Structure from Motion

Prof. Marco Marcon
Summing-up

• Stereo is the most powerful clue for determining the structure of a scene
• Another important clue is the relative motion between the scene and (mono) camera
• If scene stationary, camera moving → passive navigation
• If camera stationary, scene/objects moving → object-based (requires segmentation)
• Determining the 3-D information from the displacements of points at different time instants is called “structure from motion”
3-D Motion Models

- According to classical kinematics, 3-D motion is classified into
  - Rigid (body) motion: $X' = RX + T$
  - Non-rigid (deformable) motion: $X' = (D + R)X + T$

where $R$: rotation, $D$: deformation matrices, $T$: translation vector

- Deformation matrix does not have any constraint

- Rotation is represented using an orthonormal matrix ($RTR=I$) with various representations:
  - Euler angles
  - Euler angles with small angle approximations
  - Rotation around an arbitrary axis
  - Unit Quaternions
Modeling 3-D Velocity

- Note that $X' = RX + T$ gives a relation to represent “displacements” between two points at two time instants.
- In order to find instantaneous velocity one should find the displacement as time difference goes to zero.
- Consider the relation $X' = RX + T$ for small angles.

\[
\begin{align*}
X' &= \begin{bmatrix} 1 & -\phi & \psi \\ \phi & 1 & -\theta \\ -\psi & \theta & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} \\
Y' &= \begin{bmatrix} 0 & -\phi & \psi \\ \phi & 0 & -\theta \\ -\psi & \theta & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}
\end{align*}
\]

- Dividing both sides by $t$ (time difference) and taking limit $t \to 0$

\[
\begin{align*}
\dot{X} &= \begin{bmatrix} 0 & -\dot{\phi} & \dot{\psi} \\ \dot{\phi} & 0 & -\dot{\theta} \\ -\dot{\psi} & \dot{\theta} & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} V_X \\ V_Y \\ V_Z \end{bmatrix} \\
\dot{Y} &= \begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = \Omega \times \underline{X} + \underline{V} \\
\dot{Z} &= \Omega \times \underline{X} + \underline{V}
\end{align*}
\]
3-D Motion & Structure Estimation

- 3-D motion estimation is equivalent to “relative orientation” problem in photogrammetry
- Finding structure is relatively easy after determining motion
- There are three main approaches to 3-D motion/structure estimation problem
  - Optical flow-based methods
    - Relates the “instantaneous velocity” to optical flow (2-D motion) field to find a solution
  - Direct approaches
    - Relates spatio-temporal gradients to 3-D motion parameters
  - Point correspondence-based approaches
    - Requires initial point correspondences (conjugate pairs) in order to solve the unknown parameters
Optical Flow-based Methods

- Recall 3-D velocity vector for small angular rotation:

\[
\begin{bmatrix}
\dot{X} \\
\dot{Y} \\
\dot{Z}
\end{bmatrix} =
\begin{bmatrix}
0 & -\Phi & \Psi \\
\Phi & 0 & -\Theta \\
-\Psi & \Theta & 0
\end{bmatrix}
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}
+ \begin{bmatrix}
V_x \\
V_y \\
V_z
\end{bmatrix}
\Rightarrow
\dot{X} = \Omega \times \dot{X} + V
\]

\[
\Omega = \begin{bmatrix}
\Theta & \Psi & \Phi \\
\Psi & \Phi & \Theta \\
\Phi & \Theta & \Psi
\end{bmatrix}
\]

- In order to relate 3-D velocity to that of 2-D, use projection:

\[
u = \dot{x} = \frac{dx}{dt} = \frac{d(f \frac{X}{Z})}{dt} = f \frac{Z\dot{X} - X\dot{Z}}{Z^2} = f \frac{\dot{X}}{Z} - x \frac{\dot{Z}}{Z}
\]

\[
u = \dot{y} = \frac{dy}{dt} = \frac{d(f \frac{Y}{Z})}{dt} = f \frac{Z\dot{Y} - Y\dot{Z}}{Z^2} = f \frac{\dot{Y}}{Z} - y \frac{\dot{Z}}{Z}
\]
Optical Flow-based Methods

