The Camera

Computer Vision CMP-6035B

Dr. David Greenwood

david.greenwood@uea.ac.uk

SCI 2.16a University of East Anglia

Spring 2022

Contents

The Camera

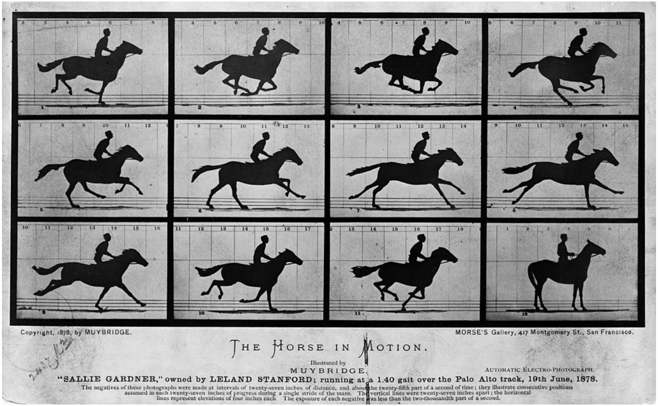

Figure 1: "Sallie Gardner," owned by Leland Stanford; ridden by G. Domm, running at a 1:40 gait over the Palo Alto track, 19th June 1878.

The Camera

Cameras measure light **intensities**.

- the sensor counts photons arriving at the pixel
- each pixel corresponds to a direction in world space

The Camera

Cameras can also be seen as *direction* measurement devices.

- we are often interested in geometric properties of a scene
- an object reflects light to a specific location on the sensor
- Which 3D point is mapped to which pixel?

The Camera

How do we get the point observations?

- *keypoints* and *features*
- SIFT, ORB, etc.
- **locally** distinct features

The Camera

Features identify points mapped from the 3D world to the 2D image.

Pinhole Camera Model

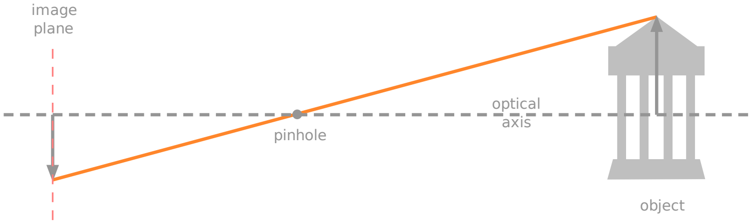

Figure 2: Light passing through a pinhole camera.

- f : effective focal length
- $\mathbf{r}_o = (x_o, y_o, z_o)$
- $\mathbf{r}_i = (x_i, y_i, f)$

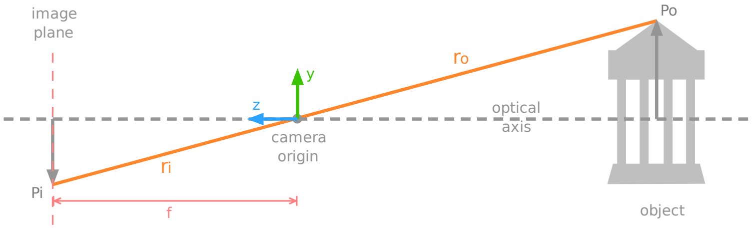

Figure 3: Camera at the origin.

Pinhole Camera Model

Using similar triangles, we get the equations of perspective projection.

$$\frac{\mathbf{r}_i}{f} = \frac{\mathbf{r}_o}{z_o} \quad \Rightarrow \quad \frac{x_i}{f} = \frac{x_o}{z_o}, \ \frac{y_i}{f} = \frac{y_o}{z_o}$$

Camera Parameters

Describe how a world point is mapped to a pixel coordinate.

Camera Parameters

Describe how a world point is mapped to a pixel coordinate.

$$x = PX$$

pixel coordinate — trans-formation — world coordinate

Figure 4: point mapping

Camera Parameters

We will describe this mapping in **homogeneous** coordinates.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Aside: Homogeneous Coordinates

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} \Rightarrow \begin{bmatrix} u/w \\ v/w \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} u/w \\ v/w \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix}$$

Coordinate Systems

We have to transform via a number of coordinate systems:

- The world coordinate system
- The camera coordinate system
- The image coordinate system
- The pixel coordinate system

