Paper Review

Article: **"Simultaneous Tracking, Tagging and Mapping for Augmented Reality"**

Summary

This article proposes a frame for Augmented Reality. The essence of AR technology is to build a virtual world that corresponds to reality by processing the images obtained from the camera. So, the core problem is that for the camera, it not only needs to recognize but also understand the surrounding environment. For example, the AR App need to recognize the wall so that it can put a virtual picture on it. This article proposes a frame called Simultaneous Tracking, Tagging and Mapping (STTM) to implement the underlying functionality of AR. This article divides the process of knowing where the camera is and what the surrounding things are into four steps corresponding to STTM, which are tracking, tagging, mapping and loop closure [1]. First, this model uses deep SORT to analyze the time series data provided by camera to detect the features of objects filmed by the camera and then tracking their location. Then, deep learning is proposed to tag each object. Finally, a virtual coordinate system is established based on the real location of each object, details of the surface and obstacles will be added to this virtual coordinate.

This process is widely used in visual SLAM, simultaneous localization, and mapping. In visual SLAM, we used to estimate the special information only on the movement of one camera. For example, if we compare the two pictures taken at t=0s and t=1s, we can use the different view angle of a landmark to calculate the distance between the landmark and the camera. However, with the development of the hardware, we have LiDAR Scanners on many of our cellphones, the process of getting the special information can be easily acquired. Therefore, this paper adds deep-SORT-based object tracking and light weight unsupervised deep loop closure to have a better performance than traditional visual SLAM but also have higher accuracy. Figure 1 shows the output of the STTM [2].

Figure 1. STTM output

Motivation

In recent years, the concept of AR and VR has gone from hot to quiet and back to hot again. The reason behind this is that AR and VR are undoubtedly a more efficient way for us to engage with the virtual world and can be a more immersive experience than text and images. However, such interaction was previously thought to be difficult to achieve because of the limited hardware technology and the complexity of AR technology to meet the demand for low latency. However, with the recent development of more new wearable devices by manufacturers such as Apple and Microsoft, hardware technology is gradually progressing. Therefore, I researched the current algorithmic framework that can be used in solving the accuracy and efficiency of AR algorithms, and recommended this article

Discussion

This article allowed me to learn how to research a problem. Often, we think that a result is out of reach, but in reality, it is formed by stacking one model and algorithm on top of another. And at different times, due to hardware limitations, one has to optimize the algorithm. When the hardware problem is solved, we need to revisit the previous algorithmic framework and improve it. After going through this process numerous times, we were able to get a good algorithmic model

References

[1]. Kang, Y., Xu, Y., Chen, C. P., Li, G., & Cheng, Z. (2021, August). 6: Simultaneous Tracking, Tagging and Mapping for Augmented Reality. In SID Symposium Digest of Technical Papers (Vol. 52, pp. 31-33).