

Development of Artificial Intelligence in Activity Recognition

Shuohan Tao

Selwyn College, University of Cambridge, Cambridge, United Kingdom

shuohantao@gmail.com

Abstract. This essay focuses on how the technology of activity recognition has developed with the development of artificial intelligence. It illustrates three major solutions in activity recognition, including sensor-based solution, visual solutions, and Wi-Fi-based solutions, and discuss the pros and cons of each of the solutions.

Keywords: Artificial Intelligence, Activity Recognition, sensor, Wi-Fi, visual.

1. Introduction

In 1915, Max Fleischer invented the technique of rotoscoping. He manually drew outlines of real-world actors captured on frames, which were then redrawn for the production of vividly realistic animation. This technique provides a short-cut to animation making for that the artist no more need to pay extra effort thinking about how real characteristics would move their body parts, they are drawing on real world characters' outlines. Many classical animation films were produced with this technique, for example "Out of the Inkwell" in 1918 and "Snow-White" in 1933. Music videos were also produced with this technique, like Dire Straits' "Money for Nothing" in 1985. (Rotoscoping. 2022. url: <https://en.wikipedia.org/wiki/Rotoscoping>) However, due to its enormous potential in fields including film-making, human-computer interaction, and sociology, computer scientists have been trying to derive on this technique to track motion of individuals with computers. In 1983, scientists in MIT developed a technique called "Graphical Marionette". It utilized two cameras to capture a series of LEDs on the subject so as to model the individual's motion. (Delle Rae Maxwell.

"Graphical Marionette: A Modern-day pinocchio". PhD thesis. 1983) The attempt, though quite avant-garde, still failed to excite any further progress in the field. Limitation of the computer speed has put the field to slow progress for a long time. Until time has come to 21st century, when the performance of computer has increased almost a thousand time. (Samuel H. Fuller and Lynette I. Millett. "Chapter 10". In: The Future of Computing Performance: Game Over or Next Level? National Academies Press, 2011) Faster computers have led to the advent of artificial intelligence, which has made tracking motions of individuals with computers possible.

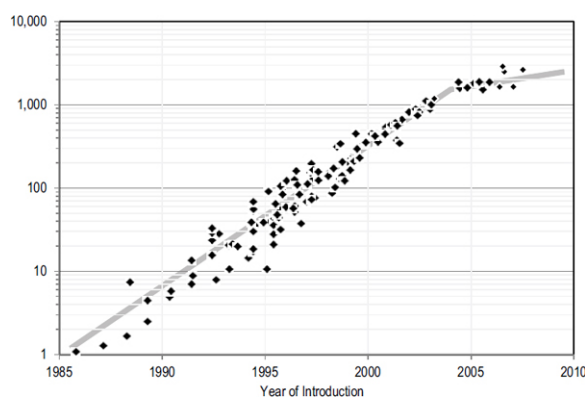


Figure 1. Clock Frequency of Computer by Year

2. Sensor-based Solutions

Sensor-based solutions is the one of the earliest approaches to track human activity. As early as 1992, attempt to track staff members within through a building has emerged. (Veronic Falcão

Jonathan Gibbons Roy Want Andy Hopper. "The active badge location system". In: ACM Transactions on Information Systems 10 (1992), pp. 91–102) The solution involved users putting on wearable "active badges" to be tracked by sensors installed across the building with infrared light. The logic in it is quite simple: any badge with less than 100% chance of showing around a certain sensor would be judged to be moving, and a dynamically updated field in the system record the last place where the badge has travelled; if the badge hasn't been detected by any sensor for 5 minutes, the field in the system will stay fixed, and the person would be classified to be "AWAY" if the badge is to stay undetected for 24 hours. The solution was primarily used by the telephone receptionist to contact the closest location within the building that a staff member could be found. Though the logic behind the solution is quite rudimentary, the scientist who proposed the solution has already shown concern for personal privacy. Later on, people have tried to record the specific physical motion of human body through accelerometers and gyroscopes, rather than just the rough locations. Gyroscopes and accelerometers record angular velocity and the accelerations in x, y, and z directions and are integrated into most smartphones, making them an ideal choice for commercial HAR (Human Activity Recognition). Making use of the data requires some preprocessing steps. Data collected from devices like mobile phones is by nature full of noise. Different users may have different habits and move in different environments, which would result in the difference in the data collected. For example, walking on a bumpy road would result in much more noisy data than walking on a flat road. To remove the noise from the data, low-pass filters with cut-off frequencies of the order of ten would usually be used to attenuate any signal with frequency larger than the cut-off frequency, as interested human behaviors generally would not exceed that frequency. Then, the data would be segregated into 2.56s intervals to generate data used for classification algorithms. The choice for classification algorithm is wide, common algorithms include K-Nearest Neighbors or simply neuron networks. The most effective algorithm in this context, however, is support vector machine, as it learns from extreme cases in the training data, which guarantees better performance than most other classification algorithms. Recently, nonetheless, voices that sensor-based solution has come to a dead end emerged. (Catherine Tong, Shyam A. Tailor, and Nicholas D. Lane. "Are Accelerometers for Activity Recognition a Dead-end?" In: CoRR abs/2001.08111 (2020). arXiv: 2001.08111. url: <https://arxiv.org/abs/2001.08111>) As shown in Figure 1 (from *ibid.*), the performance of gesture recognition has increased very little from 2013 to 2019. This is due to the fact that key challenges such as individual and environmental variations, sensor diversity and sensor placement issues are widespread and hard to overcome.

