

# 안내 로봇을 향한 관람객의 행위 인식 기반 관심도 추정

## Estimating Interest Levels based on Visitor Behavior Recognition Towards a Guide Robot

이 예 준<sup>1</sup> · 김 주 현<sup>2</sup> · 정 의 정<sup>3</sup> · 김 민 규<sup>†</sup>

Ye Jun Lee<sup>1</sup>, Juhyun Kim<sup>2</sup>, Eui-Jung Jung<sup>3</sup>, Min-Gyu Kim<sup>†</sup>

**Abstract:** This paper proposes a method to estimate the level of interest shown by visitors towards a specific target, a guide robot, in spaces where a large number of visitors, such as exhibition halls and museums, can show interest in a specific subject. To accomplish this, we apply deep learning-based behavior recognition and object tracking techniques for multiple visitors, and based on this, we derive the behavior analysis and interest level of visitors. To implement this research, a personalized dataset tailored to the characteristics of exhibition hall and museum environments was created, and a deep learning model was constructed based on this. Four scenarios that visitors can exhibit were classified, and through this, prediction and experimental values were obtained, thus completing the validation for the interest estimation method proposed in this paper.

**Keywords:** Deep Learning, Pose Detection, Multi-Object Tracking, Engagement, Interest Estimation

### 1. 서 론

신기술 및 신상품 개발, 신규 콘텐츠 제작 및 공연 등과 같은 아이터మ్은 인간의 삶과 밀접하게 연관된 요소들이며 사람의 관심에 종속된다. 사람의 관심을 이끌지 못하는 아이터మ్은 자연스럽게 도태되고 시장에서 퇴출된다. 이와 같이 사람의 관심은 시장의 경쟁력 확보를 위한 가장 중요한 요소이다.

사람이 관심을 가지는 정도를 의미하는 관심도는 사람이

취하는 행위에 따라 판단할 수 있다. 최근 사람의 행위 분석과 관심도 추정은 다양한 분야에서 연구되고 있는 주제이며 기술 및 상품의 연구개발 또는 창작 작품의 개발 방향에 많은 영향을 끼친다. 사람의 관심도는 신기술 및 신제품 개발의 기본이 되는 요소이며 관심을 유발하지 못하는 기술이나 제품은 성능이나 품질이 뛰어나더라도 소비자의 요구가 없어 경쟁력이 떨어지며, 이로 인해 추가적인 연구 개발을 위한 투자가 어려워 사장되기 쉽다. 이는 전시관이나 박물관 등 소비자가 직접적으로 접근할 수 있는 환경에 동일하게 적용할 수 있다. 소비자가 많은 관심을 가지는 대상, 관심을 가지지 않는 대상을 구분해서 분류할 수 있다는 것은 전시 대상에 대한 개선, 전시 환경에 대한 개선, 전시 환경의 배치 방법 등 다양한 요소를 보완할 수 있는 가능성을 확보할 수 있다.

사람의 행위를 결정하는 요소는 정신적, 사회적, 환경적, 문화적 배경 등과 연관된다. 사람의 행위 분석은 개인의 행위 패턴, 선호도 및 의사 결정 과정 등을 파악하는 것을 목표로 한다. 또한 유아부터 성인까지 다양한 연령대에서 나타내는 행위를 관찰하고 분석할 수 있다. 이러한 과정을 통해 개인의 니즈를 보다 정확하게 이해할 수 있고 그에 맞는 서비스나 상품을 제공할 수 있게 된다. 따라서 관심도 추정을 통해 특정 주제나 상품에 관심 정도를 예측할 수 있으며 관심도에 따라 마케팅 또는 보완 및 개선 작업 등 후속 조치를 취할 수 있다<sup>1,2</sup>.

Received : Jun. 9. 2023; Revised : Jul. 6. 2023; Accepted : Jul. 22. 2023

※ This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-00842, Development of Cloud Robot Intelligence for Continual Adaptation to User Reactions in Real Service Environments)

1. Researcher, Human-Robot Interaction R&D Center, KIRO (Korea Institute of Robotics & Technology Convergence), Pohang, Department of Future Automotive and IT Convergence, Kyungpook National University, Daegu, Korea (banily07@kiro.re.kr)
2. Senior Researcher, Human-Robot Interaction R&D Center, KIRO (Korea Institute of Robotics & Technology Convergence), Pohang, Korea (myearch@kiro.re.kr)
3. Principal Researcher, Human-Robot Interaction R&D Center, KIRO (Korea Institute of Robotics & Technology Convergence), Pohang, Korea (ejjung@kiro.re.kr)

† Head (Principal Researcher), Corresponding author: Human-Robot Interaction R&D Center, KIRO (Korea Institute of Robotics & Technology Convergence), Pohang, Korea (mingyukim@kiro.re.kr)

일반적으로 소비자의 행위를 분석하고 관심도를 추정하기 위해서는 판단하는 관찰자가 직접 행위를 관찰하고 평가하여 결과를 도출해야 하는데, 관찰자가 직접 육안으로 다른 소비자의 반응을 대략적 또는 주관적으로 판단할 수 있지만 많은 인력과 금전적, 시간적 소모가 발생하며 이는 매우 비효율적이다. 또한 소비자가 보이는 관심의 표출 정도는 관찰자의 입장에서 대략적인 판단만 가능할 뿐 정량적인 평가를 내리기 어렵다는 문제가 있다. 그 뿐만 아니라 소비자의 관심도 판단을 위한 설문 조사도 많이 활용하는데, 소비자의 심리적 상황에 따라 평가기준이 편향되는 경우가 많아 정량적인 평가의 신뢰도가 떨어질 수 있다.