- Substituting 3-D velocity relation (in terms of angles):

\[
\begin{align*}
    u &= f \left( \frac{V_x}{Z} + \frac{\Phi y}{f} \right) - \frac{V_z}{Z} x - \frac{\Theta}{f} xy + \frac{\Psi}{f} x^2 \\
    v &= f \left( \frac{V_y}{Z} - \frac{\Theta}{f} \right) + \frac{\Phi x}{f} - \frac{V_z}{Z} y + \frac{\Psi}{f} xy - \frac{\Theta}{f} x^2 \\

    \Rightarrow \quad u &= \frac{-V_x + xV_z}{Z} + \frac{\Theta}{f} xy - \frac{\Psi}{f} (1 + x^2) + \frac{\Phi y}{f} \\
    \Rightarrow \quad v &= \frac{-V_y + yV_z}{Z} + \frac{\Theta}{f} (1 + y^2) - \frac{\Psi}{f} xy - \frac{\Phi x}{f}
\end{align*}
\]

\[
\begin{align*}
    x &= f \left( \frac{X}{Z} \right) \\
    y &= f \left( \frac{Y}{Z} \right)
\end{align*}
\]
Optical Flow-based methods

- Pure Translational Case: The integral below is minimized by differentiating with respect to Z first, then \((V_x, V_y, V_z)\):

\[
\iint \left( u - \frac{-V_x + xV_z}{Z} \right)^2 + \left( v - \frac{-V_y + yV_z}{Z} \right)^2 \, dx \, dy
\]

- Pure Rotational Case: The integral below is minimized by differentiating with respect to three angular velocities:

\[
\iint \left( u - \dot{\Theta}xy - \dot{\Psi}(1 + x^2) + \dot{\Phi}y \right)^2 + \left( v - \dot{\Theta}(1 + y^2) - \dot{\Psi}xy - \dot{\Phi}x \right)^2 \, dx \, dy
\]
Optical Flow-based Methods

\[
\begin{align*}
    u &= \frac{-V_x + xV_z}{Z} + \frac{\dot{\Theta}xy - \dot{\Psi}(1 + x^2)}{rotational} + \dot{\Phi}y \\
    v &= \frac{-V_y + yV_z}{Z} + \frac{\dot{\Theta}(1 + y^2) - \dot{\Psi}xy - \dot{\Phi}x}{rotational}
\end{align*}
\]

- First eliminate \( Z \) from the two equations above to obtain a single equation.
- After defining new variables \( e_1 \) and \( e_2 \), convert this equation into a linear (matrix) equation:
  \[
  \begin{bmatrix}
    v \\
    u \\
    -x \\
    -y \\
    -xy \\
    (x^2 + y^2) \\
    (1 + y^2) \\
    (1 + x^2)
  \end{bmatrix} H = uy - vx
  \]
  
- Where
  \[
  H \equiv \begin{bmatrix}
    e_1 \\
    e_2 \\
    \dot{\Theta} + \dot{\Phi}e_1 \\
    \dot{\Psi} + \dot{\Phi}e_2 \\
    \dot{\Theta}e_2 + \dot{\Psi}e_1 \\
    \dot{\Phi} \\
    \dot{\Theta}e_1 \\
    \dot{\Psi}e_2
  \end{bmatrix}^T
  \]
Optical flow solution

\[
\begin{bmatrix}
-v & u & -x & -y & -xy & (x^2 + y^2) & (1 + y^2) & (1 + x^2)
\end{bmatrix}H = uy - vx
\]

\[
H = \begin{bmatrix}
e_1 & e_2 & \dot{\Theta} + \dot{\Phi} e_1 & \dot{\Psi} + \dot{\Phi} e_2 & \dot{\Theta} e_2 + \dot{\Psi} e_1 & \dot{\Phi} & \dot{\Theta} e_1 & \dot{\Psi} e_2
\end{bmatrix}^T
\]

\[
e_1 \equiv \frac{V_x}{V_z} \quad e_2 \equiv \frac{V_y}{V_z}
\]

