Cartoon Transfer To A 3D Model

Name: Seongsil Heo,

Home institution: Kookmin University

Proffessor/Mentor : Gopi Meenakshisundaram, Jia Chen

**Abstract -** Recently, many people like watching 3D movies and playing 3D games. However, it is difficult for non-professionals to create 3D objects. This paper will attempt to explain how to make a 3D model by transferring 2D to 3D automatically. Because it is important for 3D to be expressed in a more active way, the goal of this project is to make a more active 3D model by giving it lifelike qualities. This will be done by extracting pieces of video of poses and movement and tracking those parts in real time using TLD (tracking learning detection) algorithm in the opencv library.

The result of this project will be near total human movement in 3D thereby offering for the non- professional the opportunity to use 3D in everyday life.

Key Terms – 3D Modeling, TLD, real-time, opencv, libigl

**1. Introduction**

Recently, many people like watching 3D movies and playing 3D games because they expect them to resemble the real world. We tried to make a 2D video convert to 3D automatically using software. Non-professionals just put the object in the program and this program will change the 2D object to a 3D object. It will save time and money, and non- professionals can also use this program anytime or anywhere. This process is called 3D modelling.

However, there are still difficulties with this automatic process. It is sometimes changes the 2D model to 3D model unnaturally. For example, the objects may move unnaturally. We are going to track the object on video in real-time. This means to track an object over a sequence of images. This process will help us make the object pose like a real person, and move like a real person better than old 3D models. We applied computer vision and geometry processing to transfer pose and movement from a real person to a 3D model.

**2. Pose and Movement extraction method**

To track a specific person in video in real-time, we need tracking algorithm included in software called opencv. At first, We try to extract face and hands automatically, but it sometimes does not extract well. So we change the way to extract the part of the video that we need by creating a bounding box around the object manually instead of automatically. This makes the extraction more smooth. So its important to have accurate object tracking. There are many tracking algorithms to track the specific object such as optical flow, meanshift, TLD etc.

* 1. **Optical flow method**

Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the movement of object or camera. It is 2D vector field where each vector is a displacement vector showing the movement of points from first frame to second. Consider the image below figure 1.

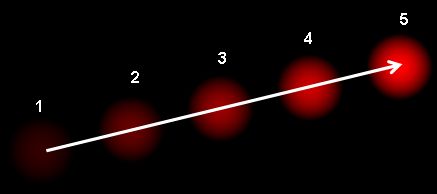


Figure 1

It uses the block matching method. Block matching method is divide the frame in fixed size, and find the blocks which is one of the most similar block compared to before blocks and then guess again using that block. They compare the before blocks and now blocks and find the most similar blocks. So it has a good accuracy. However, when we find the accurate value while guessing movement of vector, we should calculate the difference of average absolute value for every displacement surrounded in blocks. It takes long time when the window size is bigger. Also, because the size of vector is limited according to the block size, we have restriction to find movement vector.

It tracks accurately when the objects rotate, change the size, and parts of the object is hidden. However, if the object move so fast, it is hard to track.

* 1. **Meanshift method**

The mean-shift algorithm is an efficient approach to tracking objects whose appearance is defined by histogram. Consider we have a set of points. It can be a pixel distribution. We are given a small window(may be a circle) and we have to move that window to the area of maximum pixel density. It is illustrated in the simple image given below figure 2.