World to Pixels

Figure 5: World to Pixels

World to Pixels

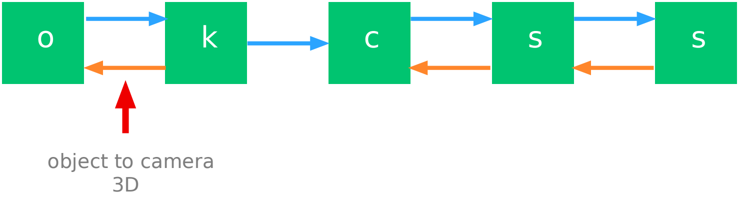

Figure 6: World to Camera coordinates

World to Pixels

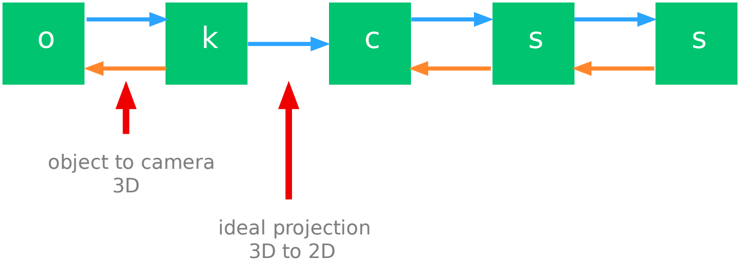

Figure 7: Projection to 2D

World to Pixels

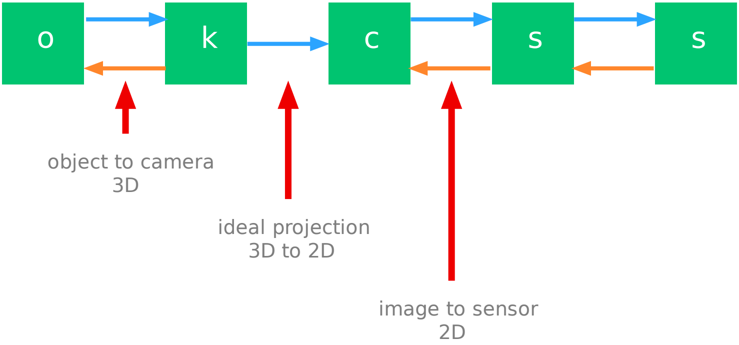

Figure 8: Convert to Sensor coordinates

World to Pixels

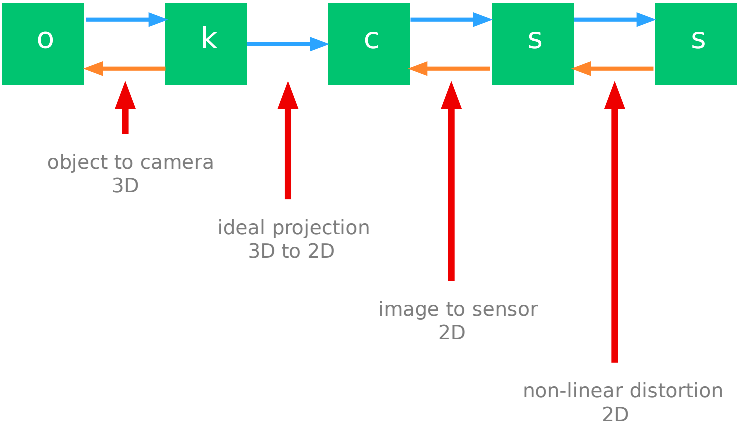

Figure 9: Lens Distortions

Camera Parameters

How do we work with these parameters?

- *extrinsic* parameters: the pose of the camera in the world
- *intrinsic* parameters: the properties of the camera

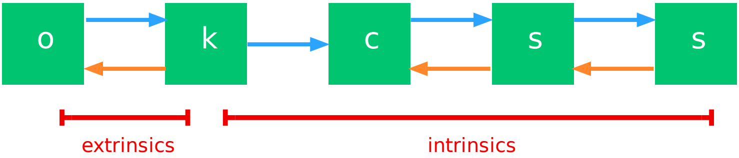

Figure 10: Camera Parameters

Extrinsic Parameters

The pose of the camera.

Extrinsic Parameters

- Describe the **pose** of the camera in the world.
- That is, the *position* and *heading* of the camera.
- Invertible transformation.

How many parameters do we need?

- 3 parameters for the position
- 3 parameters for the heading
- There are **6** *extrinsic* parameters.

