

Critical Analysis of the Paper

BY DEEPANSHU (B20MEO25)

Authors:

D. D. Lee and H. S. Seung

Brief Introduction to the paper:

Inexpensive, video-based, motorized tracking system that learns to track a head using convolutional neural network even when the other object are moving in a cluttered background.

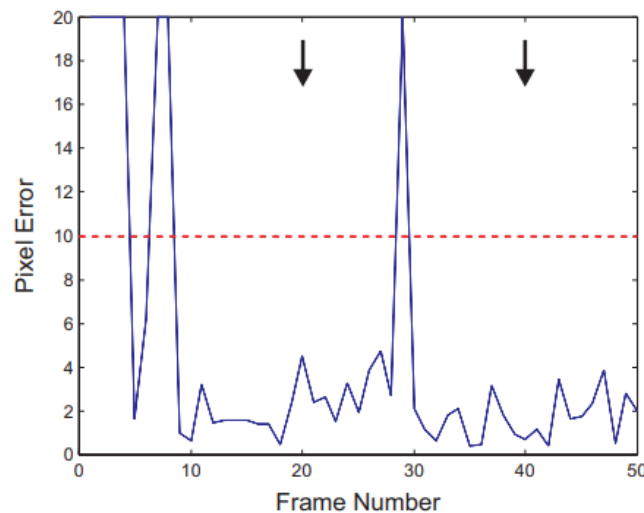
Motive of authors:

To solve the problem of video conferencing systems which makes the meeting boring and non-interactive. Without using the full space and the body language the meetings become boring and attention span is also less in this kind of meetings. User, too, is tied to a fixed location much as a corded telephone tethers one to the telephone jack.

Results of this system:

Initially, small random weights were applied and the Marvin was corrected using mouse input to look at the authors head. After only a few seconds of training with a processing time loop of around 200 ms, Marvin was able to locate the head within four pixels of accuracy. As the saccadic eye movements were initiated at the times, new environments were sampled and an occasional large error is seen. However, over time as these errors are corrected, the neural network learns to robustly discriminate the head from the office surroundings.

Given below, is an example of a learning curve showing how quickly Marvin is able to learn to track one of the authors as he moved around his office.



Advantages of this system:

1. The camera movements in this video conferencing system closely resemble the movement of human eyes.
2. Earlier it was difficult to detect the faces. Characteristics of the faces (shape and skin) were used in order to detect the face but this model uses CNN which not only learns the appropriate features of face but also learns from a **small** dataset as it learns in batch mode.
3. This system includes the microphone which allows the system to find salient object better than with either sound or visual inputs alone.
4. Saccadic camera movements are used in order to keep the maximally salient object located near the center of the field of view.
5. It can also be used at other places like music concerts where the artist is using the whole stage and the camera needs to follow them. Or any other show which require this task to be done.

Disadvantages of this system:

1. Initially the system is corrected using mouse inputs to look at the author's head.
2. The relative weighting of the different input channels shows that the luminance and color information are the most reliable for tracking the head. So, Maybe D channel is consuming some computing power which is not worth for it.
3. The whole system will be very heavy and can not be portable. It can only be used by people who uses the whole space to convey their thoughts like professors, orators, businessman, etc.

My Suggestions regarding the components:

1. If Microphone were not present then also it would have detected the face. So removing microphone will reduce the cost without affecting much on the performance of the system.
2. As we can see that the system is corrected using mouse inputs to look at the author's head, but here an Infrared Sensor can be used to locate the head and align the head at the center of the field of view. After that, infrared sensor can be turned off.
 - a. The system is like there will be an IR emitter on the person's head and an IR detector just below the CCD camera.
 - b. A heat filter renders the infrared light invisible to Marvin's video camera so that the system does not merely learn to follow this signal.
 - c. The IR detector will have a collimating lens that signals when the object is located within a narrow angular cone in the direction that the camera is pointing. This signal then can be used to train the initial weights of the neural network.

Directions for future research:

1. Neural networks are arguably the technological development with the most potential currently on the horizon. Through neural networks, we could feasibly handle almost any computational task automatically, and someday, with greater processing power than the human brain.
2. With technological advancements, we can make CPUs and GPUs cheaper and faster, enabling the production of bigger and more efficient algorithms.
3. Neural networks can be integrated with a complementary technology, like symbolic functions which is the ability to mentally represent the objects that are not in sight. Engineers are working on it.
4. The system present in the paper can be used for tracking emergency vehicles if placed on traffic lights and will alert other people around so that the roads can be cleared before their arrival and the emergency vehicle can go the location on time.

5. This system can also be used in military to keep track of enemy's planes and missiles or can be used to target them but here the speed of movement should be high.

References used:

1. https://readwrite.com/2019/01/25/everything-you-need-to-know-about-the-future-of-neural-networks/?_cf_chl_jschl_tk=__Kk4S_KGGcQk2qFbxtTuSM8hdMxTKzNOi1N1MtjoVdvc-1636208152-0-gaNycGzNO-U