Q4.

The order of the fundamental matrices is as the images submitted are numbered, first frc from plot 1 to 4, then inria from plot 1 to 3.

F = 0.0031 -0.1609 -0.1745

0.1175 -0.0207 0.8132

0.1027 -0.4006 0.3117

F1 = -0.0013 -0.1162 0.0725

0.1130 0.0006 -0.6779

-0.0540 0.7084 -0.0641

F2 = 0.0024 -0.0569 -0.0513

0.0685 0.0144 -0.7490

0.0145 0.6413 0.1294

F3 = -0.0021 -0.0380 0.0291

0.0383 0.0098 -0.7334

-0.0039 0.6736 -0.0678

F4 = -0.0006 0.0380 -0.0238

-0.0335 0.0016 -0.7226

0.0328 0.6882 -0.0117

F5 = -0.0012 -0.1452 0.0432

0.1452 0.0015 -0.6489

-0.0310 0.7289 -0.0520

F6 = 0.0002 0.1119 -0.0073

-0.1122 0.0016 0.6704

0.0053 -0.7247 0.0073

How the points were selected:

I tried to space out the points around the image so long as it is within the common view of the 2 pairs of images, and in general I tried to take any noticeable corner points that are easy to map the correspondence between the images, or sometimes center points, such as the center of the meter like object on the pipes on the left side of the frc images, and the center of the car wheel on the inria images. As far as possible I tried to include points at different depths and covering more area across the image.

To comment on the estimation of F and the accuracy:

The estimation seemed much more accurate for the inria image pair than the frc image pair, maybe because the inria image had clearer and more easily noticeable features. The epipolar lines pass much closer to the desired points in the inria images than the frc images.

The frc image seems to have more occlusion and luminance change issues which may affect the points and estimation, and the angle of the 2 cameras are much more different than in the inria images; in the frc images the 2 cameras face towards each other, while in the inria images they appear to face much more similar directions to each other.

The estimation seems more stable for n = 16; changes in the points selected affected the estimated F less for n = 16 than for n = 8; this is expected as more points should theoretically increase tolerance of noise, so long as the points are not outliers.

However somehow for my frc pairs n=8 led to better estimated F than n = 16 in terms of accuracy (as judged by epipolar lines passing through correct points)… perhaps my points for n = 16 were not all well chosen.

Q5 a. If we are looking through aperture 1, we perceive a surface moving up and down along a direction of 135 degrees counterclockwise about the z-axis, starting from the x-axis, assuming the z axis to be out of the plane of the screen and x axis to be pointing to the right. This is because all we would see is one continuous edge of a homogeneous surface with no corner points or junctions, and in such a case only the motion in the direction perpendicular to the surface edge can be perceived.

If we look through aperture 2, we perceive the true motion of the surface moving horizontally left and right. We are able to perceive the true motion in this case because we have a unique point of the object within the viewing window, which is the top corner of the lighter gray square. Having unique points such as true corners in the field of view allows the perception of the true motion of the surface to which the corner belongs, even if we cannot see the whole surface.

If we look through aperture 3, we perceive two fixed overlapping surfaces moving up and down (along the y axis direction. This is because there is a feature point in the viewing window, which is the L junction formed by the overlap between the 2 squares, and the perceived motion is the motion tracking the movement of this feature point.

Q5 b. Without occluders, all the endpoints of the stripes on the barber-pole are perceived as legitimate unique feature points. In that case, the direction of perception of motion (either vertical or horizontal) is the direction along which there are more endpoints. If the barber-pole is long and thin, there are more ‘endpoints’ along the vertical direction, the motion perceived from this larger number of endpoints dominates and vertical motion is perceived. If the barber-pole is short and thick, there are more ‘endpoints’ along the horizontal direction, and horizontal motion is perceived.

With the addition of occluders, the endpoints along the direction parallel to the occluders are now perceived as being T-junctions, produced due to the occlusion of 1 contour by another, while the ‘endpoints’ along the direction orthogonal to the occlusion boundary continue being perceived as ‘endpoints’ or reliable features to track motion. The vision system understands that T-junctions can be formed from occlusion.

Thus in the case with occluders, instead of the motion along the direction with more ‘endpoints’ being perceived as the dominant/true motion, the endpoints not occluded by the occluders are taken as being more reliable or representative of the true motion of the striped surface as compared to the T-junction endpoints along the direction parallel to the occluders (regardless of the number of endpoints in either direction). Thus the perception from the endpoints overwhelms the perception from the T-junctions and that is why the perceived motion tends to be biased in the direction along the endpoints, which is orthogonal to the occlusion boundary.

Q5 c. In Figure a and b, the perceived motion is only in the direction perpendicular to the bar, since this is the case of 2 parallel homogeneous edges in a limited viewing window, and in this case only motion perpendicular to the direction of the edges is perceived; if there was left and right motion in figure a or up and down motion in figure b, it cannot be perceived unless the end of the bar comes into view, because the bar is homogeneous and there is no optical flow generated by these motions. In Figure c, the 4 corners are perceived as T-junctions, which are unreliable. This leaves the center of intersection of the 2 bars as the reliable unique feature point favored by the vision system, and this point is perceived to be moving circularly. Since the true end points of the bars are occluded, and the cues in the viewing window are all consistent with the case if it was a cross moving circularly about the intersection of its 2 bars, this circular motion is what is perceived.

In figure d, the true ends of the bars are not occluded from view, and their motion direction can be unambiguously observed and this is perceived as the true motion direction, with the endpoints of the bars overwhelming the perception from the T-junction at their intersection.

5 d.

The equation of the form Ax = b would be as below:

Where A is of dimensions 2n x 3, vector x is of dimensions 3 x 1, and b is a vector of dimensions 2n x 1.

To solve for the x vector, just take x = b, where is the pseudoinverse of A.

To convert to homogeneous form, we would need to take u and v to the other side in both equations, and then sum the two equations.

It would not be appropriate to change it to the homogeneous form, because in that case the 2 separate components of optical flow u and v would get combined into 1 single constant in the homogeneous form of the equation, and consequently the information about direction of the flow field at the point would be lost, which is needed to be able to calculate the rotational motion components and their directions.