로 확인한 각 감정 값이 가장 높게 나온 위치

[그림 13]에 따르면 알 자지라의 사진들은 사진 하단부에 대체로 감정값이 높게 나오는 것을 알 수 있다. 다만 이 위치는 얼굴 빈도수가 낮은 편이어서 감정값이 높아도 그 효과를 기대하긴 쉽지 않을 것으로 보인다. 대신 중립이 좌측중앙에서 가장 높게 나타난다. 이 위치는 Arnheim 의 이론대로라면 눈길이 잘 가는 위치이기도 하다. 반면 로이터는 얼굴이 가장 많이 등장한 중앙

상단에서 놀람의 중앙값이 제일 높게 나타났다. 알 자지라에 비해 중앙에 출현하는 얼굴에 대해서 특히 극단적인 감정을 보이고 있다고 볼 수 있는 대목이다. 뿐만 아니라 로이터 유럽 그룹의 경우에는 역겨움과 중립도 사진 좌측 중앙과 우측 상단에서 각각 높은 값을 보였다.

분노	AE	분노	RM	분노	AE
경멸	AE	경멸	AE	경멸	AM
역겨움	AM	역겨움	RM	역겨움	AE
공포	RM	공포	RM	공포	AE
행복	AE	행복	AE	행복	RM
중립	RE	중립	AM	중립	RM
슬픔	RM	슬픔	RM	슬픔	RE
놀람	RE	놀람	RM	놀람	AE
분노	RM	분노	AM	분노	RE
경멸	AE	경멸	RE	경멸	AE
역겨움	RM	역겨움	AM	역겨움	RE
공포	RM	공포	AE	공포	RM
행복	RM	행복	RM	행복	RM
중립	AM	중립	AE	중립	AE
슬픔	RE	슬픔	RM	슬픔	RE
놀람	RE	놀람	AE	놀람	AM
분노	AE	분노	AE	분노	AE
경멸	RE	경멸	AE	경멸	AE
역겨움	AE	역겨움	AE	역겨움	AE
공포	AE	공포	AM	공포	AM
행복	RE	행복	AM	행복	AE
중립	AM	중립	RM	중립	RM
슬픔	AE	슬픔	AM	슬픔	AE
놀람	RE	놀람	AM	놀람	AM

[표 8] 사진 속 위치별 각 감정의 최대 값을 지니는 그룹

이번에는 그룹간 비교를 진행했다. 각 그룹의 위치별 중앙값을 모두 살펴본 결과, 감정별로 그 위치에서 가장 높은 값을 차지한 그룹은 [표 8]에 나온 것과 같다. 전반적으로 눈에 잘 띄는

지역(중앙중앙, 중앙상단, 좌측상단)에서 로이터 중동 그룹과 알 자지라 유럽 그룹이 높은 감정 값을 보였다. 알 자지라의 유럽 그룹과 로이터의 중동 그룹이 높은 감정 값을 보였다. 특히 이 두 그룹이 이 위치에서 긍정적인 감정 값이 상대적으로 높은 점은 살펴볼 만 하다. 만일 사진 정 중앙에서 행복 값이 유독 높다면 이는 로이터 중동 테러 사진일 가능성이 높은 것이다. 또 사진 중앙 상단에서 행복 값이 높으면 유럽테러를 보도하는 알 자지라 사진일 가능성이 높다. 그만큼 알 자지라와 로이터가 타 문화권에 대해 보도하는 사진들이, 다른 그룹에 비해 눈에 잘 띄는 곳에서 행복 값을 높게 내보내고 있었다. 실제로 사진을 살펴보면, 로이터가 내보내는 중동 분쟁 사진 가운데 행복한 모습의 어린이 사진이 종종 발견된다. 비극적인 상황을 보여주기 위한 방어적 장치로 볼 수도 있고, 또는 시리아의 일상적인 풍경 중 하나를 포착한 것일 수도 있다.

방어적 장치로 행복한 감정을 노출했을 것이라는 해석은 이 지역의 슬픔(sadness) 지수를 통해 방증할 수 있다. 사진의 좌측 상단과 중앙 상단, 중앙 중앙 부분의 슬픔이 다른 그룹에 비해서 상대적으로 높게 나타났다. 즉, 로이터가 중동 분쟁 모습을 보도할

때, 다른 그룹보다 슬픔을 더욱 부각시켰을 가능성이 있다. 이같은 경향성을 볼 때 행복 감정 또한 부정적인 감정을 돋보이게 하기 위한 장치였을 것으로 분석했다.

나아가 알 자지라 유럽, 로이터 중동 그룹이 많은 위치에서 높은 감정을 보인 점은 다시 한 번 살펴볼 법 하다. 9 곳 8 개 감정 총 72 개 포스트 가운데 알 자지라 유럽은 절반에 가까운 27 곳, 로이터 중동은 19 곳에서 가장 높은 감정 값을 보였다. 그만큼 타 문화권을 바라보는 양 언론사의 보도 사진 속 인물들이, 위치 별로 더 극단적인 감정을 표현했을 가능성이 높다. 앞서 밝혔듯 본 감정 API 는 8 개 감정의 합을 1 로 노멀라이즈 한 값이기 때문에, 어느 한 감정의 값이 높으면 다른 감정들의 값은 낮게 나온다. 따라서 어느 위치에서 특정 감정 값의 중앙값이 높게 나왔다는 것은, 타 감정 값들에 비해 훨씬 높은 값들이 더 많이 나왔을 가능성을 내포하고 있다.

