Lecture 03
Image Formation 2

Davide Scaramuzza
Lab Exercise 1 - Today afternoon

- Room ETH HG E 1.1 from 13:15 to 15:00
- Work description: implement an augmented reality wireframe cube
  - Practice the perspective projection
Goal of today’s lecture

- Study the algorithms behind robot-position control and augmented reality
Outline of this lecture

• Camera calibration
  – Non-linear algorithms: P3P and PnP for calibrated cameras
    • From general 3D objects
  – Linear algorithms (DLT) for uncalibrated cameras
    • From 3D objects
    • From planar grids

• Non conventional camera models
Pose determination from n Points (PnP) Problem

- Assumption: camera intrinsic parameters are known
- Given known 3D landmarks in the world frame and given their image correspondences in the camera frame, determine the 6DOF pose of the camera in the world frame (including the intrinsic parameters if uncalibrated)
How Many Points are Enough?

- **1 Point**: infinitely many solutions.
- **2 Points**: infinitely many solutions, but bounded.
- **3 Points**:
  - (no 3 collinear) finitely many solutions (up to 4).
- **4 Points**:
  - Unique solution
1 Point
2 Points
Inscribed Angles are Equal
3 Points

From Carnot’s Theorem:

\[ s_1^2 = L_B^2 + L_C^2 - 2L_B L_C \cos \theta_{BC} \]

\[ s_2^2 = L_A^2 + L_C^2 - 2L_A L_C \cos \theta_{AC} \]

\[ s_3^2 = L_A^2 + L_B^2 - 2L_A L_B \cos \theta_{AB} \]
Algebraic Approach: reduce to 4th order equation  
(Fischler and Bolles, 1981)

\[
\begin{align*}
    s_1^2 &= L_B^2 + L_C^2 - 2L_BL_C \cos \theta_{BC} \\
    s_2^2 &= L_A^2 + L_C^2 - 2L_AL_C \cos \theta_{AC} \\
    s_3^2 &= L_A^2 + L_B^2 - 2L_AL_B \cos \theta_{AB}
\end{align*}
\]

- It is known that \( n \) independent polynomial equations, in \( n \) unknowns, can have no more solutions than the product of their respective degrees. Thus, the system can have a maximum of 8 solutions. However, because every term in the system is either a constant or of second degree, for every real positive solution there is a negative solution.
- Thus, with 3 points, there are at most 4 valid (positive) solutions.
- A 4th point can be used to disambiguate the solutions.

By defining \( x = L_B/L_A \), it can be shown that the system can be reduced to a 4th order equation:

\[
G_0 + G_1x + G_2x^2 + G_3x^3 + G_4x^4 = 0
\]
Application to Monocular Visual Odometry: camera pose estimation from known 3D-2D correspondences
AR Application: Microsoft HoloLens
Outline of this lecture

• Camera calibration
  – Non-linear algorithms: P3P and PnP for calibrated cameras
    • From general 3D objects
  – Linear algorithms (DLT) for uncalibrated cameras
    • From 3D objects
    • From planar grids

• Non conventional camera models
Camera calibration

- Calibration is the process to determine the **intrinsic and extrinsic** parameters of the camera model.

- A method proposed in 1987 by Tsai consists of measuring the 3D position of \( n \geq 6 \) control points on a three-dimensional calibration target and the 2D coordinates of their projection in the image. This problem is also called “Resection”, or “Perspective from \( n \) Points”, or “Camera pose from 3D-to-2D correspondences”, and is one of the most widely used algorithms in Computer Vision and Robotics.

- Solution: The intrinsic and extrinsic parameters are computed directly from the perspective projection equation; let’s see how!