본 논문은 인간의 행동분석학적 특징을 기반으로 한 관심도 추정에 관한 연구로써, 전시관에서 전시 대상인 안내 로봇에 관람객이 보이는 다양한 행위를 기반으로 딥 러닝 모델을 활용한 관람객의 행위 인식 및 분석을 통한 관심도를 추정하는 방법을 제안한다. 인공지능의 기술 발전은 기존의 통계적인 방법에서 벗어나 대량의 데이터를 빠르게 처리하고 정교한 모델을 구축할 수 있으며, 이러한 특징을 바탕으로 사람의 행위를 분석하고 관심도를 추정할 수 있다. 사람의 행위를 분석하기 위해 딥 러닝 모델을 활용한 객체 인식(Object Recognition) 기술을 기반으로 행위 인식(Behavior Recognition)을 수행하고, 그 결과를 바탕으로 객체 추적(Object Tracking) 기술 기반의 연속적 행위 인식 히스토리 생성과 관심도를 추정한다. 본 논문의 2장은 관련 연구 동향, 3장은 본 논문에서 제안하는 연구에 대한 이론적인 설명, 4장에서는 실험 결과에 대한 내용으로 끝을 맺는다.

## 2. 연구 동향

딥 러닝 기반의 인간 행위 인식을 통한 관심도 추정은 다양한 분야에서 관심을 가지는 추세이다. 해당 연구 주제와 관련된 기술로는 다중 데이터 학습, 실시간 행위 인식, 시간적 모델링 등이 해당된다. 본 논문에서 제안하는 연구와 유사성이 높은 주제는 찾기 힘들지만 일부 유사 연구가 다수 진행되고 있다. 딥 러닝 기반의 인간의 행위 예측을 위한 학습 모델을 활용하는 방법은 해당 분야에서 발생할 수 있는 행위를 분류하고 이에 해당하는 데이터를 생성하여 학습 모델을 구축할 수 있다. 딥 러닝 모델 특성 상 초기 학습 모델 생성 과정에서는 자원 소모가 발생할 수 있지만 그 이후에는 빠른 처리 속도와 높은 인식률의 성능을 확보할 수 있다. Eun<sup>[1]</sup>은 실시간 행위 인식을 위해 GRU (Gated Recurrent Unit)를 확장한 IDU (Information Discrimination Unit)를 제안했다. 현재 행위에 대한 정보 획득을 위해 현재 진행중인 행위와 과거 정보 간의 관계를 모델링하여 유효한 행위에 대한 결과를 도출하는 방식이다. 시계열

데이터 방식을 활용하여 실시간으로 인간 행위를 인식할 수 있다는 특징을 가진다. Chen<sup>[4]</sup>은 실시간 행위 예측을 위해 GateHUB (Gated History Unit with Background Supression)을 제안했다. GateHUB는 GHU (Gated History Unit), FaH (Future Augmented History), Background Suppression loss를 이용하여 입력되는 영상 데이터의 유효 부분을 탐지하고 과거의 데이터를 연관지어 현재 행위를 예측하는 방식으로 동작하는 것이 특징이다. 최근 많은 관심을 가지는 딥 러닝 기반의 인간 행위 예측 기술은 시계열 데이터를 활용하여 과거 데이터 간의 유사성을 기반으로 행위를 예측한다는 것이 특징이다. Mazzia<sup>[5]</sup>는 인간 행위 인식(Human Action Recognition)을 위한 AcT (Action Transformer) 모델을 제안한다. 합성곱이나 순환 레이어가 없는 Transformer 인코더를 기반으로 한 방법을 제시했다는 것이 특징이다. MPOSE2021 2D Pose 기반의 데이터셋을 사용하여 행위 인식을 위한 알고리즘의 검증에 수행했다. Vrskova<sup>[6]</sup>는 연속되는 영상 데이터에서 인간의 행위 인식을 위한 3D Convolutional Neural Network (CNN)에 대한 연구를 수행했다. 해당 연구 결과 움직이는 행위, 서있는 행위를 포함한 다양한 행위를 인식할 수 있다는 특징을 가진다. 해당 연구 또한 관심도 추정에 응용할 수 있는 인간의 행위 인식에 관련된 연구이다. Salam<sup>[7]</sup>은 인간과 휴머노이드 로봇의 상호 작용으로 관심도를 추정할 수 있는 연구를 수행했다. 해당 연구에서는 인간의 행위를 추출하여 성향을 분류하고 행위를 기반으로 관심의 정도를 결정할 수 있다. 휴머노이드 로봇은 인간의 성향을 바탕으로 인간을 대하는 성향을 결정할 수 있으며, 이를 기반으로 관심을 유발하여 HRI 시나리오의 적극적인 참여를 유도할 수 있다는 특징을 가진다. 해당 연구는 앞서 언급한 연구들과 비교하여 인간의 행위를 인식하지는 않지만 관심도 추정 방법을 응용할 수 있다. 사람의 행위 인식을 기반한 관심도 추정과 관련된 연구는 현재까지 찾기 힘들지만 그와 관련된 연구는 다양하게 진행되는 추세이다<sup>[8,9]</sup>.

## 3. 관심도 추정 알고리즘

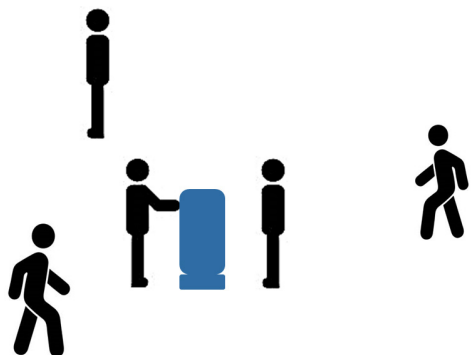
관심도 추정 알고리즘은 사람이 표출할 수 있는 다양한 행위를 바탕으로 관심에 대한 정도를 추정하는 것이다. 사람이 표출할 수 있는 다양한 행위는 몸짓, 손짓, 표정, 억양 등 다양하게 존재한다. 사람의 표정이나 억양으로 관심도를 추정하기에는 관찰하는 사람이 직접 따라다니며 기록하지 않는 이상 다양한 이유와 환경적인 요소들로 인해 부적합하다. 따라서 본 연구에서는 영상 데이터를 이용하여 사람의 몸짓, 손짓에 대한 행위를 인식하고 관심도를 추정하고자 한다. 이를 위한 과정은 총 3가지 단계로 분류할 수 있다. 첫 번째는 단일 프레임으로 대상의 행위를 인식하기 위해 딥 러닝 기반의 학습 모

델을 이용하여 사람의 행위를 인식한다. 두 번째는 인식한 대상을 지속적으로 추적하고 행위를 기록하여 시계열 데이터를 확보한다. 세 번째는 시계열 데이터를 분석하여 대상의 관심도를 추정하고자 한다.

