

# Human-Robot Interaction을 위한 모바일 환경에서의 인물 재식별화 및 추적 알고리즘

(A Person Re-Identification and Tracking Algorithm  
in Mobile Environment for Human-Robot Interaction)

한 현 영<sup>†</sup>, 박 재 현<sup>†</sup>, 최 두 원<sup>†</sup>, 김 준 광<sup>‡</sup>, 정 우 영<sup>\*</sup>

<sup>†,‡</sup>대구경북과학기술원

(Hyunyoung Han, Jaehyun Park, Doowon Choi, Junkwang Kim, Wooyoung Jung)

(<sup>†,‡</sup>DGIST)

**Abstract** : Recently, interest in personal robots has been increasing due to advances in computer vision technology and the robot industry. In this paper, we present a person re-identification and tracking algorithm in a robotic environment equipped with Android-based OS and vision devices. In addition, by applying YOLO v4 and image hash algorithm optimized for mobile environments, we improve the accuracy of object detection and tracking (about 3.52%), and the performance of real-time video analysis (tracking and person re-identification) (about 27.42 FPS). In particular, we add a filtering technique in image hash techniques to enable subtle brightness discrimination, increasing the accuracy (about 12.81%) of object re-identification. This is a key success factor that enables smooth Human-Robot Interaction (HRI).

**Keywords** : HRI, Mobile, Object Detection, Re-Identification, Object Tracking

## I. 서 론

최근 딥러닝을 이용한 객체 인식, 추적 관련 연구가 활발히 진행됨에 따라 이를 다양한 분야에 접목하려는 시도가 다수 등장하고 있다. 가령 CCTV에 위와 같은 기술을 접목함으로써 영상 감시의 효율성 향상 및 자동화 관련 연구나 [1], 자율주행 차량의 비전 센서에 컴퓨터 비전 기술을 활용함으로써 보행자, 주위 차량, 신호 등을 인식하여 운행 중 발생할 수 있는 여러 문제 상황을 해결하고자 하는 연구 등이 진행되고 있다 [2].

더불어 로봇 산업의 발전과 큰 범용성을 지닌 컴퓨터 비전 기술이 결합되고 있다. 이에 종래 산업 현장뿐만 아닌 개인의 삶에 로봇이 도래하는 것에 대한 관심도와 흥미가 증가하고 있다. 이러한 “개인

용 로봇”에 대한 사람들의 흥미와 연구자들의 개발로 인해 실제 사람들의 삶에 큰 영향을 미치고 있는 실정이다. 이를테면 시각장애인 안내견을 대체할 스마트 기기나 터치스크린을 통해 감정을 표현하고 마이크와 스피커를 이용해 가족들과 대화를 나누는 로봇 등 스마트홈 도우미를 지향점에 둔 로봇들도 개발, 출시되고 있다. 이러한 로봇에 탑재되는 컴퓨터 비전 기술은 HRI(Human-Robot Interaction)을 용이하게 하는 주된 요인이 된다. 본 논문에서는 안드로이드 기반 OS와 비전 기기가 탑재된 로봇 환경에서의 인물 재식별화 및 추적 알고리즘을 제시한다. 더불어 당 분야에서 중시되는 객체 인식 및 추적의 정확도와 실시간(Real-time) 인물 재식별화 성능 향상을 [3] 주안점으로 견지함으로써 원활한 HRI를 가능하게 하는 시스템을 제안하고자 한다.

\* Corresponding Author (wyjung@dgist.ac.kr)

H. Han, J. Park, D. Choi, J. Kim, W. Jung: DGIST

※ This work was supported by the DGIST Undergraduate Group Research Program(UGRP) grant.

## II. 본 론

### 1. 개발 및 작동 환경

본 연구에 활용된 개발 및 작동 환경은 표 1과 같다.

