C4B Computer Vision II

david.murray@eng.ox.ac.uk Michaelmas 2004
www.robots.ox.ac.uk/∼dwm/Courses/4CV/ Rev 08.11.04

Do email if you are puzzled about a question's meaning. The reply, if generally useful, will be placed on the web page.

1. The camera shown in the Figure has its x, y and z axes aligned with the world's y, z and x axes respectively. The world frame's origin is at $(0, -h, 4h)$ in the camera's frame.

   (a) Find the camera's extrinsic camera calibration matrix $[\mathtt{R}|\mathbf{t}]$, such that

   $$\mathbf{X}_C = \begin{bmatrix} \mathtt{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{X}_W$$

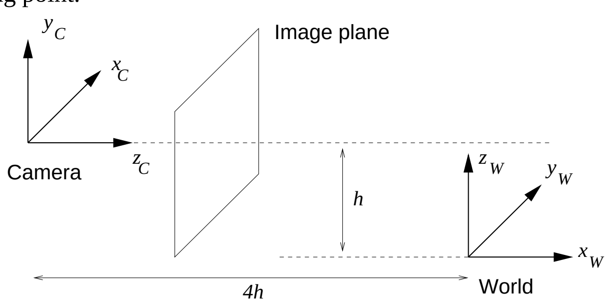
   (b) How many degrees of freedom does the extrinsic matrix possess, and what is the minimum number of points whose world and camera coordinates you would need to check the extrinsic matrix is correct?
   Check your result by testing it for this number of points.

   (c) The "ideal" image plane is placed at
   $Z_C = 1$. (Ideal means that the intrinsic camera matrix $\mathtt{K}$ is just a 3×3 identity matrix.)
   Derive the image coordinates of the vanishing point of the family of lines parallel to the following line, expressed parametrically as:

   $$(X_W, Y_W, Z_W) = (2 + 4t, 3 + 2t, 4 + 3t) \ .$$

   (d) The intrinsic calibration matrix maps ideal image positions onto actual positions in pixels. Explain the physical meaning of focal length f, aspect ratio α, and principal point u_o, v_o, and show why and where they appear in the calibration matrix. (You may assume the skew s is negligible.)

   (e) The actual camera has $f = 800$ pixels, $\alpha = 0.9$, $s = 0$, and $(u_o, v_o) = (350, 250)$ pixels. Derive the coordinates in the actual image plane of the vanishing point.



2. (a) Setting $h = 1$ in the extrinsic geometry of Question (1), determine the image positions that would be obtained from six corners of a unit cube at world positions

   $$\mathbf{X}_{1-6} = (0,0,0)\ (0,1,0)\ (0,1,1)\ (0,0,1)\ (1,0,1)\ (1,1,1)$$

   given that the camera now has the (rather implausible) intrinsic calibration matrix

   $$\mathtt{K} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} .$$

   (b) Starting with the projection equation $\mathbf{x} = \mathtt{K}[\mathtt{R}|\mathbf{t}]\mathbf{X}$, describe an algorithm which uses the measured image positions $(x, y)_i$ of at least six points with known positions in the world $(X, Y, Z)_i$ to recover the extrinsic and intrinsic calibration parameters.

   (c) Download the matlab code `calibration.m` from the web page at the top of the sheet, and annotate it with comments to confirm that it fits with your algorithm description in part (b).

   (d) Use the code to recover the extrinsic and intrinsic calibration from the image and world point data.

3. With a more realistic calibration matrix involving pixel values, such as that in Question (1), numerical values become rather large, and it is important to perform *statistical centring*.

   This involves applying a linear transformation $\mathbf{x'} = \mathtt{S}\mathbf{x}$ to the measured image points so that, in the primed coordinates, the means and variances are $\bar{x'} = 0$, $\bar{y'} = 0$, $\sigma^2_{x'} = 1$, $\sigma^2_{y'} = 1$.

   (a) Denoting the mean and variance of the measured image positions as $\bar{x}$ and σ^2_x, and similarly for y, show that the matrix to apply is

   $$\mathtt{S} = \begin{bmatrix} 1/\sigma_x & 0 & -\bar{x}/\sigma_x \\ 0 & 1/\sigma_y & -\bar{y}/\sigma_y \\ 0 & 0 & 1 \end{bmatrix} .$$

   (b) Putting $\mathbf{x}'$ rather than $\mathbf{x}$ into the calibration algorithm of Question (2) is obviously going to alter the $\mathtt{K}$ returned.
   What extra step has to be inserted into the calibration algorithm to recover the proper $\mathtt{K}$?

4. Most modestly priced lenses, particularly those with short focal length and wide-angle of view, give rise to *radial distortion*, in which the image positions $\mathbf{x}$ obtained from the projection equation are displaced radially in the image. For small κ, the distorted radial distance is

$$r_d = \frac{r}{\sqrt{1-2\kappa r^2}}$$

where $r^2 = x^2 + y^2$. For $\kappa > 0, \kappa < 0$, the distortion is call pin-cushion and barrel distortion, respectively. The effect on imagery can be seen in the images below.

(a) Using y as a parameter, derive an expression for how the straight image line $x = d$ appears under radial distortion.