- The solution steps of the previous equation are:
  - Solve it using at least 8 image point correspondences
  - Recover 5 motion parameters from the elements of \( H \)
  - Solve for depth using one of the two equations above
Direct Methods

- Direct methods utilize only the spatio-temporal image intensity gradients to estimate 3-D motion and structure parameters.
- Note that almost all optical flow-based methods can be extended as direct methods by replacing optical flow motion vectors with their estimates in terms of spatio-temporal gradients:

\[
J \equiv \sum_{(x,y) \in \mathcal{I}} (E_x u + E_y v + E_t)^2 \quad \Rightarrow \quad \sum_{(x,y) \in \mathcal{I}} (E_x u + E_y v + E_t)E_x = 0
\]

\[
\frac{\partial J}{\partial u} = 0 \quad \Rightarrow \quad \sum_{(x,y) \in \mathcal{I}} (E_x u + E_y v + E_t)E_x = 0
\]

\[
\frac{\partial J}{\partial v} = 0 \quad \Rightarrow \quad \sum_{(x,y) \in \mathcal{I}} (E_x u + E_y v + E_t)E_y = 0
\]

\[
\Rightarrow \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum_{(x,y) \in \mathcal{I}} E_x E_x & \sum_{(x,y) \in \mathcal{I}} E_y E_x \\ \sum_{(x,y) \in \mathcal{I}} E_x E_y & \sum_{(x,y) \in \mathcal{I}} E_y E_y \end{bmatrix}^{-1} \begin{bmatrix} \sum_{(x,y) \in \mathcal{I}} E_t E_x \\ - \sum_{(x,y) \in \mathcal{I}} E_t E_y \end{bmatrix}
\]
Point Correspondence-based Methods

- Point correspondence-based algorithms is similar to relative orientation methods in photogrammetry.
- Using conjugate pairs, displacement (not velocity) between frames is determined in terms of a rotation matrix and a translation vector $[X'=RX+T, (X,Y,Z)\rightarrow(x,y), (X',Y',Z') \rightarrow(x',y')]$.

\[
x' = \frac{r_{11}x + r_{12}y + r_{13} + \frac{t_1}{Z}}{r_{31}x + r_{32}y + r_{33} + \frac{t_3}{Z}} \quad y' = \frac{r_{21}x + r_{22}y + r_{23} + \frac{t_2}{Z}}{r_{31}x + r_{32}y + r_{33} + \frac{t_3}{Z}}
\]

- Solution of this equation is obtained for two different cases:
  - Planar surface: $aX+bY+cZ=1$
  - General surface
- [Note that “scaling ambiguity” exits as in relative orientation]
Point Correspondence-based Methods

- Planar Surfaces: Assume that the surface is approximated by a number of planar patches (similar to wireframe modeling)

\[
\begin{align*}
If \quad aX + bY + cZ = 1 & \Rightarrow \\
& \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \\
& \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = AX \\
A & \equiv R + T[abc] \\
\Rightarrow \quad x' = \frac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + 1} \quad y' = \frac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + 1} \\
(\text{let} \quad a_9 = 1)
\end{align*}
\]

- These set of parameters are called pure parameters and solved using at least 4 point correspondences

- Rotation, translation values, as well as plane normal, are obtained from vector \(a\) using Singular Value Decomposition of ‘\(a\)’ in a linear fashion
Point Correspondence-based Methods: the planar patches

- The definition of $A$ requires, at least, 4 point correspondences

$$\begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -xx' & -yx' \\ 0 & 0 & 0 & x & y & 1 & -xy' & -yy' \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix}$$
Point Correspondence-based Methods: general surfaces

- General Surfaces: There are two well-known linear methods for determining motion and structure of arbitrary 3-D surfaces:
  - Essential (E)-matrix method: for calibrated cameras
    - Original formulation
    - Robust solution
  - Fundamental (F)-matrix method: for uncalibrated cameras
Structure from Motion
Single view Metrology

- Is it possible to extract 3D geometric information from single images?