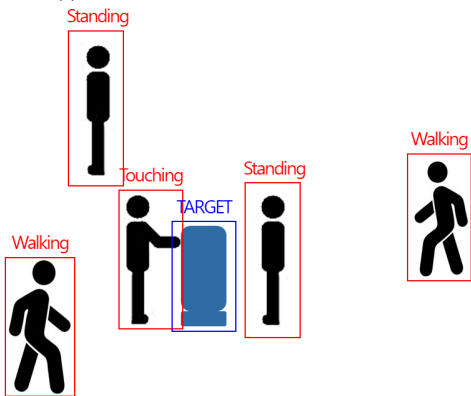
### 3.1 딥 러닝 기반 인간 행위 인식

[Fig. 1]은 전시관 내 관람객의 여러가지 행위에 대한 예시를 표현하고 있다. 전시관에서 관람객이 할 수 있는 행위는 서있기 또는 쳐다보기(Standing), 걸어가기 또는 접근하기(Walking), 접촉하기 또는 가리키기(Touching)으로 크게 3개의 레이블로 분류할 수 있다. 서있기 또는 쳐다보기(Standing)는 전시품을 쳐다보거나 찾는 행위에 포함되며 걸어가기 또는 접근하기(Walking)는 전시품으로 접근하거나 전시품을 회피하여 이동하는 행위, 접촉하기 또는 가리키기(Touching)는 전시품을 만지거나 손가락으로 가리키면서 관심을 표출하는 행위로 간주한다.

딥 러닝 모델을 활용한 객체 인식(Object Recognition) 기술은 기존에 학습된 다양한 정보를 바탕으로 다양한 소스 내에서 학습한 대상을 찾아내는 기술이다. 최근 연구 동향에 의하



(a) Visitor activities in the exhibition hall



(b) Classification results of visitor behavior in the exhibition hall

[Fig. 1] Classification of visitor behavior in exhibition halls (example)

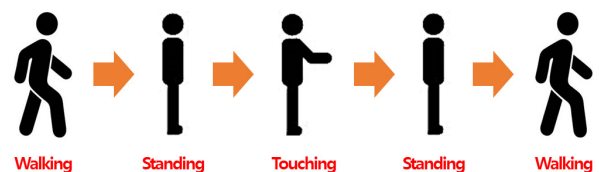
면 3차원 RGBD데이터 기반의 스켈레톤 모델을 활용하여 사람의 자세를 인식하는 경우가 많다. 해당 연구 방법은 자세 인식이 높다는 장점을 가지지만 센서의 기술적 한계로 거리에 대한 한계점과 다수의 대상을 동시에 인식하기 어렵다는 문제점이 존재한다. 하지만 본 논문에서는 2차원 RGB 영상 데이터만으로 사람의 행위를 인식하므로 이러한 문제점을 해결할 수 있다.

관람객의 행위 인식(Behavior Recognition)을 위해 딥 러닝 모델을 활용한 객체 인식 기법과 동일한 방식으로 YOLOv5 모델을 이용하여 학습 모델을 생성한다<sup>10)</sup>. YOLO는 타 딥 러닝 모델 대비 데이터 레이블링 작업이 비교적 수월하고 간단한 구조, 빠른 처리속도의 특징과 작은 객체의 인식 정확도가 높아 멀리 있는 대상에 대한 인식률도 확보할 수 있다는 장점을 가진다.

### 3.2 대상 추적 및 행위 히스토리 생성

관람객의 행위 인식 결과는 단일 프레임 시점을 기준으로 획득할 수 있다. 이를 기반으로 관람객이 화면 상에 등장한 시점부터 퇴장한 시점까지의 지속 시간과 행위 인식 결과 및 이동 방향을 확보하여 전체 히스토리에 대한 분석으로 관심도를 추정하고자 한다.

[Fig. 2]는 관심도 추정을 위해 관람객이 화면상에 등장 후 퇴장까지의 히스토리에 대한 예시를 나타낸다. [Fig. 2(a)]는 관람객이 관심을 보이는 경우에 대한 예시이며, 화면에 등장하여 이동(Walking) 후 전시품에 접근하여 관찰(Standing)한다. 전시품에 보다 더 많은 호기심이 발생할 경우 접촉(Touching)하여 조작해보거나 직접적으로 경험할 수 있다. 그 이후로도 계속해서 관찰(Standing) 후 다른 전시품을 보거나 장소 변경을 위해 이동(Walking)하여 화면에서 퇴장한다. [Fig. 2(b)]는 관람객이 관심을 보이지 않는 경우에 대한 예시이며, 관람을 위해



(a) When the visitor shows interest



(b) When the visitor shows no interest

[Fig. 2] History of visitor behavior recognition results (example)

이동(Walking)하여 화면에 등장하게 된다. 이동 과정에서 전시품을 보기위해 잠시 멈추지만(Standing) 관심이 없어 다시 이동(Walking)하여 화면에서 퇴장한다. 관심이 없는 경우 멈춰서는(Standing) 과정이 생략될 수도 있다. 이와 같이 관심의 차이에 따라 관람객의 행위가 달라질 수 있다는 것을 알 수 있으며, 히스토리의 분석으로 관심도를 추정할 수 있다.

