Lecture 7: Multiple View Vision

Adam Hawley

December 26, 2018

Contents

1 Epipolar Geometry

Considering two pinhole cameras with different projection centres, in order to relate the two images one can use *Epipolar Geometry*.

- p_l & p_r are the vectors from the centres of projections O_l in the left and O_r in the right image to the corresponding projections P_l & P_r of the same 3D point P.
- h_l & h_r are called **epipolar lines** — they are located at the interection between the image plane and the plane formed by the points P, O_l & O_r. Each point P will have an epipolar line in each image plane.
- e_i & e_r are epipoles representing the intersection points of O_lO_r with the left and right image planes. They may be located outside the actual images.

To estimate the epipolar geometry, determine the mapping between corresponding points in the two images.

$$p_r = Rp_l + t$$
$$O_lO_r = t$$

The left and right images are connected by means of a matrix representing rotation **R** in the plane PO_lO_r and translation **t**.

Epipolar geometry depends only on the cameras' internal parameters and relative pose.

$$\text{If } O_lO_r, P_rP, P_lP \text{ are coplanar } \implies \mathbf{p}_l \cdot [\mathbf{t} \times (\mathbf{R}\mathbf{p}_r)] = 0$$

This results in the Fundamental matrix.

$$\mathbf{p}_r^T \mathbf{E} \mathbf{p}_l \text{ where:}$$
$$E = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \mathbf{R}$$
$$rank(\mathbf{E}) = 2$$

We have a well defined geometry linking the epipolar vectors of right and left images by means of matrix **E**.

2 Image Rectification

Image rectification is implemented by multiplying the image matrices with a geometric transformation matrix. Epipolar lines become parallel following image rectification.

3 Depth Maps

Depth maps represent the distance from the image plane to the actual 3D scene. In computer vision depth maps are estimated from the disparity between pairs of images. If images are aligned, the disparity represents a displacement along the epipolar lines.

In computer graphics, the depth map (also called *Z-Buffer*) is calculated as the distance from the 3D scene to the image plane.

3.1 Depth Maps for Photorealistic Effects

Depth maps can be used to apply effects to specific objects/regions of the image. For example, portrait modes can use depth to apply a blurring effect to everything in an image except for a face.

3.2 Estimating Depth Maps

Depth maps can be estimated from stereo images based on fields of disparities. Each disparity represents the correspondence of a block of pixels from one image into the other. Objects which are far away will be in identical places in both images (disparity is 0). Objects which are near the camera will have a displacement according to their depth.

3.3 Block Matching

Comparing the two images and finding the disparities is done using a method called Block matching. Block matching is done using the following procedure:

1. Calculate differences between the block of pixels, pixel-by-pixel, from left image and all blocks from a search region in the right image.
2. The search region is centred at the location of the original block of pixels.
3. Choose the block of pixels from the second image which is the most similar with the orignial block and define a vector as the difference of their locations.

The size of the block of pixels $B_x \times B_y$ should not be too small (which may lead to confusion) or too large (requiring large computational complexity and low disparity resolution). Block matching works well when we have well defined features or non-regular textures. Block matching does not work well where we have regions of constant colour or regular textures.

3.4 Estimating Depth Maps ii

When estimating the depth maps from real images, local errors emerge due to:

- Left and right images are affected by the noise of the image sensors and changes in light.

- Results in wrong matches and consequently in errrs in the depth map.

3.5 3D from Multiple Views

Several images are provided from various angles and the 3D scene is reconstructed as made up of voxels. Surfaces are visible only from some cameras. For this process, the cameras have to be calibrated. Usually, this calibration is done using chessboard patterns.

4 Space Carving

The 3D scene is formed using space carving. Space carving starts with a cube in the scene. If there is no projection onto any of the images for that voxel, then it is carved away.Otherwise — get all pixels where it projects. Use all of the images to estimate the colour of the voxel from the pixels' colour.

When a voxel is seen from several images a statistical estimation takes place (averaging or median statistics).

Floating voxels and uneven surfaces emerge due to wrong correspondences and are caused by:

- Confusions in colours due to the illumination and noise
- Errors in cameras' geometry due to wrong camera calibration.

5 Conclusions

- Stereo vision
- Epipolar geometry
- Image rectification
- Depth maps
- Multi-view vision for 3D scene estimation
- Space carving method