Computer Vision

All documents are allowed. The different sections below are independent. Answers should be explained and should be concise.

1 Part I

1.1 Essential matrix (1 point)

The essential matrix between two images is given by (see course notes):

$$\mathsf{E} \sim \mathsf{R}_2 \left[\mathbf{t}_1 - \mathbf{t}_2\right]_\times \mathsf{R}_1^\mathsf{T}$$

This allows to compute an epipolar line in image 2, from a point in image 1. What is the essential matrix for the other direction (going from image 2 to image 1)?

1.2 3D Reconstruction from two images (2 points)

3D points can be reconstructed from image points determined in the two image planes, for example by intersecting the lines of sight corresponding to these image points. However, there exist 3D points which cannot be reconstructed in this manner, i.e. the intersection of the lines of sight is not unique or not defined. For which 3D points is this the case (describe their locus, with respect to the optical centers of the two cameras)?

1.3 Establishing correspondence of lines (3 points)

In the lecture, we have dealt with the epipolar geometry, which gave us a means to establish correspondence between points in two images (two points are correspondences if they lie on the respective epipolar lines).

Here, we examine the same scenario, but now for lines. Concretely, we consider two cameras, whose position, orientation, and intrinsic parameters are all known (see figure 1, right). We further consider one line $\mathbf{d_1}$ in the first image, and one line $\mathbf{d_2}$ in the second one. We assume that each of these lines is the projection of a line in 3D space.

(a) Is there a means to decide if the lines $\mathbf{d_1}$ and $\mathbf{d_2}$ are possible correspondences, i.e. if they can be projections of a single 3D line?

Try to answer using geometric arguments only, without giving formulas (they are too complicated).

(b) Suppose now that we have 3 cameras. Consider one line in each image ($\mathbf{d_1}, \mathbf{d_2}$ and $\mathbf{d_3}$). Is there a means to decide if all 3 lines are possible correspondences, i.e. if they can be projections of a single 3D line?

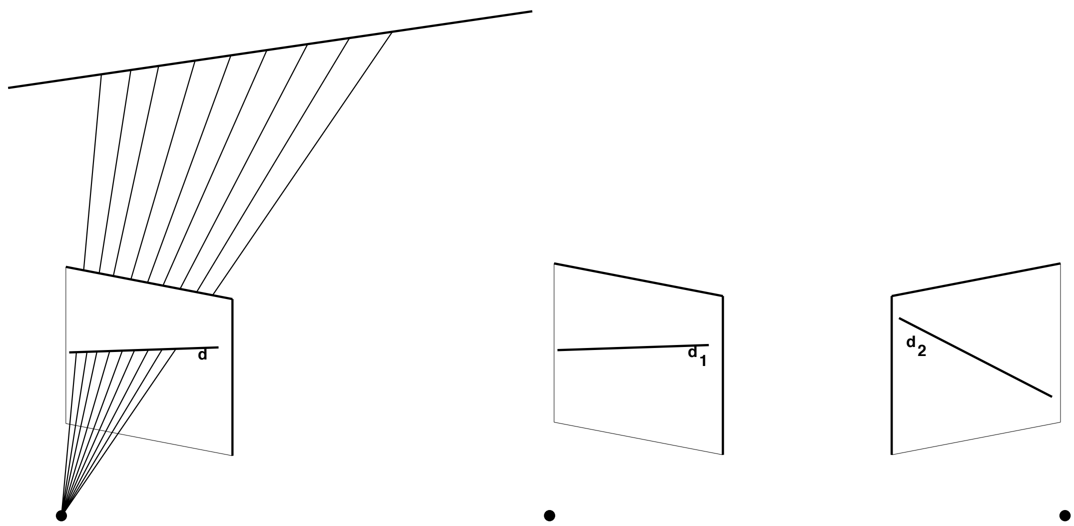

Figure 1: Left: Projection of a 3D line onto an image. Right: The scenario considered here: two cameras, with one line in each image.

1.4 Image Mosaics (4 points)

Let us consider the same setting as the one considered for the generation of image mosaics: a camera acquires images while rotating about its optical center. Consider the following special case. The camera takes **two** images. Between the two images, it carries out a rotation about the Y axis:

$$\mathsf{R} = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}$$

As for the intrinsic parameters of the camera, we suppose that they correspond to a simplified calibration matrix whose only unknown is the focal length α:

$$\mathsf{K} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

During the lecture, it has been shown that there exists a projective transformation (an homography) that links the two images, and that it can be estimated from 4 or more point matches.

The goal of this question is to estimate the homography from a single point match. To do so, write down the homography H in terms of the two unknowns, the rotation angle β and the focal length α. Then, develop a method (formulas) for computing these two unknowns, from a single point match ($\mathbf{q_1}$ in the first image and $\mathbf{q_2}$ in the second one).

2 Part II

2.1 Projective Geometry (4 points)

1. Why is projective geometry useful in vision and graphic applications ?
2. How do we go from a projective space to an affine space ?
3. How many correspondences between pairs of points are required to estimate a homography in P^n ?
4. What is an invariant for a group of transformation and why do affine transformations preserve parallelism ?

2.2 Surface modeling (3 points)

We are given a 3D modeling system composed of n cameras. Cameras are fixed and calibrated and we consider the 3D modeling process using information extracted from the images.

1. Assume a finite number n of silhouettes are available, is the topology (i.e. gender and number of connected components) of the corresponding visual hull necessarily equal to the topology of the observed object ?
2. Propose an algorithm that computes the visual hull based on a space discretization into voxels (the algorithm will be sketched only).
3. The object under consideration is not completely observed by all cameras. Modify your algorithm, if necessary, so that it handles also this situation.
4. In the case of a Lambertian (diffuse) surface can we improve such model ? how ?

2.3 Motion modeling (3 points)

Modeling motion consists in obtaining, from images, information on the motion of a moving object in the scene.

1. In the previous section object shapes are estimated from images. Do these shapes encode motion information ? What is the step required to get motion information from these shapes ?
2. Explain (sketch) how the a motion capture system, typically Vicon, works and identify the principle steps required, once images are acquired, to estimate motion with such a system.
3. Kinect: How is human motion represented with the kinect ? Which strategy, generative or discriminative, is used to determine motion ? why ?