다수의 관람객이 존재하는 전시관에서 개개인의 행위에 대한 히스토리를 획득하기 위해서는 개별 대상에 대한 지속적인 추적이 필요하다. 이를 위해 다중 객체 추적 알고리즘인 FairMOT (On the Fairness of Detection and Re-Identification in Multiple Object Tracking) 모델을 활용하여 다수의 관람객을 추적하고자 한다<sup>[11]</sup>. FairMOT는 영상 데이터에서 다중 객체 추적이 가능하고 높은 인식률, 높은 추적 지속률, 빠른 처리속도 등의 특징을 가진다. 이를 기반으로 다수의 관람객을 인식하고 각각에 ID를 부여하여 등장부터 퇴장까지 모든 상황을 기록할 수 있도록 추적할 수 있다.

### 3.3 히스토리 분석을 통한 관심도 추정

관람객의 연속적인 행위에 대한 분류 기준을 제시하고 이를 기반으로 관심도를 추정하고자 한다. [Table 1]은 전시품과 관람객의 물리적 거리 조건에 대한 행위 분류 기준, [Table 2]는 상호작용 상황에 대한 행위 분류 기준을 제시한다. 관람객이 전시품으로 접근하는 경우(Approach), 접근하지만 통과하거나 접근하지 않고 다른 방향으로 이동하여 사라지는 경우(Pass)와 같이 2개의 행위 분류 기준을 제시한다. 또한 접근(Approach) 이후 특정 범위 내에 머무르는 조건이 성립될 경우 상호작용 시도 유무에 따라서 접촉(Touch)하거나 아무런 행위를 하지 않는(None) 경우로 분류할 수 있다.

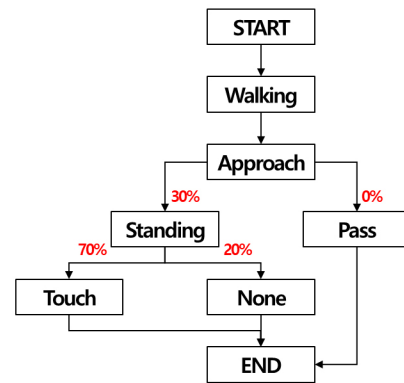
[Table 1], [Table 2]는 개별 행위에 대한 분류 기준을 의미하지만 히스토리 상에서 발생하는 일련의 연결 관계를 알기 힘들다. 관람객의 관심도를 추정하기 위해서는 개별 행위에 대한 연결 관계를 분석하여 포괄적으로 의미하는 바를 분석해야 한다.

[Table 1] Criteria for behavior classification based on physical distance

Physical distance	
Approach	Approach in the direction of the target
Pass	Pass through an object or move in a different direction

[Table 2] Criteria for behavior classification based on interaction

Interaction attempt	
Touch	Touching or Pointing to an object
None	No interaction with target



[Fig. 3] Flowchart of visitor behavior and selection of interest score

[Fig. 3]은 [Fig. 2]에서 정의한 관심 유무에 따른 관람객의 행위에 대한 히스토리를 순서대로 정의하고 그에 따른 관심도 점수 부과 방법을 정의했다. [Fig. 3]의 START와 END는 관람객의 등장과 퇴장을 의미한다. 등장 후 전시품에 접근(Approach)하여 근처에 머무른다면(Standing) 약간의 관심이 있다고 판단할 수 있어 최대 30%의 관심도를 부과한다. 이후 두 개의 경우로 분류할 수 있는데, 첫 번째는 전시품을 주시하면서 손으로 가리키거나 접촉하는 경우(Touching) 단순히 머무르는 것보다 더 많은 관심이 있다고 판단할 수 있다. 이러한 상황이 발생하는 경우 즉시 50%의 관심도를 부과하고 최대 20%의 관심도를 추가로 부과한다. 두 번째는 특정 행위를 취하지 않는 경우(None)를 의미하며 최대 20%의 관심도를 추가로 부과한다. 두 개의 경우에 따라서 최대 50~100%의 관심도를 부과할 수 있다. 그 외에 등장 후 전시품에 접근하지 않거나 접근 후 머무르지 않고 퇴장하는 경우(Pass)가 있으며, 이러한 경우 0%의 관심도를 부과한다. 현재까지 정량화된 관심도 측정 방법이나 관련된 연구 자료가 없어 본 연구에서는 저자가 직접 영상을 보고 판단하여 내린 결론으로 관심도 점수 부과 가중치를 설정했다. 차후 연구를 통해 실제 관람객의 평가 데이터를 수집하고 통계 결과를 도출하여 정량화된 관심도 측정 기준을 세우고 이를 바탕으로 개선된 관심도 추정 방법을 개발할 계획이다.

$$w_t = \begin{cases} 1 & t > 30 \\ 0.5 & 5 \geq t \geq 30 \\ 0 & t < 5 \end{cases} \quad (1)$$

$$I_t = \begin{cases} 50 & t > 30 \\ 1.6 \times t & 5 \geq t \geq 30 \\ 0 & t < 5 \end{cases} \quad (2)$$

$$I = \begin{cases} 100 & t > 30, \text{ if touch is True} \\ 50 & t > 30, \text{ if touch is False} \\ (50 + I_t) \times w_t & 5 \geq t \geq 30, \text{ if touch is True} \\ I_t \times w_t & 5 \geq t \geq 30, \text{ if touch is False} \\ 0 & t < 5 \end{cases} \quad (3)$$