### 제 3 절 사진의 통사적 구조에 대한 결론

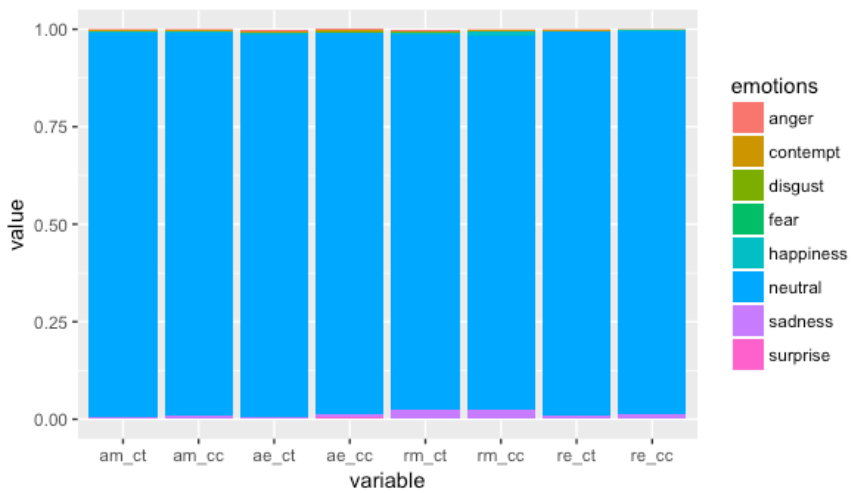
위 결과를 토대로 각 위치를 통해 사진을 그룹별로 구분할 수 있는 지 살펴 보고, 더불어 사진 속 사물의 영향도 함께 살펴봤다.

이를 위해 여러 장의 사진 가운데 그룹을 예측할 수 있는지 살펴봤다.

여러 장의 사진이 있을 때 어느 그룹에 속하는 지에 대해서는 [그림 14]와 같이 설정할 수 있다. 본 연구에서는 다른 레이블 대신 얼굴을 중심으로 진행했다. 따라서 눈에 잘 보이는 위치에 얼굴이 자리한, 인물 중심 사진을 대상으로 그룹을 분류해 봤다. 얼굴의 중심이 중앙 상단이나 정중앙에 있을 때 각 감정값들이 어떤 비율로 분포하고 있는 지 구했다. 비교를 위해 각 감정의 중앙값 합이 1 이 되도록 다시 노멀라이즈 작업을 했고, 그 결과 [그림 14]와 같이 누적 막대 그래프로 구현할 수 있었다.

눈여겨볼 만한 점은 로이터 중동 그룹의 지표다. 얼굴의 중심이 정중앙에 있을 때는 다른 그룹보다도 중립 값이 상대적으로 낮은 값을 보였는데, 이는 중앙 상단에서도 마찬가지였다. 다른 그룹에 비해 중립이 아닌 다른 감정을 더 극단적으로 잘 표출했다고 해석할 수 있다. 대신 얼굴이 중앙 하단에 중심을 두고 있을 경우, 중립적인 감정이 전체 감정의 99.98%를 차지하는 것으로 나타났다. 적어도 인물을 중심으로 하는 사진에서 만큼은, 사진의 중앙부에서 보다 살아있는(vivid) 감정을 위주로 보여주는 사진일

가능성이 높고, 하단으로 갈 경우 풍경 중심이거나 공식적인 인터뷰의 사진일 가능성이 높다고 해석할 수 있다. 역으로 말하면 사진 중심은 살아있는 감정을 보여주는 역할을, 사진 하단은 풍경을 더 살리는 역할을 한다고도 볼 수 있다.



[그림 14] 여러 장의 사진이 있을 때 얼굴 위치별(중앙상단, 중앙중앙) 감정 값의 비율에 따른 그룹 분류

앞선 연구 목표였던 통시성(일반화 가능 여부)과 위치의 특이성을 만족하는 지 살펴봤다. 결과적으로 본 연구에서 제시한 사진의 구조에 따른 편견 확인이 가능하려면 사진의 수가 많고, 인물이 등장하는 것일 수록 좋으며, 레이블 상 풍경에 대한 인지가 뚜렷해야 한다. 여러 장의 사진 가운데 중립의 비율이 낮고 슬픔의

감정이 사진 주요 부분에서 높다면, 이는 중동을 보도한 로이터의 사진일 가능성이 높다.