YES

- How?
- Why?
Background

- 2D
- 3D

Optical centre

Real or imaginary object

Painting

Real object

Photograph

Arc

Architect, Descriptive Geometry

Drawing

Projective Geometry

Reconstructed 3D model

Flat image

Painter, Linear perspective

Laws of Optics

Descriptive Geometry

Perspective Optics

3D model
Introduction

- 3D affine measurements may be measured from a single perspective image
Introduction

1. Measurements of the distance between any of the planes
2. Measurements on these planes
3. Determine the camera’s position

- Results are sufficient for a partial or complete 3D reconstruction of the observed scene
La Flagellazione di Cristo
Geometry

- Overview
- Measurements between parallel lines
- Measurements on parallel planes
- Determining the camera position
Geometry Overview

- Possible to obtain geometric interpretations for key features in a scene
- Derive how 3D affine measurements may be extracted from the image
- Use results to analyze and/or model the scene
Assumptions

- Assume that images are obtained by perspective projection
- Assume that, from the image, a:
  - **vanishing line** of a *reference plane*
  - **vanishing point** of another *reference direction*
  may be determined from the image
Geometric Cues

- **Vanishing Line \( \ell \)**
  - Projection of the line at infinity of the reference plane into the image
Geometric Cues

- Vanishing Point(s) $v$
  - A point at infinity in the reference direction
  - *Reference direction is NOT parallel to reference plane*
  - Also known as the vertical vanishing point
Geometric Cues

Vanishing point

Vertical vanishing point (at infinity)

Vanishing line

Vanishing point

Vanishing point
Generic Algorithm

1) Edge detection and straight line fitting to obtain the set of straight edge segments $S_A$

2) Repeat
   a) Randomly select two segments $s_1, s_2 \in S_A$ and intersect them to give point $p$
   b) The support set $S_p$ is the set of straight edges in $S_A$ going through point $p$

3) Set the dominant vanishing point as the point $p$ with the largest support $S_p$

4) Remove all edges in $S_p$ from $S_A$ and repeat step 2
Automatic estimation of vanishing points and lines

RANSAC algorithm

Candidate vanishing point
Automatic estimation of vanishing points and lines
Measurements between Parallel Lines

- Wish to measure the distance between two parallel planes, in the reference direction
  - The aim is to compute the height of an object relative to a reference
Cross Ratio

- Point $b$ on plane $\pi$ correspond to point $t$ on plane $\pi'$
- Aligned to vanishing point $v$
- Point $i$ is the intersection with the vanishing line
Cross Ratio

- The **cross ratio** is between the points provides an affine length ratio
  - The value of the cross ratio determines a ratio of distances between planes in the world
- Thus, if we know the length for an object in the scene, we can use it as a reference to calculate the length of other objects
Estimating Height

• The distance $|| t_r - b_r ||$ is known
• Used to estimate the height of the man in the scene
Measurements on Parallel Planes

If the reference plane is affine calibrated, then from the image measurements the following can be computed:

i. Ratios of lengths of parallel line segments on the plane

ii. Ratios of areas on the plane
Parallel Line Segments

• Basis points are manually selected and measured in the real world
• Using ratios of lengths, the size of the windows are calculated
Planar Homology

- Using the same principals, affine measurements can be made on two separate planes, so long as the planes are parallel to each other.