관람객이 머무는 시간에 따라 관심도의 차별성을 두기 위해 식 (1)의 시간에 의한 가중치  $w_t$ 와 식 (2)의 시간에 의한 관심도 누적치  $I_t$ 를 정의한다. 관람객이 최소 5초 이상 머물러야 관심의 시작 시점, 최대 30초를 초과하여 머물러야 최대 관심 상태라고 판단하고 식 (1)과 (2)를 정의한다. [Fig. 3]의 경우에 따른 관심도 추정 방법을 기반으로 관심도  $I$ 를 도출하는 식 (3)을 정의한다. 식 (3)은 접근(Approach) 이후 서있는(Standing) 상태에서 손으로 가리키거나 접촉하는 경우(Touching)와 그렇지 않은 경우(None)에 대한 관심도 추정 방법이다. 총 5개의 경우가 존재하며, 첫번째 경우는 관람객이 특정 행위(Touching)를 취하면서 30초를 초과하여 머물렀을 때이며 100%의 관심도를 부과한다. 두번째 경우는 관람객이 특정 행위(Touching)를 취하지 않았지만 30초를 초과하여 머물렀을 때이며 50%의 관심도를 부과한다. 세번째 경우와 네번째 경우는 관람객이 5초 이상, 30초 이하의 시간 동안 머물렀으며 특정 행위(Touching)를 취한 경우와 그렇지 않았을 때(None) 상황이다. 마지막으로 다섯번째 경우는 전혀 머무르지 않은 상황이며 0%의 관심도를 부과한다. 식 (3)을 기반으로 안내 로봇을 향해 손짓이나 접촉하지 않은 경우(None) 최대 50%, 손짓하거나 접촉한 경우(Touching) 최대 100%의 관심도를 부과한다.

#### 4. 실험 및 결과

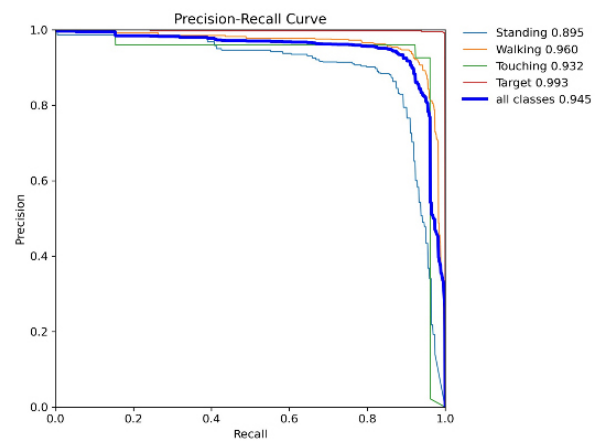
관람객의 행위 인식 모델은 자체적으로 획득한 데이터를 기반으로 학습 모델을 생성하였다. [Fig. 4]는 자체적으로 획득한 영상 데이터에서 추출한 학습용 이미지이다. 다수의 관람객이 존재하는 학습용 이미지는 총 8000개를 사용하였고 YOLOv5 모델을 사용하여 개인화된 학습 모델을 생성하였다. 학습과 검증에 사용되는 영상은 하나의 고정된 CCTV 카메라를 사용했고, 각 용도에 맞게 중복 사용하지 않았다. [Fig. 5]는 학습된 모델의 P-R (Precision-Recall) Curve와 각 레이블 별 AP (Average Precision) 수치를 보여준다. 총 4개의 레이블로 구성되며, 사람의 행위를 인식하는 3개 레이블과 안내 로봇을 인식하는 1



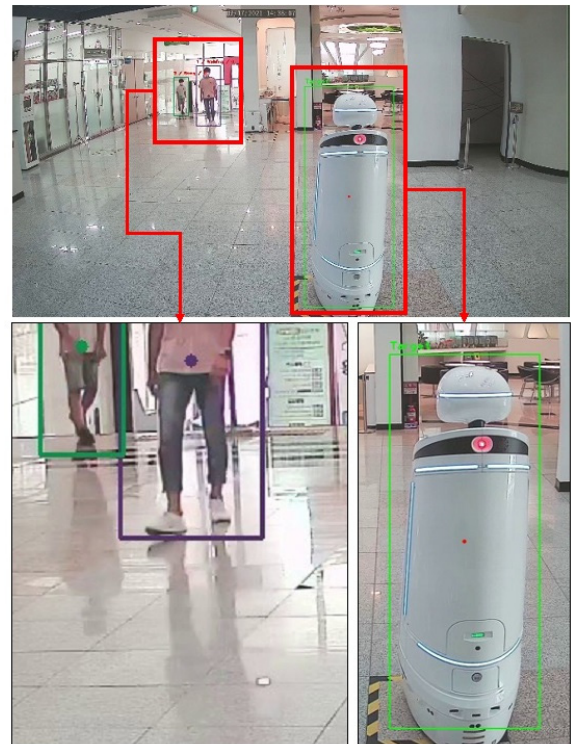
[Fig. 4] Self-acquired training images

개 레이블의 학습 성능에 대한 결과를 나타내며, 안내 로봇 인식 레이블을 제외한 행위 인식 레이블에 대한 mAP (mean Average Precision)는 0.929의 수치를 나타낸다.

[Fig. 6]은 관심도 추정 알고리즘을 이용하여 사전에 취득한 검증용 영상을 처리한 결과를 의미한다. [Fig. 6]의 우측 하단에서 초록색 박스로 체크된 안내 로봇(Target)은 이동이 가능한 안내 로봇이며 특정 지점을 지속적으로 이동하면서 관람객의 관심을 끌 수 있는 대상이다. [Fig. 6]의 좌측 하단은 관람객을 인식한 결과를 보여준다. 총 3개의 정보를 출력하게 되며, 첫 번째 정보는 각 관람객의 ID를 의미하며 두 번째 정보는



[Fig. 5] The Precision-Recall Curve of the Learning Model



[Fig. 6] Results from the Interest Estimation Algorithm

[Table 3] Experimental scenario and anticipated the level of interest

	Scenario	Expected interest
1	Visitors passing in different directions regardless of the guide robot	up to 0%
2	Visitors passing by while approaching the guide robot	up to 0%
3	Visitors pass by as they approach the guide robot and watch	up to 50%
4	Visitors who approach the guide robot, touch and gesture, and then pass by	up to 100%



[Fig. 7] Test for measuring visitor's interest 1-1

관람객의 행위 인식 결과를 의미한다. 마지막 세 번째 정보는 관람객이 대상에게 관심을 보일 때 발생하는 관심도의 점수를 의미한다.