3D position of control points is assigned in a reference frame specified by the user.
Camera calibration: Direct Linear Transform (DLT)

Our goal is to compute K, R, and T that satisfy the perspective projection equation (we neglect the radial distortion)

\[
\tilde{p} = \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R | T] \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \Rightarrow
\]

\[
\Rightarrow \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
\]

\[
\Rightarrow \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} \alpha_u r_{11} + u_0 r_{31} & \alpha_u r_{12} + u_0 r_{32} & \alpha_u r_{13} + u_0 r_{33} & \alpha_u t_1 + u_0 t_3 \\ \alpha_v r_{21} + v_0 r_{31} & \alpha_v r_{22} + v_0 r_{32} & \alpha_v r_{23} + v_0 r_{33} & \alpha_v t_2 + v_0 t_3 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
\]
Camera calibration: Direct Linear Transform (DLT)

Our goal is to compute $K$, $R$, and $T$ that satisfy the perspective projection equation (we neglect the radial distortion)

$$\tilde{p} = \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R \vert T] \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$
Camera calibration: Direct Linear Transform (DLT)

Our goal is to compute $K$, $R$, and $T$ that satisfy the perspective projection equation (we neglect the radial distortion)

$$
\Rightarrow \begin{bmatrix}
\tilde{u} \\
\tilde{v} \\
\tilde{w}
\end{bmatrix} =
\begin{bmatrix}
m_{11} & m_{12} & m_{13} & m_{14} \\
m_{21} & m_{22} & m_{23} & m_{24} \\
m_{31} & m_{32} & m_{33} & m_{34}
\end{bmatrix}
\begin{bmatrix}
X_w \\
Y_w \\
Z_w \\
1
\end{bmatrix}
$$

$$
\Rightarrow \begin{bmatrix}
\tilde{u} \\
\tilde{v} \\
\tilde{w}
\end{bmatrix} = M \cdot \begin{bmatrix}
X_w \\
Y_w \\
Z_w \\
1
\end{bmatrix}
$$

$$
\Rightarrow \begin{bmatrix}
\tilde{u} \\
\tilde{v} \\
\tilde{w}
\end{bmatrix} = M^T \begin{bmatrix}
m_1^T \\
m_2^T \\
m_3^T
\end{bmatrix} \cdot \begin{bmatrix}
X_w \\
Y_w \\
Z_w \\
1
\end{bmatrix}
$$

where $m_i^T$ is the $i$-th row of $M$
Camera calibration: Direct Linear Transform (DLT)

\[
\begin{bmatrix}
\tilde{u} \\
\tilde{v} \\
\tilde{w}
\end{bmatrix} =
\begin{bmatrix}
m_1^T \\
m_2^T \\
m_3^T
\end{bmatrix}
\begin{bmatrix}
X_w \\
Y_w \\
Z_w \\
1
\end{bmatrix} \rightarrow P
\]

Conversion back from homogeneous coordinates to pixel coordinates leads to:

\[
\begin{align*}
u &= \frac{\tilde{u}}{\tilde{w}} = \frac{m_1^T \cdot P}{m_3^T \cdot P} \\
v &= \frac{\tilde{v}}{\tilde{w}} = \frac{m_2^T \cdot P}{m_3^T \cdot P} \\
\Rightarrow & \quad (m_1^T - u_i m_3^T) \cdot P_i = 0 \\
& \quad (m_2^T - v_i m_3^T) \cdot P_i = 0
\end{align*}
\]
Camera calibration: Direct Linear Transform (DLT)

By re-arranging the terms, we obtain

\[(m^T_1 - u_i m^T_3) \cdot P_i = 0 \Rightarrow \begin{pmatrix} P^T_1 & 0^T & -u_1 P^T_1 \\ 0^T & P^T_1 & -v_1 P^T_1 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\]

For \(n\) points, we can stack all these equations into a big matrix:

\[
\begin{pmatrix}
P^T_1 & 0^T & -u_1 P^T_1 \\
0^T & P^T_1 & -v_1 P^T_1 \\
\cdots & \cdots & \cdots \\
P^T_n & 0^T & -u_n P^T_n \\
0^T & P^T_n & -v_n P^T_n
\end{pmatrix}
\begin{pmatrix}
m_1 \\
m_2 \\
m_3
\end{pmatrix}
= 
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0
\end{pmatrix}
\]
Camera calibration: Direct Linear Transform (DLT)

