

PRESS RELEASE

Researchers at the Gwangju Institute of Science and Technology Adopt Deep Learning for Multi-Object Tracking

Their novel framework achieves state-of-the-art performance without sacrificing efficiency in public surveillance tasks

Implementing algorithms that can simultaneously track multiple objects is essential to unlock many applications, from autonomous driving to advanced public surveillance. However, it is difficult for computers to discriminate between detected objects based on their appearance. Now, researchers at the Gwangju Institute of Science and Technology (GIST) adapted deep learning techniques in a multi-object tracking framework, overcoming short-term occlusion and achieving remarkable performance without sacrificing computational speed.

Keeping up Appearances: Adopting Deep Learning for Multi-Object Tracking

Tracking multiple objects simultaneously is essential for many tasks

- Autonomous driving
- Surveillance
- Industrial and scientific applications

Computers, however, cannot easily discriminate objects based on their appearance

Difficult to keep track of previous and newly detected objects

Occlusion and noise

How can deep learning help in multi-object tracking?

Introducing Deep-TAMA: Deep temporal appearance matching association

Multi-object tracking framework

- 1. Tracking and detection**
Establishing bounding boxes on tracked objects (pedestrians)
Tracking results
- 2. Geometric likelihood and gating**
Exclude unlikely object pairs based on their distance and motion
- 3. Appearance likelihood calculation with Deep-TAMA**
Each object's appearance in current frame compared with recorded appearances
The appearances of two objects simultaneously used to infer their similarity
- 4. One-to-one assignment**
- 5. Historical appearance update**
Only updated if current frame is 'favorable' for the object → Overcome short-term occlusion
- 6. Track state update and predictions**
Track initialization (new objects)
Track termination (gone objects)

Deep-TAMA offers higher accuracy than existing methods

GPU parallelization for batch acceleration → Lower computation times and higher FPS

Deep learning techniques are promising for tracking applications, which can increase public safety through real-time surveillance

Online Multiple Pedestrians Tracking Using Deep Temporal Appearance Matching Association
Yoon et al. (2021)
Information Sciences | DOI: 10.1016/j.ins.2020.10.002

GIST Gwangju Institute of Science and Technology

Computer vision has progressed much over the past decade and made its way into all sorts of relevant applications, both in academia and in our daily lives. There are, however, some tasks in this field that are still extremely difficult for computers to perform with acceptable accuracy and speed. One example is object tracking, which involves recognizing persistent objects in video footage and tracking their movements. While computers can simultaneously track more objects than humans, they usually fail to discriminate the appearance of different objects. This, in turn, can lead to the algorithm to mix up objects in a scene and ultimately produce incorrect tracking results.

At the Gwangju Institute of Science and Technology in Korea, a team of researchers led by Professor Moongu Jeon seeks to solve these issues by incorporating deep learning techniques into a multi-object tracking framework. In a recent study [published in Information Sciences](#), they present a new tracking model based on a technique they call

‘deep temporal appearance matching association (Deep-TAMA),’ which promises innovative solutions to some of the most prevalent problems in multi-object tracking. This paper was made available online in October 2020 and was published in volume 561 of the journal in June 2021.

Conventional tracking approaches determine object trajectories by associating a bounding box to each detected object and establishing geometric constraints. The inherent difficulty in this approach is in accurately matching previously tracked objects with objects detected in the current frame. Differentiating detected objects based on hand-crafted features like color usually fails because of changes in lighting conditions and occlusions. Thus, the researchers focused on enabling the tracking model with the ability to accurately extract the known features of detected objects and compare them not only with those of other objects in the frame but also with a recorded history of known features. To this end, they combined joint-inference neural networks (JI-Nets) with long-short-term-memory networks (LSTMs).

LSTMs help to associate stored appearances with those in the current frame whereas JI-Nets allow for comparing the appearances of two detected objects simultaneously from scratch—one of the most unique aspects of this new approach. Using historical appearances in this way allowed the algorithm to overcome short-term occlusions of the tracked objects. *“Compared to conventional methods that pre-extract features from each object independently, the proposed joint-inference method exhibited better accuracy in public surveillance tasks, namely pedestrian tracking,”* highlights Dr. Jeon. Moreover, the researchers also offset a main drawback of deep learning—low speed—by adopting indexing-based GPU parallelization to reduce computing times. Tests on public surveillance datasets confirmed that the proposed tracking framework offers state-of-the-art accuracy and is therefore ready for deployment.

Multi-object tracking unlocks a plethora of applications ranging from autonomous driving to public surveillance, which can help combat crime and reduce the frequency of accidents. *“We believe our methods can inspire other researchers to develop novel deep-learning-based approaches to ultimately improve public safety,”* concludes Dr. Jeon. For everyone’s sake, let us hope their vision soon becomes a reality!

Reference

| | |
|--------------------------|--|
| Authors: | Young-Chul Yoon ¹ , Du Yong Kim ² , Young-Min Song ⁴ , Kwangjin Yoon ³ , and Moongu Jeon ⁴ |
| Title of original paper: | Online multiple pedestrians tracking using deep temporal appearance matching association |
| Journal: | <i>Information Sciences</i> |
| DOI: | 10.1016/j.ins.2020.10.002 |
| Affiliations: | ¹ Robotics Lab, Hyundai Motor Company ² School of Engineering, RMIT University ³ SI-Analytics Company, Ltd. ⁴ School of Electrical Engineering and Computer Science, GIST |

*Corresponding author's email: mgjeon@gist.ac.kr

About the Gwangju Institute of Science and Technology (GIST)

The Gwangju Institute of Science and Technology (GIST) is a research-oriented university situated in Gwangju, South Korea. As one of the most prestigious schools in South Korea, it was founded in 1993 and aims to create a strong research environment to spur advancements in science and technology and to promote collaboration between foreign and domestic research programs. With the motto, "A Proud Creator of Future Science and Technology," GIST has consistently received one of the highest university rankings in Korea.

Website: <http://www.gist.ac.kr/>

About the authors

The first author, Young-Chul Yoon, is a researcher at the Robotics Lab of Hyundai Motors. This research was performed while he was pursuing a master's degree on multi-object tracking at GIST EECS under the supervision of Dr. Moongu Jeon. His research won 3rd prize from the CVPR2019 multi-object tracking challenge among 36 competitors.

The corresponding author, Dr. Moongu Jeon, is a full professor at GIST. His main research interests are in artificial intelligence, machine learning, visual surveillance, and autonomous driving. He has published over 200 technical papers in these research areas.