표 1. 개발 및 작동 환경

Table 1. Development and Operating Environment

기종	삼성 갤럭시 S9+ (SM-G965x)
프로세서	엑시노스 9810
CPU	Octa-core (4×2.7 GHz & 4×1.7 GHz)
GPU	Mali-G72 MP18
RAM	6 GB LPDDR4X RAM
내장 메모리	64 GB UFS 2.1
디스플레이	<ul style="list-style-type: none"> <li>6.2" Quad HD+ Super AMOLED (2960x1440)</li> <li>529 ppi</li> </ul>
Front Camera	8MP AF sensor <ul style="list-style-type: none"> <li>Sensor size: 1/3.6°</li> <li>Pixel size: 1.22μm</li> <li>Sensor ratio: 4:3</li> </ul>
운영체제	안드로이드 10.0

또한 본 시스템을 가정용, 개인용 로봇에 적용할 것을 고려하여 어플리케이션을 활용할 수 있는 일반 주택의 실내 적정 조도(300~700lux)에서 주로 실험을 진행하였다.

## 2. 시스템 개요

제안하는 사용자 재식별화 및 추적 시스템 개요도는 그림 1과 같다.

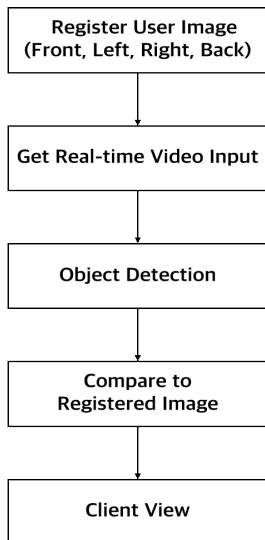


그림 1. 시스템 개요도

Fig 1. System Overview

사용자는 본 시스템이 탑재된 하드웨어를 이용하여 본인의 이미지를 등록한다 (front, left, right, back image 4장). 이후 Real-time video를 input으로 입력받는다. 본 시스템의 object detection을 통해 person object만을 추출한 후 Re-identification 과정을 거쳐 등록된 사용자만을 tracking 한다. 사용자는 이러한 과정을 display를 통해 실시간으로 관찰할 수 있다.

## 3. 네트워크 설계

### 3.1 모델 선정

본 연구의 네트워크 선정에 있어 방점을 둔 것은 모바일 기기에서도 높은 정확성과 실시간 분석을 가능하게 하는 모델을 선정, 구성하는 것이었다. 이에 모바일 환경에서도 높은 객체 탐지 성능을 보이는 YOLO v4 [4]의 연구를 참조하였다. 더불어 이미지 유사성 분석 알고리즘 중 작은 모델 크기를 가지면서도 좋은 성능을 보이는 알고리즘으로 image hash 기법을 활용하여 재식별화를 수행하였다. 궁극적으로 가벼우면서도 뛰어난 성능을 보이는 모델을 선택함으로써 모바일 환경에 적합한 네트워크를 구성하고자 하였다.

### 3.2 인물 재식별화

Re-identification 과정에 사용되는 image hash의 경우 기존 방법(Average hash) [5]을 참조하면서도 본 task에 적합하게끔 모델을 수정하여 YOLO 모델의 end 단에 post processing 하였다. 기존 방법의 경우 모든 pixel들의 평균값을 구해 각 pixel을 비교하여 평균보다 크면 1, 작으면 0으로 치환하여 이진 변환하고, 이를 hamming distance로 비교하여 유사도를 판단하였다.

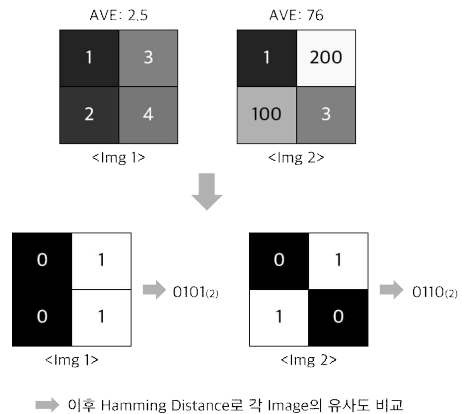


그림 2. Average hash 기법  
(2x2 픽셀 기준)

Fig 2. Average hash technique  
(Based on 2x2 pixels)