By re-arranging the terms, we obtain

\[
\begin{align*}
(m_1^T - u_i m_3^T) \cdot P_i &= 0 \\
(m_2^T - v_i m_3^T) \cdot P_i &= 0
\end{align*}
\]

\[
\Rightarrow \begin{pmatrix} P_1^T & 0^T & -u_1 P_1^T \\ 0^T & P_1^T & -v_1 P_1^T \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\]

For \( n \) points, we can stack all these equations into a big matrix:

\[
\begin{pmatrix}
X_1^w & Y_1^w & Z_1^w & 1 & 0 & 0 & 0 & 0 & -u_1 X_1^w & -u_1 Y_1^w & -u_1 Z_1^w & -u_1 \\
0 & 0 & 0 & 0 & X_1^w & Y_1^w & Z_1^w & 1 & -v_1 X_1^w & -v_1 Y_1^w & -v_1 Z_1^w & -v_1 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
X_n^w & Y_n^w & Z_n^w & 1 & 0 & 0 & 0 & 0 & -u_n X_n^w & -u_n Y_n^w & -u_n Z_n^w & -u_n \\
0 & 0 & 0 & 0 & X_n^w & Y_n^w & Z_n^w & 1 & -v_n X_n^w & -v_n Y_n^w & -v_n Z_n^w & -v_n
\end{pmatrix}
\begin{pmatrix}
m_{11} \\
m_{12} \\
m_{13} \\
m_{14} \\
m_{21} \\
m_{22} \\
m_{23} \\
m_{24} \\
m_{31} \\
m_{32} \\
m_{33} \\
m_{34}
\end{pmatrix} = \begin{pmatrix}
0 \\
0 \\
0 \\
\cdots \\
0 \\
0 \\
0
\end{pmatrix}
\Rightarrow Q \cdot M = 0
\]
Camera calibration: Direct Linear Transform (DLT)

\[ \mathbf{Q} \cdot \mathbf{M} = 0 \]

Minimal solution

- \( \mathbf{Q}_{(2n \times 12)} \) should have rank 11 to have a unique (up to a scale) non-trivial solution \( \mathbf{M} \)
- Each 3D-to-2D point correspondence provides 2 independent equations
- Thus, \( 5 + \frac{1}{2} \) point correspondences are needed (in practice 6 point correspondences!)

Over-determined solution

- \( n \geq 6 \) points
- A solution is to minimize \( ||\mathbf{Q}\mathbf{M}||^2 \) subject to the constraint \( ||\mathbf{M}||^2 = 1 \).
  It can be solved through Singular Value Decomposition (SVD). The solution is the eigenvector corresponding to the smallest eigenvalue of the matrix \( \mathbf{Q}^T \mathbf{Q} \) (because it is the unit vector \( \mathbf{x} \) that minimizes \( ||\mathbf{Q}\mathbf{x}||^2 = \mathbf{x}^T \mathbf{Q}^T \mathbf{Q}\mathbf{x} \)).
- Matlab instructions:
  - \([\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathbf{Q});\)
  - \( \mathbf{M} = \mathbf{V}(\cdot, 12); \)
Degenerate configurations

1. Points lying on a plane and/or along a single line passing through the projection center

2. Camera and points on a twisted cubic (i.e., smooth curve in 3D space of degree 3)
Camera calibration: Direct Linear Transform (DLT)

- Once we have the M matrix, we can recover the intrinsic and extrinsic parameters by remembering that

\[ M = K(R \mid T) \]

\[
\begin{bmatrix}
  m_{11} & m_{12} & m_{13} & m_{14} \\
  m_{21} & m_{22} & m_{23} & m_{24} \\
  m_{31} & m_{32} & m_{33} & m_{34}
\end{bmatrix}
= \begin{bmatrix}
  \alpha_u & 0 & u_0 \\
  0 & \alpha_v & v_0 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  r_{11} & r_{12} & r_{13} & t_1 \\
  r_{21} & r_{22} & r_{23} & t_2 \\
  r_{31} & r_{32} & r_{33} & t_3
\end{bmatrix}
\]
Camera calibration: Direct Linear Transform (DLT)