역으로 이같은 분류를 통해 편견을 구분해 낼 수 있다는 가능성도 확인했다. 로이터는 알 자지라에 비해 극단적인 감정을 보다 강조하는 경향을 보였다. 같은 시리아 내전 보도에 대해 풍경을 주로 내는 한편, 인물이 나오면 놀람과 공포를 한층 강하게 보도한 것이다. 이를 앞서 논의한 표상에 배치해 풀어보면, 결국 내전이라는 현장에서 로이터 촬영 기자 및 데스크는 인간의 비극적 감성을 조금 더 대표적으로 뽑아낸 것이라고 할 수 있다. 다만 이것이 독자로 하여금 심리적 거리감을 줄 수 있다는 점은 본 연구의 파일럿 테스트를 통해 확인할 수 있었다. 실제 로이터 기사를 통해 확인한 결과, 고객군인 타 언론사에 사진을 보내는 입장인만큼 일반적인 인터뷰 사진보다 상황이 더 부각되는 스포츠 사진에 공을 들인다고 했다.

어떤 보도가 현장을 그대로 재현해 낸 사진인 지는 현장에 있지 않는 이상 알 수 없다. 다만 로이터는 이를 감성으로 접근해 구현하고자 했고, 알 자지라는 정보 전달을 중심으로 풀어가고자 했다. 인간의 비극을 전달하는 방법에 있어서는 양측 시각에

치우침이 있을 수 있겠으나, 내전과 테러의 참혹함을 보도하고 있다는 점에 대해서는 두 매체 모두 동일하다고 볼 수 있다.



## 제 5 장 결론 및 연구 의의

본 연구에서는 보도 사진에 내포된 편견을 기계적으로 찾아내고, 그 결과를 토대로 어떤 요소가 편견을 일으키는 지역으로 찾아보고자 했다. 결론적으로 사진 위치에 따라 등장 인물의 감정값이 다르고, 이를 토대로 인물 중심 사진인지 풍경 중심 사진인지를 포착할 수 있다. 또한 인물 중심 사진에서도 언론사마다 표출하는 감정값과 보여주는 레이블에서 차이를 보이는데, 이를 바탕으로 편견을 확인할 수 있다.

### 제 1 절 연구 요약

중동 언론인 알 자지라의 중동 및 유럽 테러 사진과, 유럽 언론인 로이터의 중동 및 유럽 테러 사진을 통해 인간의 편견 여부를 기계적으로 그리고 사람 대상 설문조사로 확인해 봤다. 결과적으로 기계가 검출하는 사물과 인물의 감정에서 양 그룹의 차이가 나타났다. 공식인터뷰가 많은 알 자지라의 경우, 중동의 입장을 대변하는 경향을 드러내 보였다. 로이터는 서구의 전통적 보도 관습처럼 중동의 내전 현장에 대해서는 보다 풍경이 묘사되는

경향성을 보였고, 물리적으로 비교적 가까운 곳인 유럽 지역 테러에 대해서는 구체적인 상황 묘사를 보였다.

감정에 있어서도 그룹별 차이가 확인됐다. 로이터 중동 그룹이 타 그룹에 비해 감정이 전반적으로 높게 나타났는데, 다만 중립은 로이터 서구에서, 경멸은 알 자지라 서구에서 타 그룹에 비해 높게 나타났다. 매체 별로 보면 중립에 있어서는 알 자지라 그룹이 로이터에 비해 전반적으로 높은 편인데, 이는 인터뷰를 많이 내보내기 때문으로 파악할 수 있다. 그에 반해 역겨움이나 슬픔, 분노 면에서는 로이터가 조금 더 높은 수치를 보였다.

이를 위치 별로 따져본 결과, 전반적으로 중립의 값이 타 그룹보다 높은 것으로 나타났던 로이터 유럽 그룹은, 오히려 눈에 잘 띄지 않는 곳- 이를테면 사진 우측 상단이나 하단- 에서의 중립 값이 높았다. 이 부분의 중립값들이 평균을 끌어 올린 것이다. 즉, 사진 전체적으로 중립 감정이 로이터 유럽 그룹에서 높게 나타났다고 해서, 꼭 그 그룹에서 눈에 띄는 감정이 중립 감정은 아니었다. 잘 보이지 않는 인물이나 위치에서의 중립 값이 높았다. 반대로 로이터 중동 그룹의 경우, 절대적인 감정 값은 전체 그룹에서 가장 낮았다. 하지만 눈에 잘 띄는 곳에서는 다른

그룹보다 감정 값들이 상대적으로 높게 나타났다. 경멸의 감정 값 평균이 높게 나타난 알 자지라 서구 그룹의 경우, 실제 위치상으로도 눈에 잘 띄는 위치에서 경멸 값이 높았다. 이는 경멸 사진을 의도적으로 잘 보이는 곳에 배치했을 가능성도 제기할 만한 대목이다.

이처럼 단순히 사물의 배치와 인물의 감정을 사진 전체에서 추리는 것을 넘어서서, 사람의 시각이 닿는 위치에 따라 구조적으로 이미지를 해석하는 시도가 본 연구에서 진행됐다. 시각적으로 잘 보이는 위치에 따라 그룹별로 각기 다른 타입의 감정 배치 또한 확인했다. 본 연구의 결과물은 다음과 같은 조건 하에서 재현될 수 있다.

