Computer Vision EX3 Guy Lutsker

1. a.

Let us look at:

$$P_1X = K\cdot[I_3,0]X = \begin{bmatrix}1&0&1\\0&1&1\\0&0&1\end{bmatrix}\cdot\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}\cdot\begin{bmatrix}-1\\1\\2\\1\end{bmatrix}=\begin{bmatrix}1\\3\\2\end{bmatrix}$$

And since we know that x_1 needs to be in homogeneous coordinate, we shall divide P_1X by its last coordinate to get $\begin{bmatrix}\frac{1}{2}\\\frac{3}{2}\\1\end{bmatrix}$.

Similarly:

$$P_2X = K\cdot[R_2,,t_2]X = \begin{bmatrix}1&0&1\\0&1&1\\0&0&1\end{bmatrix}\cdot\begin{bmatrix}1&0&0&1\\0&\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}&2\\0&\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}&-1\end{bmatrix}\cdot\begin{bmatrix}-1\\1\\2\\1\end{bmatrix}=\begin{bmatrix}\frac{3-\sqrt{2}}{\sqrt{2}}\\\frac{2}{\sqrt{2}}\\\frac{3-\sqrt{2}}{\sqrt{2}}\end{bmatrix}$$

Again, dividing by last coordinate gets us $x_2=\begin{bmatrix}1\\\frac{2}{3-\sqrt{2}}\\1\end{bmatrix}$

b.

$$\begin{bmatrix}\frac{1}{2}\\\frac{1}{3-\sqrt{2}}\\1\end{bmatrix}^T\begin{bmatrix}0&\sqrt{2}&-\sqrt{2}\\-1&-\frac{1}{\sqrt{2}}&1\\0&0&0\end{bmatrix}\begin{bmatrix}\frac{1}{2}\\\frac{3}{2}\\1\end{bmatrix}=\begin{bmatrix}\frac{1}{2}\\\frac{1}{3-\sqrt{2}}\\1\end{bmatrix}^T\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{\sqrt{2}-3}{2\sqrt{2}}\\0\end{bmatrix}=\frac{1}{\sqrt{2}}-\frac{1}{\sqrt{2}}=0$$

c.

Recall that epipolar lines are actually the left are right nullities of F and so we can see that:

$$\ell_1=F^Tx_2=\begin{pmatrix}0&-1&0\\\sqrt{2}&-\frac{1}{\sqrt{2}}&0\\-\sqrt{2}&1&0\end{pmatrix}\begin{bmatrix}\frac{1}{2}\\\frac{1}{3-\sqrt{2}}\\1\end{bmatrix}=\begin{bmatrix}-6-2\sqrt{2}\\4\sqrt{2}-2\\6-5\sqrt{2}\end{bmatrix}*\frac{1}{7}$$

$$\ell_2=\begin{pmatrix}0&\sqrt{2}&-\sqrt{2}\\-1&-\frac{1}{\sqrt{2}}&1\\0&0&0\end{pmatrix}\begin{bmatrix}0.5\\1.5\\1\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{3\sqrt{2}-3}{2\sqrt{2}}\\0\end{bmatrix}$$

d.

Firstly, notice that $Fe_1 = \begin{pmatrix} 0 & \sqrt{2} & -\sqrt{2} \\ -1 & -\frac{1}{\sqrt{2}} & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} \frac{\sqrt{2}-1}{\sqrt{2}} \\ 1 \\ 1 \end{bmatrix} = 0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}^T \begin{pmatrix} 0 & \sqrt{2} & -\sqrt{2} \\ -1 & -\frac{1}{\sqrt{2}} & 1 \\ 0 & 0 & 0 \end{pmatrix} = e_2^T F$

And now observe that:

$$e_1 \ell_1 = \begin{bmatrix} \frac{\sqrt{2}-1}{\sqrt{2}} \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -6-2\sqrt{2} \\ 4\sqrt{2}-2 \\ 6-5\sqrt{2} \end{bmatrix} * \frac{1}{7} = 0$$

And that:

$$e_2 \ell_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}^T \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{3\sqrt{2}-3}{2\sqrt{2}} \\ 0 \end{bmatrix} = 0$$

2.a.

Since we know that these four points a, b, c, d are within the same plane - $Z = 1$, we know from class we can use the formula for deriving the homography $H = R + \frac{1}{d} t n^T$.

We can derive that we can use R as the given R_2, and from the same reasoning that $t = t_2$.

From the fact that the points are within the same plane as mentioned before, we can deduce that $d = 1$ and that $n = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ (since it needs to be orthogonal to the plane).

And so:

$$H = R_2 + t_2 n^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}^T = \begin{pmatrix} 1 & 0 & 1 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

b.