• Once we have the M matrix, we can recover the intrinsic and extrinsic parameters by remembering that

\[ M = K(R \mid T) \]

\[
\begin{bmatrix}
  m_{11} & m_{12} & m_{13} & m_{14} \\
  m_{21} & m_{22} & m_{23} & m_{24} \\
  m_{31} & m_{32} & m_{33} & m_{34}
\end{bmatrix}
= \begin{bmatrix}
  \alpha r_{11} + u_0 r_{31} & \alpha r_{12} + u_0 r_{32} & \alpha r_{13} + u_0 r_{33} & \alpha t_1 + u_0 t_3 \\
  \alpha r_{21} + v_0 r_{31} & \alpha r_{22} + v_0 r_{32} & \alpha r_{23} + v_0 r_{33} & \alpha t_2 + v_0 t_3 \\
  r_{31} & r_{32} & r_{33} & t_3
\end{bmatrix}
\]

• However, notice that we are not enforcing the constraint that R is orthogonal, i.e., \( R \cdot R^T = I \)

• To do this, we can use the so-called QR factorization of \( M \), which decomposes \( M \) into a \( R \) (orthogonal), \( T \), and an upper triangular matrix (i.e., \( K \)
Tsai’s (1987) Calibration example

1. Edge detection
2. Straight line fitting to the detected edges
3. Intersecting the lines to obtain the image corners (corner accuracy < 0.1 pixels)
4. **Use more than 6 points** (ideally more than 20) and not all lying on a plane

Why is this ratio not 1?

What are the «skew» and «residuals»?

<table>
<thead>
<tr>
<th>$f_y$</th>
<th>$f_x/f_y$</th>
<th>skew</th>
<th>$x_0$</th>
<th>$y_0$</th>
<th>residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1673.3</td>
<td>1.0063</td>
<td>1.39</td>
<td>379.96</td>
<td>305.78</td>
<td>0.365</td>
</tr>
</tbody>
</table>
Tsai’s (1987) Calibration example

- The original Tsai calibration (1987) used to consider two different focal lengths $\alpha_u, \alpha_v$ (which means that the pixels are not squared) and a skew factor ($K_{12} \neq 0$, which means the pixels are parallelograms instead of rectangles) to account for possible misalignments (small $x, y$ rotation) between image plane and lens
- Most today’s cameras are well manufactured, thus, we can assume $\frac{\alpha_u}{\alpha_v} = 1$ and $K_{12} = 0$
- What is the residual? The residual is the average “reprojection error”. The reprojection error is computed as the distance (in pixels) between the observed pixel point and the camera-reprojected 3D point. The reprojection error gives as a quantitative measure of the accuracy of the calibration (ideally it should be zero).

<table>
<thead>
<tr>
<th>$f_y$</th>
<th>$f_x/f_y$</th>
<th>skew</th>
<th>$x_0$</th>
<th>$y_0$</th>
<th>residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1673.3</td>
<td>1.0063</td>
<td>1.39</td>
<td>379.96</td>
<td>305.78</td>
<td>0.365</td>
</tr>
</tbody>
</table>
In this case, the camera has been pre-calibrated (i.e., K is known). Can you think of how the DLT algorithm could be modified so that only R and T need to be determined and not K?
Outline of this lecture

• Camera calibration
  – Non-linear algorithms: P3P and PnP for calibrated cameras
    • From general 3D objects
  – Linear algorithms (DLT) for uncalibrated cameras
    • From 3D objects
    • From planar grids

• Non conventional camera models
Camera calibration from planar grids: homographies

- Tsai calibration is based on DLT algorithm, which requires points not to lie on the same plane
- An alternative method (today’s standar camera calibration method) consists of using a planar grid (e.g., a chessboard) and a few images of it shown at different orientations
- This method was invented by Zhang (1999) @Microsoft Research
Our goal is to compute $K$, $R$, and $T$, that satisfy the perspective projection equation (we neglect the radial distortion).