관심도 추정 알고리즘의 검증에 위해 [Table 3]과 같이 총 4개의 시나리오를 세우고 실험을 수행하고자 한다. 첫 번째는 안내 로봇과 전혀 다른 방향으로 지나가는 경우, 두 번째는 안내 로봇에 접근하다 지나쳐가는 경우, 세 번째는 안내 로봇에 접근하여 지켜보다 지나가는 경우이며, 마지막은 안내 로봇에 접근하여 접촉하거나 손짓 후 지나가는 경우이다. 각각의 시나리오에 따라서 추정할 수 있는 관심도를 예상할 수 있으며, 실험을 통해 결과를 도출하고 비교하고자 한다. [Table 3]의 4개 실험에 대한 결과는 아래 순차적으로 그림으로 표기했다. 여기서 카메라 전체 영상이 아닌 독자들의 인식 결과에 대한 시인성을 확보하기 위해 관람객만 확대하여 표시했다. 각 실험 별로 안내 로봇과 관람객의 위치와 거리에 대해서 추가로 표기했다.

[Fig. 7]은 실험 1에 대한 첫 번째 상황으로 안내 로봇과 관계없이 다른 방향으로 지나가는 관람객에 대한 관심도를 추정하는 실험이다. 독자의 시인성을 확보하기 위해 전체 화면에서 추적하고자 하는 관람객만을 확대하여 표시하였고, 각 그림의 우측 하단에 표기된 숫자는 진행 순서를 의미한다. 해당 실험은 관람객이 안내 로봇에 전혀 관심이 없다는 것으로 간

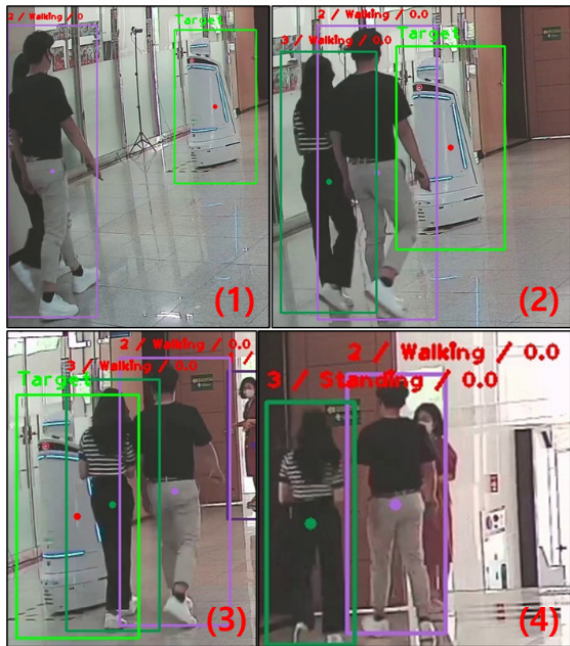


[Fig. 8] Test for measuring visitor's interest 1-2

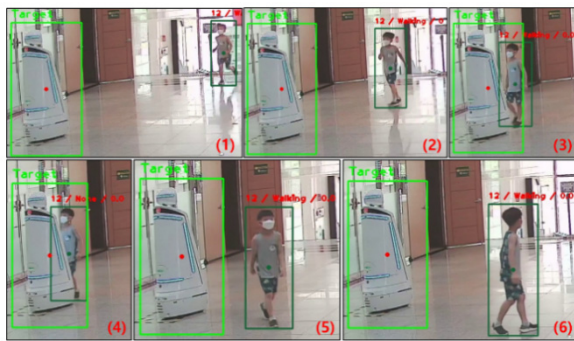
주하고 0%의 관심도를 예상한다. 안내 로봇은 관람객의 위치를 기준으로 우측에 위치하고 관람객과 로봇 간의 거리는 약 5미터 이상 떨어져 있다. ID 1의 관람객은 등장 후 이동(Walking) 상태를 유지하면서 안내 로봇에 접근하지 않고 지나가는 상황이다. 또한 관람객은 안내 로봇을 쳐다보지 않고 지나간다. 최종 결과 관람객이 화면에 사라지는 상황까지 0%의 관심도를 나타낸다. [Fig. 8]은 실험 1에 대한 두 번째 상황으로 [Fig. 7]과 동일한 경우이며, 안내 로봇은 관람객의 위치를 기준으로 우측에 위치하고 거리는 약 5미터 이상 떨어져 있다. 실험 1-2의 경우에 대한 예상 관심도는 0%를 가리킨다. ID 14, ID 15의 관람객은 등장 후 이동(Walking) 상태를 유지하며 안내 로봇을 쳐다보지 않고 지나간다. 최종 결과 퇴장하는 상황까지 0%의 관심도를 유지하는 것을 알 수 있다.