We can see the from image 1 of the question that the 2d coordinated of $e_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and in homogeneous coordinates $e1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. In addition, we know that $e_1 \propto P_1 E$. And from the givens we know we the third coordinate of E is 2, so we can denote $E = \begin{bmatrix} x \\ y \\ 2 \end{bmatrix}$, and plugging it all in gets us:

$$\Leftrightarrow \begin{bmatrix}1\\2\\1\end{bmatrix} \propto \begin{pmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\end{pmatrix}\begin{bmatrix}x\\y\\2\\1\end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix}1\\2\\1\end{bmatrix} \propto \begin{bmatrix}x\\y\\2\end{bmatrix} \to x = 2\cdot 1 = 2\,, \qquad y = 2\cdot 2 = 4$$

$$\to E = \begin{bmatrix}2\\4\\2\end{bmatrix}$$

c.

No, when applying H on e_1 do we do not get e_2. The reason is because when we calculated the homography H we did so with the assumption that the scene in within the same plane $Z = 1$, and this assumption does not hold for E, since it is in $Z = 2$.

3. a.

Let us investigate the function of this rotation matrix.

If we look at $P_i' \overset{t=0}{\triangleq} RP_i$ we would like to extract depth information (as per the hint, we should look at z_i) and so under this multiplication it is easy to see that $z_i' = r_{31}X_i + r_{32}Y_i + r_{33}Z_i$.

Converting this equation to use x_i, y_i, z_i use: $X_i = \frac{z_i}{f}x_i,\ Y_i = \frac{z_i}{f}y_i$.

And so:

$$z_i' = r_{31}\frac{z_i}{f}x_i + r_{32}\frac{z_i}{f}y_i + r_{33}Z_i.$$

Which means we can express z_i' explicitly.

Now let us look at the givens:

$$p_i = \frac{f}{z_i}P_i, \qquad p_i' = \frac{f}{z_i'}RP_i \overset{P_i=\frac{z_i}{f}p_i}{\triangleq} z_i/z_i' \cdot Rp_i$$

Which means we can express p_i' easily using the relation between $\frac{z_i}{z_i'}$ and R. This can help us build a mapping between p_i and p_i'.

And so, let's use it to get such a mapping:

$$p_i' = \frac{z_i}{z_i'}\cdot Rp_i = \frac{z_i}{r_{31}\frac{z_i}{f}x_i + r_{32}\frac{z_i}{f}y_i + r_{33}Z_i}\cdot Rp_i = \frac{r_{31}x_i + r_{32}y_i + r_{33}f}{f}\cdot Rp_i$$

As we can see this mapping does not depend on z_i information at all, and as it encapsulates the relationship between the images totally and doesn't use depth information at all we can conclude that depth cannot be recovered.

b.

In this case depth information is recoverable.

(It is not clear from the question what level of detail is required to answer this question, but I hope the following explanation is sufficient)

Intuitively, rotating an object around its center of mass gives us a second perspective upon the object, which allows us to perceive depth information. This is true because it is equivalent to keeping the object still and rotating and translating the camera such that we see the object from another angle. In this case we have seen in class that depth information can be derived, and so in this case it is computable as well.

4.

As we know that R_1, R_2, t_1, t_2 are the rotations and translations of any point P in the scene, we can represent them using our two image coordinate systems: $P_1 = R_1P + t_1,\ P_2 = R_2P + t_2$.

And so, we can write that:

$$P_1 = R_1^{-1}(P_1 - t_1)$$

And so: $P_2 = R_2R_1^{-1}(P_1 - t_1) + t_2 \rightarrow P_2 = R_2R_1^{-1}P_1 - R_2R_1^{-1}t_1 + t_2$.

And as per the essential matrix, we can express it as:

$$E = (R_2R_1^{-1})(-R_2R_1^{-1}t_1 + t_2)$$

5.

a.

Consider a standard 3×3 affine transformation matrix, consisting of rotation (upper 2×2 part), and a translation part (last column): $\begin{pmatrix} h_{11} & h_{12} & t_x \\ h_{21} & h_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix}$.

Now consider some point at infinity $p_\infty = \begin{bmatrix} p_x \\ p_y \\ 0 \end{bmatrix}$.

We need to evaluate the effect this affine transformation has on this point (intuitively, infinity should not be affected by any such transformation, but let's check it explicitly):

$$\begin{pmatrix} h_{11} & h_{12} & t_x \\ h_{21} & h_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix}\begin{bmatrix} p_x \\ p_y \\ 0 \end{bmatrix} = \begin{bmatrix} h_{11}p_x + h_{12}p_y \\ h_{21}p_x + h_{22}p_y \\ 0 \end{bmatrix}$$

Such a point is still considered to be at infinity since its last coordinate is 0.

We can conclude that points at infinity are invariant to affine transformations.

b.