1. 서로 다른 시선을 가진다고 가정할 수 있는 두 매체를 대상으로 해야 한다.
2. 동일한 사건을 어떻게 바라보는 지를 살펴야 하므로 이 두 매체는 동일한 사건에 대해 보도해야 한다.
3. 특히 인물에 대한 검출 값이 높은 만큼, 사진 속에는 인물이 한 명 이상 등장해야 한다.
4. 만일 배경 위주라면, 배경 속에 나타나는 사물 및 상황에

대한 레이블을 살펴보도록 한다.

5. 사진 화소는 적어도 가로 세로 각 80 픽셀이어야 사물 및 인물 검출이 가능하다.

이런 가운데 분석 방법은 두 가지로 나뉠 것이다. 사진에서 인물이 탐지될 경우와 탐지되지 않을 경우다. 만일 인물이 탐지된다면 인물의 얼굴이 위치한 부근의 감정값을 토대로 비교할 수 있다. 사진 중앙부에서 공포와 놀람이 다른 그룹에 비해 높다면 이는 알 자지라가 보도하는 중동 사진일 가능성이 높은 것이다. 이후 이를 둘러싼 레이블에 외투, 포스터, 골목 같은 요소가 포함돼 있다면 알 자지라 중동 그룹의 사진일 가능성은 더욱 올라간다.

인물이 탐지되지 않는 사진의 경우, 현재로서는 오직 레이블에 의지해야 하는 상황이다. 인물의 감정을 알 수가 없기 때문이다. 따라서 보다 풍경적인 부분에 집중했다면 로이터 중동 그룹일 가능성이 높고, 기자회견과 같은 보다 공식적인 사진일 경우에는 알 자지라의 사진일 확률이 높다고 봐야 한다. 차후 기술적 한계를 뛰어넘어 사물 레이블 별 위치에 따른 해석도 추가돼야 할 것이다.

## 제 2 절 연구의 시사점

이 연구는 크게 세 가지 범주에서 활용될 수 있다. 먼저 사회과학 연구다. 기존에는 사람이 직접 코딩하던 정성 연구에 기계적인 방법을 도입했다. 정치학이나 미디어 연구에서 주로 쓰이는 방식은 연구자가 직접 편견 여부를 확인하며 점수를 매기는 것이다. 여기에는 인간 코더 간 교차 분석이 필요할 뿐 아니라 다룰 수 있는 데이터의 양에서도 한계가 존재한다. 본 연구에서 다른 것과 같이 기계적인 감정 인식과 사물 인지를 활용하면 객관성이 담보될 뿐 아니라 분석 가능한 데이터의 양도 늘릴 수 있다.

나아가 사회과학 분야, 특히 미디어에서 기술적으로 활용될 수 있다. 언론에서 보도하는 이미지의 중립성에 대한 모니터링 지표로 삼거나 사진 데이터에 대해 토픽 별로 그룹화하는 것을 자동화할 수도 있다. 기존에는 사람이 레이블을 일일이 입력해야 했지만, 본 연구 방법대로라면 레이블의 구조를 활용해 어느 내용을 이야기하는지 그룹별로 분류할 수 있다. 이를 토대로 카드뉴스 자동 생성이나 영상 자동 편집 등을 기계로 구현해낼 수도 있다. 이전에는 인간이 직접 이미지와 텍스트를 째지었지만, 기계적으로

통사 구조를 통한 문맥적 이해가 이뤄진다면 충분히 자동화될 수 있는 영역이다.

두 번째로 컴퓨터 비전 연구에서의 활용가능성도 높다. 본 연구에서는 기계 학습의 피쳐 엔지니어링(feature engineering)에 있어서 예측 가능한 변수를 언어학적으로 풀어가고자 했다. 사진의 구조에 대해 딥러닝은 전반적인 패턴으로 구현하지만, 실제 어떤 변수들이 어느 영향을 주는 지에 대한 명확한 근거는 찾기가 쉽지 않다. 따라서 본 연구는 기계학습의 각 변수에 대한 구조적 인과관계를 살피고, 이를 통해 향후 오류를 잡거나 더 세부적인 예측 모델을 세우는 데 기여할 수 있을 것으로 보인다.

융합적 의의도 있다. 컴퓨터 비전을 활용해 사회과학 연구의 주제를 탐색했고, 이 과정에서 언어학적 방법론을 차용했다. 현상학 이론을 토대로 기계의 표상에 대한 연구를 고민했고, 이에 따라 인간의 시지각에 대한 기존 인지 심리학의 이론도 참고했다. 결국 인간의 편견이라는, 설명이 쉽지 않은 영역에 대해 여러 학문을 융합해 기계적으로 풀어가고자 했던 점이 가장 주요한 의의로 손꼽힐 것이다.

### 제 3 절 연구의 한계 및 제언

본 연구는 대규모 사진 분석에 있어 데이터를 구조적으로 살피고, 이에 내포된 규칙을 찾아내려 했다는 데에 의의가 있다. 다만 크게 기술적인 면과 이론적인 면에서 한계를 가지고 있다. 먼저 사물 분석 시 사용한 Amazon의 API가 각 레이블의 위치를 전송하지 않는 방침을 하에 제공되고 있었다. 사진 속 사물 위치에 따른 차이를 확인하기가 어려웠다. 이와 관련해서는 다른 딥 러닝 방법을 통하거나 혹은 표본 조사, 또는 직접 코드를 작성하는 방향으로 진행하는 편이 나올 것으로 보인다.