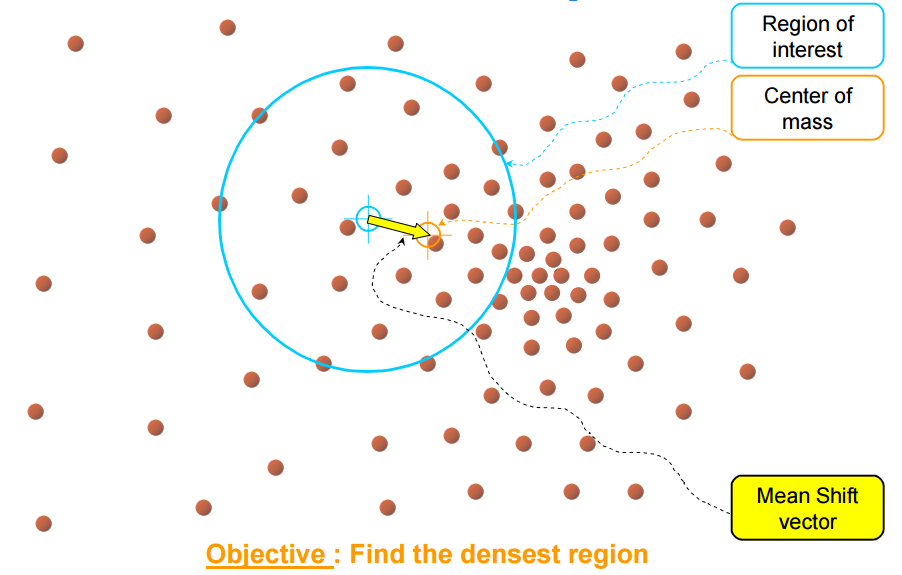
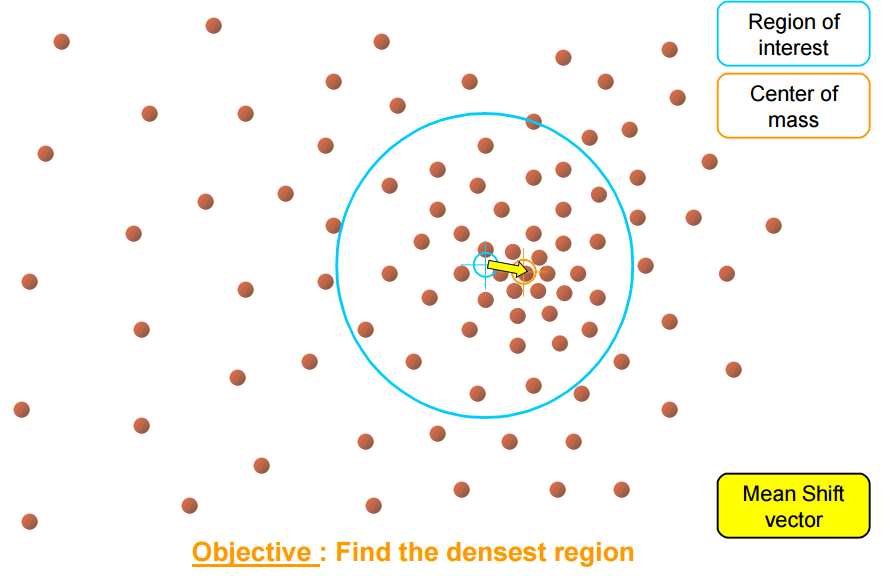
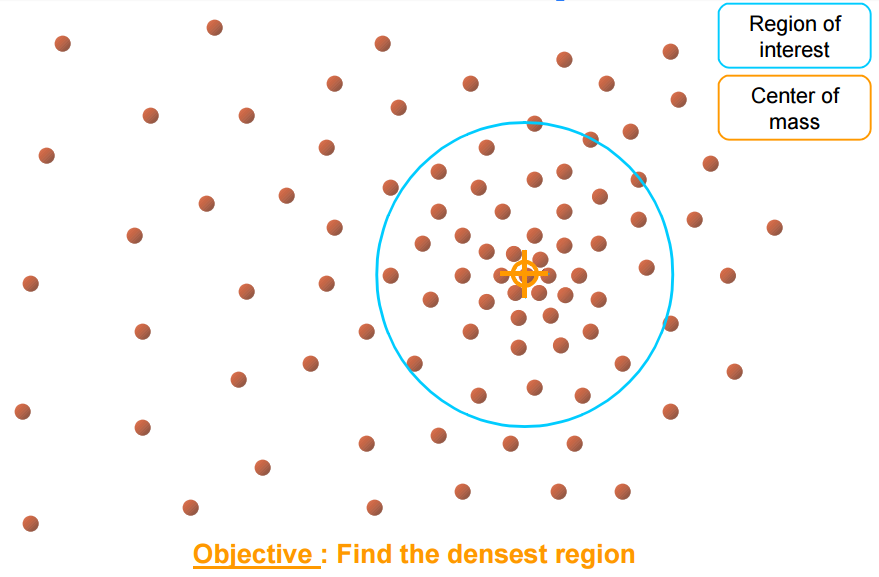
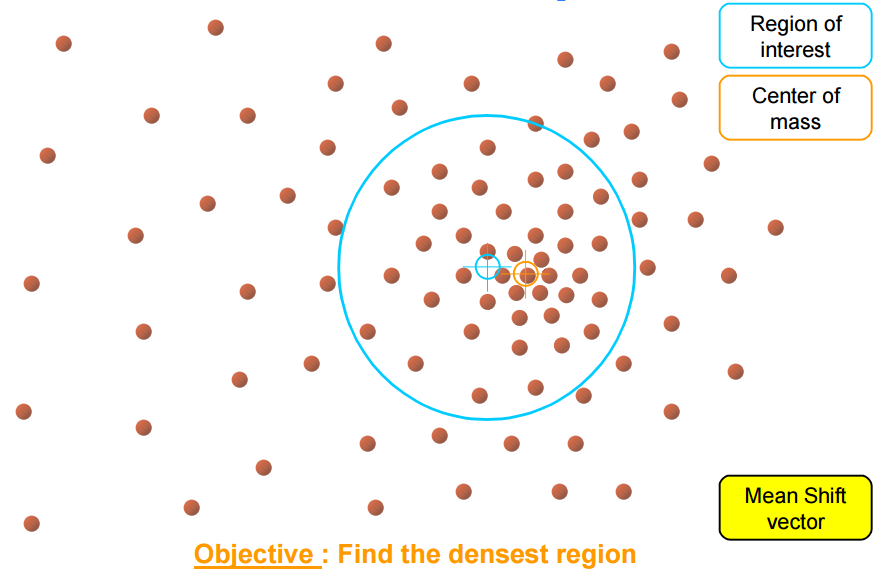
  