본 방법의 경우 이미지의 구조를 파악하기에 용이하나, pixel 값을 0 또는 1로만 변환하기 때문에 실제 pixel 값의 차이가 커도 해당 이미지의 평균 pixel 값보다 밝다면 이를 구분하지 못하는 한계가 존재했다. 가령, 그림 2와 같이 3과 200처럼 실제 매우 다른 pixel 값을 가지고 있음에도 불구하고 이러한 정보를 반영하지 못한다는 한계가 있다. 따라서 본 연구에서는 정교한 정도의 비교를 위해 pixel 자체의 절대적인 값을 추가적으로 이용한다. 제안하는 방법(Average filtered distance, 이하 AFD)은 다음과 같다.

1. Average-filtering: 이미지 픽셀의 평균값을 도출한 뒤, 각 픽셀값을 scaling하는 과정을 거친다.

$$\begin{cases} p_i \leftarrow 0 & \text{if } p_i < p_{ave} \\ p_i \leftarrow \frac{p_i - p_{ave}}{255} & \text{if } p_i \geq p_{ave} \end{cases} \quad (1)$$

2. Case 1 : 같은 위치에 있는 두 pixel을 비교할 때 한 pixel의 값이 0이고, 다른 pixel의 값은 0이 아닌 값을 가지는 경우, *mismatch*에 1을 더해준다. (Hamming Distance)

Case 2 : 만약 두 pixel 값 모두 0이 아닌 경우, 두 scaled pixel value의 차를 구하여 *distance*에 더해준다. (Absolute Value Distance)

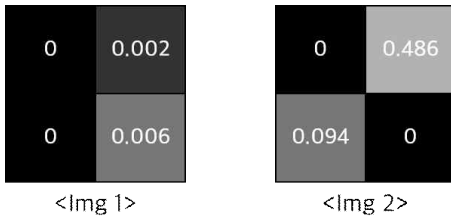


그림 3. 본 연구에서 제안하는 image hash의 적용 결과

Fig 3. image hash results proposed in this study

즉, 위의 두 이미지에서  $mismatch = 2$ ,  $distance = 0.484$  이다.

결론적으로 *mismatch*는 기존 average hash에서의 hamming distance와 일치하고, 추가적으로 *distance* 값을 활용함으로써 정밀한 유사도 계산을 가능케 하였다. 이를 통해 도출한 유사도는 다음의 식 (2)와 같다.

Similarity score =

$$100 - \left( \frac{100}{(hash\ size)^2 \times 2} \right) \times (\alpha \times distance + \beta \times mismatch) \quad (2)$$

(본 task에서  $hash\ size = 16$ ,  $\alpha = 10$ ,  $\beta = 1$  사용)

### III. 실험 결과 및 분석

#### 1. 알고리즘 실험 결과

##### 1.1 정확도

기본적으로 2.1절에서 설명한 실내 환경에서 실험을 진행하였고, 각 영상마다 주제를 조금씩 상이하게 설정하였다.

표 2. 각 영상의 주제 설정

Table 2. The theme for each video

	Video 1	Video 2	Video 3
주제	단순 객체 식별 (실내)	유사 복장 객체 식별 (실내)	단순 객체 식별 (실외)
주요 관찰 지점	단순 객체의 식별 정확도	유사 색상의 복장을 입은 대상들의 식별 정확도	조도 변화에 따른 객체 식별 정확도
영상			

각 영상에서의 정확도는 표 3과 같다.

표 3. 실험 결과: 정확도 (mAP 기준)

Table 3. Experiment result: Accuracy (Based on mAP)

	Video 1	Video 2	Video 3
Average hash	93.38% (874/936)	85.65% (555/648)	48.84% (612/1228)
AFD	93.06% (871/936)	<b>98.46%</b> (638/648)	<b>51.38%</b> (631/1228)

##### 1.2 Real-time video 분석

각 영상에서의 실시간 영상 분석 결과는 표 4와 같다.

