

**Introduction to Computer Vision**  
**Exercise 3: Stereo and Structure from Motion**

Due Date : 1.1.17

Question 1:

A camera is imaging an object at two time instances. Let  $P_i = (X_i, Y_i, Z_i)^T$ ,  $i = 1, \dots, N$  be a set of 3D points in space at the first time instance, and  $p_i = (x_i, y_i, f)^T = \frac{f}{Z_i} P_i$ . At the second time instance the points are displaced relative to the camera by  $P'_i = RP_i + t$ , where  $R$  is the rotational component and  $t$  is the translational component of either the camera or object motion. In the second image therefore the points are projected to  $p'_i = \frac{f}{Z'_i} P'_i$ . Assume point correspondences are given across the two images.

CAN depth ( $Z_i$ ) be recovered if between the two images:

1. There is only a camera rotation ( $t = 0$ )?
2. There is only an object rotation (about its center of mass)?

In case depth cannot be recovered show explicitly that it can be eliminated from the equations relating the two images. (*Hint:* show the existence of a mapping from  $p_i$  to  $p'_i$  that does not include 3D structure (i.e., depth  $Z_i$ ). Explain why the existence of such a mapping proves the claim.)

Question 2:

1. SHOW that the right epipole (the epipole in the first image) is given by  $v = \alpha R^T t$  (for some  $\alpha \neq 0$ ). *Hint:* To show this show that  $E v = 0$ .  
EXPLAIN why this implies that  $v$  is the epipole (the intersection of all epipolar lines).
2. DERIVE an expression for the left epipole (in other words, find  $u$  such that  $u^T E = 0$ ).  
*Hint:* invert the rigid transformation that relates  $P$  and  $Q$  and use the expression derived for the right epipole.

Question 3:

A calibration pattern is a set of 3D points at known positions. Such a pattern is used to calibrate a stereo rig (determine the transformation between the two cameras).

Suppose we take two images of a calibration pattern with the cameras at unknown positions and orientations (so a point  $P$  from the pattern is expressed as  $R_1 P + t_1$  in the coordinate system of image 1; and  $P \rightarrow R_2 P + t_2$  in image 2). To calibrate the cameras in this case we need a more general expression for the essential matrix.

EXPRESS the essential matrix between the two images in terms of  $R_1, t_1, R_2, t_2$ .

Question 4:

1. Show that the intersection of any two parallel lines in the projective plane  $P^2$  lies at infinity. Recall a line in  $P^2$  is defined by  $\{p|l^T p = 0\}$  s.t  $l = (l_1, l_2, l_3)$ .
2. Let  $H$  be a homography, show which line  $H$  maps  $l = (l_1, l_2, l_3)$  to.
3. Suppose  $H$  represents an affine transformation. Show where  $H$  maps the lines at infinity.

Question 5:

Let  $H$  be a non trivial homography ( $H \neq \alpha I$ ). A point is called *fixed* if it does not change location following the application of  $H$ . A line is called *fixed* if each point on the line is mapped by  $H$  to some point along the same line. A line is called *pointwise fixed* if each point along the line is fixed.

1. Derive an expression for a fixed point; what should such a point satisfy? How many isolated points (i.e. assuming a finite number of points) can be fixed under  $H$  ?
2. In the case of maximal number of isolated fixed points, identify the fixed lines (Hint: use question 4.2)
3. How many lines can be pointwise fixed?

Question 6:

Two images of the same scene are related by general rotation and translation. Corresponding points  $p_i$  and  $p'_i$  in the two images satisfy the fundamental relation  $p_i'^T F p_i = 0$ . Suppose that all the points  $P_i$  lie on some plane, and so corresponding points in the two images are related by a homography  $p'_i \propto H p_i$  (i.e.,  $p'_i = \alpha_i H p_i$  for some scalar  $\alpha_i \neq 0$ ). SHOW that in this case there exist infinitely many solutions to the fundamental matrix  $F$ .

*Hint:* A matrix  $M$  is called skew-symmetric if it has the form:

$$M = \begin{pmatrix} 0 & m_{12} & m_{13} \\ -m_{12} & 0 & m_{23} \\ -m_{13} & -m_{23} & 0 \end{pmatrix}.$$

Note that the three following conditions are equivalent:

1.  $M$  is a skew-symmetric matrix
2. For all  $x$ :  $x^T M x = 0$ .
3.  $M + M^T = 0$ .

(Bonus): Prove (2)  $\longrightarrow$  (1),(3)

- (a) SHOW that  $H^T F$  is a skew-symmetric matrix.
- (b) Taking this fact into account, how many linearly independent equations can be induced for  $F$ ? (use hints above). Is it possible, then, to obtain  $F$ ?
- (c) If the points  $P_i$  come from more than just one plane in the scene, how many planes are needed in order to uniquely recover  $F$ ?

Question 7:

Let  $p$  and  $q$  be two corresponding points in two images. Assuming the images are obtained with scaled orthographic projections

- (a) SHOW that the coordinates of the points  $p = (x_p, y_p)$  and  $q = (x_q, y_q)$  are related by the following expression (epipolar constraint):

$$Ax_p + By_p + Cx_q + Dy_q + E = 0.$$

EXPRESS the coefficients  $A, B, C, D$ , and  $E$  in terms of the components of the rotation  $R$  and translation  $t$  that relates the two images. (*Hint:* Begin with  $Q = RP + t$  and eliminate  $Z$ .)

- (b) EXPLAIN how the constraint above determines epipolar lines.
- (c) SHOW that the epipolar lines are parallel.
- (d) WRITE the epipolar constraint in the form of an Essential (Fundamental) matrix  $q^T E p = 0$  using homogeneous coordinates and DERIVE the coordinates of the two epipoles. What is the rank of  $E$ ? EXPLAIN.

Question 8:

RANSAC is a method for finding a transformation between images from point matches when some (possibly many) of the matches are erroneous. The objective of RANSAC is to identify the transformation that is supported by the largest number of matches.

Given two images, suppose we use a feature detector to identify  $m$  points in the first image and  $n$  points in the second image, and suppose we do not know in advance which of the  $m$  points in the first image matches which of the  $n$  points in the second image. Write an expression for the number of transformations that may be computed with RANSAC in the worst case and the total run-time complexity for the following cases:

1. The transformation relating the two images is a homography.
2. The two images are related by a general fundamental matrix and we use the 8-point algorithm to compute the transformation.