CSE 573: Homework Assignment 1

Name: Sokhey, Harshdeep , UB#50247213

June 04, 2018

1 Solution 1

The **Digital Emily Project** create a texture map by capturing different levels of illumination including filtered illumination using polarizing filters for different facial expression. These texture maps are combined with markers on the face that help in recreating facial muscle movement in the digital character. The light reflected is then projected onto 3D geometry, in this case, the digital face.

The light-stage capture technology allows the visual effects team to understand the illumination of characters in a scene. Object illumination changes with the light source and in order to make the character as realistic as possible correct illumination (including angle of light source, area of illumination, change of area as the character moves) is important. This technology uses multiple cameras with multiple light sources capturing illumination from multiple view points. These image captures contains all the information including what a face looks like in a particular direction of light source and light intensity. This information is then mapped on the actor's digital character, thereby giving it a more realistic look. However, this technology requires motion capture and illumination capture to be performed each time and for every character. We could improve upon this by creating standard illumination texture maps that could be re-used across characters using generative models, that can be trained to artificially illuminate objects under the assumption that similar texture maps are produced when light interacts with similar objects. For example, two faces will produce similar shadow around the nose area with light source on the side. The intensity of light reflected may vary but the interaction of light with the object will be the same. It is similar to the process of learning to draw a portrait and being able to draw any kind of face.

2 Solution 2

The cameras equipped in our mobile devices are not equipped with the compute power to perform computer vision tasks. The *Intel Movidus* can handle these computer vision tasks in the budget of a mobile/ hand-held device. Most connected devices have limited battery capacity and the main focus is on high performance and efficiency. An additional Vision Processing Unit (VPU) eliminates these hassles which giving high performance with low power consumption.

The conventional processor architectures include deep pipelines and running at several mega-hertz and gigahertz. The *Myriad 2* Vision Processing Unit(VPU) with 12 Vector VLIW processors achieves a lot of computation within a single cycle. It is ultra-low powered hence, can achieve more processing throughput in a single machine cycle.

3 Solution 3

The Project Tango is an augmented reality computing platform for mobile devices, that allows position the device using computer vision technologies. The core technologies of Project Tango are:

- Motion Tracking, allows the phone to track it position and orientation in space. This is achieved by keeping a trajectory log using the accelerometer and gyroscope measurements of the device.
- Area Learning, allows to recognize areas and places that they have already visited. This is possible by storing the meta data like posters, sign boards seen by a tango device. This information is shared among tango devices which can then recognize the visual fingerprint captured.
- Depth Perception, utilizes data from special depth camera to detect surfaces and understand metric scale of the environment.

The "Real-time meshing in Unity w/ Physics" was an interesting demo. It combines depth perception using depth camera to generate a 3D map of the surroundings and motion tracking to determine orientation and position of the device. This can be used for applications like Augmented Reality gaming, VR-based virtual tours of historical places that are sensitive or environmentally-protected areas, et cetra.

4 Solution 4

1. **Poor Locality**: Locality enables faster computation by fetching the data required for the next stage of computation into the cache while the current stage is in progress. This eliminates memory fetch delay during the computation stage. In the example, the intermediate data needed for the blur in y stage may have been evicted from the cache due to the limited size, thus enabling a fresh fetch from the slow memory device.
2. **Redundant Recompute**: We fix the issue of poor locality by computing the data as it is available. However, we are performing too many redundant calculations and wasting compute cycles in blur in x.
3. **Poor Parallelism**: The problem of poor locality and redundant re-computation can be fixed by reusing the data and not discarding it as we go. This results in a serial dependence amongst scan lines, hence, compromising the parallelism that can be achieved.

5 Solution 5

The technique uses only RGB data as input from a standard webcam. The paper mentions using training sequence to reconstruct the shape identity of the target actor and can be used to resolve geometric ambiguities in monocular reconstruction.

The concept of **DeepFakes** is based on Generative Adversarial Networks(GANs). It is capable of morphing faces of individuals in videos. These models generate quite realistic and believable fake videos with features of individuals. However, this can have serious consequences including but not limited to generating fake adult videos by morphing a face onto a adult actor's body. Another serious consequence could be creating fake extortion videos for the purpose of blackmailing. The

Deepfakes could be used to generate propaganda videos which could be a direct threat to national security. For example, generating fake videos showing public officials taking bribes or a video showing a world leader engaging in adultery. A good example that was cited in the video was that of President Obama's fake speech that was generated using this concept.