Since the points lie on a plane, we have $Z_w = 0$

$$\tilde{p} = \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R | T] \cdot \begin{bmatrix} X_w \\ Y_w \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix}$$
Our goal is to compute $K$, $R$, and $T$, that satisfy the perspective projection equation (we neglect the radial distortion).

Since the points lie on a plane, we have $Z_w = 0$.

\[
\begin{bmatrix}
\tilde{u} \\
\tilde{v} \\
\tilde{w}
\end{bmatrix} =
\begin{bmatrix}
h_{11} & h_{12} & h_{13} \\
h_{21} & h_{22} & h_{23} \\
h_{31} & h_{32} & h_{33}
\end{bmatrix}
\begin{bmatrix}
X_w \\
Y_w \\
1
\end{bmatrix}
\]

\[
\Rightarrow
\begin{bmatrix}
\tilde{u} \\
\tilde{v} \\
\tilde{w}
\end{bmatrix} = H \cdot
\begin{bmatrix}
X_w \\
Y_w \\
1
\end{bmatrix}
\]

This matrix is called Homography.

\[
\begin{bmatrix}
\tilde{u} \\
\tilde{v} \\
\tilde{w}
\end{bmatrix} =
\begin{bmatrix}
h_1^T \\
h_2^T \\
h_3^T
\end{bmatrix}
\begin{bmatrix}
X_w \\
Y_w \\
1
\end{bmatrix}
\]

where $h_i^T$ is the i-th row of $H$. 

Camera calibration from planar grids: homographies.
Camera calibration from planar grids: homographies

\[
\begin{bmatrix}
\tilde{u} \\
\tilde{v} \\
\tilde{w}
\end{bmatrix} = \begin{bmatrix}
h_1^T \\
h_2^T \\
h_3^T
\end{bmatrix} \cdot \begin{bmatrix}
X_w \\
Y_w \\
1
\end{bmatrix}
\]

Conversion back from homogeneous coordinates to pixel coordinates leads to:

\[
\begin{align*}
u &= \frac{\tilde{u}}{\tilde{w}} = \frac{h_1^T \cdot P}{h_3^T \cdot P} \\
v &= \frac{\tilde{v}}{\tilde{w}} = \frac{h_2^T \cdot P}{h_3^T \cdot P}
\end{align*}
\]

\[
(h_1^T - u_i h_3^T) \cdot P_i = 0 \\
(h_2^T - v_i h_3^T) \cdot P_i = 0
\]

where \( P = (X_w, Y_w, 1)^T \)
Camera calibration from planar grids: homographies

By re-arranging the terms, we obtain

\[(h_1^T - u_i h_3^T) \cdot P_i = 0 \quad \Rightarrow \quad P_i^T \cdot h_1 + 0 \cdot h_2^T - u_i P_i^T \cdot h_3^T = 0\]

\[(h_2^T - v_i h_3^T) \cdot P_i = 0 \quad \Rightarrow \quad 0 \cdot h_1^T + P_i^T \cdot h_2^T - v_i P_i^T \cdot h_3^T = 0\]

For \(n\) points, we can stack all these equations into a big matrix:

\[
\begin{pmatrix}
    P_1^T & 0^T & -u_1 P_1^T \\
    0^T & P_1^T & -v_1 P_1^T \\
    \cdots & \cdots & \cdots \\
    P_n^T & 0^T & -u_n P_n^T \\
    0^T & P_n^T & -v_n P_n^T
\end{pmatrix}
\begin{pmatrix}
    h_1 \\
    h_2 \\
    \cdots \\
    h_3
\end{pmatrix}
= \begin{pmatrix}
    0 \\
    0 \\
    \vdots \\
    0
\end{pmatrix}
\Rightarrow Q \cdot H = 0
\]

Q (this matrix is known)  H (this matrix is unknown)
Camera calibration from planar grids: homographies

\[ Q \cdot H = 0 \]