- A map in the world between parallel planes induces a map between images of points on the two planes.
Homology Mapping between Parallel Planes

A point $X$ on plane $\pi$ is mapped into the point $X'$ on $\pi'$ by a parallel projection.
Planar Homology

Points in one plane are mapped into the corresponding points in the other plane as follows:

\[ X' = HX \]

where (in homogeneous coordinates):

- \( X \) is an image point
- \( X' \) is its corresponding point
- \( H \) is the 3 x 3 matrix representing the homography transformation
Measurements on Parallel Planes

- This means that we can compare measurements made on two separate planes by mapping between the planes in the reference direction via the homology.
Parallel Line Segments lying on two Parallel Planes
Camera Position

- Using the techniques we developed in the previous sections, we can:
  - Determine the distance of the camera from the scene
  - Determine the height of the camera relative to the reference plane
Camera Distance from Scene

- In *Measurements between Parallel Lines*, distances between planes are computed as a ratio relative to the camera’s distance from the reference plane.
- Thus we can compute the camera’s distance from a particular frame knowing a single reference distance.
Camera Position Relative to Reference Plane

The location of the camera relative to the reference plane is the back-projection of the vanishing point onto the reference plane.
Algebraic Representation

- Overview
- Measurements between parallel lines
- Measurements on parallel planes
- Determining the camera position
Overview

- Algebraic approach offers many advantages (over direct geometry):
  1. Avoid potential problems with ordering for the cross ratio
  2. Minimal and over-constrained configurations can be dealt with uniformly
  3. Unifies the different types of measurements
  4. Are able to develop an uncertainty analysis
Coordinate Systems

- Define an affine coordinate system XYZ in space
  - Origin lies on reference plane
  - X, Y axes span the reference plane
  - Z axis is the reference direction
- Define image coordinate system xy
  - y in the vertical direction
  - x in the horizontal direction
Coordinate Systems
Projection Matrix

- If $\mathbf{X}'$ is a point in world space, it is projected to an image point $\mathbf{x}'$ in image space via a $3 \times 4$ projection matrix $\mathbf{P}$

$$\mathbf{x}' = \mathbf{P}\mathbf{X}' = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 \end{bmatrix} \mathbf{X}'$$

where $\mathbf{x}'$ and $\mathbf{X}'$ are homogeneous vectors:

$\mathbf{x}' = (x, y, w)$ and $\mathbf{X}' = (X, Y, Z, W)$
Vanishing Points

- Denote the vanishing points for the X, Y and Z directions as $v_X$, $v_Y$, and $v$
- By inspection, the first 3 columns of matrix $P$ are the vanishing points:
  - $p_1 = v_X$
  - $p_2 = v_Y$
  - $p_3 = v$
- Origin of the world coordinate system is $p_4$
Vanishing Line

- Furthermore, $v_X$ and $v_Y$ are on the vanishing line $1$
  - Choosing these points fixes the X and Y affine coordinate axes
  - Denote them as $1_{1}^\top, 1_{2}^\top$ where $1_{i}^\top \cdot 1 = 0$
- Note:
  - Columns 1, 2 and 4 make up the reference plane to image homography matrix $H$
Projection Matrix Reduction

- \( o = p_4 = \frac{1}{||1||} = 1^\perp \)
  - \( o \) is the Origin of the coordinate system

- Thus, the parametrization of \( P \) is:
  \[
P = \begin{bmatrix} 1_1^\perp & 1_2^\perp & \alpha v & 1^\perp \end{bmatrix}
  \]

\( \alpha \) is the affine scale factor
Measurements between Parallel Lines

- The aim is to compute the height of an object relative to a reference

- Height is measured in the Z direction
Measurements between Parallel Lines

- Base point B on the reference plane
- Top point T in the scene

\[
b = n(Xp_1 + Yp_2 + p_4)
\]
\[
t = m(Xp_1 + Yp_2 + Zp_3 + p_4)
\]