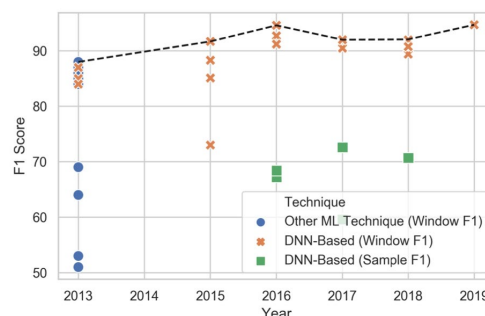


Figure 2. Gesture Recognition Accuracy by Year

3. Visual Solutions

Vision-based solution is one of the most typical and commonly used approach to behavior recognition. Such a method usually involves motion capture through a camera or optical sensors, like LiDAR. Various algorithms, like optical flow and Hidden Markov Models, could be used for different goals like single or multiple pedestrian tracking. Microsoft Kinect is a very typical successful commercial solution, and thus is suitable to be discussed here. The hardware of the product incorporates a RGB camera and a pair of infrared projector and receiver that creates depth image

based on time of flight method. The infrared projector would project a matrix of infrared rays with the receiver receiving the beams, and with the time delay and angle at which the individual ray is reflected, the orientation and distance of the surface could then be reconstructed virtually. The depth image allows for more accurate edge detection of the image captured by the RGB camera. (Kinect. 2022. url: <https://en.wikipedia.org/wiki/Kinect>) Edge detection is a kind of algorithm that identifies edges based on sudden change in feature of the pixels, like brightness, in order to separate distinctive objects in an image. As only people will move in front of the camera in most cases, it is reasonable to capture the only moving object and identify it as the user. As the product is primarily used for gaming on XBox, more accurate model of the user is required. The image of user is then processed with image classification algorithms to separate and label body parts, which then are used to virtually construct a skeletal model of the user. This way, the motion of the user is tracked with the movement of each body part measured precisely. Concerns, nevertheless, also exist for this solution. The most direct con of the solution is its invasion on privacy. In its public application like group body temperature measurement, passerby could have completely no awareness of being recorded, and situation escalates if the data recorded is not secured and are used for crime purposes.

4. Wi-Fi-based solution

Wi-Fi-based solutions is a solution that combines the pros of the two solutions above while avoid any cons of them. The approach is based on the fact that Wi-Fi signals would interact with human body when being transmitted across it. Movement of human body would cause changes in these interactions, so a Wi-Fi receiver would be able to read this information and analyse them to get the useful data for human body movement. (Activity recognition. 2022. url: https://en.wikipedia.org/wiki/Activity_recognition) This solution won't record any facial trace, so it won't interfere with users' privacy; it also provides enough data to predict the motion while works in a relatively stable environment. This also incorporates the idea of ubiquitous computing, as this technology doesn't require specific devices, all it need are Wi-Fi receivers, which exist almost everywhere. Fresnel zone is an elliptical area between the sender and receiver of electromagnetic waves. Based on the location of obstacles within the zone, the signal received may be strengthened or weakened due to diffraction, which allows the waves to travel through different path, and the combining waves may be out of phase or in phase to affect the strength of the signal received. In the case of respiration and fall detection, researchers have modelled the rib of a person as a cylinder. (Junyi Ma Yasha Wang Yuxiang Wang Dan Wu Tao Gu Bing Xie Hao Wang Daqing Zhang. "Human respiration detection with commodity Wi-Fi devices: do user location and body orientation matter?" In: UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous.

Computing (2016), pp. 25–36) As a person inhale and exhale, the size of the cylinder would change. This would cause a change in the signal phase, through which the moving distance of the rib could be calculated. Deep learning would usually be used to find the corresponding signal from all those noises. This allows for the detection of frequency and depth of respiration, and could work well for sleep quality detection. Such a technology has a promising commercial future use. For example, it could be used by large malls to analyses the pedestrian flows for better settings.

5. Conclusions

The area of activity recognition has progressed from sensor-based solution that uses accelerometer and gyroscope and support vector machine algorithm, to using vision-based solution that uses camera and various computer vision algorithms and using Wi-Fi based solution for privacy and pervasiveness. Though still primarily in theory, Wi-Fi based activity monitor has the potential to be a commercially successful solution.

References

- [1] Activity recognition. (2022). In Wikipedia. Wikimedia Foundation. https://en.wikipedia.org/wiki/Activity_recognition.
- [2] Fuller, S. H., & Millett, L. I. (2011). Chapter 10. In the Future of Computing Performance: Game Over or Next Level? National Academies Press.
- [3] Hao Wang. (2016). Human respiration detection with commodity wifi devices: do user location and body orientation matter? UbiComp'16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 25–36.
- [4] Kinect. (2022). In Wikipedia. Wikimedia Foundation. <https://en.wikipedia.org/wiki/Kinect>.
- [5] Maxwell, D. R. (1983). Graphical Marionette: A Modern-day Pinocchio [Phdthesis].
- [6] Rotoscoping. (2022). In Wikipedia. Wikimedia Foundation. <https://en.wikipedia.org/wiki/Rotoscoping>
- [7] Roy Want. (1992). The active badge location system. ACM Transactions on Information Systems, 10(1), 91–102.
- [8] Tong, C., Tailor, S. A., & Lane, N. D. (2020). Are Accelerometers for Activity Recognition a Dead-end? CoRR, abs/2001.08111. <https://arxiv.org/abs/2001.08111>.