또한 이를 이론화하기에는 많은 부분에서 가이드라인 상의 제한점이 있다. 무엇보다 알 자지라와 로이터의 언론사 특성상 이 같은 차이가 나타난 것일 수도 있다. 이를 중동 언론과 유럽 언론으로 확대해 해석하는 것에 대해 이론의 여지도 있다. 언론사만의 특징인지 여부는 타 언론사와의 교차 분석을 통해 확인하고, 테러 사진만의 특징인지 혹은 일반 사건사고 사진에서도 비슷한 성향이 나오는 지에 대해서도 확대 분석할 필요가 있다.

아직 잔혹함이나 사진에서 느껴지는 슬픔을 ‘해석’이라는 단어로 담아내기에 한계는 많다. 이를테면 알 자지라의 시리아

내전 보도 사진을 자세히 뜯어보면, 곳곳에 흩어진 시신과 잔혹한 고문, 분진에 뒤덮인 아이들의 우는 얼굴처럼 기계가 감지하지 못하는 수 많은 장면들이 존재한다. 일일이 레이블을 달 수 없는 요소들이다. 슬픈 사진, 잔혹한 사진으로 인간이 분류한 데이터를 가지고 기계에게 학습을 시켜, 그들의 어렵פות한 패턴을 읽게 하는 방식 정도가 현재로서는 가능할 것 같다.

그럼에도 불구하고 본 연구는 편견이라는 주제에 대해 사진에 내포된 시각적 요소를 구조적으로 확인하려 했다는 점에서 인지과학 및 현상학적 의의로도 발전 시켜나갈 수 있다. 향후 연구를 통해 인간의 의식에서 드러나는 행동에 대해 기계가 해석을 할 수 있는 여지를 지속적으로 확인해 갈 예정이다.



## 참고문헌

- [1] 유재연 · 서봉원. (2017) 보도사진 속 국회의원 선거 후보자의 감정과 당선의 관계. *Proceedings of HCI Korea 2017*, 146-149.
- [2] 이규임. (1987) 보도사진저널리즘의 특성과 기능. *언론중재* 22 ( ' 87.3), 8-18.
- [3] 이재현. (2009) 디지털 에크프라시스: “멀티미디어로의 전환” 과 언어적 표상. *한국언론학보* 제 53 권 5 호, 2009.10, 244-267.
- [4] 임재호. (2005) 위치통사론을 통해 살펴 본 몇 가지 통사론의 본질적 문제. *인문언어(Lingua Humanitatis)* 7, 271-289.
- [5] Francis M. Adams and Charles E. Osgood. (1973) A Cross-cultural study of the affective meanings of color. *Journal of cross-cultural psychology*, Vol. 4, No. 2, June 1973.
- [6] Gediminas Adomavicius, Alexander Tuzhilin. (2015) Context-Aware Recommender System. In *Recommender Systems Handbook*, 191-226.
- [7] Rudolf Arnheim. (1971) *Entropy and art an essay on disorder*

and order. *Entropy and Art*, 7-10.

[8] Rudolf Arnheim. (1983) *The Power of the Center: A Study of Composition in the Visual Arts*. *Journal of Aesthetics and Art Criticism* 41 (4), 448-450.

[9] Andrew W.Barrett, Lowell W.Barrington. (2005) *Bias in Newspaper Photograph Selection*. *Political Research Quarterly*, Vol. 58, No. 4 (December 2005), 609-618.

[10] Andrew W.Barrett, Lowell W.Barrington. (2005) *Is a Picture Worth a Thousand Words?: Newspaper Photographs and Voter Evaluations of Political Candidates*. *The International Journal of Press/Politics*, Vol 10, Issue 4, 98-113.

[11] Mike Berry, Inaki Garcia-Blanco, Kerry Moore. (2015) *Press Coverage of the Refugee and Migrant Crisis in the EU: A Content Analysis of Five European Countries*. Report prepared for the United Nations High Commission for Refugees.

[12] Oliver Boyd-Barrett, Terhi Rantanen. (1998) *The Globalization of News*. *The Globalization of News*, SAGE Publications, 1-14.