Denote C to be some conic, and H to be some homography. Assume p is some point in the projective plane, we see that:

$$0 = p^T C p \overset{\substack{H^{-T}H^T=I \\ HH^{-2}=I}}{\triangleq} (p^T H^T H^{-T}) C (H^{-1} H p)$$

Which means we can denote this equality as $(Hp)^T (H^{-T} C H)(Hp) = 0$

From this we can deduce that we can use the form $H^{-T} C H$ as a new conic structure under this generic homography H.

6.a.

Given that we have n points in image 2 and m points in image 1, we shall have to make some combinatoric calculation. We know that a homography needs 4 pairs of point, and so per image the number of options of choosing 4 points out of n samples is $\binom{n}{4}$ and similarly for the other image is $\binom{m}{4}$. And so, in the worst case the runtime complexity of the algorithm is $O(n^4 m^4)$.

(We could also recall that there could be several ways of distributing these 4 pairs into equations which is a constant of $4! = 24$, which is not meaningful in the asymptotic runtime calculation).

b.

Recall that by the 5-point algorithm for computing the transformation we iterate over $\binom{m}{5}$ point matches, which translates to computational complexity of $O(m^5)$ operations of transformations asymptotically.

c.

Observe that from a randomly chosen set of 5 pairs of points, the probability that all 5 are matches correctly is $\frac{1}{32}$, while the completing probably of having at least one unmatched is $\frac{31}{32}$.

We can see that a geometric RV that describes the problem in question can be analyzed for its expectation, where $x \in Geo(\frac{1}{32})$. Recall that the expectation of a geometric RV is 1/ its probability, and explicitly by the formula of expectation:

And so $\mathbb{E}[X] = \sum_{i=1}^{\infty} \frac{i}{32} \left(\frac{31}{32}\right)^{i-1} = \frac{\frac{1}{1}}{\frac{1}{32}} = 32$.

7.a.

Recall that as learned in class the problem of structure from motion is to derive the depth information of all points – namely reconstruct the original 3d points from the scene P_i.

In addition, in order for us to calculate then we are looking to solve for all R_i, t_i the rotations and translations of the points with respect to the image planes. We also assume that all transformations are orthographic.

Let us look at the objective:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(r_i^T P_j - x_{ij}\right)^2 + \left(s_i^T P_j - y_{ij}\right)^2$$
$$s.t \ ||r_i|| = ||s_i|| = 1 \ \wedge r_i^T s_i = 0$$

Firstly recall that since we are requiring orthographic projection here, and as mentioned in the question $R_i = \begin{bmatrix} r_i \\ s_i \end{bmatrix}$ we will impose that r_i, s_i will be an orthonormal set - $||r_i|| = ||s_i|| = 1 \ \wedge r_i^T s_i = 0$. This explains the constraint given in the objective.

As seen in class we can alleviate the search of the translation as the translation can be acquired by looking at the centered points $p_{ij} = \widetilde{p_{ij}} - mean(p_{ij})$. This leaves us with looking for the rotations alone.

By this formulation we can see that by trying to minimize $\left(r_i^T P_j - x_{ij}\right)^2$ and $\left(s_i^T P_j - y_{ij}\right)^2$

With respect to r_i, s_i should give us the best fit – meaning the obtained transformation R_i will give more accurate reconstruction of the 3d information.

Overall, we have seen both why the objective has the form $\sum_{i=1}^{m}\sum_{j=1}^{n}\left(r_i^T P_j - x_{ij}\right)^2 + \left(s_i^T P_j - y_{ij}\right)^2$

And why the constraints are $||r_i|| = ||s_i|| = 1 \ \wedge r_i^T s_i = 0$.

b.

By changing the projection type to be scaled orthographic we add an additional scale factor to our formulation (uncentered). This means that we can write that instead of $\widetilde{p_{ij}} = R_i P_j + t_i$ we get that here $P_{ij} = c_i(R_i P_j + t_i)$, where c_i is the multiplicative constant per image.

This means that we cannot just replace t_i with the mean $\overline{p_{ij}}$ right away we need to check the new mean. Although, it's quite easy to see that if each image was multiplied by some constant, so will the mean, and so we can write that by the same logic as presented in class we can write that t_i could be alleviated from the objective if we will replace the points by their centered values – $p_{ij} = \widetilde{p_{ij}} - \overline{p_{ij}}$.

Now, as before, we need to take care of the rotation. Here it should again be quite similar since we just got a multiplicative constant the objective shall be:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(c_i r_i^T P_j - x_{ij}\right)^2 + \left(c_i s_i^T P_j - y_{ij}\right)^2$$
$$s.t \ ||r_i|| = ||s_i|| = 1 \ \wedge r_i^T s_i = 0$$

The objective is almost the same for the same reasons as mentioned in (a), and the multiplicative constant is added to represent the scaled orthographic projection, and the constraints stay the same.

Altogether minimizing this objective should give an accurate solution to the problem.