[Fig. 9]는 실험 2에 대한 첫 번째 상황으로 안내 로봇에 접근하다가 멀어지는 경우에 대한 실험 결과를 나타낸다. 실험 2는 관람객이 안내 로봇의 방향을 향해 쳐다보면서 접근하지만 전혀 관심이 없다는 것으로 간주하고 0%의 관심도를 예상한다. ID 2, ID 3의 관람객이 좌측에 위치한 안내 로봇을 지나 문으로 이동하는 상황이다. 안내 로봇을 향해 접근하면서 잘못된 판단으로 관람객의 관심도가 증가할 수 있지만 안내 로봇 근처에서 일정 시간 이상 머무르지 않아 결과적으로 0%의 관심도를 유지하면서 퇴장하는 것을 알 수 있다. [Fig. 10]은 실험 2에 대한 두 번째 상황으로 [Fig. 9]와 동일하게 관람객의 관심도는 0%로 예상한다. ID 12의 관람객은 안내 로봇에 접근했다





[Fig. 9] Test for measuring visitor's interest 2-1



[Fig. 10] Test for measuring visitor's interest 2-2

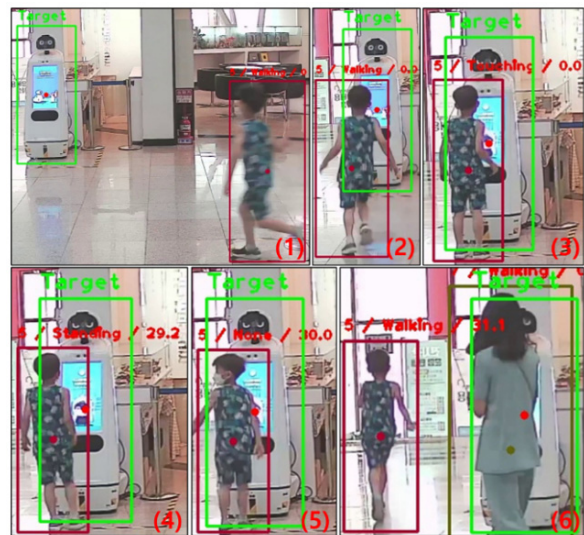
가 크게 관심이 없어 즉시 다른 곳으로 이동한다. 관람객은 안내 로봇 앞에 서서 약 4초간 머무르다 즉시 이동하였으므로 결과적으로 0%의 관심도를 유지한다.

[Fig. 11]은 실험 3과 실험 4의 경우가 동시에 발생한 상황으로 안내 로봇에 접근하여 지켜보거나 접촉하는 경우에 대한 실험을 나타낸다. 노란색 원피스를 착용한 ID 32의 관람객은 안내 로봇에 접근 후 특별한 행위 없이 지켜보다가 퇴장하는 상황이며 흰색 반팔 티셔츠를 착용한 ID 29의 어린이 관람객은 안내 로봇에 접근 후 손으로 안내 로봇을 만져보고 머무르다 퇴장하는 상황이다. ID 32의 관람객은 안내 로봇에 접근하여 머무른 시간이 약 11초, 19.2%의 관심도를 나타내며, ID 29의 관람객은 안내 로봇에 접근하여 머무른 시간이 약 16초, 75.2%의 관심도를 나타낸다.

[Fig. 12]는 네 번째 경우인 안내 로봇에 접근하여 접촉 후 퇴장하는 경우에 대한 실험이다. ID 5의 관람객이 안내 로봇을 향



[Fig. 11] Test for measuring visitor's interest 3



[Fig. 12] Test for measuring visitor's interest 4

해 접근한 뒤 쳐다보다가 스크린을 터치한 뒤 다시 서서 쳐다보는 상황이다. 관람객은 안내 로봇에 접근하여 머무른 시간이 약 7초이며 0.5의 가중치가 적용되어 31.1%의 관심도를 나타낸다.

[Table 4]는 앞에서 실험한 4개 경우의 실험 결과를 정리하였다. 실험 1, 2는 관람객의 행위에서 안내 로봇에 관심이 없다는 것을 명확하게 판단할 수 있다. 이러한 근거를 바탕으로 0%의 관심도 추정 결과를 도출하였다. 실험 3, 4는 1, 2와 달리 안내 로봇에 접근하여 일정 시간 이상 머물렀으며 접촉과 같은 행위를 통해 관심이 있다는 판단에 대한 근거를 제공하였다. 실험 3에서 ID 32의 관람객과 ID 29의 관람객이 머무른 시간은 비슷하지만 ID 32의 관람객은 접촉이나 손짓과 같은 특정 행위를 취하지 않아 50%의 관심도가 적용되지 않았으며 최종 19.2%의 관심도가 도출되었다. 반면에 ID 29의 관람객은 접촉

[Table 4] Experimental Result of Estimating the Level of Interest

Experiment sequence	The result of interest estimation	Time (second)	Weight	Expected interest
1-1	ID 1 : 0%	-	0	0%
1-2	ID 14 : 0%	-	0	0%
	ID 15 : 0%	-	0	
2-1	ID 2 : 0%	-	0	0%
	ID 3 : 0%	-	0	
2-2	ID 12 : 0%	4	0	0%
3	ID 29 : 75.2%	16	1	~100%
	ID 32 : 19.2%	11	1	~50%
4	ID 5 : 31.1%	7	0.5	~100%

행위를 취하여 50%의 관심도가 적용되었으며 최종 75.2%의 관심도 추정 결과가 도출된 것을 알 수 있다. 또한 두 관람객 모두 10초 이상 머물러 시간에 대한 가중치는 1로 적용되었다. 실험 4의 ID 5 관람객은 접촉 행위를 취하여 50%의 관심도가 적용되었으나 7초간 머물러 시간에 대한 가중치가 0.5로 적용되어 최종 31.1%의 관심도 추정 결과가 도출되었다.