Extrinsic Parameters

Point in world coordinates:

$$\mathbf{X}_p = [X_p, Y_p, Z_p]^T$$

Origin of camera in world coordinates:

$$\mathbf{X}_o = [X_o, Y_o, Z_o]^T$$

Transformation

Translation between origin of world and camera coordinates is:

$$\mathbf{X}_o = [X_o, Y_o, Z_o]^T$$

Rotation R from world to camera coordinates system is:

$$^k\mathbf{X}_p = R(\mathbf{X}_p - \mathbf{X}_o)$$

Homogeneous Coordinates

$$\begin{bmatrix} {}^k\mathbf{X}_p \\ 1 \end{bmatrix} = \begin{bmatrix} R & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} I_3 & -\mathbf{X}_o \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_p \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} R & -R\mathbf{X}_o \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_p \\ 1 \end{bmatrix}$$

or:

$${}^k\mathbf{X}_p = {}^kH\mathbf{X}_p, \quad \text{where} \quad {}^kH = \begin{bmatrix} R & -R\mathbf{X}_o \\ \mathbf{0}^T & 1 \end{bmatrix}$$

Intrinsic Parameters

Projecting points from the camera to the sensor.

Intrinsic Parameters

- projection from camera coordinates to sensor coordinates
- central projection is **not** invertible
- image plane to sensor is invertible
- linear deviations are invertible

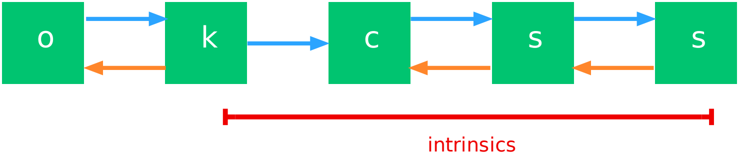

Figure 11: Camera Intrinsics

Recall for our pinhole model:

$$ {}^{c}x_p = c\frac{{}^{k}X_p}{{}^{k}Z_p} $$
$$ {}^{c}y_p = c\frac{{}^{k}Y_p}{{}^{k}Z_p} $$

where c is the focal length, or *camera constant*.

Homogeneous Coordinates

$$\begin{bmatrix} U \\ V \\ W \\ T \end{bmatrix} = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} {}^kX_p \\ {}^kY_p \\ {}^kZ_p \\ 1 \end{bmatrix}$$

Drop the 3rd row:

$$\begin{bmatrix} {}^cx_p \\ {}^cy_p \\ 1 \end{bmatrix} = \begin{bmatrix} {}^cu_p \\ {}^cv_p \\ {}^cw_p \end{bmatrix} = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} {}^kX_p \\ {}^kY_p \\ {}^kZ_p \\ 1 \end{bmatrix}$$

Ideal Camera

The mapping for an ideal camera is:

$$^{c}\mathsf{x} = {}^{c}P\mathsf{X}$$

with:

$$^{c}P = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & -R\mathbf{X}_o \\ \mathbf{0}^T & 1 \end{bmatrix}$$

Calibration Matrix

We can now define the *calibration matrix* for an **ideal** camera.

$$ {}^{c}K = \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix} $$

The mapping of a point in the world to the image plane is:

$$ {}^{c}P = {}^{c}KR[I_3 | -\mathbf{X}_o] $$

Linear Errors

The next step is mapping from the image plane to the sensor.

- Location of principal point in sensor coordinates.
- Scale difference in x and y, according to chip design.
- Shear compensation.

Location of Principal Point

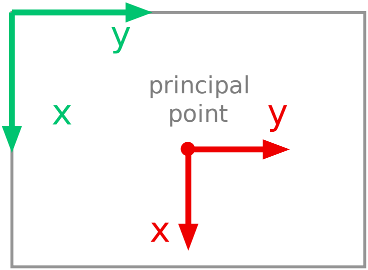

Figure 12: Principal Point

Origin of sensor space is not at the principal point:

$$
{}^{s}H_{c} = \begin{bmatrix} 1 & 0 & x_H \\ 0 & 1 & y_H \\ 0 & 0 & 1 \end{bmatrix}
$$

Compensation is a *translation*.

Scale and Shear

- Scale difference m in x and y.
- Sheer compensation s.