Figure 2

It is good to track even if the shape of the object changes. However, if the color of the object is similar as surrounded object, it is hard to track the object.

* 1. **TLD method**

TLD is a tracking learning detection. It is a real-time algorithm for tracking of unknown objects in video streams. The object of interest is defined by a bounding box in a single frame. TLD simultaneously tracks the object, learns its appearance and detects it whenever it appears in the video. When the object disappear and appear again, Optical flow and Meanshift method can’t find the object anymore. However, TLD has both detector and tracker. So TLD method can find the object again even if they disappear before. In our thought TLD algorithm track accurately than other algorithm. So we will use TLD algorithm in our project. You can see the result in the below figure 3.

They use P-N tracker. When user assigns the domain that they want to track, tracker and detector initialize simultaneously. After the frames input, they find the object using both tracker and detector. If tracker succeed to find the object, the data will be updated and makes detector’s reliability more powerful. Also, if the detector find some scope, and if it isn’t correspond to the scope that the tracker found, it will be added on the negative examples. So it will also makes detector’s reliability more powerful.

And if the tracker fails tracking, wait until the detector succeeds. If the detector succeeds, initialize the tracker to the detector’s scope, and start tracking again.



Figure 3

**2. 3D shape manipulation**

After extracting the part of the video using TLD algorithm, we store the center of the bounding box’s coordinates using matrix. Look at the below figure 4. We store the coordinate of the vertices in V matrix, and store the triangle connectivity in F matrix in real time.

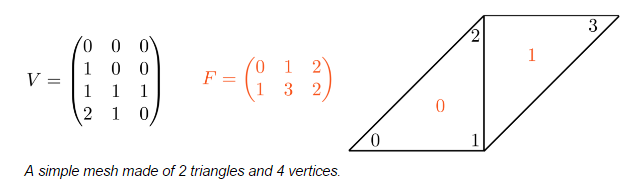


Figure 4

And, we can change it to 3D As Rigid As Possible in the libigl library. The shape is represented by a triangle mesh. We can move several vertices of the mesh naturally using the information in matrix. The system will computes the positions by minimizing the distortion of each triangle. The key idea is to use non-linear deformation techniques so that each they can minimize the error of changing the complicated shapes such as non-trivial bending and twisting.

**3. Result**

When we input the video, we can extract the information that we want by ourselves. See the Figure 5. We extract the face, hands, and foots by ourselves using bounding box. And when a woman change her position, it still track the position that we extract in real time.

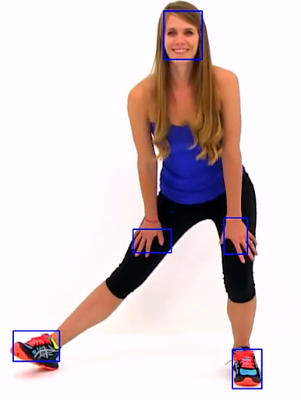
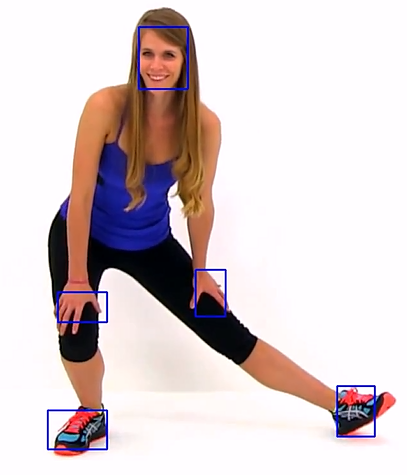


Figure 5

Next, we make 3D using libigl library. See the Figure 6. That is the result of our project. They can pose and move in various direction like a real person in 3D. When they move, we can see they seem like moving smoothly.



Figure 6

We still sometimes can’t track the object. We need to research tracking the object more accurately, so we can change them to 3D more accurately.

**4. Works Cited**

http://publications.lib.chalmers.se/records/fulltext/136882.pdf

http://libigl.github.io/libigl/tutorial/tutorial.html

http://docs.opencv.org/3.1.0/

http://cvpr.uni-muenster.de/teaching/ss10/BildverarbeitungundComputerVisionSS10/script/ BVCV-12-Tracking.pdf