COL780: Project(2D to 3D conversion)

Jaskirat Singh(2013EE10460), Nitish Agarwal , Pawan Sharma

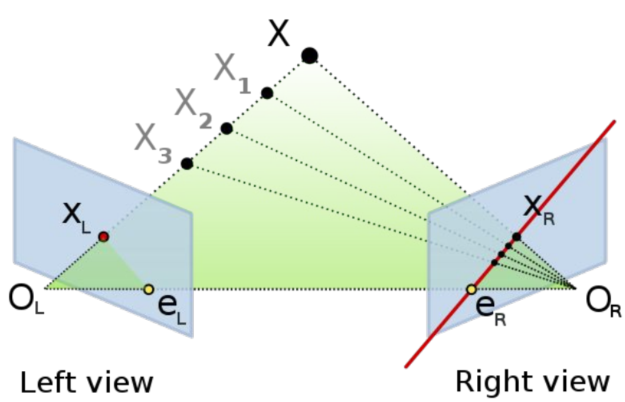
November 2016

1 Problem Statement

Given a 2d video we have to generate the corresponding 3d video of the same scene. For our project we choose the video of a tennis match which helps us to have a lot of constraints for camera calibration.

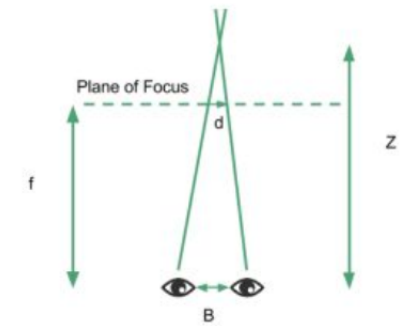
2 Basic Problem

The basic problem we face is that the triangulation from the right view (given left view) is not directly possible, since each point maps to a ray in 3d and hence cannot be uniquely located unless we gather some other information.

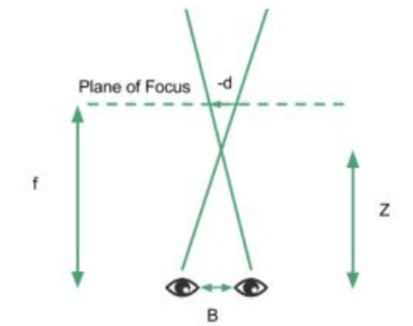


3 Need for depth map

So the solution to this problem is to somehow find the depth map for the input image or dimension map along any direction.This is because different points will have different disparity based on their depth.



Object further than imaginary screen plane



Object nearer than imaginary screen plane

4 Using depth maps

If D is disparity between the left and right views then given one view the other view can be calculated as :

$$D = \frac{B(Z - f)}{Z}$$

$$O_{i,j} = \sum_{d} I_{i,j}^{d} D_{i,j}^{d}$$

$$O_{i,j+D_{ij}} = I_{i,j}.$$

$$L = |O - Y|$$

5 Implementation 1:Approximate 3D

It included the following steps:

- Perform camera calibration using points on the court and their real world co-ordinates.
- Right view of ground is calculated using simple homography corresponding to translation of camera center by inter eye distance.
- The players are segmented out using background subtraction
- All points on the player are assumed to be on a plane and a similar homography is performed.
- Thus we finally get our approximate 3d conversion which is quite good.