- [13] Howard Chapnick. (1994) Finding an Ethical Structure in Photojournalism. In: Truth Needs No Ally: Inside Photojournalism, University of Missouri Press, 305-312.
- [14] Noam Chomsky. (1957) Syntactic Structures. The Hague/Paris: Mouton, ISBN 978-3-11-021832-9.
- [15] Culler, J. (1990) Framing the sign: Criticism and its institutions. Norman: University of Oklahoma Press.
- [16] Hubert L. Dreyfus. (1992) What Computers Still Can't Do: A Critique of Artificial Reason, Cambridge, Mass: MIT Press.
- [17] Eco, Umberto. (1973; 2009) 서문. In 가짜전쟁(원제 *Semiologia Quotidiana*), 열린책들, 5-10.
- [18] Eco, Umberto. (1983; 2009) 투명성의 상실. In 가짜전쟁(원제 *Semiologia Quotidiana*), 열린책들, 229-255.
- [19] Harvey V. Fondiller. (1979) The Best of Popular Photography. Watson-Guptill Publications, 272.
- [20] Chambers Harrap. 1993. The Chambers Dictionary. Edinburgh.
- [21] Baotian Hu, Zhengdong Lu, Hang Li, Qingcai Chen. (2014)

Convolutional Neural Network Architectures for Matching Natural Language Sentences. *Advances in Neural Information Processing Systems* 27 (NIPS 2014)

[22] Ali Jahanian, S.V.N. Vishwanathan, Jan P. Allebach. (2017) Colors —Messengers of Concepts: Visual Design Mining for Learning Color Semantics, *ACM Transactions on Computer-Human Interaction*

[23] Lu C., Krishna R., Bernstein M., Fei-Fei L. (2016) Visual Relationship Detection with Language Priors. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016*. ECCV 2016. *Lecture Notes in Computer Science*, vol 9905. Springer, Cham.

[24] George B. Karev. (2000) Cinema Seating in Right, Mixed and Left Handers. *Cortex*, 36, 747-752.

[25] Haewoon Kwak, Jisun An. (2016) Revealing the Hidden Patterns of News Photos: Analysis of Millions of News Photos through GDELT and Deep Learning-based Vision APIs. *News and Public Opinion (NECO' 16, collocated with ICWSM' 16)*.

- [26] Y Liu, D Zhang, G Lu, WY Ma. 2007. A survey of content-based image retrieval with high-level semantics. *Pattern recognition*, Vol 40, Issue 1, 262- 282.
- [27] Lydia Manikonda, Munmun De Choudhury. (2017) Modeling and Understanding Visual Attributes of Mental Health Disclosures in Social Media. *ACM Transactions on Computer-Human Interaction*
- [28] MacCannel, D. (1976) *The tourist*. New York: Schocken.
- [29] S. Magnussen, M.W. Greenlee. 1999. The psychophysics of perceptual memory. *Psychological research* no. 62.
- [30] Manovich, Lev. (2011) *From reading to pattern recognition*. <https://manovich.net>
- [31] Manovich, Lev. (2011) *Trending: The Promise and the Challenges of Big Social Data*. In M. K. Gold(Ed.), *Debates in the Digital Humanities*, Minneapolis, MN: The University of Minnesota Press, 460-475.
- [32] Lev Manovich. 2014. *Watching the world*. <https://manovich.net>
- [33] Daniel McDuff, Rana El Kaliouby, Evan Kodra, Rosalind

Picard. (2013) Measuring Voter' s Candidate Preference Based on Affective Responses to Election Debates. Humaine Association Conference on Affective Computing and Intelligent Interaction.

[34] Annabel McGoldrick. (2006) War Journalism and 'Objectivity' . Conflict & Communication online, Vol. 5, No. 2.

[35] Jean-Claude Milner. (1989) Introduction a une science du langage.

[36] Hans Moravec. (1988) 마음의 아이들 (원제 Mind Children), 78-93.

[37] Stig A. Nohrstedt & Rune Ottosen. (2008) War Journalism in the Threat Society: Peace journalism as a strategy for challenging the mediated culture of fear?. Conflict & communication online, Vol. 7, No. 2.

[38] David Park. (1999) Picturing the War: Visual Genres in Civil war News. The Communication Review, Vol. 3(4), 287-321.

[39] R. W. Picard. (1995) Affective Computing. M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 321, 3.

- [40] Katharina Reinecke, David R. Flatla, Christopher Brooks. (2016) Enabling Designers to Foresee Which Colors Users Cannot See. CHI' 16.
- [41] Keiran M. Rump, Joyce L. Giovannelli, Nancy J. Minshew, Mark S. Strauss. (2009) The Development of Emotion Recognition in Individuals with Autism. *Child Development*, September/October 2009, Volume 80, Number 5, 1434-1447.
- [42] Naomi Sakr. (2006) Challenger or lackey?: The Politics of News on Al-Jazeera. *Media on the Move: Global Flow and Contra-Flow*. Communication & Society Series. Routledge, London, UK.
- [43] Liangcai She, Bo Long, Weiyi Meng. (2009) A latent topic model for complete entity resolution, IEEE international Conference on data engineering, 880-891
- [44] Christian Szegedy, Alexander Toshev, Dumitru Erhan. (2013) Deep Neural Networks for Object Detection. *Advances in Neural Information Processing Systems 26 (NIPS 2013)*.
- [45] William Uricchio. (2011) The algorithmic turn: photo synth, augmented reality and the changing implications of the image, *Visual*

Studies, 26:1, 25-35

[46] Catalin Voss, Peter Washington, Nick Haber, Aaron Kline, Jena Daniels, Azar Fazel, Titas De, Beth McCarthy, Carl Feinstein, Terry Winograd, Dennis Wall. (2016) Superpower glass: delivering unobtrusive real-time social cues in wearable systems. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16), 1218-1226.