\(n\) and \(m\) are unknown scale factors
Affine Scale Factor

- If $\alpha$ is known, then we can obtain $Z$
- If $Z$ is known, we can compute $\alpha$, removing affine ambiguity

$$\alpha Z = \frac{-||b \times t||}{(\hat{l} \cdot b)||v \times t||}$$
Representation

**Image plane**

- Reference object
- Height: ?
- Vanishing line

**Image plane**

- Reference object, height: $Z_r$
- Vertical vanishing point, $v$
- Height: $Z_x$: ?
- Vanishing line, $l$
Measurements on Parallel Planes

- Projection matrix $P$ from the world to the image is defined with respect to a coordinate frame on the reference plane.
- The translation from the reference plane to another plane along the reference direction can be parametrized into a new projection matrix $P'$. 
Plane to Image Homographies

\[ P = \begin{bmatrix} \mathbf{l}_1 \mathbf{l}_2 \alpha v \mathbf{l}^\top \end{bmatrix} \]

\[ P' = \begin{bmatrix} \mathbf{l}_1 \mathbf{l}_2 \alpha v \alpha Z v + \mathbf{l}^\top \end{bmatrix} \]

where \( Z \) is the distance between the planes
Plane to Image Homographies

- Homographies can be extracted:
  \[ H = \begin{bmatrix} p_1 & p_2 & \alpha Z v + 1^T \end{bmatrix} \]
  \[ H' = \begin{bmatrix} p_1 & p_2 & 1^T \end{bmatrix} \]

- Then \( H'' = H' H^{-1} \) maps points from the reference plane to the second plane, and so defines the homology
Generic Algorithm

1. Given an image of a planar surface estimate the image-to-world homography matrix $H$

2. Repeat
   a) Select two points $x_1$ and $x_2$ on the image plane
   b) Back-project each image point into the world plane using $H$ to obtain the two world points $X_1$ and $X_2$
   c) Compute the Euclidean distance $\text{dist}(X_1, X_2)$
      i. $\text{dist}(A, B) = ||A - B||$
Camera Position

- Camera position $C = (X_c, Y_c, Z_c, W_c)$
- $PC = 0$
- Implies:

$$PC = l_1^X X_c + l_2^Y Y_c + \alpha \mathbf{v} Z_c + \mathbf{h} W_c = 0$$

- Using Cramer’s Rule:

$$X_c = -\det \left[ l_2^Y \mathbf{v} \mathbf{h} \right], \quad Y_c = \det \left[ l_1^X \mathbf{v} \mathbf{h} \right],$$

$$\alpha Z_c = -\det \left[ l_1^X l_2^Y \mathbf{h} \right], \quad W_c = \det \left[ l_1^X l_2^Y \mathbf{v} \right]$$
Camera In Scene
Uncertainty Analysis

- Errors arise from the finite accuracy of the feature detection and extraction
  - ie- edge detection, point specifications
- **Uncertainty analysis** attempts to quantify this error
Uncertainty Analysis

- Uncertainty in
  - Projection matrix $P$
  - Top point $t$
  - Base point $b$
  - Location of vanishing line $l$
  - Affine scale factor $\alpha$

- As the number of reference distances increases, so the uncertainty decreases
Ellipses are user specified.

\[ \mathbf{t} \] and \( \mathbf{b} \) are then aligned to the vertical vanishing point.

Alignment constraint:
\[ \mathbf{v} \cdot (\mathbf{t} \times \mathbf{b}) = 0 \]
Applications

- Forensic Science
  - Height of suspect
- Virtual Modeling
  - 3D reconstruction of a scene
- Art History
  - Modeling paintings
Forensic Science
Virtual Modeling
Art History
Conclusions

- Affine structure of 3D space may be partially recovered from perspective images
- Measurements between and on parallel planes can be determined
- Practical applications can be derived
Non Rigid Structure from Motion