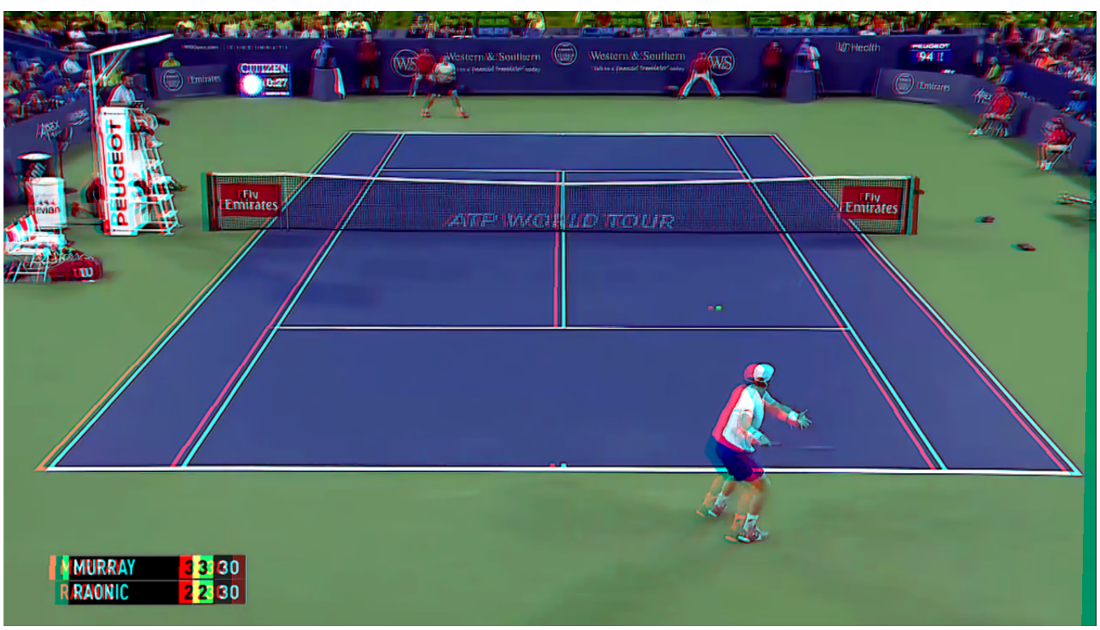


Figure 1: anaglyph 3d output for one of the frames

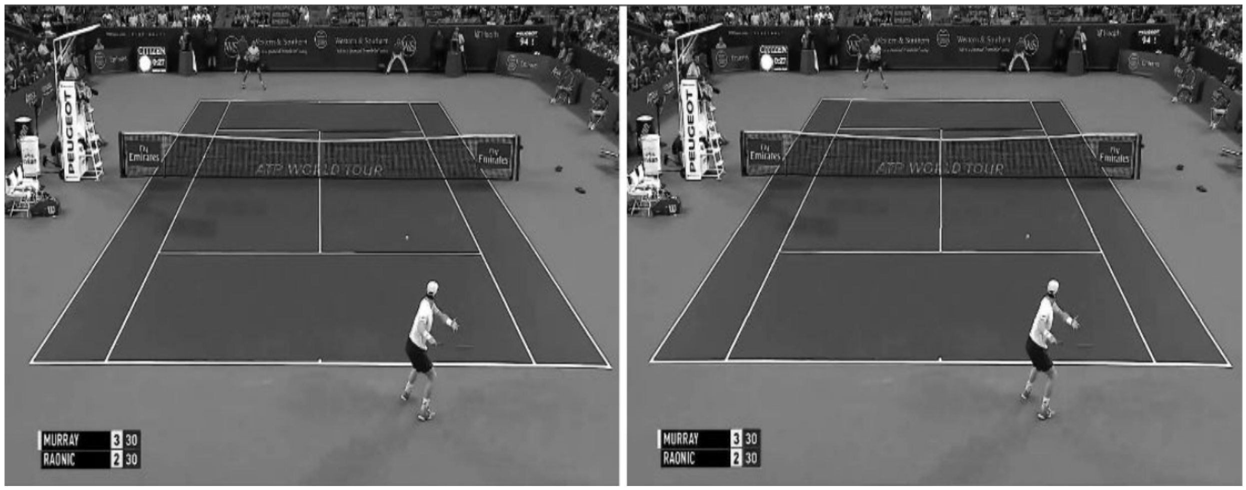


Figure 2: left and right views side by side for one of the frames

6 Deep 3D Implementation

The basic problem in the previous case was that we did not know the exact depth map for all of the points.Recently Xie at al,2016 showed a deep 3d architecture that can be directly trained on the stereo pairs and generates the right view given the left view. Here is the basic architecture:

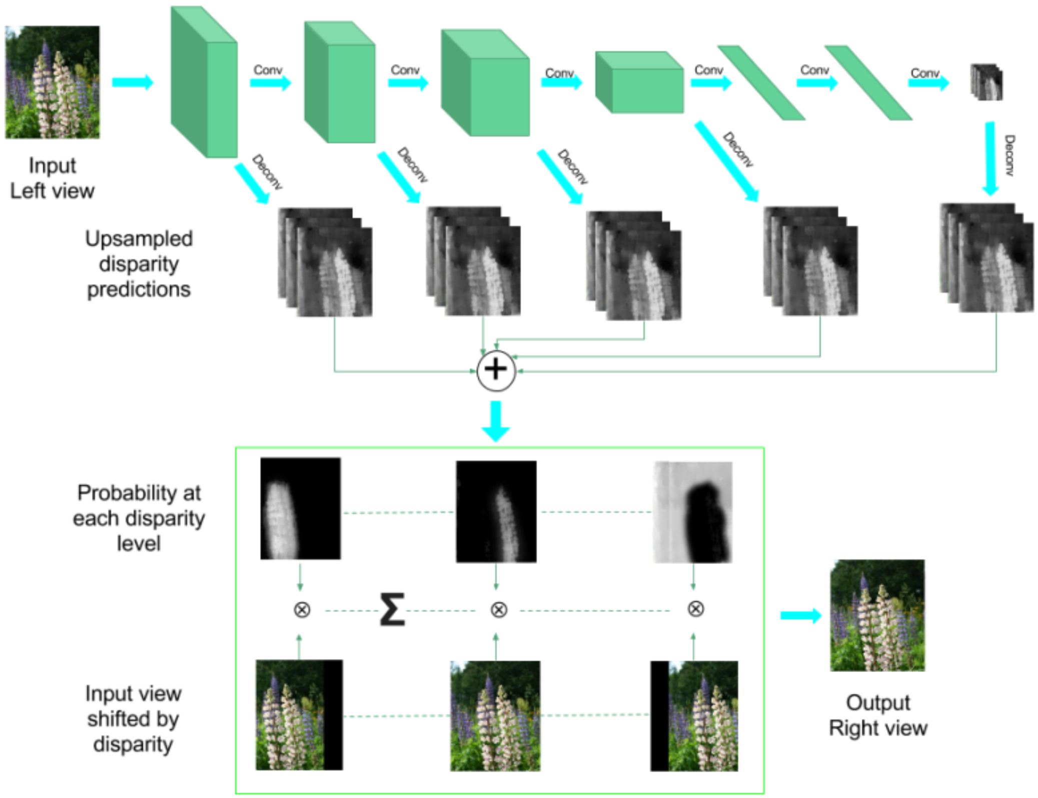


Figure 3: basic architecture for deep-3d network

Please refer to the submission folder for 3d video output using this model.

7 References

[1]Xie, Junyuan, Ross Girshick, and Ali Farhadi. "Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks." arXiv preprint arXiv:1604.03650 (2016).