[47] Tina Walber, Ansgar Scherp, Steffen Staab. (2014) Smart Photo Selection: Interpret Gaze as Personal Interest. CHI '14.

[48] Hartmut Wessler, Manuel Adolphsen. (2008) Contra-flow from the Arab world? How Arab television coverage of the 2003 Iraq war was used and framed on Western international news channels. Media, Culture & Society, Vol. 30(4), 439-461.

[49] Michael Wheeler. (2010) The Problem of Representation: In Handbook of Phenomenology and Cognitive Science, Springer Science+Business Media, 319-335.

[50] Semir Zeki. (1999; 2003) 이너비전: 뇌를 보는 그림 뇌로 그리는 미술 (원제 Inner Vision: An Exploration of Art and the



Brain), 14-22.

[별첨 1] 로이터와 알 자지라의 사진 속 사물 분석 결과 (상위 20 개)

로이터 중동				로이터 서구			
label	신뢰도점수	Freq	score1	label	신뢰도점수	Freq	score2
Person	0.964075345	4168	4018.266038	Person	0.971821291	2136	2075.810278
People	0.985233489	3937	3878.864246	People	0.984541713	2061	2029.14047
Human	0.991202308	3886	3851.812169	Human	0.991064574	2032	2013.843214
Town	0.599563143	2930	1756.720009	Crowd	0.683825476	871	595.6119896
Rubble	0.763457988	1953	1491.033451	Vehicle	0.78569382	659	517.7722274
Road	0.593332328	2497	1481.550823	Downtown	0.56549676	713	403.1991899
Street	0.59872834	2230	1335.164198	Road	0.559915263	671	375.7031415
Alley	0.606408618	1954	1184.92244	Plant	0.870285851	430	374.2229159
Alleyway	0.607744569	1926	1170.51604	City	0.549629408	659	362.2057799
Vehicle	0.779143358	1415	1102.487852	Military	0.71876207	503	361.5373212
Downtown	0.577064161	1871	1079.687045	Town	0.557635605	643	358.559694
Building	0.584119859	1825	1066.018743	Parade	0.676795749	482	326.215551
Military	0.727085103	1211	880.5000597	Potted Plant	0.921622097	350	322.567734
Child	0.568604252	1421	807.9866421	Car	0.786124594	405	318.3804606
Army	0.758978597	1057	802.240377	Carnival	0.615406518	509	313.2419177

Kid	0.569987702	1388	791.1429304	Street	0.556274368	556	309.2885486
Outdoors	0.586877971	1301	763.5282403	Festival	0.616884659	500	308.4423295
Soldier	0.73926818	1030	761.4462254	Urban	0.542556207	549	297.8633576
City	0.562688182	1274	716.8647439	Architectur e	0.665202986	440	292.6893138
Architectur e	0.601621398	1188	714.7262208	Clothing	0.62834173	445	279.6120699
알 자지라 중동				알 자지라 서구			
	신뢰도점수	Freq	score1		신뢰도점수	Freq	score2
Person	0.957443341	7367	7053.485094	Person	0.972437369	609	592.2143576
People	0.966603234	7149	6910.246516	People	0.98329378	587	577.1934486
Human	0.990966412	6782	6720.734203	Human	0.99121035	578	572.9195823
Vehicle	0.769710434	2098	1614.852491	Vehicle	0.782871482	184	144.0483527
Rubble	0.785168032	1833	1439.213002	Crowd	0.702879117	194	136.3585487
Crowd	0.708524555	1909	1352.573376	Selfie	0.556573151	188	104.6357523
Army	0.781996418	1615	1262.924215	Clothing	0.683288213	148	101.1266555
Military	0.73876999	1640	1211.582783	Car	0.740950442	126	93.35975574
Selfie	0.54486665	2060	1122.4253	Portrait	0.584433619	154	90.00277725
Clothing	0.670442117	1616	1083.434461	Overcoat	0.721330185	110	79.3463203
Town	0.572941257	1863	1067.389561	Speech	0.814635811	94	76.57576622

Soldier	0.748731336	1414	1058.706109	Parade	0.683110809	112	76.50841064
Outdoors	0.597047739	1688	1007.816583	Military	0.728600975	102	74.31729946
Downtown	0.577409426	1738	1003.537582	Suit	0.75386491	95	71.61716648
Building	0.570958434	1757	1003.173968	Plant	0.82264789	87	71.57036644
Road	0.570224478	1729	985.9181216	Van	0.670532879	105	70.40595234
Portrait	0.567031968	1567	888.5390937	Face	0.561947535	123	69.11954685
Car	0.693633421	1275	884.3826115	Officer	0.845702497	81	68.50190224
Face	0.546233485	1465	800.2320561	Downtown	0.571710617	119	68.03356342
Poster	0.765209472	994	760.6182152	Automobil e	0.770513388	84	64.72312458