We need to add 4 additional parameters to our calibration matrix:

$$
{}^{s}H_c = \begin{bmatrix} 1 & s & x_H \\ 0 & 1+m & y_H \\ 0 & 0 & 1 \end{bmatrix}
$$

Calibration Matrix

Normally, we combine these compensations with the ideal calibration matrix:

$$K = \begin{bmatrix} 1 & s & x_H \\ 0 & 1+m & y_H \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} c & s & x_H \\ 0 & c(1+m) & y_H \\ 0 & 0 & 1 \end{bmatrix}$$

Calibration Matrix

$$K = \begin{bmatrix} c & s & x_H \\ 0 & c(1+m) & y_H \\ 0 & 0 & 1 \end{bmatrix}$$

There are **5** intrinsic parameters:

- camera constant c
- scale difference m
- principal point offset x_H and y_H
- shear compensation s

Projection Matrix

Finally, we have the 3×4 homogeneous projection matrix:

$$P = KR[I_3 | -\mathbf{X}_o]$$

It contains **11 parameters**:

- 6 extrinsic parameters
- 5 intrinsic parameters

Direct Linear Transformation

$$x = PX$$

pixel coordinate — trans-formation — world coordinate

Figure 13: point mapping

Control Points

Figure 14: known points in the world

We have *control points* of known coordinates in the world. We want to estimate the camera parameters, given these points.

Parameter Estimation

- **Goal**: camera parameters, P.
- **Given**: control points in the world, X.
- **Observed**: coordinates (x, y) in the image.

Mapping

Direct Linear Transformation (DLT) maps a point in the world to a point in the image.

$$
\begin{aligned}
x &= KR[I_3| - \mathbf{X}_o]\mathbf{X} \\
&= P\mathbf{X}
\end{aligned}
$$

Camera Parameters

$$x = KR[I_3| - \mathbf{X}_o]\mathbf{X} = P\mathbf{X}$$

- Intrinsic parameters K
- Extrinsic parameters $\mathbf{X}_o$ and R.
- Projection matrix P contains intrinsic **and** extrinsic parameters.

Direct Linear Transformation

Compute the **11** *intrinsic* and *extrinsic* parameters.

How many points are needed?

Homogeneous projection:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = P \begin{bmatrix} U \\ V \\ W \\ T \end{bmatrix}$$

Normalised homogeneous projection:

$$\begin{bmatrix} u/w \\ v/w \\ 1 \end{bmatrix} = P \begin{bmatrix} U/T \\ V/T \\ W/T \\ 1 \end{bmatrix}$$

Euclidean coordinates:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

We can expand the multiplication by P to get the following:

$$x = \frac{p_{11}X + p_{12}Y + p_{13}Z + p_{14}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}}$$

$$y = \frac{p_{21}X + p_{22}Y + p_{23}Z + p_{24}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}}$$

Each point gives **two** observation equations, one for each image coordinate.

How many points are needed?

Each point gives **two** observation equations, one for each image coordinate.

We need at least **6 points** to estimate *11 parameters*.

Rearrange the DLT Equation

$$\mathbf{x}_i = P\mathbf{X}_i$$

$$\mathbf{x}_i = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \mathbf{X}_i$$

$$\mathbf{x}_i = P\mathbf{X}_i = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \mathbf{X}_i$$

Define three vectors:

$$A = \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \end{bmatrix}, \quad B = \begin{bmatrix} p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \end{bmatrix}, \quad C = \begin{bmatrix} p_{31} \\ p_{32} \\ p_{33} \\ p_{34} \end{bmatrix}$$

$$\mathbf{x}_i = P\mathbf{X}_i = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \mathbf{X}_i$$

Rewrite the equation as:

$$\mathbf{x}_i = P\mathbf{X}_i = \begin{bmatrix} A^T \\ B^T \\ C^T \end{bmatrix} \mathbf{X}_i$$

$$\mathbf{x}_i = P\mathbf{X}_i = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \mathbf{X}_i$$

Rewrite the equation as:

$$\begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} = \mathbf{x}_i = P\mathbf{X}_i = \begin{bmatrix} A^T \\ B^T \\ C^T \end{bmatrix} \mathbf{X}_i = \begin{bmatrix} A^T X_i \\ B^T X_i \\ C^T X_i \end{bmatrix}$$

$$\mathbf{x}_i = \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \quad \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} = \begin{bmatrix} A^T X_i \\ B^T X_i \\ C^T X_i \end{bmatrix}$$

$$x_i = \frac{u_i}{w_i} = \frac{A^T X_i}{C^T X_i}, \quad y_i = \frac{v_i}{w_i} = \frac{B^T X_i}{C^T X_i}$$