**Minimal solution**
- \( Q_{(2n \times 9)} \) should have rank 8 to have a unique (up to a scale) non-trivial solution \( H \)
- Each point correspondence provides 2 independent equations
- Thus, a minimum of 4 non-collinear points is required

**Over-determined solution**
- \( n \geq 4 \) points
- It can be solved through Singular Value Decomposition (SVD) (same considerations as the case before apply)

**Solving for K, R and T**
- \( H \) can be decomposed by recalling that

\[
\begin{bmatrix}
  h_{11} & h_{12} & h_{13} \\
  h_{21} & h_{22} & h_{23} \\
  h_{31} & h_{32} & h_{33}
\end{bmatrix}
= \begin{bmatrix}
  \alpha_u & 0 & u_0 \\
  0 & \alpha_v & v_0 \\
  0 & 0 & 1
\end{bmatrix}
\cdot
\begin{bmatrix}
  r_{11} & r_{12} & t_1 \\
  r_{21} & r_{22} & t_2 \\
  r_{31} & r_{32} & t_3
\end{bmatrix}
\]
### Types of 2D Transformations

<table>
<thead>
<tr>
<th>Name</th>
<th>Matrix</th>
<th># D.O.F.</th>
<th>Preserves:</th>
<th>Icon</th>
</tr>
</thead>
<tbody>
<tr>
<td>translation</td>
<td>$\begin{bmatrix} I &amp; t \end{bmatrix}_{2\times3}$</td>
<td>2</td>
<td>orientation $+$ $\cdots$</td>
<td></td>
</tr>
<tr>
<td>rigid (Euclidean)</td>
<td>$\begin{bmatrix} R &amp; t \end{bmatrix}_{2\times3}$</td>
<td>3</td>
<td>lengths $+$ $\cdots$</td>
<td></td>
</tr>
<tr>
<td>similarity</td>
<td>$\begin{bmatrix} sR &amp; t \end{bmatrix}_{2\times3}$</td>
<td>4</td>
<td>angles $+$ $\cdots$</td>
<td></td>
</tr>
<tr>
<td>affine</td>
<td>$\begin{bmatrix} A \end{bmatrix}_{2\times3}$</td>
<td>6</td>
<td>parallelism $+$ $\cdots$</td>
<td></td>
</tr>
<tr>
<td>projective</td>
<td>$\begin{bmatrix} \tilde{H} \end{bmatrix}_{3\times3}$</td>
<td>8</td>
<td>straight lines</td>
<td></td>
</tr>
</tbody>
</table>

This transformation is called Homography.
Camera calibration from planar grids: homographies

- Demo of Camera Calibration Toolbox for Matlab (world’s standard toolbox for calibrating perspective cameras):
  http://www.vision.caltech.edu/bouguetj/calib_doc/
Application of calibration from planar grids

• Today, there are thousands of application of this algorithm:
  – Augmented reality

AR Tags: http://april.eecs.umich.edu/wiki/index.php/April_Tags
Application of calibration from planar grids

• Today, there are thousands of application of this algorithm:
  – Augmented reality
  – Robotics (beacon-based localization)
• Do we need to know the metric size of the tag?
  – For Augmented Reality?
  – For Robotics?

AR Tags: [http://april.eecs.umich.edu/wiki/index.php/April_Tags](http://april.eecs.umich.edu/wiki/index.php/April_Tags)
If the camera is calibrated, only R and T need to be determined. In this case, should we use DLT (linear system of equations) or PnP (non linear)?

**DLT vs PnP: Accuracy vs noise**

Lepetit, Moreno Noguer, Fua, EPnP: An Accurate $O(n)$ Solution to the PnP Problem, IJCV’09
DLT vs PnP: Accuracy vs number of points

If the camera is calibrated, only R and T need to be determined. In this case, should we use DLT (linear system of equations) or PnP (non linear)?