표 4. 실험 결과: 실시간 영상 분석 (Frame 수)

Table 4. Experiment result:  
Real-time video analysis (Based on Frames)

	Video 1	Video 2	Video 3
Average hash	34.82 FPS	34.70 FPS	54.45 FPS
AFD	<b>23.81 FPS</b>	<b>22.73 FPS</b>	<b>35.71 FPS</b>

## 2. 분석 및 평가

3.1.1의 결과에 따라 본 연구에서 제시하는 기법이 비슷한 색상의 복장을 입은 대상들의 재식별에 대해 강건함을 알 수 있다. 더불어 기존 방법의 정확도에 비해 상당히 높은 식별 정확도를 나타내며 안정적인 시스템 작동이 가능함을 알 수 있다. 또한 3.1.2의 결과에 따라, 실시간 영상을 무리 없이 분석할 수 있고 실제 Client View까지 원활하게 제공할 수 있다. 다만 실외 영상의 경우 조도의 차이에 의해 Video 1, 2에 비해 낮은 정확도를 나타내었다. 그러나 해당 상황에서도 기존 기법에 비해 높은 정확도를 가지는 것으로 판단할 때, 본 연구에서 제시한 방법이 높은 성능으로 재식별 기능을 수행할 수 있다는 가능성을 확인하였다.

## 3. 시스템 구현

본 연구에서 구축한 시스템은 2.2절에서 제시된 시스템 개요에 따라 그림 4와 같이 구현된다.

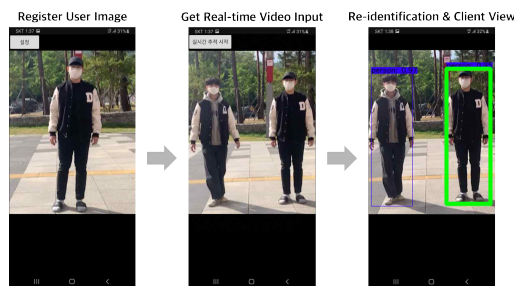


그림 4. 본 연구의 시스템 구현 화면

Fig 4. System Implementation

## IV. 결 론

본 연구는 최근까지도 활발히 이루어지고 있는 객체 탐지와 인물 재식별화를 모바일 기기 환경에서 수행하여 가정, 사무실과 같은 환경에서 인물을 tracking 하는

시스템을 구현하고자 하였다. 모바일 환경에 적합한 가벼운 네트워크를 선택하면서도 각 모델을 본 task에 맞게 최적화하는 과정을 거침으로써 실시간 영상에서도 등록된 사용자의 이미지를 높은 성능으로 식별 및 추적할 수 있었다. 한편, 조도 변화에 따라 정확도가 민감하게 변화한다는 것은 본 연구의 한계로 드러났다. 후행 연구를 통해 이러한 한계를 극복하여 안정적인 네트워크를 구축한다면, 폭넓은 분야에 본 시스템을 적용할 수 있을 것이라 생각된다. 본 어플리케이션을 적용하기 위해 요구되는 기기의 사양이 높지 않고, 복잡하지 않은 과정을 거치기 때문에 비교적 낮은 사양의 기종에서도 본 시스템이 작동될 수 있음을 시사한다. 이는 저렴한 비용으로도 객체 탐지 로봇을 제작 및 배포할 수 있는 가능성을 열며 최근 사람들의 관심을 끄는 취미용 로봇, 청소년 교육용 로봇 등의 제작에 있어 널리 활용될 수 있을 것으로 기대한다.

## 참 고 문 헌

- [1] Byon, Sungwon, et al. "A study on location information aided re-identification in CCTV environment." *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2017.
- [2] Feng, Di, et al. "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges." *IEEE Transactions on Intelligent Transportation Systems* (2020).
- [3] Wang, Robert J., Xiang Li, and Charles X. Ling. "Pelee: A real-time object detection system on mobile devices." *arXiv preprint arXiv:1804.06882* (2018).
- [4] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).
- [5] Chamoso, Pablo, et al. "A hash based image matching algorithm for social networks." *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer, Cham, 2017.

한현영 외 4인 : Human-Robot

Interaction을 위한 모바일 환경에서의 인물 재식별화 및 추적 알고리즘