## 5. 결 론

본 논문은 딥 러닝 모델을 활용하여 전시관에서 관람객의 행위를 분석하고 추적하여 관심도를 추정하는 연구에 대한 내용을 제안한다. 딥 러닝 기반의 인식 모델인 YOLOv5를 활용하여 다수의 관람객에 대한 행위를 인식하였고 FairMOT를 활용하여 다수의 관람객을 추적, 행위에 대한 히스토리를 생성하였고 이를 분석하여 관심도를 추정할 수 있다. 사용자의 행위 분석을 통해 관심도 수치에 따른 정량적 평가 기준을 결정할 경우 보다 객관적인 시점으로 판단할 수 있는 조건을 만들 수 있다. 향후 연구 계획으로는 관람객의 행위 인식의 범주를 다수로 늘리고 관심도를 추정하는 방향에 대한 세부적인 조건을 수립하여 보다 세분화된 관심도를 추정할 수 있는 연구를 수행하고자 한다. 또한 본 논문에서 제안하는 알고리즘의 인식과 추적 성능에 대한 고도화를 진행할 예정이다. 그 외에도 관람객의 정성적 평가 결과에 기반하여 관심도를 추정하는 방법에 대한 검증은 수행할 예정이다.

## References

[1] S. Yim and M. Kim, "A Study on Attitudes toward Man's Appearance Management and Cosmetics Purchasing Behavior," *Journal of the Korea Fashion & Costume Design Association*, vol. 16, no. 4, pp. 79-98, Dec., 2014, [Online], [https://www.kci](https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artid=ART001943726)

[go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artid=ART001943726](https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artid=ART001943726).

[2] J. W. Jung and S. Y. Cho, "The Effects of High School Students' Environmental Concerns and Subjective Norms on Purchase Intentions of Eco-Friendly Products : Mediating Effect of Attitude toward Eco-Friendly Products and Services," *Korean Journal of Environmental Education*, vol. 32, no. 4, pp. 475-487, Dec., 2019, [Online], <https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE09298315>.

[3] H. Eun, J. Moon, J. Park, C. Jung, and C. Kim, "Learning to Discriminate Information for Online Action Detection," *Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, pp. 809-818, 2020, DOI: 10.1109/CVPR42600.2020.00089.

[4] J. Chen, G. Mittal, Y. Yu, Y. Kong, and M. Chen, "GateHUB: Gated History Unit With Background Suppression for Online Action Detection," *Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, USA, pp. 19925-19934, 2022, DOI: 10.1109/CVPR52688.2022.01930.

[5] V. Mazzia, S. Angarano, F. Salvetti, F. Angelini, and M. Chiaberge, "Action Transformer: A Self-Attention Model for Short-Time Pose-Based Human Action Recognition," *Pattern Recognition*, vol. 124, Apr., 2022, DOI: 10.1016/j.patcog.2021.108487.

[6] R. Vrskova, R. Hudec, P. Kamencay, and P. Sykora, "Human Activity Classification Using the 3DCNN Architecture," *Applied Science*, vol. 12, no. 2, Jan., 2022, DOI: 10.3390/app12020931.

[7] H. Salam, O. Celiktutan, I. Hupont, H. Gunes, and M. Chetouani, "Fully Automatic Analysis of Engagement and Its Relationship to Personality in Human-Robot Interactions," *IEEE Access*, vol. 5, pp. 705-721, Sept., 2016, DOI: 10.1109/ACCESS.2016.2614525.

[8] B. Fernando, E. Gavves, M. J. Oramas M., A. Ghodrati, and T. Tuytelaars, "Modeling Video Evolution for Action Recognition," *Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA, pp. 5378-5387, 2015, DOI: 10.1109/CVPR.2015.7299176.

[9] Z. Gao, Y. Zhang, H. Zhang, Y. B. Xue, and G. P. Xu, "Multi-dimensional human action recognition model based on image set and group sparsity," *Neurocomputing*, vol. 215, pp. 138-149, Nov., 2016, DOI: 10.1016/j.neucom.2016.01.113.

[10] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael, TaoXie, J. Fang, imyhxy, Lorna, Z. Yifu, C. Wong, A. V. D. Montes, Z. Wang, C. Fati, J. Nadar, Laughing, UnglvKitDe, V. Sonck, tkianai, yxNONG, P. Skalski, A. Hogan, D. Nair, M. Strobel, and M. Jain, ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation, [Online], <https://zenodo.org/record/7347926#.ZDziMnZBYuc>, Accessed: Jan. 06, 2023.

[11] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking," *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3069-3087, Sept., 2021, DOI: 10.1007/s11263-021-01513-4.





**이 예 준**

2015 동아대학교 전자공학과(학사)  
2018 경북대학교 전자공학부(석사)  
2019~현재 경북대학교 미래자동차·  
IT융합학과(박사과정 수료)  
2018~현재 한국로봇융합연구원  
HRI 연구센터 주임연구원

관심분야: 딥러닝, 영상처리



**정 의 정**

2006 한양대학교 전자컴퓨터공학부(학사)  
2008 한양대학교 전자컴퓨터공학부(석사)  
2013 한양대학교 전자컴퓨터공학부(박사)  
2014~2015 Carnegie Mellon University,  
The Robotics Institute, Postdoctoral  
Fellow  
2018~현재 한국로봇융합연구원  
HRI 연구센터 책임연구원

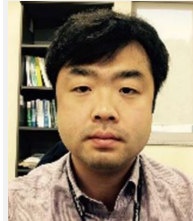
관심분야: 모바일 로봇, 자율주행, 로봇 기구학, 객체 인식



**김 주 현**

2012 동명대학교 로봇시스템공학과(학사)  
2014 동명대학교 로봇시스템공학과(석사)  
2015~현재 한국로봇융합연구원  
HRI 연구센터 선임연구원

관심분야: 로봇 제어, 시뮬레이션



**김 민 규**

2012 University of Tsukuba(박사)  
2013 성균관대학교 인터랙션사이언스연구소  
(박사후연구원)  
2014 아인트호벤공과대학 산업 디자인  
(박사후연구원)  
2015~현재 한국로봇융합연구원  
HRI 연구센터 센터장(책임연구원)

관심분야: Human-Robot Interaction, Data Analytics, Artificial Intelligence