Lepetit, Moreno Noguer, Fua, EPnP: An Accurate $O(n)$ Solution to the PnP Problem, IJCV’09
DLT vs PnP: Timing

Lepetit, Moreno Noguer, Fua, EPnP: An Accurate $O(n)$ Solution to the PnP Problem, IJCV’09
Outline of this lecture

• Camera calibration
  – From 3D objects
  – From planar grids

• Non conventional camera models
Omnidirectional Cameras

Rome, St. Peter’s square
Overview on Omnidirectional Cameras

Omnidirectional sensors come in many varieties, but by definition must have a wide field-of-view.

- **~180° FOV**
  - Wide FOV dioptric cameras (e.g. fisheye)

- **>180° FOV**
  - Catadioptric cameras (e.g. cameras and mirror systems)

- **~360° FOV**
  - Polydioptric cameras (e.g. multiple overlapping cameras)
Catadioptric Cameras
Dioptric Cameras (fisheye)

- **Nikon Coolpix FC-E9 Lens**
  - 360° × 183°

- **Canon EOS-1 Sigma Lens**
  - 360° × 180°
Example:

Same scene viewed by three different camera models:

- Perspective
- Fisheye
- Catadioptric

[http://rpg.ifi.uzh.ch/fov.html](http://rpg.ifi.uzh.ch/fov.html)
Z. Zhang et al. (RPG), Benefit of Large Field-of-View Cameras for Visual Odometry, ICRA 2016
Applications
Applications

• Daimler, Bosch: for car driving assistance systems
Applications

- Daimler, Bosch: for car driving assistance systems
- Meteorology: for sky observation
Applications

- Daimler, Bosch: for car driving assistance systems
- Meteorology: for sky observation
- Endoscopic Imagery: distortion removal (for the surgeon)

(Courtesy of Endo Tools Therapeutics, Brussels)
Applications

- Daimler, Bosch: for car driving assistance systems
- Meteorology: for sky observation
- Endoscopic Imagery: distortion removal (for the surgeon)
- RoboCup domain
Applications

- Daimler, Bosch: for car driving assistance systems
- Meteorology: for sky observation
- Endoscopic Imagery: distortion removal (for the surgeon)
- RoboCup domain
- Google Street View
Catadioptric cameras

- mirror
- perspective camera
Catadioptric cameras

Central catadioptric cameras

- mirror (surface of revolution of a conic)
- camera
- single effective viewpoint
Catadioptric cameras

- hyperbola + perspective camera
- parabola + orthographic lens
Why is it important that the camera be central (i.e., have a single effective viewpoint)?

- We can unwrap parts or all omnidirectional image into a perspective one.
Why is it important that the camera be central (i.e., have a single effective viewpoint)?

- We can unwrap parts or all omnidirectional image into a perspective one
- We can transform image points normalized vectors in the unit sphere
- We can apply standard algorithms valid for perspective geometry.
Omnidirectional camera calibration toolbox for Matlab (Scaramuzza, 2006)

- World’s standard toolbox for calibrating omnidirectional cameras (used at NASA, Daimler, IDS, Volkswagen, Audi, VW, Volvo, ...)

- Main applications are in robotics, endoscopy, video-surveillance, sky observation, automotive (Audi, VW, Volvo, ...)

https://sites.google.com/site/scarabotix/ocamcalib-toolbox
Equivalence between Perspective and Omnidirectional model
Equivalence between Perspective and Omnidirectional model

Measures the ray direction (angles).
Equivalence between Perspective and Omnidirectional model: the Spherical Model

Measures the ray direction (angles).
Representation of image points on the unit sphere

Always possible after the camera has been calibrated!
Summary (things to remember)

- P3P and PnP problems
- DLT algorithm
- Calibration from planar grid (Homography algorithm)
- Readings: Chapter 2.1 of Szeliski book

- Omnidirectional cameras
  - Central and non central projection
  - Dioptric
    - Catadioptric (working principle of conic mirrors)
- Unified (spherical) model for perspective and omnidirectional cameras
- Reading: Chapter 4 of Autonomous Mobile Robots book