[별첨 2] 각 그룹의 사물별 환산 점수

AM	New Score	AE	New Score
Person	0.61937874	Person	0.80354729
People	0.60680071	People	0.78316614
Human	0.59015930	Human	0.77736714
Vehicle	0.14180299	Vehicle	0.19545231
Rubble	0.12637978	Crowd	0.18501838
Crowd	0.11877181	Selfie	0.14197524
Army	0.11089956	Clothing	0.13721391
Military	0.10639118	Car	0.12667538
Selfie	0.09856210	Portrait	0.12212045
Clothing	0.09513825	Overcoat	0.10766122
Town	0.09372932	Speech	0.10390198
Soldier	0.09296681	Parade	0.10381059
Outdoors	0.08849811	Military	0.10083758
Downtown	0.08812237	Suit	0.09717390
Building	0.08809044	Plant	0.09711040
Road	0.08657517	Van	0.09553046
Portrait	0.07802415	Face	0.09378500
Car	0.07765916	Officer	0.09294695

Face	0.07026976	Downtown	0.09231148
Poster	0.06679120	Automobile	0.08781970
Parade	0.06454618	Soldier	0.08396665
City	0.06386055	Road	0.08179597
Overcoat	0.06280447	City	0.08059743
Smoke	0.05984068	Police	0.07257986
Suit	0.05981991	Army	0.07229007
Plant	0.05885045	Potted Plant	0.06903048
Architecture	0.05855512	Military Uniform	0.06317636
Military Uniform	0.05792811	Town	0.06252233
Street	0.05709049	Carnival	0.06202511
Flyer	0.05670667	Festival	0.06056799
RM	New Score	RE	New Score
Person	0.58566769	Person	0.72631570
People	0.56534969	People	0.70998616
Human	0.56140681	Human	0.70463373
Town	0.25604431	Crowd	0.20840167
Rubble	0.21732013	Vehicle	0.18116593

Road	0.21593803	Downtown	0.14107739
Street	0.19460198	Road	0.13145666
Alley	0.17270404	Plant	0.13093873
Alleyway	0.17060429	City	0.12673400
Vehicle	0.16068909	Military	0.12650011
Downtown	0.15736584	Town	0.12545825
Building	0.15537366	Parade	0.1141412
Military	0.12833407	Potted Plant	0.11286484
Child	0.11776514	Car	0.11139974
Army	0.11692761	Carnival	0.10960179
Kid	0.11531014	Street	0.10821852
Outdoors	0.11128527	Festival	0.10792243
Soldier	0.11098181	Urban	0.10422090
City	0.10448400	Architecture	0.10241053
Architecture	0.10417231	Clothing	0.09783487
Hat	0.09901726	Military Uniform	0.09117430
Urban	0.09139809	Hat	0.08817483
Car	0.08534302	Officer	0.08663419
Crowd	0.08409645	Automobile	0.08376080

Ruins	0.07871840	Selfie	0.08187211
Soil	0.07662301	Soldier	0.08001191
Baby	0.07014925	Building	0.07429721
Rock	0.06528790	Outdoors	0.06760960
Clothing	0.06434085	Portrait	0.06535835
Neighborhood	0.06113458	Army	0.06360991



[별첨 3] 신뢰도 확인을 위한 파일럿 테스트 설문조사지

1) 다음 단어군이 ‘얼마나 내 주변 풍경과 가까운지’를 1~5 척도로 써 주세요.

<폐허, 흙, 연기, 모래, 성, 들판>

[가깝다] 1 ----- 2 ----- 3 ----- 4 ----- 5 [멀다]

2) 다음 사진 속 인물(들)은 어떤 감정을 보이고 있나요?

(1) 분노 (2) 경멸 (3) 역겨움 (4) 공포

(5) 행복 (6) 중립 (7) 슬픔 (8) 놀람



3) 다음 사진에서 제일 먼저 보이는 객체(사물이나 배경, 또는 생물이나 상황)를 차례대로 3개 써 주세요. (예시: 강아지, 축제, 모자)



4) 다음 사진이 얼마만큼 내 이야기 같은지를 1~5 척도로 써주세요.



[내 이야기 같다] 1 -- 2 -- 3 -- 4 -- 5 [다른세상 이야기 같다]

Abstract

**A Study on the Syntactic  
Structure and Media Bias of  
Press Photographs:**

**Focusing on the analysis of visual data on the  
media reports of Europe and the Middle East**

유재연 (Jaeyoun You)

디지털정보융합전공 (Digital Contents and  
Information)

The Graduate School of Convergence Science and  
Technology

Seoul National University

This study addresses a methodology for the machine to adapt grammatical interpretation of images. Contrary to humans, who can 'closely-read' the photo, machine interprets the image based on large data set, so the deduction is unavoidable. We propose that applying syntax analysis to the image could offset losses and improve machine to perform contextual interpretation. To prove this assumption, we've measured the bias between European and the Middle East media with deep-learning api, which detects objects and emotions in photos. Accordingly, we could statistically calculate differences on views of both media. We also have examined the weight and the position of each object, to trace the syntax structure. Thus, we identified the grammatical substance in news images. Our goal is to build up a basis of structural analysis on photographs, thereby expecting to contribute to the development of computer vision.

**Keywords: Syntactic Methodology; Media bias; Image Recognition; Affective Computing; Data Journalism**

**Student Number: 2015-26101**