System of equations

$$x_i = \frac{A^T X_i}{C^T X_i} \quad \Rightarrow \quad x_i C^T X_i - A^T X_i = 0$$

$$y_i = \frac{B^T X_i}{C^T X_i} \quad \Rightarrow \quad y_i C^T X_i - B^T X_i = 0$$

Leading to a system of linear equations in A, B, and C:

$$-X_i^T A + x_i X_i^T C = 0$$
$$-X_i^T B + y_i X_i^T C = 0$$

let:

$$\mathbf{p} = \begin{bmatrix} A \\ B \\ C \end{bmatrix} = vec(P^T) = \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \\ p_{34} \end{bmatrix}$$

$$
\begin{aligned}
-X_i^T A \qquad\qquad\quad &+x_i X_i^T C = 0 \\
- X_i^T B \quad &+y_i X_i^T C = 0
\end{aligned}
$$

rewrite as:

$$a_{x_i}^T \mathbf{p} = 0, \quad a_{y_i}^T \mathbf{p} = 0$$

with:

$$
\begin{aligned}
\mathbf{p} &= vec(P^T) \\
a_{x_i}^T &= (-X_i, -Y_i, -Z_i, -1, 0, 0, 0, 0, x_i X_i, x_i Y_i, x_i Z_i, x_i) \\
a_{y_i}^T &= (0, 0, 0, 0, -X_i, -Y_i, -Z_i, -1, y_i X_i, y_i Y_i, y_i Z_i, y_i)
\end{aligned}
$$

for each point we have:

$$a_{x_i}^T \mathbf{p} = 0, \quad a_{y_i}^T \mathbf{p} = 0$$

stacking all the points vertically:

$$\begin{bmatrix} a_{x_1}^T \\ a_{y_1}^T \\ a_{x_2}^T \\ a_{y_2}^T \\ \dots \\ a_{x_n}^T \\ a_{y_n}^T \end{bmatrix} \mathbf{p} = M\mathbf{p} \stackrel{!}{=} 0$$

Where M is a $2n \times 12$ matrix.

Solving the Linear System

Solving a system of linear equations of the form $Ax = 0$ is equivalent to finding the null space of A.

- Apply the Singular Value Decomposition (SVD) to solve $M\mathbf{p} = 0$.
- SVD returns a matrix U, S, and V such that $M = USV^T$.
- Choose **p** as the singular *vector* belonging to the singular *value* of 0.
- Solution is the last column of V.

Direct Linear Transformation

Does it always work?

Critical Surfaces

No solution if all points X_i are on a **plane**.

Decomposing the Projection Matrix

From P to K, R, $\mathbf{X}_o$

Decomposing the Projection Matrix

We have P, how do we obtain $K, R, \mathbf{X}_o$?

Structure of P:

$$P = [KR| - KR\mathbf{X}_o] = [H|\mathbf{h}]$$

with:

$$H = KR, \quad \mathbf{h} = -KR\mathbf{X}_o$$

Decomposing the Projection Matrix

$$H = KR, \quad \mathbf{h} = -KR\mathbf{X}_o$$

We can obtain the projection centre by:

$$\mathbf{X}_o = -H^{-1}\mathbf{h}$$

Decomposing the Projection Matrix

$$H = KR$$

What do we know about these matrices?

Decomposing the Projection Matrix

Exploit the structure of $H = KR$

- K is a triangular matrix
- R is a rotation matrix

There is a standard method to decompose a matrix to a rotation and triangular matrix.

- **QR** decomposition

Decomposing the Projection Matrix

We perform a QR decomposition on H^{-1}, given the order of rotation and triangular matrices.

$$H^{-1} = (KR)^{-1} = R^{-1}K^{-1} = R^{T}K^{-1}$$

Decomposing the Projection Matrix

The Matrix $H = KR$ is homogeneous, therefore so is K, so we must normalise.

$$K \leftarrow \frac{1}{K_{33}} K$$

DLT recap

1. Build the matrix M.
2. Solve using SVD; $M = U\ S\ V^T$, solution is last column of V.
3. If individual matrices are required, we can use QR decomposition.

Summary

- Camera Model
- Intrinsic and Extrinsic Parameters
- Direct Linear Transformation

reading:

- Forsyth, Ponce; Computer Vision: A modern approach. Section 1.3
- Hartley, Zisserman; Multiple View Geometry in Computer Vision