(b) You have an edge detector and straight line fitter at your disposal, and can point the camera at any scene you wish. Develop the outline of an algorithm to determine κ.

(c) Derive an expression for the undistorted r as a function of r_d and κ.

Pin cushion

No distortion

Barrel

5. (a) What is the aperture problem, and why does it arise in gradient-based schemes for computing motion?

(b) Show how the analysis of the image intensity that underpins the Harris method of corner detection can be extended to compute image motion.

(c) Explain how the image motion so computed would differ from that computed by matching of discrete corner features.

(d) The aperture problem may also exist when matching straight lines between views taken from widely separated viewpoints. If the problem does exist, what information about the image lines must be lacking?

6. (a) A scene $\mathbf{X}$ has instantaneous translational velocity $\mathbf{V}$ and angular velocity $\mathbf{\Omega}$ relative to a perspective camera with focal length f.

Derive an expression for the projected image motion, assuming the image plane to lie in front of the optic centre.

(b) A camera with focal length of 1000 pixels (and otherwise ideal with $\alpha = 1$, $s = 0$, and $(u_o, v_o) = (0, 0)$) is translating and also panning with a rotation about the y-axis such that $\mathbf{\Omega} = (0, 0.1, 0)$ rad/s. Image motion at the pixel positions $(x, y) = (500, 0)$, $(0, 500)$ pixels is found to be $(\dot{x}, \dot{y}) = (25, 0)$, $(84, -40)$ pixels/frame where the video frame rate is 25 frames/s.

i. In which direction is the camera translating?

ii. What are the relative depths of the two points?

(c) A camera with focal length $f = 10$ mm is known to be translating along its optic axis, such that $\hat{\mathbf{V}} = (0, 0, -1)$mm/s, with an additional but unknown rotation $\mathbf{\Omega}$. Image motion at the points $(x, y) = (10, 0), (0, 10), (5, 0)$mm is measured to be $(\dot{x}, \dot{y}) = (-4, 2), (-2, +4), (-4, -1)$ mm/s respectively

i. Show that $\mathbf{\Omega}$ can be fixed in general by measuring the motion field at 3 points.

ii. Find $\mathbf{\Omega}$.

7. A camera with focal length of 1000 pixels (and otherwise ideal with $\alpha = 1$, $s = 0$, and $(u_o, v_o) = (0, 0)$) is mounted on a vehicle looking forward with its optic axis horizontal at a height h above the ground. The vehicle moves over a planar ground surface $\mathbf{X} \cdot \hat{\mathbf{N}} = -h$ with unit normal $\hat{\mathbf{N}} = (0, 1, 0)^\top$ with speed s, so that the scene's motion relative to the camera is $\mathbf{V} = (0, 0, -s)^\top$ and $\mathbf{\Omega} = (0, 0, 0)^\top$.

(a) Derive an expression for the motion field from the ground plane, and state in which region of the image the expression holds.

(b) An obstacle of height αh appears on the ground a distance D in front of the camera.

i. Derive and expression for the projected motion from the topmost point of the obstacle.

ii. Find the value of α that would make the magnitude of the image motion from the top of the obstacle 10% larger than that which would have been measured had the obstacle not been there.

8. (a) Suggest, with reasons, which method of visual tracking might be employed for the following tasks:
   i. Visually guiding a robot repair craft to dock with the Hubble space telescope.
   ii. Tracking cardiac walls in ultrasound imagery.
   iii. Tracking a speeding car from a police helicopter.

   (b) Explain the operation of the Lucas-Kanade tracker for a general warp $\mathbf{p}$.

   (c) For the case of a 3-dof Euclidean transformation, derive $\partial\mathbf{f}/\partial\mathbf{p}$, and show how the updated warp is derived from $\mathbf{f}_{\Delta\mathbf{p}}$ and $\mathbf{f}_{\mathbf{p}}$.

9. A design question. Use reasonable guesses if a quantity is not given.

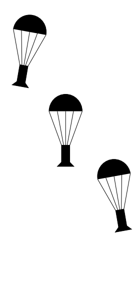
   A colleague from JPL recently gave a seminar to our vision group describing how his team had used visual tracking from a downwards looking camera to determine whether the *lateral* speed of Spirit (and Opportunity) was greater than 3 ms^{-1} during its final parachute descent to the surface of Mars. It *was*, and a lateral motor was fired to reduce the speed so that the cushioning bags would not be split by being dragged over sharp rocks.

   The lander had a radar altimeter, and an inertial attitude sensor to give the orientation of the camera relative to the Martian surface as it swung beneath the parachute. The parachute slowed the lander from Mach 2 to around 100ms^{-1} before releasing the heat shield, and then to a vertical speed of 50ms^{-1}. The lateral measurements were made at around 1500m when the speed was around 75ms^{-1}

   (a) Describe, with appropriate equations, how might the system might work, and give consideration to the visual features that might be used in the tracking system.

   (b) In the last seconds before hitting the surface, retrorockets were fired to slow the craft from 50ms^{-1} to rest some 20m above the surface, at which point the bags deployed and the parachute was cut off.

   The retrorockets were actually fired using the altimeter. How could the divergence of a motion field be used instead? Comment on whether it would have been practical to do so.



Some Answers and Hints

See web page given at start of sheet.